A MODEL FOR INFORMATION RISK MANAGEMENT IN ECONOMIC INTELLIGENCE SYSTEMS

THESIS FOR THE AWARD OF CO-SUPERVISED DOCTORATE DEGREE

Université Nancy 2, France & University of Ibadan, Nigeria

By

Olufade Falade Williams ***ONIFADE***

Members of Jury

Evaluators: Samuel AJILA: Professor, Carleton University Ottawa, Canada

Eric BOUTIN: Professeur, Université du Sud Toulon Var, France

Examiners: Odile THIERY: Professeur, Université Nancy 2, France (Supervisor)

Adenike, O. OSOFISAN: Professor, University of Ibadan, Nigeria, (Supervisor)

Fabrice PAPY: Université Paris 8, France

Gérald DUFFING: ICN-Ecole de Management (Nancy 2) France

Laboratoire Lorrain de Recherche en Informatique et ses Applications – UMR 7503

(Equipe SITE)

ACKNOWLEDGEMENT

To my supervisors both home and abroad Professor (Mrs.) Odile Thiery and Professor (Mrs.) Adenike Oyinlola Osofisan, I am very grateful. Their candid encouragement, patience and the pain in reading over and over this work is note worthy.

To the duo of Professor Samuel Ajila (Carleton University, Ottawa, Cananda) and Eric Boutin (Universite du sud Toulon var, France) for accepting to read through and serve as external examiners during the defence of this work, thank you for your laudable comments.

I equally want to extend sincere appreciation to Professor Amos Abayomi David who took it upon himself to give the opportunity to me. His doggedness and resilience, words of wisdom and encouragement, readiness to serve, listen and advice is very uncommon. Thank you very much sir. His households are very wonderful; I cannot forget the scintillating meal of "Mrs. Prof".

The contributions of colleagues and friends at SITE LORIA, France are well appreciated. First and foremost, to my friend and co-supervisor Gerald Duffing, who assisted tremendously in the success of this work. Sincere greetings go to Victor Omoyibo of France, Y'aman, Bros SOJ, Baba, Hannen, Chedia, Pierre, and Stefan who on many occasions assisted in French translation.

To the French government ambassador to Nigeria that facilitates the scholarship for this work. To the manager of foreign students both CNOUS and CROUS and wonderful people like Rojo, Jeanick and Benedict, thank you very much.

To my sister and friend, whom I closely got acquainted with in the last lap of the program, Rudzz Rakotomalala surnamed Abake. Her supports, endearing words makes me feel at home in France, how I wish I could turn back the hands of time? Regard to her daughter Miangaly (Miangy).

To my students and colleagues at the University of Ibadan, particularly my soul mate Olumide Longe, Junbor, Dokita Robert, Dr. Akinkunmi, Alake Daramola and others too numerous to mention, thank you.

To all my family members, holding the forth while am not at home, prayerfully supporting me in this work, I am greatly indebted to you all. I really love and appreciate you. Thanks for being there. Abraham-Marian Onifade dynasty shall enlarge in grace.

And to Him, Eternal, who has made ALL things beautiful in His time, Merci Seigneur

DEDICATION

This work is dedicated to all those that have taken worthwhile risks, those taking it presently, and those who will take it in the future,

It is also dedicated to Afolabi, Anjolaoluwa and Abimbola and the ONIFADE's

And to Olubunmi Adetutu Anike for taking the risk of marrying me

Contents

List of Figures and Tables

Lists of Figures

Lists of Tables

General Introduction

Context of the Study

The advances in computer systems and communication technologies have given rise to the concept known as information technology. It has become a new tool that facilitates and shapes application of information on strategies, business models, critical success factors, and risk factors amongst others. The increasing important of information technology is revealed as a key facilitator of business competitiveness, emphasizing the importance in the development of skills such as relationship management, business process analysis and design, project and programme management, business change management, systems architecture, networking, risk management and security amongst others.

Economic Intelligence (EI) engulfs the duo of Information Systems and Decision theory. The objective amongst other of EI is to provide information to facilitate right decision in the right place, at the right time, with the right cost in order to support and enhance the overall functionality of the organization which is synonymous to an Informing Systems. Consequently, EI must be able to link Strategic management with Technological usage to produce and enhance hitherto herculean task of decision making. This will however necessitates the need for adaptation, followed by the introduction of concepts necessitating integration of means which may be human, techniques, organization or information can be appropriately harnessed. The management of the above depicts some level of heterogeneity which must be fused-out to synchronize resultant decisions from such operation.

This notion will provide us with better insight into the complexities involved in the process of decision making. Pinpointing the possible risk accruable from the interactions of the actors found in EI processes alone will be grossly inadequate because the whole process rest on the available information, which we considered in this research to resides in various database, information delivery systems or data warehouse. Therefore, identification and classification of error that could result into risk in decision making will be grossly inadequate without taking into consideration the quality of information from the information delivery systems.

With generic actors already identified in the context of economic intelligence, it has been established that decision rationale can be measured/determined based on the awareness of the challenges i.e. risk and threat incurable by the decision. Consequently, the EI process stages assume operational individuals are knowledgeable about finding, evaluating, analyzing, integrating, managing, and conveying information to others efficiently and effectively with the aim of facilitating strategic decision making. Since strategic decision is relying on available information, it is thus imperative to note that the quality of data of acceptable level is important in ensuring reliable decision.

The problem of information risk management thus forms a broad base that engulfs the decision maker, the information specialist (watcher) and other economic intelligent actors coupled with the information. The interpretability, accessibility, timeliness, adequacy amongst other information quality attributes will in no small measure equivocally affect the nature of decision accruable for it. The ability of the decision maker to adequately identify and present what is at stake commences the success or failure of any strategic decision. This is followed by his or her ability to inform the information specialist on the decision problem, the stakes from which laudable hypothesis can be formed. While the above is human intensive, another debilitating factor is the quality of information based on search operation. The following pertinent question thus forms the basis of this research: what is the importance of economic intelligence and its actors in the process of delivering strategic decisions? Is the quality of information (accessibility e.t.c) of any threat to strategic decision? These among others formed the basis for this research as shown in the problem definition below.

Research Problem Definition & Hypotheses

The main problem identified in this research stems from the need to reduce the risk of decision problem definition, leading to information need definition, information sources identification, information retrieval process, result validation, and interpretation of indicators amongst others. The factors enumerated above constitute what can be described as information risks. Consequently, the research question is on

How best can we manage these information risks factors within the context of economic intelligent systems?

To assist in realizing the state goal, the above can be further broken down into simpler and manageable chunks as follows

- How can we determine what is at risk and what constitute risk in economic intelligent systems
- How can we identify and localize these risks among the essential components of economic intelligence (EI process, EI actors, and Information)
- How do we manage accruable risks in "reconciling knowledge" among participating actors and processes
- How do we manage "technical risks" inherent in information acquisition, processing, presentation e.t.c.

The research hypothesis thus results from the above and can be stated as follows: ***With appropriate risk factor definition and modeling, information risk can greatly be reduced if not eliminated towards the process of taking strategic decision in economic intelligent systems***. Consequently, our means of achieving this was succinctly enumerated below.

The existing EI architecture at SITE –LORIA consists of multi-dimensional components which involve the main components of EI systems i.e. EI actors, EI processes and information. The processes via which these components are brought together constitute the mode via which this research evolved. At a glance, we can identify the following distinct but interrelated phases which were managed in this research: actors' interactions, information sourcing and collection, information processing, information utilization for decision making. The onus from the above is such that it will be inadequate to formulate a monster model that takes care of all identified risk factors in these stages.

Consequent upon the above, we proposed four different models to operate on these divergent components but in a harmonized manner. The first is tagged KNOWREM, this assists in the reconciliation of knowledge amongst EI actors. We followed this by FuzzOntology, it helps in resolving the imbroglio that can ensue during knowledge reconciliation and decision problem definition. We made bold to say that the volume of available information whether of quality or not can grossly affects the overall decision. To ameliorate the effect of information overload, we proposed a dual layer filtration called decisionability. The last of these models is FuzzyMatch. It assists the information specialist to combat the problem of missing or non-missing but wrong data.

Largely, we employed the soft computing paradigm (Fuzzy Logic) tremendously in the implementation of most of our models. This is sequel to the fact that most of the activities under study cannot be precisely acquired. Again we note that human language cannot be adequately captured because of the inherent ambiguity in its classification. This among other factor necessitates our choice of fuzzy logic.

Thesis arrangement

The presentation of this write up commences with a general introduction to the need for undertaking this research. This is followed with the enumeration of the problem statements and the hypothesis. The chapter arrangements are as follows:

The first chapter discusses the concept, various models and architecture hitherto proposed in economic intelligence and its processes. We take a look at the interactions between the actors and the processes; we define the concept of actor as employed in the scope of the work. The notion of decision making and decision problems was considered taking into cognizance the various modes and their distinguished attributes, we also compared the notion of EI as informing systems. We made a brief incursion into information and its effect on decision making. Views about information were reviewed and this is articulated with approaches to decision making. Lastly, we recap the major submissions of this chapter and pinpoint our submissions and support amongst the reviewed and presented information.

Chapter two featured principally contributions in the realm of information retrieval, data and information quality and the taxonomies of data quality dimensions. Our first attention is drawn to information retrieval algorithms, models and tools which are employed by user in meeting their information needs. We reviewed existing information retrieval tools and taxonomies to establish the fundamental components in retrieval tools: information representation and reasoning strategies. We featured uncertainty, vagueness, and incompleteness as factors debilitating against information retrieval. The effect of these tools alongside other factors in contribution to data/information quality is also presented. We also reviewed various submissions on what constitute quality. The effect of poor data quality on decision making was included to present a robust and all-encompassing view on the subject. Lastly in the chapter, we look at data/information quality issues in data warehouse concept.

In the third chapter, the main focus is on decision theories and possible risk involved. This enabled us to discuss various decision models and their proponents. These facilitate robust articulations which were later employed in the development of models employed for the research. We also present the concept of risk and risk factor, with understanding based on the provisions between hard and soft computing paradigm. This point onward depicts our introduction of the fuzzy logic paradigm. The notion of risk was investigated for the development of the possible risk factors employed in the next chapter.

We commenced the process of modelling towards implementation on chapter four. The rationale is to facilitate the accomplishment of our earlier stated problem definitions and research hypothesis. The chapter features the concepts of risk and risk factors (RFs) in economic intelligence which were later employed to formulate different models. We also expound on concepts like Translation Credibility (TC) and Decisionability. These two assist in the act of knowledge reconciliation (KNOWREM) between actors. Lastly, we present formal definitions for the KNOWREM model and hybridize ontology and fuzzy in what we called fuzzontology. FuzzOntology was presented here in this chapter as a simple framework but the implementation was actually performed in chapter five.

Chapter four creates the take off for the full implementation of chapter five. We thus implement two major models in this chapter namely FuzzOntology and FuzzyMatch. We have earlier on defined fuzzontology in the previous chapter, but herein we present the operational mode and compare with two other analytical tools. the implementation of the fuzzontological model assisted in articulating the manner via which it can assist a decision maker and watcher in the cause of delivering their duties. We featured information retrieval and discuss the design and implementation of our search tool called FuzzyMatch. We made comparison between its functionalities and existing tools. Examples and case studies were employed to determine the efficacy of the system.

The last of the chapters in this research is the conclusion, here we feature an all round discussion of every activities performed during the course of this research. These were juxtaposed with our earlier stated research problem definition and hypothesis to see how far we fared. In all, we were able to confirm our hypothesis that given an appropriate definition and model for the management of information risk, decision makers will enjoy the full benefit of adequate information acquisition and usage with minimum risk.

Deciding not to decide, is not the same thing as indecision, however all decision bears alongside a level of risk, for if it is risk free, no one will carry it out

Chapter 1
Economic Intelligence and Decision Problems

1.1 Fundamentals of Economic Intelligence

In this section we take in depths look at the notion referred to as economic intelligence (EI). We begin with the general notion and definition to facilitate a robust understanding of the concepts and draw out the major concept to systematically analyze the essence of their interactions, the effect on the decision ability of a decision maker alongside available information (from several information delivery sources). We shall attempt to define the various concepts in the best manner as seen in different literatures we have reviewed and in line with the definitions relevant to the research team SITE at LORIA, where this research is taking place. The best starting point for discussing the fundamentals of economic intelligence is through its various definitions, thus we shall provide various definitions by different authors and harness their baseline for the purpose of this research.

Economic intelligence (EI) according to Martre (1994) is defined as a set of coordinated actions of search, processing and distribution for exploitation of useful information for economic actors. The author opined that these actions were carried out legally having required protection aimed at safeguarding the company's patrimony, alongside operational quality ensuring minimum delay and cost. Revelli (1998) in similar manner defined EI as the process of collection, processing and distribution of information aimed at reducing uncertainties during the process of strategic decision making. Again, EI has been defined based on the focus on the particular level of an enterprise as the decision maker's capacity at exploiting the knowledge and the new experiences, while reinvesting already acquired towards solving a newly identified decision problem (Kislin et al. 2002). Another consideration put forward by Carayon (2003) considers EI as an encompassing process through which decision makers have a good and clear understanding of the ground on which it operates, thereby making him/her more familiar with the terrain and the operations.

Aggregating the above mentioned opinions and definition of EI, one can easily identify concepts like “information”, “actors/user”, and “decision maker” frequently recurring in them. The implication is thus that the three concepts formed the basis of Economic Intelligence. Consequently, discussion shall be central on them individually and collectively as the case may be. We can also note that within each of these entities, intra-operation is a possibility which is worth considering too.

In the above section, we have introduced the major definitions to be considered for this research. While noting that, there is a possibility of adapting the definition above to the focus of the research, no significant differences are expected. Thus, the importance of EI based on the provided definition and its identified processes will be the focus of the next section.

In the rest of this chapter we shall discuss the various models and architecture hitherto proposed in economic intelligence and its processes. We take a look at the interactions between the actors and the processes; we define the concept of actor as employed in the scope of the work. In second part of this chapter, the focus was on decision making and decision problems. Various mode were considered and their distinguished attributes, we also compared the notion of EI as an informing systems. Third section has its focus on information and its effect on decision making. Views about information were reviewed and this is articulated with approaches to decision making. Lastly, we recap the major submissions of this chapter and pinpoint our submissions and support amongst the reviewed and presented information.

1.1.1 Economic Intelligence & Its Processes

In the section above, we attempt to present various views based on prevailing definitions of economic intelligence. Thereafter we opined that an aggregation of the definition will pinpoint the major concepts inherent in EI. In this section, efforts will be geared at expanding the notion of concepts based on defined processes.

A systematic approach has been developed for EI process by SITE team at LORIA. Therein, progressive stages referred to as (EI Process) for decision making was established. In Bouaka & David (2004) which is widely supported by other authors for example (David & Thiery, 2002; Kislin, et al., 2002,) presented these stages as follows:

1. Identifying needs in the form of problems to be solved (threat, danger, consequences) 2. Translation of decision problem into information search problem. 3. Identifying the types of information needed to obtain the result, 4. Identification of appropriate sources; 5. Validation of sources, 6. Information gathering, 7. Validation information, 8. Processing of information collected for the calculation of indicators, 9. Interpretation of indicators, 10. Decision-making for solving the problem.

Table 1.1: Stages in Economic Intelligent Process

It is worth noting that the content of table 1.1 needs not necessarily follow a sequential order at all times, there can be need for backtracking from various level or stages during the process of fine-tuning to arrive at a considered optimum solution to the problem. Consequently, the various stages as presented in table 1.1 above is a clear indication of the concerted effort geared towards the realization of timely, cost effective and strategic decision making. The process stages commences with the identification of what need be solved – the decision problem (DP) in term of threat, danger and consequences, followed by the translation of the DP into best forms to facilitate information retrieval (IRP) (Bouaka &David, 2004). This stage helps to determine the nature and types of required information cum the identification of sources where it could be obtained. Once the sources are identified, then it remains the validation of such sources to reduce the number of non-quality data, or data with errors, or unreliable data. The source validation is directly followed by the information gathering and its validation. Thereafter, the next task will be processing of the information aimed at calculating necessary indicators, which is followed by its interpretation. The interpreted data will be presented to the decision maker who is expected to amongst other things attempt to draw reasonable inference from the data and some other factors as we will be presenting in the subsequent sections.

The definitions presented in section 1.1 described EI as a set of *coordinated actions* of *search*, *processing* and *distribution for exploitation* of *useful information* for *economic actors* (Martre, 1994). A proper juxtaposition of the definition and the content of table 1.1 depict a set of mappings. Thus, the whole process can be described as a set of "*coordinated actions of search and processing"* – these activities are not limited to search from the information delivery systems alone but also encompass the brain tasking activities for the acquisition of related experiences and other cognitive factors (Wang et al. 2004). Similarly, the "*distribution for exploitation of useful information"* although literarily simple, but involved a herculean task of rigorously defining and redefining of the earlier stages to guarantee a proper understanding for the rationale for the decision making.

Thus, it is important to note that, while table 1.1 present the process in a simplified manner, the sequence is not usually guaranteed to be sequential. There could be several instances where some stages will be visited more than once for the optimum functionality of the process.

The three principal concepts drawn out of the various definition of EI are presented in figure 1.1. The trios of user, process and information is inherent in all forms of human interactions necessitating decisions. It therefore implies that the understanding of their individual operation cum interaction will go a long way to enhance the overall objective as depicted in the definitions earlier presented. While table 1.1 itemized the EI processes, other forms of processes exit between and amongst the participating entities which must be taking into cognizance. Thus, we listed in the processes box, not only the EI processes, but also possible processes, that can occur, between the other two concepts: users and information.

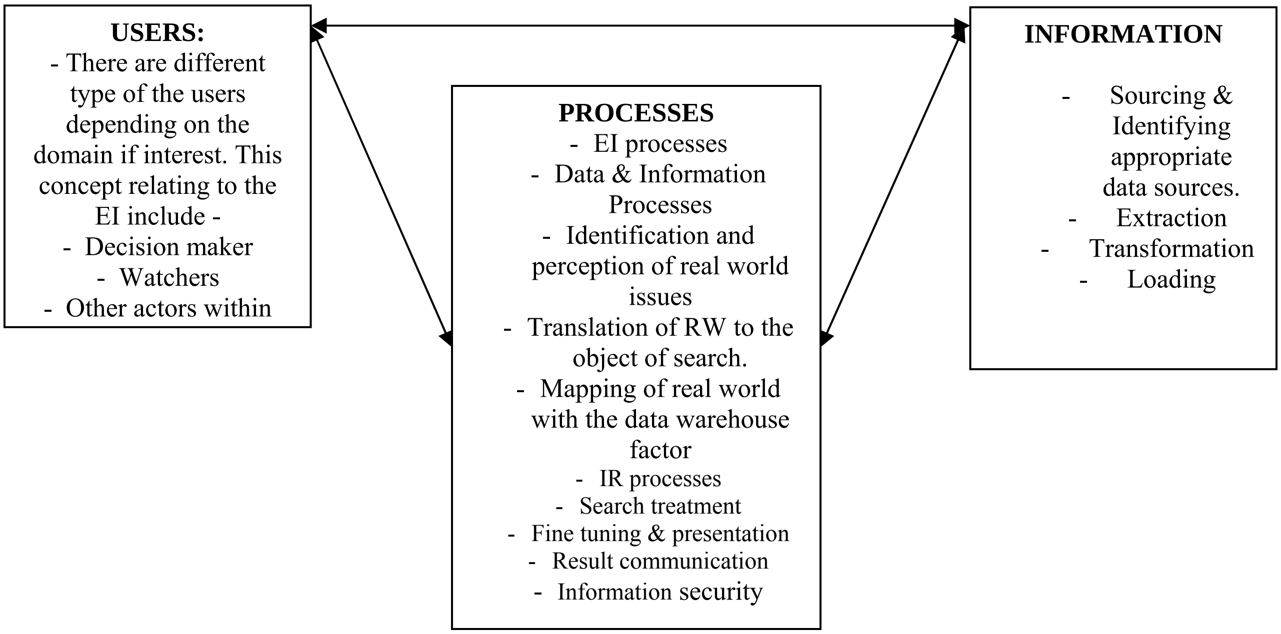


Figure 1.1: The Three Principal Concepts in EI

The above presented represents the proposition of the SITE team at LORIA in dealing with a decision problems. This section has thus taken a look at some definitions of EI and how this is adapted into a process hierarchy at SITE LORIA. In the next section, we shall present various attempts to further corroborate the importance of the above mentioned process based on model propositions and formulations.

1.1.2 Existing Models, Interactions & their Operations

In a bid to arrive at a shared conceptualization of different terms employed by different people in any community there is need to have predefined terms with general agreement on its connotations (Kuhne, 2005). A model is usually credited to have three distinct features:

- Mapping features – a model is based on the original
- Reduction features – it only reflects a (relevant) selection of the original's properties, and
- Pragmatic features – it needs be usable in place of the original with respect to some purpose

The above thus presents a model as a projection based on the abstraction of a whole entity. *A model provides information on something (content, meaning), created by someone (sender), for somebody (receiver), and for some purpose (context of use)*. While there exist several definitions to what a model is, none of them ever shy away from its importance as a mini representation of a global picture.

Economic intelligence team at LORIA has also developed some models aimed at assisting with the *coordinated sets of actions* of *search*, *processing* and *distribution for exploitation* of *useful information* for *economic actors.* Amongst the models already developed in the research team are: MEPD – which assists in the explanation of decision making problem in the best possible manner to make it comprehensive (Bouaka, 2004). In relation to the expression of the search problem for adequate information in connection with intelligence, model MIRABEL (Goria, 2006) was proposed. The above two models deal with definition for appropriateness, the next model which was termed RUBICUBE assists in multidimensional analysis of data warehouse information using a university data warehouse as a case study (Peguiron, 2006). For the contextualization of information via annotation in a business intelligence systems, AMIE was developed (Robert, 2007). This was followed by an attempt to inculcate the users' need into the definition of economic intelligent systems with a model named MORPRIE (Afolabi, 2007). While the above mentioned are saddled with definition and expression of information, CADRIE focused on the user of the information by specifying the roles of the moderator-coordinator of a regional economic intelligence (Knauf, 2007). WISP, thus based its design on one of the user/actor of EI responsible for information retrieval. The model is particularly focused on the specification of the problems of information search Kislin, 2007). Below is a tabular representation of the major contribution of the above mentioned models.

Model	Purpose	EI phases
MEPD/ DPM : Bouaka Najoua (2004)	This model combines both the context of the problem, the representation of the decision-maker and the challenges of the decision-making problem. Its objective is to facilitate the identification and the representation of the decision problem on the one hand and preparation of the information research project on the other**.** It aids the understanding of the problem through identification of the user's characteristics and assessment of the level of identified stakes of the problem.	Identification and definition of a decision problem
MIRABEL: Goria Stephane. (2006)	Model for the expression of the information problem to initiate searching and analysis of its context in relation to the inference from the understanding of a user and watcher. It helps to show the interpretation of variations that exist between two persons about the expression of a concept.	- Identification and definition of a decision problem. b) Translation of the decision problem to an information search problem
WISP: Kislin Phillip. (2007)	Translation of the decision problem to informational search problem and the specification of the problems of finding information by the watcher.	- Translation of the decision problem to an information search problem in relation with - Identification and definition of a decision problem
MORPRI^{2}E: Afolabi Babajide. (2007)	To represent users in relation to their specific needs, depending on the context and to model an information system which adapts to the changing needs of Users.	- Identification of relevant information sources - Collection of relevant information
AMIE : Roberts Charles(2007).	For the contextualization of information by annotations in a business intelligence system as well as validation of information relevance and reliability.	- Analysis of the information collected to extract indicators for decision
CADRIE: Audrey Knauf. (2007)	For the specification of the roles and the competences of the animator-coordinator of a regional economic intelligence setup.	All Phases
RUBICUBE: : **Frédérique Peguiron (2006)**	It adapts information system design to users' types of needs in the upstream design of a multidimensional data warehouse applied to a teaching environment.	- Identification of relevant information sources - Collection of relevant information
Contextualization for Multimedia representation: Hanène Maghrebi	Adding context parameters to multimedia information representation	- Collection of relevant information

Table 1.2: Models of EI at SITE LORIA

Table 2.1 above gave a concise snap shot of the various models already proposed by the team SITE at LORIA. Earlier on, we deduced from several available definitions that a model is a mini representation of a global picture enabling an all-round development of a concept or idea having being broken down into various fragment representing the entirety. Similarly, table 1.2, in an attempt to present a robust understanding into the whole process engulfed in economic intelligence, proposed and developed these models, whose operations are distinct but complementary, attempting to cover the EI process earlier mentioned in table 1.1. These models amongst other thing have enabled the understanding that, although the processes were hierarchically listed, their operations are interwoven and integrating these models in real life operation will result into recursive operations. Thus, each of the models, instead of being confined to a particular stage of the process, could also trigger other stages not closely linked with the model. This robust interoperability has further corroborates the importance and efficacy of the EI process.

In this section, we have outlined the various models already proposed by the team SITE at LORIA, depicting their operational interdependency, thus showcasing the rationale for their development. It is important to note that, while it cannot be said that model propositions has ended, or that the above models are adequate for the proper economic intelligence operations, a clear distinction of the human factor effect is important. Thus, in the next section, we shall be looking at the definitions and roles of EI actors.

1.1.3 Definition of EI Actors & their Roles in Economic Intelligence Process

The role is seen as a character or part played by a performer. It is also viewed as the characteristic and expected social behavior of an individual, his functions or position. With these in mind, we take a look at the definition of an actor and their roles within the context of economic intelligence. The player of these roles is described as an actor/user. In a similar manner, EI has identified and designate different task or roles to different set of people to allow for a much more encompassing operation amongst the participating entities. This section illuminate on the notions above in relation to EI process.

In table 3 below, we make an attempt to further depict the interactions obtainable from the various actors, process and information with a bid to clarify the activities of each of the participating actors at the various levels in EI processes. Thus starting from the identification of decision problems (DP), followed by the transformation of the problem into an information retrieval (search -IRP) problem unto the final stage of making available a set of information (premises) upon which the decision maker can conveniently base his judgments of the situation, the above thus buttress the earlier proposition of the trios – Object, Signals, Hypothesis <o, s, h> as the fundamental concept before the commencement of any decision activity David & Thiery, (2003). The table consists of three major sections: Users – that represents the various classes of user in the operation, each probably involved in one or more well defined stage of the operation; the EI stages – corresponds to the standard operational sequence as enumerated within the economic intelligence group. The Information systems – this corresponds to the repository of information used by the organization. For the purpose of this research, our focus is on the data warehouse (DW) system, which houses a standardized, consistent, clean and integrated form of data sourced from various operational systems in use in the organization, structured in a way to specifically address the reporting and analytic requirements.

	USERS	EI STAGES	INFORMATION SYSTEMS (Data warehouse/ Issues)
1	Decision makers, Watchers & Coordinators	Identification of problems to be solved in terms of dangers, risk & threats	
2		Identify the indicators to watch (research information)	
3	Watchers & Coordinators	Identification of appropriate (validated) source of information	Upstream issues
4		Collection and Validation of relevant information	
5		Necessary transformation on the collected information	Downstream issues
6		Indicators calculation	
7	Decision makers	Result interpretation	
8		Decision based on the inference from supplied information	

Table 1.3: Differentiation of actors' activities within EI processes.

The table above presents a brief differentiation of various activities amongst other taking place within an EI process. The importance of the interdependencies and the proficiencies of individual actors to reflect the global goal is thus imperative. From the foregoing, we make a little definition based on the data warehouse issues that can mar or make the effectiveness of the activities of the decision maker, i.e. Upstream and the Downstream.

- Upstream – this is the formation of data warehouse from heterogeneous sources of data, the following questions amongst others can thus be asked: how adequate and reliable are the sources? is the method of collection, formatting and presentation standardized? All these questions tend to achieve what we refer to as data reconciliation.

- Downstream – as depicted from figure 1.1, it is taking place at the lower level of the table, which concerns the major actors in the decision processes. The question could thus be asked: what is the quality of the calculated indicators? How adequately sufficient are they to avoid the risk of:
    - Wrong Utilization and/or
    - Wrong Interpretation

Several generic actors have been developed in the context of economic intelligence, amongst these are the decision maker, watcher and the coordinator. Below we take a look at their roles as defined within the research context.

Decision maker: this is the individual in the organization that is capable of identifying and posing a problem to be solved in terms of stake, risk or threat that weighs on the organization. In other words, he is expected to be familiar with the *needs of the organization*, *the stakes*, the *eventual risks and the threats* the organization can be subjected to (Bouaka, 2004).

Information watcher: this user/actor otherwise referred to as "watcher" is the person within the organization that specializes in the methods of collection and analysis of information. His objective is to obtain indicators (using information) or value added information that the decision makers depend on for his decision process. After receiving the problem to be solved as expressed by the decision maker based on request or deliberations, the information watcher must translate it into information attributes to be collected and which are used to calculate the indicators (Kislin, 2007).

Coordinator: this actor was introduced based on the rationale that questioned the adequacy of the existing actor responsible for the operation of search in the possibility of being overloaded with time, thereby reducing efficiency and possible introduction of errors into his work. Although the nomenclature was formally known as an infomediary, coined from "information and mediator", it was later changed to reflect a broader role in EI process defining his status; mission, place and influence on other participants in the process to a coordinator (Goria et al. (2005), Knauf & David, 2004). In distinguishing between the work of a watcher and that of a coordinator, the following were noted.

The watcher's operation was characterized as follows:

- characterizing the stakes involved in a decision problem
- translation of these stakes into indicators and associated informational problems
- drawing-up of available sources of information
- the proper retrieval process is carried out, and validated using various indicators
- present the information in the best format required by the decision maker (raw or processed).

Coordinator is however responsible in 2 broad ways

- It can exist as a neutral entity, a third party, provider of unbiased information. Not necessarily hired by an organization/company but help consumers at large to determine the suitability of their choices.
- It can also provide vendors with consumer information that enables the vendors develop/manufacture goods tailored towards consumers' specification. This type usually protects consumers' interests.

The classification and the need to distinguish roles become imperative going by the fact that within the concept of Economic Intelligence, division of labour has resulted into specialized classes of users dealing with information acquisition and its presentation, decision making and the acting agents. Consequently, delivering timely information for strategic decision making is a collective responsibility of the various actors in EI process whose activities must be neatly integrated.

This section takes a look at the definitions of various actors already identified in the EI systems. It highlights their responsibilities and interactions based on the earlier defined EI process. Examples of generic actors were given and their role within the overall objective of economic intelligence. In the next section, we shall take a look at the architecture of EI systems alongside the roles of actors hereby defined.

1.1.4 Architecture of an EI System

The development of several models still need be rightly and tightly fit into a framework in which they can operate as cooperating entities for the smooth operation of the whole systems. David & Thiery, (2003) developed a model for the EI system which will be presented with the earlier notions of the three principal concepts in EI.

- **Selection** permits the constitution of the information system (IS) of the organisation that can be
  - (i) the production database (that allows current usage of the organisation),
  - (ii) all the information support for an information retrieval system (in documentation for an example) or
  - (iii) a strategic information system (SIS) based on a data warehouse. This information system is constituted from heterogeneous data and from heterogeneous sources with the aid of a filter.
- **Mapping:** mapping permits all users an access to the data in the IS. Two methods of access are opened to the user: access by exploration and access by request. The exploration is based on a system of hypertexts. The requests are expressed with the aid of Boolean operators. The result of the mapping is a set of information.
- **Analysis:** in order to add value to the information found, techniques of analysis are applied on the results. For instance, the assistant of a head of department that we consider as the information watcher can present a summary to his head of department.
- **Interpretation:** this means in general, the possibility of the user of the system being able to make the right decisions. It does not mean that the sole user of the system is the decision maker; it can include the information watcher. One can see then the interest in capturing the profile of the decision maker in a metadata stored on the data warehouse which can be used to build a specific data mart for a group of decision makers or even better a particular user.

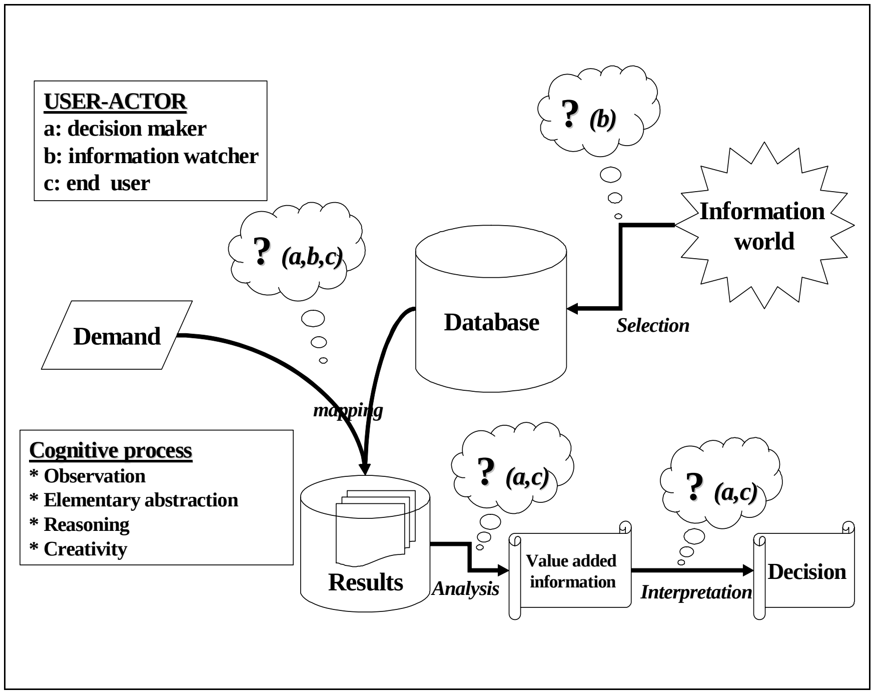


Figure 1.2: Architecture of an Economic Intelligence System.

David & Thiery, (2003) highlighted the above architecture in an attempt to set a framework which will accommodate the models earlier discussed in this chapter. Literally or informally, architecture is considered to consist of “framework”, some well defined concepts and rules governing the operations to be set forth. This architecture is a foresight on the importance to integrate the models that will follow it. Thus, we can see from the definition describing the architecture how it closely follows the EI process earlier discussed with attempts to identify the actors and their roles as pinpointed in previous section. The above described is a common place in software development, where most, if not all the concepts earlier discussed are present. Similarly, EI defined its processes, develop its models (may be there are more to be developed), defined its actors and relationships (in form of rules, but not rigid) and thus it will be required to integrate these in a properly developed architecture or framework similar to that shown in figure 2. Next in figure 1.3 we present the architecture with the various models imposed to clearly depicts its importance as an anchor for the entirety of the IE process and the models earlier discussed

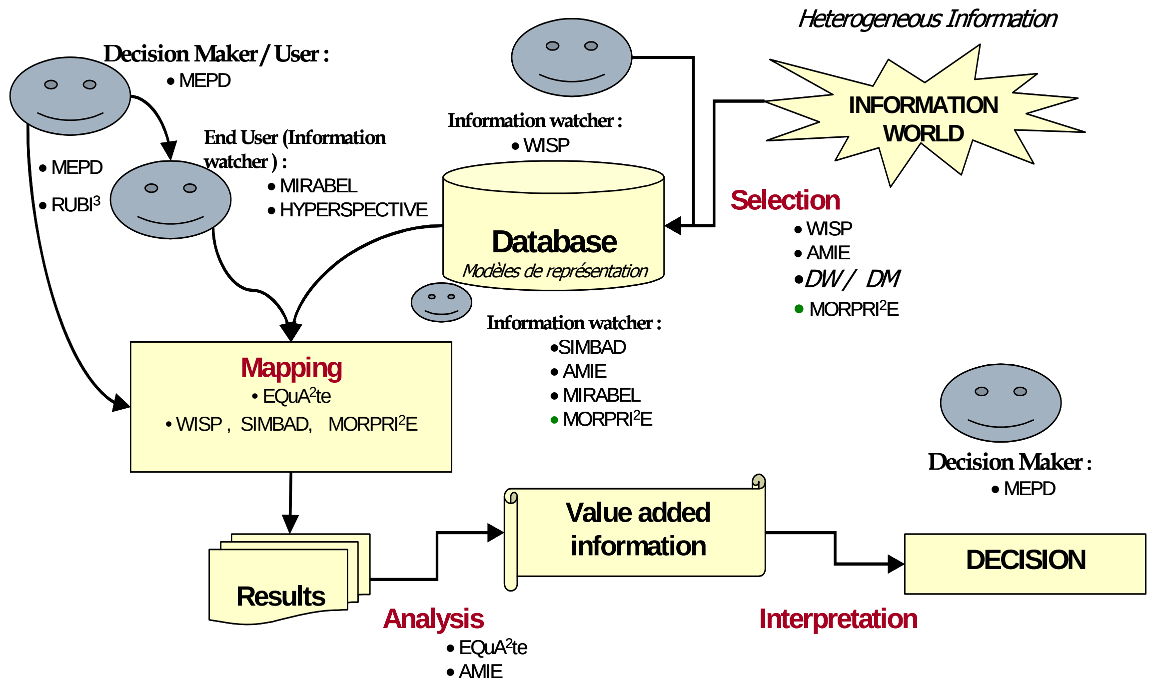


Figure 1.3: Architecture of an EI system according to research team SITE (SITE 2006)

The triad of users/actors, process and information formed the bedrock of economic intelligence. It is aimed at evolving a process of collection, processing and distribution of information aimed at reducing uncertainties during the process of strategic decision making taking into consideration the fact that decisions are based on available information amongst other things. We have therefore presented major concepts in EI, facilitates their understanding through rigorous definition, elucidating their interactions and supporting models that were developed for the purpose. We have also presented the various economic actors, including generic types already recognized and their roles in EI process. Lastly, we showed the architecture in which all the discussed concepts tightly fit.

1.2 Decision Making & Decision Problem

Overview

With the sporadic growth in human knowledge stemming from experiences and reasoning abilities to create order in the mass of available information, it becomes imperative that decision resulting from such must be well guarded in the right direction.

Decisionability of individual is determined by many factors. The list includes sense of judgment, experience in related situation and cognitive abilities amongst others. While decisions making are based on available information, the mode and method of getting such information is as important as the decision. Human are limited in expressive ability of their perception of the world which is the result of lack of lexical impression, incompleteness and inaccuracy in measurement Sivanandam, et al., (2007). While decision making has become an everyday phenomenon with every human, its formulation and implementation have a great impact on the outcome of the decision. In this section, we take a look at the concept of decision making and decision problems. Other sub-sections within in the section discuss the rationale for decision problems, taking a typical example of the process in EI. We also highlight EI as an informing system and rap up with approaches to decision making.

1.2.1 Rationale for Decision Problem

There could be different reasons for embarking on a particular decision, thus the rational for a decision problem could be possibly inferred from the various definitions we provide below from various literatures. We categorize the act into two broad groups:

- Based on choice of alternative – decision making can be considered as a study of identifying and choosing alternatives based on the values and preferences of the person concerned (Harris, 1998). The implication form the above is that, there are several alternatives choices in the process of decision for consideration. It therefore involves not only *identifying several alternatives*, but also making a *choice in selecting* a best-fit for the already *stated desire, goals, values, lifestyle* e.t.c.
- Reduction of uncertainty from choice – above, we underlined the factors we considered important in the process. While *alternative, choice and values/goals* are central to the definition above, these factors result into the next factor to be considered in our definition. Decision making in this regard is the process of appropriately reducing *uncertainty and doubt* from available alternatives to facilitate the selection of best alternative from available choices. It is important from the above that very few decisions are made with absolute certainty, sequel to the fact that

complete understanding of all alternatives are not usually possible. Thus, emphasis is more on the reduction of uncertainty rather than total elimination.

In another dimension, we can consider decision problem rationale from the conceived kinds of decisions. Although all animate being are involved in decision making, the decision process and ability of humans are the most developed, involving complex operations aimed at arriving at a logical conclusion. In figure 1.4 below, we represent some kinds of decisions based on the utilized expressiveness of language.

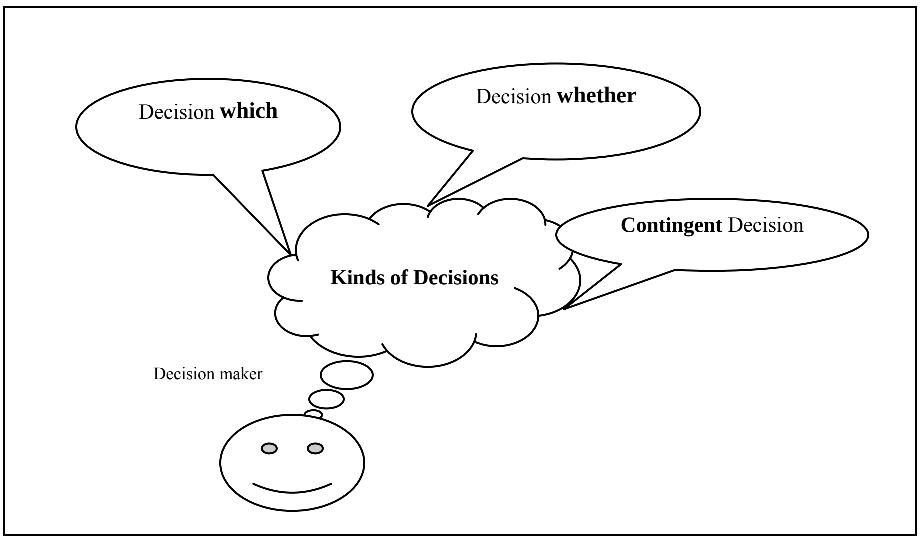


Figure 1.4: Kinds of decisions

Considering the premise stated above while attempting to define decision making, we stated there are several alternatives which necessitate choice. "*Decision which"* involves a choice of one or more from the available alternatives relative to a set of possibilities. On most occasions, the alternatives selection is based on the level of satisfiability presented by each of the alternatives based on sets of other predefined criteria. "*Decision whether"* involves a bivalue answer yes/no. This is an either/or decision that must be resolved before considering the sets of alternatives e.g. whether to vote, or abstain. This type of decision usually involves weighing the pros and cons before any attempt is made. The last type shown in figure 4 is what we referred to as *"Contingent decision"*, this is a provisional decision, based on some condition. They are believed to have been made but put on hold until some other conditions are met (Harris, 1998), for example, "*I shall* *return for my university education if I can raise enough money for my tuition"*. The above described is vividly based on the fact that humans generally carry around a set of already made, contingent decision and are just waiting for appropriate environment, trigger or opportunity to decide.

In this section, we have elucidated the definitions and various concepts that anchor decision making. Goals, Alternatives, and Uncertainty were identified and these were used to define the notion referred to as decision making. Lastly, we highlighted three kinds of decisions to include: decision whether, which, and contingent decision. In the next section, we shall be looking at the various components of decision making.

1.2.2 Decision Making & Cognitive Ability

Human instincts, subconscious belief, conscious belief, values and intuition have been identified as the major emphasis that determines the mode of decision making. Thus the differences displayed by the various modes are the reflections of the level to which the above mentioned are displayed (Barrette 2006). In describing the above statement, we present below figure 1.5, to depict the various forms of modes, the factors affecting the stated modes amongst other things.

- Instinct-based decision making – the origin of this form of decision can be traced to the cellular structure of human beings (DNA). It is principally associated with issues related to survival, or what can be described as the survival instincts. A typical example is the act of suckling and crying in newly born babies. Distinguishing feature if these types of decision making are: actions usually precede any deep thought, decisions are made based on recognizable past experiences (maintaining internal stability and equilibrium). It is said that human have little on control over these types of decisions, but instead, they control us.
- Subconscious belief-based decision making – decision making in this context is similar to the above described, but unlike the above, decision is based on personal memories rather than (DNA) memories. Thus, actions precede thoughts. This is usually expressed with emotional charges that can be positive or negative expressed

in form of blame, internal competition, rivalry, and caution in the negative sense, and openness, trusts e.t.c in the positive dimension (Barrette 2006).

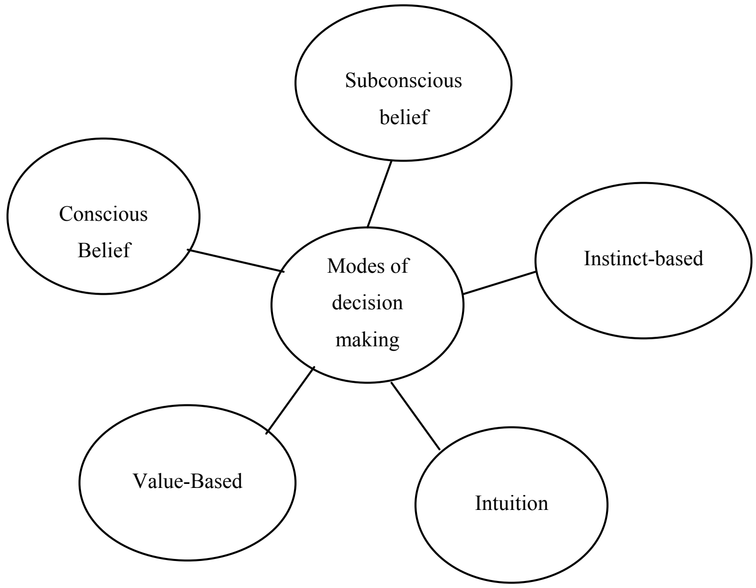


Figure 1.5: Components of decision making that defines its modes.

Popular examples of sub-conscious belief-based decision making are what are described as tears of joy, feeling of pride and spontaneous happiness. We can therefore summarize the main features as follows: actions always precede thoughts, decision making are based on past experience, we are not rightly in control of our actions and behaviours i.e. emotional stability and rationality, it is very personal. It can also be found that operations are from the first three levels of consciousness (Harris, 1998).

- Conscious belief-based decision making – the subconscious is not endowed with ability to make rational decisions, rather there must be a shift to conscious based decision making. This involved inserting a period of pause between meaning making and decision making. This period allows for reflection, reconsideration and thoughtful logical process aimed at understanding the rationale for the job at hand. It is therefore obvious that, decision making actually follows thought in this mode. In this mode, ample time is created for what to decide on, discuss with others before final inference. There is however, a similarity between the conscious and the subconscious-based decision making – they both employ information based on past experiences (what is assumed to be known) to attempt addressing the future.

- Value-based decision making – although mention is made of attempt to create future in conscious based decision making. The future therein discussed is still saddled with some level of emotional attachment which may impair the whole process. A major question under this type is "is *this decision rational and is it in alignment with our values?"* The answer to the above would either call for a rethink or go ahead with the decision making operation. This is sequel to the fact that any decision that is not in alignment with the values of an organization is said to lack integrity, if it contradicts personal values, it lacks authenticity. Thus, the lack of integrity and authenticity is an indication that no personal or group cohesion is obtainable. Value based decision allows making decisions that is targeted towards already stated missions. It has a stack difference from conscious based belief decision making since there it de-emphasize meaning making. In value based decision making, attempt is made to allow values guide our behaviour, not our belief, this is sequel to the fact that values are judged to be universal concepts transcending all contexts. Beliefs however are local and contextual Collins & Porras, (1994).
- Intuition-based decision making – this is best described via its characteristics i.e.: data gathering and information processing takes place in the normal pattern, judgment is suspended – no meaning making takes place, subconsciously or consciously, thought and belief agenda are suspended, the mind is allowed to make a deep dive into mind space of collective unconscious, thought arises after a period of reflection based on deep sense of knowing and the thoughts reflects wisdom – focusing on common good, deeply held values and consideration for long term. Implicitly, they reflect what is about to emerge.

1.2.3 Decision Making in EI Context

Decision makers can employ intuition to solve a problem or rationalize the problem. Usually, the more challenging the situation, the higher (likelihood) is the associated monetary values of the consequence, which is directly synonymous with the rationale of expectation from the decision maker. Decision making in EI context followed a well laid out procedure, ranging from the identification of the need for decision making, redefining the decision need between the watcher and the decision maker, the watcher thereafter searching for appropriate information based on the understanding gathered from the redefinition upon which the final inference is based with other factors alongside to be considered. Thus, decision making in EI realm is a robust process taking into cognizance all possible factors laid out properly for an effective decision making.

EI from its definition is set to present a coordinated action of search and information utilization for timely, effective and strategic decision making (Martre, 1994 & Revelli, 1998). Bouaka & David, (2004) presented a model that combines both the context of the problem, the representation of the decision-maker and the challenges of the decision-making problem. The objective is to facilitate the identification and the representation of the decision problem on the one hand and preparation of the information research project on the other**.** It aids the understanding of the problem through identification of the user's characteristics and assessment of the level of identified stakes of the problem. Sequel to the above, Bouaka & David (2004) presented a proposal aimed at assisting the decision maker in explicitly defining his decision problem. Their focus was on the environmental, organizational and personal data. These were used to juxtapose the relationship amongst the EI actors to determine who poses the question and why the question was posed. They assumed amongst many things that there is an atmosphere of trust and confidence between the parties involved in the information need definition and acquisition.

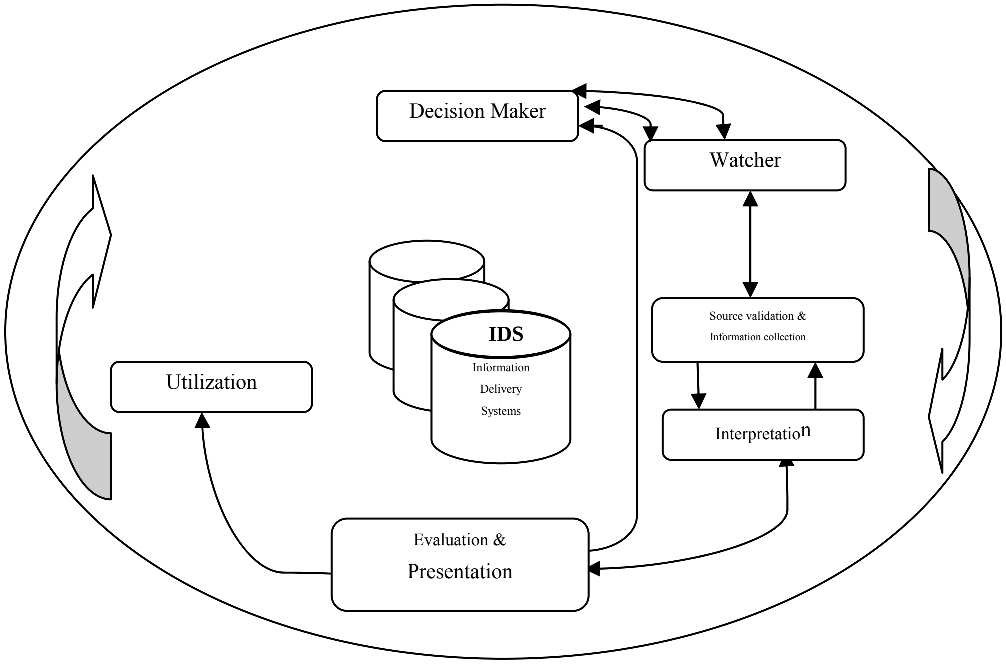


Figure 1.6: Circular representation of the principal concepts in EI

Figure 1.6 above presents the activities of generic actors, information and various processes ensuing towards decision making in EI. David & Thiery (2001) present a model capable of adapting to different actors involved in the EI process. EI, from their point of view involves the act to understand the process involved in the production of *interpretable indicators* for decision making based on internal and external information available to it. Their opinion is formed from the fact that decision rationale can be measured/determined based on the awareness of the challenges: risk and threat incurable by the decision. The focus is thus on the team leader and the watcher – an information specialist. Their requirement and responsibility were outlined and this is employed for the development of the information base which involves dual filtration. Following this, David & Bueno, (2001) developed a personalized information retrieval system that is code named METIORE. Characteristically, the first thing in *Metiore* is to capture the user objectives done for a session, these are thereafter formulated in natural language. From the above described, it is evident that the operation of the system is based on the systems approach and user's interactions. The next stage allows the user to make simple or complex queries- search function, using some attributes. The first thing to note is that, the user is expected to have a little information about the desired information to effectively use this program.

Another trend in decision making pattern of EI is based on the Duffing et al. (2005). With the focus stemming from economic monitoring, which is enhanced by information system and data warehouse? It was submitted that data quality of appropriate level is important for accurate decision making. Recalling that the origin of the information is as important as the processes taking place on the information that are residing in the data warehouse. The process of selection, cleaning, storage and retrieval are very important to the decision made out of them. The decision maker must base his actions on available indicators (Thiery & David, 2001) relevant to his problems. The interaction, compositions and roles of EI actors were examined to determine the possibility of risk, its type and source for any decision taken. They arrived at two types of risks: Trades risk - reconciliation of knowledge and processes related to a specific area, and Technical risk – which involves information processing on the data warehouse. The pattern of risk modeling will thus be from structural and behavioural axis. They collapsed the EI processes into four for the risk identification which include: Identification; Extraction; Transformation and Utilization.

1.2.4 Economic Intelligence and Informing Systems

1.2.4.1 Informing Systems and Her Processes

According to *Nadler's (1970)* terminology on work systems, *informing systems are defined as classes of work systems whose basic output is information that affects the actions of its recipients. A model satisfying the above thus consist of three basic components: informing source(s), communication channel(s), and information receiving entities, which are sometimes referred to as informing clients within the realm of informing science*. The term informing science applies to disparate fields sharing common goal of providing their clients with information in a form, format, and schedule that enables the maximization of its effectiveness. From the above, three interrelated concepts can again be deduced: the clients (saddled with a task that requires information for its completion), the delivery system (aimed at providing information), and the informing environment that creates information to assist the clients in completing their works (Cohen, 1997 & 1999).

With the above notion of informing, Gackowski, (2006) provides a provisional schema of informing. Therein, the work of Nadler (1970) and Cohen, was thoroughly displayed in a diagrammatic format to arrive at the provisional schema of informing. Two distinguished types of informing were identified in figure 1.1 i.e. direct informing and indirect informing. The direct informing proceeds with the observation of realities, gathering of necessary information and the decision making process follows the act of comprehension, qualification and quantification. These acts will trigger the memorization and representation process before the search for adequate and related information of the observed issues (Yingxu, et. al 2004). Direct informing is devoid of the use of information delivery systems in acquitting itself with necessary information for the goal of decision making. Consequently, his information search is dependent on the interpretation given to what he observed in reality before attempting to present the results to the entities informed i.e. clients, users, e.t.c.

Indirect informing is more complex than the direct counterpart. It involves distinguished sources of information with different characteristics, information delivery systems and the entities informed as shown in figure 1.7. Each of these concepts and factors introduce different types of what we had earlier called risk factors that can culminate into risk in the process of informing towards decision making. The indirect informing has resulted into division of labour that involves differences in point of operations of both the informing entities and the entities informed. With the distinctions above, cognitive ability of individuals concerned are brought into play and can contribute in no small measure to the success or otherwise of the result of the informing operations.

Interpretability of information implies that it matches any states with some attributed or associated meaning in the mind of the receiving individual, or with any state that automatically triggers a design sequence of state transition in the receiving numerically controlled device. It is contextual, based on the knowledge of the interpreter, conditional, level of training and experience amongst other cognitive ability of human (Gackowski, 2005).

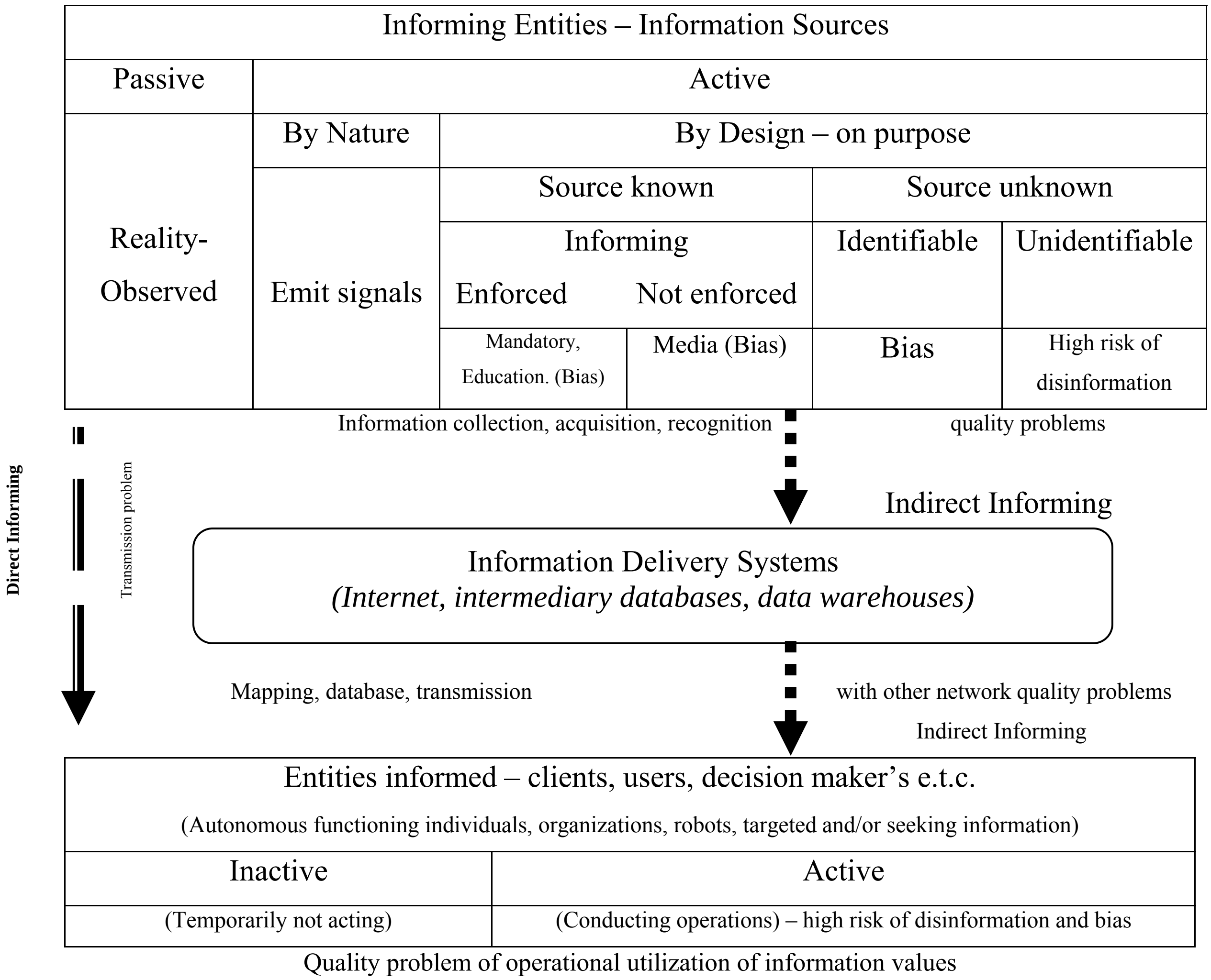


Figure 1.7: The Provisional Schema of informing *(Gackowski, 2006).*

Thus, we submit that, apart from the requirement of interpretability of gathered information which is the first direct primary requirement as a prerequisite for further examination of any aspect of its information quality, it is equally expedient that the credibility of understanding between the entities informed and the informing entities be ensured to discourage misconception. Our submission is sequel to the fact that different actors are involved at different stages of the EI process aimed at facilitating strategic decision making.

1.2.4.2 Economic Intelligence as Informing

Economic Intelligence (EI) is defined as a set of coordinated actions of search, processing and distribution for exploitation, of useful information for economic actors. These actions are carried out legally with all the necessary protection for the safeguard of the company's patrimony, and with the best quality, delay and cost (Martre, 1994). Similarly, it is the process of collection, processing and distribution of information with the goal of reducing uncertainty in taking strategic decisions (Revelli, 1998).The concept of EI must be able to link strategic management with technological usage to produce and enhance hitherto herculean task of decision making. Usually, there will always be the need for adaptation, followed by the introduction of concepts necessitating integration of means which may be human, techniques, organization or information (Thiery, & David, 2002). The management of the above depicts some level of heterogeneity which must be fused together to synchronize resultant decisions from such operation. It is thus imperative to note that data quality of acceptable level amongst others is important for appropriate decision making.

As earlier discussed, the major concepts in both informing systems and economic intelligent revealed a lot of similarities even though the terminologies employed differ. In the realm of economic intelligence, three major concepts were defined to include: actors, information and processes, while informing science identified the clients, the delivery systems and the informing environment creating information to assist the clients in the discharge of his operations. Without mixing word, we can draw a one – to – one relationship amongst the concept to facilitate further discussions.

Centrally in EI context, the focus is on the *Information*, *Actors*, and the research *Processes* of information. These concepts were graphically depicted in table 1.1. It consist of the defined actors in EI on the left hand side, the EI process stages in the middle and the 'faces' of the information delivery systems used in securing information for their operations. Economic intelligence has defined some generic actors which include: the decision maker – this actor is responsible for determining the organizational needs and formulating them into a decision problem (DP). The others are the Watcher and the Coordinator. The watcher is a specialist in information management and retrieval.

He is expected to thoroughly understand the problem-need of the decision maker (decision problem) and make concerted effort to rigorously define the translation into the object of search i.e. information retrieval problems (IRP). The coordinator as the name implies oversees the functionality of the operations within the systems. The table amongst other things shows the major concepts and their interactions to depict the level and point of involvement of each actor in the EI process stages. The processes have been well organized into various stages of operations referred to as the EI process stages. Starting with the identification of decision problem (DP) that requires or demand for solution, it proceeds through to the attempt to formulate the decision problem into information retrieval problem (IRP) together with other actors as defined in the EI context (Bueno, & David, 2001). The success of the process above will in no small measure determine the level of acceptability in terms of quality of the information result which is to be employed for the decision making process. It must be stressed that the origin of the information is as important as the processes taking place on the information. A peculiar meeting point of all information from different information delivery systems is the data warehouse, after which the information are selected, cleaned, stored and utilized by the application. The above processes enhances decision making, it therefore behooves that, the need for comprehensible, reliable and easily accessible information resulting from a properly resolved IRP cannot be overemphasized.

The concept of Actors as listed in the rightmost part of table 1.1 has a broad definition in EI process: the set of actors include the decision maker, the watchers, coordinators and others yet without nomenclature. However, it is interesting to note that, the interdependency of these actors result into division of labour which attracts both good and bad outcomes. In the formal case, the good side is to enhance the overall operation of decision making, and in the later case, it can constitute some level of risk into the overall decision made. We submit that the interactions among the actors resulting from the decision of labour as depicted in indirect informing process and their individual cognitive abilities can result into the risk of non-quality data. Risk factors which include: personal preference, environmental factors, misinformation and other roots of biases coupled with the data warehouse quality constraints were considered to have negative impact on the overall quality of the decision.

EI is involved in the act of understanding the processes involved in the production of interpretable indicators for decision making, based on the information available to it (Thiery & David, 2002). The leader is expected to identify the problem in terms of risk, challenge and threat, through the observation of certain parameters within the environment, and also inform of assumptions which requires verification. Decision maker in EI, wanting to resolve a DP, derived from the combination of Objects, Signal and Hypothesis <O, S, H>, resulting from both the internal and the external views is acquainted with some aspect of the decision problem - ***Objects***, which were perceived/generated by some events – ***Signal***: (gesture, circumstances, perceptions, deductions which communicates something; cue, sign which initiates a certain action, conveying of information with resultant meaning) from where certain deductions can be made – ***Hypothesis***. However, the knowledge is limited based on the available data currently at his disposal. Therefore, *Data Values* are symbolic representation of aspects of reality that are known, given, or assumed true. *Thus, we can see the decision maker in EI as an entity to be informed*. The required informing process however commences with the ability to adequately transform the decision problem into information retrieval problem.

The Watcher specializes in the collection of information using different methods for collection and analysis of information, geared towards value-added information. This is an example of informing entities, i.e. based on the presented DP, appropriate information retrieval problem has been arrived at towards the commencement of search operation. Thus, he is expected to translate the problems in term of attributes and indicators on information used, identify and verify relevant sources, collect relevant indicator and perform some pretreatment on the information***. As a specialist in information systems, his first task will be to adequately comprehend the need for the information retrieval problem based on the decision problem.*** The Coordinator, another actor in EI, on many occasions also assists in the transformation of the decision problem (DP), supervises the search and other relevant processes towards the presentation of most relevant information for the decision maker. It most often serves as the bridge between the decision maker, and the sources of information, to identify/notify on appropriate location of relevant information to aid search process (Knauf & David, 2004).

As earlier presented in table 1.3, the interesting thing to note will be the level of participation and interaction of each of these concepts. As expected, the operation begins with the decision maker identifying a decision problem in terms of the need for problem to be solved (danger, threat and risk). Thereafter, he alongside other actors engage in the process of identifying the necessary indicator to watch (research information) before the commencement of the information retrieval process.

The importance of the interdependencies and the proficiencies of individual actors to reflect the global goal are thus imperative. From the foregoing, we make definitions based on the data warehouse issues that can mare or make the effectiveness of the activities of the decision maker, i.e. Upstream and the Downstream.

Upstream – this is the formation of data warehouse from heterogeneous sources of data, the following questions amongst others can thus be asked: how adequate and reliable are the sources? Are the methods of collection, formatting and presentation standardized? All these questions tend to achieve what we refer to as data reconciliation. Downstream – as depicted in figure 1.3, it is taking place at the lower level of the table, which concerns the major actors in the decision processes? The question could thus be asked: what is the quality of the calculated indicators? How adequately sufficient are they to avoid the risk of wrong utilization and/or wrong interpretation.

Consequently, we present in figure 1.8 below our proposed provisional schema for informing in economic intelligence process. Using the earlier presented decision making process stages, we partitioned the whole decision making process to manageable chunks with a bid to adequately identify the inherent risks and the manner towards resolving them. To this end, each of the partitions are accorded the name of the model proposed towards the resolution of risk in decision making process.

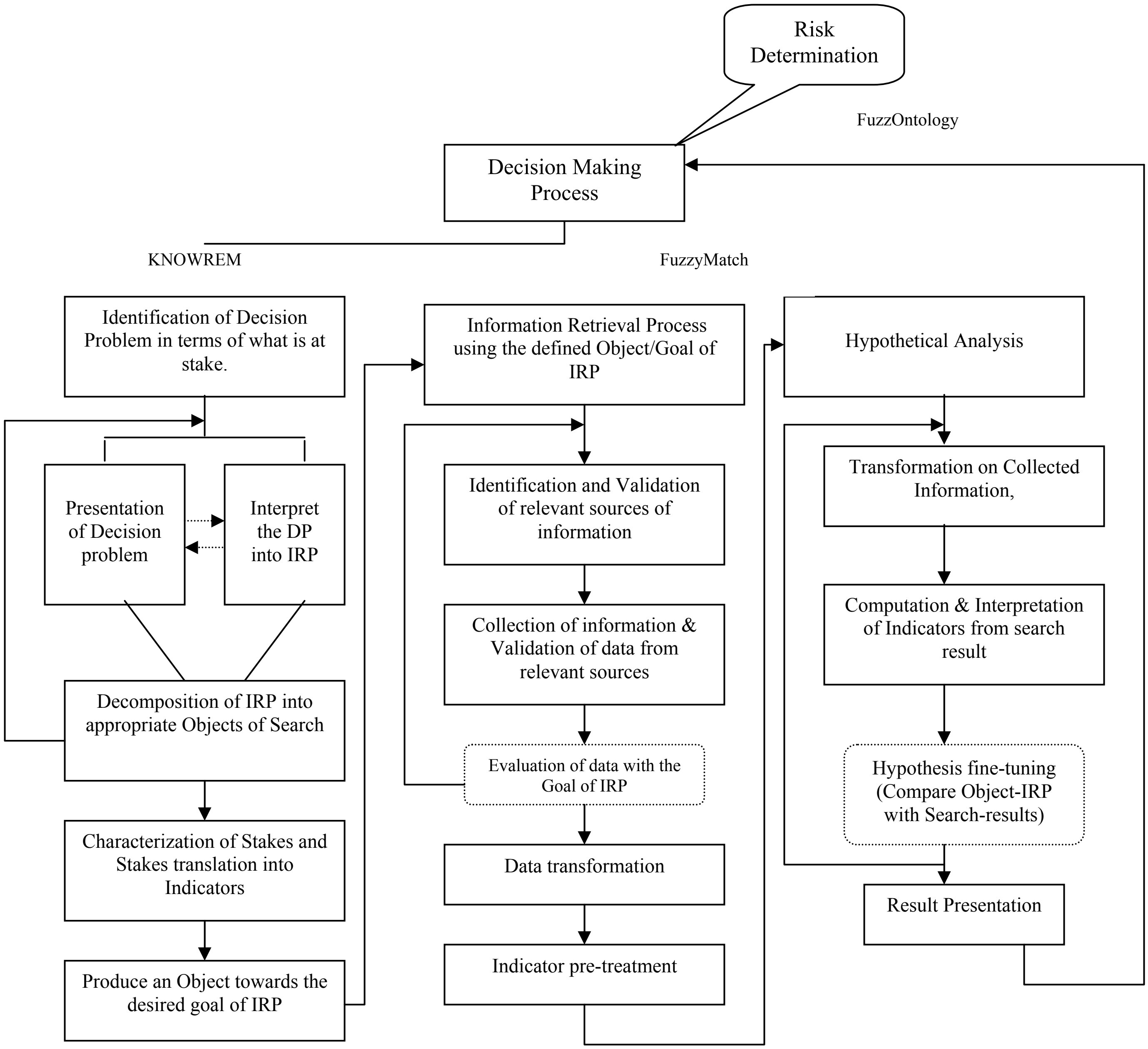


Figure 1.8: Schema for Indirect Informing in Economic Intelligence

Our first attempt towards identification and management of decision problem identification and resolution between the decision maker and the watcher is tagged KNOWREM. KNOWREM employs ontological principles to capture the decision problem resolution between the EI actors involved.

A measurement criterion tagged Translation Credibility will be used to determine the level of reconciliation between the actors and the readiness to proceed into information retrieval stage. The second stage thus follows and our proposition for this is called FuzzyMatch – a search engine that employs fuzziness level to resolve ambiguous and incongruous queries.

The last of our model is tagged FuzzOntology – succinctly, we attempt to interpret whatever ontology capture by fuzzy inference system. The rationale is born out of the problem of interpretation that usually ensure in human usage of languages of communication.

It is thus obvious from the above that the *focus of this research is the management of associated risk that can result for the definition, interpretation, retrieval and utilization of such information towards decision making.* We shall subsequently take a simple look at what denotes information, possible types and its relevance in decision making in the next section.

1.3 Importance of Information in Decision Making

Overview

There have been several attempts to facilitate quality of information employed in decision processes, but much has not been done in evaluating the importance of information in decision performance (we shall synonymously refer to information and data). With conflicting notions of information and data not helping the evaluation, it behooves to assert that "What we understand is based on what we already know, and what we already know comes from being able to understand".

Baron & Hershey (1988) opined that information plays an important role in decision making. Their submission was based on evaluation to gauge the relevancy and irrelevancy of information on decision. In other to make high quality decisions, it is deemed important to have access to information that is complete and relevant to decision task as much as possible (Stevenson, 1985). The importance of information in the delivery of appropriate decision cannot be over emphasized. However, another factor that is of importance is the volume of available information. In any decision situation, there can be variation in the amount of information available for the job. Keren & de Bruin (2003) submitted that the first key component of decision is “obtain relevant information”.

On one extreme, the information may be grossly inadequate, while the other is the possibility of too much information sequel to the fact that the tendency for the decision maker to seek for more information in discharging his duty is very high. It thus becomes imperative to note that the amount of information for decision is very crucial. If there is more than enough information, the following could result:

- Increase in time needed to process information thereby resulting to delay in decision making
- Information or Cognitive overload could ensue. This is a state where there is so much information which consequently declines the performance of the decision maker. Possible decline results into ‘forgetfulness’ e.g. a crash course on international affairs for a presidential candidate.
- There results a selective use of information. The danger in this lies in the fact that the choices are based on preconceived ideas or solutions.
- Result into mental fatigue
- The last in this list is Decision fatigue. This is the state at which decision making ability decline into arriving at fast, careless decision or ultimately into decision paralysis.

Considering from the perspective of a typical enterprise, Redman (1998) itemized the possible impacts of poor data/information quality to include customer dissatisfaction, increased operational cost, less effective decision making and reduced ability to arrive and execute strategy. The impact of poor data quality was viewed from three different dimensions: Operational impact, Typical impact and Strategic impact. The operational impact perspective has to do with customers’ dissatisfaction, increased cost and lowered employee job satisfaction. Similar effect is seen at the decision making level where tactical impact is felt. Poor data quality compromises decision making and in another dimension, it makes implementation of data warehouse aimed at facilitating better decision making cumbersome (Celko, 1995). Hammer & Stanton (1995) consider the effect from the effect it has on the reengineering process, while mistrust can also ensues amongst the workers. Thus making available appropriate and adequate information for decision making is a sine qua non to organizational performance.

The third level identified by Redman (1998) has not been extensively researched into, however the result of the above mentioned (operational and tactical impacts) clearly point to the possible effect that could result in decision making. It has thus been identified that lack of quality data makes setting and execution of strategy difficult. This results also into management distraction and possibility of compromising organizational alignment.

The importance of information to decision making cannot be complete without examining an aspect or attribute of information which is its adequacy in volume available for the process of decision making. In the view of Huang et al. (1999) and Deming (1982), it is practically impossible to achieve high quality decision but only via an improved IQ in a systematic manner. Keller & Staelin (1987) made a submission in their proposed model for measuring decision effectiveness and unequivocally state that information quality and available quantity is of great importance. Their model state that decision effectiveness can be expressed as

$$\text{Decision effectiveness} = -g_1(N) + g_2(TQ) + \varepsilon_{2i}$$

In the above equation, N represents the quantity of information, TQ denotes the total quality of the information, and $g_1(.)$, $g_2(.)$ are functions reflecting the effect of the duo of information quality and quantity on decision effectiveness. Using a two-strategy approach to decision which are: conjunctive decision making and weighted additive decision making, Chengalur-Smith et al. (1999) conducted several experiments to describe the importance of information to decision making. Fisher et al. (2003) explored the importance of experience and time in relation to the issue of data quality cum the ability to effectively utilize same for decision making.

Ge & Helfert (2006) developed a framework to assess the quality of decision based on information quality dimensions. Their experiment was conducted in three categories: in the single-dimension situation with IQ improvement on a single dimension, in the dimension-combination situation, and in the total-IQ situation without IQ assessment and improvement. The authors assert that without improvement and assessment, the quality of decisions might even be lower than gambling.

The general notions from most information system (IS) literatures were that data are a prerequisite for information and information can be created from raw data. Evaluating the success of an IS follows two dimensions of which information quality is one, whereby decision quality is a function of the former (Stephenson, 1985). In Todd & Benbasat (2000), a comprehensive presentation of review on the impact of information technology (IT) on decision making was given, however, it was adjudged to be grossly inconclusive (Eierman et al. 1995). Several other factors were proposed to be taking into consideration before the impact on decision making can be appropriately measured. Thus decision performance was considered from the effects of moderating and mediating variables such as decision-maker capability (Benbasat & Taylor, 1982), decision strategy (Silver, 1990). The last of these is the on decision performance which was based on the effect of data quality on decision performance using as the yardstick the conceptual framework of data quality in Wand & Strong (1996).

Another dimension to view the importance of data in decision making is the fact that decisions involves choices from possible group of preselected alternatives, which could be the results of an earlier decision. Consequently, the volume of available information for the decision making is as important as the accruable decision. This does not however implies that “the more the information, the better the decision” as another factor known as cognitive overload could results. Caution should therefore be exercised in the usage of data for decision purposes since decision left unmade becomes decision by default, or better still a decision made for you (Harris, 1998).

This section has highlighted the importance of quality data in the process of decision making. We have evaluated various contributions of several authors without necessarily distinguishing between the usage of the term “data and information”. On the overall, we state clearly with respect to decision making that – a good decision should be logical, representing the available information cum the carefully selected choice out of the available alternatives. In the next section, we shall attempt to clearly distinguish between the terms usage and the danger in synonymous reference of the two concepts.

1.3.1 Definitions and Types of Information

Strictly in this section, we shall look at the concept of information from different views to determine how it affects this research and help to provide adequate ground for the utilization and adoption of the term. We also look at information from different perspectives amongst which are: information as a Product and information as a Process.

Shannon & Weaver (1949), the earliest proponent of information concept presented their view from a content-neutral perspective. Information is regarded as that which reduces uncertainty, thus making presupposition on knowledge of a priori probabilities. With this notion, it is impossible to adequately accommodate various objects within the limit. Chmielecki, (1994) described the submission of Shannon to be anthropocentrically bias for its inability to accommodate both lower and higher animals. The fundamental feature of animate systems (which can be regarded as informational system) is their ability to discriminate and select from available stimuli. This he judged from basis that some systems must cope with varieties of environmental stimuli. The above thus implied that, the important thing to be considered for this category is the detected differences between distinguishable entities. “Differences” and “Detection” becomes a sine qua non for establishing the notion of information in this set. Gackowski, (2006) opined that information represent the aspect of reality that are yet unknown, to be collected, acquired, and recognized. In another view, information is considered as anything that one is capable of perceiving: these include written communication, spoken communication, photographs, arts, music, and nearly anything that is perceivable (Alderman, 2007).

The immanency of information has brought so much controversy into the subject. Chmielecki, (1994) described information is being "out there". While this is ambiguous in its meaning, the author explained that information should be defined in terms appropriate to the systems that are making use of it, which is in terms of the resources available to it. Information is described as an abstract entity. It is considered to possess no separate existence on its own, this is sequel to the fact that there is no considerable differences to determine the real states of affairs between which the differences can be established or constitute its code. The rationale is drawn from his view that same information can be represented/encoded in various manners e.g. a piece of music can be stored compact disc, encoded on magnetic tape e.t.c.

The surge in the usage of the term leading to what can be described as 'infomania' has brought about various concepts among which are: the information age, the information society, the information worker, information management and information economy (Knox, 2007). The onus is thus that information need not be seen as computer oriented or generated. The importance of the concept in academic literatures and business environment has necessitated the need to address the generic view of the concept of information. In the following subsections, we shall present different types of information as given by various authors and the generic grouping of information as a product or as a process.

1.3.2 Types of Information

Again the categorization of information failed to present a uniform front. While the numbers presented by the authors differs, there are fundamental similarities between some which makes the issue more of nomenclature than opinion differences. In what follows, we shall briefly present the types as seen in major literatures and various institutions.

We shall consider the submission of Clark (1999) as the first types of information to be presented. Clark identified five types of information: as a Fact, Concept, Procedure, Process and Principle. It is consider a fact if it uniquely present or identify an object, a person, place or date. As a concept, the author said it must present a category of items or ideas sharing common features. This type allows for discrimination between and amongst concepts to dissuade ambiguity. Information as a procedure takes into cognizance various steps that depict how to make or do something. Typical example can be procedure for booting or shutting down of a plant or system. Another type of information considered is information as a process. Here, we are concerned with the description of the mode of operation of something, e.g. the process of photosynthesis. As a principle, information presents rules, heuristics, guidelines, criteria which enable the prediction of an outcome e.g. writing a research report. All the above mentioned types of information are types we encounter in everyday living.

In the submission of the California State University, department of library Information, six types of information were identified. These include Fact, Opinions, Objective, Subjective, Primary and Secondary Information. The categorization as a fact is similar to that of Clark (1999), however, it was added that it has to be concrete, and should be something that can be proven. They also consider information as an opinion based on what seems to be true. This type of information is a based on personal views or judgments. Objective information has elements of fact in it, usually reported by experts or authorities in a particular field. It is expected to present all sides of an issue. Information like this are those found in the encyclopedia articles, books and periodic articles. As opposed to the above, subjective information is usually based on the opinion of individual or groups. It presents usually analysis of hitherto stated fact, it is preceded by phrase like “from my understanding, in my own opinion, I felt, e.t.c”. The other types are primary and secondary information. If it is created at the point of occurrence, or considered as the original document or research being conducted in a field of study, then it is regarded as primary information. It is expected that the information is in its original form and has not been published anywhere, neither translated nor interpreted. However, secondary information stems from the primary counterpart. It is usually the result of analysis performed on primary information leading to interpretation, translation or examination in a manner different from the original presentation.

Chmielecki, (1994) described three types of information – parainformation, structural information and metainformation. Parainformation was considered the first-order information, it is elemental, primordial type of information. This is sequel to fact that it presents the simplest form which serves as the building block of any other type of information. Structural information was considered the second-order information as it is composed of several pieces of information to create information. Metainformation is the highest level of all the above mentioned types. In the first instance, it comprises collection of collections, and usually comes after preliminary processing has been carried out on earlier types. Chmielecki, (1994) concluded that there is a stepwise emergence during the evolutionary process. This starts with the organism being able to handle parainformation solely (cells, multicellular organisms, plants, primitive animals, e.t.c), followed by developed creatures being able to deal with structural information and finally by higher creature that can deal with metainformation.

1.4 Views of Information

The various types of information by various schools of thoughts can again be distinguished into two main camps. The first camp viewed information as a resource while the other opined it should be seen as a process enacted by human. Davy (1998) described information as the centre of all businesses both private and public. The author asserts that the manner of information utilization by an organization is a determinant of how competitive, efficient and ultimately profitable the organization is. Subsequently, we shall consider the submission of the two camps (product versus process) based on different submission from literatures.

1.4.1 Information as a Product

Wang et al. (1998) was amongst the earlier proponent of the concept via which information as a product was greatly elaborated. Present a myriad of information quality investigation on problems encountered by many organization, the author submitted that information should be treated as a product or an end deliverable satisfying consumer's need. Information product (IP) is defined as a collection of data element instances meeting the specified requirements of a data consumer, these requirement are usually employed for business decision making, legal reporting or government reporting (Lee et al. 2006). In treating information as a product the following must be taking into cognizance: understand the consumer's information need, manage information as a product of well-defined production process, manage information as a product with life cycle, and appoint an information product manager to manage the information product. A similar notion to the above is to view information as a tangible resource or a product which is manufactured (Chaffey & Wood, 2005).

1.4.2 Information as a Process

A contrasting view to the notion of information a product is the consideration of information as a process. It emphasizes human role in the creation of information. Popular proponents are Mutch (1996), Davenport (1997) and the work of Davenport & Prusak (1998). Ajila & Wu, (2007) made a survey on the effect of open source adoption in which the operational process was highlighted with statistical evidences. Shankaranarayan, et al (2003) developed a framework for data management using information product approach. The authors paralleled the process of gauging and managing data quality with manufacturing and consequently tagged data as "information product (IP)". The model thereafter developed was named IP-MAP. Similar provisions but not expressly stated was the notion of Redman (1998). The IP-MAP concept was then compared with the dimension stated in Redman which are: Conceptual, Data value and Data representation, with the main focus on the data value.

The above mentioned approaches to information have been a subject of deliberation among researchers. While each proponent and their supporter are doing all it takes to establish their point and believes, we are in support of the inclusion of human factors in the creation and utilization of information thus seeing it as a process than a product. In the next section, we take a look at various approaches to decision making.

1.5 Approaches to decision making

There are several approaches to decision making however, two of these stands out remarkably. Thus decision making in organization either follows an authoritarian or a group method. In the formal, the executive figure utilizes his/her prerogative to enforce the decision, while the later involves ability for a group to unanimously reach a conclusion. Below we examine with examples the operations of the above mentioned.

Authoritarian Approach: in this type, the person at the helm of affairs makes decision based on acquired knowledge. This will subsequently diffused to other members of the team. In most cases, the operation usually follows that, while its possible to 'make decision' in 5 minutes, it will take like 30 minutes to explain the rational and another 30 minutes for the decision to gain acceptance amongst the participants. Figure 1.9 below gives the interpretation of this fact (Harris, 1998).

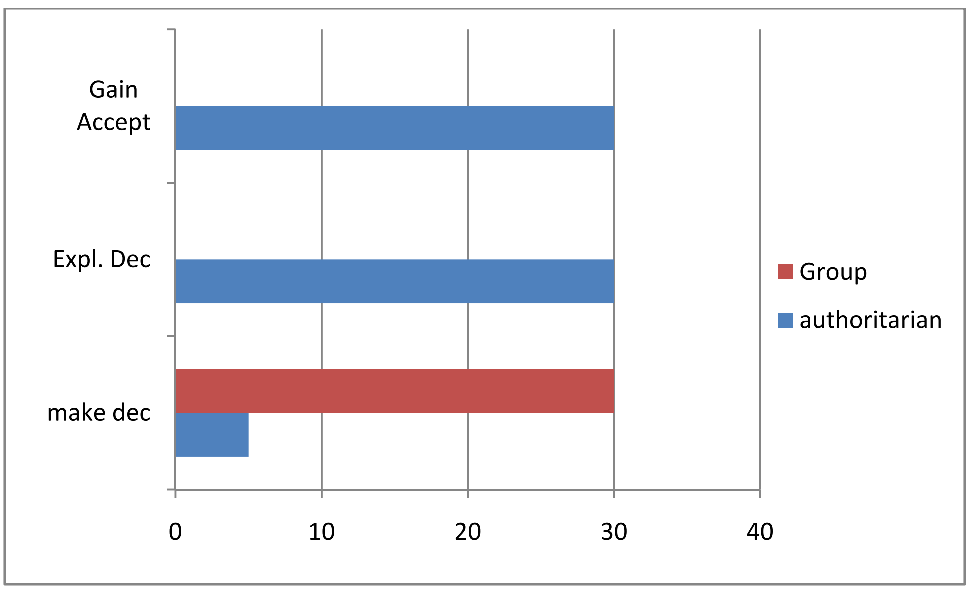


Figure 1.9: Authoritarian vs. Group approach to decision making

Group decision Approach: in contrast to the above the group usually shares and analyses ideas before coming to an agreeable logical conclusion. This is followed by the implementation of the decision. The group decision pattern has two main types. Free discussion is the first, in this pattern, the target for decision is made the object of discussion, and thus everybody participates, e.g. promotion of an officer in an organization. The second type is the developmental or structured discussion (Harris, 1998). The target problem here is broken down to various steps with specific intermediate goals before the overall decision making. The focus of this type is on specificity rather than creating impressions. A time-break down for this type will be as follows: a group will likely make decision in 30 minutes, and there will be negligible time for explanation and acceptance by the members.

It is interesting to note that from the point of efficiency, group decision performs better than the authoritarian, however, delivering critically timely decision within a group can be difficult as the time it takes to arrive at a logical conclusion is usually long. In whichever forms or approach employed towards decision, one factor that facilitates worthwhile decision is the available information. The next section thus elucidate on the importance of information to decision making.

1.6 Recap

In this chapter, we have deliberated on concepts like economic intelligence and decision problems, decision making and decision problems, importance of information to decision making and some popular views of information. We expounded EI and its processes alongside various models that have been developed for the operation.

We also presented the EI architecture with attempts to fix the existing models appropriately within the architecture. We identified and defined generic actors like the decision maker, coordinator, watcher and their roles in bringing about strategic decision making. Within the realm of decision making and decision problems, we established the similarities between the principle of informing and EI. We made bold to assert that, the professionalism employed in EI leading to the distinguished roles performed by different actors made it a typical example of informing systems. Thereafter, we looked at various approaches to decision making and the effect of cognitive ability of individuals involved in the process. In the last part of the chapter, we explore various notions presented as regards information and its effect on decision making. Two major views were of interest here from different school of belief. These are information as a process and information as product.

As shown in the schema for informing in economic intelligence process, most of our propositions have been introduced but not into detail. These tit-bits shall represent the major part of subsequent chapters as we shall embolden their importance and relationship to the overall goal of this research. Our convictions follows from our research focus and bias that we shall adopt the notion of information as a process and that decision making is not mechanical but need to follow a set of laid down procedure which are not however easily determined nor can it be easily or totally automated. We are strong of the opinion that human involvement in creation, manipulation and utilization can only bring effective decision making. Thus our allegiance disagrees with treating information as a product.

In all, this chapter is meant to lay the foundation for this research work in presenting the basic/necessary information about major concepts to be employed in the course of this research. Subsequently, we shall begin to see the interrelationship amongst them, their effects and importance towards determination and management of risk factors, for delivery of strategic decisions in economic intelligent process. We have therefore amongst other things depicts the relationships between the identified actors in economic intelligence systems. While it is almost impossible for them to work without information, we also show the definitions of information and how various models of decision making employ information in carrying out their daily tasks.

Having established the notion of information in this chapter, our focus in the next chapter will be the mode and manner via which the actor (information specialist i.e. the watcher) acquires information, with consideration for the tools and methods. We shall also be interested in what constitute quality of data and information.

"There is no institution more prototypically committed to the systematic application of information to decisions than the modern bureaucratic organization… Information is not simply a basis for action. It is a representation of competence and a reaffirmation of social rule"

(Feldman & March, 1981)

Chapter 2
Information Retrieval and Data/Information Quality

2.1 Overview of the main concepts

Information retrieval is inherently predicated on users searching for information from their "information need" that result from the interpretation of decision problem. Apart from the quality of data and information in the data warehouse, the volume, timeliness of the information to the decision maker is equally important. While the information need might be right, inherent errors resulting from dirty data are detrimental to the overall goal of information retrieval. Since most search operations are performed on the internet or corporate organizations expensively constructed and maintained data warehouses, Broder (2002) submitted that the main difference with the classic model for IR and the one augmented for the web lies in the replacement of "Matching Rules" in the former with "Search Engines" in the latter. The term has many definitions found in the literatures, but we support the fact that "data fit for use is information" and "decision relevant data is also information".

Data and information quality engulf highly complex and huge magnitude of data quality issues based on the comprehensive context and the organizational context. Data quality problems are usually not in isolation, they comprise accumulated, lengthy, and hidden process, and signals root constituting data consumers' experience of difficulties with using data (Lee, et al. 2006). It is worth noting that the existence of data quality problem is not limited to automated computer environments, thus attempt at improving data quality problem must consistently and carefully diagnose and improve not only the data but alongside the enabling environment in the specific context. Data environment refers to issue related to collection, storage and usage. Alongside the above are the database systems, information systems infrastructure, related task process mechanism, rules, methods, actions, policies and culture representing a typical organizational composition.

In chapter one, we presented the notion of information and various notions from different authors, again we shall look at the misconception amongst data, information and knowledge as portrayed in the literatures.

The notion of quality has been described as been “fungible” – the same information can be used by different consumers with widely variant purposes and grossly dissimilar domain of interest (Bovee, et al. 2002). This factor necessitates a high level of flexibility and consistency in the definition. Juxtaposing the above two, we have information quality which has suffered from multiplicity of definitions and views with vivid examples as found in FASB, (1993), Wang, et al. (1995), Wang & Strong (1996).

This chapter elucidates the various contributions in the realm of information retrieval, data and information quality and the taxonomies of data quality dimensions. Our first attention is drawn to information retrieval algorithms, models and tools which are employed by user in meeting their information needs. We reviewed existing information retrieval tools and taxonomies to establish the fundamental components in retrieval tools: information representation and reasoning strategies. We featured uncertainty, vagueness, and incompleteness as factors debilitating against information retrieval. The effect of these tools alongside other factors in contribution to data/information quality is also presented. We also reviewed various submissions on what constitute quality. We shall also consider the submissions from literatures on the conflicting notions of information, data and knowledge. The effect of poor data quality on decision making will also be included to present a robust and all-encompassing view on the subject. Lastly in this chapter, we look at data/information quality issues in data warehouse concept. This is because most of the data employed for decision problems resolution are kept in the warehouse, where search operations are carried out. We then present a recap to establish again, what we adopt, support and utilize for our own research.

2.2 Information Retrieval Algorithms, Methods, Technologies and Tools

The exponential growth in the available information witnessed in the last decade has resulted in the proliferation of information retrieval objects, which consist of algorithms, methods, technologies and tools. These objects are saddled with the responsibility of ensuring user access to prompt and adequate information, however, the story is not always as expected. Information retrieval (IR) is the scientific discipline concerned with the analysis, design and implementation of computerized systems aimed at addressing the representation, organization of, and access to vast amount of heterogeneous information already encoded in digital format (Rijsbergen, 1979).

The expectation from any Information Retrieval System (IRS) is to make available such information considered pertinent to a user's query (formally expressed in the system's query language). It therefore behooves that, the effectiveness of any IRS is based on the parameters which reflects the ability of such system in accomplishing the user's goal. Unfortunately, these goals are not deterministic sequel to the presence of uncertainty and vagueness embedded in many parts of information retrieval process (Crestani & Pasi, 1999). Canfora & Cerulo (2004) opined that a key feature of an IRS is the retrieving the document satisfying the information need of a user from amongst a huge collection of documents. These systems, in web context are referred to as search engines. In a bid to facilitate ease of search, user information request are represented by keywords or phrases that are indexed. These representations are known as queries, and the indexing can assume diverse term depending on the tools, however, ranked IR methods are popular i.e. documents are ranked based on measure of relevance as compared to user's request.

Users' expressions of their information need are often represented with a query in an uncertain and vague manner (Crestani & Pasi, 1999). This is sequel to the fact that the representation of a document informative content is often uncertain, and the story is true for the process by which query representation is matched with a document representation.

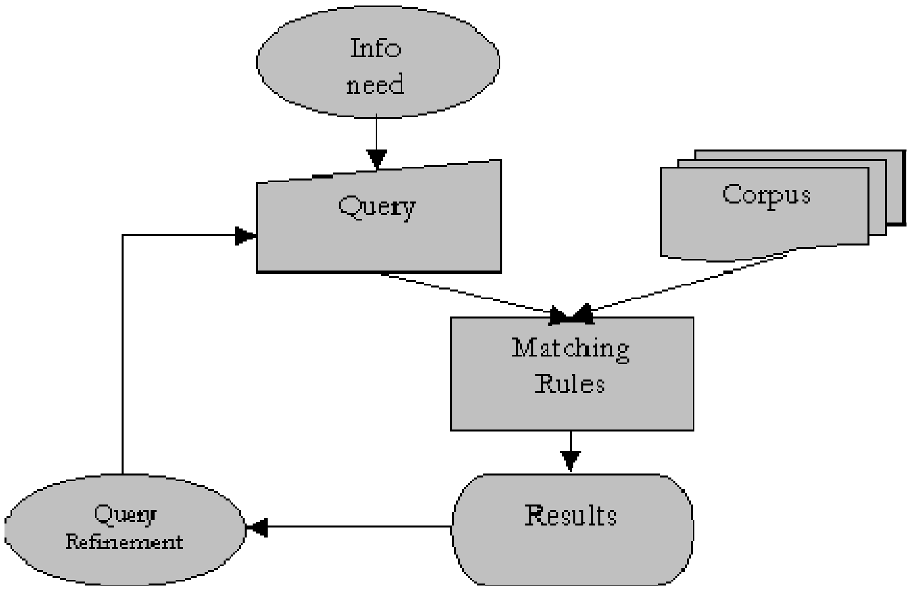


Figure 2.1: Classic Model for Information Retrieval (Broder, 2002)

The expectation is thus in the IRS to appropriately deal with the concept of uncertainty and vagueness which has been majorly ignored in commercial IRS. Another user expectation is the expected flexibility in IR process. Flexibility in this regards implies the capability of the system to manage imperfect (vague and/or uncertain) information, and also to adapt its behaviour to the user context. Centrally, the main goal in IR is the quest to the set of ***relevant*** documents, amidst large collection in a bid to satisfy the information need expressed in form a query by a user (Crestani & Lalmas, 2000). We note that these large documents can be in form of texts, images, video, audio, mediums, or sometime in a multiple format of any combination of the above mentioned.

Figure 2.1is a simplified model for the classic view of standardized information retrieval system. This model has been very popular until the incursion of the popular World Wide Web (WWW) causing a strong resurgence in information retrieval focus and thus necessitating a shift into several other models to be discussed in later part of this work (section 2.6.2). Search engines represent an outgrow of IR with utmost attention (Glover, et al., 1999). Many of these search engines however employ the traditional retrieval models with Boolean query pattern which was the first formal language for IRS. The Boolean method characteristically forces the user to precisely express their information need as a set of un-weighted keywords, consequently the user to expresses her IR need with vague requirement (Pasi, 2008). In subsequent sections, we shall discuss the various components of IRS, chronology of web search engines and the various model of information retrieval system. This discussion is aimed at setting the background for the concept of information retrieval problem resulting from missing data/information and the consequential effects (risks) on decision making.

2.2.1 Chronology of Web Search Engines

The issue of missing data continue to linger not only at the level of design but also at the point of retrieval towards usage. Whenever users either do not find the required data or link as the case may be or they have to search endlessly through the results which are usually bulky, the interpretation is usually – the data might be missing. This factor on most occasions has to do with the robustness of the search engine employed for the retrieval process and the dexterity of the user at information search operation.

The Internet and computer technology have immeasurably increased the availability of information. However, as the size of information systems increases, it becomes harder for users to retrieve relevant information. Search engines have been developed to facilitate fast and efficient information retrieval. There are many software packages for search engine construction on the Internet. The website searchtools.com alone lists more than 170 search tools, many of which are free or free for noncommercial use.

In information retrieval, a common problem known as the *missing data issue* commonly arises. This problem is due to common human errors at database creation and management or on the part of the user.

Archie

Prior to the existence of web search engines there was a complete list of all webservers. This list was edited by Tim Berners-Lee and hosted on the CERN webserver. One historical snapshot from 1992 still remains. As more and more webservers went online the central list could not keep up. On the NCSA Site new servers were announced under the title "What's New!" but no complete listing existed any more. The very first tool used for searching on the (pre-web) Internet was Archie. The name stands for "archive" without the "v." It was created in 1990 by Alan Emtage, a student at McGill University in

Montreal. The program downloaded the directory listings of all the files located on public anonymous FTP (File Transfer Protocol) sites, creating a searchable database of file names; however, Archie did not index the contents of these sites. Sample query form is shown in figure 2.9.

Goper, Veronica & Jughead

The rise of Gopher (created in 1991 by Mark McCahill at the University of Minnesota) led to two new other search programs, Veronica and Jughead. Like Archie, they searched the file names and titles stored in Gopher index systems. Veronica (Very Easy Rodent-Oriented Net-wide Index to Computerized Archives) provided a keyword search of most Gopher menu titles in the entire Gopher listings.

Figure 2.2: Archie Query Form

Jughead (Jonzy's Universal Gopher Hierarchy Excavation And Display) was a tool for obtaining menu information from specific Gopher servers. While the name of the search engine "Archie" was not a reference to the Archie comic book series, "Veronica" and "Jughead" are characters in the series, thus referencing their predecessor (Wall, 2007).

Wandex, Aliweb, & Jumpstation

In June 1993, Matthew Gray, then at MIT, produced what was probably the first web robot, the Perl-based World Wide Web Wanderer, and used it to generate an index called 'Wandex'. The purpose of the Wanderer was to measure the size of the World Wide Web, which it did until late 1995. The search engine Aliweb appeared in November 1993. Aliweb did not use a web robot, but instead depended on being notified by website administrators of the existence at each site of an index file in a particular format.

JumpStation (released in December 1993) used a web robot to find web pages and to build its index, and used a web form as the interface to its query program (Broder, 2002). It was thus the first WWW resource-discovery tool to combine the three essential features of a web search engine (crawling, indexing, and searching) as described below. Because of the limited resources available on the platform on which it ran, its indexing and hence searching were limited to the titles and headings found in the web pages the crawler encountered.

Webcrawler

One of the first "full text" crawler-based search engines was WebCrawler, which came out in 1994. Unlike its predecessors, it let users search for any word in any webpage, which has become the standard for all major search engines since. It was also the first one to be widely known by the public. Also in 1994 Lycos (which started at Carnegie Mellon University) was launched, and became a major commercial endeavor.

Soon after, many search engines appeared and vied for popularity. These included Magellan, Excite, Infoseek, Inktomi, Northern Light, and AltaVista. Yahoo! was among the most popular ways for people to find web pages of interest, but its search function operated on its web directory, rather than full-text copies of web pages. Information seekers could also browse the directory instead of doing a keyword-based search (Craswell, et al., 2001).

The Big Five Search Engine

In 1996, Netscape was hoping to give a single search engine an exclusive deal to be their featured search engine. There was so much interest that a deal was struck with Netscape by 5 of the major search engines, where for $5Million per year, each search engine would be in a rotation on the Netscape search engine page. These five engines were: Yahoo!, Magellan, Lycos, Infoseek and Excite.

Search engines were also known as some of the brightest stars in the Internet investing frenzy that occurred in the late 1990s. Several companies entered the market spectacularly, receiving record gains during their initial public offerings. Some have taken down their public search engine, and are marketing enterprise-only editions, such as Northern Light. Many search engine companies were caught up in the dot-com bubble, a speculation-driven market boom that peaked in 1999 and ended in 2001 (Wall, 2007).

Google

Around 2000, the Google® search engine rose to prominence. The company achieved better results for many searches with an innovation called PageRank. This iterative algorithm ranks web pages based on the number and PageRank of other web sites and pages that link there, on the premise that good or desirable pages are linked to more than others. Google also maintained a minimalist interface to its search engine. In contrast, many of its competitors embedded a search engine in a web portal.

Figure 2.3: Google Query Page

By 2000, Yahoo was providing search services based on Inktomi's search engine. Yahoo! acquired Inktomi in 2002, and Overture (which owned AlltheWeb and AltaVista) in 2003. Yahoo! switched to Google's search engine until 2004, when it launched its own search engine based on the combined technologies of its acquisitions. Microsoft first launched MSN Search® (since re-branded Live Search) in the fall of 1998 using search results from Inktomi. In early 1999 the site began to display listings from Looksmart blended with results from Inktomi except for a short time in 1999 when results from AltaVista were used instead. In 2004, Microsoft began a transition to its own search technology, powered by its own web crawler (called msnbot). As of late 2007, Google was by far the most popular Web search engine worldwide. A number of country-specific search engine companies have become prominent; for example Baidu is the most popular search engine in the People's Republic of China and guruji.com in India (Wall, 2007).

Search engines are resources to assist users in information retrieval. Information retrieval is inherently predicated on users searching for information from their "information need" that result from the interpretation of the decision problem. Apart from the quality of data and information in the data warehouse, the volume, timeliness of the information to the decision maker is equally important. Thus, the importance of the search engine is vividly expedient in this regard. While the information need might be right, inherent errors resulting from dirty data are detrimental to the overall goal of information retrieval. In this research, we adopt the definition of data quality describing it as "data fit for use by data consumer (EI actors)" based on the inability to have a uniform model for what constitute a quality data. In next section, we shall take a look at different model of information retrieval employed, with their composition and pattern of operation.

2.2.2 Information Retrieval Models

Since most search operations are performed on the internet or corporate organizations expensively constructed and maintained data warehouses, the focus of this section will be to understudy common models employed in the search operation. Although Broder (2002) a prominent researcher in information retrieval building experience with names like Alta Vista, IBM amongst other submitted that the main difference with the classic model for IR and the one augmented for the web lies in the replacement of "Matching Rules" in the former with "Search Engines" in the latter. This is shown in figure 2.4 below. Thus we shall refer to IR models and search engines in this research synonymously while search operation can be attempt to retrieve text, audio, video, images (both still and moving) or a combination of any of these. Our focus here as earlier mentioned lies in the textual information retrieval and that is what we are sticking to throughout this research.

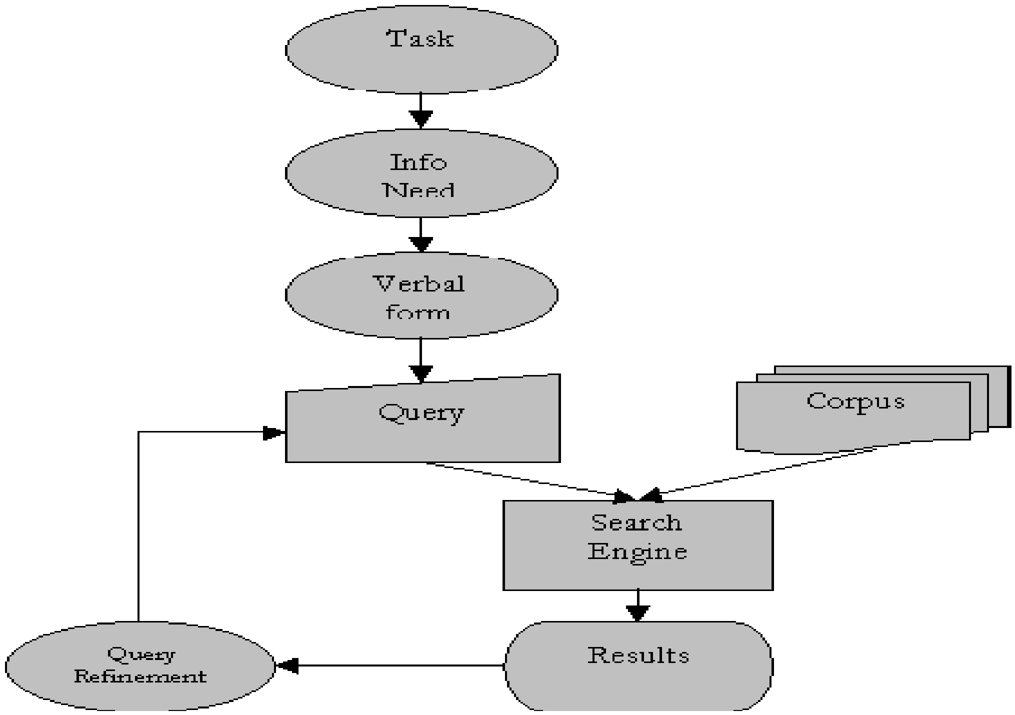


Figure 2.4: Web-Augmented Classic Information Retrieval Model (Broder, 2002)

The web-augmented model was born out of the fact that there is need for integrating human-computer interaction and the cognitive aspect of the user. An information retrieval object is described as an artifact solving a more or less general IR issue. IR object was described as composed of three components, Tasks, Forms and the Context (Canfora & Cerulo, 2004). With this notion, Broder, (2002) linked information need with some set of tasks.

Information need is usually verbalized (silently, mentally, not loud) and this is translated into query submitted to the search engines. This information need determine the nature of the queries submitted for selection from a collection of documents (corpus) based on the matching rules. This background allows recognizing different sets of users. A user can either be experienced, versed, or inexperienced. The manner via which each of these users constructs their query goes a long way to determine the probable result from the databases. In the area of cognitive model for web search we have Navarro-Prieto, et al., (1999), Muramatu & Pratt (2001) explore user's mental model for search engines with other related results presented in Choo, et al., (1999). In all these models, there is an agreement that web searches are sequel to user's information need.

There are established models of IR which are the Boolean, Vector space, Probabilistic, and Fuzzy models. Various implementations of these models are in existence, but more and more surfaces as their limitations become apparent for retrieval purposes. Another popular approach to IR is based on the method of analysis of natural language (Smeaton, 1992). It was however found out that this method is limited in the level of deepness of the analysis of the language, and their consequent range of applicability i.e. *satisfying interpretation of the documents' meaning needs a too large number of decision rules even in narrow application domains* (Crestani & Pasi, 1999).

The main components of IRSs are: collection of documents, a query language allowing the expression of selection criteria synthesizing the user's needs, and the matching mechanism which estimates the relevance of the documents to the query (Passi, 2008). Attempt to estimate the relevance of each document with respect to a specific user need is based on a formal model which provides a formal representation of both documents and the user queries. Using the trio of documents collection, query language and the matching mechanism in an IRS, the input represents the user's query while the corresponding output reflects the relevance estimation of the user information need (query) and the information collection.

Existing IRSs and search engines provides simplified modeling of IR put privileges efficiency at the expense of effectiveness (Passi, 2008). Researchers have shown that the effectiveness of an IRS is a function of the query language that is expected to represent in a more accurate and faithful manner, the user's information needs (Bordogna & Pasi, 2001).

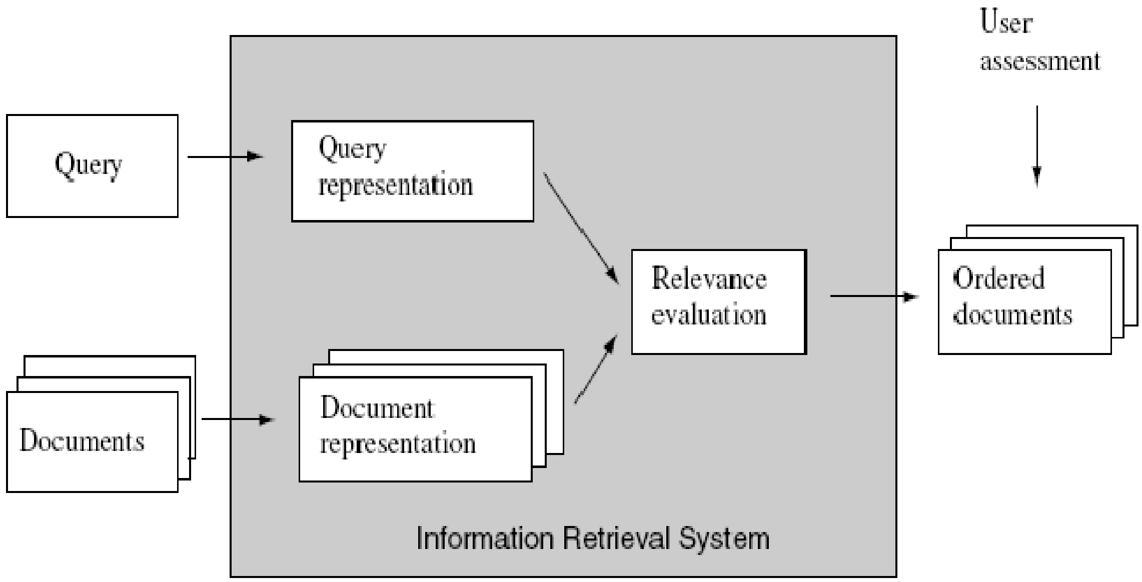


Figure 2.5: Schema for Storing and Retrieving information (Crestani & Pasi, 1999)

Most popular query languages are based on keywords specification. The mode fails to encourage expressing uncertainty and vagueness constraints that are relevant to information need. Apart from the query language, of importance is the representation of the document's information content, this takes the form of keyword extraction and weighing. This is shown in figure 2.5 schematically. It is however worth noting that documents representation are done without taking into cognizance the subjective view of the users on the documents (Bordogna & Pasi, 1995).

Attempt at making sure that a user retrieve documents "*relevant*" to her query necessitate a formal representation of the documents contents known as "*Indexing*". With this method, analysis of a document is followed by a surrogate describing the document in the index. With this in mind, a query to an IRS provides either an exact answer or a ranking of document with highest possible relevance (Fuhr & Buckley, 1991). The result thus is a function of the formal model adopted in designing the system. Subsequently, we take a look at some common models like the Boolean model, Vector space model, the Probabilistic model and the Fuzzy model.

Boolean Model

The Boolean model is a popular example of formal representation employed in defining both the documents and the user query in a bid to adequately model the IRS. Unlike other models, the Boolean model provides an exact answer. It is based on mathematical theory where documents are represented as a set of index-terms (Lee, 1994). It is however unfortunate that the index terms cannot be used to differentiate the information content (Moens, 2000). A Boolean model query consist of logical formula made up of index terms, and logical connectives (AND, OR, NOT,..). No document is thus relevant except it satisfies the logical formula representing the query (Crestani & Passi, 1999).

Vector Model

Vector representation considers each document as being described by a vector of components that are representative of the semantic content of the document. Traditional vector employs keywords also known as index term, other popular version is the n-grams (Canfora & Cerulo, 2004). Vector model is also said to be based on spatial interpretation of both document and queries. As opposed to the Boolean model, it associate with each index term a numeric value referred to as *index term weight*. This weight helps expresses the variable degree of significance accruable from the synthesis of the information content (Crestani & Passi, 1999). Vector categorization can further be presented as Binary – text documents is represented by a binary vectors of terms, weighted – either by Term Frequency (TF) or Inverse Documents Frequency (IDF) (van Rijsbergen, 1979).

Probabilistic Model

In this model, documents are ranked in decreasing order based on the probability of user's relevance ratings alongside her/his information needs. Formal theories of probability are very common in IR research using estimation or evaluation of the probability of relevance (Crestani, et al., 1998). In probabilistic IR models, documents are ranked according to their estimated probability of being relevance (R) via $P(R \mid q, d)$. However, accurately predicting the relevance of every document in a collection is difficult because there are large numbers of variables involved in the representation of documents in comparison to small document relevance information at its disposal.

Probabilistic model employ concepts and techniques related to logic and artificial intelligence (Crestani & Passi, 1999).

Soft Information Retrieval Models

The need for accommodating vagueness and uncertainty in database representation has been the basis for the introduction of fuzzy systems into the field. Popular soft information retrieval models are based on fuzzy set theories and the connectionist (neural networks) theory. Considering the Boolean query where achieves are partitioned into two i.e. the relevant documents and the irrelevant documents. This crisp partitioning are liable to reject relevant items because of strict queries and sometime bring out irrelevant results (Salton, 1989).Fuzzy modelling IR approach is based on the use of linguistic information at various level in the retrieval process (Bordogna & Passi, 2001). Another approach employing fuzzy method is for defining flexible query languages capable of capturing the vagueness of user needs as well as simplifying user-system interaction.

The problem of uncertainty in decision problem formulation, interpretation and information retrieval problem has necessitated the need for a model capable of accommodating the vagueness and uncertainty that persist in human-computer interactions. This section has taken a look at the various models of information retrieval and search engines. We shall hereafter consider the representation and organization of information as a means of reducing the effect of missing data/information.

2.2.3 Information Representation and Reasoning Strategies

Information representation has been treated to some extent in the previous section. In this section, we hope to pinpoint the interaction between the representation and reasoning strategies as being employed in IR and other search engines. Modeling the process of information retrieval is generally complex consequent upon the fact that they are multifaceted and inherently endowed with vague concepts difficult to formalize. Human component has been of focus lately considering the issue of relevance, information need and other subjective factors. The importance of reasoning strategy employ alongside the representation of information shall be considered here. Reasoning strategy facilitates the representation of similarity problem in computing the relevance of a document with regards to submitted queries (Canfora & Cerulo, 2004). The bulk of our discussion in this section will therefore be based on the classifications and taxonomy presented by the above mentioned authors.

Our first consideration in this section will be the representation of an information retrieval model by Naeza-Yates & Riebeiro-Neto, (1999). Information retrieval model was characterized by a set of quadruple given as {D, Q, F, R(q, d)} and these factors were defined as follows:

- D is a set of *logical views* for the documents in the collection, it is a *representation component;*
- Q is a set of *logical views* for the user information needs, it is a *representation component;*
- F is a *framework for modeling* *document representation*, *queries* and *their relationships*, it is a *reasoning component*;
- R(q, d) is a *ranking function* which *associates a real number with* *a query* $q \in Q$ and *a document* $d \in D$. It is a *reasoning component*.

The above characterization forms the basis upon which the design and models of this research is anchored. While there are many of such characterization, we submit that this adequately capture the necessary ingredients in understanding the properties of any information retrieval model. Thus, we shall employ these properties and compare the provisions of major search engines with our model to establish their strength and weaknesses in latter chapters.

Considering the submissions of Naeza-Yates & Riebeiro-Neto, (1999), we can make some mapping and clarifications. Principally, information retrieval models can be partitioned into its representation and reasoning components. The representation involves the document collections and the user's information needs, while the reasoning consist of the framework that models the representation (documents and query) and the relationship between them, and a ranking function that helps determine the relevance of the documents with the query. The figure below presents the vertical taxonomy of information retrieval models.

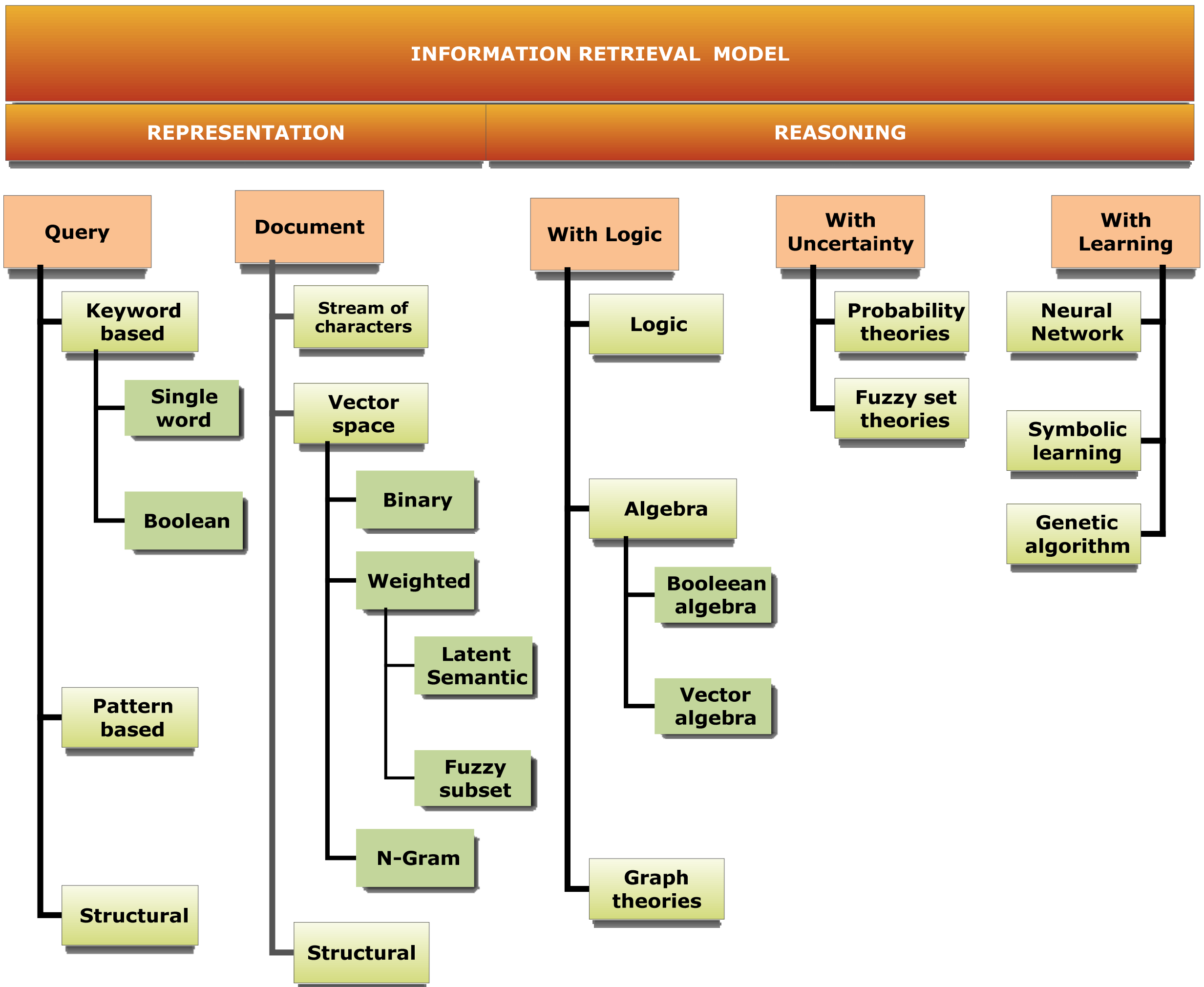


Figure 2.6: Vertical Taxonomy of IR models (Canfora & Cerulo, 2004)

The taxonomy presented in figure 2.6 is an expanded view of the characterization done by Naeza-Yates & Riebeiro-Neto, (1999), however, Canfora & Cerulo, (2004) made an explicit documentation of these characteristic into their atomic nature. Most of the existing work on information retrieval object focused on the representation aspect. There is no subset under both query and document representation that has not been taking into consideration in the design of retrieval tool (Sebastiani, 1998). Unfortunately, the story is not the same for reasoning as there appeared sparse consideration for most of the patterns hereby mentioned. Reasoning with logic has a fairly large attempt in retrieval tools design, but others were sparsely featured.

The above disparities in consideration of both components of information retrieval models have been the basis for the lopsided query results for users' information needs. As an example, imagine after the deliberation between a decision maker and the watcher, the information specialist attempting to make available information he deemed fit for the decision problem formulate his query based on many factors (cognitive ability, experience, understanding of the decision problem, e.t.c.) and expect a result. Unfortunately, the available tools for retrieval is based majorly on representation and less on reasoning, thus result for the query will be based on either relevant or irrelevant and not necessarily accommodating some vagueness and uncertainty in the retrieval operation. In subsections, we shall deal with the two partitions (representation and reasoning) and their components.

Representation in IR Models

The mode and manner of representing information determines the level, rate and accuracy of access to it. As the focus of this research is in textual information, it representation engulfs both the documents and the queries. A document is the representation of the information to be encoded, while a query is the representation of information need by the user. We can characterize a text as a quadruple: syntax, structure, semantic, and the style. The syntax and the structure are determined by the application and the creator of such information. The semantics are specified by the authors. The styles could be a result of the pattern or manner we want the document to appear e.g. Microsoft word or Latex format. It was however said that, it is generally sufficient to characterize queries by the structure and semantics.

Query Representation for Information Retrieval Models

Decision problems usually elicit information retrieval problems which in turn constitute the information need that forms the query. Information need is inherent in the user, it is implicit in the user's mind with a sole aim of bridging the knowledge gap (Canfora &

Cerulo, 2004). There are principally three types of information needs: *known item information need*, *conscious information need, and confused information need*. In the first case, it involve search to verify the existence of known documents. In the second cased, the user search for documents they do not know, but it pertains to the subject they know. The third category implies the user is familiar with neither the documents nor the subject (Mizzaro, 1996).

Query representation can follow *keyword-based, pattern-based or structural*.

- In the keyword based, the queries are searched for some keywords and documents with such keywords are made available as retrieved objects. Keyword queries are popular because they are intuitive and easy to express. The implementation of keyword-based can be in single word or Boolean. The single word is the most elementary, forcing a match between the user query and the keyword. The Boolean is the oldest; it uses a combination of keywords. Usually, a Boolean query constitute an expression whose elements are keyword, present Boolean operation employs the NEAR operator which is a move away from the canonical AND and OR.
- Pattern based query presentation provides a more specific query formulation that facilitate specification of text with some properties. Pattern based employs syntactic features which are match with the segment specification of the documents.
- Structural queries are tailored towards structured information. It is usually an embedded mechanism on basic queries with facilities for expressing structural constraints with containment, proximity, or other restriction on the structural elements in a document. It is usually categorized into three: fixed structure which are very restrictive. The fixed structure restricts the search to a particular document field. The second is the hypertext, this is about the most flexible using a directed graph with nodes holding texts and links representing connections between nodes. The main drawback of this structure is that, it is not easy to query the hypertext structure connectivity, except the text content of each node. This structure represents the browsing saga for navigational purposes. The third type is

the hierarchical structure. It is a representation of natural decomposition of many text collections (books, articles, structural programs etc.).

Documents Representation for Information Retrieval Models

We have earlier discussed about document representation in our previous section. Consequently, we shall not flog the cat in this section again as most of the concepts are already discussed. Document representation has been via keyword from time immemorial. These keywords are extracted from the text or inserted by the authors. Document representation follows three main forms: *Stream of characters, Vector space, and Structural.*

The stream of characters had no interpretation on its structure and semantic content. The vector space method has some varieties. It describes each document as a component that is representing the semantic content of the document. Vector representation can be any of binary, weighted (latent semantic or fuzzy subset), and N-Gram. The composition of the structural representation is similar to that earlier discussed about query representation. Popular example is this regards is the XML for information modelling. In the next section, we shall be looking at the second broad division in information retrieval model – reasoning.

Reasoning Methods in Information Retrieval Models

Our earlier sections have dealt with the importance of documents and query representation towards a successful information retrieval process. Traditionally, information retrieval model depicts query and document representation as the duo that determines the results of any search operation (figure 2.12). The result of the query is thus from the judgment on the relevance established by the retrieval engine. In this section, another important dimension in information retrieval is reasoning which comprises of the methods, models, and technology used in the matching process (Crestani & Lalmas, 2000).

Attempt to understand the concept of reasoning will facilitate a better appreciation of the reasoning component in information retrieval (Canfora & Cerulo, 2004). Cooper, (1971) was about the first recorded authors that considered the logical definition of relevance in IR. The understanding of relevance in IR is as follows:

> "*A stored sentence is logically relevant to (a representation of) information need if and only if it is a member of some minimal premise set of stored sentences for some component statement of that need*"

A more popular definition of relevance was given in (Saracevic, 1975). Therein *"Relevance is the* (A) *of a (B) existing between a (C) and a (D) as determined by an (E)". Where:*

(A) - measure, estimate, judgment …

(B) - utility, matching, satisfaction …

(C) - document, document representation, information provided…

(D) - question, question representation, information need …

(E) Request, intermediary, export …

The above considerations show some of the views adopted on the concept of relevance. Now relating reasoning to relevance, we deduced that reasoning involved the set of the methods, models and technologies employed in matching the document and query representation of an information need. Canfora & Cerulo, (2004) submitted that reasoning strategy in information retrieval consist of one or combination of the following: reasoning with logic, reasoning with uncertainty, and reasoning with learning. These were figuratively displayed in figure 2.13. Subsequently, we take a brief look at these models.

Logical Approach to Reasoning

- Logic – models of these forms are usually mathematically inclined. Its approach can be formulated in terms of P(d → n), P is the predicate i.e. representation of document d is relevant to the representation of information need n. The arrow serves as the conditional connective based on formalized logic. A vivid description of logic-based information retrieval is presented in Sebastiani, (1998).

- Algebra – this is based on algebra calculus, it has been popularly adopted. Its operations are based on the definitions in an algebraic domain. Other variations of this model are the Boolean algebra, which employ Boolean expression to verify the relevancy of a condition and the Vector algebra. Vector algebra employs weighting scheme for both documents and query representation. It also facilitates ranking similar to the document vector space (Salton, 1989). Other variants of this model were covered in van Rijbergen, (1979).
- Graph theories – this model operate based on the structures formed by the vertices and edges of the graph. Its popularity booms with the advent of the web as it models web resources perfectly. Flake, et al., (2002) added the maximum flow method aimed at identifying the web communities as opposed to hitherto approach used for bibliographies. Other examples in web context are PageRanking algorithm (Brin, et al., 1998) which is now employed by Google search engine (Google, 2006).

Uncertainty Approach to Reasoning

Dealing with uncertainty in information retrieval has the following approach

- Probability theories – information retrieval with uncertainty based on probability premise its operations on the probability that the user will find a relevant document amidst the collection. It was introduced by Robertson & Sparck, (1976). Popular variants include the Bayesian network, inference network (Turtle & Croft, 1990). Inference network forms the spine of INQUERY (Broglio, 1995).
- Fuzzy set theories – this model gains acceptance due to the need to manage vagueness, incompleteness and uncertainty in users query formulation. It operates contrary to the exact Boolean match for query and the documents resulting from the crisp nature of Boolean operations. There are several version of its application based on the work of Zadeh, (2002). It facilitates a

robust manner of handling query formulation and treatments with concepts like linguistic quantifiers using "at least k" or "about k".

Learning Approach to Reasoning

The user dimension to information retrieval has seen the introduction of learning to the IRP. The operations are based on neural networks, symbolic and inductive learning amongst others. We take a look at some these approaches.

- Neural networks – this operates based on the body system neurons. Its development involves layers as found in Belew, (1989) among the first proponent using a three-layer (author, index term, and documents). In this model, user relevance feedback was used to alter document representation with time. Other variation for specific task can be found in Lin, et al., (1991), and Macleod, & Robertson, (1991).
- Symbolic learning – this model has limited application range i.e. limited with respect to other learning techniques. It represents the numeric classification of the IF-THEN rules. Variations can be found in Chen & She, (1994) and Flake, et al., (2002).
- Genetic algorithm – this has several applications developed in this regard. It features in documents indexing (Gordon, 1988), and document clustering (Gordon, 1991). In the design, a keyword represents a gene (bit pattern), a document is a vector of keywords (bit string) represents individuals, collection of documents initially adjudged by the user as the initial population.

A complete reference of information retrieval model can be found in Canfora & Cerulo, (2004). The taxonomy presents a robust opportunity to cross-match existing models (objects of retrieval) with new design, focusing on the provision by the models as regards representation and reasoning. In this section, we have reviewed several provisions in information retrieval world, juxtaposing their functionality and compositions to determine their effectiveness. The goal of this work is to determine how best the problems of missing data/information can be minimized and this has provided ample background for understanding this problem.

In next section, we shall take a look at the effect data/information resulting from the various object of search employed by many information specialists for decision making. The quality of such retrieval exercise is our focus here

2.3 Data and Information Quality

Considering from different perspectives, the term information has several faces to its merit. It is sometimes used to succinctly imply data. The distinctions between the duos thus become imperative, or better still require a research into the proper definition from literatures about the two concepts. We shall attempt to gain clarity by juxtaposing the various definitions of the term information, data and the third party which is knowledge. It is important to note that, the differences in the conflicting notions are sequel to divergent perspectives and understanding of the term.

Three main reasons were identified to ginger the study in the differences namely: the consideration from the business environment where information technology usage results into production of something that is useful and of value i.e. information. The mode of identification from various disciplines which include information management, library studies e.t.c and finally the colloquial use of the term information present divergent and conflicting opinion on the same term being used by different people to connote different things (Knox, 2007). In the opinion of Davenport, (2000), information has been neglected at the expense of more tangible resources. The consequence is that, historically the information system arena has focused on the “T” technology in information technology at the expense of the ‘I’ information.

In this section, we intend to rigorously review various submissions on these concepts, while pinpointing the basis for the misconceptions if there are any, and also present our opinion for this research as regards the three concepts. These will be followed by the distinction between data and information quality, enumeration of various quality metrics as seen in literatures and their effect in delivering strategic decisions. We shall also take a look at the accruable risks from the usage of such data.

2.3.1 The Misconceptions Amongst Data, Information & Knowledge

Information was referred to as 'the knowledge about the decision', 'the effect of its alternatives', 'the probability of each alternative' amongst others (Harris, 1998). It is sometimes referred to as meaning, drawing upon data interpreted (created) by human sense making processes, including application of bias by its producer (Bednar & Welch, 2008). We note the inclusion of the word "bias" in the constitution of information definition as given by the authors, and stressed that, its going to constitute major part of risk factor determination in this research. Dumas, et al., (2005) conducted variety of research on the need to assert if there is any notable differences between communication and information. Popular knowledge management literatures attempts to make distinction among data, information and knowledge. Tuomi, (1999) opined that the general notion is that, data is viewed as a facts that becomes information as they are combined into meaningful structures, which would subsequently result into knowledge through the application of meaningful information in a desired environment towards the goal of making a prediction. Data again was referred to as a set of discrete, objective facts about events (Davenport & Prusak, 1998).

O'Brien

O'Brien (2003, 2004) defines "*information as data placed in a meaningful and useful context*", the Oxford English Dictionary (OED) (1989) described *information as knowledge concerning some particular fact, subject, or events* i.e. information is seen as knowledge. The impact of human element in the "learning process" prior to transformation of information into knowledge was properly presented in Davies & Ledington (1991). The Oxford English Dictionary (OED, 1989) again reiterated that knowledge communicated concerning some particular fact, subject, or event, whether be it appraised or told, intelligence, news especially contrasted with data. The above implied, the reader/user is expected to make appropriate referencing from available information. Thus, verbs like "communicated, appraised, told" formed the basis for the work of Bednar & Welch (2008) in their work titled bias, misinformation and the paradox of neutrality.

The popular reference and usage of the term "information" in everyday life can almost present a universally accepted view of the concept. Unfortunately, this understanding or supposed acceptance is more of 'grey' mix message sense type of acceptance similar to the same way the concept of jealousy is acknowledged but differently interpreted by individuals (Knox, 2007). This unclassified reference of terms is a basis for ambiguity and confusion which in turn can result into drastic consequences (Wilson, 1996). This again was corroborated in the work of Checkland & Howell (1998) while exhibiting the confusion and the basic differences between the basic concept of data, information and knowledge and their relationship.

Gackowski

There is need for a rigorous distinction between data and information (Gackowski, 2005) to address the general notion and their usage synonymously. Whatever is known, given, or available constitute the data parts, while everything not presently available, missing, and thus requires proper intelligence or other means constitute the information parts towards decision making. Although the difference seems innocuous, it has resulted in the qualitative and quantitative differences in decision situation and operations. Consequently, it was stressed that, data value- (known already) never changes decision situation for they have or should have been already accounted for, however, information value, if only relevant, of significant impact or materiality, always changes decision situation model and/or implementation of the decision made and/or the result of the affected operations (Gackowski, 2006).

Davenport

In the view of Davenport (1997), he described data as simple observation of states of real world, while information is data endowed with relevance and purpose, and knowledge is described as valuable information. As quoted in Knox (2007), referring to the opinion of Sveiby (1997) where information was described as meaningless, and only becomes meaningful when it is interpreted. There seems to be no end to the inter-usage of these concepts, in another submission two sets of equation was given in an attempt to show the relationship and their interdependency.

Data + Information = Knowledge

Data + Information + Knowledge = Wisdom

Lueg, (2001) opined that knowledge is often viewed as information with a specific properties, while information serves as a preliminary stage to knowledge. The opinion above elicits some hierarchical transition in the development and understanding of the concepts we have been discussing. The view of English (1999) is such that: information is data in context. Information is usable, it is the meaning of data, and thus fact becomes understandable. An attempt to translate the above into knowledge appropriate for deriving decisions from the view of Chaffey & Wood (2005) can only result from the addition of expert's opinion, skills and experience to the combination of data and information.

Harper

Harper, (1999) argued that it is not enough to say that information is somewhere out there in the world brought about via some processes of human agency and technology. Rather, information was relevant when it is seen by all the parties involved. Fox, (2004) was of the opinion that data are stored in many formats and encoding via increasingly sophisticated conventions and standards. Considering from this perspective, information with meaning was referred to as knowledge (O'Leary & Selfridge, 2000). Bringing in again the human dimension into the scene, Davenport, (1997) opined that information and knowledge are characteristically human creations, thus their management will continue to pose a problem unless we give people a primary role.

Knox

It has been widely supported in many literatures that there is a misunderstanding and a poorly discussed issue regarding the notion of what is tagged information. The problem of distinction is again noticeable in the submission of Kroenke, (2007) where information is ambiguously defined as "a difference that makes a difference". Knox, (2007) thus submitted that, much of the problem revolves round gaining a clear, workable definition of the term as well as an issue of how to manage and create value from information. The table below presents several notable propositions on these concepts as adapted from various authors.

Author	Data	Information	Knowledge
Drucker (1989)	Data is relatively easy to capture and does not necessarily require analysis	Information is data endowed with relevance and purpose	Knowledge by definition is specialized
Davies & Ledington (1991)	Data consists of many individual bits/ pieces/items or facts that can be simultaneously, or sequentially, processed to support the learning process	Information is not some object that exists in the world – information is part of the learning process – information has meaning according to the interpretation which is happening	No actual reference to the term knowledge
Boddy, Boonstra and Kenndy(2002)	References to recorded descriptions of things, events, activities and transactions	Information is data that has been processed so that it has meaning and value to the recipient	No clear definition is offered except to state certain information systems help people to make decisions by incorporating human knowledge into the system
Davenport (1997)	Simple observation of the states of the world - Easily structured - Easily captured on machines - Often quantified - Easily transferred	Data endowed with relevance and purpose - Requires some unit of analysis - Need consensus on meaning - Human meditation necessary - People turn data into information	Valuable information from the human mind, includes reflection, synthesis, context - Hard to structure - Difficult to capture on machines - Often tacit - Hard to transfer
Checkland & Holwell (1998)	Data are checkable, facts that can be agree, disputed both of which allow evidence to be brought forward	This is data – capta that then has been enriched. I.e related to other things, seen as part of a larger whole – gains significance	Larger structures of related information – expected to have longevity
Chaffey & Wood (2005)	Decrete, objective facts about events. Data are transformed into information by adding value through context, categorization, calculations, corrections, and condensation	Organised data, meaningful and contextually relevant. Used for decision making	The combination of data and information to which is added expert opinion, skills and experience to result in a valuable asset which can be used to make decisions

Hislop (2005)	One could see data as being raw numbers, facts, images, words, sounds based on observation or measurement	Information represents data arranged in a meaningful pattern, data where some intellectual input has been added	Means to analyse / understand information / data, belief about causality of events / actions, and provide the basis to guside meaningful action and thought. That is one could say knowledge can be understood to emerge from the application, analysis and productive use of data and/or information
Gackowski, (2005)	Whatever is known, given or available	Not personally available, missing	No reference to Knowledge
Oliveira et al. (2005)	Something that one can take inventory about	Yet known & must be acquired	Application of the knowledge reasoning
Gackowski, (2006)	Never changes decision situation	Have significant input in decision situation if relevant	No reference to knowledge

Table 2.1: Data, Information, Knowledge Distinction (Adapted from Knox 2007)

A pertinent question can thus be: "at what instance does data transits into information, and information into knowledge". Knox, (2007) in an attempt to answer the above raised question posed some other questions as follows: is there a definite definition of the term information? How is information created or is it simply a 'step' in the traditional hierarchical structure of data, information and knowledge? The foregoing intimates us of the probability of a transition from one of these entities to the other, thereby implicitly pinpointing inherent differences but showcasing their interdependency. Marchand, (2000) was strong of the opinion that information is the way via which business people express, represent, communicate, and share their knowledge attempting to establish an extremely delicate relationship between information and knowledge.

Callaos & Callaos

The pattern of usage and adoption employed by many researchers on the term information is obviously divergent. This factor thus contributes in no small measure to present and reflect different perspectives (Callaos & Callaos, 2002). In the "Infological equation" proposed by Langefors (1966) and shown in figure 2.7, information was paralleled with meaning brought about via interpretation. This was however in contrast with the submission of Shannon & Weaver, (1949) in the model of communication where the authors reiterated that "information must not be confused with meaning". It has however being argued that Shannon (1938) gave a mathematical definition and not a conceptual one. This stems from the fact that it is wrong to confuse a measure of a thing with the thing measured, let alone confused with metrics, with the thing measured by it (Callaos & Callaos, 2002).

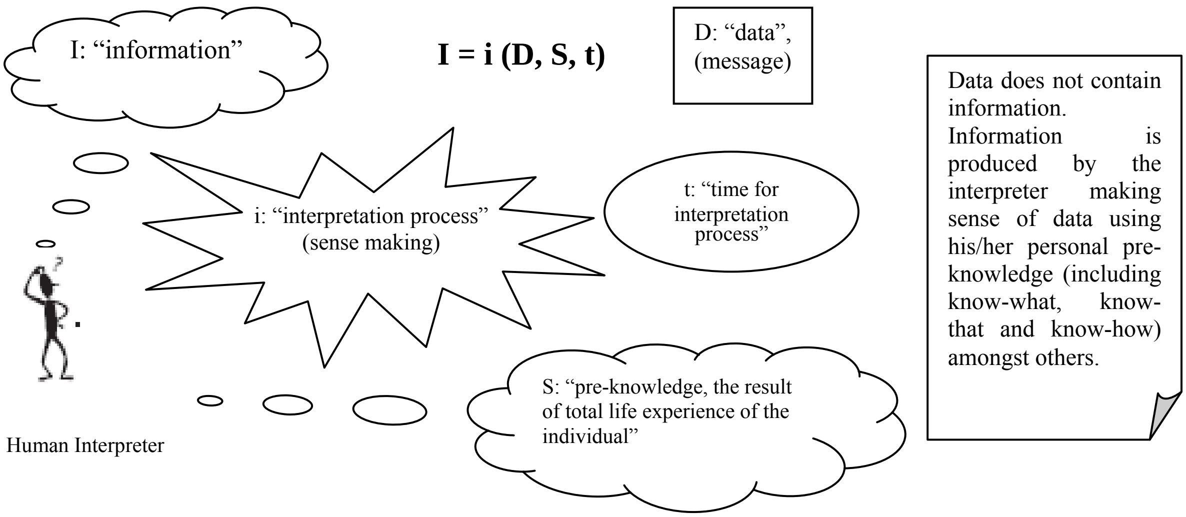


Figure 2.7: The Infological Equation (Badnar & Welch, 2007)

Langefors

In demonstrating the workability of the 'infological equation', Langefors (1966, 1995) was of the opinion that those who are to interpret data in order to inform themselves must be viewed as part of the system. Using the equation I = i (D, S, t), where 'I' is the information (knowledge) produced by a person from the data 'D' alongside with pre-knowledge 'S' through an interpretation process 'i' in interval't'. Bednar & Welch, (2007) reflecting on the opinion of Langefors made by Schutz, (1967) reiterated the impossible nature of communicating "meaning" between people. This was sequel to the fact that simply transmitting data will not lead to communication of shared understanding knowing fully well that 'i' and 'S' cannot be assumed to be common. To this end, communication can only be seen to approach success most closely where individuals interpreting the same data belong to a group with possible vested professional interest. This is sequel to the fact that every act of interpretation does not necessarily invoke the entire 'S' attribute to every individual, thereby creating room for some aspect of shared experience leading to similarities in the 'i' among group members (Bednar & Welch, 2007).

It is not a common place for a piece of data to generate similar 'factual' meaning when interpreted by different individuals. However, derivable inferences would be likely different more widely in 'meaning' of the data for different individual based on his/her associations, and/or possible consequences depending on the uniqueness of 'S'. Communication and intention is context-dependent. Interpretation of context continually evolves with time thus having great influence on sense-making and communication (by Wittgenstein, (1963) and quoted in Bednar & Welch, (2007)).

Floridi

In the opinion of Floridi (1999) "information is provided when data answer an explicit or an explicit question made by the data receptor". To this end, for data to be informative, it should be associated with a relevant question i.e. information consists of datum and relevant questions. Following the submissions of Floridi, Knox, (2007) made the following submissions that:

- A datum is a "given" thing, not "given" thing, but the one that makes a difference. So the gendre of datum is "to be given" and the characteristics that makes it specific.
- Information is a cognitive content, not any cognitive content, but one related to the association of data and a relevant question, be it implicit or explicit.
- Data and information are two sides of the same coin: datum is the objective side of the coin and information is the subjective side.

It is possible to ask the question: why go through this stress in determining what constitute a data, information and knowledge? The reason lies in the fact that if the misconception is not properly treated, then the conflicting notion will continue to disturb the way and manner by which they are employed and can therefore constitute a problem in the informing processes.

2.3.2 Discussion

Callaos & Callaos, (2002) typified information conception from two different perspectives: subjective and objective. This was necessitated by the authors concern for the incoherent homonyms resulting into divergent and dissimilar references in the usage pattern of the term.

The infological equation as seen in the preceding section distinctively identify differences between data and information, however, information was synonymously referred to as knowledge brought about via the interpretation and pre-knowledge of the individual in question. Information has been widely defined as “interpreted data” thus bringing in a reference to the ‘i’ as depicted in the infological equation. However, if this is anything to follow, then different people viewing same data will come up with different interpretations, i.e. propose different meanings to the same data (Callaos & Callaos, 2002). The subjective nature, i.e. related to a subject – “a mind, ego, or agent of whatever sort that sustain or assumes the form of thought or consciousness” of information was established. Another similar view to the one expressed above is the description of information as “data plus meaning” or “meaningful data” (Checkland & Scholes, 1990; Mingers, 1997). With the following, an etymological perspective of the term ‘data’ and ‘meaning’ was undertaken for Callaos & Callaos (2002) and the authors made bold to say that “information” as “meaningful data” would be defined as “significant data”, “data full of meaning”, “data having a meaning or purpose”, similarly, “data plus meaning” would be defined as “data plus significance”, “data plus the thing conveyed by it in the mind”.

The views of different authors in the nature of information list are almost unending. This led to the opinion of Dervin, (1989) that since all information producing is internally guided, and also accepted that all human observing is constrained, sense-making corroborates the subjective view of information.

Relating information and knowledge, Neil (1992) emphasized that “knowledge representation are not knowledge but rather representation of knowledge”. This submission further geared the notion that information is generated inside the mind of a person or a subject. It is dependent on the person where it is generated by the data stimulus, coupled with his/her individual experience. Information in the view of Koshen (1983) is “decision-relevant data”, thus remarking the subjectivity therein embedded.

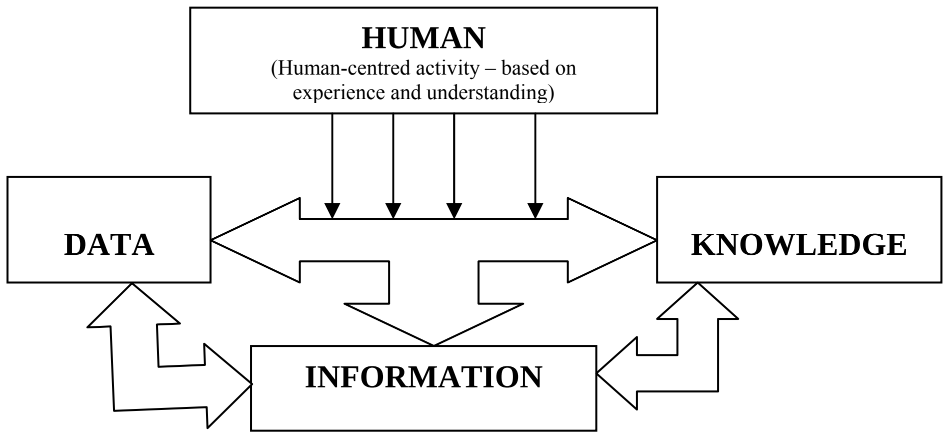


Figure 2.8: The Circular relationship amongst data, information & knowledge (Knox, 2007)

It is no gain saying that the manner by which individual interprets data, information and knowledge will have tremendous impact on what course of action they will undertake in collecting, managing and sharing such information within an organization. The interdependency of the trio again was the vocal point in Knox (2007) where several other authors’ opinions were juxtaposed leading to the formation of what is referred to as the circular relationship between data, information and knowledge. As opposed to the synonymous reference to data and information, Davies & Ledington, (1991) disagreed with this stating that information is more than data and thus requires some form of human interaction/involvement relating to interpretation (Minger, 1997). The element of interpretation is subjective (Merriam-Webster, 1999) and thus requires human involvement in making sense of something via their unique attribute with which they are endowed. Tuomi (2000) argued that there is a reverse hierarchy of data of data – information and knowledge as data emerged last only after knowledge and information are available. Consequently, Knox, (2007) asserted that data does not exist in isolation but is a result of human intervention by creating data via their knowledge and understanding and this was used to arrive at figure 2.8 above.

With various submissions from different authors presented in this section, we have attempted to dislodge the various conflicts arising from the indiscriminate use of these words: data, information and knowledge. Earlier attempt provides hierarchical view saying that, data leads to information and information to knowledge while some actions take place before the transformation. Lueg, (2001) defines knowledge as information with specific properties i.e. a preliminary stage to knowledge. In the view of Stewart, (2000), knowledge is a conclusion drawn from data and information. Maqsood, et al. (2004) unequivocally stressed the importance of human endeavour in determining the differences amongst the trio. Overall, we have been able to disambiguate the synonymous reference to data and information, and also go a step further to include knowledge from the presentation of popular literatures. In the next section, we shall be taking a look at the various information quality attributes and their taxonomical presentations.

2.4 Taxonomies of Data and Information Quality

Information quality has in recent times been in the forefront of major researches in many fields ranging from information systems (IS), data warehousing, data modeling, human computer interaction and data transmission to mention a few. The term quality itself is not an easily definable term, it is not absolute. To this end, it has many facets, and its meaning varies across different situations, users (Firquin, 1992; Burgess, et al., 2007) and queries (Bouch, 2000). Several authors have given the chronology of data quality based on different attributes and dimension. There exist several reasons for a data to be of poor quality, this include not reflecting real world condition, not easily understood, its timeliness amongst others. The cost of poor quality data must be measured in terms of users' requirement (Ishikawa, 1985). This is sequel to the fact that in the face of accurate data, if the user can not adequately interpret or have access in a timely manner, then it's almost valueless. The above present the major focus of this section. Hitherto, we have attempted to distinguish amongst data, information and knowledge and the way by which several authors have referred to them, however, in this section, the term "data quality" and "information quality" will be used synonymously. We hope to review attributes of information quality and their dimensions from several authors and possibly establish the relationship and interdependency amongst them. Research into information quality has many focus, these include:

- Data of high quality is seen as valuable asset.
- Data of high quality can increase customer's satisfaction.
- Data of high quality can improve revenues and profits.
- Data of high quality can provide a strategic competitive advantage

The above were the submissions of Lee, et al., (2006) and this is sequel to the general belief this day that accurate and timely-available information is strong weapon for any organization. It is note worthy to mention that the essence of this work is not to measure quality of information, but importantly the effect of such attributes in the overall performance of employing particular information for delivering strategic decisions. In the preceding section, we shall take a look at the popular chronological order of data quality as presented by different authors and thereafter we shall look at the various propositions offered by them

2.4.1 Chronicle of Data/Information Quality

A consistent and accurate chronology of the work on information quality might be difficult to present because of the diversity in focus, or more appropriately the mode of achieving quality presented by different authors. This becomes more cumbersome due to the varieties and the clumsy manner with which reference is made to the various dimensions of information quality. In this section we present a simple chronicle of data/information quality and examine various submissions made by each or group of the chronicler.

Wand & Wang

One of the most referred works on information quality is the work of Wand & Wang, (1996). The authors identified and derived four DQ attributes. Their work was based on distinctions between the external and internal views of an information system described in Wand & Webber, (1995). The external view is concerned with the use and effect of an information system, addressing and justifying the purpose for its development. Internally, it was considered “given” as a black box endowed with the functionality capable of representing real world system. It is the belief of the authors that data should be in an exhaustive mapping with the real world, i.e. there is a possibility to map a real world state into more than one state in an information system however, no two or more states in the real world can be simultaneously represented in an information system. With this notion, data is viewed as incomplete whenever there is no such one-to-one correspondence between information systems and the real world, or ambiguous if the information system’s state corresponds to more than one such real world states. Data, from their point of view is meaningless if a state in the information system fails to correspond to any real world states. The above ontological foundation based quality dimension yielded four dimensions: complete, unambiguous, meaningful, and correct. These attributes however turned out to be a part and not the whole of deemed fit attributes for intrinsic data quality. Liu & Chi, (2002) therefore concluded that, the ontological mapping approach excludes many other important attributes unspecified. Table 2.2 below gives the summary of the ontological foundations to data quality dimensions.

D.Q. Dimension	Nature of Associated Deficiency
Complete	Improper representation: missing IS states
Unambiguous	Improper representation: multiple RW states mapped to the same IS state
Meaningful	Meaningless IS state and Garbling (map to a meaningless state)
Correct	Garbling (map to a wrong state)

Table 2.2: Intrinsic Data Quality Dimensions (Wand & Wang, 1996)

Wang & Strong

Another popular proponent in the same year is the work of Wang & Strong, (1996). This research work employed 179 data quality dimensions, selects 15 dimensions and grouped them into 4 distinct categories. The aim is to develop an empirical study capable of collecting data quality dimensions from data consumers, fix important ratings and there after establish a hierarchical structure for data quality dimension.

Data quality was defined as “data fit for use by data consumers”. A quality dimensions is defined as a set of data quality attributes which represents a single aspect or construct of data quality. In this research, data is treated as a product. With this background, a two-stage survey with two-phase sorting was conducted, and this led to the development of the hierarchical framework for organizing data quality dimension.

The outcome of the research defined “Intrinsic data quality (DQ)” to imply that, data have quality in their own right. C contextual DQ reflects the requirements that data quality must put into consideration within the context of the task. Representational data quality includes aspects related to the format of the data, it’s meaning amongst other. The last of these is the “Accessibility DQ”. This factor is closely related to above mentioned, and it basically emphasize the importance of the role of the system (Wang & Strong, 1996). Figure 2.3 below give the conceptual framework developed for this research.

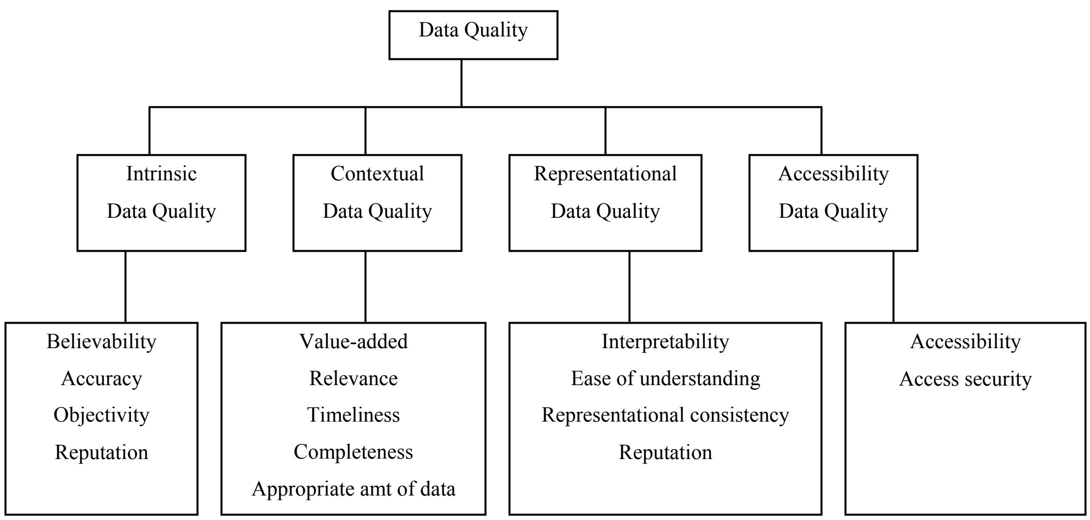


Figure 2.3: Conceptual Framework of Data Quality (Wang & Strong, 1996)

This framework enjoyed wide citation and recommendations. It also formed the basis for some other empirically oriented researches which stemmed up thereafter.

Redman & others

The multiplicity of the dimensions has always been a bone of contention amongst data quality researchers. While some of the attributes are repeated and new ones included, some only redefine existing dimension to fit their work. Cykana, et al., (1996) proposed six characteristics of data quality, Matsumura & Shouraboura, (1996) also work with two categories and four attributes. The work of Miller, (1996) was based on ten dimensions and Redman, (1996) has three categories and twenty-seven dimensions. The work is an intuitive grouping which resulted into data quality dimensions into three distinct views, these include: Conceptual (6 dimensions); Values (4 dimensions) and Representational (8 dimensions). It should be noted that, each dimension may also include sub-dimension for richness of definition. Table 2.3 gives the summary of the set of data quality dimensions proposed by Redman, (1996).

Information Quality Category	**Information Quality Dimension**
Conceptual View	Content, Scope, Level of details, Composition, View Consistency, Reaction to change
Data View	Accuracy, Completeness, Currency, Value Consistency
Data Representation	Appropriateness, Interpretability, Portability, Format precision, Format flexibility, Ability to represent null values, Efficient usage of recording media, Representation consistency

Table 2.3: The 3 Views & 27 dimension of Data Quality by Redman (1996)

Other sample definitions considered were Gardyn, (1997) with five dimensions from the research titled "A Data Quality Handbook for a Data Warehouse" and Abate, et al., (1998) in "A Hierarchical Approach to Improving Data Quality" having four categories and 15 dimensions.

Jarke, Dedeke, & others

The year 1999 witnessed a new incursion into data/information quality. Jarke, (1999) proposed a project aimed at guiding data warehouse design activities. This resulted into specifying data quality dimension tailored for data warehouse environment, and dimensions classified according to the roles of users in the same environment. The author came up with four distinct categories and twenty-three dimensions. Table 2.4 gives the summary of the proposition.

Along the same vein, Dedeke, (2000) proposed 5 dimensions and 28 categories. These were used to develop a conceptual framework aimed at developing quality measures for information systems.

CATEGORIES	DIMENSIONS
Design & Administration Quality	Correctness, Completeness, Minimality, Traceability, Interpretability, Metadata Evolution
Software Implementation Quality	Functionality, Reliability, Usability, Software efficiency, Maintainability, Portability (ISO 9126)
Data Usage Quality	Accessibility, Availability, Security, Usefulness, Interpretability
Data Stored Quality	Completeness, Credibility, Accuracy, Consistency, Data Interpretability

Table 2.4: Sets of Dimensions by Jarke (1999)

Major contributors in the year 2001 are the works of Eppler, presented with four quality levels and sixteen criteria, Bovee, et al., (2001) also with four criteria and ten components. Data quality concept of "fitness for use" was the focus of Bovee, et al., (2001) the four dimensions proposed comes with other sub-dimensions. These include: Accessibility – ability to get information, Interpretability – being able to understand it, Relevance – being applicable to a specific domain and purpose of interest and Credibility – that is, belief to be credible. Naumann, (2002) proposed four dimensions and 22 metrics/sub-dimensions in the paper titled "Quality-Driven Query Answering for Integrated Information Systems".

Burgess, et al

Similarly, Burgess, et al., (2002) made frantic effort to harmonize the various propositions geared at establishing taxonomy of quality appropriate for information filtering. In their work, three main dimensions were proposed to engulfed the earlier proposition based on user-centric information filtering. This was arrived at from a total of thirty-six widely used definitions and a hundred a seventy-nine criteria. These dimensions are cost, utility and time. They were further broken down into eight sub-dimensions and these also were broken further. One confusing point in this taxonomy is the possibility of a dimension like “Time” reoccurring as a sub-dimension under another dimension. It thus implies that there is possibility of circular relationship between the stated dimension and the sub-dimensions.

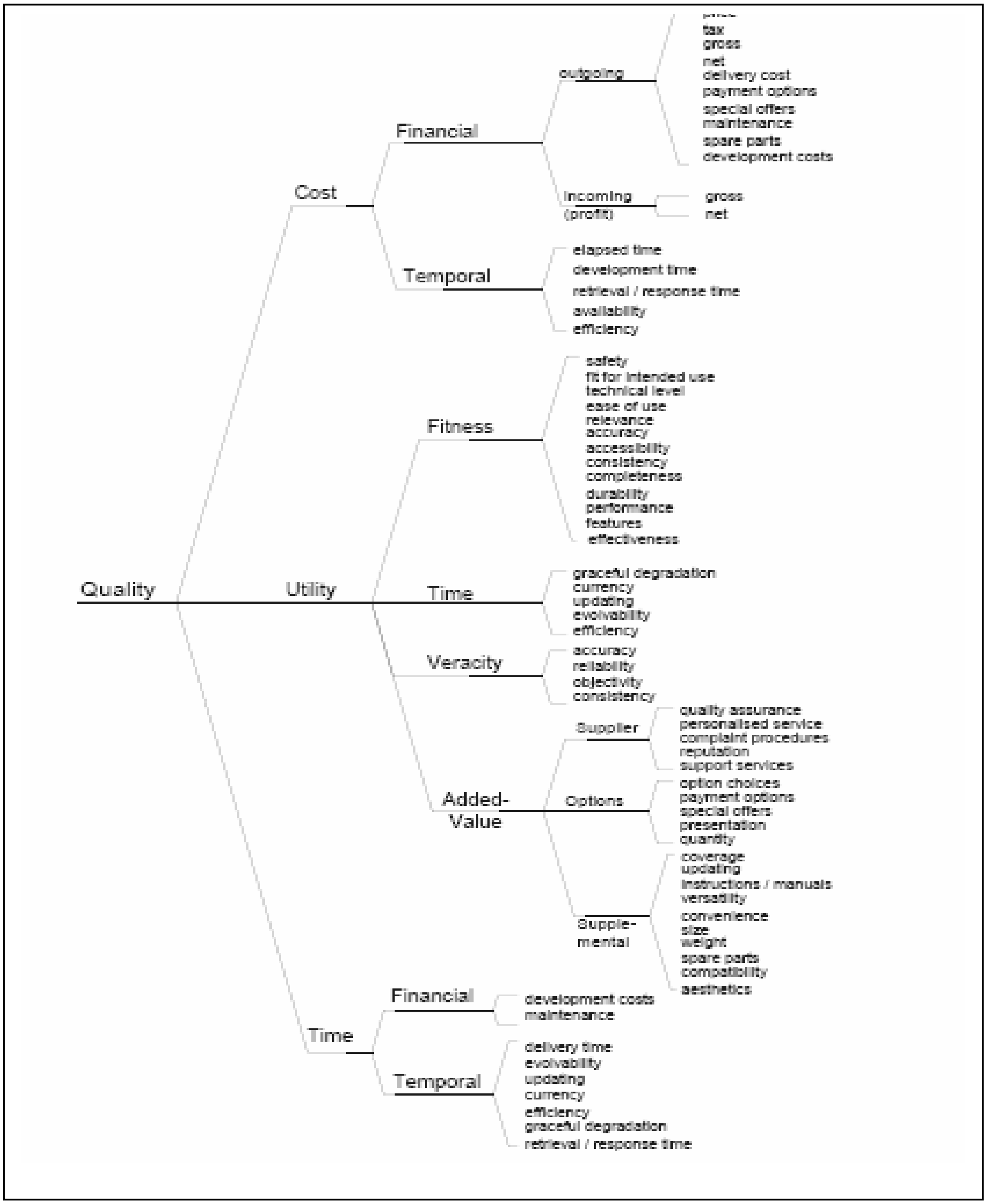
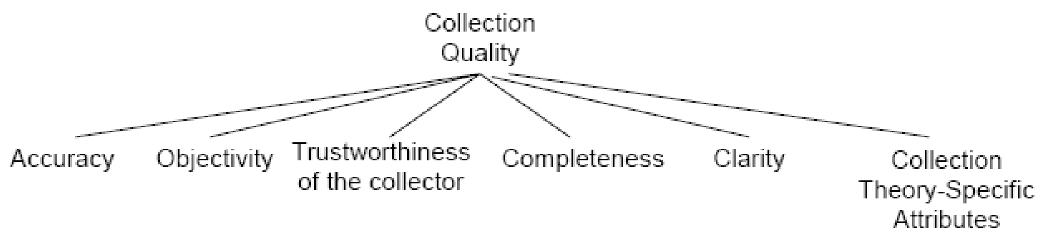
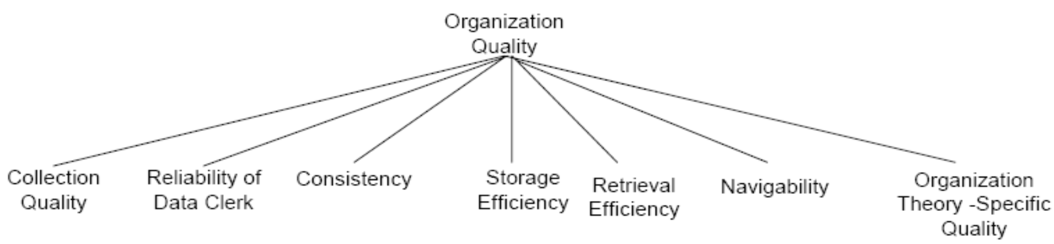


Figure 2.9: Taxonomy of Information Quality (Burgess, et al., 2002)

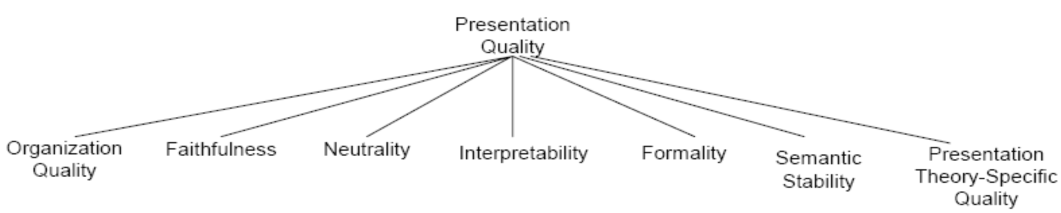
The argument of this work was based on the earlier notion and proposition of "High Quality = Data that is fit for purpose", this was countered by another notion that "fit for purpose today ≠ fit for purpose tomorrow" thereby revealing the importance of time in their proposition. Above in figure 2.9, we presented the major contribution of Burgess, et al., (2002)



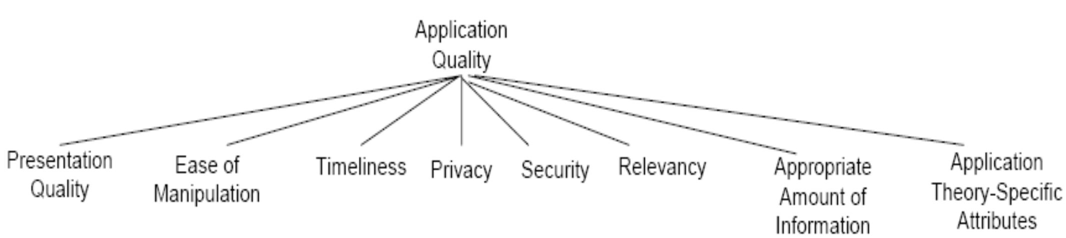
Collection Quality



Organization Quality



Presentation Quality



Application Quality

Figure 2.10: Evolution of Theory Specific Approach to Data Quality (Lui & Chi, 2002)

Liu & Chi

Another popularly referred work of the year is the proposition of Liu & Chi, (2002). The work follows from the observed weaknesses in the product analogy based approach and the narrowness of the ontological approach proposed in Wand & Wang (1996).

Figure 2.10 gives the summary of the propositions of Lui & Chi (2002). The authors maintained that, data evolve through stages of data evolution life cycle (DELC). This involves series of transformation which independently introduce some type of errors in the proposed development life cycle. Four main themes were identified: collection quality, organization quality, presentation quality and utilization quality. These were further sub-divided as shown in figure 2.10 above. Pipino, et al., (2002) presented an overview of popular dimensions and their definitions to describe the subjective and objective assessment of data quality. This led to the presentation of three functional forms for developing objective data quality metrics to give a view of how data quality is employed in practice. Similar contribution was made in "Evolutional Data Quality- A theory specific" where Data evolution was defined in four stages: Collection; Organization; presentation and Application. These stages were used to define data quality (Lui & Chi, 2002). It was shown that DQ is both multidimensional and hierarchical. In their submission, they stated that data are reflections of real world objects through a theory designating the sets of modules, methods, techniques, approaches and heuristics employed for the data collection, organization, presentation and application.

2.4.2 Juxtaposition of various Views

Several definitions of information presented by different fields were outlined in Bovee, et al., (2003). The authors however stressed that, quality is usually inclined from the point of view of user-centric and product-oriented. The 2 perspective makes information quality relative to the domain of use (David & David, 2001). They presented definitional approach to quality (Intrinsically or Extrinsically), as well as Model of Information (theoretically, system or process output, or product). Shankaranarayanan, et al., (2003) opined that the process of gauging and managing DQ was parallel with manufacturing by tagging data as "Information product (IP)". The authors claimed that the IP approach facilitates a comprehensive, intuitive and visual representation of the manufacture of an

IP, the model was thus tagged IP-MAP. It allows for visualization of the wide distribution of data and other resources, and also the flow of data elements and the sequence of process towards the creation of required IP in the process. It affords the decision maker the ability to understand the resources; processes; systems; business units and the organizations involved.

Eppler, et al., (2004) evaluate the effect of organizational, technological and legal perspectives on information quality. The plethora of information problems was said to include misinformation, information overload, paralysis by analysis, wrong decision, scrap and re-work or distrust. It further consider the dimension of Wang, et al., (1996), IQ as multidimensional construct and a more balanced approach taking into cognizance the comprehensive management of product, process cum the technical and social issues. A formal definition of information quality problems was the focus of Oliveira, et al., (2005). The author believed that this approach will assure clear and precise definition for each DQ problem. It is useful in specifying what is required to detect automatically the problem i.e. metadata knowledge needed; mathematical expression defining the DQ problem and the function that is required to perform some transformation. Adir & Shankaranarayanan, (2005) stated that existing efforts like: treating data as product; capturing metadata to manage data quality; statistical techniques; source calculus and algebra; data stewardship and dimensional gap analysis all failed in inculcating the contextual factors. It was thus opined that, once the concept of contextual perspective is generally accepted, there might be need to re-evaluate the current data quality assessment methods. The authors thereafter differentiate between impartial (objective) and contextual (subjective) data quality.

Lima, et al., (2006) focused on the thematic and methodological aspects, the author reviewed 170 articles and classified them based on 5-major classifications. The result was thereafter used to develop a framework that falls into 3 views of research: Organizational; Behavioural and Operational.

The disparity in the usage and adoption of data quality attribute has continued to contribute to the inability to have a uniform model. While some proposition had gained popularity, others continue to emerge while several authors define and re-define existing dimension and attributes. Below we shall take a look at the way and manner by which several reference was made to similar attribute making it appear as if we are in circular mode. In Bovee, et al., (2001) quality is implied from the point of user-centric or product-oriented view, several references were also made to the notion that, quality of information is regarded as a global assessment of its fitness for use in (Bovee. et. al. 2001) and (Burgess. et. al. 2002, 2004, 2007). Several dimensions therefore exit attempting to define what information quality connotes with little or no success as they mostly continue to use existing dimensions to define the new ones. With this situation, we shall adopt the submission of Gackowski, (2005 & 2006) where a proposition was made on operation approach to quality within the context of decision making. The result of these was a hierarchical result-oriented taxonomy dividing the quality attributes into direct and indirect ones which were directly mapped into primary and secondary ones. Their operational completeness was thus based on either being effective or economically effective. With this notion, five, "first time" or initial evaluations of operation quality of data/information was proposed as it relates to their usability in the context of *Direct Informing* and these include:

- Interpretable during acquisition,
- Of significance impact,
- Operationally timely available,
- Actionably credible and
- Task-specific effectively operationally complete

The above present a strict way (hierarchy) of evaluating the quality of data/information. The order of examination shows their importance and once a level cannot be adequately satisfied, it was considered irrelevant to continue with the evaluation process. The above, but for the second factor was again used to determine the economic nature of the secondary attributes.

In another dimension referred to as repetitive routine operation Gackowski, (2005 & 2006), asserts that data/information quality present different perspectives to decision makers and other users of information values. Mostly in this scenario, information technology is employed to store useful data items in common databases and/or data warehouses for user's access. The above is referred to as *Indirect Informing*. In a situation like this, *data/information acquisition, entry, verification, validation, storing, making them available and treatment into presentable and interpretable format* process is usually undertaking by professionals (watchers in the context of EI). With this arrangement, testing for "actionable credibility" (i.e. how truthful, reliable, and consistent it is. Actionable credibility is dependent on whether the data have valid information, misinformation, or disinformation. It extends further to include the credibility of sources, quality of mapping within the delivery system (data warehouse), and the credibility of its presentation when we are talking about indirect informing). The above operation is not performed directly by the user or clients. *This separation between the "acquisitors" and data users (involving the use of complex data and information delivery systems) can result into a scenario where data/information considered to be interpretable during acquisition is no longer interpretable to a decision maker or users not present during the process of acquisition or that the product of search was not actually what the decision maker expects -misconception*. The above could be sequel to *different mindsets, languages, cultural backgrounds, conventions, experience and other cognitive ability*. For example, if the convention of writing names in country-W is [Surname, other-names], then for another country whose convention is [other-names, Surname], it implies that every information search based on either or combination of the following may not match the desired intention. This is because "Amos David" ≠ "David Amos" in all cases. The above leads to another factor referred to as *Presentation Interpretability*: it can either be *operationally presentation interpretable* or *economically presentation interpretable.*

<u>Operationally presentation interpretable</u> – this is viewed from the perspective of indirect informing, where data/information acquisition is separated from the users or decision makers. However, presented data/information must be interpretable and understandable e.g. legibility; of preferred language; measurement units; conventions e.t.c. This factor

thus becomes another cogent one in any organizational circumstance and its thus tagged primary factor but not universal.

Economically presentation interpretable – this factor which is closely tagged with its form, format and mode of delivery determines the rate at which the user or decision maker *read, perceive, interpret, comprehend, analyze, absorb, infer conclusions, react and finally act upon it*. Factors of importance here include clarity, consistency, order, media used, level of summarization, user preferred form of presentation such as text, graph, diagram, picture, esthetics e.t.c.

The operational quality of data/information presented above is of huge importance in the operation of any organization. The two broad divisions of direct informing and indirect informing adequately represent the position and characteristics of data/information, the users or the decision maker, and the various processes amongst them. However, the distinction constitutes serious risk factors e.g. misinforming, disinforming, and other root of biases (based on direct and indirect informing characteristics) in organizations employing information technologies despite the professionalism of the people involved.

As a form of rap-up, the submissions of Tejay, et al., (2006) shall serve as our guide. It gives a summarized format for most of the considered quality dimensions and deduced a semiotic analysis of the concept to arrive at four levels: empirics, syntactic, semantics and pragmatics as shown in the table 2.5 below.

Semiotic levels	DQ Dimension	Work
Empirics	Accessibility	Delone, et al., (1992), Goodhue, (1995), Miller, (1996), Wang, et al., (1996), Bovee, (2001)
	Timeliness	Ballou, et al., (1985), Caby, et al., (1995), Fox, et al., (1994), Goodhue, (1995), Hilton, (1979), Miller, (1996), Wang, et al., (1996), Zmud, (1978), Wand & Wang, (1996)
	Locatability	Goodhue, (1995)
	Portability	Caby, et al., (1995)
	Security	Miller, (1996), Wang, et al., (1996)
Syntactic	- Accuracy	Ballou, et al., (1996), Wang, et al., (1996), Caby, et al., (1995), Fox, et al., (1994), Goodhue, (1995), Hilton, (1979), Miller, (1996), Zmud, (1978), Delone,, et al., (1992), Doernberg, et al., (1980), Norman, (2002)
	- Appearance , Comparability, Freedom from bias, Precision, Redundancy, Uniqueness, Usable	Delone, et al., (1992)
	- Arrangement, Readable	Zmud, (1978)
	- Clarity, Ease of use, Presentation	Goodhue, (1995)
	- Coherence, Format	Miller, (1996), Redman, (1996)
	- Compatibility	Goodhue, (1995), Miller, (1996)
	- Composition	Caby, et al., (1995)
	- Flexibility, Robustness, Conciseness	Delone, et al., (1992), Wang, et al., (1996)
	- Consistency	Ballou, et al., (1985), Caby, et al., (1995), Fox, et al., (1994), Wang, et al., (1996), Wand, et al., (1996)
	- Correctness	Wang, et al., (1996)
	- Ease of operation, Objectivity	Brodie, (1980)
	- Integrity	Caby, et al., (1995), Goodhue, et al., (1995)
	- Level of detail	
Semantics	- Ambiguity	Doernberg, et al., (1980), Wand, et al., (1996)
	- Believability, Understandabilty	Wang, et al., (1996)
	- Content, Informativeness	Delone, et al., (1992)
	- Factual, Reasonable	Zmud, (1978)
	- Interpretability	Wang, et al., (1996), Caby, et al., (1995), Bovee, (2001)
	- Meaningful	Goodhue, (1995), Wand, et al., (1996)
	- Reliability	Brodie, (1980), Delone, et al., (1992), Goodhue, (1995), Zmud, (1978)
	- Validity	Miller, (1996)
Pragmatics	- Appropriate amount of data, Reputation, Value-added	Wang, et al., (1996)
	- Appropriateness	Caby, et al., (1995)
	- Completeness	Ballou, et al., (1985), Caby, et al., (1995), Fox, et al., (1994), Miller, (1996), Wang, et al., (1996), Wang, et al., (1996),

	- Relevance - Importance, Sufficiency, Usefulness	Doernberg, et al., (1980), Norman, (2002) Delone, et al., (1992), Hilton, (1979), Miller, (1996), Wang, et al., (1996), Bovee, (2001), Norman, (2002) Delone, et al., (1992)

Table 2.5: Semiotic Analysis of Data Quality Dimension (Adapted from Tejay, et al., 2006)

Table 2.5 is a presentation of what is referred to as the data quality dimension, i.e. set of data quality attributes that represents a single aspect or construct of data quality (Wang & Strong, 1996). Data quality problems range from its definition, measurement, analysis, and improvement to tools, methods and processes (Wand, et al., 2001). Tejay, et al., (2006) defined semiotic interpretation of data quality dimensions as attempt to address the definition, measurement and analysis aspect of data quality. Alongside, the improvement aspect is implicitly mentioned. One important fact stressed was that, dealing with quality attributes such as metrics will ultimately lead away from the main goal into the field of networking, thus the need for the semiotic study. Semiotic broadens the understanding of the interdependencies amongst data, information and knowledge vis a vis data quality.

The semiotic analysis depict that the pragmatic level is associated with knowledge, semantic level is associated with information while only the syntactic level is associated with data (Tejay, et al., 2006). The diagram below further buttresses this submission.

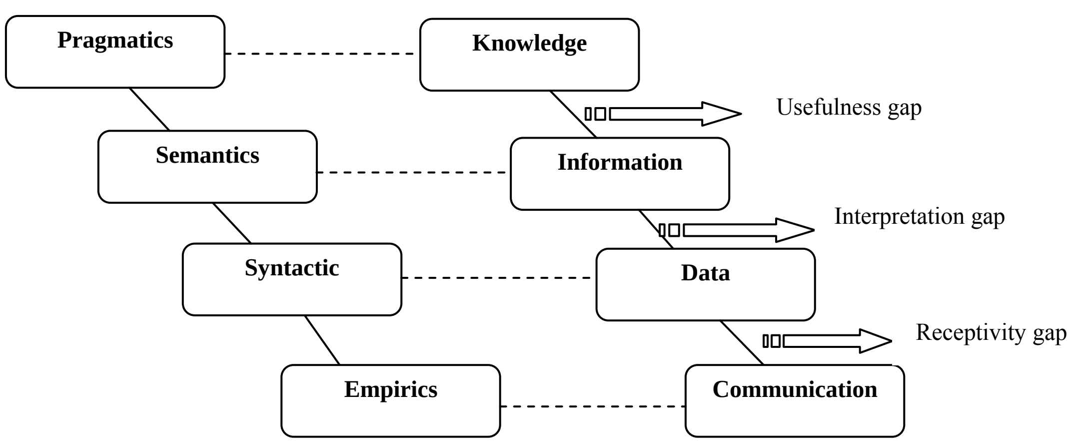


Figure 2.11: Semiotics, data-information-knowledge and the gap (Tejay, et al., 2006)

Figure 2.11 thus depict that the dimensions operating at the pragmatic, semantic and the syntactic level pertains to knowledge quality, information quality and data quality respectively. Consequently, attempt to improve the data quality will focus attention on dimensions operating at the syntactic level. However, it is almost impossible to neglect dimensions associated with knowledge quality and information quality. The trio need be properly harnessed to arrive at a laudable conclusion.

In this section, we have evaluated the contributions of major authors on the concept of data/information quality. While various attributes and dimensions exit, their interdependencies in usage and interpretation need be scrutinized to facilitate proper understanding towards universally accepted dimensions. It therefore behooves that there is an urgent need for a framework that integrate these active components (data/information, users or clients, and processes) with a bid to determine and arrest possible risk factors before they result into risk in the process of utilization. The semiotics level approach is what we shall adopt for our view about the differences in data quality dimensions. The next section will focus on the importance of data quality in the data warehouse concept. This is sequel to the fact that most information used for decision purposes end up in corporations' data warehouses. Thus, its provision for ensuring quality should be considered.

2.5 Data Quality in Data Warehouse Concepts

In this section, our focus is on data warehouse quality (DWQ). We shall take as a case study the work under the DWQ which is a cooperative project in the ESPRIT program of the European Communities. The transition and rationale for the establishment of data warehouse will be considered cum the reason for the enforcement of quality in a data warehousing environment.

Historically, the DWQ results from the observation of E.F. Codd and W. Immon which dated back to the early 90's. Their concern was based on stated fact that operational level in on-line transaction processing (OLTP) where current data were maintained in great detail in their immediate environment and decision support application (OLAP) which is concern with lightly aggregated and often globally reconciled historical data cannot coexist in the same database and still deliver the same efficiency. A Data Warehouse (DW) can be defined as a collection of technologies aimed at enabling the knowledge workers (executives, managers, analyst, and decision makers) to make better and faster decision. The onus is thus to present the right information in the right time with the right cost aimed at supporting the right decision (Jarke & Vassiliou, 1997). Operationally, the DW caches selected data of interest to a customer/user or group to facilitate easy and faster access in a cheaper and effective manner.

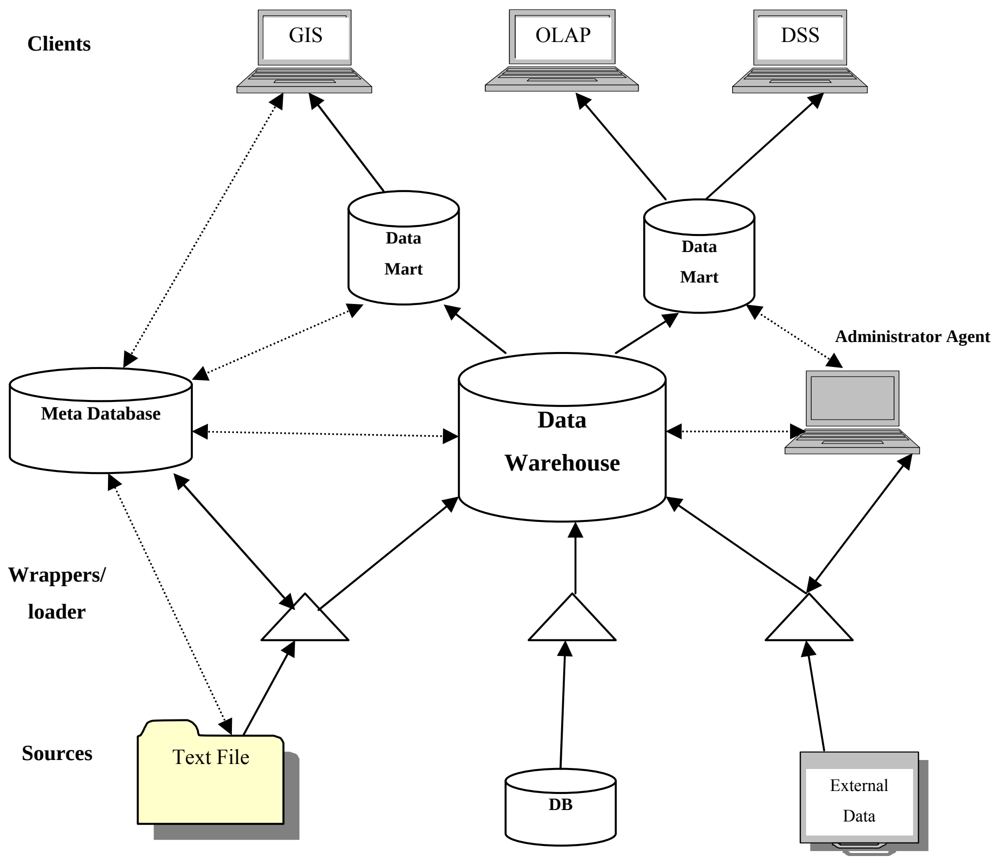


Figure 2.12: Sample structure of a data warehouse (Jarke & Vassiliou, 1997)

A data warehouse comprises of the following components: sources – any data store whose content is subject to be materialized in a data warehouse, wrappers – to load the source data into the warehouse, destination databases – this consist of data warehouses and data marts, meta database – serves as repository for information about other components (the schema of the source data), agent for administration which include the data warehouse design, scheduler for update initiation e.t.c and the client which displays the data in a desired mode e.g. statistical package. The above components' interaction is shown diagrammatically in figure 2.12 above. (Refer figure 2 in Jarke & Vassiliou, 1997)

The data warehouse quality management presents the interaction between the quality of goals and warehouse stakeholders (Vassiliadis, 2000). The doctoral research featured the need to organize the design, administration and the evolution of the data warehouse in a manner that different and sometimes opposing quality requests from a user can be simultaneously resolved/attended. Vassiliadis, et al., (2000) highlights the interrelationship between warehouse object and process and quality metrics. Five different user categories were enumerated with different functions. They are discussed based on the issue versus object with several examples. This factor allows further insight into the design and administrative quality (DAQ). DAQ involves

1. schema quality (correctness, completeness, minimability, traceability, interoperability)
2. metadata evolution

Other factors considered include: Data loading quality (analyzability, transactional availability), Data usage quality (accessibility, usefulness), Data quality (completeness, credibility, accuracy, consistency, data interpretability).

Considering the data warehouse evolution, the complex nature of the data warehouse causes the components to frequently evolve independent of each other (Tayi & Ballou, 1998). There is possibility of new materialized views creation and old ones updated. Some earlier data sources may be out of use, new sources may be added. There could be product evolution or update, changes in design choices is also a possibility. The qualities of a data warehouse therefore depend on the quality of the sources; the quality of the extraction process and the quality of the data warehouse components (Gatziu, et al., 1999).

Data quality from the perspective of Jarke & Vassiliou, (1997) consist of the organizational structure, responsibilities, procedures, processes and resources for implementing data quality management, the following were of particular interest: Interoperability (syntax, version control, semantics aliases e.t.c), Usefulness (relevance to data warehouse, timeliness e.t.c), Accessibility (system availability, transaction availability, privileges), Believability (completeness, consistency, credibility, accuracy). The above mentioned were considered under the notion of data quality in data warehouse. A diagrammatic representation of the above is given in figure 2.13 as presented in Jarke & Vassiliou, (1997).

Quality factors like accessibility and timeliness, believability and understandability, design and usage-flexibility was mentioned to be important in considering the success of any data warehouse (Gebhardt, et al., 1998). According to Raisinghani, (1999) the rationale is the fact that heterogeneous sources feeding a major source will usually present similar data stored in different formats.

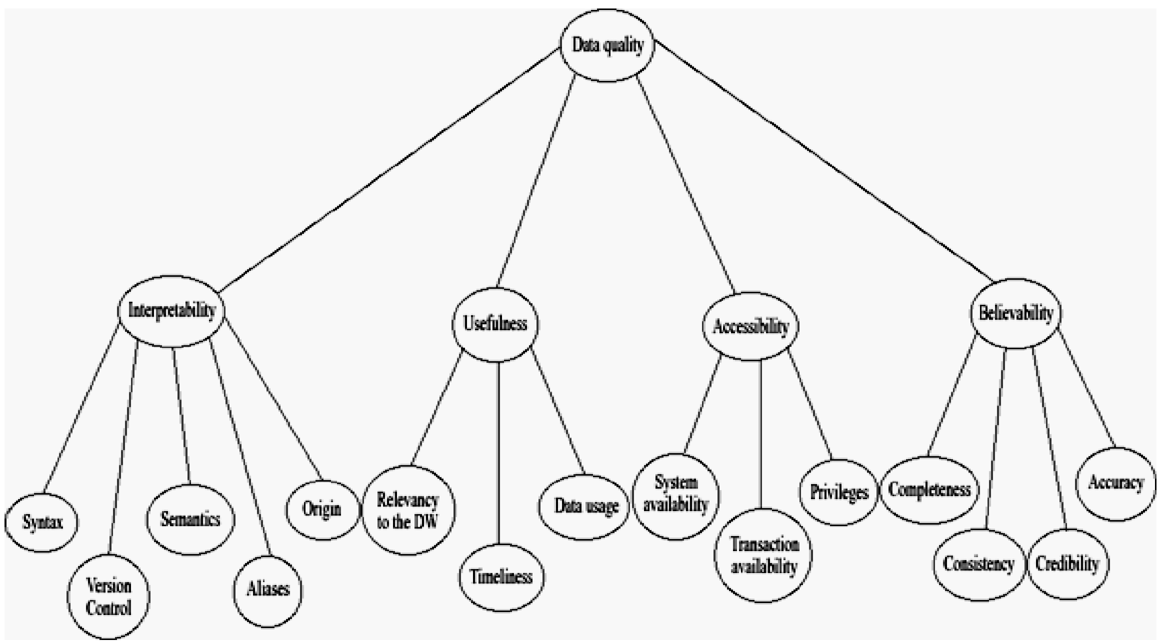


Figure 2.13: Quality Factors in Data Warehousing (Jarke & Vassiliou, 1997)

This fact thus requires the process of cleaning on the data, these involve: integration; transformation and cleaning. Attempt was made to classify errors inherent in the process of cleaning. Listed amongst these are: incomplete data; incorrect entries;

incomprehensible entries, and inconsistent entries. The author also dived into issues like Naming Conflicts: homonyms and synonyms, Structural Conflicts: these include type conflicts; dependency conflicts; key conflicts and behavioural conflicts. With these arrays of inconsistencies, data cleaning operation is imperative in the maintenance of a data warehouse (Raisinghani, 1999).

2.6 Taxonomy of Dirty Data

Prior to this section, we have dealt with various submissions based on the taxonomies of data and information quality. We also extended the review to include quality issues in the warehouse arena. It is important to note that by consolidating data from disparate sources into a "central" position (warehouse) facilitates running of data analysis across application to obtain information that are strategic and tactical towards taking cogent decisions (Inmon, 1999). It is however unfortunate that most of the data kept in the data warehouses for strategic decisions are 'dirty'. By dirty data we imply that data is either missing or wrong, or it is in a non-standard representation (Williams, 1997). The concept of missing data is encountered in every information retrieval system in existence today. Generally, as the complexity and size of data increases, the issue of missing data becomes expedient.

There are several reasons why the data may be said to be missing the reasons stem from both human related faults and machine malfunctions. Examples are: data may be missing because equipment malfunctioned, the weather was terrible, people got sick, or the data were not entered correctly.

Missing Completely at Random

In the above cases, the data are said to be *missing completely at random (MCAR).* When we say that data are missing completely at random, we mean that the probability that an observation (X_i) is missing is unrelated to the value of X_i or to the value of any other variables. Thus data on family income would *not* be considered MCAR if people with low incomes were less likely to report their family income than people with higher incomes.

Similarly, if Whites were more likely to omit reporting income than African Americans, we again would not have data that were MCAR because missingness would be correlated with ethnicity. However if a participant's data were missing because he was stopped for a traffic violation and missed the data collection session, his data would presumably be missing completely at random. Another way to think of MCAR is to note that in that case any piece of data is just as likely to be missing as any other piece of data (Dunning, & Freedman, 2008).

Notice that it is the value of the observation, and not its "missingness," that is important in this regard. If people who refused to report personal income were also likely to refuse to report family income, the data could still be considered MCAR, so long as neither of these had any relation to the income value itself. This is an important consideration, because when a data set consists of responses to several survey instruments, someone who did not complete the Beck Depression Inventory would be missing all BDI subscores, but that would not affect whether the data can be classed as MCAR. This nice feature of data that are MCAR is that the analysis remains unbiased. We may lose power for our design, but the estimated parameters are not biased by the absence of data.

Missing at Random

Another dimension considered by Howell, (2009) known as *missing at random*. Often data are not missing completely at random, but they may be classifiable as missing at random (MAR). For data to be missing completely at random, the probability that Xi is missing is unrelated to the value of Xi or other variables in the analysis. But the data can be considered as missing at random if the data meet the requirement that missingness does not depend on the value of Xi after controlling for another variable.

Missing Not at Random (MNAR)

If data are not missing at random or completely at random then they are classed as *Missing Not at Random (MNAR).* For example, if we are studying mental health and people who have been diagnosed as depressed are less likely than others to report their mental status, the data are not missing at random. Clearly the mean mental status score for the available data will not be an unbiased estimate of the mean that we would have obtained with complete data. The same thing happens when people with low income are less likely to report their income on a data collection form (Dunning, & Freedman, 2008).

When we have data that are MNAR then, the problem is significant. The only way to obtain an unbiased estimate of parameters is to model missingness. In other words we would need to write a model that accounts for the missing data. That model could then be incorporated into a more complex model for estimating missing values. This is not a task anyone would take on lightly.

Kim, et al., (2003) provide hierarchical refinement approach to arrive at their taxonomy of dirty data. Their taxonomy was based on the premise that dirty data manifest itself either as missing data, not-missing but wrong, and not-missing and not wrong but unusable. The hierarchy decompose these manifestations and also represents dirty data resulting from more than one type of dirty data. Below we present a comprehensive listing of their taxonomy.

- Missing data
    - Missing data where there is no Null-not-allowed constraint
    - Missing data where Null-not-allowed constraint should be enforced
- Not-missing, but
    - Wrong data, due to
        - Non-enforcement of automatically enforceable integrity constraints
            - Integrity constraints supported in relational database systems today
                - User-specificable constraints
                    - Use of wrong data type (violating data type constraint, including value range)
                    - Dangling data (violating referential integrity)
                    - Duplicated data (violating non-null uniqueness constraint)
                    - Mutually inconsistent data (action not triggered upon a condition taking place)

            - Integrity guaranteed through transaction management
                - Lost update (due to lack of concurrency control)
                - Dirty read (due to lack of concurrency control)
                - Unrepeatable read (due to lack of concurrency control)
                - Lost transaction (due to lack of proper crash recovery)
        - Integrity constraints not supported in relational database systems today
            - Wrong categorical data (e.g., wrong abstraction level, out of category range data)
            - Outdated temporal data (violating temporal valid time constraint; e.g., a person's age or salary not having been updated)
            - Inconsistent spatial data (violating spatial constraint; e.g., incomplete shape)
    - Non-enforceability of integrity constraints
        - Data entry error involving a single table/file
            - Data entry error involving a single field
                - Erroneous entry (e.g., age mistyped as 26 instead of 25)
                - Misspelling (e.g., principle instead of principal, effect instead of affect)
                - Extraneous data (e.g., name and title, instead of just the name)
            - Data entry error involving multiple fields
                - Entry into wrong fields (e.g., address in the name field)
                - Wrong derived-field data (due to error in functions for computing data in a derived field)

            - Inconsistency across multiple tables/files (e.g., the number of Employees in the Employee table and the number of employees in the department table do not match)

- Not wrong, but unusable data
    - Different data for the same entity across multiple databases (e.g., different salary data for the same person in two different tables or two different databases)
    - Ambiguous data, due to
        - Use of abbreviation (Dr. for doctor or drive)
        - Incomplete context (homonyms; and Miami, of Ohio or Florida)
    - Non-standard conforming data, due to
        - Different representations of non-compound data
            - Algorithmic transformation is not possible
                - Abbreviation (ste for suite, hwy for highway)

There has been however various methods for treatment of dirty data (Dunning, & Freedman, 2008, Inmon, 1999, Kim, et al., 2003). Among these are commercial software tools for creating data warehouses or transforming data for multidimensional analysis or data mining with several ways to replace missing data in a field with mean arithmetic values. The effectiveness of this 'guess' could sometimes be detrimental. Incomplete text, use of abbreviation and other forms of missing data can be handled in a better manner as we shall see later on in this write up. In the next section, we take a look at the operations of information retrieval model and tools, popularly referred to as search engines on the web platform that perform the retrieval and cross analysis of data towards delivering strategic decisions.

2.7 Effect of Data & Information Quality on Decision Making

"Not All Data are Equally Created" was the submission of Even & Shankaranarayanan, (2006) when examining the accruable benefits from complex data resources believed to have been obtained at a cost which is difficult to evaluate. Most researches cannot adequately quantify the value of information used for a particular decision directly, but subjectively based on the outcome of the decision made from the information. Intuitively, almost everyone knows what the word ‘quality’ connotes, however attempt to define it usually result into struggle (Burgess, et al., 2006). Quality has been defined and employed in different environment: software quality, data quality, information quality, and web quality. The submission was that, although there are multitudes of research leading to the creation of divergent definition of quality, “no single definition or standard of quality exists” (Smart, 2002).

Potential capital losses and heightened exposure are inherent in the usage of poor data quality management (Even & Shankaranarayanan, (2005). Existing efforts like treating data as product; capturing metadata to manage data quality; statistical techniques; source calculus and algebra; data stewardship and dimensional gap analysis all failed in inculcating the contextual factors (Pipino, et al.). It was thus opined that, once the concept of contextual perspective is generally accepted, there might be need to re-evaluate the current data quality assessment methods. Conducting review research on the impact of data quality on decision performance, Jung, (2004) evaluate: contextual, representational and accessibility of data quality and their influences on decision making. The rationale is such that high quality decision is based on access to information which is complete and relevant to the scope under consideration. Distinction was made amongst data, information and knowledge, from which the relationship between data quality and decision making was established. Contextually, the requirement is that data quality must be viewed in the context of the task at hand, i.e. data must be relevant, timely, complete, and appropriate in terms of amount to be able to add value. Representational and Accessibility data quality implies the importance of the role of the system i.e. the system must be accessible but secured, and present data in a way that they are interpretable, easy to understand, and represented concisely and consistently.

It has been founded that major problem faced by a decision maker are based on how best to improve consumer satisfaction, lower high cost, and complete ongoing projects amongst others. In coping with the above, inappropriate data quality will spell doom for the enterprise by causing: consumers' dissatisfaction; less effective decision making; difficulties in executive strategies. It can also lead to disrupt in employees morale, organizational mistrust and makes it difficult to align the enterprise (Redman, 1998).

The value of information was given as the difference between the benefit in the presence of full-information versus the possibly reduced benefit in the presence of partial or no-information. This value only materializes with usage and experience. It was therefore established that, the value is greatly influenced by the design of information products and the IS/IT environments managing them (Even & Shankaranarayanan, 2005). Judging from the fact that value is defined based on contextual usage of data, the focus is more on the context rather than the structure, again, values contributed varies also with usage. The above thus presents Context driven view of data management versus structural view (model characteristics, field structure, and record counts e.t.c.). The evaluation features the Lorentz & Gini coefficient statistical tool to demonstrate the possible implication of utility inequality for data management decision via the examination of the effect on utility-cost trade-off and the contribution to the net benefit.

Organizational performance was viewed from operational risk perspective by Raneses, et al., (2006). The authors defined operational value at risk OPVAR – as how much is expected to be lost if an event in the tail of the loss probability distribution does not occur. The author lament the inability of organizations to distinguish between Internal risks (operational risks) and the External risks (regulations, credit and market controls). OPVAR is not the optimal measure of risks because aggregating individual risks does not increase the overall risk, or is not a coherent risk measure. The alternative to OPVAR was expected shortfall (ES), which are the average value of losses that can be expected if a loss in excess of OPVAR is observed. The rationale was based on the usage of inappropriate data for decision making. Ge & Helfert (2006) also research into the relationship between information quality and decision quality. While the authors maintained that even in the presence of sterile information quality other factors like personal preferences, decision maker's experience, and operating environment amongst other factors need be put into consideration (Gackowski, 2006). The above did not delimit the importance of organization's inability to determine the effect and relationship between decision quality and IQ dimension.

It is a common understanding that creating awareness about a problem and its impact is critically the first step in resolving the problem (Kotter, 1996). The above possibly amongst other factors necessitated the need for various evaluation processes to determine the impact, importance and possibly the consequences of non-quality data in decision processes. In this section, we have reviewed the submission based on the views about the importance of data quality on organizational ability to possess and deliver competitive advantages. This impact consist in operational, typical and strategic impact in the view of Redman, (1995, 1996), operational risk, Raneses, et al., (2006), Bias, Misinforming and Disinforming, Gackowski, (2005, 2006). Poor data quality compromises decision making and in another dimension, it makes implementation of data warehouse aimed at facilitating better decision making cumbersome (Celko, 1995). Hammer & Stanton (1995) consider the effect from the effect it has on the reengineering process, while mistrust can also ensues amongst the workers. Thus making available appropriate and adequate information for decision making is a sine qua non to organizational performance. In all the need for this type of evaluation cannot be overemphasized, it therefore becomes imperative that the awareness in the data/information quality research should be encouraged to engulf a broader sense to be able to adequately inform on the importance of information quality on decision making.

2.8 Recap

The ability to perform exceptionally by a decision is dependent on the definition of the decision problem from which information retrieval process stems. While there are extensive progresses in database and data warehouse design, we believed that running data analysis on these information delivery system is a function of the user queries, the cleanliness of the data and the flexibility of the information retrieval tool. It is however unfortunate that defining user's information need is saddled with subjectivity and other factors that make the whole process vague, uncertain and incomplete. In this research, our approach supports the notion that *"data fit for use by a consumer is the quality of the* *data",* however, dirty data will either result in bad decision if it is not missing, or inability to take strategic decision by the decision maker if data is missing. Consequent upon this, the importance of adaptive information retrieval tool was stressed at the beginning of this chapter. In all, we believed that, the retrieval tools, ability to appropriately define the information need from the decision problem and the quality of available data constitute the bedrock of strategic decision making.

Accruable risk from non quality data can better be imagined than allowed to happen. An error of parallax in thermometric reading immediately renders the following judgments wrong. It is therefore important that, efforts at delivering strategic decisions, rest on one part on the available relevant information, the retrieval tools and the human being involved. The last on this list is human being, and this will be the focus of the next chapter. We shall explore theories of decision making and the possible risk in the face of data and uncertainty. Our focus in chapter three will be centered on decision from information, its models and the accruable risks.

Ontological imprecision *appears when the exact truth cannot be achieved, then the question is about imprecise object of the reality.* ***Epistemological imprecision*** *occurs in situations where a human being can not identify an object exactly; the reason may be for example foggy weather.* ***Linguistic imprecision*** *means that the verbal expression is imprecise and can have several meanings* (Niskanen, 1989, 1998).

Chapter 3
Fundamentals of Decision Theories & Risk

3.1 Overview

Decision making is essentially an integral and important factor of human life. Decision process involves consideration of available information about the outcome and choice amongst alternatives for subsequent action. The fact remains that, a good decision facilitates a good outcome. The gambling paradigm which can be traced to the early 17^{th} century studies in gambling has been widely supported as the predominant framework within decision making. Tacitly, the assumption is that any decision can be represented as a choice expressed as "gambles" with options depicted in terms of probabilities and utilities. *Decision here is seen as a choice of option(s) which offers the most advantage – rational decision making*. The unprecedented support gained by this paradigm was based on monetary gambles as stimulus (Passam, et al., 2003). Human decision making behaviour violates the fundamental axioms of economic models of choice in predictable ways. These violations were reflected on the decision maker's reliance on heuristics, biases and strategies deem to perform reasonably well in most cases but sometimes result into systematic errors (Dawes, 1998, Kahneman, et al., 1982).

Several other paradigm has thereafter being proposed each with different views but the major components expressed or highlighted were practically the same with very little variations. According to Karen & de Bruin (2003) most decision making are based on three main components: obtaining relevant information (from memory or external world), construction of the decision or problem space followed by attempt to fix the acquired information appropriately into the decision problem structure, and assessing the values and likelihoods of different outcomes. A drawback of the above is the blind omission of the methodology for achieving the first two stages. In the view of Barrette (2006), there are four major components to decision making. This author listed the following: data gathering, information processing, meaning making and finally, decision making in relation to the set goal. Bullen & Sacks (2003) reported several work on the various components of decision making and the complexities involved in the operations. Harris (1998) identified decision environment as the most singular factor from which other possible components stem. A decision environment was defined as a collection of information, alternatives, values and preferences available during the time of decision. Thus in an ideal environment for decision making, it's expected that all possible information will be available, in possible best accurate proportion, and alternatives. The above is however constrained based on the fact that time and efforts to gain information or identify possible alternatives are limited. Decision environment is presented to be of incremental rate, since new information and alternatives may spring up even after the decision might have been made. The identification of the above is followed by the attempt to formulate a response to the situation that fall in line with stated goals. The level of the decision maker's consciousness is brought into play at this point.

The notion of utility model of decision making implicitly assumes that a person's goal are important for determining their preferences (Kahneman & Tverskey, 2000). Several other demonstrations have followed to depict that people's choices do not follow the predictions of normative models. Consequently, concepts like heuristics and biases were introduced into decision making theories (Kahneman, et al., 1982, Kahneman & Tverskey, 1974, 1986,). Heuristic is considered as a fairly simple rule that can be employed in performing some cognitive task. The above is sequel to the fact that economic models of choice were deemed to be computationally expensive to achieve. The above require the decision maker to consider all available information and weights. Some of these models will still be considered later on, but in the meantime, we shall consider decision making and the influence of uncertainty.

In the rest of this chapter we discussed various decision models and their proponents. The essence is to articulate and later come up with a model to be employed for the research. We also present the concept of risk and risk factor, with understanding based on the provisions between hard and soft computing paradigm.

3.2 Decision Models & Theories

The gambling paradigm which can be traced to the early 17th century studies in gambling has been widely supported as the predominant framework within decision making. Tacitly, the assumption is that any decision can be represented as a choice expressed as "gambles" with options depicted in terms of probabilities and utilities. *Decision here is seen as a choice of option(s) which offers the most advantage – rational decision making*. Several other paradigm has thereafter being proposed each with different views but the major components expressed or highlighted were practically the same with very little variations. According to Karen & de Bruin (2003) most decision making are based on three main components: obtaining relevant information (from memory or external world), construction of the decision or problem space followed by attempt to fix the acquired information appropriately into the decision problem structure, and assessing the values and likelihoods of different outcomes. A drawback of the above is the blind omission of the methodology for achieving the first two stages. In the view of Barrette (2006), there are four major components to decision making. This author listed the following: data gathering, information processing, meaning making and finally, decision making in relation to the set goal. Bullen & Sacks (2003) reported several work on the various components of decision making and the complexities involved in the operations.

Harris (1998) identified decision environment as the most singular factor from which other possible components stem. A decision environment was defined as a collection of information, alternatives, values and preferences available during the time of decision. Thus in an ideal environment for decision making, it's expected that all possible information will be available, in possible best accurate proportion, and alternatives. The above is however constrained based on the fact that time and efforts to gain information or identify possible alternatives are limited. Decision environment is presented to be of incremental rate, since new information and alternatives may spring up even after the decision might have been made. The identification of the above is followed by the attempt to formulate a response to the situation that fall in line with stated goals. The level of the decision maker's consciousness is brought into play at this point.

Hunt's Decision Making Model

There has been sharp contrasting view on whether decision is "Analytical" or "Intuitive". Analytics are concerned with details while intuitive focus on patterns. Consequently, human instincts, subconscious belief, conscious belief, values and intuition have been identified as the major emphasis that determines the mode of decision making. Thus the differences displayed by the various modes are the reflections of the level to which the above mentioned are displayed Barrette (2006). Hunt et al. (1989) on the structure of decision identified concepts like the decision maker (DM) viewed as a stable person endowed with certain belief, predisposition, skills, and experience amongst others that describe his personality. The decision task (DT) calls for attention. Decision situation (DS) connotes contextual and ecological factors both conceptually and circumstantially. Others are the decision process (DP) and the decision outcome (DO). The focus of this study was to determine the relationship between DM's characteristics and the DP, given an ill defined DT. Figure 1.1 below summarizes the relationship of these concepts.

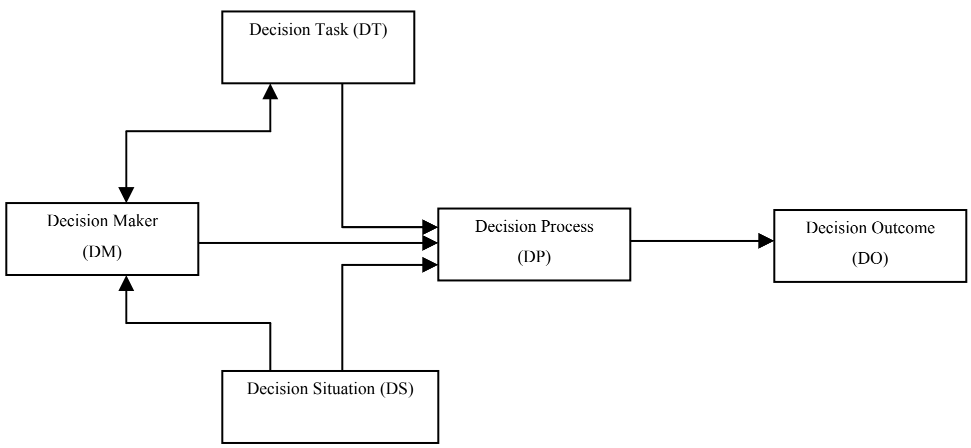


Figure 3.1: Hunt's et al. model of decision making (Hunt, 2000)

Several schools of thought agreed to the fact that decision making is a process that chooses a preferred option or a cause of actions from amongst a set of alternatives on the basis of given criteria or strategies (Wilson & Keil, 2001, Wang, et al., 2004). Similarly, studies on decision theory is a multidisciplinary one which encompasses cognitive informatics, cognitive science, computer science, psychology, management science, economics, sociology, political science, and statistics (Berger,1990).

In the submissions of Wang, et al., (2004) decision theories can be broadly grouped into two paradigms: Descriptive and Normative theories. A common example of the descriptive theory is the expected utility paradigm and Bayesian theory (Wald, A. 1950). This is based on empirical observation and other experimental studies of choice behaviour (Osborne & Rubinstein, 1994). The second theory is opposed to the descriptive theory and assumes a rational decision maker follows a well-defined preferences conforming to some axioms of rational behavior. In another dimension, set out procedures or steps are laid out for decision purposes. Economic Intelligence consist of such well defined process stages (David & Thiery, 2005), Edward & Fasolo (2001) developed a 19-steps decision making process. The above probably followed closely from the personal construct theory which proposed that experience and perception heavily influence decision making process (Kelly, 1955).

Decision Making: Mechanical or Judgmental?

Decisions commonly studied are called “programmed” and “structure”. They are decision for which the notion of process is considered irrelevant, it thus imply that decisions is an immediate mechanical connection of behavioural outcomes with environmental conditions (Gackowski, 2006a). Figure 3.2 presented below is far from making the act of decision mechanical. It is clearly evident that, the process which stems from problem solving ability proceeds in two parts to involve

(i.) The processes of comprehension; qualification and quantification of the magnitude of the decision to be made. The above becomes expedient based on the fact that, misrepresentation of any of the stages above listed could result into wrong perception/judgments which in turn would mar

(ii.) The subsequent processes of search; representation and memorization. In a situation where division of labour has brought about specialization of duties and operation, stage (i.) above could be undertaken by different user/actors.

Sequel to the above, the diversity in human personality, intuition and experiences, organizational factors, root of biases and misinforming unequivocally demand for reliable knowledge reconciliation between participating actors to forestall possible risk of misinformation in the process of delivering strategic decisions. Presented below in figure 3.2 is our submission and opinion on the complexities involved in decision making which could not possibly qualify it to be a mechanical operation?

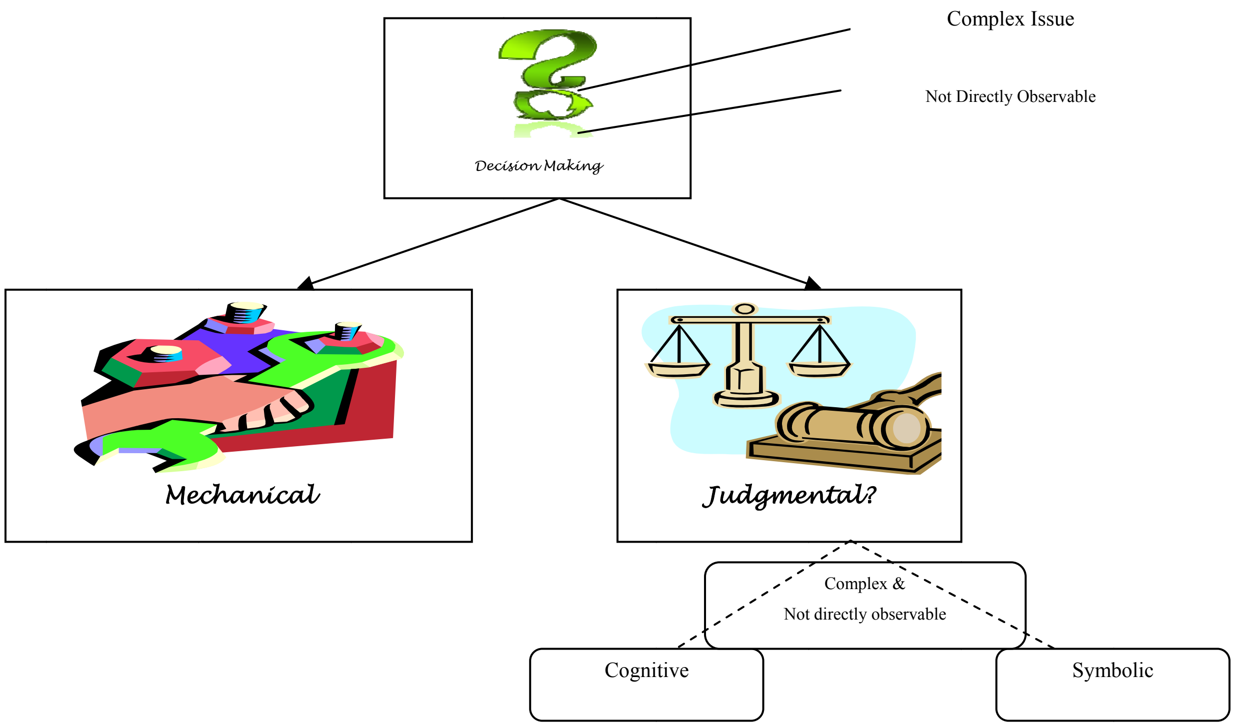


Figure 3.2: The complexities in decision making process (Onifade, et al., 2008)

There are popular schools of thought that consider decision making as being mechanical. Their arguments were based on the processes involved in the operation which was interpreted as the mechanics of the operation. Contrary to above opinion, most human decisions are judgmental than mechanical, which implies and support the various stages proposed/operated upon in the economic intelligence context. It was therefore supported that decision is a shorthand way of speaking about something which is quite complex and not directly observable i.e. cognitive and symbolic processes by which participating actors build and communicate model of their realities, apply them to cases or use them to organize action. Ruble & Cosier (1990) defined cognitive style as the way in which people process and organize information and subsequently arrive at judgment or conclusion based on their observation (Eden, 1994).

Mintzberg and Associated Model

Mintzberg, (1979) develops a theory based on the organizational structure postulating five basic parts constituting an organization. Our interest in this model is the part relating to the decision maker which Mintzberg described as the strategic apex. The model for managerial role has earlier on been presented in Mintzberg, (1973) and this is shown in figure 3.3.

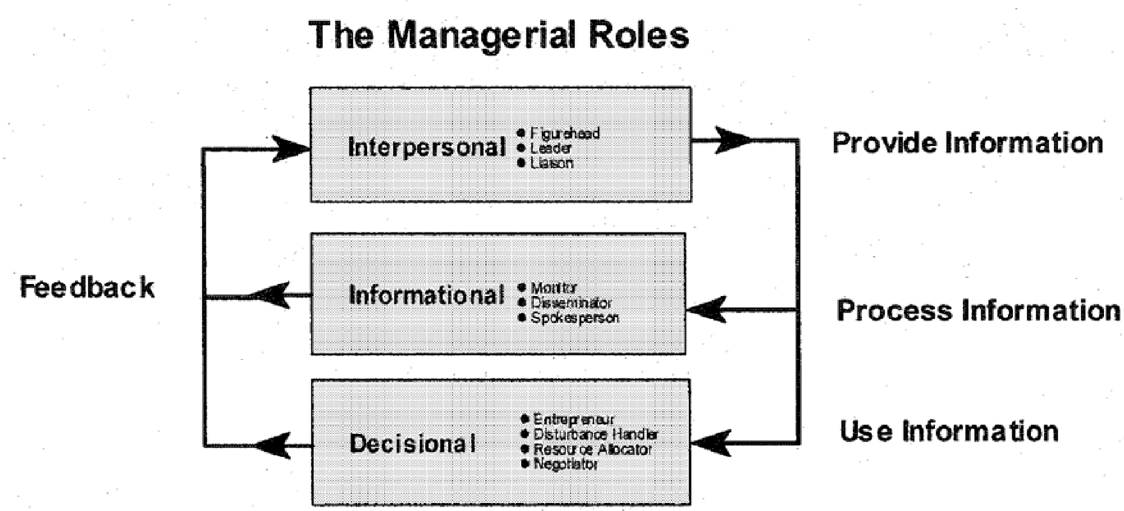


Figure 3.3: Mintzberg's model of Managerial roles (Mintzberg, 1994)

The description of the above model is almost similar to what we call EI process stage at SITE-LORIA research team. The organizational structure hitherto mentioned can thus be argued alongside the role of the manager or decision maker in our context. We can vividly see that the decision maker (manager by Mintzberg nomenclature) can provide information in form of 'decision problem', process information as regards 'information retrieval operation' and indicators calculation, and use the information to decide. Other factors considered important are listed alongside in the boxes for each of the interpersonal, informational, and decision function of a decision maker.

Mintzberg's school of thought on "strategy formation" posits that there are 10 schools of thought associated with strategic planning. The first three are related and are designated: "Design, Planning, and Positioning." These three are, according to Mintzberg, prescriptive in nature, that is, they strive to elucidate—or to put into plain words—the appropriate methods or procedures of making strategy (Mintzberg, 1994). The Design and Planning Schools are almost identical with the exception that the Planning School is more "formal" in its prescribed processes than that of the more conceptual and "informal" processes connected with the Design School. Both are essentially SWOT analyses and, therefore, their strength lies in the assessment stages of strategic planning as shown in figure 3.4. The Positioning School places greater emphasis on "content" rather than process and for this reason Mintzberg terms it as being more "analytical" in the rational or methodical sense (Young, 2002).

The remaining seven schools of thought Mintzberg categorizes are descriptive by comparison. That is to say, they are strategic approaches or viewpoints that take on thematic interpretations. For instance, Mintzberg describes the Cognitive School as one which focuses on mental or deliberative processes (Mintzberg, 1990). In the case of the Entrepreneurial School of strategic planning, the mainstay is a "strong person" with a compelling vision of the future. The Learning School places emphasis on collective or "group-think." The other schools of thought regarding strategy formation and their accompanying thematic views or processes include: Political = Power; Cultural = Ideological; Environmental = Passive; and, Configurational = Episodic (Young, 2002).

Decision making models consider here are not necessarily a comparison between an expert and a novice, but attempt is geared towards determining the various dimension employed by decision makers and possible researches that has been done in such direction. As regards the various model considered thus far, another important result is the work of Klein, (1998) reported by Horn, (2006) where a number of military commanders and firefighters were studied and the results is as follows. 46 – 96 percent of their decisions commences with a spontaneous recognition of the situation at hand.

Thereafter, additional investigative steps aimed at representing the validity of their assumption will be carried out.

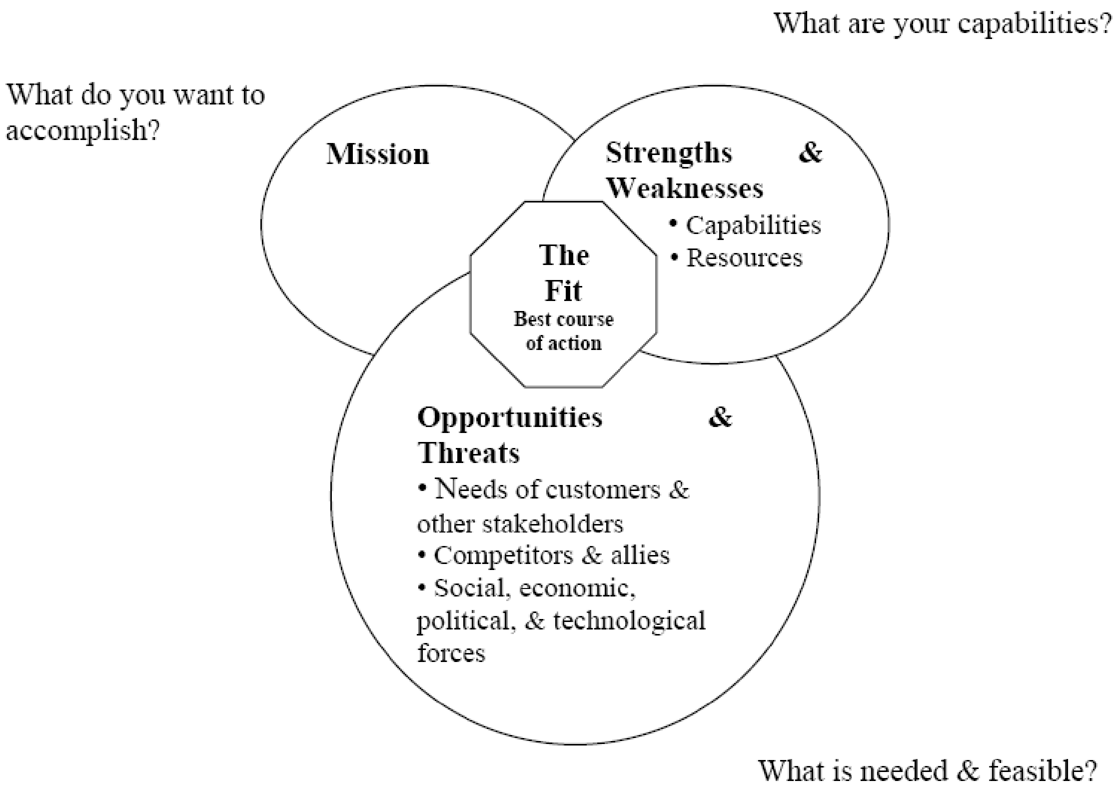


Figure 3.4: Composition of Strategic Planning with SWOT (Barry, 1997)

There has been various research submissions on the common "head vs. formula" concept aimed at establishing the superiority of rational-analytical approaches over the soft judgmental or intuitive approaches. While this notion has been reported successful in laboratory studies, it incursion into strategic decision making suffers a deathblow. Strategic decisions are characterized by incomplete knowledge, especially in dynamic business environments as of today. Thus identifying quantitative equations among variables or finding numeric values for initial state are extremely difficult if not impossible (Sinclair & Ashkanasy, 2005). Corroborating the above, Mintzberg, et al., (1976) reiterated that strategic problems are ill-structured and hence cannot be programmed.

Integrated Decision Making Model

The wealth of information available to decision makers in today's technologically driven world necessitated the need to be versed in many fields. It therefore behooves that multidimensional approach to decision making encompassing bounded rationality, heuristics, insight and intuition would be a better suited approach than mono approaches hitherto considered. In the submissions of Sinclair & Ashkanasy, (2005), an integrated model of analytical and intuitive decision making where both approaches are employed both iteratively and complementarily should suffice. In figure 3.5, this integrated approach is presented and its operation is based on the determination of the disposition and contextualization of either decision approaches.

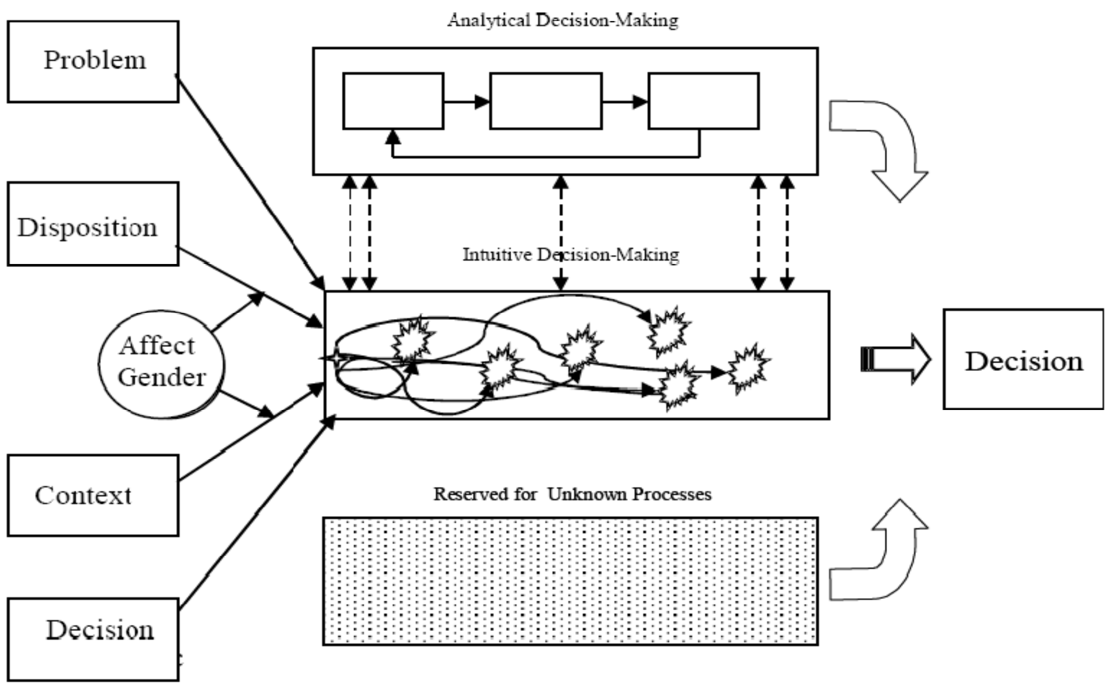


Figure 3.5: Integrated model of Analytical and Intuitive Decision Making
(Sinclair & Ashkanasy, 2005)

The model provides a representation for cognition and affective components within a single construct thus facilitating experience-based and affect-based perspectives. This is more robust than the experience-based perspective where emotion is regarded as detrimental to intuitive process (Simon, 1987). It also surpasses the functionality of the analytic approach which adheres to strict abstract rules of analysis and logic thus yielding precise answer to complex factual problems. In the integrated model therefore, the dominance of either approaches is a factor of personal disposition and the decision making context as shown in figure 3.5. Juxtaposing the submission of Mintzberg et al., (1998) and Epstein, et al., (1996) where each group of authors stressed factors like personal factors, emotional involvement, and organizational context, Sinclair & Ashkanasy, (2005) concluded that intuitive decision making is affected by four broad categories of factors: problem characteristics, decision characteristics, personal disposition, and decision making context.

Gambling Paradigm

In the work of Karen & de Bruin, (2003) in reviewing the assessment of decision quality mentioned the Gambling paradigm, the Conflict model, and the Accountable model. The authors raised a fundamental question on whether decisions should be judged by the process (via which they are derived, e.g. EI) or by the outcome (and its associated consequences). The process-biased draw their inspiration from the Utility theory discussing on the need for a well structures decision (Lipshitz, 1989). This view amongst other things believed that decisions are made under uncertainty, corroborating the fact that decision making is a bet (Edwards, et al., 1984). The structure is not as important as probably the outcome, this is the view of the normative adherents i.e. decision making should incorporate some level of probable outcomes as these affects the fulfillment of the decision maker (Baron, 1994).

The gambling paradigm dominates the metatheory in the field of decision making for a long time. The focal point is that every decision problem can be translated into a choice between gambles (Schoemaker, 1982). This theory relies heavily on utility theory (Keren, 1991, Keeny & Raifia, 1976). The conflict model considers decision making as the resolution of emotional conflicts. The rationale is such that, the choice of one option is automatically implying the discard of others (Janis & Mann, 1977; Coombs, 1987). The third approach reviewed is the accountability model. This model pinpoints that the decision maker is expected to convincingly defend his choice of actions – being held accountable (Tetlock, 1991; Lerner & Tetlock, 1999).

Decision making is the process of constructing the choice criteria (or functions) and strategies leading into the usage to select a decision from a set of possible alternatives. Different decision theories provide different choice functions (Wang, et al., 2004). Notable among these paradigms are the Game theory, and the Bayesian theory. The later is closely related to the utility theory because both of them provide alternative decision criteria from different point of view.

In sum, different theories and models of decision making exist, and the performance of these are difficult to judge based on different intangible factors inherent in human factors involved in decision making. Consequently, the judgment of decision quality depends on the views/perspective taken by the judge, i.e. whether the focus is on the outcome or the process, the decision or the decision maker, and a statistical or a clinical judgment amongst others. It therefore behooves that environment of application would be another factor to be considered in a process like this. In the following section, strategies employed in taking decisions were reviewed.

Strategies for Making Decisions

It is not uncommon for decision maker to perceive a genuinely complex situation as being simple. In another dimension, a merely complicated situation may be viewed as being complex (thus employing to resolve via deductive reasoning). In a bid to delimit the problems of inappropriate sense of judgment of individual decision maker, team-based decision making could be employed (Bullen & Sacks, 2003). However, decision team is a complex system not only because of the ensuing relationship amongst participants, but again the misperception, attempts at misinforming, disinforming and biases form another debilitating blow to the process.

The key to better decision making lies in obtaining relevant, accurate and timely information and employing the cognitive ability of individuals to assist in translating information into knowledge and lastly, decision making (Wilson, 1995). There are stack differences in ways via which individual structure and employ information for learning and problem solving purposes (Ford, 1995). With several researches in the field of memory structure aimed at describing the process of knowledge formation, acquisition and utilization, we can classify memory into both short and long term memory. However, we are more concerned with the long term memory in this section. This memory part consists of both declarative and procedural knowledge. Declarative knowledge has been described as 'knowledge that' while procedural knowledge is regarded as 'knowledge how' (Best, 1989). Declarative knowledge is flexible and can often be reorganized to suit purposes. Procedural knowledge is not however flexible (Maqsood, et al., 2004).

There are wide varieties of decision making strategies developed in the traditional decision and game theories, cum cognitive science, system science, management science, and economics. Research has proven that the outcome of a decision making process can be inferred from the decision making strategies employed by the decision maker. It is no boast to assert that these strategies require different decision selection criteria (Wang, et al., 2004).

Harris" Decision Making Strategy

Harris, (1998) highlighted some decision making strategies and the procedures for achieving them. The rationale was based on the fact that there is no single, definite way to solving a given problem, it is therefore the prerogative of the decision maker to choose out of such alternatives. The process of choice can be simple or complex necessitating several stages of adjustment based on importance, time, and available resources amongst other. With this background, a decision maker can employ:

- Optimizing – a strategy of choosing the best possible solution to a problem, while discovering many alternatives and choosing the best. This is dependent on: the importance of the problem; available time for the operation, cost of other alternatives, resources and knowledge at his/her disposal; and other human factors like personal traits, psychology and cognitive ability.
- Satisficing – this word as can be seen is not a true English word. It is coined from 'satisfactory' and 'sufficient'. The operation is such that, the decision maker choose the first satisfying alternative, and not necessarily the best one.

- Maximin – this strategy comes from ‘maximize the minimums’. This is a strategy of the pessimist, where in the worst possible outcome of each decision is considered and the decision with the highest minimum is chosen. The maximum becomes important when the possible consequences of a failed decision are particularly harmful or undesirable. Maximum is saddled with the salvage value of a decision, or the guaranteed returns of the decision, simply put – ‘a bird at hand is better than two in the bush’ strategy.
- Maximax – this stands for ‘maximize the maximums’. The focus is on the evaluation and subsequent choice of alternatives based on their possible payoff. It is an optimist strategy, since favourable outcomes and high potential areas are the main target. It is a welcome strategy when risk taking is most acceptable i.e. the ‘go-for-broke’ philosophy is in reign (Harris, 1998).

Wang et al. Strategies based on Cognitions

In table 3.1, Wang, et al., (2004) declared that the taxonomy of strategies and corresponding criteria for decision making can be classified into four categories i.e.: Intuitive, Empirical, Heuristic, and Rational. From the table, it can be deduced that existing decision theories provide a set of criteria (C) for evaluating alternative choices for a given problem.

Intuitive and Empirical are in line with human intuitive cognitive psychology, in which there is no specific rational model towards explaining the decision criteria. Under the rational decision category, we can have two distinct sub-divisions: static and dynamic strategies and criteria. Heuristic decision making are commonly and widely employed by human beings, detail of which can be gotten from (Hastie, 2001 and Wang & Wang, 2004).

No.	Category	Strategy	Criterion
1	Intuitive		
1.1		Arbitrary	Based on the most easy or familiar choice
1.2		Preference	Based on propensity, hobby, tendency, expectation
1.3		Common senses	Based on axioms and judgment
2	Empirical		
2.1		Trial and error	Based on exhaustive trial
2.2		Experiment	Based on experiment results
2.3		Experience	Based on existing knowledge
2.4		Consultant	Based on professional consultation
2.5		Estimation	Based on rough evaluation
3	Heuristic		
3.1		Principles	Based on scientific theories
3.2		Ethics	Based on philosophical judgment and belief
3.3		Representative	Based on common rules of thumb
3.4		Availability	Based on limited information or local maximum
3.5		Anchoring	Based on presumption or bias and their justification
4	Rational		
4.1	Static		
4.1.1		Minimum cost	Based on minimizing energy, time, money
4.1.2		Maximum benefit	Based on maximizing gain of usability, functionality, reliability, quality, dependability
4.1.3		Maximum utility	Based on cost-benefit ratio
4.1.3.1		-Certainty	Based on maximum probability, statistic data
4.1.3.2		-Risks	Based on minimum loss or regret
		-Uncertainty	
4.1.3.3		-Pessimist	Based on maximin
4.1.3.4		-Optimist	Based on maximax
4.1.3.5		-Regretist	Based on minimax of regrets
4.2	Dynamic		
4.2.1		Interactive events	Based on automata
4.2.2		Games	Based on conflict
4.2.2.1		-Zero sum	Based on $\sum$ (gain + loss) = 0
4.2.2.2		-Non zero sum	Based on $\sum$ (gain + loss) $\neq$ 0

Table 3.1: Taxonomy of Strategies and Criteria for Decision Making (Wang, et al., 2004)

If indeed a process approach is taking for considering decision making then, the procedure might as well be as important as the outcome of the resultant decision. Decision making is considered as one of the basic cognitive processes by human via which a preferred option or choice of actions are made from several other alternatives. With several varieties of strategies and criteria as revealed by the reviewed literatures, it pertinent to say that, decision making is multifaceted and thus, there is no single method that can adequately accommodate the complex nature of human being as shown in categorization made in table 3.1.

Decision theory usually offer framework that enables one to think about decision making under uncertainty (Pomerol, 2001). States are generally identified to comprise of the past, the present and the future fused together in comprehensive manner. Harris, (1998) in similar vein mentioned decision about decision, a state resulting from the need to decide within decision thus creating a hierarchical structure to decision strategy. With divergent practical situations, information and knowledge between the past, present and the future, there is need to distinguish between diagnosis and look ahead strategy (Pomerol, 1997).

In classical decision theory, a decision maker not only chooses an alternative but a policy according to various events whose possible occurrence is bounded with time. Thus, the decision maker makes a choice amongst alternatives, wait for the nature`s move and the corresponding result of his goal, a situation known as scenario development (Pomerol, 1998).

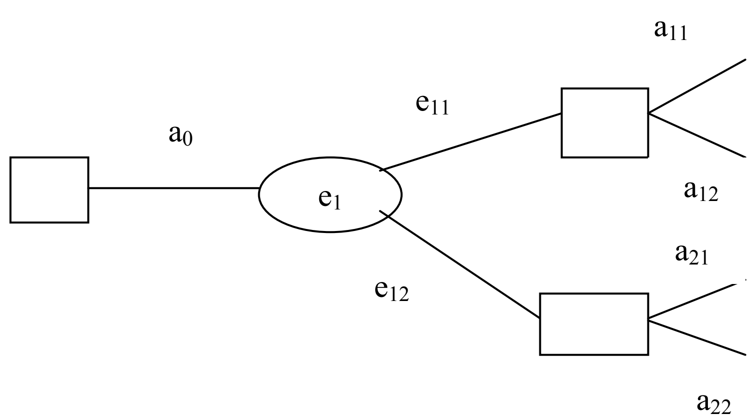


Figure 3.6: Scenario development for decision making (Checkland, 2000)

A typical decision scenario leaves a decision maker with countless number of choices. With multiple choices, the decision maker faces a combinatorial explosion of the number of branches which can be detrimental to the successful development of the decision tree. Consequently, several attempts have been proposed to alleviate this problem. Among which include tree-reduction or pruning method, what-if analysis and scenario thinking. We can therefore define from figure 3.6 a scenario S = {a_0, a_{11} if e_{11}, a_{22} if e_{12}} comprising of sets of alternatives and events transpiring between them. The development of this scenario will result into a tree, thus two events e_{11} and e_{12} are possible events at node e_1. Although the example of figure 3.6 is simple, scenario based decision making resulting into decision tree usually lead to combinatorial explosion. Several solutions have been proposed to deal with this situation, amongst which are: Time-table evaluation (Pomerol, et al., 1995), Privatization Law Systems (Levine & Pomerol, 1989), and "Extremeness Aversion" (Tverskey & Simon, 1993).

The multiplicity of decision problem is something individuals or group decision makers have to cope with at different stages, time and circumstances. Resolving it is best described as unraveling a riddle. This section has reviewed a number of strategies and listed taxonomy of decision strategies to depict various mode employed in decision making. In next section, we take a look at the differences between decision making via the traditional hard computing and soft computing paradigm.

3.3 Risks in EI Process & EI actors

Risk management is an important branch of knowledge management which is concerned with the quality of the knowledge models as essentials for insight and analysis. Knowledge is seen as legitimate and meaningful resources that strengthens the overall management performance, as a result, knowledge management is viewed as a sine qua non towards creation, storage, sharing, and reusing of the organization's knowledge, employing advances in today's technology(Allan, et al., 2007).

Centrally, the focus in EI context is on the trio of ***Information***, ***Actors***, and the research ***Processes*** of information. Their interactions have been graphically depicted in table 1.1 of the previous chapter. Economic intelligence has defined some generic actors which include: the decision maker – this actor is responsible for determining the organizational needs and formulating them into a decision problem (DP). The others are the Watcher and the Coordinator. The watcher is a specialist in information management and retrieval. He is expected to thoroughly understand the problem-need of the decision maker (decision problem) and make concerted effort to rigorously define and present the translation into the object of search i.e. information retrieval problems (IRP).

The coordinator as the name implies oversees the functionality of the operations within the systems. The EI process stage amongst other things shows the major concepts and their interactions to depict the level and point of involvement of each actor in the EI process stages. The processes have been well organized into various stages of operations referred to as the EI process stages. Starting with the identification of decision problem (DP) that requires or demand for solution, it proceeds through to the attempt to formulate the decision problem into information retrieval problem (IRP) together with other actors as defined in the EI context (Bueno, & David, 2001). The success of the process above will in no small measure determine the level of acceptability in terms of quality of the information result which is to be employed for the decision making process. While it is a known fact that such explicit definition and division of labour is expected to bring about laudable result, it's equally important to note that such interaction cannot be devoid of frictions which we referred to as risk.

A distinctive work on trends in decision making pattern and accruable risk within EI stems from the work of Duffing et al. (2005). With the searchlight directly on economic monitoring, which is enhanced by information system and data warehouse, it was submitted that data quality of appropriate level is important for accurate decision making. Recalling that the origin of the information is as important as the processes taking place on the information that are residing in the data warehouse. The process of selection, cleaning, storage and retrieval are very important to the decision made out of them. The decision maker must base his actions on available indicators (Thiery & David, 2001) relevant to his problems. The interaction, compositions and roles of EI actors were examined to determine the possibility of risk, its type and source for any decision taken. The authors arrived at two types of risks: Trades risk - reconciliation of knowledge and processes related to a specific area, and Technical risk – which involves information processing on the data warehouse. The pattern of risk modeling was therefore considered from both structural and behavioural axis.

Earlier on, we have made a role of existing actors in EI with their nomenclature. Herein we are concern about one of the actors known as the decision maker. It is the prerogative of this actor to be the watch-eyes for the organization in terms of its needs and operations. To achieve this, the decision maker has to be up and doing. Amongst the important attribute consider for adequate performance of this actor is his/her cognitive ability. While it cannot be argued that the cognitive capacities of decision makers may be of various capacities, based on their level of exposure, environment, and other factors, it is a common understanding however that, the core cognitive processes of human brain share/exhibit similar and recursive characteristics and mechanisms (Wang, et al., 2004). Personal construct hypothesized that experience and perception heavily influenced decision making process (Allan, et al., 2007). Delivering strategic decisions, high rate performance and complex decision making are cogent features in Economic Intelligence; it is therefore no gain saying to assert that appropriate method of information-sourcing for knowledge application is of utmost importance.

This section serves as a preview into the main features of EI. With the principal actors already identified and possible roles, it is evident that the interactions between these actors coupled with the intrinsic and extrinsic factors that can constitute risk factors would be some of the militating factors against delivery of strategic decision. These factors can be summed up as uncertainties. Next section will thus introduce us to these issues as a main challenge to decision making processes.

Risk Factor & EI Actors

Sequel to the distinction made by Gackowski (2006) distinguishing between direct and indirect informing. EI systems have been previously identified to be an example of indirect informing based on the established actors like the decision maker, watcher, and coordinator amongst others. The roles of individual have been hitherto stated in earlier chapters as found in Bouaka & David, (2004), Knauf & David, (2004), and Duffing, et al. (2005).

The concept of EI must be able to link Strategic management with technological usage to produce and enhance hitherto herculean task of decision making. Usually, there will always be the need for adaptation, followed by the introduction of concepts necessitating integration of means which may be human; techniques; organization or information (Simier, et al., 2004). The management of the above depicts some level of heterogeneity which must be fused-out to synchronize resultant decisions from such operation.

It is thus imperative to note that the data quality of acceptable level is important for appropriate decision making. Centrally in EI context, the focus is on the Information; Users, and the research Processes of information. It must be stressed that the origin of the information is as important as the processes taking place on the information. The meeting point of all information is the data warehouse, after which the information is selected, cleaned stored and returned to the application. The above processes enhances decision making, it therefore behooves that there is need for comprehensible, reliable, and easily accessible information (Duffing, et al., 2005).

One of such users in EI context is the decision maker, who must base his actions on available indicators relevant to the problem, therefore, all forms of strategic decision making rest on him. Other users have different roles which shall be outline later on. With the three central focus presented above , we present a simple diagram to represent them, while their personal, intra-personal and person-processes interaction will form the basis of discussion from which we hope to formulate the inherent risks that can result from such associations.

The concept of Actors, a kind of user has a broad definition in EI process: the set of users/actors include the decision maker, the watchers, infomediators/coordinators and others yet without nomenclature. However, it is interesting to note that, the interdependency of these actors result into division of labour which attracts both good and bad outcomes. In the formal case, the good side is to enhance the overall operation of decision making, and in the later case, it can constitute some level of risk into the overall decision made.

EI should involve the act of understanding the processes involved in the production of interpretable indicators for decision making, based on the information available to it (David & Thiery, 2004). The leader is expected to identify the problem in terms of risk, challenge and threat, through the observation of certain parameters within the environment, and also inform of assumptions which demand/requires verification.

The Watchers specialize in the collection of information using different methods for collection and analysis of information, geared towards value-added information. He is expected to translate the problems in term of attributes and indicators on information used, identify and verify relevant sources, collect relevant indicator and perform some pretreatment on the information. The Infomediator/coordinator on many occasions assists in the transformation of the decision problem; supervises the search and other relevant processes towards the presentation of most relevant information for the decision maker. It sometimes servers as the bridge between the decision maker, and the sources of information, to identify/notify on appropriate location of relevant information to aid search process (Knauf & David, 2004).

The overall objective is to facilitate minimum error in the discharge of the duties of the actors in EI processes. Thus, it will be required that a critical analysis of inherent risk be made on each phases as an entity, and their resultant interaction towards a common goal. The diagram below represents an interaction amongst the various actors participating in EI processes. The following are the definition followed in this research as it affects the usage and consideration of the terms ‘information’, ‘users’, and ‘processes’.

1. Information = { database technology & distribution; interoperability & migration and information extraction, sharing & utilization}
    - Database Technology & Distribution: Database systems, Distributed systems
    - Interoperability & Migration: Heterogeneous database systems, Migration legacy databases, Multimedia database systems and Client server databases

- Information Extraction, Sharing & Utilization: Internet databases, Collaborative database management, Data warehousing and Data mining

The above is meant to determine the quality of data which in turn determines the appropriateness of decision that emanates from such environment

2. Users = {decision maker, watchers, info-mediators, e.t.c.}. The list of users which is referred to as actors in the context of economic intelligence is not yet exhausted.

3. Processes = processes to be considered include the processes in Economic Intelligence concept, which are listed directly below, and the processes in data warehouse which will be discussed in the later part of the write up. {The set of all possible activities undertaken by the classes of users stated in the EI process stages}.

Starting from the lower part of figure 3.7, the major material (raw or processed) is presented as the information, from which there could exist various forms of errors ranging from non-availability; partial-representation; disparity in data format e.t.c. We can therefore represent risk at this level with $\mathcal{R}_i$, - risk accruable from information acquisition, representation and extraction.

The second level has two different actors which might not necessarily be at the same functional level. Thus, from the work of Kinslin, (2007), the Watcher is responsible for sourcing for information based on the decision problems observed and defined by the decision maker, which is now translated into an information retrieval problem. We therefore implies that, from this part, we have $\mathcal{R}_w$, - that is the risk resulting from the mapping of the decision problems and its corresponding information retrieval work undertaking by the watcher.

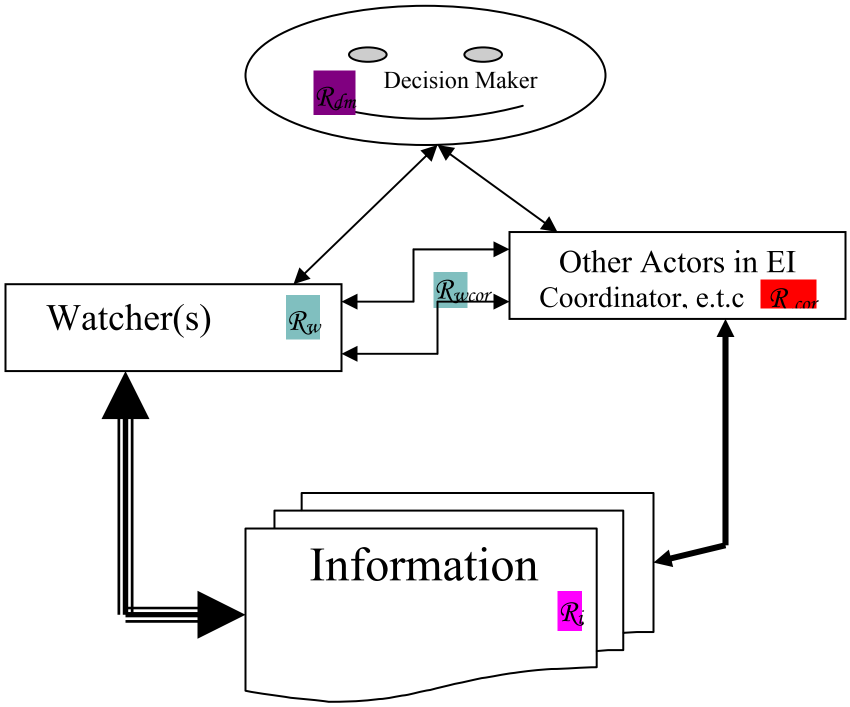


Figure 3.7: Relationship and Accruable Risks amongst EI Actors

The newly identified actor in the EI processes is the Infomediator, Audrey in her presentation and other related journal elucidate the activities and importance of the Infomediator/coordinator in the execution and supervision of various tasks with the organization. This actor, among other functions coordinates and supervises the activities of the watcher(s), with a view to present a common view to the decision maker. Thus, we use $\mathcal{R}_{im}$, - to represent the possible inadequacies that could result from her activities and constitute a risk to the discharge of the duties of the decision maker.

We also opined that since there is interrelationship between watcher(s) and the Infomediator/coordinator, there's the possibility of some type of risk that cannot be directly grouped on either of the earlier discussed actors. We therefore use $\mathcal{R}_{wim}$, - to denote the risk which can result from the joint action/inaction of both the watcher(s) and the Infomediator/coordinator.

A particular type of risk that is not yet expressly stated is the one based on the decision-ability of the decision maker (Onifade, et al., 2009). There are several works on the effect of individual personality on the corresponding act/decision made by such individual. This type also includes factors like the actors attributes; his behaviour and the context of information sought. We represent this type by $\mathcal{R}_{dm}$, - that is, resulting risk from the action/pronouncement of the decision maker. It should however be noted that, this is not the only risk that could result from a particular type of decision made by the actor; instead, it is the summation of the total error present within the process of any decision to be made. Thus, for any cogent decision to be made, there exist the maximum of the summation of all inherent risk with the EI processes. Mathematically, we have

$$\mathcal{R} = \sum \left(\mathcal{R}_{i} + \mathcal{R}_{w} + \mathcal{R}_{im} + \mathcal{R}_{wim} + \mathcal{R}_{dm}\right) \qquad (3.1)$$

The onus of the above is that, dealing with risk in a dynamic environment like EI processes, involves the modeling of the activities with user preferences alongside other factors which will be deemed fit in the course of this research. While noting the interaction amongst the EI actors as a source of risk, these risks are the result of uncertainties, ambiguities and incompleteness that characterizes decision making. In the next section, we take a look at uncertainty and its effects on decision making.

Risk of Uncertainty as Challenges to Decision Making

Decision making often relies on relevant information supposedly extracted from data. Obtaining such information involves many data analysis techniques like statistical analysis, clustering algorithms, and modeling via neural network or machine learning (Hewett, 2004). It has however been observed that most analysis techniques does not particularly apply to incomplete data or missing data. Decision makers are prone to mirage of problems militating against the delivery of their duties. Despite the avalanche of richly interrelated information and incorporate technical knowledge, there is little or no room for error cum less time to rationalize available information (Wherry & Hopson, 1981). It is therefore no gain saying to assert that decision making is plagued with various factors amongst which uncertainty stands out. Uncertainty can be of different mode, however, we make bold to assert here that the kind of uncertainty dealt with in this research has to do with decision made before all relevant parameters are know with every certainty (Wallace, 1998).

A popular analogy is the one relating uncertainty to risk. Risk is generally interpreted as the existence of unpalatable occurrence brought about by the pair of threat and vulnerability. It implies that, risk is not necessarily a danger, but an indication of a dangerous situation which could result from the willful or unwillful act of exercising /playing on the weakness (vulnerability) of the system in question.

Decision theories offer several frameworks towards arriving at logical conclusion under uncertainty. States towards decision making were categorized to include the past, the present and the future from which deduction must be made in a very comprehensive concept (Pomerol, 1995). With ever changing nature of the information and other factors, decision maker is faced with uncertainty. The use of "what if" analysis and/or scenario thinking consequently suffice to allow for distinction between diagnosis and look-ahead (Pomerol, 2001). The implication of the above is the fact that the decision maker is saddled with a combinatorial explosion of the number of branches resulting from the decision tree. The next stage involves evaluation of the choices (pruning) via which several alternatives would be considered for acceptability.

Assessment of risk is a complex and ambiguous exercise, it is therefore not unconnected with the fact that many if the existing techniques utilize the probability distribution function of threat and impact loss (Wang, et al., 2006). While it is almost impossible to eliminate all forms of risk with the sporadic growth in information technology usage and adoption, the need for efficient and adequate risk assessment cannot be overemphasized. Risk assessment has been described as an important part of risk management (CORAS, 2002).

Since most risk assessment often hold under uncertainty resulting from incomplete or partial information, denial of operation or its lateness, it is always difficult for users to collect and utilize precise and complete evidence in estimating the probability of threat and impact losses. The shortcoming of the above mentioned lies in the fact that such approach requires data to be presented in numerical format, which is not usually possible in decision situation.

The classification work performed by Koller, (2000) made a broad grouping of the assessment techniques into five parts: Discriminant Function Analysis (DFA); Bayesian Analysis; Decision Tree Techniques; Factor Analysis (FA) and Neural Nets (NN). Work by individuals and organization on risk assessment dated by to time immemorial. Popular amongst which is the work of Carroll, (1983) where the author proposed a 'familiar risk analysis model' employing the Annual loss Expectancy (ALE)' to calculate the risk of an alternative via multiplication of the Annual Rate Occurrence (ARO) with the single loss Expectance (SLE), and Exposure Factor (EF) with the value of assets. A qualitative risk assessment method based on fuzzy similarity measurement in determining alternative ranking was presented in Chen, (2003). Other corporate bodies involved in this process include: the Consultative Objective and Bi-functional Risk Analysis (COBRA), and Centre of Risk Management of Engineering Systems (CRMES, 1991), and Consultative Objective Risk Analysis System (CORAS, 2000), whose objectives are

- to develop a base framework applicable to security critical systems that will supply customizable, component based road maps to aid the early discovery of security vulnerabilities, inconsistencies and redundancies
- to assess the applicability, usability and efficiency of the framework by extensive experimentation in the fields of e-commerce and telemedicine.
- To investigate its commercial viability and pursue its exploitation within relevant market segments, while playing influential role in standardization of organizations.

The above is to draw a correspondence between the limitation of existing infrastructures and the design of new systems with satisfying degree of security. This is sequel to the increasing complexity of IS urges the improvement of existing design, thus we have: complementary security risk analysis methods and System architecture (CORAS, 2000). The methods employed are as followed. Exploitation of the risk analysis methods with semi formal specification supported by adaptable tool integration platforms: Markov analysis, Hasop e.t.c. Implemented in XML technology, the internal representation is similar to internal data representation formalized in XML/XSL schema. The methods for risk analysis include: failure modes, effects and critical analysis (FMEA/FMECA); fault tree analysis (FTA); hazard and operability analysis (HAZOP) and cause consequence analysis (CCA).

In this section, we have attempted to unravel the mystery behind the colossus named uncertainty and risk in decision making. We made bold to assert that there is high risk arising from the need for result validity due to human factor in decision making. Practical decision making based on formal models and methods which are applied to complex, ill-structured situation control, formalization of representation of people (expert, analyst, decision makers) on situation, its problem, and also about people's goal and interest unequivocally turns to be a cogent aspect of the problem solving (Abramova & Kovriga, 2008). Consequently, most decision, information retrieval systems are largely characterized by various forms of uncertainty viewed as risk. We have distinguished between the diagnosis and look-ahead theories thus concluding that uncertainty increases because of the vague nature of prevailing circumstances. In the next section, we shall be looking at different decision models and their characteristics.

3.4 Decision Making viewed between Hard & Soft Computing Paradigm

"Basically, soft computing is not a homogeneous body of concepts and techniques. Rather, it is a partnership of distinct methods that in one way or another conform to its guiding principle. Consequently, the dominant aim of soft computing is to exploit the tolerance for imprecision and uncertainty to achieve tractability, robustness and low solutions cost (Zadeh, 2003). The principal constituents of soft computing are fuzzy logic, neurocomputing, and probabilistic reasoning, with the latter subsuming genetic algorithms, belief networks, chaotic systems, and parts of learning theory. In the partnership of fuzzy logic, neurocomputing, and probabilistic reasoning, fuzzy logic is mainly concerned with imprecision and approximate reasoning; neurocomputing with learning and curve-fitting; and probabilistic reasoning with uncertainty and belief propagation" (Zadeh, 1994).

In the soft computing framework, the basic idea which has been developed so far were consisted in supposing that there is a set of resolving agents (Pelta, et al., 2006) which are basically algorithms for solving combinatorial optimization problems, and to execute them cooperatively by means of a coordinating agent to solve the problem in question, taking the generality based on minimum knowledge of a problem as a fundamental premise.

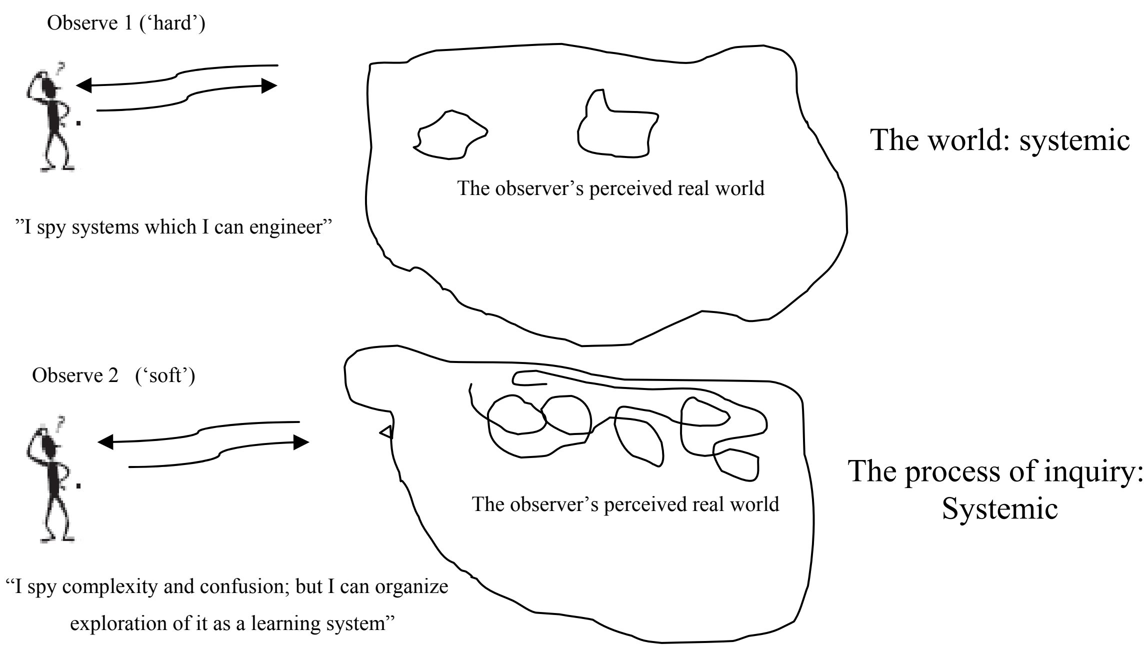


Figure 3.8: The hard and soft system instances (Checkland, 2000)

Each solving agent acts autonomously and only communicates with a coordinating agent to send it the solutions as it finds them and to receive guidelines about how to proceed. The coordinating agent receives the solutions found by each solving agent for the problem, and following a fuzzy rule base to model its behaviour, it creates the guidelines which it then sends to them, thereby taking total control of the strategy.

There are many criterions how to divide systems – static and dynamic ones, deterministic and stochastic ones (Pešl, & Hřebíček, 2003). System theories (Jančarová et al., 1996, Získal et al., 1996 and 1998, Hall et al., 1962, Gault et al., 1987) define the system mathematically as a set of elements P and a set of interactions R between them ($R \subseteq P \times P$). In this section, we shall employ the succinct partition of soft and hard systems. In the case of hard systems, the problem is well structured, clearly defined and we often already have known algorithms for finding the solution. These problems are reproducible. The soft systems are ill-structured with many factors interacting with each other, with randomness and more evaluating criteria. Moreover, we must often consider a human intervention into the system. The problems are mostly unique.

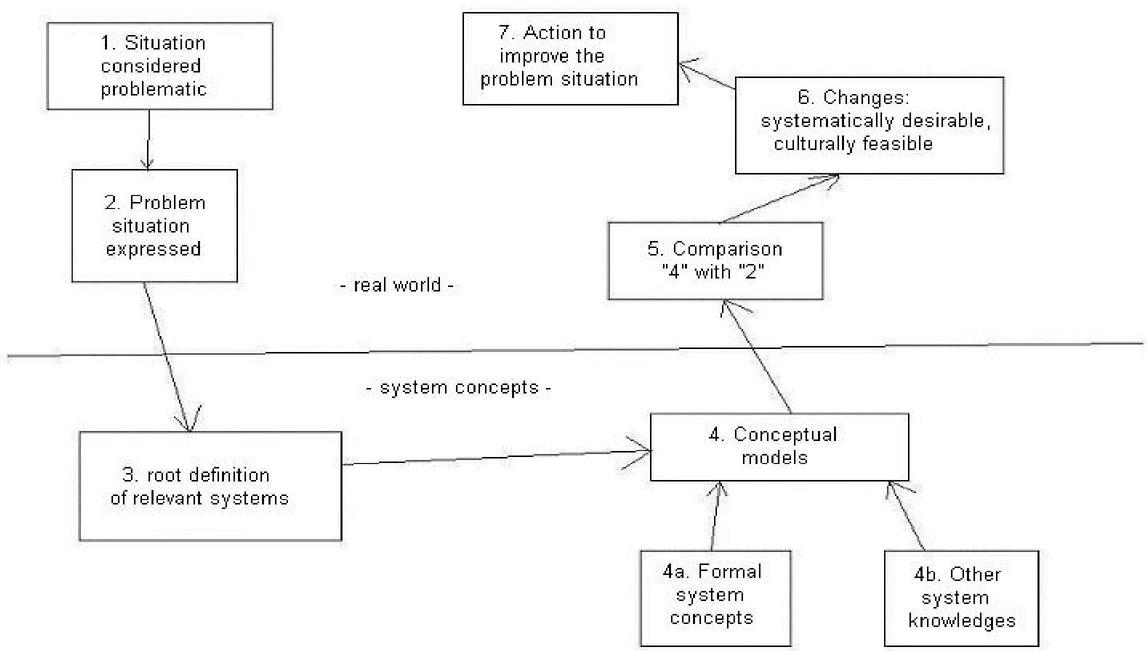


Figure 3.9: Problem solving steps in Soft Systems Methodology (Pešl, & Hřebíček, 2003)

Whenever available information is too imprecise to justify the use of numbers and there is tolerance for imprecision that can be exploited towards a tractable, robust, low solution cost and better rapport with reality, Zadeh (1996) opined that computing with words (CW) is the answer. CW involves the fusion of natural languages and computation with fuzzy variables. Gary, et al. (1999) referred to it as a new mathematical approach capable of accommodating inherent complexities in modeling human behaviour. These techniques includes: Fuzzy logic, Neural networks, Probabilistic reasoning – Genetic algorithms, Chaos theory, Belief nets and Learning theory. Unlike the 'hard computing' the difference as earlier mentioned is in the tolerance for imprecision, uncertainty and partial truth. It has as its role model the human mind.

Fuzzy inference systems have been applied successfully in many fields which include automatic control, data classification, decision analysis, expert systems and computer vision to mention a few (MathWorks, 2008). With this wide range of application spanning several fields that are interdisciplinary in nature, it has been associated with various names. These include: fuzzy-rule-based systems, fuzzy modeling, fuzzy associative memory, fuzzy logic controller, and simply (and unambiguously) put fuzzy systems. In fuzzy logic, the truth of any statement becomes relative to the matter of degree of participation. Fuzzy inference is the process of formulating the mapping from a set of given inputs to a set of given outputs employing fuzzy logic. This resultant mapping enables a basis on which robust decisions can be based, made or help in discerning a pattern from available information.

Fuzzy algorithm has been described as an ordered set of fuzzy instructions, which upon execution yield an appropriate solution to a specified problem (Zadeh, 1973). It is therefore no gainsaying to assert that fuzzy algorithm pervade almost all our endeavour. We employ it either consciously or unconsciously when we talk, walk, drive a car, search for an object, tie a knot, park a car, cook a meal, find a number in a telephone directory or even attempt to search for information for decision making process.

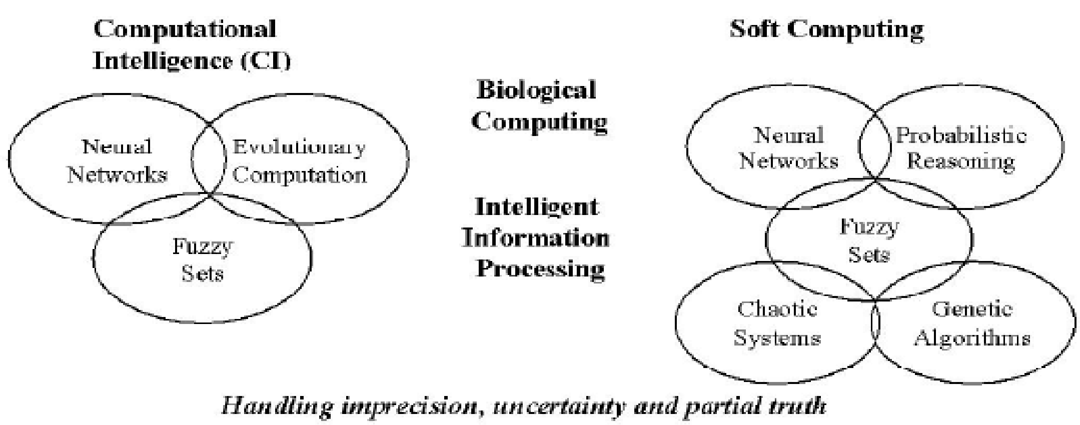


Figure 3.10: Methodologies for handling uncertainties (Sterritt, 2000)

Principally, the basis for soft computing as earlier mentioned is the need to accommodate imprecision and uncertainty amongst others, figure 3.10 depict various methodologies for handling uncertainty (Sterritt, 2000). This research is employing the fuzzy logic paradigm and its properties in resolving and management of information risks in EI. Zadeh, (2002) submitted that computing with words is in inspired by the remarkable human ability to perform a wide variety of physical and mental tasks without any clear-cut measurement or computation e.g. driving inside a heavy traffic. The contributing factor to this phenomenal capability is the brain's ability to manipulate perceptions – perception of distance, size, weight, colour, speed, time, direction, force, number, truth, likelihood and other physical and mental objects. Consequently, we can infer that the difference between measurement (hard computing) and perception (soft computing) is the crispy nature required in the formal, while the latter is fuzzy.

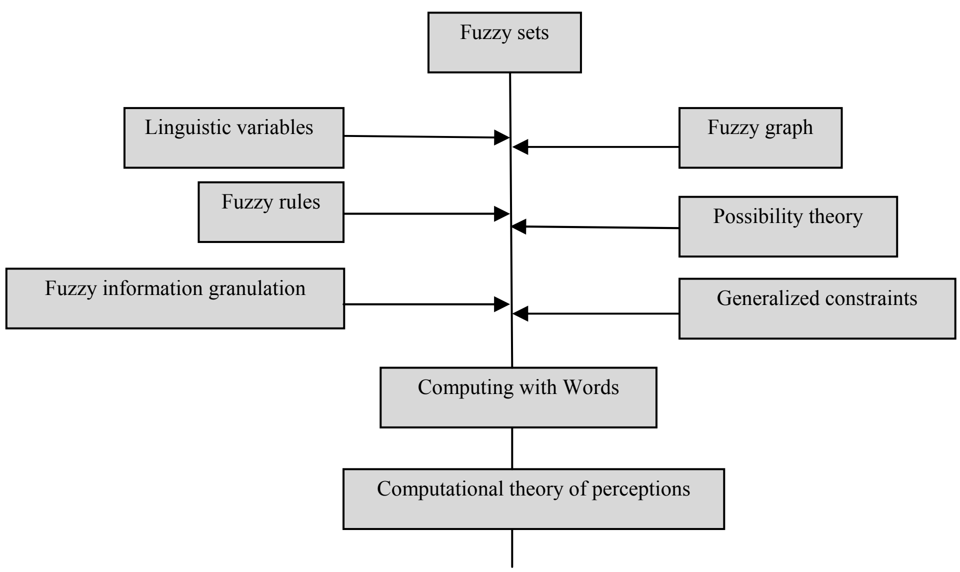


Figure 3.11: Conceptual structure of computational theory (Zadeh, 2002)

A granule is a fuzzy set of points having the form of a clump of elements drawn together by similarity (Zadeh, 1996). Consequently, a word *w* is a label of granule *g* and, conversely, *g* is the denotation of *w*. Zadeh (1996) defines a word as being atomic (as in strong) or composite (as in not very strong). Within the concept of computing with word

(CW), a granule *g* which represents the denotation of a word *w* is viewed as a fuzzy constraint on a variable. This is usually propagated from premises to conclusion.

Soft system methodology has several important advantages: it is only a recommendation for the process and the researcher needs not to hold it properly. Moreover, it offers backtracking and iteration processes when we would like to improve the output from some of the stages. The work can be done simultaneously at variously detailed levels and in more epochs.

The soft and hard system specifications seem that there is a strong boundary between these two categories and that we can consider the area of policy and sociologic sciences as soft systems and the area of physics, biology, geology etc. as hard ones (because here it is only a description of the material world). But this point of view is very inaccurate – e.g. in the economy we can reflect the mathematical formulas and empirical data as a part of hard system and on the opposite side the market dynamics and economic process behavior as a soft system. It means we should not try to fix such boundary and the system scientists should not apply only one methodology without looking for connections to other methodologies.

Frequently asked in criticism is the connection between soft and hard systems. The differences and specifications have been mentioned above, but there are also many correspondences and common characteristics. Most important of them it is the common goal – decision based on the model solution – and incidental output (in both cases we would like to have the model with solution). There is also common planning and searching for alternative solutions. In all, the importance of soft and hard systems is the result they attempt to produce – aiding decision making. It is therefore a factor of convenience and appropriateness in the choice of system approach employ for a particular operation.

3.5 The Crux amongst Uncertainty, Vagueness and Imprecision

Uncertainty management is the major crux between ‘hard’ and ‘soft’ computing methodology. There are various submissions to the concepts of vagueness/impreciseness/uncertainty as depicted in different article (Virrantaus, 2003). However, it is not always clear whether the question is about real world object or the conceptualization of it, the entity in the data base or its schema. Considering from the perspective of knowledge engineering applications dealing with fuzzy information and originally based on the Zadeh´s paper in 1965 (Zadeh, 1965). Vagueness, inexactness and recently imprecision has been one of the core themes in knowledge engineering associated with decision making and reasoning. Niskanen (1989) provides taxonomy of the core concepts; in the following these definitions are used. Virrantaus, (2003) briefly introduce the taxonomy as follows: In knowledge engineering the special interest has been in linguistic imprecision, it means the human verbal way of describing things (Niskanen, 1986, 1998). Words like “big”, “small”, “short”, “tall” are not precise, but imprecise expressions (Zadeh, 2002). However, there are some other types of imprecision, which maybe even more interesting from our point of view. Niskanen (1989, 1998) gives an analysis of the uncertainty and imprecision of the source data sets for a military terrain analysis application and taxonomy of terms related to imprecision. On the highest level of concept taxonomy he defines division into two main themes: imprecision and uncertainty.

Uncertainty is a concept associated with probability and the classical accuracy and error definitions. Imprecision is divided into three subparts: ontological, epistemological and linguistic imprecision. ***Ontological imprecision*** *appears when the exact truth cannot be achieved, then the question is about imprecise object of the reality.* ***Epistemological imprecision*** *occurs in situations where a human being can not identify an object exactly; the reason may be for example foggy weather.* ***Linguistic imprecision*** *means that the verbal expression is imprecise and can have several meanings* (Niskanen, 1989, 1998).

Uncertainty is inevitable at all levels of humans' interaction with their environment. At the lowest level, biological processes are never clear-cut and without noise. At the cognitive level, uncertainty results from *inadequate information sources*, *limited information processing capacity*, or *ambiguities in natural language* (Wang, et al., 1998). It was consequently asserted that an important aspect of uncertainty research is how to quantify or measure the uncertainty Barto, (1990) & Bar-Hillel, (1980). One classical approach is probability theory in general and the Bayesian approach in particular (detail discussion in Robert, 1994). In the Bayesian approach, uncertainty is usually represented by a probability, (P, a number between 0 and 1), interpreted as degree of belief based on all available knowledge. When new evidence comes, belief is updated or revised based on Bayes' Theorem, which is generally regarded as the normative model of belief revision. This approach has been founded theoretically and is easy to understand, thus it has been successfully applied to a wide range of domains.

An alternative approach to uncertainty management is Dempster-Shafer (D-S) theory (Shafer, 1976), which represents the belief about a proposition as an interval [Be*l*, P*l*]. Be*l* is interpreted as the degree of *belief* (or *credibility*) and P*l* is the degree of *plausibility*. The interval between the two represents ignorance, which is difficult and cannot be represented by probability theory. Both Be*l* and P*l* are numbers between 0 and 1, and the two are related in the following way: P*l*(H)=1-Be*l*(~H). Consequently, the degree of plausibility of H is the complement of the credibility of not H. Probability theory is claimed to be a special case of D-S theory, when the [Be*l*, P*l*] interval degenerated into a point, i.e., Be*l*=P*l*=P. It has been founded that D-S theory has its limitations. Majorly, it is harder to understand and implement than probability theory. On many occasion it is therefore assumed that the set of hypotheses is exclusive and exhaustive and requires an independent body of evidence, which is usually unrealistic (Wang, et al., 1998).

There have been sizable numbers of studies with a bid to identify different types of uncertainty based on the psychological sources of uncertainty. For example, Kahneman & Tversky (1982) distinguish between external uncertainty and internal uncertainty.

External uncertainty refers to disposition or randomness of external events, which are something people cannot control. An example is the outcome of throwing a coin. External uncertainty can be assessed in either a relative frequency mode or by subjective judgment of a single event. Internal uncertainty refers to ignorance, which results from incomplete knowledge. A typical example is that one is unsure Paris is the capital of France.

Living involves decision making which comes with various degree of uncertainty. These become inevitable with increased level of human interaction with the environment. A decision maker thus have to cope with diverse forms of uncertainty ranging from subjective estimation and perception, complexity of studied systems cum their environment, interactions amongst subsystems, lack of precise values, missing values, limited information processing capability, and ambiguity in natural languages (Passam, et al., 2003). Wang, et al., (1998) thus submitted that the notions of uncertainty, imprecision and ambiguity which are inherently present in natural system cannot be adequately handled with probability theory dealing with randomness. Consequently, attempts at imposing precision or theoretical assumption will only fuel the complexity of the model. In accordance to the above, we quote expressly from the work of Bouyssou (1989) where four main sources of uncertainty, imprecision and inaccurate determination in decision model were expounded. The original presentation of these factors was in Roy (1998).

The "map" is not the "territory".

Locating a plant, choosing equipment, investing in new activities is crucial decisions for a firm. The purpose of decision-aid is to compare such complex alternatives. If one wants to use a formal model of decision-aid, the complexity of these alternatives and their consequences makes it often impossible to compare them directly. This comparison is made possible through the use of "maps" of these complex "territories". For an alternative, a map consists of a model of the consequences of its implementation (in order to describe an alternative it is possible to use several maps of different "scale" using e.g. a hierarchical model).

These maps create a tractable language that allows an effective communication between the various actors of the decision process and provides an adequate basis for the comparison of the alternatives. However the establishment of the maps inevitably involves many simplifications, omissions and distortions which introduce in the model an important source of arbitrariness (Bouyssou, 1989).

Indeed, there are often several and equally valid ways of building these maps. While forced to use maps in order to compare territories, the analyst has to make a tradeoff between the richness and the readability of the maps: the "richer" one map is, the closer it is to the territory, but the more difficult it may be to compare it to other maps.

The "future" is not a "present" to come

The alternatives that are to be compared will only be implemented in a more or less distant future. Thus, at the time of the study, the consequences of the implementation of an alternative are very often unpredictable for they depend on environmental factors and/or the strategy of other actors that are still unknown and may well be influenced by the implementation of that alternative. This is the most classical source of Imprecision, Uncertainty, and Inaccurate Determination (I.U.I.D.) that is mentioned in every textbook on decision models. Many efforts have been devoted to cope with this unpredictability using, e.g., probability distributions, plausibility measures, scenarios, etc (Bouyssou, 1989).

Roy (1988a) mentioned the unpredictability of the consequences of implementing an alternative also stems from the fact that the alternatives are not completely specified at the time of the study. When a firm tries to compare several sites for locating a new plant, the precise characteristics of each site may not have been completely investigated yet. Furthermore, the precise draft of the plant to be built may not be available and may well depend on the site chosen. Thus, even if one could predict with a very high precision the consequences of an alternative, an element of inaccurate determination would remain since the alternatives are still "projects".

"The data are not the result of exact measurement"

The establishment of a map usually involves the consideration of two types of data. Data of type I, are closely linked to the territory that the analyst wishes to describe. The modeling of uncertainty mentioned in the above paragraph will apply to this first type of data. For instance, suppose that an analyst has to evaluate the human consequences of building a polluting plant on a given site. He will have to cope with uncertainty since he will be forced to envisage various scenarios for the growth of the population in that area. He will also have to deal with imprecision since the present number of people living close to the projected plant is far from being perfectly known. Counting houses on a map or on-site studies do not lead to precise evaluations. Thus, it is important to realize that many figures used in decision-aid models are only "order of magnitudes". This imprecision is often seen as stemming from the measurement techniques that are used. It also comes from the fact that, in many situations, the very definition of what "should" be measured is very imprecise. Using the same example as above, it is not clear how the analyst should take into account schools, hospitals, second homes, etc. This inaccurate determination of what is to be measured is certainly at least as important as the imprecision inherent to any kind of measurement.

Data of type II concern the way the first type of data is used in the construction of the map. Parameters like discounting rates or utility functions designed to capture an attitude towards risk are examples of this second type of data. They are more linked to a particular value system than to an alternative. In our cited example, such data could consist of the weight assigned to each inhabitant that is function of the distance between his residence and the projected plant. The attitude toward risk of the firm concerning the amount of nuisance created for the riparian, etc. Though techniques have been created to assess these data, it is important to keep in mind that they are very often "created" as well as "measured" (typical example is presented in the work of McCord and de Neufville (1983) concerning utility functions).

"The model is not the description of a real entity independent of the model"

Data of the second type are connected with certain aspects of the preference system(s) of the actor(s) involved in the decision process. It is well known that the questioning process used by the analyst in order to obtain these data may significantly influence the answers

(Bouyssou, 1984). This is all the more true since the preference system of an actor may not be completely structured at the time of the study: areas of firm conviction may well coexist with areas of hesitation and ambiguity in which the influence of the model on what is to be "captured" is overwhelming.

Furthermore, the various actors may well disagree and, as a result of a discussion, some actors may change their mind on some point thus creating some "inconsistencies" with previously stated judgments. In such cases the management of these hesitations, contradictions and conflicts seems a prerequisite to any convincing decision-aid model. This is linked to what Roy and Bouyssou (1986) called a constructive attitude towards decision-aid, as opposed to a descriptive one, in which the role of the analyst is not to describe as accurately as possible supposedly pre-existing preferences but to provide information and tools that are useful for justifying, building and arguing preferences.

3.6 Recap

In this chapter, it has been founded that most decision making are based on three main components: obtaining relevant information (from memory or external world), construction of the decision or problem space followed by attempt to fix the acquired information appropriately into the decision problem structure, and assessing the values and likelihoods of different outcomes (Karen & de Bruin, 2003). Consequently, it's no gain saying to assert that decisions are rarely what we perceived them to be, partly because we sometimes misrepresent it, or the influence and presence of other intangible factors like biases, misinforming, and disinforming constitute a barrier in modeling the decision appropriately. It therefore behooves that emotion is an intrinsic component of individual's decision ability. Bullen & Sacks, (2003) succinctly concluded that any decision framework devoid of emotion is doomed to fail. The power to decide thus lies in inherent adaptability and not the accuracy of its predictions.

Decision maker are daily faced with ill-structured business environment worsened with fast-paced changes and rising uncertainty. Consequently, this research work shall adopt the integrated decision making model and inculcate the earlier discussed factors like uncertainty, vagueness and incompleteness as factors that derail the decision making process. These factors has contributed in small means to decision risk in many organization based on the pattern of decision making. Our choice of decision making model for the rest of this research shall be based on the recognitions stemming from the integration of both analytical and intuitive-based decision making. This follows closely from one of our propositions in this chapter where the rationale between mechanical and judgmental decision pattern was considered. In all, decision cannot be made in the absence of quality information, thus it is evident that the last two chapters are buildups to model propositions in this research towards minimizing information risk in economic intelligent systems. In next chapter, we shall enumerate our models proposed to reduce the effect of all the aforementioned.

"*So far as the laws of mathematics refer to reality, they are not certain. And so far as they are certain, they do not refer to reality*". —Albert Einstein

"*I believe that nothing is unconditionally true, and hence I am opposed to every statement of positive truth and every man who makes it*".

—H. L. Mencken

"*As complexity rises, precise statements lose meaning and meaningful statements lose precision*". —Lotfi Zadeh

Chapter 4
Our Models for Risk Management in Economic Intelligent Systems

4.1 Overview

In the previous chapters of this write up we have reviewed extensively the issues relating to the major components that make up economic intelligence system. These include the human-actors, information and the processes that take place in delivering strategic decision. We emphasize on the result of such interaction with a possible consequence of risk. It is also important that some types of risk are intrinsic to the particular actor under study. It is therefore important that, to adequately manage these risk, we need to model adequately both intrinsic and extrinsic factors of risk and their possible effect in decision making process. The other feature of this research is our attempt to capture intangible factors stated in earlier chapters which really affect the decisionability of a decision maker. We hope to bring to notice the importance of determining primarily the inherent risk factors in the EI process stages and how each of these culminate into overall risk that can hinder the delivery of strategic decision.

The word model has found its way into the heart of many fields in human endeavour. It is not uncommon to find words like ‘software models’, ‘system development model’, ‘model driven development’ amongst others. In this research, our focus will be based on such definitions that suit our intent as regards the word model. Thomas Kühne, (2005) in describing what a model is quoted Stachowiak (1973) in his proposition states that a model must consist of three main features: Mapping feature i.e. a model is based on an original, Reduction feature i.e. a model only reflects a (relevant) selection of the original’s properties, and Pragmatic features i.e. a model is an information on something (content, meaning), created by someone (sender), for somebody (receiver) and for some purpose (usage content)

Note that a good model need not be simple. However, the simplicity sometimes facilitates faster understanding but must not become a reason for not capturing the important behaviours of the system. For example, the models that are used to predict global warming are very complicated, as they must be to describe the very complicated Earth climate system. A model need not be mathematical, but it is a common phenomenon. A wind tunnel model is a physical model of a larger object. Animal models (e.g., lab rats) are often used to model human physiology.

Models developed for decision-making support represent the part of knowledge relevant to a decision making process, which can be used in a more efficient way in a computerized form than provided by other means. The meaning of this statement implies that the models, and the way they are used cannot be defined a priori. The knowledge useful for model-based representation is typically heterogeneous, therefore it is often initially represented by several models developed and tested separately before they are combined into one model, or in several larger models that are used as a system of models. In fact most of successful applications show that the requirements for model-based support change along a (typically long) process of development and application of a model. In some cases the modeling process (which starts with structuring the modeled problem) is even more important than the final version of a model, and the end-product of this process, which is a DSS into which the developed model is included.

A modeling process is typically composed of the following interlinked elements:

- Model (symbolic) specification,
- Collection and verification of data to be used for model instantiation,
- Model verification and testing,
- Generation of various instances of the model,
- Comparative analysis of each of these instances.

The modeling process is in fact also a learning process for both the model developers and future users. During this adaptive process the assumptions and requirements for the model are modified until a best possible fit between possibilities offered by the modeling technology, available knowledge, and data on one side and the needs of the users on the other side is achieved (Makowski, 2000). Different methods and tools are used for different elements of modeling process, which typically requires a substantial amount of resources. Therefore it is typically rational to reuse software that was developed (and tested) as a model for other applications.

The importance of this section lies in the fact that most natural occurrence to man are not always realistically feasible. Consequently, we create a model – which is a symbolic representation of the focus of research and use it to determine the suitability. This can be verified and tested if there are data, or simulated to have some experimented results. Described below are various models proposed in the course of this research. The arrangement however is not in any strict manner, but efforts are geared at presenting the models in the most orderly manner to facilitate maximum comprehension. We recalled that the main focus of this research is in information risk management in economic intelligence. It is therefore pertinent to identify what can be tagged to be 'at risk' with minimum confusion, point of occurrence of such risk and the possible impact on the overall process of decision making.

The rest of this chapter features the concepts of risk, risk factors (RFs) in economic intelligence which were later employed to formulate different models. We also expound on concepts like Translation Credibility (TC) and Decisionability. These two assist in the act of knowledge reconciliation (KNOWREM) between actors. Lastly, we present formal definitions for the KNOWREM model and hybridize ontology and fuzzy in what we called fuzzontology.

4.2 Risk Factor Model

With the notion of risks identified in previous chapter, we hereby look at what we tagged risk factor and formulate a model for its operation in EI systems. In this model, we tried to identify and group related tasks in the process of decision making and drawing from the earlier listed processes in EI. Attempt is also made to employ various stages either implicitly or explicitly taking into consideration an existing data warehouse, thus it is imperative to state here that, the focus is not principally towards designing a new data warehouse, but probably through the pointing out of certain inadequacies that can constitute risk, future design could be enhanced. We tagged figure 4.1 as the provisional schema for indirect-informing in economic intelligence.

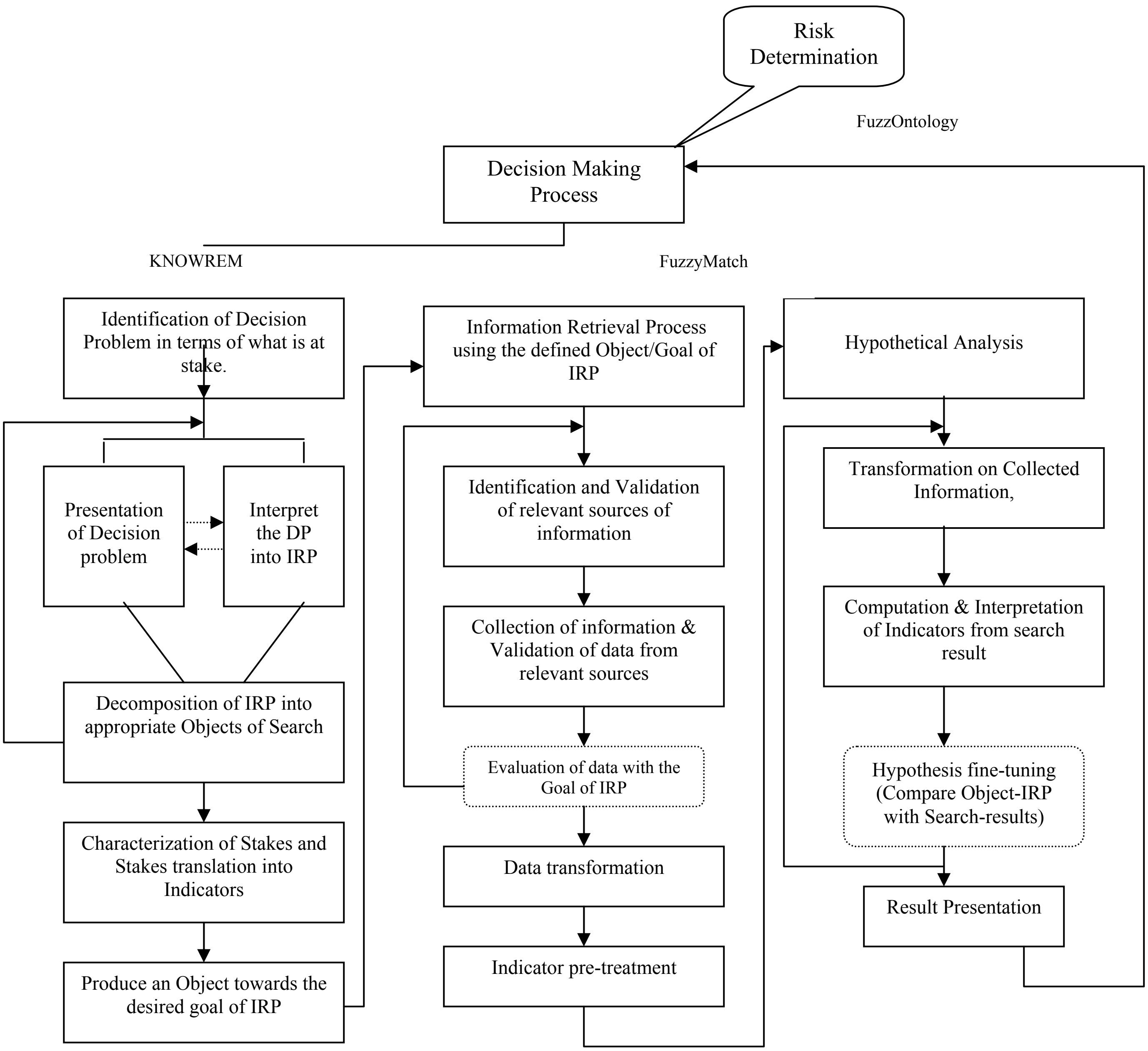


Figure 4.1: Provisional Schema for Indirect-Informing in EI processes.

We tried to partition figure 4.1 above into four distinct modules, to take the queue from the proposition in Bouaka & David, (2004), described as the Object – Signal – Hypothesis. Therefore we will refer frequently to the representation of the major concepts

(Users, Information and Process); and the mode of representing the interaction amongst the actors of EI (Decision maker, Watchers, Coordinator, and the Information (Data warehouse)) as presented in chapter three.

The naming convention employed in figure 4.1 below follows from what will be developed in the later section of this research report. We tagged the first stage as the knowledge reconciliation and this is determined by a quality factor we called translation credibility (TC). The information retrieval process employs as our proposition a fuzzy-based approach to resolve ambiguity in search resulting from dirty data. We expounded the ontological framework with fuzzy inference system to form fuzzontology which implies resolving what is captured via ontology the fuzzy way. The schema will thus be dissected and models proposed for each section will be discussed in this chapter.

To this end, figure 4.1is partitioned into modules and each of them is taking in turn while possible occurrences (not necessarily exhaustive) were listed. We partitioned the figure above into four distinct parts, described as the Object – Signal – Hypothesis. Therefore we will refer frequently to figures 1 – representing the major concepts (Users, Information and Process); 2 – representing the interaction amongst the actors of EI (Decision maker, Watchers, Infomediators, and the Information (Data warehouse)) and 3 – where we presented the global model of the interaction of the concepts in figure 4.1.

KNOWREM Model

In this module, a decision problem is already identified by the decision maker, and following, he is trying to liaise with other actors who are responsible for providing him with appropriate information towards the goal of making worthwhile decisions. We tagged this stage as the Knowledge reconciliation phase (KNOWREM). In figure 4.1, we represent possible errors from each actor’s interactions, thus, for the decision maker, we have $R_{dm.}$

Risk (R_{dm})

- Individual characteristics in terms cognitive style, personality traits and background.

- Wrong assumptions in the definition of objectives and stakes
- Inability to satisfactorily express himself to the understanding of the watchers and the infomediators
- Improper decomposition of the decision problems (DP) into information retrieval problem (IRP).
- Wrong characterization of stakes in terms of threat, vulnerability and consequences.

It is noteworthy that other actors involve in the process shared some of the possible acts/sources of risk along side with him.

FuzzyMatch Model

The main activity taking place in this module is the process of retrieval. However, this process is the result of the activities in module KNOWREM, thus, some of the possible source of risk hitherto listed in 1 will still be repeated for the actors here. In another way, this module houses the data that will be used for the process of decision making (data warehouse), consequently, it inherits the lapses that can result from design, administration and evolution of the data warehouse. Major actors of interest here will be the Watchers and the Infomediators, and the classification of risk will be R_{im} and R_{w}. We have hitherto enumerated the activities of the watchers and the infomediators, and that qualifies them for these operations.

Risk R_w and R_{im}:

- Inability to appropriately comprehend the notion of the decision maker in translating the decision problems into information retrieval problem.
- Wrong decomposition towards the formulation of the search objectives
- Partial or faulty characterization of stakes and its translation
- Search based on wrongly defined objectives
- Wrong information sources for even right goals
- Inability to validate sources of information
- Inaccessibility of data sources (classified information)
- Lack of validation of gotten information which could result from:
- Incomplete data (missing records/ fields; records of fields which by design are not captured

- Incorrectness (wrong, but sometimes right codes, wrong calculation and aggregation, duplicate records, wrong information captured into the source system).
- Incomprehensibility (multiple fields with one field, weird formatting to conserve disk spaces, unknown codes, spreadsheets and word processing files, many-to-many relationships and hierarchical files that allows multiple parents)
- Inconsistency (inconsistent use of codes, meaning of codes, overlapping codes, different codes with same meaning, inconsistent business rule, inconsistent aggregation, inconsistent grain of the most atomic information, inconsistent timing, use of attributes, date cut-offs, inconsistent use of nulls, spaces, empty values, lack of data referential integrity, and out of synch data)

Other problems could result from

- Naming conflict - i.e. people in different part of the same organization employing different terminology and names. These inconsistencies can results into *Homonyms* – same name for different concept (e.g. equipments to denote computer systems and somewhere else furniture) and *Synonyms* – same concepts referred to by different names: clients/customers.
- Structural conflicts - this type of conflicts result from different modeling construct in different schemas. Examples include: type conflicts; dependency conflicts, key conflicts and behavioural conflicts.

We can quickly notice that the bulk of the errors that can constitute risk in this module is the resultant operation from the data warehouse processes which, majorly results from the cleaning operations that consist of integration; transformation aimed at monitoring and presenting the most recent view of the data warehouse at every point in time.

The result of the search carried out based on the decision problem transformed into informational retrieval problem by the watchers and the infomediators to define the object of goals for the search has now been executed in module 2. The next stage of operation is the data transformation and treatment towards advising and presenting the result to the decision maker. Involved again in this module are the watchers and the infomediators that are required to complete the task of search with stepwise refinement towards interoperation with the indicators, this task precedes the decision making, and thus requires serious/consistent monitoring to ensure completeness or validity of the operation, with confidentiality of information and also guarantee its relevance of its applicability. Again we would have three types of risk here, some earlier stated with the risk that can result from the interactions of the two actors (watchers and infomediators), we thus have $R_{im,}$ R_w and $R_{wim.}$

Risk $R_{im,}$ R_w and $R_{wim.}$

- Wrong hypothetical analysis
- Inappropriate transformation of data (rounding-up of decimal places)
- Limited access to classified information
- Inappropriate processing (data mismatch, wrong context of application)
- Faulty interpretation of results due to lack of experience, personal intuition, environmental factors, e.t.c
- Inaccurate or partial presentation of facts -this is the intersection of the activities of both actors, intuitively, likely risk will result from undoing categorized as:
- omission of necessary facts
- commission on individual notions (not discussed but out of feelings)
- Assumption – things that either party should have done, but overlooked

FuzzOntological Model

Decision problems with defined objectives and stakes emanates from this module. He attempts to present his findings to the other actors in module KNOWREM towards the formulation of information retrieval problem aimed at solving the decision problem. Several other activities have taking place and the result of such is now coming back to him to assist him in arriving at the best decision based on the primary factor of the information available at his disposal, and other factors like his individual characteristics (cognitive style, personality, background e.t.c.), environmental parameters i.e. the total system composition for the information need and the organizational parameters i.e. signals or prompts necessitating the need for the information. Consequently, the decision maker has possibility of adding to the risk through the trio of individual characteristics, environmental parameters and organizational parameters. This type of inadequacies that can constitute risk is tagged R_{dm}. Apart from this personal risk, since all his decisions will be base on the information made available through the activities of other actors, he became the inheritor of all the hitherto mentioned risk. This is sequel to the fact that strategic management revolves round the environment; organization; information and the technology. *While some of the sources of risk earlier mentioned in module one still holds, the major focus in this module will be the inference(s) that can be drawn from the presented information to present right decision in the right place, at the right time, with the right cost in order to support and enhance the overall functionality of the organization.*

Risk R_{dm}

- Lack of experience and intuition
- Faulty cognitive ability with environmental and organizational parameters
- Wrong sense of judgments
- Wrong data analysis and presentation

On the whole, the overall risk from his act of decision will be the result of equation 1, which is given as $R = \sum (R_{i,} + R_w + R_{im} + R_{wim} + R_{dm})$. This equation however represents the maximum accruable risk from an act of decision making. With this notion, it implies that, the maximum level is not always the expected level of risk based on the activities of each module, and the result which is passed to the next.

4.3 Cognitive-Based Risk Factor Model for Decision making in EI

A fundamental issue is the question of whether decision quality should be judged by the processes (by which they are derived, e.g. EI) or by the outcome (and the associated consequences – the Introduction of Risk in strategic decision making). From the view point of the process-based approach – "almost all decisions are made under uncertainty". It therefore behooves that a decision is a 'bet', its consideration must therefore include the stakes and odds and not the outcome. However, it is believed that a good decision structure should inculcate possible outcomes as these affects the fulfillment of the decision maker's goal – "consequentialism". A typical example is referring to an operation as being successful after the death of the patient. This is unsatisfactory to most people concerned, because at the end of the day, the consequences (death) are more salient than the processes that produce them (Bullen & Sacks, 2003).

The challenges in identifying what constitute or otherwise of risk and the possible level of its effect are always difficult to identify and quantify. The assertion of Duffing, (2005) categorically depicts that there are different pattern of occurrence. Consequently, a one stop approach will be grossly inadequate for a dynamic environment like EI. With the above assertion on the importance of cognitive style in decision making process, coupled with the notion of risk presented in EI context by Onifade, (2008), we make bold to say that the effectiveness of such decision is a measure of the level of corresponding risk accruable from such process. In another dimension however, to effectively determine the level of risk inherent in a decision taken, we introduce another concept known as Risk Factor.

A Risk Factor (RF) as considered in this research is a concept, being, circumstance or a factor that increases the chances of being vulnerable to threat and consequences in a particular situation. These can be consequent upon things we do usually; inferred; or sometimes it could be situations and circumstances beyond our control e.g. ailment peculiar to old age. A risk factor here can thus be defined as any individual action(s), interaction(s), process, or inter/intra processes capable of undermining the overall performance/goal of an operation. RF can result from the action or inaction of a user(s), any of the processes hitherto stated, or on any of the activities performed during the life-cycle of data/information.

Several reasons have been brought forward for the justification that decision quality is a complex task that is difficult to be measured. Amongst these are the fact that it is viewed as being too broad and ill defined, making it almost impossible to derive a systematic or concise manner of attending to it. In another view, decision making is seen as multi-dimensional and thus the multiple classes requires different judgment criteria. Also of importance is the issue of "decisions about decisions" leading to what can be termed as the second order decision making. This could possibly escalate into the third, fourth-order decision until ad infinitum (Eden, 1994).

The fact that decision makers never act on the basis of actual reality, for it is elusive and never fully known, makes decision a complex issue. They act on the available purported representation of reality. We therefore agree with the submission that a decision situation involves taking inventory of what is known by the decision maker (data); what is yet known and must be acquired (information) and the application of rules of reasoning (knowledge) need be adequately reconciled for optimum efficiency (Hunt, et al., 1989, Wang, 2004). In an attempt to capture the importance of knowledge reconciliation in decision making process, we present a linear, qualitative cause-effect diagram known as the Fish – Bone architecture (Onifade, 2008).

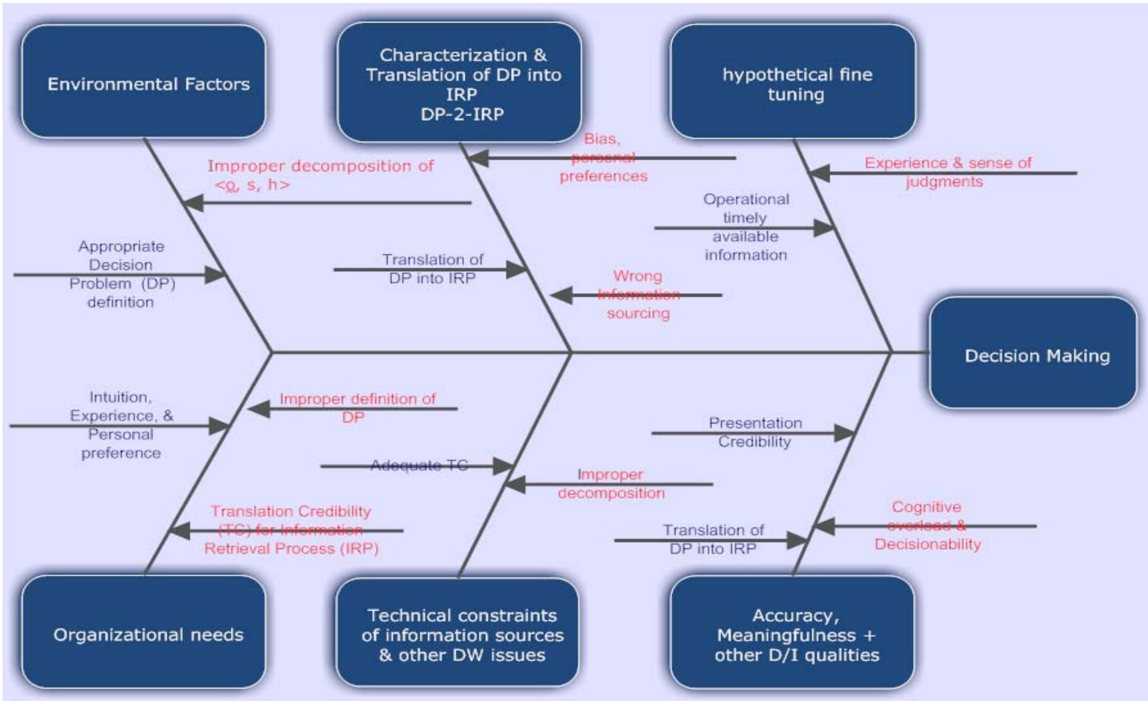


Figure 4.2: Cognitive-Based Risk Factor architecture for decision making in EI

Represented by a typical fish skeleton, each bone of the fish identifies a RFs (list not necessarily exhaustive) that can impair the quality of the result of the next stage under consideration (intrinsically – as a fundamental requirement of the result; contextually – judging from the circumstances and event under which the operation is performed, and possibly philosophically – concerned with or given to relative thoughts and resultant meaning) and/or required actions to implement the decision made. The multiplicity in occurrence of these concepts (Risk Factors) is based on the imprecision resulting from the difficulty to adequately transform and translate the decision problem (DP) as perceived and presented by the decision maker into its equivalent information retrieval problem (IRP) which would be utilized in the information search process. The most complex universal direct primary quality requirement of data/information values is Actionably Credible: that is the degree of credibility of a data/information value at which the user/decision maker is willing to take action (Bullen & Sacks, 2003).

Presentation of decision problems (as perceived by decision maker) to the other actors of EI towards reduction of risk in the whole process, must ensure an unquestionable level of knowledge reconciliation between and amongst participating entities (Onifade, 2008). Knowledge reconciliation (details in section 4.5 below) process must match some states with their equivalent attributes or associated meaning or otherwise with any state resulting for example into automatic triggering of a sequence of operation (state transitions) in a receiving numerically controlled device. From figure 4.2 above, attempt was made to integrate both the tangible (e.g. data warehouse issues) and intangible (e.g. root of biases) factors in a bid to properly highlight the importance of the knowledge reconciliatory process towards the goal of effective decision making. Hitherto, the two factors were taken in isolation and it was thus difficult to understand the integrated effect on the quality of data emanating from such process. The inclusion of these factors will facilitate a balanced approach in the detection and management of risk, identification of risk factors and its reduction based on appropriate integration of the cognitive factors, environmental factors, organizational needs, technical constraints of the information delivery systems and the root of biases and disinformation.

In the above, we have stressed the importance in realizing the urgency in the ability to adequately study the inherent risk factors in an attempt to facilitate decision making. Hitherto, major study in decision making take in isolation the various factors we looked at, but we are of the opinion that a robust result cannot be achieved in this regards.

In this section, the cognitive-based risk factor model that identifies possible point of occurrence of suck risk factors in the interaction between and amongst the participating actors in economic intelligence was presented. This was intertwined by the importance of the cognitive ability if the decision maker based on his ability to adequately discern the pattern of events necessitating decision making. On the whole, we present a cognitive based risk factors architecture for decision making in economic intelligence. The rational is to identify and factor both tangible and intangible components that can culminate into risk for decision maker. We are quick to point out that these followed closely from our submissions in chapter three about models of decision making and the effect of emotion and cognitive abilities of the decision maker.

4.4 Translation Credibility and Decisionability

Identifying the quality factor tagged "Interpretable during acquisition" among the direct primary quality requirement in all situations as the first mandatory quality factor is important. While we are in agreement with this submission, we make bold to say that since another actors are involved in the operation of search for information, it becomes expedient that the rationale and the requirements for the decision problem (entities informed) be well comprehended by the actor involved in the search operation (informing entities). Inability to fully comprehend the need behind the DP can constitute a grave consequence to the overall decision made. Consequently, before any consideration for interpretability during acquisition in indirect informing e.g. EI, the preceding process of deliberation between the entities informed and the informed entity must be ensured.

In the following subsections, we look at two fundamental concepts tagged translation credibility and decisionability. It is our belief that the level of understanding displayed by the decision maker in defining the decision problem with that which the watcher acquires during deliberation will determine the success of the operation. We therefore attempt to measure the level of understanding before the commencement of information retrieval process.

4.4.1Translation Credibility

With the sporadic growth in human knowledge stemming from experiences and reasoning abilities to create order in the mass of available information, it becomes imperative that decision resulting from such must be well guarded in the right direction. Decisionability of individual is determined by many factors. The list includes sense of judgment, experience in related situation and cognitive abilities amongst others. While decisions making are based on available information, the mode and method of getting such information is as important as the decision. Human are limited in expressive ability of their perception of the world which is the result of lack of lexical impression, incompleteness and inaccuracy in measurement (Sivanandam, et. al., 2007).

Several attempts have been made in the last decades to formulate a decision making model aimed at solving decision making issues. Ge & Helfert (2006) reviewed a long list of decision making models based on different perspectives. Gackowski, (2006) described *actionably credible* as the factor that helps check for the presence or options amongst disinformation, misinformation and valid information. Disinformation – intentionally misinforms. Misinformation – unintentionally misrepresents reality. Valid information faithfully represents or reflects reality. He described credibility of data/information as a complex function of 19 indirect attributes of quality, these include: credibility of the source from where they are derived, variety of independent sources available, quality of mapping, credibility of their presentation e.t.c. With the above submission, we reiterate the fact that, the process of information retrieval which is an important stage in EI process and informing systems should not commenced until the validity of the translation credibility can be ascertained.

We define Translational Credibility (TC) *as the level to which the decision problem being identified by the decision maker in terms of* ***object****,* ***signal*** *and* ***hypothesis*** *can be adequately represented and translated into a proper object of search (IRP). In another word, we say, the degree to which the notion of the decision maker's mind set can be appropriately mapped towards the specification for the retrieval processes.*

<u>*Definition:*</u> We define a property called *translation credibility* as the result of an evaluation function γ that determines whether the information search should proceed based on the attempt by the decision maker to inform the other actors (watcher and others) responsible for the process of information retrieval. It is not a rigid function; however, the 'satisfiability' as determined by the concerned user triggers the next stage of information retrieval. Thus:

Suppose there are elements τ_i (i = 1, 2, …, n) of $\Re$ (set of Information Retrieval Problem - IRP) which are together a subset of the universal set Ɗ (Decision Problem – DP) represented for all elements $\partial \in$ Ɗ by its characteristic function, then:

$$\varphi_{\Re}(\partial) \quad \begin{cases} 1, & \partial \in \Re \\ 0, & \partial \notin \Re \text{ (otherwise)} \end{cases} \tag{4.1}$$

Where the decision problem is represented by ∂, and $\Re$ is the corresponding information retrieval problem (IRP). $\varphi_{\Re}(\partial)$ is the membership of IRP, and Ɗ is the entire decision problem set.

The above becomes important due to the imprecision and information granularity. With the introduction of the degree of membership of the element of the IRP, we can therefore implies the set Ɗ contains an object τ to the degree d(τ), i.e. d(τ) = Degree($\tau \in$ Ɗ). A typical example will be the recruitment exercise for "tall people". The phrase will have different meaning to different individual considering the query. To Mr. X, it can connote a height below or equals to 4 feet, another Mr. Y can assume the phrase implies somebody with height 5.5'' feet. Mr. Z however, may believe that if you are not up to 6 feet, you are not tall. From the above, it is apparent that the same phrase "tall people" does not convey the same meaning to different individuals X, Y, and Z, i.e. there is variability in the interpretation given to it. The phrase "tall people" can thus be referred to as the linguistic variable, commonly employed by human beings in the presentation and description of situation. The result is different level of uncertainty and vagueness in comprehension.

The value of the translation credibility γ assists in monitoring the level of inference the other actors in economic intelligence concept have been able to deduce from the problem posed by the decision maker. As the equation (4.1) above shows, it is the determinant level for further processes towards a credible decision making based on the available data all other factors being equal. We submit again that the omission of this stage in the evaluation of D/I quality makes the operation incomplete based on the fact that information derived are usually employed for the furtherance of the organization via decision making.

This becomes imperative going by the fact that w*ithin the concept of Economic Intelligence, division of labour has resulted into specialized classes of users dealing with information acquisition and its presentation, decision making and the acting agents.* Unfortunately on most occasions, there is little or no control on information obtained from such compartmentalized operation for information search. Thus, information provided are usually of varying degree of quality and can therefore represent "out of data", "unavailability", "incorrect", or sometime gives a potentially dangerous information e.g. expiring date of a drug or beverage that is meant for consumption. In figure 4.3, we present a conceptual diagram of the model for TC.

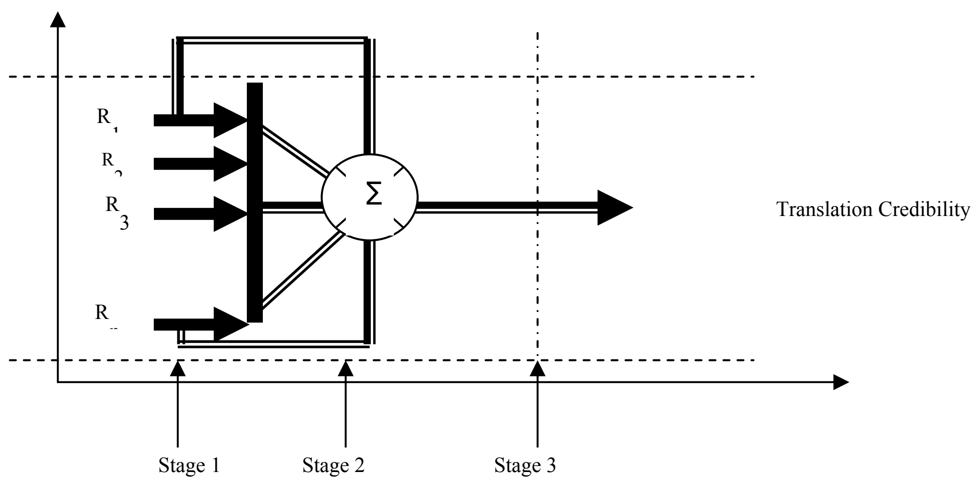


Figure 4.3: Conceptual diagram for Integrated Fuzzy Algorithm

Identification of Concepts for TC/KNOWREM

We present a typical example from our research thus far using the stated principles of computing with words to define our concepts and partitioning them into appropriate world of discourse. We recalled that concepts identified are key factors or characteristics of the system to be modeled. It consists of the following amongst others: the inputs, outputs, variables, states, events, actions, goals and trends which define the would be system. The following were identified at the first part of the research in determining the translation credibility between the decision maker and the watcher in the process of knowledge reconciliation (KNOWREM).

Actors	&	Risk Factors (RFs)
Decision Maker		Organizational needs (ORN)
Watcher, etc		Environmental factors (ENF)
		Intuition & experience (INE)
		Personal preference (PEP)
		Disinformation (DIS)
		Misinformation (MIS)
		Level of Understanding (LOU)

In the listing above, we have identified some important RFs which can undermine the process of knowledge reconciliation. They are termed the fuzzy input variables. We referred to them as risk factors because, they are not risk per se until they have fully exercised the vulnerability of the system to cause it to behave contrary to expectations. Two other concepts itemized are the actors which include: decision maker and the watcher. The fuzzy input variables (RFs) will be assigned membership functions (MF) to determine their inclusion or belongingness into the world of discourse. We shall represent these MF with linguistic values or descriptors of "High", "Medium" and "Low". These three take the abbreviation (H, M, L). Thus the sets of the RFs and MF identified are:

RF = {ORN, ENF, INE, PEP, DIS, MIS, LOU}

MF = {H, M, L}

Recalling that we are using Mamdani model which is rule based, then our rules formulation will follow the general pattern of the 'if' specifying the antecedents and the 'then' which also connotes the consequents. Below, we present a tabular representation of this model for the rule generation.

Antecedents					Consequent (TC)
ORN	ENF	INE	BIA	LOU	
H	H	H	H	H	GOOD
M	M	M	M	M	AVERAGE
L	L	L	L	L	BAD

Table 4.1: Tabular representation of the antecedents and the consequent

A quick look at the fuzzy input variables in table 4.1 shows an alteration in the number earlier listed. The rationale is born out of the fact that, while we want the listing to be very explicit, using same in the modeling can generate some level of complexities on one part, and in another could lead to a situation where there are overblown numbers of rules. While table 4.1 is kept simple, it is interesting to note that, there is no confusion in the consequent section, as they are just listed to give the representational view. In the real sense of it during the operation, each of the fuzzy input variables is utilized in the selected partition style for the model. This partition style assists in defining the fuzzy region and the corresponding membership functions. Consequently, we show that a 10 – input fuzzy model having 2 membership functions will result into $2^{10} = 1024$ fuzzy if-then-rules. This phenomenon is known as "course of dimensionality". It derails the performance of the system gracefully. To this end, we represent the trio of 'PEP', 'DIS', and 'MIS' as "ROOT OF BIASES", shortened as 'BIA'. The choice of 'H', 'M', "L' is an attempt to simplify the delineation of the membership functions used for the work.

Flowchart of the operations

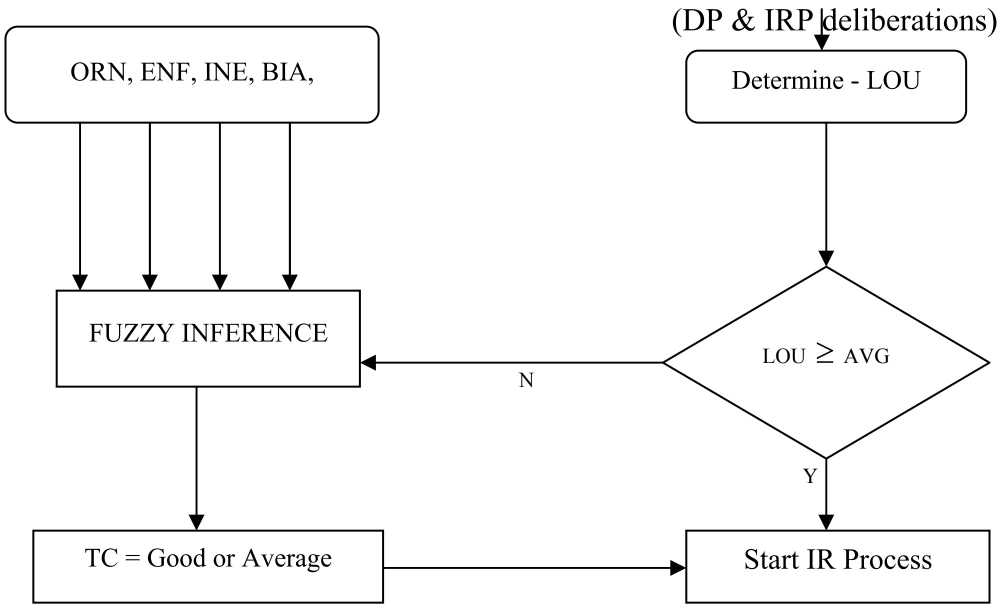


Figure 4.4: Determining the level of TC before IR process through the RFs effects.

In describing the above, we try to present a picture or example of its operation in real life situation. This scenario involves a newly trained dentist (Dental doctor) attending to a patient in the hospital. The following is based on probable dialogues which can ensue between them.

Scenarios Description

In what follows, we present the modus operandi for the transitions between TC determination and the commencement of information retrieval process for decision making.

Elderly people (EP) usually have problem describing their dental ***pain*** experience to the doctor (Dr). Classification of the pain is usually of the form "***very painful***", "***moderately painful***" and "***not painful***".

Conversation

Dr: what is the problem, madam?

EP: I have problem with my teeth

Dr: what type of problem do you have and around where?

EP: I have “***pain***” at the “***upper part***” on the right “***towards the back***”

Dr: how would you describe the pain?

(The Dr begins to use some instrument to touch the suspected areas and asked the EP to describe the nature of pain experienced in those areas touched.)

Dr: how is it here?

EP: no, there is “***no pain***” there

Dr: what about here?

EP: it’s “***moderately painful”*** there

Dr: and here?

EP: oh! Yes, it’s “***very painful***” there, yes, yes // EP almost screaming

The Dr. representing the decision maker (DM) in the above scenario employed the dialogue to arrive at the formulation of decision problem (DP) to be solved in terms of <o, s, h>, and the necessary information that need be acquired for the resolution of the problem. The above is best described as a decision problem, thus the Dr. must be able to:

1. understand the problem by dialoging with the EP
2. use the gathered information to establish the nature of the problem
3. diagnose and prescribe to remedy the situation

Step 1:

We can thus see that (1) represents problems identification in term of <o, s, h>. This step is very important because it affords the Dr the opportunity to understand the problem without which he cannot move successfully ahead. Here, his competence, sense of judgment, intuition and experience is brought into play. In relation to our earlier research, for the first step, we tagged the first step as Knowledge Reconciliation (KR), having identified some risk factors (RFs). Thus, determining the level of understanding and the conceptions gathered during the process. Again, we made bold to say that the success or otherwise of this stage is crucial to the overall diagnosis of the problem, identification of the decision needs, and subsequent solutions to be proffered. Furthermore, we designate a term which we called Translation Credibility (TC) to determine the level of understanding acquired during the KNOWREM process. It satisfiability alone is recommended to trigger the next stage which is the information retrieval process. We submitted that, improper decomposition, lack of understanding and inappropriate object of search can be highly detrimental to the decision process.

Step 2:

The Dr. employs the conviction based on the level of TC, and the formulated object of search for the information retrieval process. In the formulation of the search object, the signals derived from the dialogue will be employed to formulate and refine his subsequent queries. In all, we believed the inference formulation will take an "if rule" format which is peculiar to fuzzy inference systems. The form is such that:

*If "**pain**" is "**moderate**" and "**age**" is "**elderly**" and "**located**" on the "**far right**" corner of the mouth, and "…", **then** apply …, (and other solution).*

Step 3

Step 3 deals particularly with the utilization of gathered information to arrive at a logical conclusion to the hitherto identified problem.

Human mode of communication inherently habours high risk of ambiguity and imprecision as clearly shown in the discussion between the elderly person and the doctor. Unfortunately, it is very difficult for conventional systems to handle these factors as their design can only accommodate accurate information devoid of imprecision and possibly expressed in bi-logic format. It therefore behooves that dealing with this issue with a bid to minimize resultant risk can be achieved if there is provision for definition of the above described concept – Translation Credibility. Although it does not singularly eliminate these risks, but it will serve as a right step in the right direction to determine the TC following the pattern of chart in figure 4.4. In subsequent section, we take a look at another factor that can undermine the process of decision making tagged 'decisionability'.

4.4.2 Decisionability

The importance of appropriate and adequate information toward decision making cannot be overemphasized. Considering the sheer volume of available information to information consumer and the rapidly advancing information management technologies facilitating corporate firms and organization to manage large and complex data sources aimed at exploring new opportunities towards strategic decision making. The age of inadequate information is gradually fading off while unconsciously introducing another concept known as "Information Overload". Using adequate and timely available information for delivering strategic decisions is the focus of economic intelligence, however with ever increasing volume of information made available to the decision maker, the risk of indecision, forgetfulness, amongst others result. The consequent could be disastrous as the need arises for decision based on the interrelationship amongst the trio of decision situation, decision maker and the decision process in the face of information overload. This section attempts to establish the need for usage of 'fine-tuned' information in taking strategic decisions

"*Deciding not to decide is not the same as indecision*" goes an old saying. There could be several factors that undermine the process of decision making. One of such factors earlier identified was lack of access to adequate information, but the present story has changed with the advent of the internet and other private and corporate information delivery sources. O'Brien (2003, 2004) defines "*information as data placed in a meaningful and useful context*", the Oxford English Dictionary (OED) (1989) described *information as knowledge concerning some particular fact, subject, or events* i.e. information is seen as knowledge. A decision environment can be defined as a collection of information, alternatives, values and preferences available during the time of decision. Thus in an ideal environment for decision making, it's expected that all possible information will be available, in possible best accurate proportion, and alternatives (Harris, 1998). The above is however constrained based on the fact that time and efforts to grain information or identify possible alternatives are limited. Decision environment is presented to be of incremental rate, since new information and alternatives may spring up even after the decision might have been made. The identification of the above is followed by the attempt to formulate a response to the situation that fall in line with stated goals. The level of the decision maker's consciousness is brought into play at this point.

We describe *Decisionability as the performance of the decision maker based on utilization of available information*. The impact of human element in the "learning process" prior to transformation of information into knowledge was properly presented in Davies & Ledington (1991). Its dynamic and circular properties were also established in Knox, (2007). Burgess, et al., (2007) evaluated the impact of information overload in determining quality criteria to assist in information search. "*The term information or cognitive overload describes situations in which the individual is no longer able to integrate new information for decision-making, due to the great amount of information he or she is exposed to. He or she can no longer productively use the quantity of information in the available time scale.* Consequently, decision quality, efficiency, and even well-being may be reduced" Eppler, (2001). The above phenomenon has been referred to as 'cognitive overload' or 'information fatigue syndrome'. A succinct summary of the above thus implies, "even with highest form of information quality, information flow should be adequately controlled to dissuade cognitive overload".

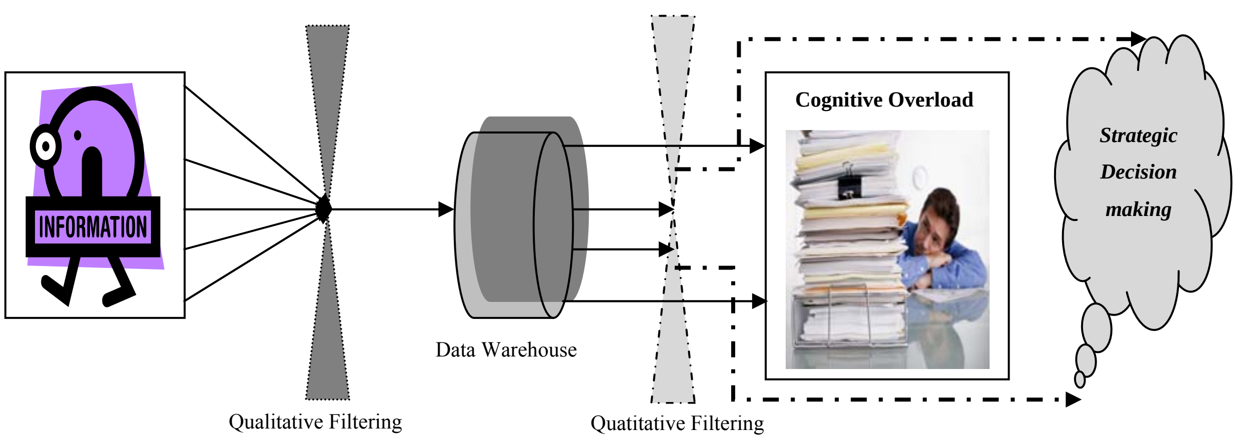


Figure 4.5: Effect of Information Flow, Quality, and Overload on Decision Making

The tendency for extensive information search is the most probable option in decision making. However, such attempts can results into one or more of the following scenarios:

- *Delay in decision because of time required to obtain and process the extra information*. This delay could impair the effectiveness of the decision or solution
- *Information overload or cognitive overload could result*. This situation is the phenomenon in which so much information is available to the extent that "decisionability" declines because of the information in its entirety cannot be appropriately managed. A consequence of this is forgetfulness if information are crashed into memory in short period of time e.g. "*crashed course towards an important debate like US presidential election debate*"
- *Selective use of information will occur* – the decision maker will employ only facts supporting his preconceived ideas or solutions
- *Mental fatigue results* – slower work or low quality decision
- *Decision fatigues* – a state of worn-out, not wanting to take decision again. This finally results into fast, careless or total decision paralysis.

The quantity of information that can be handled meaningfully by human mind is limited. It is therefore pertinent that information be consciously selected, otherwise processing will result into biases towards the earliest information, thereafter, the mind attempts to ignore subsequent information or forget earlier ones (Harris, 1998). In figure 4.5, we present a two-staged filtrations aimed at supporting earlier submission of Harris (1998), Bullen & Sacks, (2003) and Burgess, et al., (2007).

The first stage features what is described as information flow, a phenomenon that is constantly on the increase sequel to the advent and adoption of information technology. This is followed by a *qualitative filtering* into organization's data warehouses based on desired quality attributes (Lee, et al., 2002; Strong, et al., 1997 & Gackowski, 2006, 2007). The complex nature of the data warehouse causes the components to frequently evolve independent of each other. There is possibility of new materialized views creation and old ones updated. Some earlier data sources may be out of use, new sources may be added. There could be product evolution or update, changes in design choices is also a possibility. The qualities of a data warehouse therefore depend on the quality of the sources; the quality of the extraction process and the quality of the data warehouse components all of which affect the process and quality of retrieved information.

The figure possesses some colourations which are not meant for aesthetic but to clearly distinguish the operations and the results. Thus, if the second filtration is not performed, the tendency is pronounced to have information or cognitive overload. The second stage in more refined shade depicts another filtration to dissuade cognitive overload.

With increasing volume of information to users come problems like specificity of database queries, requirements for users to explicitly state their information need and information or cognitive overload. Several researches have looked at information flow and information quality but not much work was recorded to determine the effect of the trio. Considering the flow, quality and the overload in EI is a serious challenge. This is sequel to the fact that strategic decision making is the sole aim of EI, thus in a bid to forestall this, we juxtaposed the trio and depict the effect of information/cognitive overload on decision maker through a two-stage filtrations. The ensuing risk is better imagined than allowed to occur, as it can practically result into both mental and decision fatigue in which case the decision maker's act becomes irrational.

As more and more information is made available for decision process, the possibility will be on the increase for fatigue resulting into wrong, delayed decision or sometimes indecision. We believed the two stage filtrations will go a long way to address the issues raised in this regard. In the next section, we present the model that facilitates deliberation between concerned actors in EI systems tagged KNOWREM.

4.5 Knowledge Reconciliation and Ontological Framework in Economic Intelligence

Knowledge is seen as legitimate and meaningful resources that strengthens the overall management performance. As a result, knowledge management is viewed as a sine qua non towards creation, storage, sharing, and reusing of the organization's knowledge, employing advances in today's technology (Allan, et al., 2007). Economic Intelligence

(EI) is saddled with usage of timely availability of information towards strategic decision making. However, while decision making is based on available information, it has been observed with concern that reconciling the "need for decision" and subsequent "search for relevant information" poses a serious threat to the overall decision because of some intangible factor that are difficult to be expressed culminating into non-quality of retrieved data, and sometime time taken to adequately mapped the decision maker's "mind-set" into an appropriate object for information retrieval. Ontology potentially enable automated knowledge sharing and reuse among both human and computer agents; this is facilitated based on their ability to interweave human and machine understanding through formal and real-world semantics.

Decision making is seen as one of the basic cognitive process of human behaviour, it involves the choice of preferred options or courses of actions from amongst sets of alternatives defined by certain criteria (Redman, 1998). Wang, et al., (2004) presents the process of decision making from the basis of Layered Reference Model of the Brain (LRMB) which identifies decision making as one of the 37 fundamental cognitive processes. There are two principal categories:

- Descriptive theories are based on empirical observation and on experimental studies of choices behaviour.
- Normative theory assumes a rational decision maker that follows well defined, preferences that obey certain axioms of rational behaviour; these are the Utility paradigm and Bayesian theory.

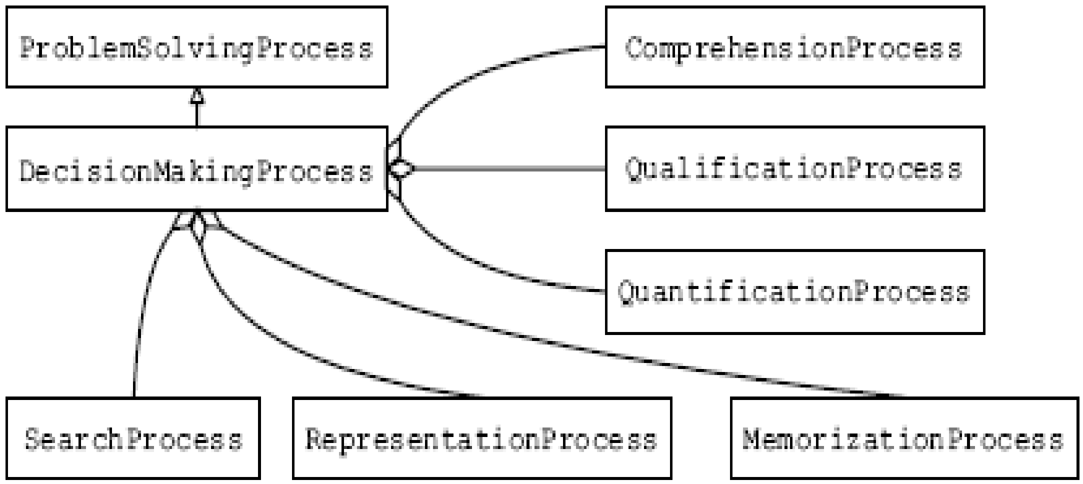


Figure 4.6: The relationship between decision-making process and other processes in LRMB (Wang, et al., 2004)

In figure 4.6 above as the caption indicates, it shows the relationship between the act of decision taking and other processes in LRMB (Wang, 2004). The cognitive capacities of decision makers may be of various capacities, based on their level of exposure, environment, and other factors. However, it is no gain saying to assert that the core cognitive processes of human brain shares/exhibits similar and recursive characteristics and mechanisms (Onifade, 2008).

There is complexity of problem definition derives from the differences in interpretation of each key actor in a given situation of the event and to the explanations of the influences between events. In the formal part, it is important to recognize that personal values play a part in interpretation, and the latter part opined that individuals brings to bear different experiences and wisdom that has created different belief systems (Gackowski, 2006b). Majority of work on decision making has been normative and dominated be rationalist perspectives.

With the above assertion on the importance of cognitive style in decision making process, coupled with the notion of risk presented in EI context by Onifade, et al., (2008) we make bold to say that the effectiveness of such decision is a measure of the level of corresponding risk accruable from such process. In another dimension however, to effectively determine the level of risk inherent in a decision taken, we introduce another concept known as Risk Factor in (Onifade, 2008). We have earlier on described a *Risk Factor* (RF) is a concept, being, circumstance or a factor that increases the chances of being vulnerable to threat and consequences in a particular situation. These can result from things we do usually; inferred; or sometimes it could be situations and circumstances beyond our control e.g. ailment peculiar to old age. We therefore identify a risk factor as any individual action(s); interaction(s); process; or inter/intra processes capable of undermining the overall performance/goal of an operation. RF can result from the action or inaction of a user(s), any of the processes hitherto stated, or on any of the activities performed during the life-cycle of data/information. These risk factors are embedded in the "Information" as represented by $\mathcal{R}_i$; "the Watcher" represented by $\mathcal{R}_w$; "Coordinator and other actors" given by $\mathcal{R}_{cor}$; the interaction between the "Coordinator and Watcher" given by $\mathcal{R}_{wcor}$; and the "Decision maker" which is also represented by $\mathcal{R}_{dm}$ as presented in section I. consequently, we can defined the following:

$\mathcal{R}_i$ - risk accruable from information acquisition based on sources heterogeneity, representation, extraction and utilization. The RFs will lie in the transformation, validation and utilization processes. E.g. different date formats (mm/dd/yy Vs dd/mm/yy)

$\mathcal{R}_{w'}$ – that is the risk resulting from the mapping of the decision problems and its corresponding information retrieval work undertaking by the watcher. E.g. "we need a car?' what other properties helps qualify the car?

$\mathcal{R}_{cor}$ – to represent the possible inadequacies that could result from coordinator activities and constitute a risk to the discharge of the duties of the decision maker. E.g. "Operationally timely availability of information or task execution".

$\mathcal{R}_{wcor}$ – to denote the risk which can result from the joint action/inactions of the watchers and the coordinators? Caution must be taken not to duplicate the effect of the RF already noted. E.g. Assumption, Omission and Commission

$\mathcal{R}_{dm}$ - risk accruable from the point of definition and presentation of decision problem, to its interpretation and the final process of taking inference from available information. E.g. Improper observation & presentation of the decision problem

Delivering strategic decisions, high rate performance and complex decision making are cogent features in Economic Intelligence; it is therefore no gain saying to assert that appropriate method of information-sourcing for knowledge application is of utmost importance.

Ontology has been defined as formal vocabularies, shared by a group of people/individual, interested in a specific domain. Ontology potentially enable automated knowledge sharing and reuse among both human and computer agents; this is facilitated based on their ability to interweave human and machine understanding through formal and real-world semantics. The operation is further enhanced with a solid methodology that guarantees the collaborative engineering process. In summary the key advantages of ontology are: knowledge sharing; reusability of knowledge and their being machine understandable.

It was argued that the basis for ontologically driven information system was the desire to incorporate knowledge from different domain into a single frame work. Ontological application in risk simulation brings about the conceptual requirement of "reusability". This is sequel to the fact that models are specified by humans to embody domain knowledge, characterized by ambiguity, thus ontology provides a profound solution in bridging the semantic-gap between the knowledge-space and the simulation model (Jung, 2004).

Presented below in figure 4.7 is our Ontological Framework for knowledge reconciliation (KNOWREM Model aimed at facilitating improved classification of stakes. The figure represents the first four major steps in the distinguished but interleaved process stages of economic intelligence (David & Thiery, 2003). However, their interrelationship can sometimes results into non- quality data and further constitute a risk factor to the whole process. In a bid to nip this at the bud, the knowledge reconciliatory process which was highlighted as the first major process precedes every other operations.

The framework attempt to map this with the EI domain alongside the domain ontology developed for the framework. The essence of the domain ontology is to facilitates the act of reusability and enhance the speed of processing vis a vis the credibility of the overall operation. The next stage which involves decomposition of the identified objects to facilitate translation into appropriate object of search, cross-matching it with existing facts and knowledge, once again for reliability of reconciliation.

The characterization and translation of stakes hitherto identified take into cognizance the logical and the user defined views aimed at presenting a balanced interpretation devoid of misrepresentation and minimum ambiguity. At this point, ill-defined, ambiguous and fuzzy expressions will be handled by fuzzy decision support to aid the characterization and translation of stakes as identified, presented and expected to be translated (reconciliation) to facilitate appropriate data retrieval, on time, credible and of required quality.

KNOWREM metadata repository is expected to store the historical events of previous translation and reconciliation processes. On one hand, it will assist in speeding up related tasks that has been once successfully conducted and also allow for reusability, bringing about great gain in time for the global operation of decision making. Consequently, we present our ontological framework for knowledge reconciliation process (KNOWREM) in the figure below.

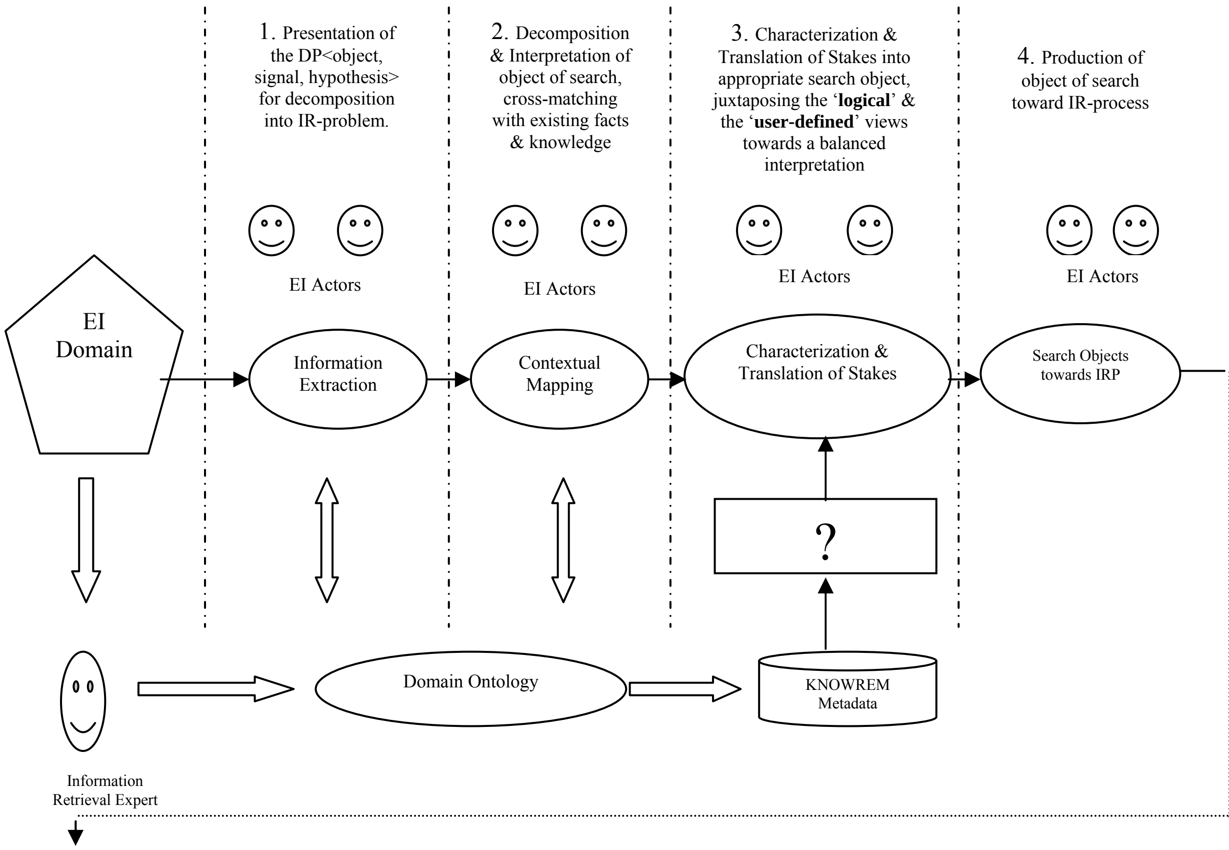


Figure 4.7: Ontological Framework for KNOWREM Model

The above presents a flexibly-defined and reusable framework, pedaled by ontology, and capable of capturing knowledge and modeling purposes. The framework consists of a metadata repository, where all earlier definitions and risk-factor patterns earlier captured with interpretations were kept, allowing periodical update as more risk factors are identified in the knowledge reconciliation process. It also has the domain ontology, the EI domain, and possible interactions within and amongst these entities. In the following section, we continue with an attempt to buttress the result hitherto presented above. The attempt lies in the formalization of some of the terms employed for KNOWREM.

Formal Definition of Knowledge Reconciliation

Earlier on we define what constitute a model and its attributes. We also made mention of the fact that a model could be pictorial or mathematical. Mathematical models are aimed at creating enduring foundation for hitherto proposed models e.g. knowledge reconciliation (KNOWREM) model. This is amongst other justifications expected to present a much deeper understanding than the contextual provisions of other types, and facilitates ease of automating the framework. In the meantime, we adopt the definitions and models presented in (Wang, et al., 2004). However, rigorous attempts were made to improve, extend these definitions to include other factors deemed necessary for knowledge reconciliation in economic intelligence.

The axiom of selection (or choice) (Wang, et al., 2004) states that – there exists a selection function for any nonempty collection of nonempty disjoint alternatives. Based on this, we have:

Definition 1.1 Suppose $\{A_i \mid i \in I\}$ be the collection of disjoint sets $A_i \subseteq U$ and $A_i \neq \theta$, a function

$$f\text{: } \{A_i\} \rightarrow A_i,\ i \in I \qquad (4.2)$$

Is the *choice function* if $f(A_i) = a_i$, $a_i \in A_i$. or an element $a_i \in A_i$ may be chosen by f. On the basis of the choice function and the axiom of selection, a decision can thus be defined:

Definition 1.2: A decision d, is a selected alternative a_i, from a nonempty set of alternatives A_i, $A_i \cap A_j = \theta$, $i \neq j$, $i, j \in I$, $A_i \subseteq U$, based on a given set of criteria, i.e.:

$$d = f(A_i, C) \qquad (4.3)$$

$$= f\text{: } A_i \times C \rightarrow A_i,\ i \in I,\ A_i \subseteq U,\ A_i \neq \theta \qquad (4.4)$$

Where C is the set of criteria for the selection of alternatives, and $\times$ represents a Cartesian product. It should be pointed out that the criteria C can be a simple one or a complex one. The latter is the combination of a number of joint criteria depending on multiple factors.

Definition 1.3: Decision making is the process of decision selection from available alternatives against the chosen criteria for a given decision goal.

With respect to definition 1.2, the number of possible decisions, n, can be determined by the sizes of Ai, and C, i.e.:

$$\#(Ai) \bullet \#(C) \tag{4.5}$$

In this regards, # is the cardinal calculus on sets. Considering from the above equation, in case $\#(Ai) = 0$ and/or $\#(C) = 0$, implies there is no derivable decision for such occasion. In relation to the above mathematical definitions, we present the cognitive process of decision making and use the fundamental of Object-Attribute-Relation (OAR) model to describe our interest on the concepts of decision problem; Decision Problem Translation (DPT) into Information Retrieval Problem (IRP). Sequel to the diagrammatic representation of the relationship between decision-making process and the other major processes in LRMB as shown in figure 4.6, it was implied that the cognitive model of human memory, particularly the long-term memory (LTM), can be described by two fundamental artifacts.

1. Objects: The abstraction of external entities and internal concepts. There are also sub-objects known as attributes, which are used to denote detailed properties and characteristics of an object.
2. Relations: Connections and relationships between object-object, object-attributes, and attribute-attribute.

The above discussion led to the development of what was referred to as, an Object-Attribute-Relation (OAR) model of the memory which was described as follows:

$$OAR = \langle o, A, R\rangle \tag{4.6}$$

Where o is a given object identified by an abstract name, A is a set of attributes for characterizing the object, and R is the set of relations between the object and other objects or attributes of them.

Decision making process has been listed among the processes in the LRMB; thus, we adopt the pattern of definitions above presented at first to present our notion of decision problem, the translation process and the criteria we called translation-credibility.

Definition 1.4: A decision problem within the context of Economic Intelligence (EI) can be defined in term of three principal attributes: the object, signal, and the hypothesis.

The *Objects* are the perceptions and abstractions of both internal concepts and external entities, brought about by the presence of some *Signal(s)* generated by intuition, perception, and experience; which could be generated internally or externally, and from which some inferences in form of *Hypothesis* could be made as to the effect of the perceived objects. We can therefore generate an *OSH* model, representing the decision problem.

$$OSH = <\text{o, S, H}> \qquad (4.7)$$

Suppose we represent the decision problem by δ, o as the *object* whose identification results from the presence of some internal and external set *signals*, and by which, some sets of possible *hypothesis* could be deducted.

Let $\{o \in O,\ \alpha \in S, \text{ and } \beta \in H\}$, then we can define a function Φ as follows:

$$\Phi : \{\text{OSH}\} \rightarrow <\text{o}, \alpha, \beta> \qquad (4.8)$$

As the function that maps the prompt for decision to its equivalent properties in the OSH-model.

It therefore suffices to define decision problem from the basis of (vii) to have:

$$\Phi : \{\Psi\} \rightarrow <\text{o}, \alpha, \beta> \qquad (4.9)$$

Translating the decision problem into an information retrieval problem (object of search) is the next task. We reiterate here, that, this is usually a stage which has not been given much attention, and which could mare or make the success of any meaningful decision. This activity becomes imperative noting that sizes of many organizations has brought about division of labour resulting into specialized actors in every stage of the organization, which parallels with our concept in EI. Thus, the success of this exercise is very crucial to the decision making process.

Definition 1.5: Decision problem translation or knowledge reconciliation is defined as the transformational/reconciliatory mapping, which attempt to generate appropriate objects of search from the decision problem. Hence, let a function Γ be the transformational/reconciliatory mapping, we write:

$$\Gamma : \Psi \rightarrow \Re \qquad (4.10)$$

Where $\Re$ is the corresponding translated decision problem (object of search). We declare that Γ is a bijective mapping. i.e.

$$\text{For}\, \Gamma : \Psi \rightarrow \Re ,\, \exists\, \partial , \tau ,\, \ni\, \partial \in \Psi \wedge \tau \in \Re$$

$$\Rightarrow \Gamma(\partial) = \tau ,\, \forall\, \partial \in \Psi \wedge \tau \in \Re \qquad (4.11)$$

The importance of equation (4.10) is to remove any form of ambiguity in the process of reconciliation. We believed that this process could be detrimental if not properly handled. It is therefore without gain saying that the formal definition will provide a better platform for establishing a lasting model for the process.

Interpretation: A Basis for Misconception in Knowledge Reconciliation

It is not a common place for a piece of data to generate similar 'factual' meaning when interpreted by different individuals. However, derivable inferences would be likely different more widely in 'meaning' of the data for different individual based on his/her associations, and/or possible consequences depending on the uniqueness of 'S'. Communication and intention is context-dependent. Interpretation of context continually evolves with time thus having great influence on sense-making and communication (by Wittgenstein, (1963) and quoted in Bednar & Welch, (2008). Callaos & Callaos, (2002) typified information conception from two different perspectives: subjective and objective. This was necessitated by the authors concern for the incoherent homonyms resulting into divergent and dissimilar references in the usage pattern of the term.

In demonstrating the workability of the 'infological equation', Langefors; (1966) was of the opinion that those who are to interpret data in order to inform themselves must be viewed as part of the system. Using the equation I = i(D, S, t), where 'I' is the information (knowledge) produced by a person from the data 'D' alongside with pre-knowledge 'S' through an interpretation process 'i' in interval 't'. Bednar & Welch, (2008) reflecting on the opinion of Langefors made by Schutz, (1967) reiterated the impossible nature of communicating "meaning" between people. This was sequel to the fact that simply transmitting data will not lead to communication of shared understanding knowing fully well that 'i' and 'S' cannot be assumed to be common. To this end, communication can only be seen to approach success most closely where individuals interpreting the same data belong to a group with possible vested professional interest. This is sequel to the fact that every act of interpretation does not necessarily invoke the entire 'S' attribute to every individual, thereby creating room for some aspect of shared experience leading to similarities in the 'i' among group members.

Information has been widely defined as "interpreted data" thus bringing in a reference to the 'i' as depicted in the infological equation. However, if this is anything to follow, then different people viewing same data will come up with different interpretations, i.e. propose different meanings to the same data. The above assertion will usually spell doom during knowledge reconciliation process between the decision maker and the watcher who is charged with the responsibility of providing necessary information for decision purposes. The subjective nature, i.e. related to a subject – "a mind, ego, or agent of whatever sort that sustain or assumes the form of thought or consciousness" of information was established. Another similar view to the one expressed above is the description of information as "data plus meaning" or "meaningful data" (Mingers, 1997). With the following, an etymological perspective of the term 'data' and 'meaning' was undertaken for Callaos & Callaos, (2002) and the authors made bold to say that "information" as "meaningful data" would be defined as "significant data", "data full of meaning", "data having a meaning or purpose", similarly, "data plus meaning" would be defined as "data plus significance", "data plus the thing conveyed by it in the mind".

The aim of soft systems methodology (SSM) in this research would be to take seriously the subjectivity which is the crucial characteristic of human affairs and to treat this subjectivity in a way characterized by intellectual rigor. The SSM is variously characterized as a system of enquiry, enquiry process, learning system, reflection in action, an organized version of doing purposeful thinking or structured way of thinking.

Common way of expressing misconception are – ''I thought you meant…, do you intend to say…, I suppose…," e.t.c. All these expression (list not exhaustive) are result of misconceptions which could arise in an attempt to inform or be informed. The consequent may be grave or mild, but it will be absurd to actually wait for the result before they are nipped in the bud. Usually in decision making process, one or all of the following will be true:

- there is a problem to be properly identified and resolved
- there is need for dialogue and deliberation between parties concerned
- several factors are involved and need be identified discretely, assessed individually and also collectively to facilitate proper understanding.
- Problem descriptions employ “word” that engulfs uncertainty and vagueness.

Soft information retrieval benefit from the above because the deepness of the analysis of language, and their consequent range of applicability implies a satisfying interpretation of the documents’ meaning requires a large number of decision rules even in narrow application domains. It therefore behooves that the resolution of the problem requires the approximation of the vague descriptions and their distribution as earlier mention. CW model earlier discussed comes in handy with the first step of precisiating imprecise, uncertain and ambiguous information. This is sequel to the fact that the imprecision and vagueness of term capable of resulting into misconception could be treated in a quantitative manner in CW.

Although fuzzy inference systems do not necessarily perform better than human being in handling imprecision and uncertainty in interpretation, it has however assist in accommodating hither subjective factors which are impossible in traditional hard systems. It is therefore no gain saying that knowledge is indispensable in effective reading and comprehension; however accruable knowledge is a factor of derivable meaning (interpretation) from the informing sources.

4.6 FuzzOntology

Natural language is employed by human beings every day, their languages as means of communication has been shaped by thousand years of human history, making it more convenient and effective (Zadeh, 1991). This is due partly to the fact that human behaviour and expressions represent highly complex, non linear and adaptable systems. It is therefore no gain-saying that, sentences writing in everyday languages portrayed an efficient form of communication. However, uncertainty, imprecision and ambiguity are inherently present in natural systems. Probability theory dealing with randomness therefore fails to adequately model many situations; consequently, artificially imposed precision or theoretical assumptions only result in increasing the complexities of the models (Wang, et al., 1998). In fuzzy system, rules are formulated employing these linguistic expressions and applying them to human behavioural problem since languages are employed in the formulation of the rules (Jang, 1996, Abraham, 2001).

The conventional manner of using information requires discrete values which are precise and devoid of ambiguity, however, this is not realizable as human being employs imprecise expression with high level of uncertainty or no clear boundaries to describe a situation e.g. I am *very* hungry, it is going to be cloudy today. Thus, design of hard computing systems fall short and are always at the mercy of the user to bail them out of such confusion. The above mentioned, pave the need for design methodology that can take as input vague descriptions from human being and use same to compute or operate successfully. This singular factor is the basis for the emergence of soft computing paradigm.

Delivering strategic decisions, high rate performance and complex decision making are cogent features in Economic Intelligence; it is therefore no gain saying to assert that appropriate method of information-sourcing for knowledge application is of utmost importance.

In earlier section we defined ontology as formal vocabularies, shared by a group of people/individual, interested in a specific domain. It potentially enable automated knowledge sharing and reuse among both human and computer agents; this is facilitated based on their ability to interweave human and machine understanding through formal and real-world semantics. The operation is further enhanced with a solid methodology that guarantees the collaborative engineering process. In summary the key advantages of ontology are: knowledge sharing; reusability of knowledge and their being machine understandable.

It can therefore be deduced that the basis for ontologically driven information system was the desire to incorporate knowledge from different domain into a single frame work. Ontological application in risk simulation brings about the conceptual requirement of “reusability”. This is sequel to the fact that models are specified by humans to embody domain knowledge, characterized by ambiguity, thus ontology provides a profound solution in bridging the semantic-gap between the knowledge-space and the simulation model (Eden, 1994, Wang, 2004). Earlier on we presented the knowledge reconciliation model and ontological framework, however further research has shown that accruable knowledge by any actor is determined largely by the meaning derivable from available information. While ontology prided itself in bridging the semantic gap and fostering reusability of knowledge, it has been noticed with dismay that even capture knowledge can still be misapplied based on the fact that natural language which serves mainly as human means of expression abounds with perceptive information and perception are intrinsically imprecise and fuzzy.

Fuzzontology (Onifade, et al., 2010) therefore is our attempt to create a fusion between ontology and fuzzy inference system i.e. whatever is captured by ontology, can be interpreted by fuzzy. Indeed it is noteworthy to assert here as earlier pinpointed by Zadeh, (2009) that incursion and operation of fuzzy logic otherwise referred to as computing with word (CW) model is not an extension of language processing rather, the transition from computing with numbers to computing with word is endowed with potential to cause a paradigm shift capable of opening the doors to wide ranging enlargement of the role of natural languages in scientific theories. Consequently, fuzzontology attempt to employ soft computing paradigm to precisitate the meaning of words and propositions drawn from natural language (captured by ontology), thereafter reason and compute via the precisiated word and propositions (fuzzy inference system).

4.6.1 Soft Computing Paradigm

Soft computing differs from conventional computing in that, unlike hard computing, it is tolerant of imprecision, uncertainty and partial truth. Therefore, it appeared to be that, the role model of soft computing is the human mind (Zadeh, 2002). The guiding principle of soft computing are: the tolerance for imprecision, uncertainty and partial truth aimed at achieving tractability, robustness and low solution cost (Cox, 1999). While several participating entities have been listed, it is imperative to assert that soft computing (SC) is not a scene of uncoordinated combination of fuzzy logic, neural network and probability reasoning (genetic algorithm, chaos theory, belief nets, and learning theory). Contrariwise, it is consist in a partnership within which participating entities contribute usually at different organizational levels providing and resulting into a hybrid system. Consequently, their combination results into a complementary effect rather than being competitive with a bid to ameliorate the constricted ways hitherto operational.

The fundamental solving techniques for soft systems were described by the Checkland's soft system methodology (SSM). It has been first published in 1981 and extended in the 90's (Checkland et al., 1981, 1990 and 1998, Rose, 2000). It is a philosophical system concept formulated specially as a tool for solving management problems. Logical information stream compares the conceptual models with reality perception and the complementary cultural stream does the cultural, policy and intervention analysis. Pešl, & Hřebíček, (2003) describes a methodology which specifies seven phases of the problem solving process as described in figure 3.4. The third and the fourth step is realized using system concepts, all other are done in real world.

In our earlier discussion on the need for translation credibility during indirect informing process, we identified some inherent risk factors (RFs) and employed same for the dialogue between the elderly person and the medical doctor. Relating these to our EI decision problem, we refine the earlier identified RFs membership function, redefining and assigning fuzzy membership functions as follows:

<ORN>: what is the need, at stake, (SWOT): '***well understood'***, '***understood***', '***not-understood***'.

<ENF>: Govt. policies, competitors, customers: "***Highly favourable***', '***favourable***', '***not-favourable***'.

<INE>: level of education, cognitive experience, and intuition: '***Advance***', '***intermediate***', '***low***'.

<BIA>: Personal preference, misinformation, disinformation: '***prominent***', '***average***', '***low***'

<LOU>: interpreting the decision problem: '***high***', '***average***'.

We notice the similarities in the above RFs identified during the KNOWREM process and the Dr. & EP earlier discussed. The linguistic descriptions are common phenomenon in dealing with human communication. Thus, the above is determined based on their belongingness to particular world of discuss i.e. membership functions (MF). MF determines and indicate the level at which a condition is true and believable. Figure 4.8 below describe the importance and showcase the pattern of utilizing membership functions derived from the descriptive attributes in language of discussion.

It is expected that the membership values for one input variable add to unity everywhere. This is achievable by dividing the membership values m_i (x) by the sum of all membership values which results into normalized MF. The good thing about this method of problem resolution is that, while individual problem items are considered, fuzzy inference system can simultaneously monitor and adapt the effect of each of the identified factors to be accommodated in the overall consequences to present a very robust solution to the problem.

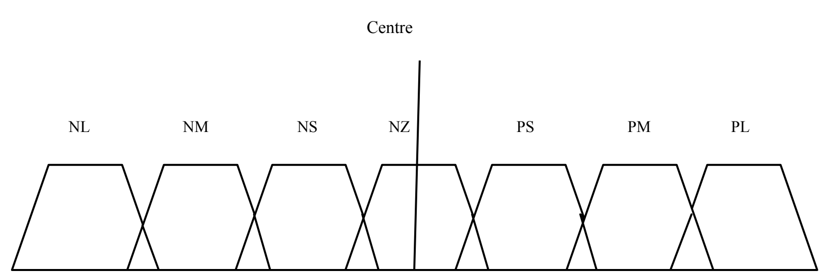


Figure 4.8: A Trapezoidal example for Fuzzy definition of Membership function

As a key to the above representation, the following suffices:

NL – Negatively Large	PS – Positively Small
NM – Negative Medium	PM – Positive Medium
NS – Negative Small	PL – Positive Large
NZ – Near Zero	

There is no restriction to the number of membership definition for any occurrence in fuzzy inference system (FIS). However, strict attempt must be made to employ the most reasonable and representative number which will not lead to messy rules definition in the course of the main operation. The figure below presents a typical example (using triangular membership function) to depict the ways by which this method can be used to determine the belongingness of risk factors (RFs).

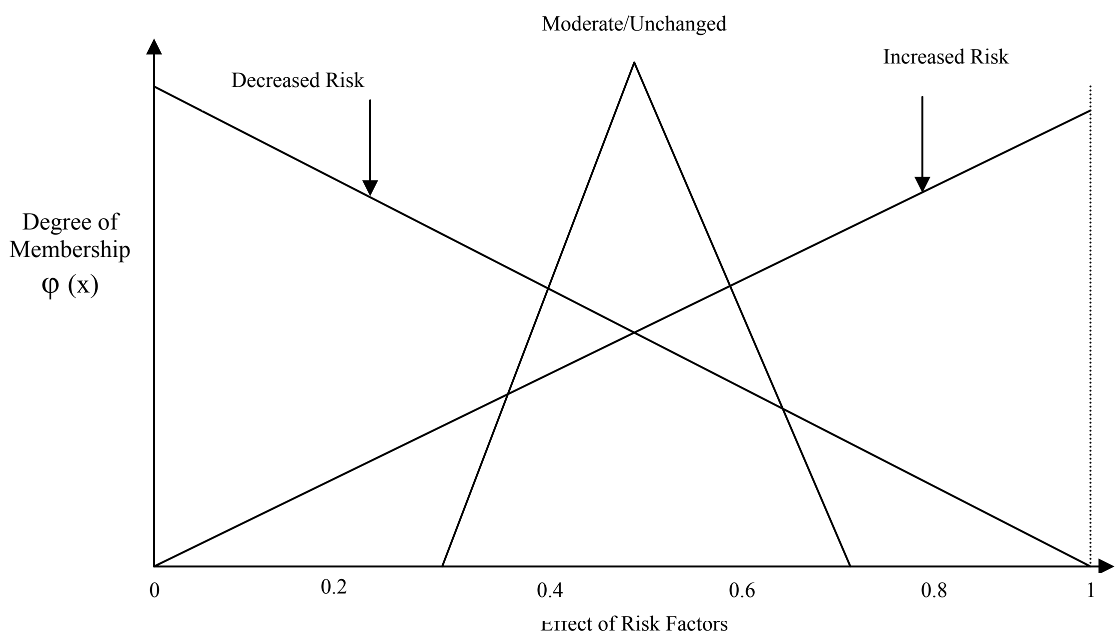


Figure 4.9: Explaining the concept of Membership Functions with RFs

Figure 4.9 presents a method of explaining the treatment of linguistic variables in fuzzy systems. The above figure exemplifies the pattern of risk growth in a particular situation. It is noticeable at first glance that gradual increase in the value of any of the identified risk factors will result into appropriate level of risk in the consequent part. A similar result is obtainable for a decrease in the RF values. It again assists in determining the intermediate interval based on similar expression.

We note again that it is expected that membership values for one input variable add to unity everywhere. This membership example of being able to generalize concepts of interval of real numbers to fuzzy intervals or real numbers and vice versa i.e. real numbers to fuzzy intervals and fuzzy numbers is the basis and the strength of fuzzy systems. The advantage of using fuzzy numbers and fuzzy intervals is enshrined in the ability to adequately express the effect of the measurement error more sincerely and faithfully in a pure, simple linguistic manner devoid of unnecessary precision and at the same time depicting the vagueness of the result in the employed linguistic manner. Fuzzy membership functions (MFs) associates linguistic labels (e.g. ‘tall’) with a particular area of one of the input or output variables (e.g. ‘secretary’).

4.6.2 Computing with Word – the Fuzzy paradigm shift

Fuzzy set theory was first introduced in 1965 by Lotfi Zadeh at the University of California, Berkeley and fuzzy logic was derived from it for use in applications. It has been applied to diverse fields from control theory to artificial intelligence, yet still remains controversial among most statisticians who prefer probabilistic logic and some control engineers who prefer traditional two valued logic. Though it spots an unconventional name, fuzzy logic is a field that has risen to the forefront of modern technology. Fuzzy technology was first developed in the United States, but it has truly bloomed into a billion dollar industry in Japan. The U.S., by some estimates, trails Japan by ten years in developing fuzzy logic applications.

Fuzzy logic is a form of multi-valued logic derived from fuzzy set theory to deal with reasoning that is approximate rather than precise. Just as in fuzzy set theory the set membership values can range (inclusively) between 0 and 1, in fuzzy logic the degree of truth of a statement can range between 0 and 1 and is not constrained to the two truth values {true, false} as in classic predicate logic. And when linguistic variables are used, these degrees may be managed by specific functions.

Fuzzy logic system is a powerful problem-solving methodology with a myriad of applications in embedded control and information processing. Fuzzy provides a remarkably simple way to draw definite conclusions from vague, ambiguous or imprecise information. In a sense, fuzzy logic resembles human decision making with its ability to work from approximate data and find precise solutions. As opposed to classical logic which requires a deep understanding of a system, exact equations, and precise numeric values, Fuzzy logic incorporates an alternative way of thinking, which allows modelling complex systems using a higher level of abstraction originating from our knowledge and experience.

The concept therein is all about the relative importance of precision: "*How important is it to be exactly right when a rough answer will do*?"

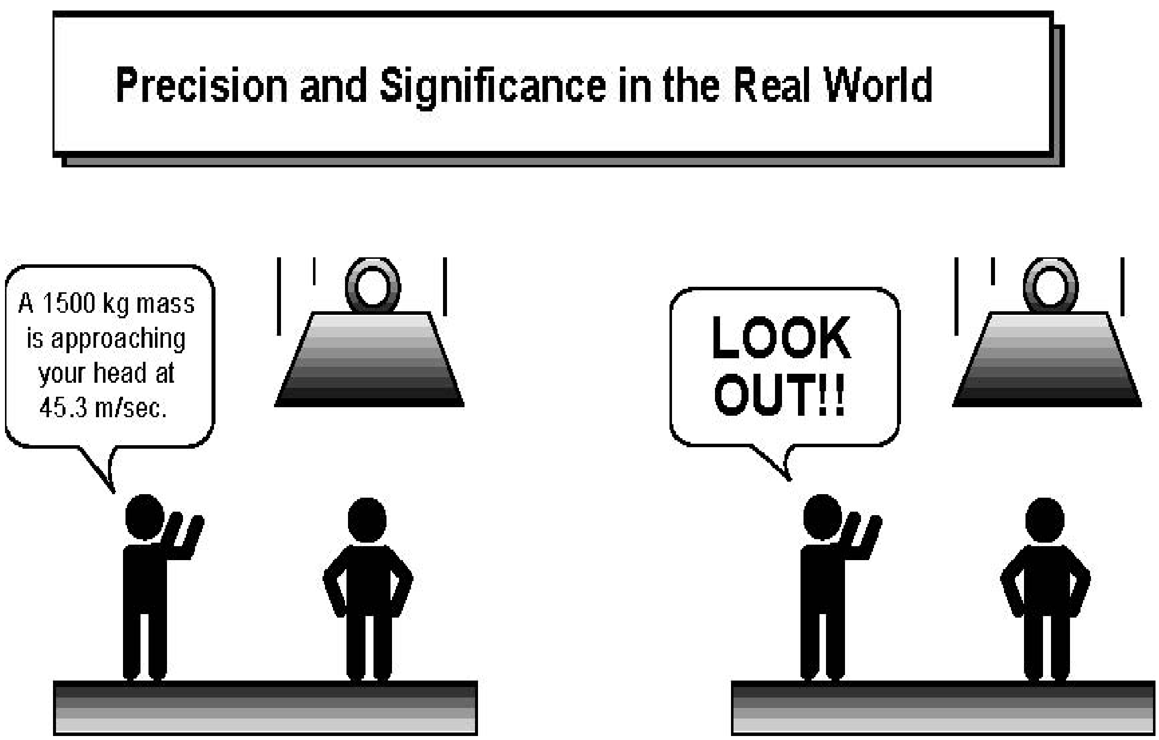


Figure 4.10: Dealing with precision and significance

The ability to formally represent and reason with perceptive information is the bedrock of computing with words model (CW-model). In figure 4.10, an informing scenario was presented but being interpreted by two different actors – one employing the conventional method (precision) and the other informing in a fuzzy manner. In the formal, the actor attempts to inform by describing the impending danger to the precise, accurate weight and the rate at which it is descending. Little wonder if the load will not crush the man before the whole process is completed. In the latter however, the actor informs the other actor by a danger warning 'look out!!' In these two scenarios, we can see the need to accommodate imprecision towards effective communication and comprehension. Below, we consider some popular conclusion of some notable people on the need to accommodate imprecision consequent upon the need to acknowledge the sharp contrast and mismatch between the formality and precision of classical logic and the flexibility and variation of natural languages.

Precision is not truth. —Henri Matisse
"*Sometimes the more measurable drives out the most important*". —René Dubos
"*Vagueness is no more to be done away with in the world of logic than friction in Mechanics*". —Charles Sanders Peirce
"*I believe that nothing is unconditionally true, and hence I am opposed to every statement of positive truth and every man who makes it*". —H. L. Mencken
"So far as the laws of mathematics refer to reality, they are not certain. And so far as they are certain, they do not refer to reality". —Albert Einstein
"*As complexity rises, precise statements lose meaning and meaningful statements lose precision*". —Lotfi Zadeh

There abound many misconceptions about what Computing with Words (CW) is and what it has to offer. The following are direct excerpt from the submissions of Lofti, Zadeh (2009) in an attempt to disambiguate the prevailing misconception about the relationship between CW and natural language processing. According to the author, the common misconception that CW is closely related to natural language processing in reality does not hold. More importantly, at this juncture what is widely unrecognized is that moving from computation with numbers to computation with words has the potential for evolving into a basic paradigm shift—a paradigm shift which would open the door to a wide-ranging enlargement of the role of natural languages in scientific theories.

In essence, CW is a system of computation which adds to traditional systems of computation two important capabilities:

(a) The capability to precisiate the meaning of words and propositions drawn from natural language; and

(b) The capability to reason and compute with precisiated words and propositions.

As a system of computation, Zadeh (2009) described a CW-based model, or simply CW-model to comprise of three principal components.

- A question, Q, of the form: What is the value of a variable, Y?
- An information set, $I=(p_1, ..., p_n)$, where the p_i, $i=(1, ..., n)$, are propositions which individually or collectively are carriers of information about the value of Y, that is, are question-relevant. One or more of the p_i may be drawn from world knowledge.

A proposition, p_i, plays the role of an assignment statement which assigns a value, v_i, to a variable, X_i, in p_i. Equivalently, p_i may be viewed as an answer to the question: What is the value of X_i? X_i and v_i, may be explicit or implicit. A proposition, p_i, may be unconditional or conditional, expressed as an if-then rule (Zadeh, 2009).

Basically, an assignment statement constrains the values which X_i is allowed to take. To place this in evidence, X_i and v_i are referred to as the constrained variable and the constraining relation, respectively. More concretely, what this implies is that the meaning of a proposition, p, may be represented as a generalized constraint, X isr R, in which X is the constrained variable, R is the constraining relation and r defines the modality of the constraint, that is, the way in which R constrains X. When v_i is a word or a combination of words, X_i is referred to as a linguistic variable, with v_i being its linguistic value. When it is helpful to stress that p_i assigns a value to a variable, p_i is referred to as a valuation. Correspondently, the information set, I, is referred to as a valuation system, V.

- The third component is an aggregation function, f, which relates Y to the X_i.

$$Y=f(X_1, ..., X_n) \tag{4.12}$$

It was therefore asserted that the principal difference between CW and conventional systems of computation according to Zadeh, (2009) is that CW allows inclusion in the information set, I, of propositions expressed in a natural language, that is, linguistic valuations. Legalization of linguistic valuations has important implications. First, it greatly enhances the capability of computational methodologies to deal with imperfect information, that is, information which in one or more respects is imprecise, uncertain, incomplete, unreliable, vague or partially true. In realistic settings, such information is the norm rather than exception. Second, in cases in which there is a tolerance for imprecision, linguistic valuations serve to exploit the tolerance for imprecision through the use of words in place of numbers. And third, linguistic valuations are close to human reasoning and thus facilitate the design of systems which have a high level of machine intelligence, that is, high level of MIQ (machine IQ).

The question can thus be asked "*What does Computing with Words have to offer*?" The answer rests on two important tools which are provided by the machinery of fuzzy logic. The first tool is a formalism for mm-precisiation of propositions expressed in a natural language through representation of the meaning of a proposition as a generalized constraint of the form X isr R, where as noted earlier X is the constrained variable, R is the constraining relation and r is the modality of the constraint (Zadeh 1986). The second tool is a formalism for computing with mm-precisiated propositions through propagation and counter-propagation of generalized constraints. The principal rule governing constraint propagation is the Extension Principle (Zadeh, 1965, 1975). In combination, these two tools provide an effective formalism for computation with information described in a natural language. And it is these tools that serve as a basis for legalization of linguistic valuations.

The above excerpt becomes expedient to juxtapose the submission of Liu, (2007) based on the possible application of CW in automatic text summarization. Conclusively, Zadeh pointed out that what is important to note is that the machinery of fuzzy if-then rules — is a machinery which is employed in almost all applications of fuzzy logic — and this is a part of the conceptual structure of CW.

4.6.3 Description of Fuzzy Systems

The most fascinating aspect of fuzzy logic lies in trading off between significance and precision (something that humans have been managing for a very long time). Fuzzy logic sometimes appears exotic or intimidating to those unfamiliar with it, but once you become acquainted with it, it seems almost surprising that no one attempted it sooner. In this sense fuzzy logic is both old and new because, although the modern and methodical science of fuzzy logic is still young, the concepts of fuzzy logic reach right down to our bones. Fuzzy logic is a convenient way to map an input space to an output space.

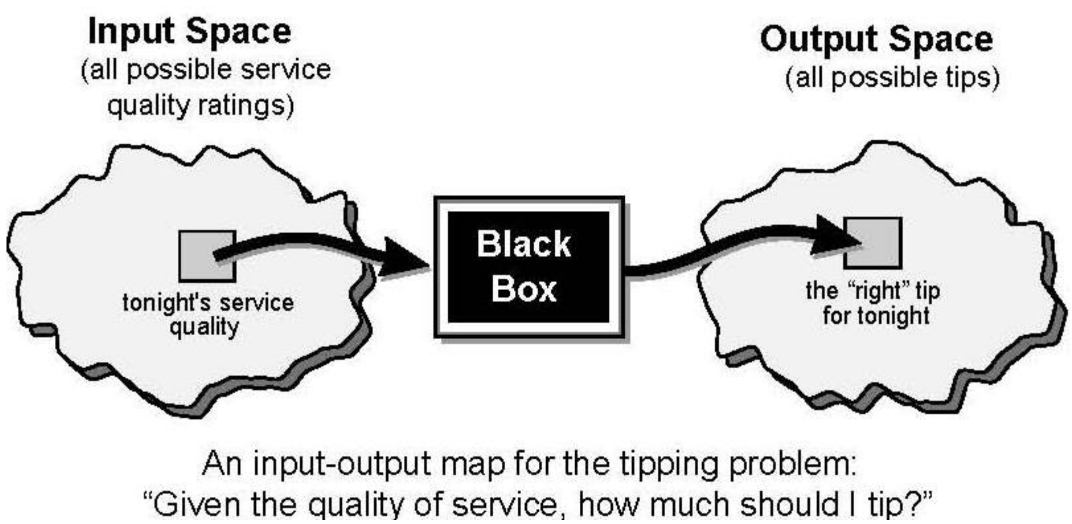


Figure 4.11: Mapping operation in fuzzy systems

The rationale for the adoption of fuzzy lies in what figure 4.11 depicts. This is the starting point for everything else, and the great emphasis here is on the word "convenient." We can therefore ask - *what is meant by mapping input space to output space*? The following examples suffice: *you tell me how good your service was at a restaurant, and I will tell you what the tip should be. You tell me how hot you want the water, and I will adjust the faucet valve to the right setting. You tell me how far away the subject of your photograph is, and I will focus the lens for you. You tell me how fast the car is going and how hard* *the motor is working, and I will shift the gears for you*. All the above coined examples involved the ‘appropriate’ mapping of inputs to the ‘appropriate’ outputs.

Located between the input and the output space is the Black Box (BB) where the major operation takes place. Does anyone has an idea of what could go in the black box? Any number of things:

- Fuzzy systems
- Linear systems
- Expert systems
- Neural networks
- Differential equations
- Interpolated multidimensional lookup tables and host of others

There are dozens of ways to make the black box work, however, it normally turns out that fuzzy is often the very best way. In the remark of Lotfi Zadeh, “*In almost every case you can build the same product without fuzzy logic, but fuzzy is faster and cheaper*.” Below are some of the accruable benefits from the use of fuzzy systems.

- *Fuzzy logic is conceptually easy to understand* - the mathematical concepts behind fuzzy reasoning are very simple. What makes fuzzy nice is the “naturalness” of its approach and not its far-reaching complexity.
- *Fuzzy logic is flexible* - with any given system, it’s easy to massage it or layer more functionality on top of it without starting again from scratch.
- *Fuzzy logic is tolerant of imprecise data* - everything is imprecise if you look closely enough, but more than that, most things are imprecise even on careful inspection. Fuzzy reasoning builds this understanding into the process rather than tacking it onto the end.
- *Fuzzy logic can model nonlinear functions of arbitrary complexity* - you can create a fuzzy system to match any set of input-output data. This process is made particularly easy by adaptive techniques like ANFIS (Adaptive Neuro-Fuzzy Inference Systems).

- *Fuzzy logic can be built on top of the experience of experts* - in direct contrast to neural networks, which take training data and generate opaque, impenetrable models, fuzzy logic lets you rely on the experience of people who already understand your system.
- *Fuzzy logic can be blended with conventional control techniques* - fuzzy systems don't necessarily replace conventional control methods. In many cases fuzzy systems augment them and simplify their implementation.
- *Fuzzy logic is based on natural language* - the basis for fuzzy logic is the basis for human communication. This observation underpins many of the other statements about fuzzy logic.

A fuzzy subset A of a set X is a function $A: X \rightarrow [0,1]$ from X to the unit interval $[0,1]$. The value of A(x) is thought of as the degree of the membership of x in A. This function is sometimes called the *membership function* of A. In the special case that the membership function takes on only the values 0 and 1, A is called an *ordinary or crisp subset of X* and its membership function coincides with its characteristic function. The set X is sometimes called the *universe of discourse*. A fuzzy subset is often simply referred to as a Fuzzy set.

Other notations in common use for the membership function of the fuzzy subset A of X implies the

$$\mu_A : X \rightarrow [0,1] \text{ and sometimes } \int_{x \in X} \frac{\mu_A(x)}{x} \qquad (4.13)$$

or, if the domain is discrete,

$$\sum_{x \in X} \frac{\mu_A(x)}{x}. \qquad (4.14)$$

These more complicated notions do not convey additional information; consequently, we do not make a notational distinction between a fuzzy subset A and its *membership function* $A: X \rightarrow [0,1]$.

The support of a function

$$A: X \rightarrow [0,1] \text{ is the set } supp(A) = \{x \in X \mid A(x) \neq 0\} \qquad (4.15)$$

The notion of membership function as hitherto depicted in figures 4.8, 4.9 and equation 4.13 – 4.14 is central to the effectiveness of any fuzzy system. There are many types of membership functions employed in fuzzy systems, notable among them are:

1. Triangular Functions
2. Trapezoidal Functions
3. Gaussian Functions

Triangular and Trapezoidal Functions are both piecewise-linear functions, are often used in applications. The Triangular function A with endpoints (a,0) and (b,0), and high point (c, α) is defined by

$$A(x) = \begin{cases} \alpha\left(\frac{x-a}{c-a}\right) & if\ a \le x \le c \\ \alpha\left(\frac{x-b}{c-b}\right) & if\ c \le x \le b \\ 0 & Otherwise \end{cases} \quad (4.16)$$

The Trapezoidal function B with end points (a,0) and (b,0), and high points (c, α) and (d, α) is defined by

$$A(x) = \begin{cases} \alpha\left(\frac{x-a}{c-a}\right) & if\ a \le x \le c \\ \alpha & if\ c \le x \le d \\ \alpha\left(\frac{x-b}{c-b}\right) & if\ d \le x \le b \\ 0 & Otherwise \end{cases} \quad (4.17)$$

The Gaussian functions, the familiar bell-shaped curve are of the form

$$A(x) = e\hat{}(-((x-c)^2)/2\sigma^2) \quad (4.18)$$

Basically, there are five parts to the fuzzy inference systems: *fuzzification* of the input variables, *application of the fuzzy operator* (AND or OR) in the antecedent, implication from the antecedent to the consequent, *aggregation* of the consequents across the rules, and *defuzzification*. In the subsequent sections, we briefly take a look at these cryptic and odd names which help convey specific meaning in fuzzy inference system.

Fuzzy Inputs

Sequel to the discussion in the preceding section, we shall take a look at the major components in a fuzzy logic system or the fuzzy logic toolbox. The first of this component is the *fuzzy inputs*. In the Fuzzy Logic Toolbox, the input is always a crisp numerical value limited to the universe of discourse of the input variable and the output is a fuzzy degree of membership in the qualifying linguistic set (always the interval between 0 and1).The first step is to take the inputs and determine the degree to which they belong to each of the appropriate fuzzy sets via membership functions. Fuzzification of the input amounts to either a table lookup or a function evaluation which thereafter facilitate the inference process. This is clearly shown in figure 4.12. This figure captured the whole components to be discussed in this section and went further to show the interactions between and amongst these components.

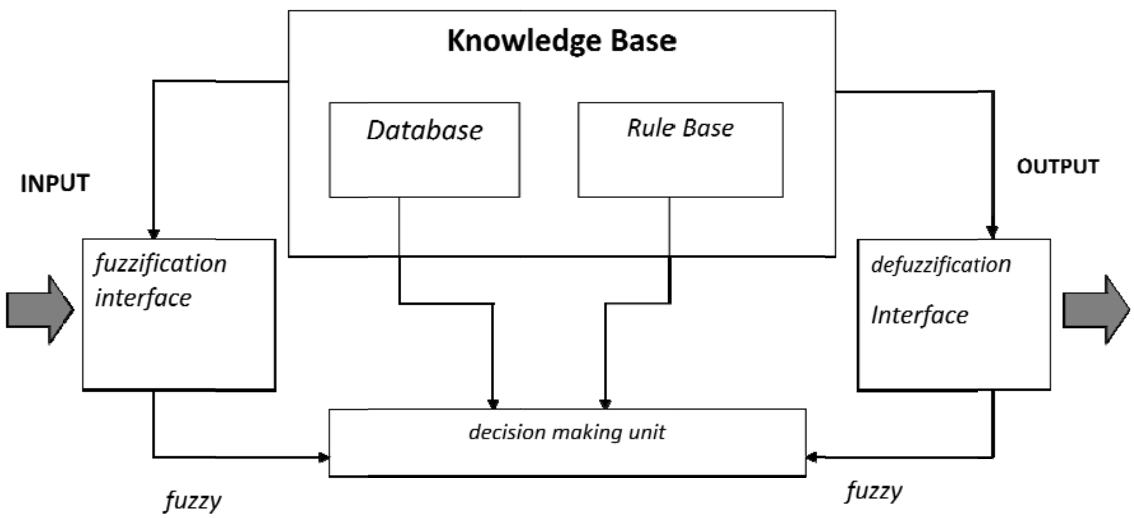


Figure 4.12: Major components of Fuzzy System (Zadeh, 2000)

Fuzzy Operators

Following the fuzzification of the inputs, we are informed of the degree to which each part of the antecedent has been satisfied for each rule – "belongingness". If the antecedent of a given rule has more than one part, the fuzzy operator is applied to obtain one number that represents the result of the antecedent for that rule. This number will then be applied to the output function. The input to the fuzzy operator is two or more membership values from fuzzified input variables. The output is a single truth value.

It is important to note that any number of well-defined methods can fill in for the AND operation or the OR operation. In the Fuzzy Logic Toolbox, two built-in AND methods are supported: min (minimum) and prod (product). Two built-in OR methods are also supported: max (maximum), and the probabilistic OR method probor. The probabilistic OR method (also known as the algebraic sum) is calculated according to the equation

$$\text{probor}(a, b) = a + b - ab \tag{4.19}$$

In addition to these built-in methods, it is possible to create personalized methods for AND and OR by writing such function and setting it be the method of choice.

Implication Method

In fuzzy systems, every rule has a weight (a number between 0 and 1), which is applied to the number given by the antecedent. However, before applying the implication method, effort must be geared to monitor the rule's weight. Generally if the weight is 1 it has no effect at all on the implication process. From time to time there can be variations in the weight of one rule relative to the others by changing its weight value to something other than 1.

Once proper weights have been assigned to each rule, the implication method can be implemented. A consequent is a fuzzy set represented by a membership function, which weighs appropriately the linguistic characteristics that are attributed to it. The consequent is reshaped using a function associated with the antecedent (a single number). The input for the implication process is a single number given by the antecedent, and the output is a fuzzy set. Thereafter, implication is implemented for each rule. Two built-in methods are supported, and they are the same functions that are used by the AND method: min (minimum), which truncates the output fuzzy set, and prod (product), which scales the output fuzzy set.

Aggregation of Outputs

Again, as shown in figure 4.12, decision making is made possible via the combinatorial effect of the database and the rule base that both constitute the knowledge base. This is sequel to the fact that decisions are based on the testing of all of the rules in an FIS, the rules must be combined in the best manner in order to make a decision. Aggregation is thus the process by which the fuzzy sets that represent the outputs of each rule are combined into a single fuzzy set. Aggregation only occurs once for each output variable, just prior to the fifth and final step, i.e. defuzzification.

The input of the aggregation process is the list of truncated output functions returned by the implication process for each rule. The output of the aggregation process is one fuzzy set for each output variable. Notice that as long as the aggregation method is commutative (which it always should be), then the order in which the rules are executed is unimportant. Three built-in methods are supported: max (maximum), probor (probabilistic OR), and sum (simply the sum of each rule's output set).

Defuzzification

The input for the defuzzification process is a fuzzy set (the aggregate output fuzzy set) and the output is a single number. As much as fuzziness helps the rule evaluation during the intermediate steps, the final desired output for each variable is generally a single number. However, the aggregate of a fuzzy set encompasses a range of output values, and so must be defuzzified in order to resolve a single output value from the set. This unique feature is amongst others responsible for the success of fuzzy logic

Defuzzification process can be achieved via various methods. There are five built-in methods currently supported: centroid, bisector, middle of maximum (the average of the maximum value of the output set), largest of maximum, and smallest of maximum. However, the most popular defuzzification method is the centroid function calculation, which returns the center of area under the curve.

4.7 Embedded SITE-LORIA Economic Intelligence System's Architecture

A similar figure to figure 4.13 was presented in chapter 2. However, it was not yet embedded with the models proposed in the course of this research. Figure 4.13 followed from the submission of David & Thiery, (2003) in an attempt to set a framework which will accommodate the existing models at SITE-LORIA. Literally or informally, architecture is considered to consist of "framework", with some well defined concepts and rules governing the operations to be set forth. This architecture is a foresight on the importance to integrate the subsequent models. Thus, we can see from the definition describing the architecture how it closely portrayed and corroborates the EI process earlier discussed with attempts to identify the actors and their roles. The above described is a common place in software development, where most, if not all the concepts earlier discussed are present. On the same vein, EI defines its processes, develop its models (probably there are more to be developed), defined its actors and relationships (in form of rules, but not rigid) and thus it will be required to integrate these in a properly developed architecture or framework similar to that shown in figure 4.13.

Now embedded with the existing architecture are four sets of red-edged, blue coloured rectangles representing our proposed models. Following will be discussion about each of these models but not necessarily in any sequential manner. This is sequel to the fact that, there exist close relationship amongst these models, and thus, their discussion can only be logically attained if back and forth method is employed.

Figure 4.13: Architecture of EI Systems according to SITE-LORIA embedded with our Models

These models include: Knowledge Reconciliation and Ontological Framework tagged (KNOREM). This model attempts to determine the level of understanding that results from the interaction between some set of actors (decision maker and watcher) towards the need to deliver strategic decision. Decisionability succinctly describes the problems associated with increase volume of information on the ability of a decision maker to adequately perform his roles. Fuzzontology is a compound word derived from Fuzzy Inference System (FIS) and Ontology. This followed from the result of KNOREM taking into cognizance the opinion of Langefors, (1966) dealing with 'interpretation'. The model thus interprets what ontology captures with fuzzy inference system. The last of the proposition is tagged FUZZYMATCH, again, it combines the tolerance in FIS to address the issue of missing data that could result from missing data, more important missed/spelling errors. This will be the main focus of the next chapter.

4.8 Recap

This chapter has featured principally the focal points of this research and this is unequivocally depicted in the title of the chapter. We begin by discussing what constitute a model and what does not, followed by a need for model development. Since the focal point of this research is risk management, we reflect on our submission from the previous chapter. This is followed by attempt to present a chronology of accruable risk factors that can result from the activities and relationship of EI actors. We went a step further to capture alongside other risk factors the cognitive ability of the actor involved in decision process. Building around the existing EI system architecture at SITE-LORIA, we identified and proposed other models that are deemed necessary in facilitating the decisionability of the actors, determining is understanding of the decision problem – translation credibility, measuring the success of informing process between actors – knowledge reconciliation via ontology and the final contribution is attempt to employ soft system methodology in resolving ambiguity and imprecision inherent in human conversation – fuzzontology. Other focus of this chapter is introduction into soft system methodology, formalizing the knowledge reconciliation process and the importance of proper interpretation between informing and informed entities.

The full implementation and simulation of Fuzzontological model will be done in next chapter. We shall also discuss one of the models depicted on the embedded SITE-LORIA and was yet mentioned i.e. FuzzyMatch.

“*The only real valuable thing is intuition; there is no logical way to the discovery of these elemental laws. There is only the way of intuition, which is helped by a feeling for the order lying behind appearance*”.

Einstein, 1909

Chapter 5

Design and Implementation of the Proposed Solution

5.1 Hypothesis Recall, Summary of Related Works and Proposals

Adequate access to quality forms the bedrock of delivering strategic decision. In previous chapters, we have dealt with the concept of data/information quality and their effect and also went ahead to review various forms of decision making models. We made bold to say that our research support the definition of data quality which state "data fit for customer or consumer purpose". This shows that the user information need is been met with some level of relevance. Relevance has been described as a situation where there is a minimum satisfaction between defined information needs and stored documents. All these diverse forms of subjectivity have been a clog in the wheel in information risk management. The research hypothesis thus results from the above and can be stated as follows: ***With appropriate risk factor definition and modeling, information risk can greatly be reduced if not eliminated towards the process of taking strategic decision in economic intelligent systems***. Consequently, our means of achieving this was the proposition of various models aimed to fit one or more stages in the EI process.

The existing EI architecture at SITE –LORIA consists of multi-dimensional components which involve the main components of EI systems i.e. EI actors, EI processes and information. The processes via which these components are brought together constitute the mode via which this research evolved. At a glance, we can identify the following distinct but interrelated phases which were managed in this research: actors' interactions, information sourcing and collection, information processing, information utilization for decision making. The onus from the above is such that it will be inadequate to formulate a monster model that takes care of all identified risk factors in these stages.

Consequent upon the above, we proposed four different models to operate on these divergent components but in a harmonized manner. The first is tagged KNOWREM, this assists in the reconciliation of knowledge amongst EI actors. We followed this by FuzzOntology, it helps in resolving the imbroglio that can ensue during knowledge reconciliation and decision problem definition. We made bold to say that the volume of available information whether of quality or not can grossly affects the overall decision. To ameliorate the effect of information overload, we proposed a dual layer filtration called decisionability. The last of these models is FuzzyMatch. It assists the information specialist to combat the problem of missing or non-missing but wrong data.

Our approach for this research has been the focus of previous chapter, in this chapter; we shall show some of the design and implementation of our models fitting into the SITE-LORIA economic intelligence architecture. Again, in this research we adopt the assertion that data does not exist in isolation but is a result of human intervention by creating data via their knowledge and understanding.

The escalation in changes and globalization has led organizations with no better option than searching for new management approaches to decision-making Sinclair & Ashkanasy (2005). Hayward & Preston (1998) argue that linear rational models fail to perform satisfactorily for businesses operating under rising pressure and ambiguity (Andersen, 2000; Kuo, 1998). Considering from the managerial decision-making point of view, Nutt (1999) reported that rational decision-making strategies struggle to reach the 50% success mark. Sinclair & Ashkanasy (2005) concluded that many of the requirements for bounded rationality are becoming more difficult to satisfy based on the evidence of Eisenhardt & Zbaracki, (1992), Langley, et al., (1995).

Eisenhardt (1989) and Wally & Baum (1994) suggested that organizations consequently opt to embrace more holistic approaches to non-programmed decisions. In particular, their new openness to investigate alternative decision-making methods has been facilitated by the threat of high decision costs (Tomer, 1996). The above is further exacerbated by the exigency in time demand (Kuo, 1998), inadequate information (Agor, 1984; Goodman, 1993), and fast-paced change (Andersen, 2000), along with other factors triggered by emerging economic and technological forces since the 1980s (Hunt, 2000). These factors have led management researchers to question the effectiveness of rational decision-making as the only viable alternative (Sinclair & Ashkanasy2005).

New conceptual frameworks, such as the irrationality of the 'garbage-can theory,' however, also fail to provide comprehensive solutions (Langley et al., 1995). Researchers like Andersen, (2000); Eisenhardt, (1999) consequently resorted to exploration of less tangible concepts, such as intuition, however these work has not progressed to the point of developing a workable model. We, in resonance with Sinclair & Ashkanasy (2005) argue that the answer might lie in complementing the management tools that have withstood the test of time with new approaches, adequately responsive to today's changing business environment.

The above described necessitates the need for the proposition of knowledge reconciliation (KNOWREM) model in our earlier chapter. KNOWREM is based on the principle of ontology with the aim of facilitating common repertoire for people with similar goal. This is to facilitate reusability which in turn hastens the process of decision making. However, reusable knowledge still suffers from risks based on divergent opinions/views resulting from the process of interpretation (Onifade, et al., 2008). There abound perceptive information and perceptions which are intrinsically fuzzy and imprecise in natural languages. This has in no small measure hinders the progress towards universality of natural language content representation and reasoning because there exist significant limitation in expressing uncertainty and imprecision in information and knowledge.

Liu, (2007) in comparing the existing automatic text summarization applications with the concept of computing with words (CW) stressed that even with the wide adoption and application of first order predicate logic (FOPL) it's limitations in expressing qualitative quantifiers, modifier, or propositional attitudes (associated with words like believe, wants, opinion, knowledge, should, e.t.c.) still fall short of being perfect. Fuzzy logic however provides means of capturing, representing and operating on various quantifiers and predicate modifier thus providing a robust and closer solution to natural language (Zadeh, 1978, 1983, 1999).

We extend the earlier proposed ontological framework for knowledge reconciliation to accommodate the judgment and interpretation based on perceptions with fuzzy inference system. This is tagged FuzzOntology. The essence is to assist in resolving the ambiguity resulting from logical and rational mind in delivering decisions. It is however important to note that no worthy decision can result in the absence of necessary information. In this regards we developed a fuzzy based information retrieval system that extends the string matching function both at the representational and reasoning level. With this, we were able to adequately guide against the concept of missing and non-missing but not retrievable partly due to the dirtiness of the data.

In the rest of this chapter, we present the implementation of the fuzzontological model proposed in previous chapter and articulate how it can assist a decision maker and watcher in the cause of delivering their duties. We featured information retrieval and discuss the design and implementation of our search tool called FuzzyMatch. We made comparison between its functionalities and existing tools. Examples and case studies were employed to determine the efficacy of the system.

5.2 Implementation of FuzzOntological Model

It has been established that satisfying interpretation of the documents or deliberation needs a very large number of decision tree even in narrow application domain (ontology based systems). Our notion of fuzzontology (Onifade, et al., 2010) was based on the above described – whatever ontology captures, can be interpreted by fuzzy logic to provide a means of capturing and representing the inherent ambiguity in natural language employed by human in communication. We reiterate again here that fuzzy does not and cannot outperform human in handling imprecision and uncertainty but rather provide a means of reducing accruable risk and computation based on perceptions which were hitherto impossible in discrete operation of computer systems. Consequently, our approach in this regards will be to assert that data does not exist in isolation but as a result of human intervention brought about via intuition, rational analysis of their knowledge and understanding.

While there still exist philosophical discussions on whether a (symbolic) language is necessary for consciousness and thinking abilities, it is cannot be over emphasized that language is still humans most effective tool to structure his intuition, experience and judgmental abilities aimed at employing these to model his environment. Consequently, the ambiguity, vagueness, and imprecision can be modeled only with technology that has such provision (Onifade, et al., 2008).

Complexity of problem definition derives from the differences in interpretation of each key actors in economic intelligent process usually results into different level of comprehension of the event and to the explanations of the influences between events. Ontology potentially enable automated knowledge sharing and reuse among both human and computer agents. Ontology can facilitate human-machine interaction and understanding through the use of formal and real-world semantics (Wand & Weber, 1995). However, whenever ontology is faced with 'interpretation' there is always possibility of ambiguity which can highly impair the information search operation and subsequent utilization of such information for decision purposes.

It is no gain saying that the manner by which individual interprets data, information and knowledge will have tremendous impact on what course of action they will undertake in collecting, managing and sharing such information within an organization. The interdependency of the trio again was the vocal point in Knox, (2007) where several other authors' opinions were juxtaposed leading to the formation of what is referred to as the circular relationship between data, information and knowledge. The fact that information is more than data is consequent on the need for human interaction/involvement relating to interpretation before any logical derivative can be achieved. The element of interpretation is subjective and fuzzy thus requiring human involvement in making sense of something via their unique attribute with which they are endowed. Tuomi, (1999) argued that there is a reverse hierarchy of data – information and knowledge as data emerged last only after knowledge and information are available.

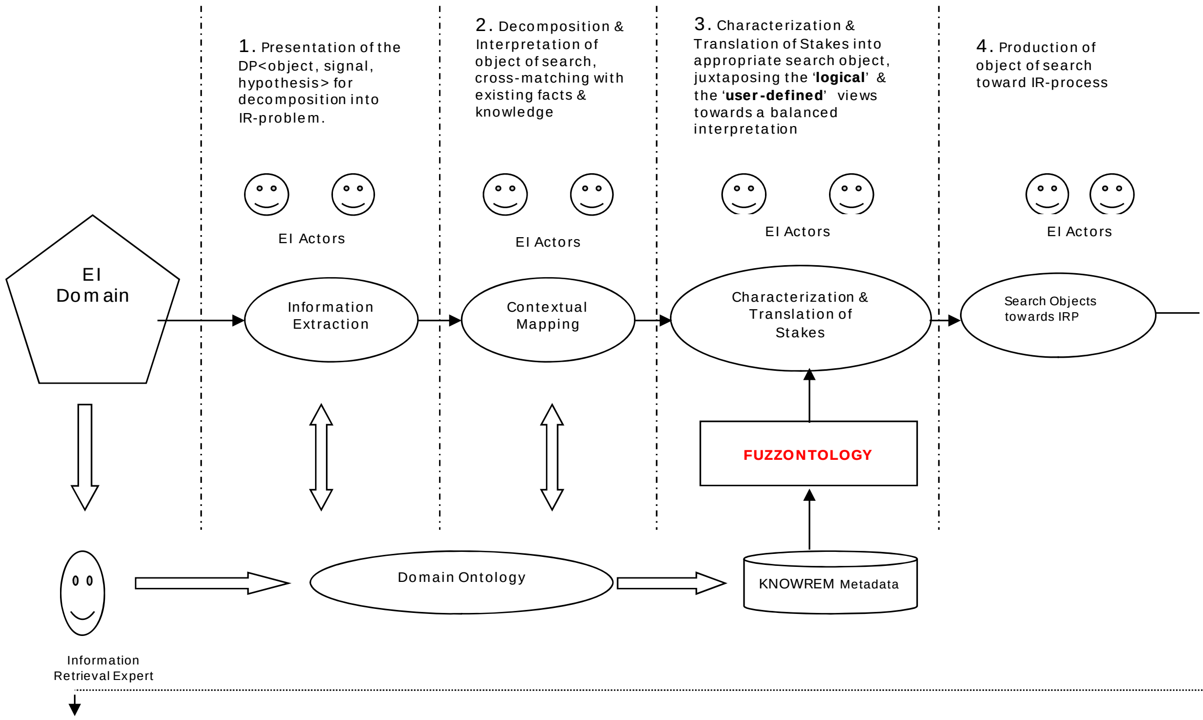


Figure 5.1: Embedded fuzzy in Ontological framework – Fuzzontology (Extended fig 4.7)

The importance for the combination of fuzzy with ontology again can be reasoned and argued from existing tools like the SWOT analysis tool and the Structured Analysis of Competing Hypothesis (SACH). In the later part of the discussion, we hope to compare the operations of these two analytical tools based on some parameters with our designed fuzzontological tool. The above mentioned tools made no candid provision for handling the inconsistencies in human interpretation and subjective sense of judgment.

The arguments above were the rationale for the attempt to embed fuzzy in the existing ontological framework. In figure 5.1, we present the embedded framework with the point of fuzzy operation clearly distinguished by marking it in red. The framework has been properly defined in chapter four alongside other fuzzy concepts like membership function, rule generation, Fuzzification and defuzzification. This section will employ earlier presented notion to present development and our simulated results.

5.2.1 Design of Membership Functions for Linguistic Variables

Earlier on, we described the importance of our choice of fuzzy logic in the resolution of inherent ambiguity and imprecision amongst other factor which is capable of resulting into risk in decision making. Consequently, we have designated four factors representing the linguistic variables to be employed in this regard. We employed four principal factors (although the list is not exhaustive) for this purpose and these include:

Organizational Need (OrgNeed)

With this factor we hope to evaluate or represent the general notion of the principle of SWOT. Thus for any decision purposes, the decision maker is expected to understand the major components discussed below based on SWOT as credited to Humphrey, (1970).

SWOT analysis is a strategic planning method usually employed to evaluate the

- Strength – the attributes of the person or organization that can facilitate achieving the objectives of the organization;
- Weaknesses – these are attributes of the person/actors or organization that are inimical to achieving desired objectives;
- Opportunities – these represents external conditions that can assist in achieving the objectives; and
- Threats – are the external factors that are inimical to the overall objectives as regards the decision process.

Environmental factors (EnviFac)

This factor is on many occasions considered alongside the SWOT analysis. Herein however, we decided to split it to drive home the importance of this factor. We reiterate that none of these factors considered is less important and therefore requires a proper understanding of them as an individual and a group. Presented in figure 5.2 is a combined membership function for the above mentioned factors. Earlier on we mentioned that partitioning of the world of discuss is not limited to any number, but here again for simplicity, we partitioned them into three (3) each.

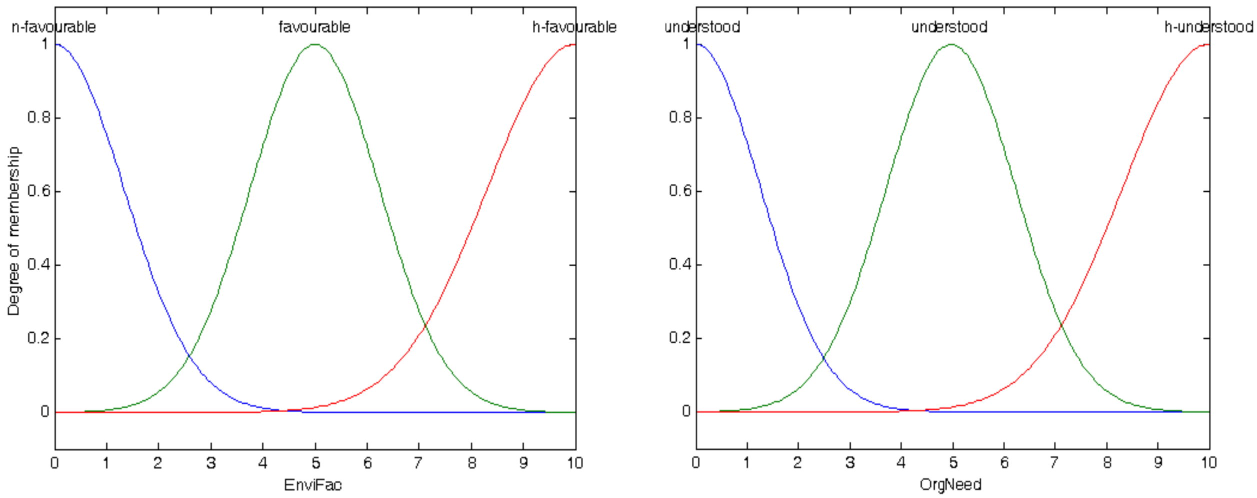


Figure 5.2: Membership functions for 'Organization's Need' & 'Environmental Factors'

We viewed the environmental factors from both internal and external factors. Internally, we are interested in the strengths and weaknesses inherent and internal to the organization, while externally we are interested in the opportunities and threats facing the organization as posed by the environment in which they found themselves. The design of our fuzzontological system thus employs the following categorization for the membership function: for the organizational need we adopted: *well-understood, understood and not-understood* as the three partitions for the membership functions. Environmental factors have: *highly-favourable, favourable and not-favourable.* The understanding and exploitation of the opportunities and threat is extremely important and highly determinant on some other factors which we shall shortly consider as the other two designated membership functions. As an example in an attempt to analyze the market position of a medium consultant group the following were observed and taking into consideration as shown in table 5.1 below

Strength	**Weaknesses**	**Opportunities**	**Threat**
Reputation in market place where they operate	Lack of consultants at operating level rather than partner level	Established and well positioned for medium jobs	Established larger firms operating at medium level
Expertise at partner level in HRM consultancy	Inability to undertake multi-faceted assignments due to staff size, fund, technical know-how etc	Identified other functional area in consultancy other than HRM	Incursion of other upcoming medium-sized or smaller firms into the scene
Existing proofs of successful track records in handling similar jobs			

Table 5.1: SWOT analysis of a medium consultancy firm

Analyzing the above factors might not be a serious problem if the necessary tools are on ground. In fact, our ontological framework can adequately capture these facts during the knowledge reconciliatory stage, however, the problem usually arise when these factors are to be taking holistically towards resolving decision problem. As an example, we employ similar background to SWOT analysis in specifying our factors. These are employed by the decision maker and the watcher during the decision problem deliberation to produce the information retrieval problem. Via ontology, each phase of our model is linked to the domain ontology, where similar problem in the past were kept for future access. The phases also interact with the EI domain to understand the problem definition, phases concerned and possible risk factors. We took cognizance of these factors because in the analysis of decision situation, *it is not impossible that strengths might be underestimated and threat over-bloated thereby resulting into a loop-sided evaluation*.

This model assists the decision maker and his watcher to model and present their understanding in a manner that is robust and adaptive. It has been clearly demonstrated in earlier models proposed at SITE-LORIA, France that converting signals to stake requires serious understanding and insight, thus, this model assist the decision maker to act based on available information and other factors. A glaring and recent example is the available information to the US intelligence about the purported suicide bomber. With available information at the disposal of the intelligence body (the watchers), it was impossible for them to adequately translate the evidences into stake, and thus great risk.

Fuzzontology can assist based on the two incorporated domains and the fuzzy interpretations employed. We earlier commented on the importance of guiding against the accruable risk that could result from high influx of information for decision process. It is a common understanding that decision maker will continue to seek for more information in a bid to resolve a decision problem. However, the danger in this approach is that, at a point, the human mind cannot adequately process this information. Cognitive overload was earlier identified as an inhibiting factor in decision making, this factor thus lead the way into the inclusion of intangible factors like root of biases and intuition as other examples of linguistic variables capture in our model. Below in figure 5.3 we present the other two factors.

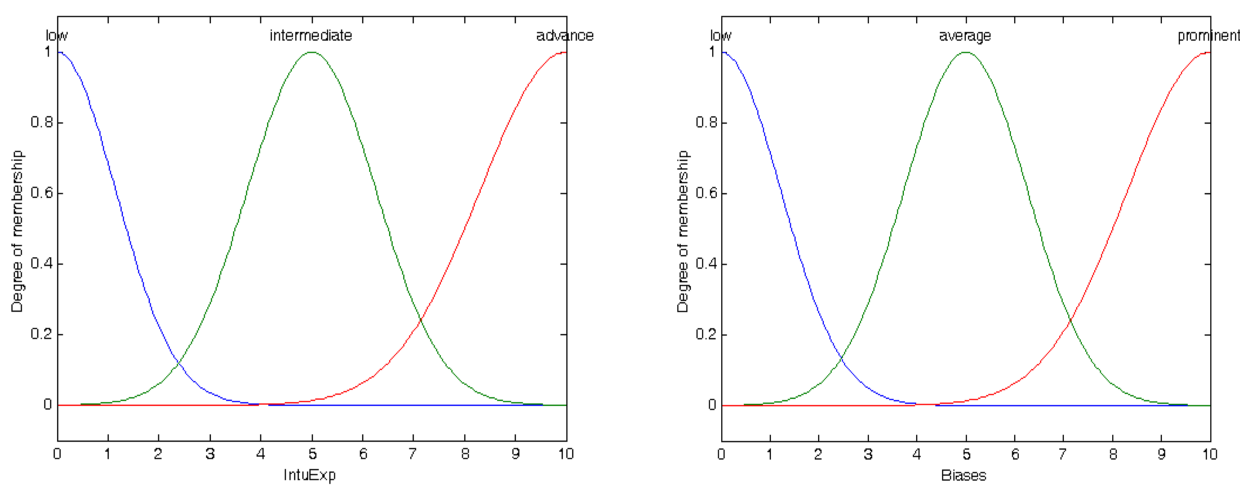


Figure 5.3: Membership functions for 'Intuition & Experience' & 'Biases'

Intuition and Experience (IntuExp)

The application of rational reasoning is brought to question once subjective judgment is involved. To this end, intuition play unequivocal role in our ability to decide. Issues like selecting a meal to eat, what to wear or choice of a life partners are examples of the role of intuition and inadequacy of rational reasoning. Einstein, (1909) declared that "*the only real valuable thing is intuition; there is no logical way to the discovery of these* *elemental laws. There is only the way of intuition, which is helped by a feeling for the order lying behind appearance*". There abound many coined out names for the use of intuition – it is considered as 'gut feelings', 'sixth senses', 'inner sense', 'instinct', 'inner voice', and 'spiritual voice amongst other.

Harper, (1990) stressed that intuition is not the opposite of quantitative analysis, nor is it an attempt to eliminate quantitative analysis. The need to understand and employ intuition results from the fact that strategic business decisions are devoid of complete, accurate, and timely information. We agreed and stressed that the focus of our work is in tune with Harper's opinion. Thus, the inclusion of the factor tagged intuition and experience can be associated with the submission of Isenberg (1984) and Simon (1987) summarizing and linking the two as – intuition being a non-conscious, quick pattern recognition and synthesis of past professional experience and expertise i.e. it is possible to circumvent analysis in favour of holistic scanning of memory for similar event or situations based on experience. We categorized "IntuExp" as either 'advanced', 'intermediate', or 'low' depending on the level of education and cognitive experience.

Root of Biases

The last of the linguistic variable employed for this work is 'biases'. Pretty well its inclusion in decision analysis has not receive popular support, however, whether we like it or not, it is a factor that greatly influence the way we decide and also a difficult one to adequately measure. When judging the likelihood of potentially positive outcomes, human beings have an overwhelming tendency to be overoptimistic or overconfident: they think that the future will be great, especially for them. Almost all of us believe ourselves to be in the top 20 percent of the population when it comes to driving, pleasing a partner, or managing a business (May, 2006). In the making of strategic decisions, optimism not only generates unrealistic forecasts but also leads managers to underestimate future challenges more subtly – for instance, by ignoring the risk of a clash between corporate cultures after a merger.

It is common belief that most people have been around people long enough to have a pretty idea of what drives their decisions, and when decision makers deny the obvious people get miffed (Gilbert, 2006). There are almost unending lists of biases to the list of cognitive biases (Wiki, 2009) the effect of these are usually a major point of concern for evaluation of decision making. Although many would argue otherwise, but it's important to know that bias beclouds the sense of objectivity in decision making and endears conclusion based on feelings than reality e.g. a doctor scoffs at the notion that a gift from a pharmaceutical company can adequately motivate him to prescribing it to a patient, or a justice is more disposed to favour a client where he or she have a stake. Succinctly put, bias is failure to fully inform about the true state of a situation. This factor forms a base for our decision making and thus must be included in evaluating the rationale behind decision. Unfortunately, it is intangible, imprecise, and difficult to qualify, thus we employ fuzzy principle to include its effect on the overall operation. We designate it by 'prominent', 'average' or 'low' (figure 5.3).

Translation Credibility

The last of all the linguistic variables is the 'Translation Credibility'. Earlier on in chapter four, we stressed the importance of this factor as the determinant of whether or not information retrieval operation will commence based on the understanding of the information specialist – watcher. This factor is of importance in the whole retrieval process. We believed that, there is no point commencing retrieval except a thorough understanding of the problem definition and the information needs have been adequately considered and refined. It is factor based on fuzzy values rather than precision. It can be asked 'how much understanding do you garner from the discussion'? this type of question are not as easy as they may appear, more so if it involves real life critical decision like controlling missiles, administering anesthetics e.t.c.

In figure 5.4, we described TC membership function differently from the existing one and provide the reasons for this.

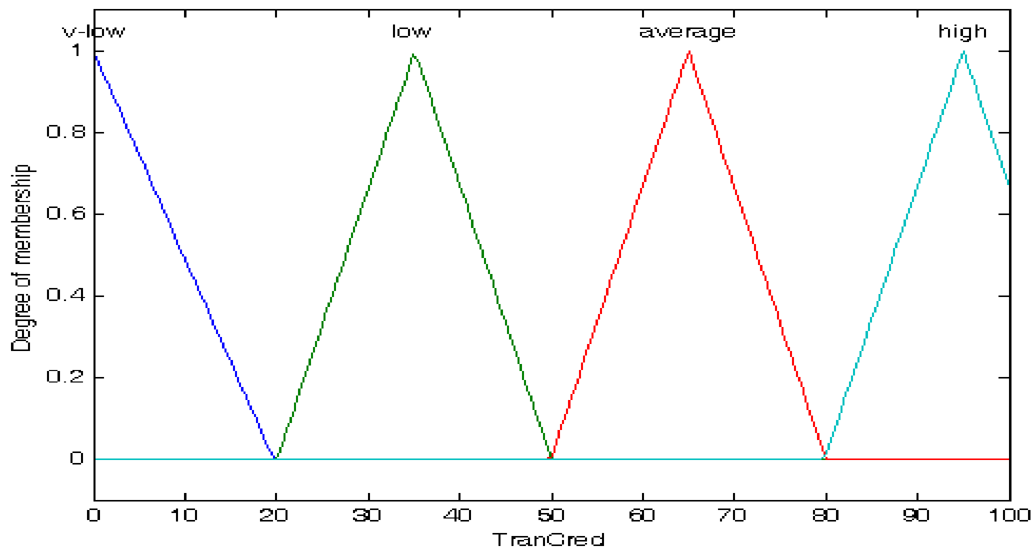


Figure 5.4: Fuzzified result – Translation Credibility

One important thing to note is the difference between the membership function types for 'TransCred' which is "Triangular" in shape. Hitherto, the type employed for earlier design was "bell shaped" with a hump usually referred to as Gaussian membership function *(gaussmf)*. In *gaussmf*, the deciding factor is the distribution of the factor under consideration, i.e. if there is no clear cut point or sharply divergent position, then *gaussmf* will represent the scenario better than *trimf* where there is a pointing edge. The essence of *trimf* is to be able to adequately pinpoint the point at which the effect is noticeable. Thus decision making here is partitioned into a range of '100' in which '50' marks the average i.e. the minimum requirement of understanding based on the multi dimensional variables under consideration.

In figure 5.4, we represent the effect of the fuzzified linguistic variable earlier mention and tagged the result as either 'very low', 'low', 'average', or high. This result trigger of full information retrieval process as the case may be.

5.2.2 Generated Rule Viewer

Fuzzy rule viewer provides ample opportunity to manipulate the rules used in designing the fuzzy system. Mamdani fuzzy based systems are predominantly rule based system. In the design of this system each of the four linguistic variables (y) has three membership functions (x). We can represent this as x^y and this result into 81 numbers of such rules developed via decision tree. Figure 5.5 below give an example of the rule viewer resulting from the Fuzzification of the four inputs.

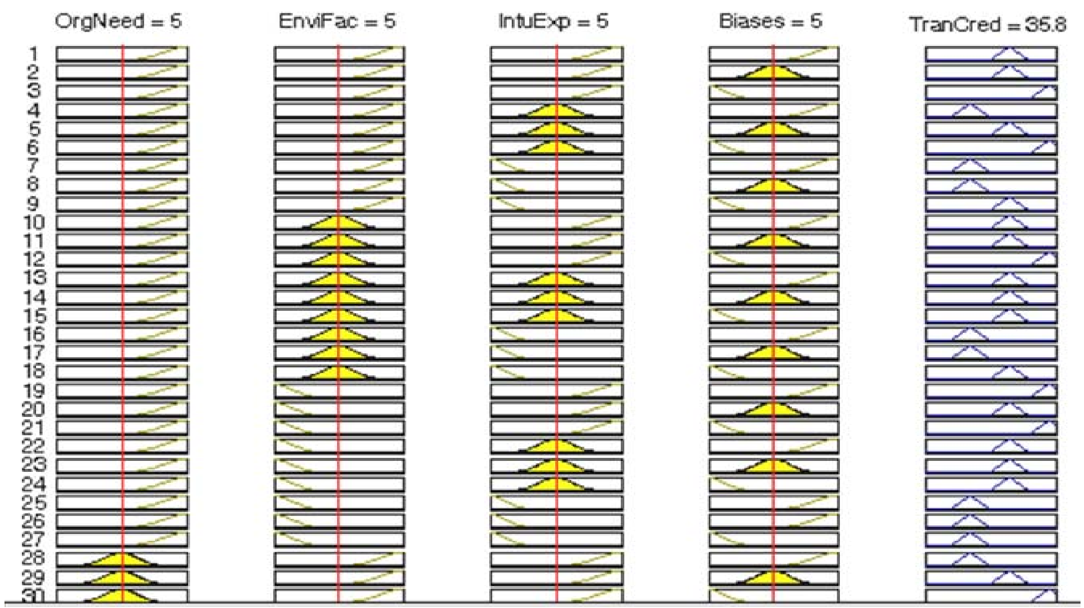


Figure 5.5: Rule viewer for the input variables and the defuzzified output

Fuzzy inference system interpretation is facilitated by the rule viewer. It presents the manner by which the shape of a particular membership function influences the overall result. Figure 5.5 captures the input variables with the intermediate value representing the default evaluated value for the system. The red line in between the four variables provides opportunity for adjusting the value of each of the factors to view the effect globally on the whole system. The last column on figure 5.5 represents the defuzzified result of the input variables – translation credibility (TC). With the default value for all the input variables, the TC is barely 35.8 out of 100. This showed a complete view of all the factors and the 81 rules represented in the systems. With this, we can at every point in time have a holistic view of the whole system which we consider a herculean task for human based on the dimension under consideration. The rule viewer assists in the generation of the surface viewer which is the next presentation.

5.2.3 Manipulation of Linguistic Variables to generate Surface Viewers

Apart from the simplicity involved in fuzzy systems, another exciting factor is the pictorial representation of its components which brings about better view of the concepts. Surface viewer is a provision in fuzzy logic system which is equipped with popup menu for selecting any two inputs and one output for plotting. There is however a limit to the dimension it can adequately handle without trouble in displaying result.

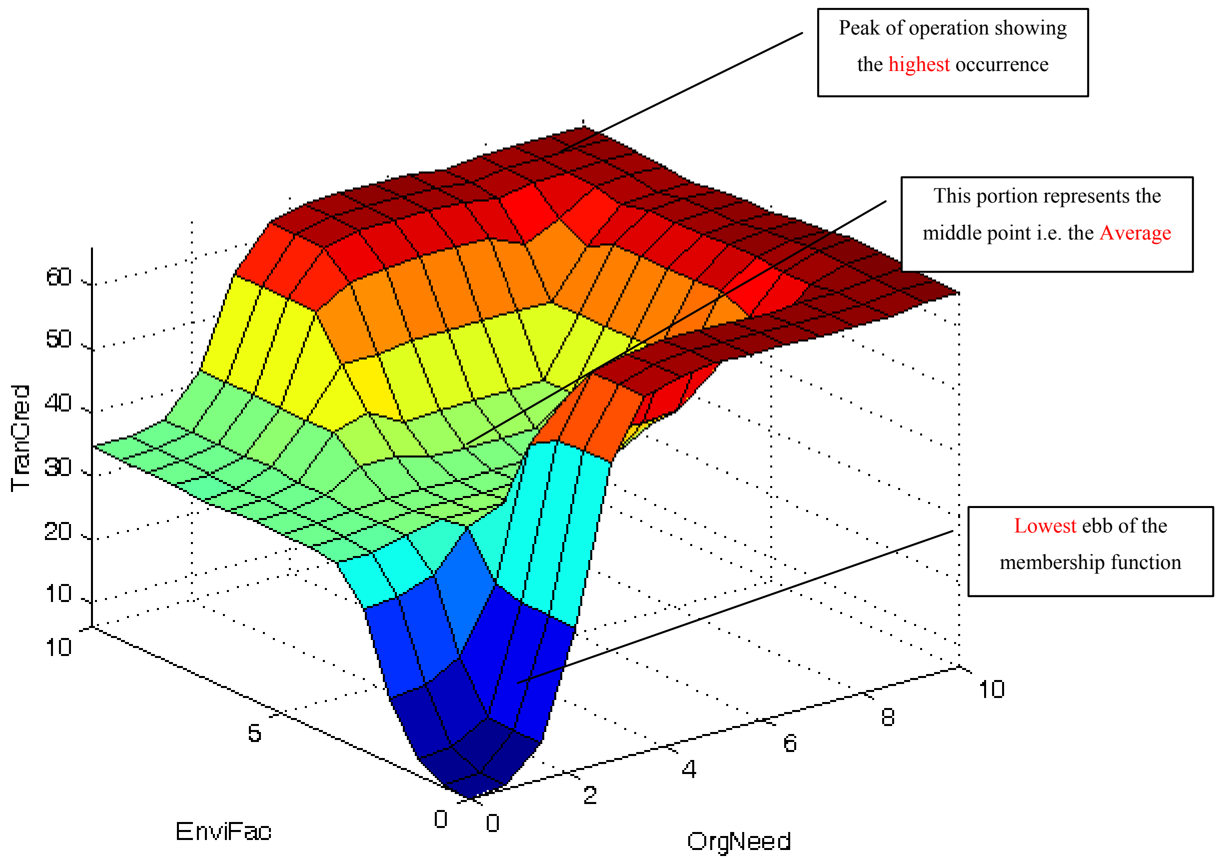


Figure 5.6: Surface viewer result for EnviFac and OrgNeed

In figure 5.6, we present the result for a two-inputs, one-output plotting giving a three-dimensional figure. It is interesting to note that any of the axes can be grabbed and repositioned for a different view of the three-dimensional view. The three labels at the right side of the figure are actually for extra information towards discussing the result –

they are not part of the surface viewer. Thus figure 5.6 depicts the result of environmental factors and the organizational need as the two affect the translation credibility (TC) – the output of the fuzzy system. The colouration of the surface viewer provides ample opportunities to discuss the result in a progressive manner. Reflecting from the discussion on membership functions, we have hitherto designated each input variables with three membership function for simplicity. This is clearly shown in the figure as the blue colour at the lower part show the duo of *EnviFac* and *OrgNeed* fussing at zero and resulting into a low level for the corresponding *TranCred* that hints the decision maker and the watcher that the knowledge reconciliation process is not successful. This warns the watcher ahead of time that, information retrieval process cannot start at this point. The duo thus has to reconsider or reappraise their positions to facilitate a better result. With different shades of blue, surface viewer depicts that there cannot be considerable progress until the two input factors reach average value.

At the point of average, the green colours are not evenly spread but wavy and the various shades of green results from the incursion of the yellow colours preceding the highest peak at red colour. It can be observed also that the value of *TranCred* is still at 35.8 which was the calculated average for all the values when all inputs were at average. As the values of the factors increases, we see a domelike shape and the colouration moving from yellow to orange, to red and a maroon. As it was the basis for the adoption of fuzzy in this research, we can see how imprecision can affect the overall act of decision making. It is not impossible to find a formal manner of achieving this operation, but fuzzy provides a mean for quick recalculation, and accommodation of intangible factors.

Our next discussion is centred on *IntuExp* and *OrgNeed* as the input factors considered in figure 5.7 while the result is still remain *TranCred.* Following the discussion above, we see a slight difference in the result here based on the importance and the designated membership functions for the factors. A sound knowledge of the organizational requirements is important for any decision making, but will it be adequate based on the level of intuition and experience of the decision maker? This is a million dollar question which is difficult to answer, unfortunately rational decision model will fail in this regard too because measuring decision makers level of experience is difficult. This is very evident in the ways managers deliver in corporate organizations. Intuition and experience are not thought in the four corridors of university's classroom, instead it is gathered as one grows in the delivery of his/her duties. This same factor is depicted in the classification of ranks within an army (Captain, Major, General etc). In all these examples, the intuition and experience which corroborate our supported model for decision making (integrated model) will be very handy.

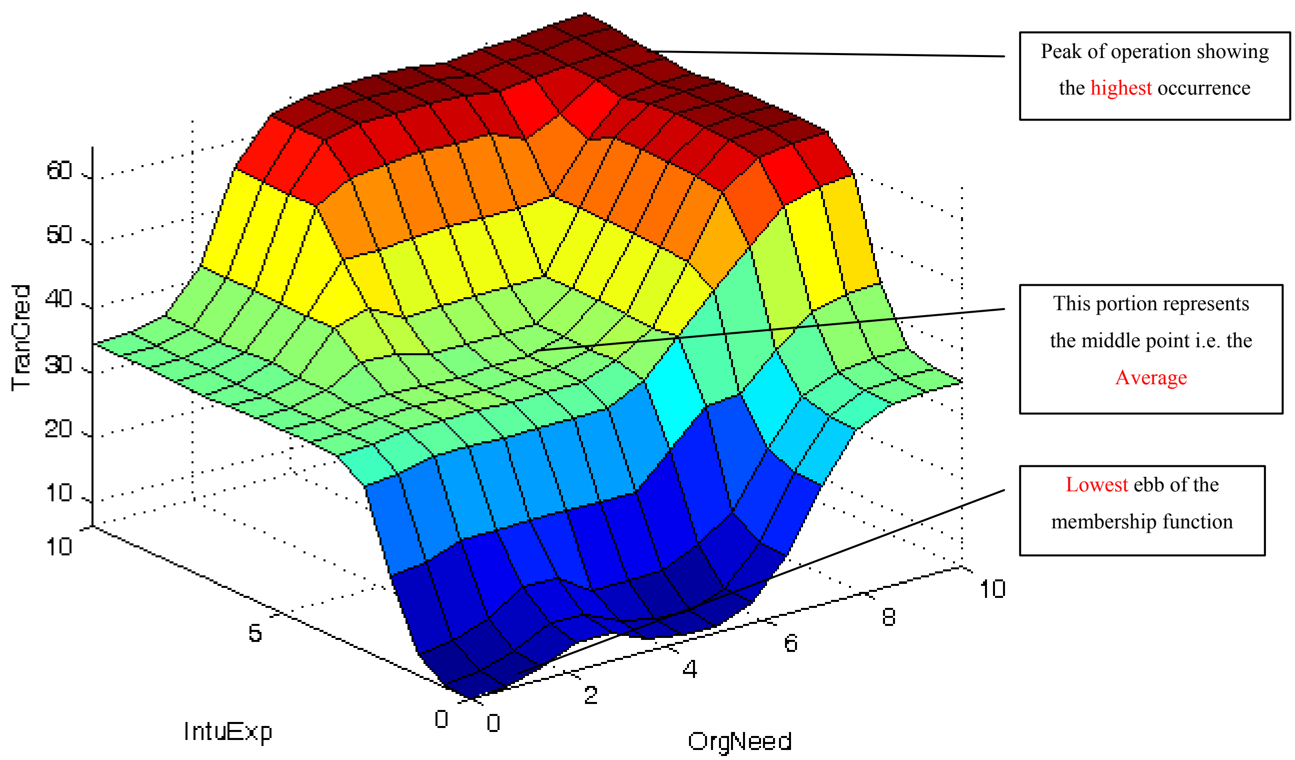


Figure 5.7: Surface viewer result for IntuExp and OrgNeed

The role of intuition is brought to play in figure 5.7. Similar to what obtains earlier in figure 5.6; we agreed that a low value for both of the factor cannot gear appropriate TC. We observed however that with a gradual growth in the knowledge of the *OrgNeed*, corresponding growth was not witnessed on the part of *IntuExp*. This result further corroborates the inadequacies of rational decision making in the face of uncertainty. We found out that at a value around the average, the *TranCred* did not have a corresponding growth because *IntuExp* has not adequately improved. This helps warn the decision maker or the watcher to understand and take into cognizance (as possible risk factor) the experience and intuition of either party when decision are to be made. At the core middle, we see the result of *TranCred* at 35.8 grossly inadequate for decisive action, but requiring a more rigorous deliberation to improve the value of *TranCred* before embarking on the next stage in the economic intelligence process stages – information retrieval. The importance of intuition is further depicted in the colour distribution of figure 5.7; the blue colour persists longer although with a different shade into the higher realm of the membership function for *OrgNeed*.

The above factor led to the presence of green colouration (average) even at the peak of *OrgNeed* definition. This colour changes to yellow at the same position delaying the appearance of orange and red colours. Using the foiled bombing of the US bound plane as an example, it is glaring that what the *intelligence authority lacks is ability to determine the level of translation credibility from the available information*. Several factors might be responsible, in this case, taking experience and intuition into focus, we can say that if the decision maker in this case has been experienced in this regard and using high level intuition to map available information, the passenger should have been given a thorough check. This simulation thus resonate our support for the importance of intuition in decision making. This surface viewer can be adjusted appropriately until a desire result is gotten or even change the dimension or the value at the axis. With some adjustment and the signal (name already in the database), appropriate manipulation can adequately warn on the need to appropriately perform proper screening on available signal corresponding to a signal. Next we look at the effect of intangible factor like bias and its effect in delivering strategic decision making.

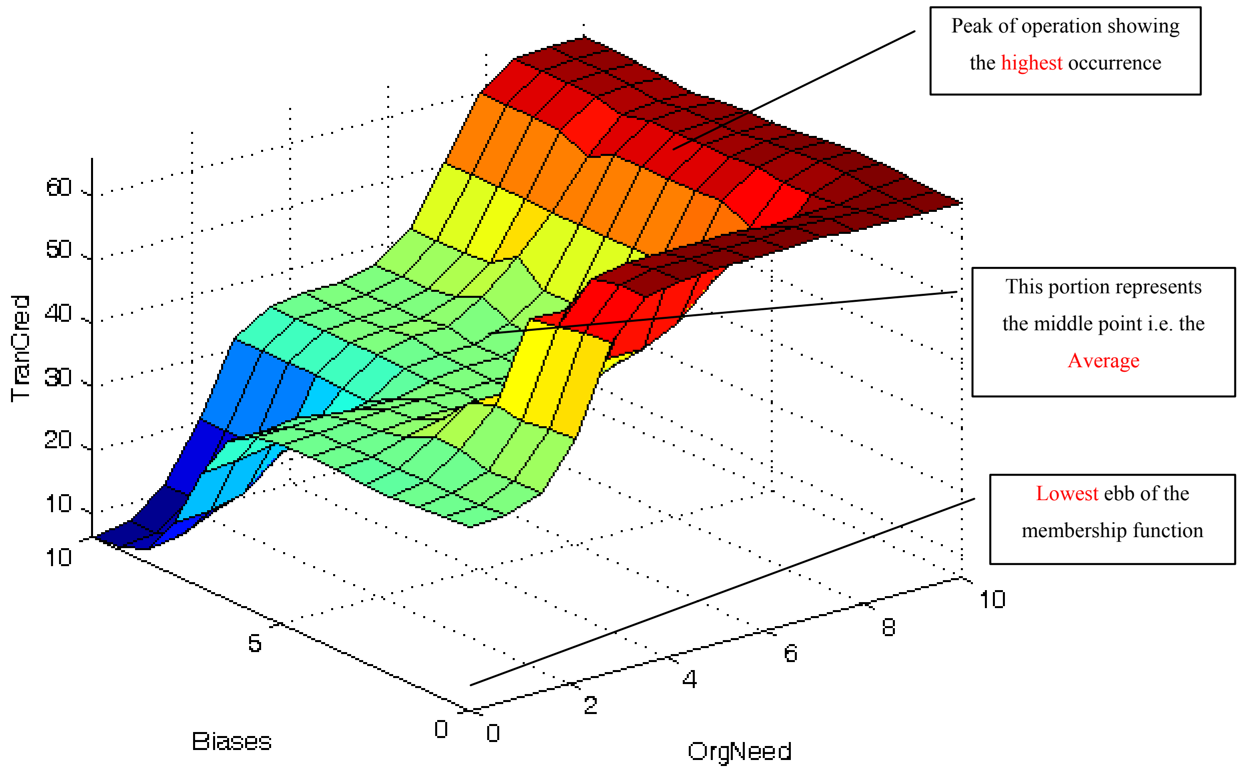


Figure 5.8: Surface viewer result for EnviFac and OrgNeed

Most decision process and result evaluation fails to include the effect of biases and personal preferences. While many of these procedures recognized the importance of these intangible factors in effective delivery of decision, capturing and utilizing them becomes a serious challenge. Experiments showed that a set of judges evaluating student moral and academic performance usually attempt to find something good to say about their student once they are familiar or know the student – bias, but to others, they tend to be more focused and accurate in their judgment.

Root of biases and personal preferences are integral part of human decision making. This includes other factors like favouritism whose effect are not only difficult to measure but also to capture. Bias is naturally taken as being negative, and towing this line of reasoning we can see a sharp turning in figure 5.8 as opposed to the other two figures. The prominent rate of bias displayed forces the blue colour to the high side of the chart. Clearly captured is the fact that, not minding the level of understanding of the organizational need presence of bias really distorts the decision making process. The prevalent rate of bias could not allow the knowledge about the organization displayed to have any significant effect. It is again interesting to see that if it is almost possible to remove bias, the *OrgNeed* stay put at half the chart shown in green colour. The gradual increase in the value of the green (moderate/average) factors is observable with its inclusion around 7 on the bias axis.

Fuzzontology (Onifade, et al., 2010), attempts to capture the lapses encountered in the ontological framework earlier proposed for knowledge reconciliation in economic intelligence. This is sequel to our earlier identified factor called translation credibility which assists in monitoring the derivable understanding between the decision maker and the watcher prior to the information retrieval stage. This we believed will help monitor the understanding gathered from the piece of information (decision problem definition) emanating from the decision maker with what the watcher can make out of it. To further exemplify the operation of the design, we shall thereafter make some comparison among SWOT, SACH and FuzzOntology.

5.2.4 Evaluating SWOT, SACH and FuzzOntology

Interpretation is the driving factors in the delivery of any logical decision. Following closely the infological equation; we observed that no two individuals can interpret the same information the same way. SWOT, (acronym for Strength, Weakness, Opportunities and Threat) is a popular analytical tool for comparing an organization with its environment. It employs information gathered from the environment and partitions it into external and internal issues. The Structured Analysis of Competing Hypothesis was developed by the Palo Alto Research Center (PARC). Its operation is based on the formulation of simplest hypotheses, testing them, and increasing the complexities until clearer estimates about the hypotheses are achieved. It thereafter resolves inconsistencies and increase complexity if the need arises.

SWOT prided itself in ability to distinguish between internal and external issues based on information gathered from the environment. It is however important to stress our point here that comparing organization with its environment is not always presenting a clear cut facts. To this end, it was concluded that SWOT is as good as the information it contain like any analytical tool. Most of available information is subjective if not biased. How then can there be proper resolution in the subjective judgment if it is not treated fuzzily. Fuzzontology employs similar factors and developed a membership function for the environmental factors which might be difficult to be expressed precisely. Apart from this tangible factor, we also model and simulate the inclusion of intangible factor like biases and personal preferences which cannot be captured in existing tools.

With regards to the context of economic intelligence, fuzzontology will assist in resolving the ambiguity introduced during knowledge reconciliation and the point when the decision maker is faced with many alternatives from where a choice must be selected. During the knowledge reconciliation stage, the actors can employ FuzzyOntology like the SWOT analysis to perform all round analysis of a clumsy phenomenon. The impressive thing about this approach is the ability of the decision maker to define their needs in linguistic term and at the same time being computable. Fuzzy takes in the linguistic variables and formulate through defuzzification a combined result which is crispy for use.

Factors / Tools	Problem assessment process	Mode of problem resolution	Factor representation	Refinement method	Flexibility	Provision for intangible factors
SWOT	Partitioned into 4-quadrants	Matrix	Discrete / Probability	Piecewise	Not-flexible	No
SACH	8-steps process method	Matrix	Figurative (+, -, N/A)	Piecewise	Flexible	Yes
FuzzOntology	Flexibly defined by the users	Fuzzy inference method	Membership functions with (linguistic var.)	Wholesome	Flexible	Yes

Table 5.2: Comparing SWOT, SACH and FuzzOntology

Table 5.2 represents a summary of operations among SWOT, SACH and our designed FuzzOntology tools. It is evident from the above that, while the representation is wholly a human affair, it is also important to assist in the manipulation of available information in the best manner to adequately present the picture close to reality. Pattern of representation of both SWOT and SACH are too strict to facilitate flexibly define decision problems. SACH figuratively represents the identified factors as being consistent or not by positive or negative (+ / –) sign. SWOT however employs either discrete numbers or calculated probability, again from human judgment which has been subjectively created. This information generated a matrix structure which are resolve to inform the user/decision maker. The first thing to note here is the mode of refinement (pair-wise). Attempts like this can result in omission of important attributes in the final computation. Human ability to retain a lot of information concurrently has been a point of discussion. This was equally pinpointed during our discussion on decisionability, i.e. with increasing information at the disposal of a decision maker; the likelihood is high to do with whatever comes in handy and not necessarily considered in the best proportion to reflect the real state of the situation.

In SWOT analysis, there was no direct provision for intangible factors like preferences and biases since it employs probabilistic approach to designate occurrence or otherwise. SACH employs analysis drawn from scientific methods, cognitive psychology, and decision analysis thereby inculcating human factors into the overall analytical process. Fuzzontology employs linguistic variable to facilitate humanly expressed word like "fairly favourable, very rugged" e.t.c. which cannot be adequately quantified either by SWOT or SACH. In other to guide against possible risk factor like misinterpretation or misrepresentation, we employed fuzzy logic techniques to capture the intangible and imprecision constituting ambiguity in interpretation based on hitherto common vocabularies via ontology.

With the success of this stage in decision making, the stage is set for exclusive information retrieval process aimed at assisting the decision process. It should be pointed out that, it does not necessarily imply that there has been some level of information search during the reconciliation stage. However, retrieval done at that stage we believed were used to further fine tune the problem definition and information need definition.

In the next section, we believed that the success of this stage triggers the next – information retrieval process, our focus is in guiding against the concept of "no information in the presence of information". A situation that may arise based on outright wrong spellings and/or alphabetical transposition which form part of the missing data problems.

5.3 Information Retrieval, Search Engines Operation and Query Matching

The concept of "*no information in the presence of information*" is actually a rat race which can be described as a fool being thirsty in the abundance of water. Large corporations and government based their functionality on the volume of information available to them. It is however a sad story that despite the volume of available information, they sometimes experience wrong results or no result from their database. We take a look at information retrieval process based on query matching ability of search engines in this section. Typical search engines retrieve information based on keywords given by users and return the information found as a list of search results. Keyword-based search engines have a weakness despite their popularity, in that they often return a large list of search results with many of the top list of search results being irrelevant. This problem can be trivially avoided if users know exactly the right query terms to use (Widyantoro & Yen, 2001). These terms represent unique or at least very specific ones, causing the search engines to bring only the relevant information to the top list of search results. However, such query terms are often very hard to find considering the mirage of dirty data possibility (Kim, et al., 2003), and in most occasions, they do not even exist. In order to get the information needed, finding the right query terms can become additional task during information seeking activity (Burgess, et al., 2007).

Endowed with various levels of functionalities depending on the size of the organization, search engines facilitate easy and fast information retrieval. There exist software packages for search engine construction on the Internet. The website searchtools.com alone lists more than 170 search tools, many of which are free or free for noncommercial use. Most search engines operate on the principle that pre-indexed data is easier and faster to search than raw data. The form and quality of the index created from the original documents is of paramount importance in determining the pattern of searches. The commonly used indexing method is the full text inverted index. It takes a large amount of disk space and the indexing process is slow, because it keeps most of the information in a document. A variety of this method is to index only the title, keywords, description, and author parts of a document. With this, the indexing process can be faster with smaller resulted index. Another important feature of search engines is Relevance Ranking. This ranking method refers to the method that decides a document's relevance to a query. Factors such as word frequency in the document, word position in the text, and link popularity are usually considered. PIPER-project, (1998) also include information classification, query interface and query construct amongst other.

Web search engines work by storing information about many web pages, which they retrieve from the WWW itself. These pages are retrieved by a ***Web crawler*** (sometimes also known as a ***spider***) — an automated Web browser which follows every link it sees. Exclusions can be made by the use of robots.txt. The contents of each page are then analyzed to determine how it should be indexed (for example, words are extracted from the titles, headings, or special fields called meta tags). Data about web pages are stored in an index database for use in later queries. Some search engines, such as Google, store all or part of the source page (referred to as a cache) as well as information about the web pages, others like AltaVista stores every word of every page they find. This cached page always holds the actual search text since it is the one that was actually indexed, so it can be very useful when the content of the current page has been updated and the search terms are no longer in it. This problem might be considered to be a mild form of linkrot, and Google's handling of it increases usability by satisfying user expectations that the search terms will be on the returned webpage. This satisfies the principle of least astonishment since the user normally expects the search terms to be on the returned pages. Increased search relevance makes these cached pages very useful, even beyond the fact that they may contain data that may no longer be available elsewhere.

When a user enters a query into a search engine (typically by using key words), the engine examines its index and provides a listing of best-matching web pages according to its criteria, usually with a short summary containing the document's title and sometimes parts of the text. Most search engines support the use of the boolean operators AND, OR and NOT to further specify the search query. Some search engines provide an advanced feature called proximity search which allows users to define the distance between keywords.

Considering from the point of utilization, the most important thing to any user is the relevance of the result provided for the search query. While there may be millions of webpages that include a particular word or phrase, some pages may be more *relevant, popular, or authoritative* than others. Most search engines employ methods to rank the results to provide the *"best"* results first. How a search engine decides which pages are the best matches, and what order the results should be shown, varies widely from one engine to another. The methods also change over time as Internet usage changes and new techniques evolve.

Most Web search engines are commercial ventures supported by advertising revenue and, as a result, some employ the practice of allowing advertisers to pay money to have their listings ranked higher in search results. Those search engines which do not accept money for their search engine results make money by running search related ads alongside the regular search engine results. The search engines make money every time someone clicks on one of these ads. Revenue in the web search portals industry is projected to grow in 2008 by 13.4 percent, with broadband connections expected to rise by 15.1 percent. Between 2008 and 2012, industry revenue is projected to rise by 56 percent as Internet penetration still has some way to go to reach full saturation in American households. Furthermore, broadband services are projected to account for an ever increasing share of domestic Internet users, rising to 118.7 million by 2012, with an increasing share accounted for by fiber-optic and high speed cable lines.

The focus of this section lies in understanding the risk factors in information retrieval process, and probably proposes a novels manner for information need representation and reasoning. We have identified basic manner via which risk of "no information in the presence of information" could result, these include: *Query representation* (how best do we represent the queries for internal mapping), *Interpretation* (knowledge reconciliation among the actors in EI), *Lexicographic* (issue of misspelling or transposition), *Syntactic* (arrangement of components, streams of character representation), and *Phonology* (how do we pronounce, a societal, cross-cultural and environmental factor). Thus, the next section discusses query syntax as regards to existing search engines and we build the discussion from that point.

5.3.1 Querying Syntax and Search Engines Efficiency

The method of query submission usually follows the prompt for users' query. These query input determines the success of retrieval operation. It is therefore important that strict attempt should be made to the supply input which includes one or more keywords describing required information. This method many times does not supply desired result. It is therefore important to carefully construct users' query or provide modes via which user inputs can be refined towards adequate retrieval process. The listed factors represent some of the important consideration in determining the efficiency of a search engine.

- The speed of the search engine especially with regards to an increase in the *SIZE* and *COMPLEXITY* of the information system in question?
- *Boolean Search* - Can the search engine look up pages containing some word and not containing some other word? Does the search engine support the AND and OR logic among query words?
- *Phrase Matching* - Can the search engine match only those documents that contain words in exactly the same sequence as that of the query?
- *Attribute Search* - Can search engine perform search within only the body, title, description, keywords, URL, or other parts of documents?
- *Fuzzy Search* - Can the search engine match documents that contain words that are similar to requested query? Are search by soundex, metaphone, or substring supported?

- *Word Forms* - Is word stemming supported?
- *Wild Card* - Is there a wild card character that can be used in search to match any one or more character or symbol?
- *Regular Expression* - Regular expressions are symbols that users add to their queries to describe complex patterns to match. Is regular expression search supported?
- *Numeric Data Search* - Can the search engine deal with numeric queries such as "Quantity > 300"?
- *Case Sensitivity* - Is the search engine case sensitive, or can it be configured as case sensitive?
- *Nature Language Query* - Does the search engine support nature language queries?
- How often is the ***missing data*** issue encountered and treated with the search engine? This question goes in phases:
  - Is the appropriate information found more often than not?
  - Is the information found relevant to the user?

The above listed factors provide that using efficient words and query string coupled with the functional provisions of the search engines; possibility of correct-hit improves for the required information. The disparity is the way via which information seeker bind their query and the supplied volume of returned information is of concern to most users. Many of the supplied links do not even exist any longer. Listed among the factors above is missing data, the problem occurrence and effect on database query shall be the focus of the next section.

5.3.2 Designation of Retrieval Activities

In the above section, we focus on the technicalities of the retrieval objects (search engine). It is however important as earlier mentioned in chapter two that understanding and resolving information problem involve a critical consideration of the representation and reasoning of both the documents and the user's information needs. Earlier on we identified user information need to be of three types: *known item information need, conscious information need, and confused information need.* These three encompasses every form of user queries and must be adequately dealt with to resolve the risk information acquisition and utilization.

In Naeza-Yates & Riebeiro-Neto (1990), information retrieval model was defined as a quadruple {D, Q, F, R(q, d)}. Again we have dealt with this in chapter two but it is also important to form a broad based discussion on this subject. With these notions, we shall thus take a look at the following information retrieval.

- Ad hoc retrieval – this is characterized by arbitrary subject of search and short duration. Common example is a researcher doing literature search in a library. In this type, retrieval system knows the sets of documents to be searched, but cannot anticipate the topic that will be investigated (Voorhess & Harman, 2001)
- Known item search – this is similar to the ad hoc search, the searcher has an idea that the desired documents exist. Typical example is the retrieval operation for all articles by "Odile Thiery". Precise query language are normally implemented for this type of search task
- Interactive retrieval – this model is based on the popular user modelling where by the system attempt to perceive user's interaction with the system and use this to modify current search strategies (Robin, 2000). Other variation of this is the classical relevance feedback approach (Rocchio, 1971).

Other task identified in Canfora & Cerulo, (2004) are Filtering, Browsing, Clustering, Gathering, Crawling and Mining. Detailed information can be found in the above cited author. We are more interested in the above three because they fit directly into the pattern of our work. In subsequent section, we shall employ these factors to depict the functionality of our developed systems and existing retrieval objects. Designated retrieval activities have formed the basis on which contextual information access, seeking and retrieval is founded. Its importance lies in the fact that, "if you can know your user, you are likely going to treat her/him in a special manner". We believed and agreed with other proponents of user centred information access that there should be a level of flexibility incorporated into retrieval activities to guide against the problem of "no result in the presence of information". These inadequacies have been linked to the kind of query and documents representation alongside mode of relevance determination employed by the system.

This section is to briefly introduce the various designations of information retrieval patterns and their importance in the functionality of the system and the users. It is a buildup for the implementation of our fuzzy model in subsequent sections. Next, we take a look at the concept of missing, non-missing data and associate problems of dirty data.

5.3.3 Missing Data, Non-Missing Data and the Problem of Dirty Data

It is not uncommon for system data to degrade rapidly, this can commence with customers information, for example names, addresses and missing information. The rate at which errors like these accumulate can be in matter of days, few weeks or might even take longer time. The unfortunate is that information from such databases becomes unreliable. Error is not limited to the size of a database or the organization, even in professionally designed, implemented and operated with strict data control, there exist errors which constitute risks inimical to the organization. The focus of this research is not in detection and removal or what is referred to as data cleaning, but improve search operation despite the level of dirtiness of the database.

We referred to dirty data as a term employed to refer to information/data that is *misleading*, *incorrect or without generalized formatting*, that has been collected by any data-capture means. This could be in form of *spelling mistake or punctuation*, *incomplete or outdated data*, or even data that has been *duplicated* in the database (Mike, 2009). Wilmes, (2009) made a categorization of possible problem of data quality resulting from dirty data in order of increasing difficulty these are validity, completeness, consistency and correctness.

Impossible phone number, nonexistence postal code, a future birth date are examples of invalid data. This type can easily be fixed than other types of dirty data. Detecting incomplete data is more difficult than invalid data, however inconsistent type may prove much more difficult to detect since it requires more inside knowledge (substitute “rules” or “metadata”. The most worrisome of these is the incorrect data. This is sequel to the fact that it is valid, complete and consistent, yet it is just wrong. Thus it will not be detected by validation, completeness, or consistency check. They are almost intractable.

Although dirty data taxonomy has been presented earlier in our discussion, we here again make a diagrammatic representation in the next section to allow vivid access to each level of dirtiness, and possible treatment. We made bold to say that, we have no intention to propose all encompassing solutions to these problems. We shall therefore delimit the problems that are paramount to this work and others will be left for future work as they are not directly in line with the objectives of this research. The next section presents our hierarchical taxonomy of dirty data.

5.3.4 Hierarchical Representation of Dirty Data

The taxonomy presented in Kim, et al., (2003) is based on the premise that manifestation of dirty data comes in three broad ways: *missing data, not missing but wrong data, and not missing and not wrong but unusable*. The last occurrence is more pronounced whenever there is database integration or when representation standards are not consistently pursued in inputting data. The taxonomy also represent dirty data that are manifested based on the combination of more than one type of dirty data (e.g. wrong order in data concatenation, misspelling – “Amos David” instead of “David Amos”).

Without such taxonomy or metric, it will remain difficult to know with a high degree of confidence the quality of business intelligence derived from data warehouses and the quality of decisions made on the basis of such business intelligence. Consequently, the primary objective and contribution from Kim, et al., (2003) is to develop a comprehensive taxonomy of dirty data. Below in figure 5.9 is the hierarchical representation of dirty data.

This taxonomy is aimed at providing a framework for understanding the origins of a complete spectrum of dirty data and the impact of dirty data on data mining, and sheds light on techniques for dealing with dirty data and for defining a metric for measuring data quality. In resonance with the submissions of Kim, et al., (2003), we define dirty data and sources of dirty data as follows.

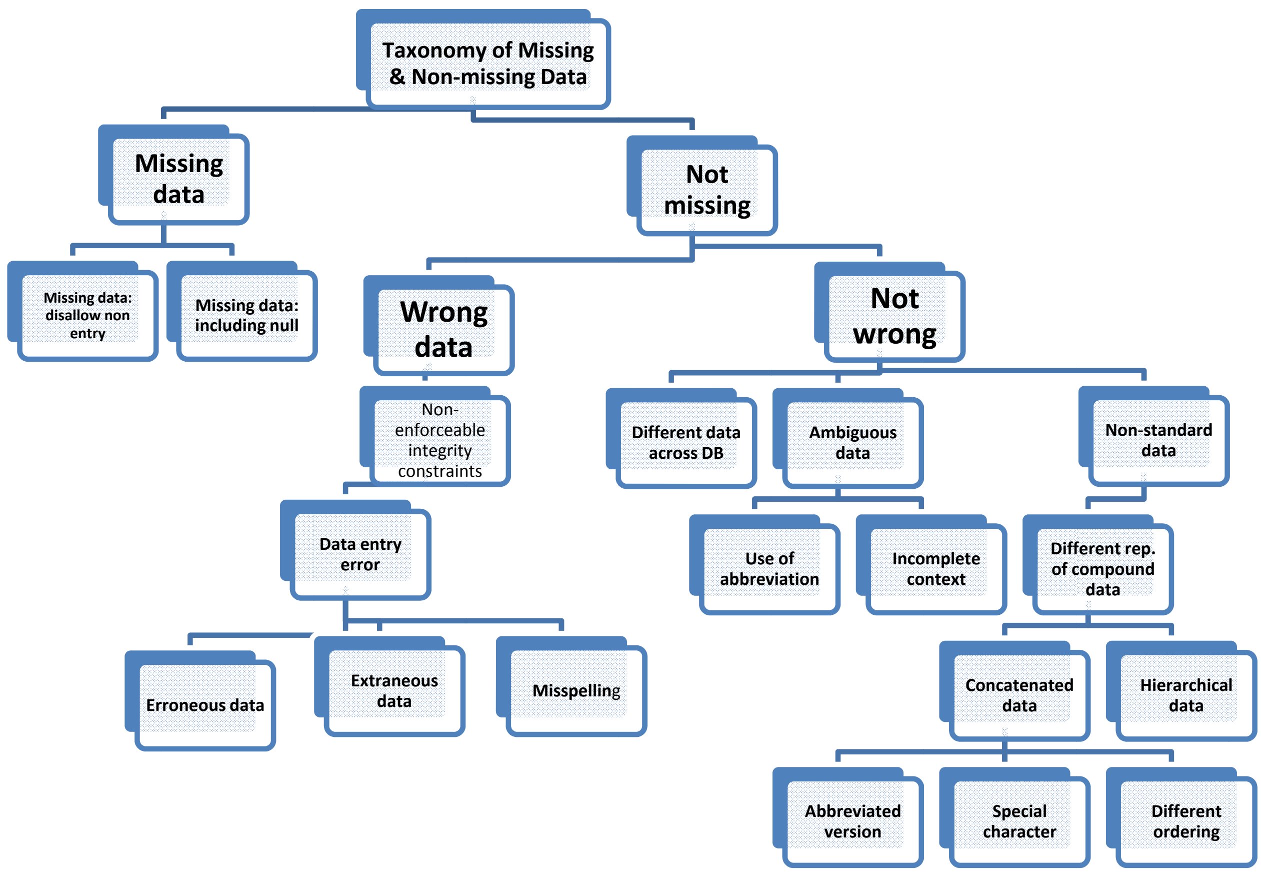


Figure 5.9: A tree structure for Data Missingness & Non-missingness (Adapted from Kim, et al., 2002)

- The life cycle of data includes its capture, storage, update, transmission, access, archive, restore, deletion, and purge. The focus of our research is on the access aspect by a user or application that operates correctly. *As such, we say that data is dirty if the user or application ends up with a wrong result or is not able to derive a result due to certain inherent problems with the data.*

- The sources of dirty data include data entry error by a human or computer system, data update error by a human or computer system, data transmission error by a computer system, and even bugs in a data processing computer system.

Again, the above is a buildup towards our proposed solution for designated classes of dirty data. In the following, we introduce our proposed soft methodology (Fuzzy) to information retrieval process. We take a look first at the operational model and from it build up gradually our proposal with some real life tests.

5.4 Fuzzy Search

The basis for information retrieval is to facilitate decision. The process of information retrieval can either be implicit or explicit based on the decision maker's 'understanding of the problem', 'familiarity with the terrain' and the 'size of the problem' among others. Decision making requiring explicit information search usually bears various forms of risk which makes accessibility to qualitative information an important factor in decision making. Sadly, with exponential growth in the volume of information in digital format, existing information retrieval algorithms, methods, technologies and tools are too constricted for adaptive and robust information retrieval activities. This factor has rubbished many attempts to inform because of inability to access adequate information on timely manner. On the part of the users, incomplete and partial understanding exist making the process cumbered with imprecision and ambiguities. On the part of the system, lack of flexible representation of queries and documents exist. These factors constitute high risk to decision problem resolution. We therefore propose in the rest of this chapter, a fuzzy based information retrieval model capable of accommodating some of inherent problems in retrieval process.

Fuzzy searching is much more powerful than exact searching when used for research and investigation based on its accommodation for imprecision and ambiguity. Fuzzy searching technique comes in handy when researching unfamiliar, foreign-language, or sophisticated terms, the proper spellings of which are not widely known or asserted.

Fuzzy searching can also be used to locate individuals based on incomplete or partially inaccurate identifying information in an attempt to deal with dirty data. A fuzzy search is done by means of a fuzzy matching program, which returns a list of results based on likely relevance even though search argument words and spellings may not exactly match. Exact and highly relevant matches thus appear near the top of the list while subjective relevance ratings, usually are expressed in percentages form.

A fuzzy matching program can operate like a spell checker and spelling-error corrector. For example, a user can types *"Misissippi"* into *Yahoo* or *Google* (both of which employ some level fuzzy matching), a list of hits is returned along with the question, *"Did you mean Mississippi?"* Alternative spellings, and words that sound the same but are spelled differently, are given. A fuzzy matching program can compensate for common input typing errors, as well as errors introduced by optical character recognition (*OCR*) scanning of printed documents. The program can return hits with content that contains a specified base word along with prefixes and suffixes. For example, if *"planet"* is entered as a search word, hits occur for sites containing words such as *"protoplanet"* or *"planetary"*. The program can also find synonyms and related terms, working like an online thesaurus or encyclopedic cross-reference tool. In the *Ask Jeeves* search engine, entering word like *"galaxy"* returns entries like *"Galaxy Photography", "Milky Way"* and *"The Nine Planets Solar System Tour".*

Fuzzy matching programs usually return irrelevant hits as well as relevant ones. Superfluous results are likely to occur for terms with multiple meanings, with one or few corresponding to the user's request. If the user has only a vague or general idea of the topic, or does not know exactly what to look for, the ratio of relevant hits to irrelevant hits tends to be low (the ratio is even lower, however, when an exact matching program is used alongside). The above serves as introduction to fuzzy search operation that will be employed for the design of our model. In the subsequent section we present a generalized model description for the search operation and discuss the system operation.

5.4.1 Operational Model Description

We pointed out earlier that our intention is to come up with a robust and flexible search engine based on fuzzy paradigm to accommodate the presence of dirty data and provide useful result instead of no match found that characterize popular search engines. In this section, we represent the search operation diagrammatically and discuss the main expectation of the user from the system. This is shown in figure 5.10 below.

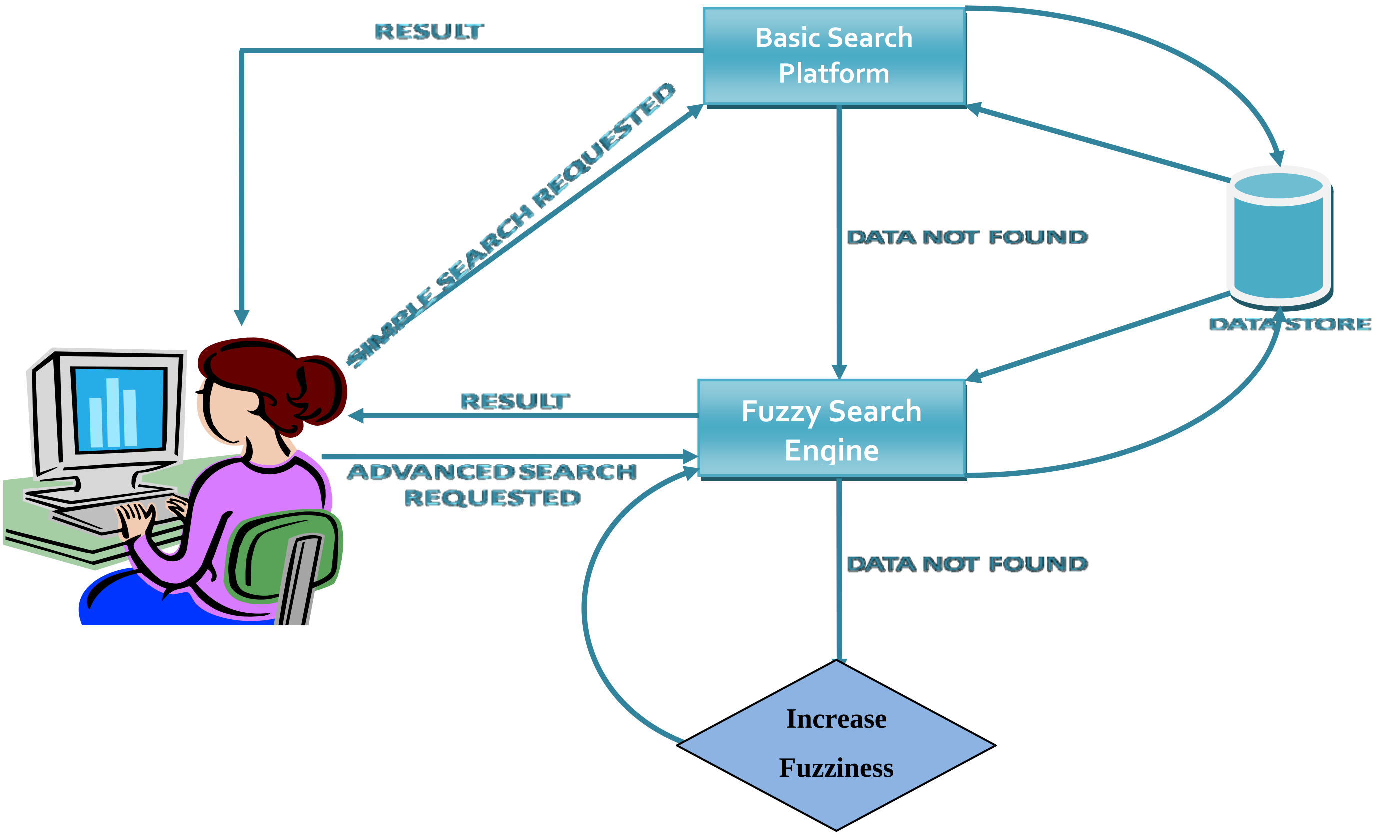


Figure 5.10: Operational model for fuzzy search operation

In figure 5.10 above, a given search operation involve the user given a particular search parameter from which the search engine is expected to provide either *exact* or *fuzzy* results. The exact results would be similar to results provided by a typical search engine, while the ability for the user to manipulate her search operation by increasing the fuzziness of the result constitutes the focus of this work. In order to accomplish this, several algorithmic submissions and earlier work on string matching by Seller, (1980) and Navarro, (2001) were considered amongst others. The proposed fuzzy search engine is expected to perform the forms of string matching listed below:

Strong Matching: In this case the user input is matched directly to an input or some inputs in the database. These matches when present should be placed higher up in the ranking of the outputs.

Weak & Fuzzy matching: In this case, the user input cannot be matched directly to the contents of the database. Herein lays the strength of the search algorithms. It is expected that in this situation, the engine should provide the best weak match or in worst case a fuzzy match and go a stretch further to provide the *'fuzzy match percentage'* which would signify the similitude between the user input and the database value being considered.

The weak matching is considered to be a '***subset match'*** or a ***'fuzzy match'.*** The earlier signifies that the content of the user input can be found (in whole) in the database input. e.g. '*Oberman'* can be found in whole in '*Hoberman'***.** The later signifies that there is no direct relationship between the user input and the database value. When this occurs, a fuzzy match is performed from the internally generated membership function by analyzing the supplied strings. A fuzzy match completely ignores the language and construct of the user input and database value. It concentrates on the actual character content and arrangement in order to establish the slightest relationship between the two values being compared. In order to provide reasonable results, adequate sorting algorithms were employed to sort the results of the fuzzy search.

User input manipulation: The strength of this work is dependent on the ability of the engine to manipulate user's inputs in order to establish some relationship with the content of the database. Probable operation for the manipulation includes the following:

- Splitting
- Concurrent character comparison
- Unmatched character comparison
- Fuzzy likeness evaluation
- Fuzzy rank evaluation

- Best match evaluation

Sorting and Display: the results gotten from the fuzzy search must be sorted and presented to the user. This is the most critical requirement of the system due to the fuzzy nature of the search. We employ a form of fuzzy-percentage calculation to adequately categorize the ranking and display the results in a manner such that the most relevant results are shown at the top-down manner as the relevance decreases.

Approximate string matching: Traditionally, approximate string matching algorithms are classified into two categories: on-line and off-line. With on-line algorithms the pattern can be preprocessed before searching but the text cannot. In other words, on-line techniques do searching without an index. Early algorithms for on-line approximate matching were suggested by Wagner & Fisher, (1974) and Sellers, (1980). Both algorithms are based on dynamic programming but solve different problems. Sellers' algorithm searches approximately for a substring in a text while the algorithm of Wagner and Fisher calculates *Levenshtein distance*, being appropriate for dictionary fuzzy search only (Dyke, 2006). On-line searching techniques have been repeatedly improved. Perhaps the most famous improvement is the *bimap algorithm* (also known as the shift-or and shift-and algorithm), which is very efficient for relatively short pattern strings (Baeza-Yates, &Navarro, 1998). The Bitap algorithm is the heart of the UNIX searching *utility agrep*. An excellent review of on-line searching algorithms was done by Navarro, et al., (2001).

Although very fast on-line techniques exist, their performance on large data is grossly inadequate. Text preprocessing or *indexing* makes searching dramatically faster. Today, a variety of indexing algorithms have been presented. Among them are *suffix_trees*, *metric trees* and *n-gram* methods. In computing, *approximate string matching* provide a means for finding approximate matches to a pattern in a string. The closeness of a match is measured in terms of the number of primitive operations necessary to convert the string into an exact match. This number is called the *edit distance* — also called the *Levenshtein distance* — between the string and the pattern. The usual primitive operations are:

i. *Insertio**n*** (e.g., changing *cot* to *coat*),

ii. *Deletion* (e.g. changing *coat* to *cot*), and

iii. *Substitution* (e.g. changing *coat* to *cost*).

Some approximate matchers also treat *transposition*, in which the positions of two letters in the string are swapped as a primitive operation e.g. "transposition of *cost* to *cots*". Different approximate matchers impose different constraints. Some matchers use a single global un-weighted cost, that is, the total number of primitive operations necessary to convert the match to the pattern. For example, if the pattern is *coil*, *foil* differs by one substitution, *coils* by one insertion, *oil* by one deletion, and *foal* by two substitutions. If all operations count as a single unit of cost and the limit is set to one, *foil*, *coils*, and *oil* will count as matches while *foal* will not. Other matchers specify the number of operations of each type separately, while still others set a total cost but allow different weights to be assigned to different operations. Some matchers allow separate assignments of limits and weights to individual groups in the pattern.

Most approximate matchers used for text processing are regular expression matchers. The distance between a candidate and the pattern is therefore computed as the minimum distance between the candidate and a fixed string matching the regular expression. Thus, if the pattern is *co.l*, using the POSIX notation in which a dot matches any single character, both *coal* and *coil* are exact matches, while *soil* differs by one substitution.

5.4.2 Mathematical Model Description

Search operation involves a lot of string manipulations. Their efficiency is thus closely linked to the performance of the algorithm upon which they are implemented. We employed the Transfer function (Tanino, 1984) employed in fuzzy preference ordering in group decision making to evaluate each of the sub-strings with the alternatives toward the matching of a query as shown in eqn. [5.1].

$$p_{ij}^{k} = f\left(x_{i}^{k},\ x_{j}^{k}\right) = \frac{1}{2}\left(1 + \left(x_{i}^{k} \ominus x_{i}^{k}\right)\right) \qquad [5.1]$$

p_{ij}^{k} characterizes the match-preference degree between alternative sub-strings a_i and a_j expressed via $\mu_f(x)$ and Θ is the subtraction operation on two fuzzy sets.

Again, from the principle of Pseudo-Order Preference Model (POPM) for determining the preference of one or more pseudo-criteria, (Wang, et al., 2006) three fundamental preference relations in classical preference structure suffices. The last was manipulated to have the following: Strict match ($\mathscr{M}$), Weak match ($\mathscr{W}$) and Fuzzy match ($\mathscr{F}$) which is capable of generating inferences even if the order/arrangement of the sub-string confuses the retrieval system. To this end, instead of generating no match-found, the fuzzy match operates based on the predefined membership function and the comparison of the sub-strings. We adapt these to arrive at the following equations:

Strict match relation ($a_i \mathscr{M} a_j$)

$$\mathrm{M}_{ij}^{k} - m_{ji}^{k} > m \qquad [5.2]$$

Weak match relation ($a_i \mathscr{W} a_j$):

$$\mathscr{W} < \mathrm{M}_{ij}^{k} - m_{ji}^{k} \leq m \qquad [5.3]$$

Fuzzy match relation ($a_i \mathscr{F} a_j$):

$$|\mathrm{M}_{ij}^{k} - m_{ji}^{k}| \leq \mathscr{W} \qquad [5.4]$$

(where k = 1, …, m; i, j = 1, …, n)

Equation 5.4 represents the major drive away from popular search engines. The fuzzy match relation permits us to accommodate a high level of ambiguity which would have hitherto generated a 'no match found' to users query. It is however important to state that the following assumptions were made.

- Access to stored data is accomplished by presenting a sample data in a query condition.
- Sample data will be in the same national language and notational standards as those used for the stored data. For example, we assume that the user or application will not look for French date notation in an English date field.

With the above assumptions and mathematical model for the work, we are set to discuss the design and implementation with possible test-result for the fuzzy based search engine tagged FuzzyMatch.

5.5 FuzzyMatch Modelling

FuzzyMatch architecture in this section will feature a host of modelling techniques. These techniques include component technology, visual programming, patterns and framework. The rationale is in presenting a robust method for model representation. In particular, solving recurring architectural problems such as physical distribution, concurrency, replication, security, load balancing and fault tolerance requires initial proper design. The development of the World Wide Web (WWW), while making some things simpler, has exacerbated these architectural problems. Our intention is to make the operations and composition of the system understandable with little or no assistance. The first attempt is the presentation of a conceptual framework on the modality of FuzzyMatch. This section describes the developed search tool and also made some comparison with existing technologies.

5.5.1 Conceptual Diagram for FuzzyMatch

A conceptual model is a collection of conjured diagrams and text that are combined to document the functionality of a particular system. The emphasis is on what a system does rather than how. In order to develop good software, a developer must capture and understand the customer's requirements and needs, and make a product or system that they want. The sequential analysis supports the gathering and analysis of user-centric requirements by starting with your users' goals.

The conceptual diagram was employed in the development of the Fuzzy Search Engine (FuzzyMatch). Subsequently, we present the documentations and operational modalities in the following section.

In the stand-alone design of the system, there is basically only one actor involved in the system and this is the user of the system. The user friendliness of the system is of utmost importance and this is reflected in the simplicity of its interface. Users of search engines are often not highly literate and so it is imminent that the interface is easy to understand

and navigate. This has been the basis for many human computer interactions work aimed at reiterating the importance of interface design.

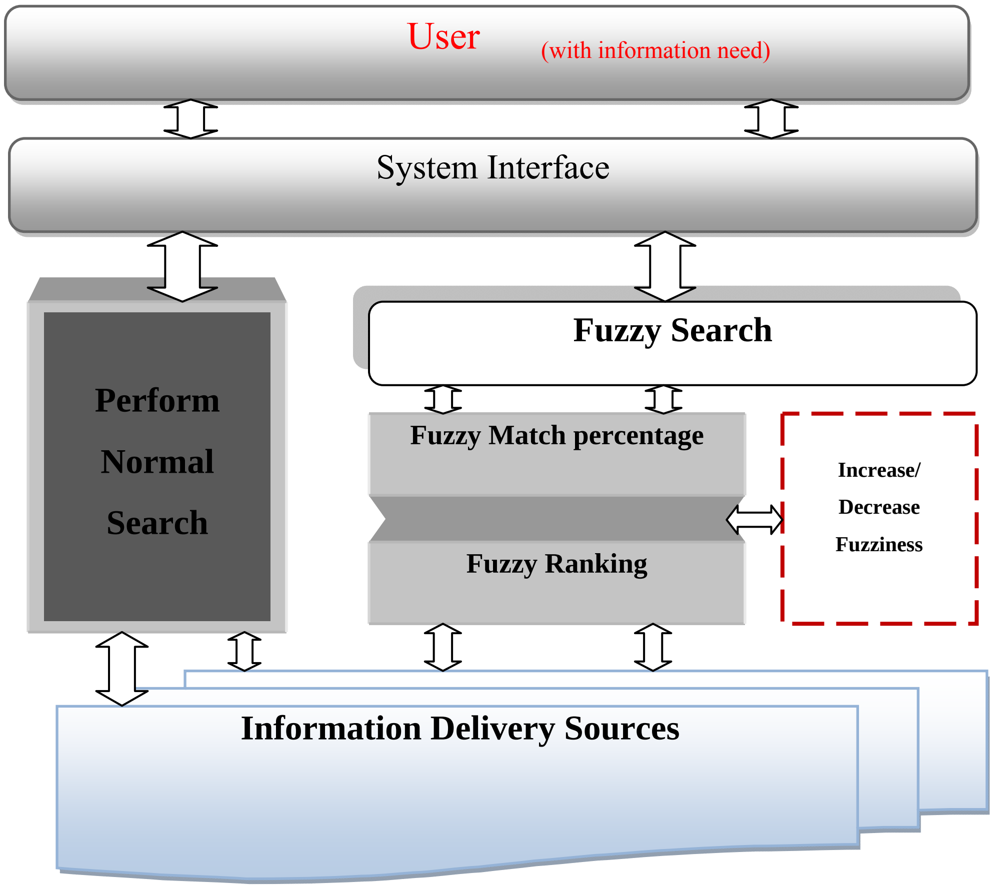


Figure 5.11: Conceptual Diagram for the Proposed System

We can categorize two distinct classes of users in the system, these are the normal and the advance users. This categorization is based on the experience and dexterity of the user with search operation. We believed that if a user is familiar with search operation, the mode of framing queries will be quite different thus improving the possibility of correct-hits. Most websites have search engines that can be used to locate data and general information. However, research has shown that most of this search engine are not useful because an inadequate knowledge of the 'search keys' directly implies that a user cannot locate the desired data. On a commercial phone retail site tagged www.slotsltd.com. Locating a phone on the site through the search engine requires the user to provide the phone name. However, there are customized or code-names for phones. Attempt to locate a phone known as '*Nokia N97'* requires one enter the search key *'NK N97'.* Typing '*Nokia N97'* would result in an empty result. This is a common phenomenon on most sites and is a major problem for users and site owners.

Subsequently, we explain the interaction among the connections and relationships between the use cases and the user.

- *Initiate Search:* in the system a search is initiated by simply typing the search query in the text box provided in the interface and pressing the enter key. A search will also be initiated if the fuzziness of the system is altered while the text box is not empty.
- *Increase or Decrease Fuzziness:* the user can increase or decrease the fuzziness of the system by using the slider provided in the interface. Altering the fuzziness basically affects how strict the system is in searches. It affects just how much the user input is manipulated. An increase in fuzziness would imply that the results may be less like the search query provided by the user. A decrease would imply that the results are more like the search query. Having a zero percent fuzziness level would imply that the results would contain database values that have a perfect match to some or all of the contents of the users query. Testing has shown that at a 50% fuzziness level, the results are highly relevant to the search query with a low occurrence of the missing data issue. Hence the system will be set to a fuzziness level of 50% by default.
- *Perform Natural Search:* A natural search is one that is done with zero percent fuzziness. It is similar to searching using **'direct matching'** since the results have to be exactly the same as the user input. When performing a natural search, the fuzzy algorithms are not implemented in full and the search basically becomes a typical search engine.
- *Perform a Fuzzy Search:* this is initiated when the fuzziness level is set above 0%. When performing a fuzzy search, the fuzzy algorithms are implemented in full.
- *Query Database:* the contents of the database are loaded at the initiation of a search so as to cross reference the user's input with the database contents.

- *Calculate Fuzzy Likeness Percentage:* Since the algorithm does not use direct matching, the user's input is cross referenced with the contents of the database in a fuzzy manner. After the implementation of the necessary algorithms, a ***'fuzzy likeness percentage'*** is obtained. This value implies just how alike the user input is to the database content being considered.
- *Calculate Fuzzy Ranking:* The ***'fuzzy rank'*** is used to rank the final output. It is a vital part of the search process. The rank is calculated with regards to how many of the inputs were *matched perfectly*, how many were matched as subsets and how many were matched fuzzily.
- *Sort Results:* After the search is completed, the results have to be sorted so that the most relevant result is shown at the top and the least relevant at the bottom. This is another crucial part of the search process. The sorting done based on the *fuzzy ranking*, *percentage likeness*, and the result itself, in this order, thus the result with the largest *fuzzy rank* stays at the top regardless of its *percentage likeness* or *actual value*. If two results have the same *fuzzy rank* then they are sorted based on the *percentage likeness* with the largest percentage likeness coming at the top. Finally if two results have the same *fuzzy rank* and *percentage likeness*, then they are sorted alphabetically based on their *actual value*.

5.5.2 Class Diagram for FuzzyMatch

The figure below shows the class diagram which contains a pictorial representation of the classes composed in the system as well as their interactions, dependencies and relationships. Figure 5.12 shows the class representation of our model. Principally, we have the FuzzyMatch main that coordinated other modules. The important sub-modules for technical functionalities are the ones situated on the right hand side – *"fetchvalidstring", "fuzzystringmatch"*, and *"fuzzystringfunction".* The first of these sub modules utilizes a dynamic buffering system for classification and reordering function as shown in subsequent section. FuzzyStringMatch is based on equation 5.4 presented in section 5.5.2. It operates closely with the earlier mention modules and employs our developed fuzzy ranking model.

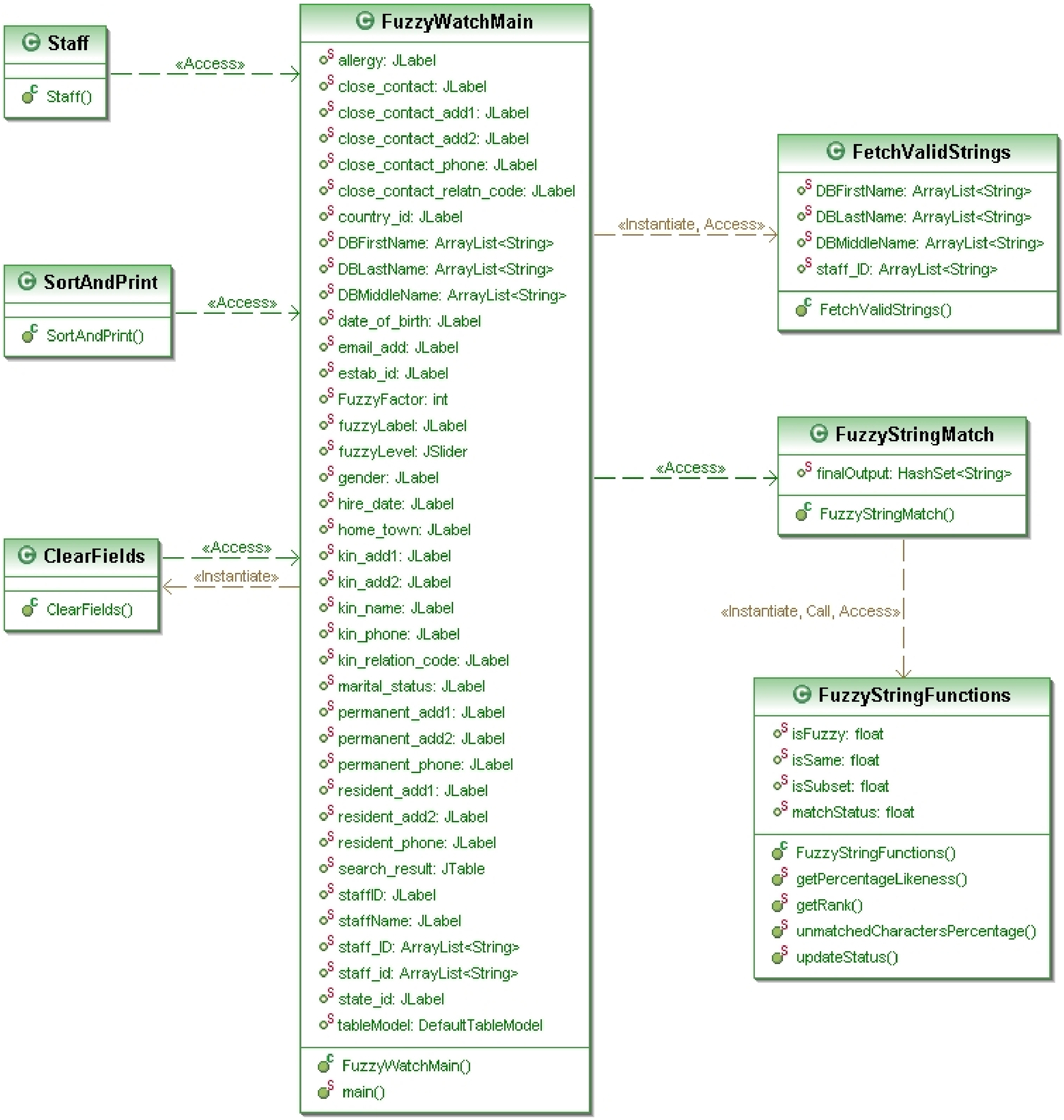


Figure 5.12: Class Diagram for Proposed System

The class diagram assist in modelling the retrieval activities alongside the content of the database. It can be seen from figure 5.12 that the main (FuzzyMatchMain) houses the operational objects for intelligent manipulation between the user query and the matching operation from the database. It is therefore little wonder that we listed the main components necessary for this operation. Other sub-modules shown in the FuzzyMatch diagram were briefly represented and highlighted. They shall be discussed in due time.

5.6 FuzzyMatch Architecture and Implementation

FuzzyMatch diagram presents a generalized model ("Architecture") for the whole search operation. It depicts the overview of the search process and the activities that transcend from the point of query submission by the user, to the point where the result is displayed. The whole operation is highly modularized but again compact enough to present a robust well grounded operation for the user. Figure 5.13 is our designated FuzzyMatch architecture.

- *Splitting User Input Based On Spaces*: the first thing after the submission of users query is splitting. At this stage, the input from the user is split into individual strings based on spaces. For example, ***'Onifade Olufade'*** becomes two distinct strings: ***'Onifade'*** and ***'Olufade'***. String splitting is important in this operation to facilitate the treatment of input independent of the context. The result of the splitting would be stored in a dynamic array containing the sub-inputs 1-n (see Figure 5.13).
- *Matching With Database Values:* Each of the database values retrieved is split into sub-strings and matched with the users input in order to locate the best matches.
- *Getting Best Match*: In order to locate the best match, each string in the user input matched with all the sub-inputs in the database string. To accomplish this matching, the two strings to be matched are fed to the ***'fuzzy membership function'.*** The function returns a value that signifies just how alike the strings are. The user input sub string being considered is matched to the database sub string that gives the highest fuzzy membership value after all the sub inputs have being considered. The pointer is also used to map the user's sub-input to the particular database value.
- *Back Tracking to Replace Match Statistics (Fuzzy Selection):* due to the manner in which the matching is performed, it is possible for two user sub-input to point to the same database substring. However, only one of the sub-inputs is the *best match.* In order to ensure that the current match is the best match, the algorithm

performs a scan of the pointers to check if any user sub-input is pointing to the database substring being considered.

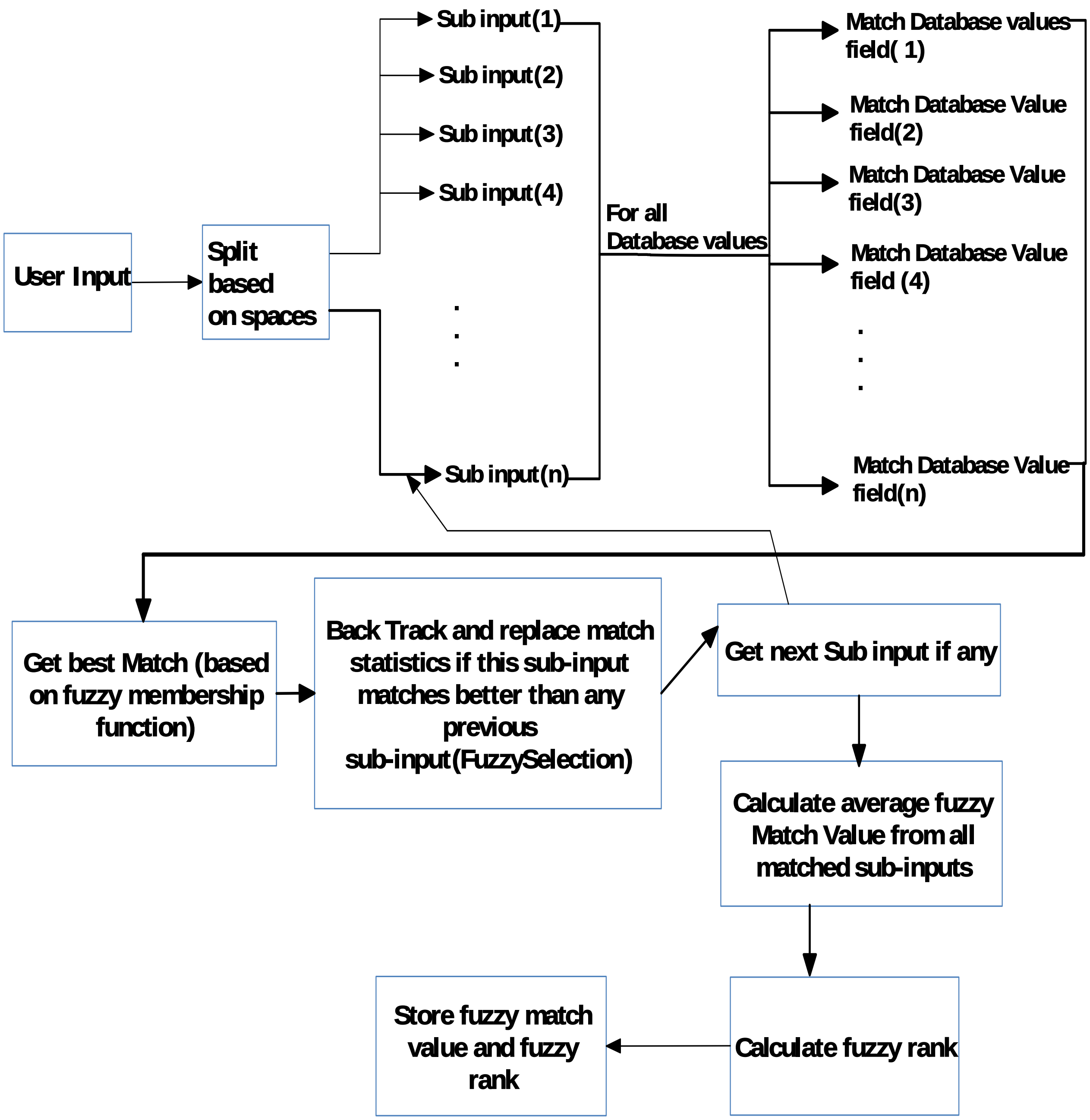


Figure 5.13: FuzzyMatch Diagram

If any exist, the match statistics are retrieved (the fuzzy membership value). This value is compared to the value gotten from the current match. If the previous value is greater, then the current match is *deleted* and *ignored* completely. However, if the current value is greater, the previous match statistics are deleted and replaced with the current one. This process ensures that at all times, the matches that are stored are the best possible matches, and that less relevant matches are ignored. This process is hence tagged ***'fuzzy selection'***.

- *Get Next User Sub-Input:* After the matching is done, the next user sub-input is retrieved and the entire process is re-iterated. This continues until all the sub inputs have been treated.
- *Calculate Average Fuzzy Match Value:* After all user sub inputs have been treated, the average fuzzy match is then calculated from the individual matched values. Note that not all the user sub-inputs will have matches and thus the average is calculated from only the fuzzy match values of inputs that have matches.
- *Store Fuzzy Rank:* Whilst the fuzzy matching is taking place, flags that signify the type of match that took place are set. In this model there are three flags namely:
    - Perfect_Match flag
    - Subset_Match flag
    - Fuzzy_match flag

  Based on the value of this flags, a fuzzy rank will be assigned dynamically to the particular matching process. The fuzzy rank will be assigned based on how many of the sub inputs were matched perfectly, as subsets, or fuzzily.
- *Store Fuzzy Match Value and Fuzzy Rank:* After the entire matching process has been completed, the results obtained are stored in a dynamic list and made to point to the particular database field being considered. These values would be used to rank the outputs at the end of the entire search.
- *Get Next Database Value And Redo Process***:** After the match statistics have being stored, the variables and flags are reset and the next database field is retrieved. The entire process is then reiterated.

5.6.1 The Match String Operation

This model illustrates how two strings are matched with the fuzzy matching algorithm. Figure 5.14 below illustrates the fuzzy string matching process involving a user sub-input and a substring from a database field.

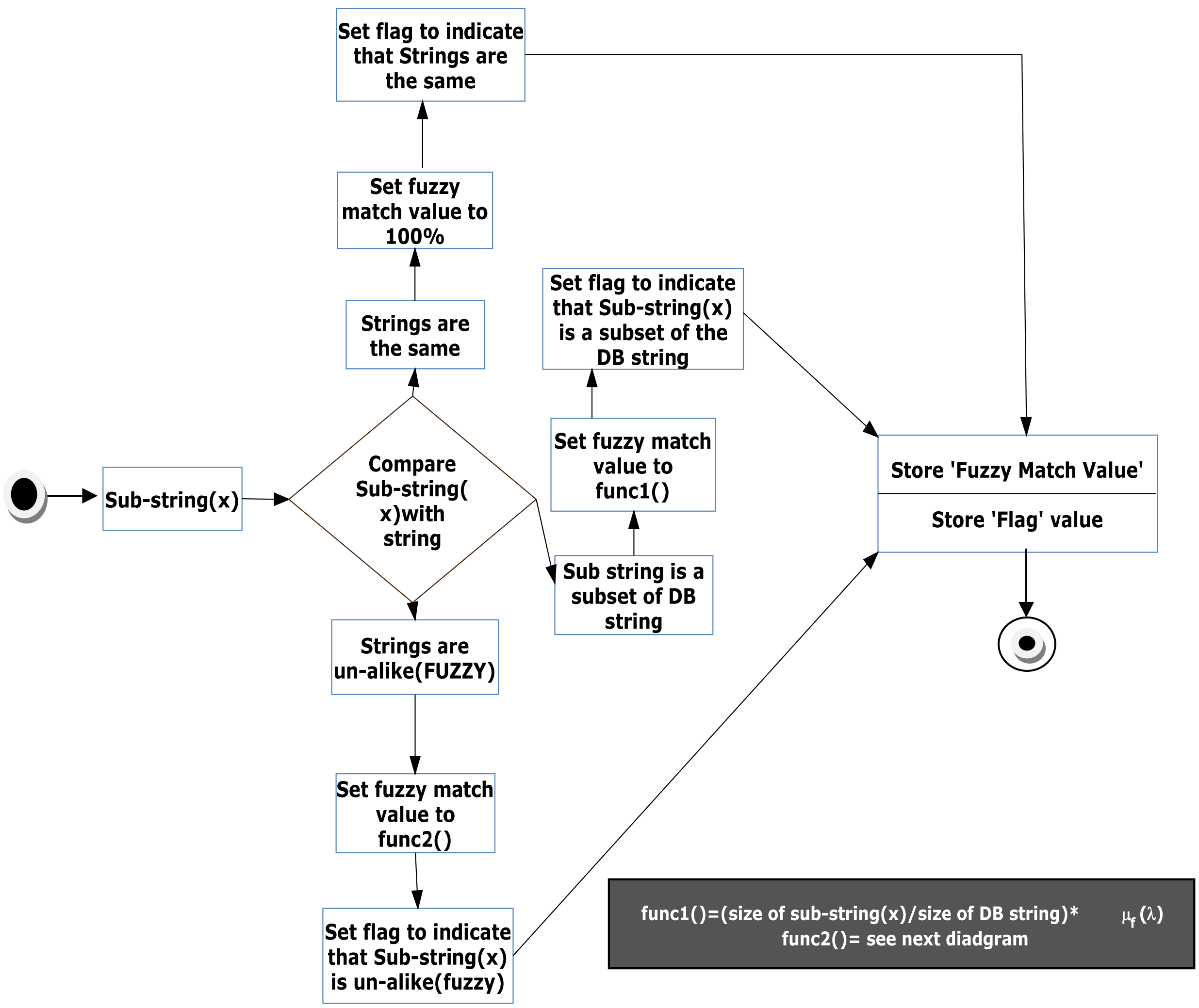


Figure 5.14: String Matching Operations

When the user sub-input is compared to the database input, there are three scenarios that may arise:

- The user sub-input is a "*perfect match*" to the database substring: In this case, the fuzzy match value is set to 100% to indicate that the strings are the same. The flag indicating that the strings are the same (Perfect_Match flag) is set. This is necessary for the calculation of the fuzzy rank.

- The user sub-input is a *perfect substring* of the database substring: This indicates that the user sub-input can be found in the database sub-input, but is not a perfect match. For example *'fade'* can be found in *'olufade'*. Matching sub strings is necessary so that a match is produced even if a name is typed incompletely. When this scenario arises, the fuzzy match value is calculated as the percentile fraction of the user sub-input size and the database substring size (see Figure 5.13). Also, the ***'Subset_Match flag'*** is set to indicate the type of match performed.
- The user sub-input has *"no direct relationship"* with the database input i.e. the user entry is ambiguous and thus fuzzy. This represents a more complex scenario. When this occurs, the fuzzy match value is calculated by matching the two strings based on the concurrent arrangements of their characters and their actual character content. We made bold to say that this is the major contribution to our work in this regard. We adopted the POPM earlier discussed to determine the choice of database hit with the provided query. This is made surer via the fuzzy membership determined by the system and the fuzzy match equation. The subsequent model presents a full description of this process. We also set the *'Fuzzy_Match flag'* to indicate the type of matching being performed.

5.6.2 The Fuzzy String Match Model

The FuzzyMatch string matching operation is extremely important whenever the search algorithm encounters two strings that are unalike. Our rationale for this task is born out of the fact that whenever there exist no direct relationship between two strings, the strings may still have some things in common. FuzzyMatch string matching is our attempt to guide against the risk accruable form some class of dirty-data as shown in figure 5.9. The hierarchical representations include strings that are miss-spelt, inconsistent entries, incomplete context, different ordering and ambiguous data. Consider the strings '*oni**f**ade'*' and '*oni**t**ade'*. The two strings are practically the same, but for the character 't' in the later. The problem arises when a typical matching algorithm encounter this entry, once there no direct relationship can be established, it would be ignored.

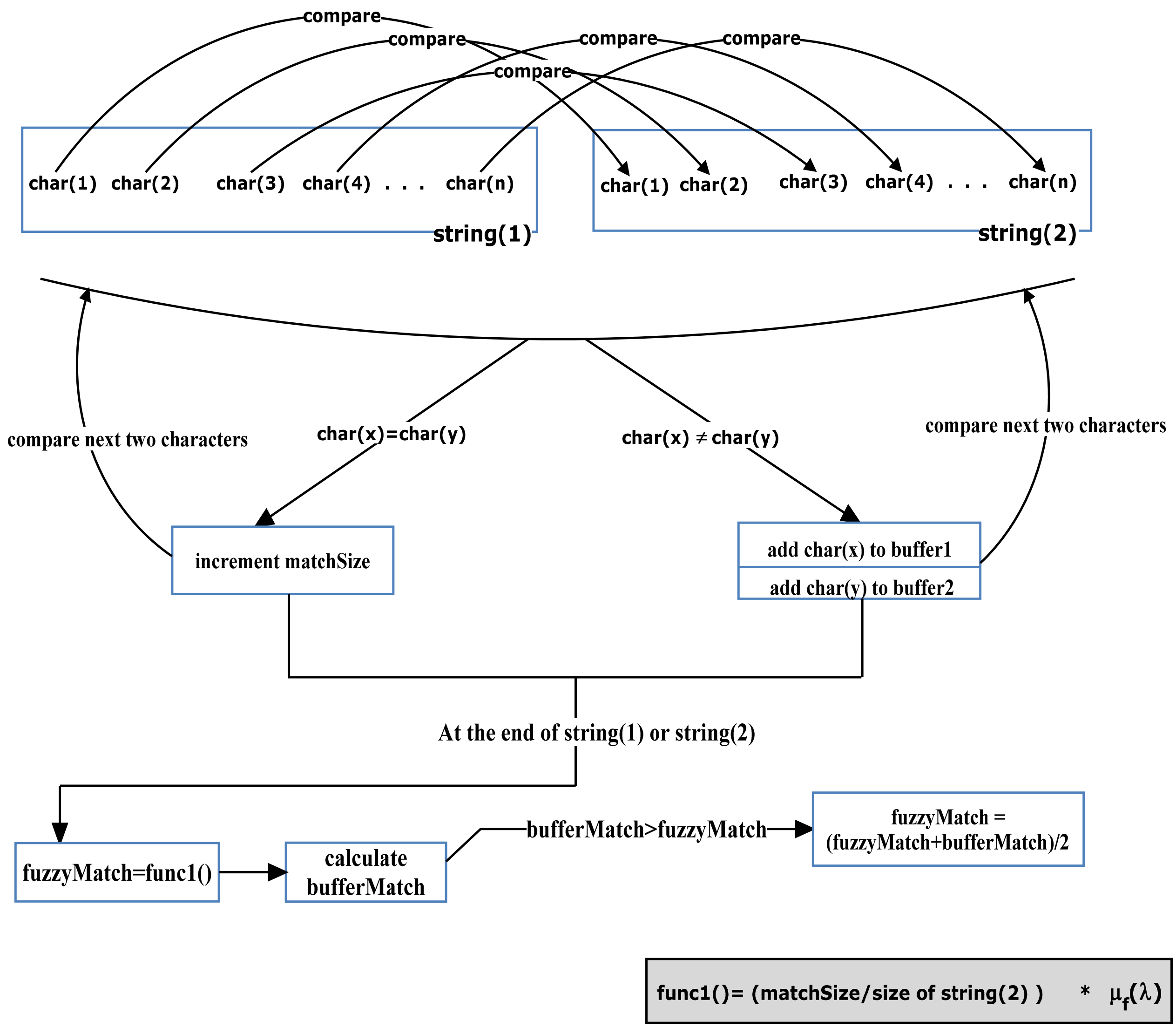


Figure 5.15: FuzzyMatch String Matching Model

However, when viewed fuzzily, the two strings have a lot in common. Firstly, we can establish that the substring ***'oni'*** and ***'ade'*** are in the same position when the two strings are analyzed concurrently. Another point is that they both have the same number of character and thus the main problem is misspelling, substitution or transposition.

The above described scenario formed the basis for the FuzzyMatch string matching algorithm analyses shown in figure 5.15. In other to favourably and concurrently compare the user's string and the database contents, two dynamic buffers were created at the commencement of the operation. One holds the unmatched characters of the user sub input *'buffer1'* and the other holds the unmatched characters of the database substring *'buffer2'*. The algorithm then scans the character content of the two strings concurrently. When the characters are similar, the variable indicating how many characters were matched is incremented. If the characters are dissimilar, the two characters are stored in *buffer1* and *buffer2* respectively. After all the characters might have been compared, it gets to the end of one of the strings (in the case where the size of the two strings are not the same), the fuzzy match value is calculated based on the level of containment or belongingness (via fuzzy membership function) of the *matched character size* and the *size of the database substring* (see Figure 5.15). The above operation does not do away with the unmatched characters, instead they are considered to generate some other entries to be displayed alongside the retrieved entries.

While this could generate a high volume of redundant entries, the user has the opportunity to decrease the level of fuzziness and thus reducing the number of entries. We considered the above as exigent for two reasons, extreme cases of misspelling as in the cases of dyslexia, and when the supplied query forms a subset of the database content but not a whole e.g. *'Oberman'* and *'Hoberman'*.

Dyslexics for example could spell a word with the same character content but in most cases, the characters are muddled up. For example a dyslexic could spell '*clement'* as 'elcmten'. In order to trap cases like these, the algorithm analyses the character content of the two strings even if their characters do not match concurrently. To do this, the unmatched characters placed in buffer1 and buffer2 described above are analysed to check for similarity.

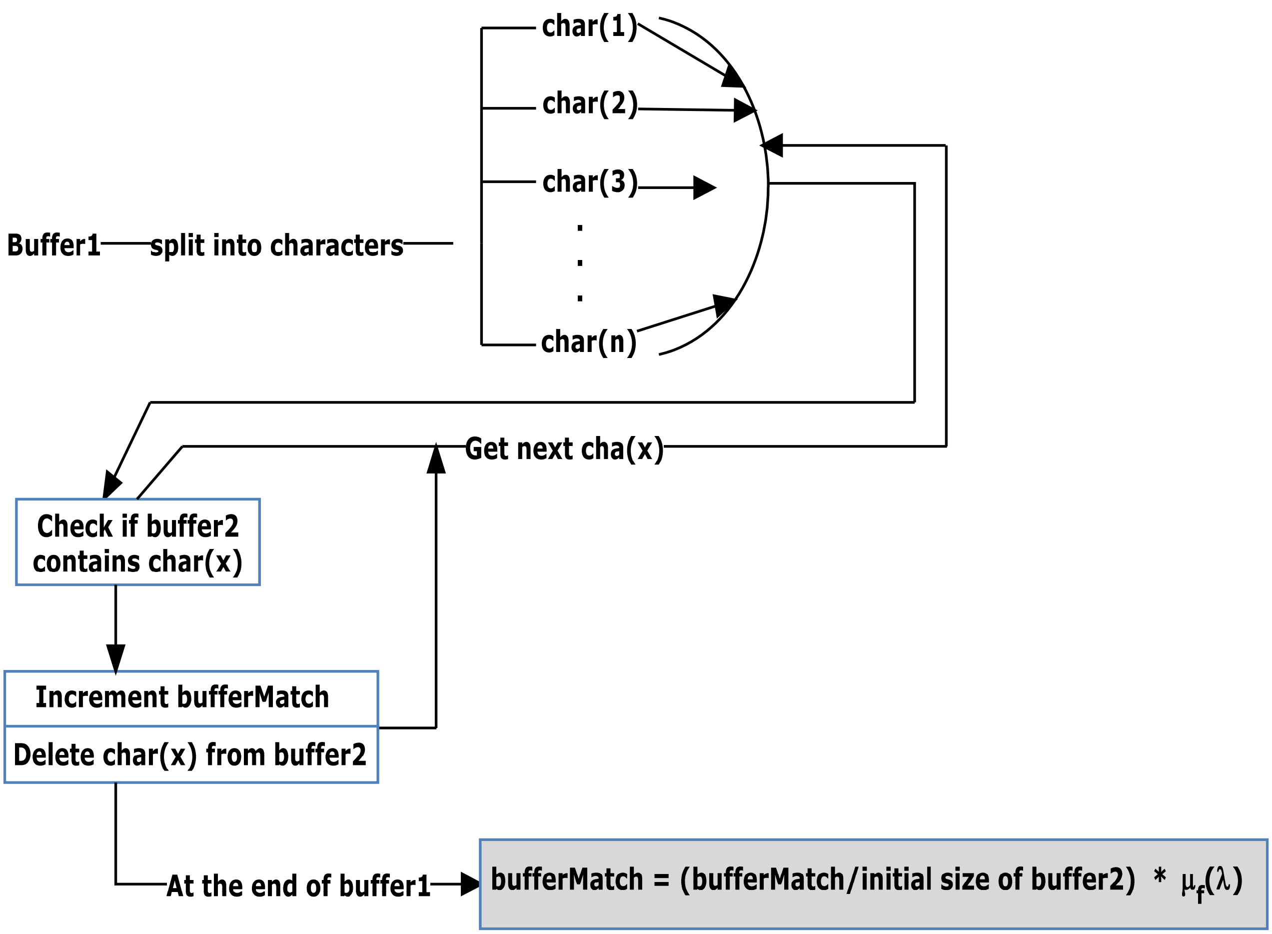


Figure 5.16: FuzzyMatch Buffering Model

The case of dyslexics could be considered as an extreme case, but research has shown that most of failed-hit in retrieval operation are due to misspellings. Google and Yahoo search have propose some level of fuzziness to such problems, but the operation is not as robust as what is present in FuzzyMatch.

In figure 5.16, we present the buffering operation which another important model included in FuzzyMatch. The character splitting operation is not complex, but it is important to feed the buffering model for character analysis and fuzzy decision making. After the analysis, a ***buffer Match*** value is produced. This buffer value is then compared to the *fuzzy match* value. If it is larger, then the new fuzzy value becomes the average of the *fuzzy match* and the *buffer match*. However, if it is less, then the buffer value is discarded and the fuzzy value remains unchanged. When this process is applied to the strings '*clement'* and 'elcmten', a fuzzy match value of comparable to 50% is recorded which we considered fair enough considering that the strings cannot be matched concurrently and could have been otherwise discarded in other search engines.

5.6.3 Fuzzy Ranking and Fuzzy Sorting

Since fuzzy search usually returns some redundant data, we attempt to monitor the level of redundancy with what we call fuzzy ranking. FuzzyMatch search is done fuzzily, it therefore behoves that the ranking and eventual sorting should also be fuzzy in nature. In other to keep track of the ranking dynamically, we established the following search flags *Perfect_Match flag, Subset_Match flag', Fuzzy_Match flag* (please note that subset_match flag does the operation of weak match). The first attempt at search results into raw, disorganized form. In order to rank the fuzzy results, a customized ranking algorithm called *fuzzy ranking* was developed. This algorithm is aimed at ensuring fairness and efficiency of the output. To achieve this, the ranking must present foremost the most reasonable, relevant and valid result while less relevant results are placed further down.

At every instance of complete search operation, the three search flags would have been dynamically adjusted to reflect the corresponding level of how much of the search query was matched perfectly, weakly, or fuzzily. The algorithm uses the value of this flags to determine the ranking points to assign to the result being considered.

Fuzzy Ranking Model Description

The ranking model employs a two-level membership association. Figure 5.17 depicts the sample model employed for FuzzyMatch. It is shown from the left-most part of the model that the function "Get Perfect Match, Fuzzy Match and the Buffer Match values" results from some other operations as shown in the FuzzyMatch architecture presented in figure 5.13.

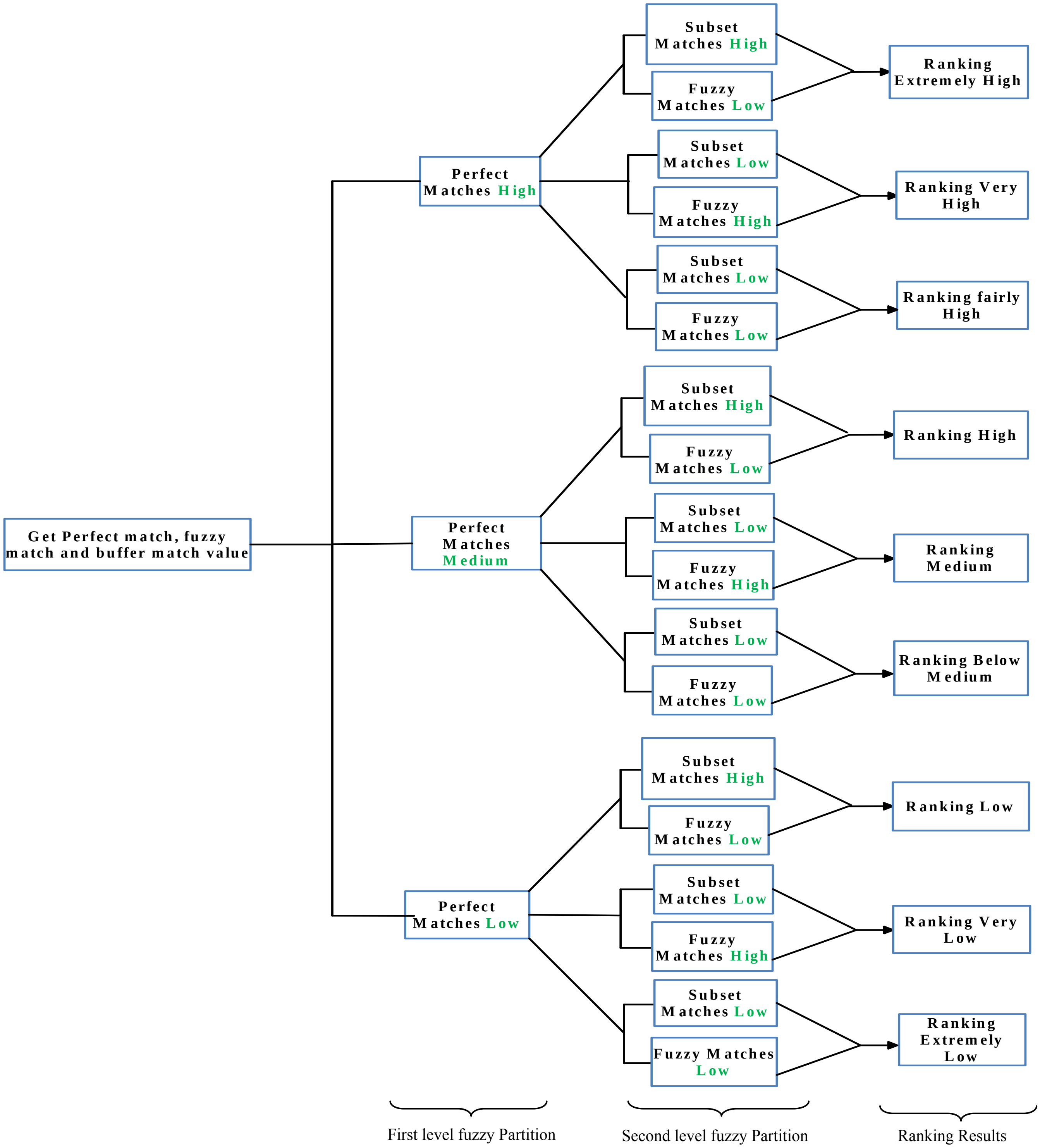


Figure 5.17: Ranking Model for the FuzzyMatch Architecture

The first level fuzzy partition assigns three divisions to the perfect matches i.e. high, medium and low. This is followed by another level of partitioning referred to as the second level. In this stage, the partitioning is more complex than the previous one. Comparison is made between the subset match (the weak match) and the fuzzy match. This follows from the fact that once we can establish the level of match in the first level, there will be no need for the second level, however if we cannot adequately establish this, then the onus is to move from identifiable subsets (weak substring match) to the fuzzy matches. The two levels partitioning resolves what could have resulted into a difficult, unmanageable fuzzy association. The algorithm assigns higher values for the ranking as the amount of *'perfect matches'* increases. The ranking points are reduced gradually as the amount of '*subset'* and '*fuzzy matches'* increases.

The ranking result thus buildup from the fuzzy association between "if the match is weak or fuzzy". With this, the ranking can forced into another realm of belongingness tagged extremely high, very high, fairly high, medium, below medium, low, very low, and extremely low. These associations facilitate the operation of the fuzzy slider which shall be discussed during implementation of the model. Another issue put into cognizance is the mode of result display to reflect the relevance as obtained in other search engines. In the next section, we discuss our provision for sorting the results based on relevance – level of fuzziness.

Fuzzy Sorting Model Description

As obtained in most search engines, part of the displayed result is the number of pages and the time taken for the search operation. It has been observed that relevance decreases as the number of pages increases. FuzzyMatch does not have number of pages and the time taken for the operation, but we provide a fuzzy slider to reduce or increase redundant entries. Sorting helps to display the output of users' request, it is thus important to manipulate the manner of our display based on fuzzy ranking, fuzzy match and common alphabetic ordering. Sorting is expedient because no user wants to peruse the whole output length to locate the desired result. Also in most cases, the result would contain a large number of outputs which would make a manual search difficult.

Bearing this in mind, the algorithm uses the following search characteristics:

- The Fuzzy Rank
- The Fuzzy Match Value
- Alphabetical Ordering

These values are stored in parallel lists pointing to the list containing the search result. Hence sorting would involve sorting all the lists simultaneously so that the pointers still reference the right values. Initially, the lists are sorted using the value of the first list which contains the fuzzy rank of all the results. After sorting using this, the resulting lists are sorted using the fuzzy match percentage and then they are finally sorted alphabetically (this step is however unnecessary but recommended). Figure 5.18 below shows an overview of the entire process.

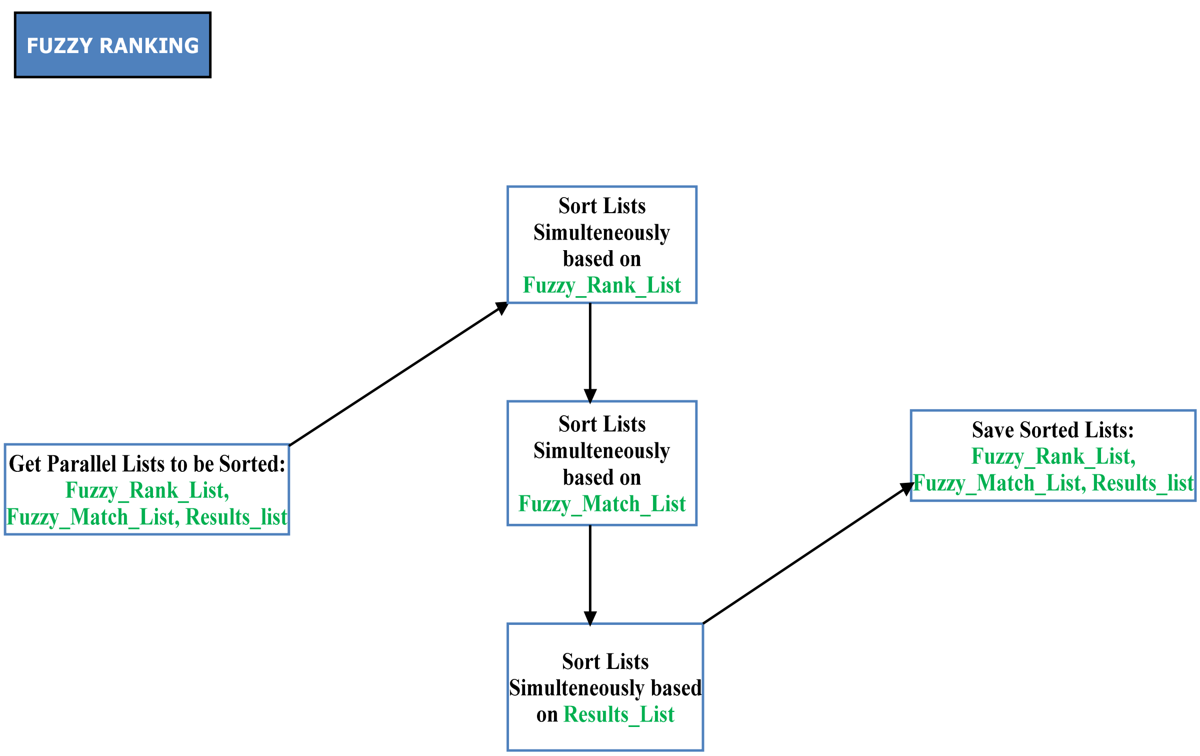


Figure 5.18: Sorting Model for the FuzzyMatch Architecture

The sorting model for the FuzzyMatch architecture is the result of multi-sorting operations. We earlier on pointed out the search characteristics and thus sorting results from “fuzzy rank lists”, “fuzzy match lists”, and the “result lists”. These three are simultaneously performed and the global result is saved for display. While we agreed that the operation is clumsy, it affords us the opportunity to have a global view simultaneously.

Most design starts from model or architectures presented in an abstract manner prior to their implementation. In the section above, we have presented a global FuzzyMatch architecture, and also some of the major models that constitute the whole design. While the design has its drawback, it has contributed immensely in the treatment of some classes of dirty data. In the next section, we present the implementation of FuzzyMatch with some examples to test its efficacies.

5.7 FuzzyMatch Implementation

The design and implementation with examples are the major focus of this section. Presented earlier can be considered to be abstract thoughts, thus we shall present real life examples on the usage of FuzzyMatch as a search engine. We employed a database of a prominent international institute in Nigeria. This database has close to five thousand (5000) entries which are cleaned to represent extracted, transformed, and loaded data example. We also for convenience and example purpose include other entries that were not in the original database to depict the robust nature of the system. The system was rigorously tested with various databases both as stand alone or a network environment. The results are similar. However, we have not been able to test on very large database to determine the length of time and other factors. We are however convinced that, even if the search time increases because of the complexities, the accuracy of match would not be compromised.

We have tested the system on systems with 2.0 GHz, Intel Pentium IV minimum system requirement. The implementation tools are JAVA^{TM}, MATLAB version 7.3.0. and WAMP5. JAVA^{TM} language is of choice because of its portability and robust inbuilt functions for string manipulation. MATLAB provides excellent environment for implementing fuzzy logic and its inference system. The JMATLINK provide the link to integrate the functionality of fuzzy system with Java. The WAMP5 houses the database employed for the search tests. In what follows, we shall describe the screen shots and their mode of operation.

5.7.1 FuzzyMatch Home Screen

FuzzyMatch home screen shows the main features available to the user at startup. These include the Query Input Box (QIB), the Logo, FuzzyMatch slider (FS), the result viewer space, and the detailed viewer space as shown in figure 5.19. We explain each of these in the next subsections.

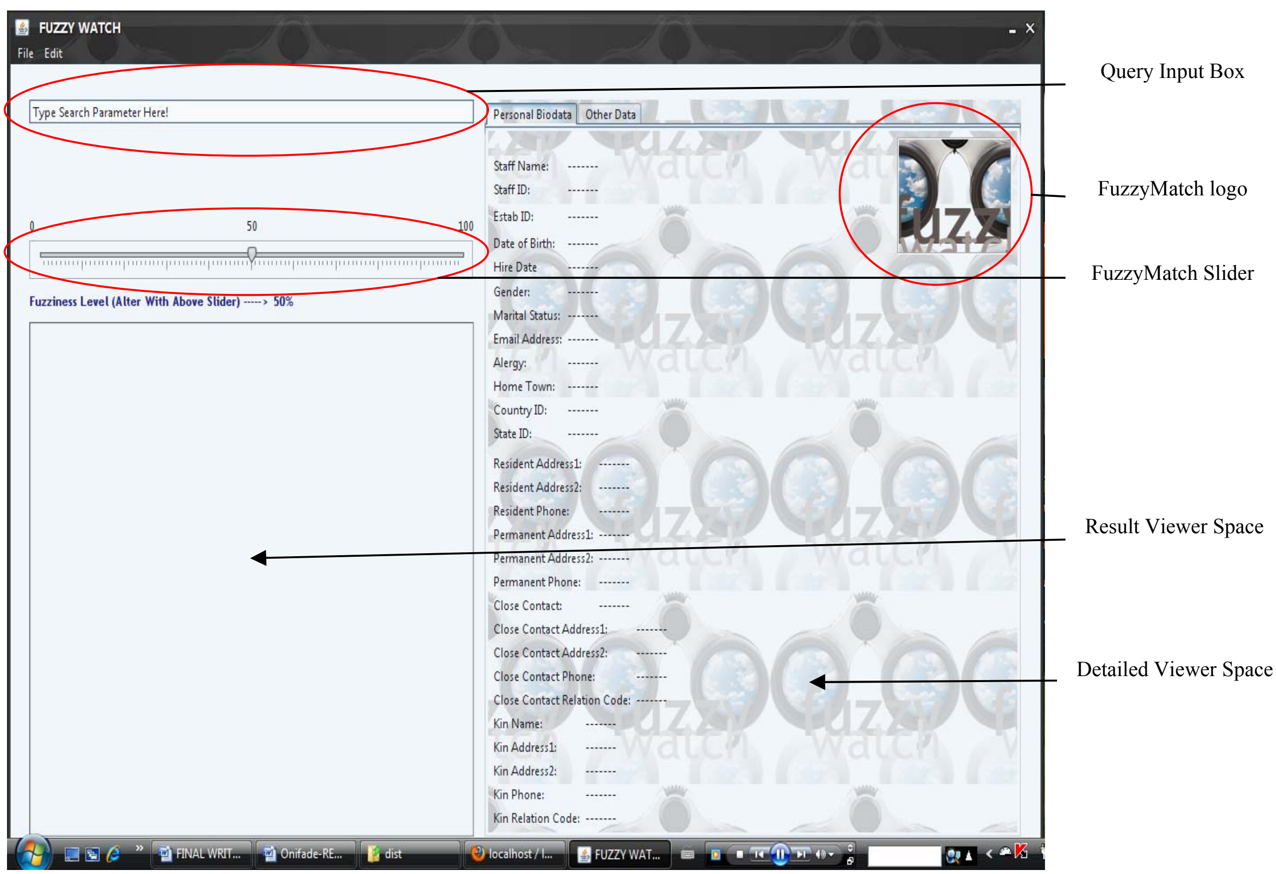


Figure 5.19: FuzzyMatch Home Screen

Query Input Box

This is similar to other search engine's query acceptance mode. It serves as the input space for the user. Our test bed database is of personnel management, thus the type of query accepted or tested are the employee's information predominantly. Consequently, the user can enter the desired query which is in this case is the name of the person. The QIB has been configured to search simultaneously with the user key press using java multithreading technology. To this end, as the user type results are shown at real time. The user could also press the ENTER key to finalize entry.

Interface Logo

The logo here present can serve two functions, in the captured screen we have the FuzzyMatch logo displayed. It can however serve a more detailed role by displaying the passport confirming record particulars of the entry if present. When the passport is not available, it displays a default logo.

FuzzyMatch slider

The FuzzyMatch slider (FS) enables the user to increase or decrease the fuzziness of the search engine and in essence reduce or increase the system's strictness with regards to searching. By fuzziness we imply the level of belongingness of the query submitted to the world of discourse. Moving the slider forward (to the right) increases the system's fuzziness and in essence increases the engine's consideration of fuzzy matches. It can be noticed that sliding the bar forward increases the results shown. FS is handy if the user cannot find his or her desired record in the current result, user thus have the next option of increasing the fuzziness (i.e. move the slider forward) to allow for more matches which could be redundant anyway. The good thing about the FS is that, queries can have several pictures (results) based on the level of the slider i.e. you can fine-tune.

In another dimension, moving the slider backward reduces the systems consideration of fuzzy matches i.e. it reduces the fuzziness. It can be seen that moving the slide bar backwards reduces the results. This could result from the user's desire to reduce the results if they are considered too voluminous to be perused. Reducing the fuzziness gradually excludes fuzzy matches until the result is comprised of only perfect matches (this happens when the fuzziness level is zero, the system reverts to a normal search engine operation and functionality at this stage). The user can use a zero fuzziness level if he/she is absolutely sure of his or her search query.

Result Viewer

This presents tabular information relating to the results of the query submitted by the user. The default nature is blank, but it assumes population as soon as search operation commences.

Detailed viewer's space

Result from search queries many times are made up similar entries which might confuse the user. The detailed viewer's space provides ample opportunity for the user to quickly have a firsthand glance at his/her choice before finally closing the search area. The user can view record details by simply clicking the record in the results viewer's space.

With this system discussion, we are set to discuss the operation of the model and present performance evaluation of its operation with other information retrieval models and tools in the next sections. The system's operation discussed will provide a rich background for a comparative analysis between our retrieval tool and others. The essence of this is to determine how well we can manage the risk factors in information retrieval activities of the economic intelligent actors.

5.8 Comparative Analysis of FuzzyMatch with other Search Tools

It is difficult to determine the efficacy of a newly developed tool without necessarily comparing with existing tools. While in earlier sections we have shown the flexibilities in the operations of FuzzyMatch tool, it becomes expedient to also compare the functionalities and its composition with other tools. Interestingly, Canfora and Cerulo,

(2004) made a taxonomy of information tools and methods, while Broder, (2002) discussed the taxonomy of web search tools. These two amongst others formed the basis for this comparison.

Information Retrieval Objects	**Vertical Taxonomy**													
	Representation						**Reasoning**							
	Query			Document			With Logic			With Uncertainty		With learning		
	Keyword-based	Pattern-based	Structural	Streams of characters	Vector space	Structural	Logic	Algebra	Graph theory	Probability theories	Fuzzy set theories	Neural network	Symbolic learning	Genetic algorithm
CORA	×					×			×					
TACHIR						×			×					
SIFT	×				×			×						
NewsWeeder	×				×			×						
Grep		×		×				×						
LEXA		×		×				×						
OCP		×		×				×						
INQUERY	×				×					×				
SMART	×				×			×						
ILA	×		×			×	×							
Weblearner	×				×					×				
Isearch	×		×		×			×						
Google	×				×				×	×	×			
FuzzyMatch	×			×	×				×	×	×			
ResearchIndex	×				×	×		×	×					
Glimpse, Agrep	×			×	×			×						
Scatter/Gather	×				×			×	×					
Amalthaea	×				×				×					×
WEBSOM	×				×			×				×		

Table 5.3: Vertical projections of popular Information Retrieval Objects

(Adapted from Canfora & Cerulo, 2004)

An information retrieval model can be represented via the representation and reasoning strategy employed. Thus if we designate R_p as the representation model for the documents and queries, and R_s as the framework for modelling the relationship between document and query representation which constitutes the reasoning strategy, then we can write any of such representation as a couple < R_p, Rs>.

We made concerted effort to discuss each of these components in chapter two. This was followed by a vertical taxonomy of information retrieval models based on two broad based classifications – Representation (documents and query) and Reasoning (framework and Ranking). These two "Rs" as we popularly refer to them constitute the fundamentals of all the tools and methods we reviewed as well others from the literatures.

In table 5.3, we present an adapted vertical projection of information retrieval model based on Canfora & Cerulo, (2004). We evaluate the vertical taxonomy of each search tools alongside our newly developed tool – FuzzyMatch.

Comparative analysis of information retrieval models and tools will feature the dissection of the "black box" which performs the retrieval operations. In table 5.3, there are nineteen (19) objects composed from different sources. Although similar composition was done by Canfora & Cerulo, (2004), Broder, (2002), amongst others, we also took concerted efforts to understudy the models and tools whose "black-box" is accessible from research lab's technical reports, companies and institutions. The table thus present what is referred to as the combination of both the vertical and horizontal taxonomy which were described as not disjoint?

The vertical taxonomy classifies information retrieval models based on two components view, which is representation and reasoning. In a similar manner, the horizontal classification represents information retrieval objects with regards to the application areas. As earlier mentioned, each of these retrieval models and tools can be classified and described via the pair of < R_p, Rs>. We can write

R_p (query) = {keyword-based}

R_p (document) = {vector space}

R_s (with uncertainty) = {keyword-based}

This is a representation of “Weblearner”, one of the search tools considered in this research, and any of the tools and models considered or not can be represented by this simple classification. We reiterate again that our focus on this research is basically on text retrieval as there can be many forms of retrieval as earlier mentioned e.g. video, audio, multimedia, graphics amongst other.

We took concerted efforts to juxtapose the functionalities of common search tools and models. The importance of this is the need to showcase the efficacy of our system in text retrieval as compared to others. We are quick to mention here that there are other functionalities embedded in most of the search tools that our own does not have. Concepts like browsing, clustering, mining, gathering, and crawling which characterizes most of the web search tools. However, the fundamental operations of their black-boxes were examined and partitioned into either the representation or reasoning strategy.

Google being one of the most popular search tools had our candid focus. Search tools like *Yahoo Directory* is one of the earliest internet web search engine whose operation is based on manual indexing and categorizing process. Every web source is first evaluated by experts and thereafter inserted into the list of known sites. Unfortunately, it falls under the tools which we could not adequately access the functionalities of the black-box.

Isearch is an open source project developed in the 90s. It forms the basis for many of the available commercial search engines. The composition is as shown in figure 5.28. Its reasoning is based on probabilistic principles. *Isearch* employ four factors in its black-box configuration, these are: *keyword-based, structural, vector space, and algebra* which represent the reasoning strategy employed by the search tool. All the other three factors were for query and documents representation.

ResearchIndex is a scientific literature search tool. It is also open source. Users can browse the database with citation links. It uses full text of the article alongside the citations to facilitate full Boolean, phrase and proximity search. Again from our analysis in figure 5.28, the black-box composes of *keyword-based, structural, vector space, algebra, and graph theory*. It is interesting to note that that inclusive nature of the tools in accommodating the full text and citation is facilitated by the inclusion of two distinct logical reasoning components.

Glimpse and Agrep are not the same tool, but they have many things in common. *Agrep* is a powerful version of Grep, which is also part of *Glimpse*. Its strength lies in handling misspelled words and Boolean queries, as well as limited forms of regular expression. It is an indexing and query tool that facilitates search via local and/or web file systems. Its compositions are *keyword-based, streams of characters, vector space, and algebra*. The only reasoning component here found is in the logical sense. However, the strength of *Glimpse* and *Agrep* is in the manipulation of streams of characters which allows it to adequately handle misspelled words.

Scatter/Gather as the name implies is a search tool that permits the user to scatter documents into different clusters or delimiting groups. The scattering and regrouping is a continuous exercise until a desired group is found. It is a tool which employs text clustering predominantly to group documents based on content similarities. The components of its black-box are *keyword-based, vector space, graph theory and algebra.* It can be seen again here that, it employs Boolean logic as a form reasoning to resolve search operation. *LEXA, Grep, and OCP* are not the same thing, we only group them together because they are made up of the same components. *GREP* is the acronym for Global Regular Expression Print. Particularly suited for ad hoc retrieval, it employs the ed command to print all lines matching a certain pattern – g/re/p where "re" is a regular expression.

LEXA on the other hand is corpus processing software, employed for mining task, it permits automatic lemmatization of any input ASCII texts, frequency list creation and tokens that appears in any loaded text. It also generates lexical density tables. *OCP* stands for the Oxford Concordance Program. It is a gathering tool used for generating concordance, word lists, and indexes from text in any language or alphabets. *OCP* functions based on ASCII file of the text with facilities for up to eight (8) characters to represent a letter. These three tools have the same components which are *streams of characters, pattern based, and algebra*.

Amalthaea is ad hoc tool suited for filtering, mining and gathering. It is an experimental tool for information discovery and filtering. It has its foundations on the knowledge of autonomous agent and artificial intelligence with cooperating and competing multi agent systems. The components consist of *keyword-based, vector space, graph theory and genetic algorithm*. It forms an example of search tools using genetic algorithm in our research.

Google is a household name. It is a general purpose service which is suitable for information gathering via crawling principle. *Google* is a web search engine that is known for its usage of page ranking. Links popularity determines the accorded importance in the search service. A page is highly ranked if it is referenced by lot of other pages. In recent time, *Google* enhances its search operation via the inclusion of fuzzy AND expansion. The main components of *Google* black-box are *keyword-based, vector space, graph theory, probability theory and fuzzy set theories.*

The newly developed search tool for this research is tagged *FuzzyMatch*. As earlier mentioned, it is suitable for ad hoc retrieval, known item search, conscious information search, and confused information search. *FuzzyMatch* is however still a research pet project and therefore has not include task like clustering, mining, gathering and crawling at this stage of the research. The focus presently, is in making available information representing a close match to user`s information need despite the level of uncertainty, incompleteness and ambiguity inherent in the query. Another rational for this development is the need to guide again the problem associated with missing and non-missing data during information retrieval process.

In retrieval activities, *FuzzyMatch* will compete favourably with most of existing search services, tools and models. *FuzzyMatch* is composed of the following: *keyword-based,* *streams of characters, vector space, graph theory, probability theory and fuzzy set theories.* It was observed from the other reviewed search tools and services that but for the exception of Glimpse, Agrep and others with character manipulation, the problem of gross misspelling was not properly tackled. Although Google introduce fuzzy AND to facilitate robustness in resolving ambiguous queries, it is still not capable of handling issues relating to Dyslexia, Phonological and Lexicographic errors. As depicted in our previous examples, the limitation of Google whenever the level of misspelling is higher than certain degree would unavoidably result into the risk of non-retrieval even when the information is resident in the database.

5.9 Results Discussions

In chapter two, we examined the functionalities and technicalities of information retrieval (IR) tools and models. Taxonomy of information retrieval models and tools were examined and discussed. The essence is to depict the importance of this in information retrieval process. We bring to mind again the focus of this research i.e. the model for information risk management in economic intelligence. In chapter four, we presented various models developed for this research and the last one is FuzzyMatch which deals directly with information retrieval operations.

Our focus in this work is on text document retrieval i.e. information can be represented by text documents, thus it will not be out of place when we employ information and documents synonymously. An important aspect of any IR system is the ability to provide documents capable of satisfying the desire, yearnings of the information needs of its user from available document collection. With this foregoing, it is apparent that indexing plays important role in the IR process. Information needs are represented by keywords or phrases which are indexed (Canfora & Cerulo, 2004). Unfortunately, this representation (user's query) causes inevitable loss of information because inherent ambiguity and limited means of expressing desired information.

Users information needs can take any of the following dimensions – known item information need, conscious information need, and confused information need. This implies that the representation (both documents and query) and the reasoning strategy employed by the search tools are of utmost importance. Below in figure 5.20 we present our model for integrating representation and reasoning for information retrieval process.

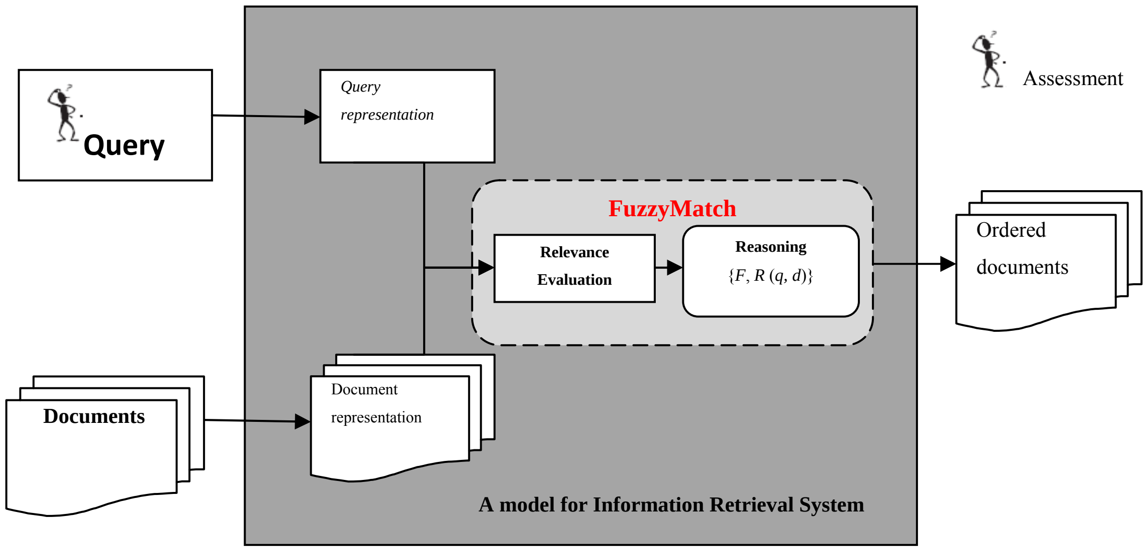


Figure 5.20: Integrated Model for Information Retrieval

Information retrieval model can be characterized by a set of quadruple given as {D, Q, F, R(q, d)} and these factors were defined as follows:

- D is a set of *logical views* for the documents in the collection, it is a *representation component;*
- Q is a set of *logical views* for the user information needs, it is a *representation component;*
- F is a *framework for modeling document representation, queries* and *their relationships,* it is a *reasoning component*;
- R(q, d) is a *ranking function* which *associates a real number with a query q ∈ Q* and *a document d ∈ D*. It is a *reasoning component*.

Vivid study of existing IR models and tools showed that the differences in their composition and operation are determined by these quadruple representations. It therefore implies that any improvement can only result from the manipulation of these components. The first important fact we noticed is the ranking function R(q, d) associating real number from *a query $q \in Q$* to *a document $d \in D$*. The "real number" has force a crisp nature to the ranking values which limits resultant ability to retrieve. In figure 5.20 we employed our earlier discussed FuzzyMatch model to compute the relevance and the ranking of available documents. All the models developed in this research are meant to eliminate or reduce the effect of risk in information sourcing and utilization.

A decision maker who observed a decision problem employs various means to formulate the decision problem. He delimits this into corresponding object, signal and stakes. Working with his counterpart the watcher, he attempts to inform the watcher and a form deliberation ensues. During this process, the first thing we noticed is the need to determine the level of understanding that the watcher can garner from the decision maker. We tagged this "Translation Credibility", thus we attempt to capture/reduce the "*risk of interpretation"*. Once this stage has been successfully completed, the information retrieval process commenced. At this point there are various risks militating against the watcher. These include – "*risk of Information Need Representation*", "*Lexicographic risk*", "*Syntactic risks*", "*Phonological risks*", and difficulties in segmenting words into individual sounds, or blending sound to make words "*Phonemic awareness risk"*.

Composing the appropriate query for information retrieval posses the risk of representation resulting into little or no retrieval, more so if the retrieval exercise falls under the confused information search category. Again, it is therefore important that the relevance and reasoning strategies to be employed must be flexible to take into cognizance these risk factors. A lexical definition is descriptive, reporting actual usage within speakers of a language, and changes with changing usage of the term, rather than prescriptive, which would be to stick with a version regarded as "correct" regardless of drift in accepted meaning. They tend to be inclusive, attempting to capture everything the term is usually referring to, and as such are often too vague for many purposes (Wikipedia, 2010). Lexicographic risk as we put it in this research thus introduces another form of ambiguity to query formulation, word usage and inferred meaning. The risk is not limited to confused information need this time, it also include the conscious information category whereby user search for documents they do not know, but regarding a subject they a familiar with (e.g. Paris is the capital of France, and Paris street in U.K).

Syntactic risk has to do with the organization of words in sentences, the ordering of and relationship between the words and other structural elements in phrases and sentences.
The syntax may be of a whole language, a single phrase or sentence, or of an individual speaker (Encarta 2008). This type of information risk is predominant in cross-language information retrieval exercise, e.g. a French speaker composing in another language where he/she is not very fluent. Another occurrence is the support for commutative representation of information e.g. "Odile Thiery" may not be equal to "Thiery Odile" in all information retrieval exercise. The last mentioned of these risks is the Phonological risk. This includes the way and manner by which words are pronounced base on location, ethnic origin or environmental factors. Pronouncing a name like "Duffing" by a French man will sound to a listening Englishman like "Duffan". If this involves a search operation, it will be difficult to retrieve an entry like "Duffing" which is the intended document.

The focus of the information retrieval model presented in figure 5.20 above is to guide against such risk factors and other related types. With this model, we develop a fuzzy-based system that base relevance determination on fuzzy-values instead of the real value employed by existing retrieval tools and methods. This singular factor as trivial as it might appear has reduced the occurrence of the above mentioned risk factors in no small measure as we shall begin to show in the subsequent sections. Firstly, we consider real life scenarios and examples and lastly, we made comparative analysis of FuzzyMatch functionalities with other existing information retrieval tools and methods.

Typical Search Operations

We have pointed out in model discussion about the FuzzyMatch ability for concurrent character matching and buffering operation. This factor was buttressed with a dyslexic's example of string ***'ONIEADF'.*** The next example depicts how FuzzyMatch handles this problem with another real life example that will be provided.

The pattern of matching in FuzzyMatch has been discussed extensively in earlier sections, we however present the same example employed in showing the deficiencies of most search engines via the mistyped query ***'ONIEADF'.*** This string has various classes it fits in the taxonomy of dirty data. Among these are the '*different ordering*' and '*ambiguous*' under the not-missing, not wrong, and '*misspelling*', '*erroneous*', and '*extraneous*' under not-missing, but wrong classification. The cases of gross misspelling and transposition have been important factor which all search engine has been battling with to enhance user's satisfaction. Little wonder Yahoo and Google has introduced some level of fuzziness into their search engines. Their provisions have not however been able to handle more than three character strings of displaced or misplaced search strings.

It is important to note that this example is amongst the early stated risk factors identified above. While dyslexia is a natural phenomenon, people who are not suffering from these inadequacies also make mistakes with spellings in form of transposition problems. Example include insertion e.g. instead of "cot" user write "coat", deletion e.g. instead of "coat" we have "cot" and the last is substitution whereby we have "coat" instead of "cost". In these examples, it might appear insignificant because of the single words, but in real life retrieval operations, we have found such problem to constitute grave consequences. With this information in mind, a user's submission of ***'ONIEADF'*** results into the following results as shown in figure 5.21 below.

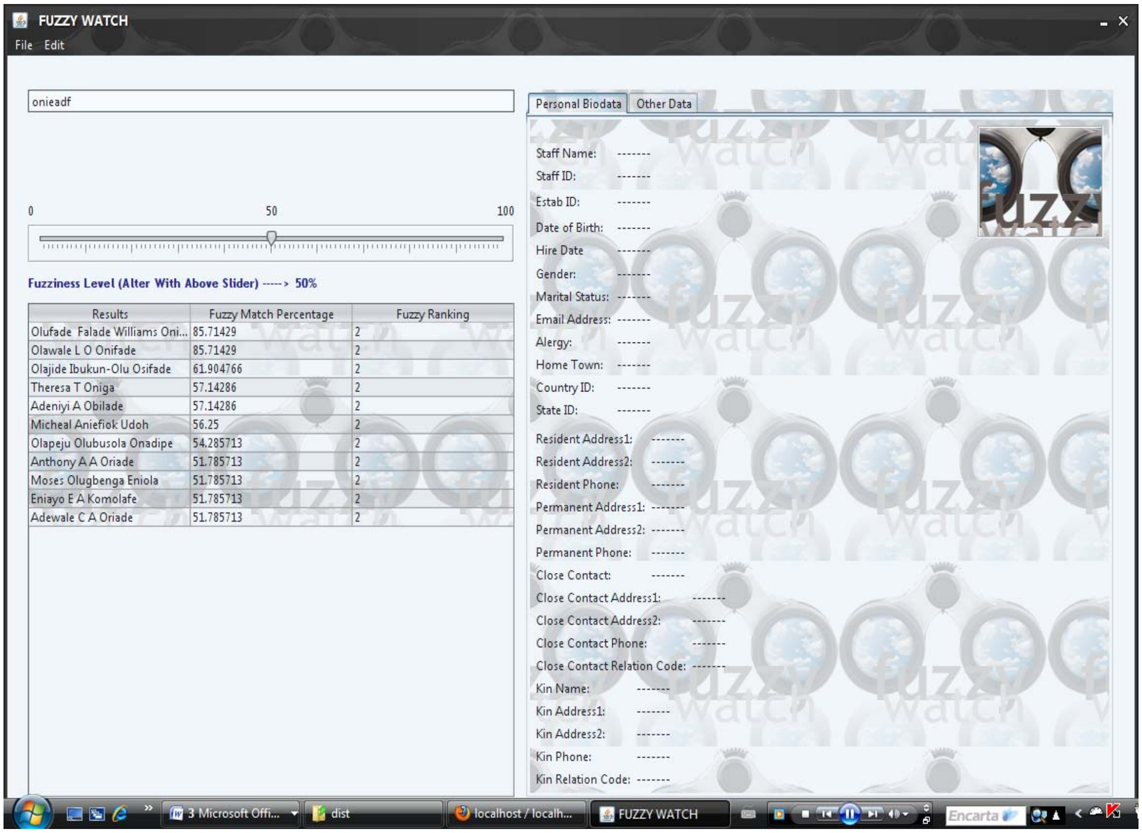


Figure 5.21: FuzzyMatch search for *'onieadf'*

The QIB upon the reception of the input triggers a search operation simultaneously. The results are displayed accordingly, and it will be interesting to discuss the varieties of matches brought forward on the result viewer's space. Although the results are not numbered, the first two have the desired results string in them i.e. 'Onifade'. The third, eighth and eleventh are 'Osifade', 'Oriade', and 'Oriade' respectively. The first of these three was listed higher than the other two because the fuzzy match percentage is greater i.e. 61.9%. This entry can be taken for a single spelling error whereby's' was wrongly typed in place of 'n'. This poses little problem for search engines these days, but its interesting that our model also capture it. The other two entries represent missing characters which could result from user's poor comprehension of the name based on tribe and country of origin e.g. an English man calling "Duffing" and using the pronunciation for search. The fuzzy ranking of all the results are very low. This is because, from the ranking model, the search results from 'perfect match_low', 'weak match_low', and 'fuzzy match_high' combination. This is clearly reflected in the fuzzy ranking value as low as 2.

Example

"A detective in search of a serial killer found out the names of the suspected killer. In an attempt to confirm some other information visited the suspect's school of graduation and demanded for the name search. The detective produces the names as "Liza Oberman". With the index being on the surname, no match was found for Oberman with Liza as first name. He went back frustrated to later found out that, it is accented "Hoberman". The effect of this error is better imagined if there had been another L. Oberman".

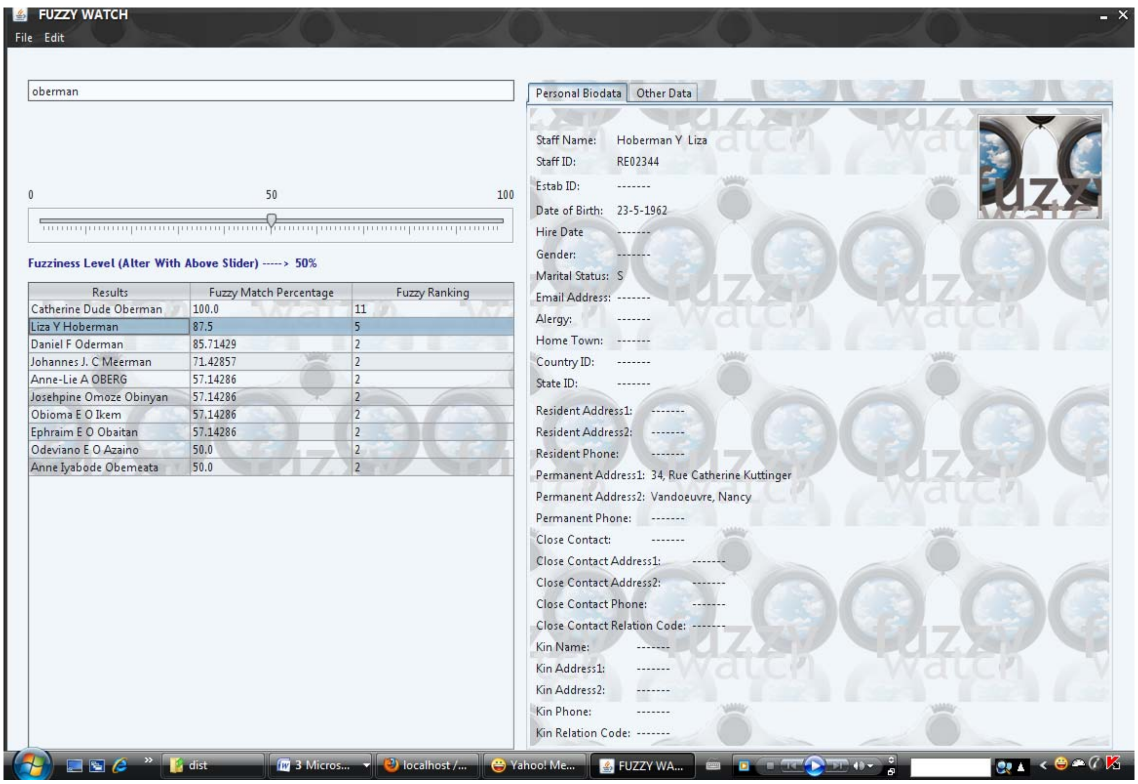


Figure 5.22: Oberman vs. Hoberman

In response to the above scenario, we can see again the efficacy of FuzzyMatch as compare to traditional search engines. The fuzzy belongingness was able to accommodate this inconsistency and reduce possible errors in the discharge of the detective's duties. It visible from figure 5.22 that there was an entry with the search keyword, but it's not the desired result. FuzzyMatch however went a step further to display other relates entries either with excess characters or even misspelled cases as in the entry corresponding to 'Oderman'.

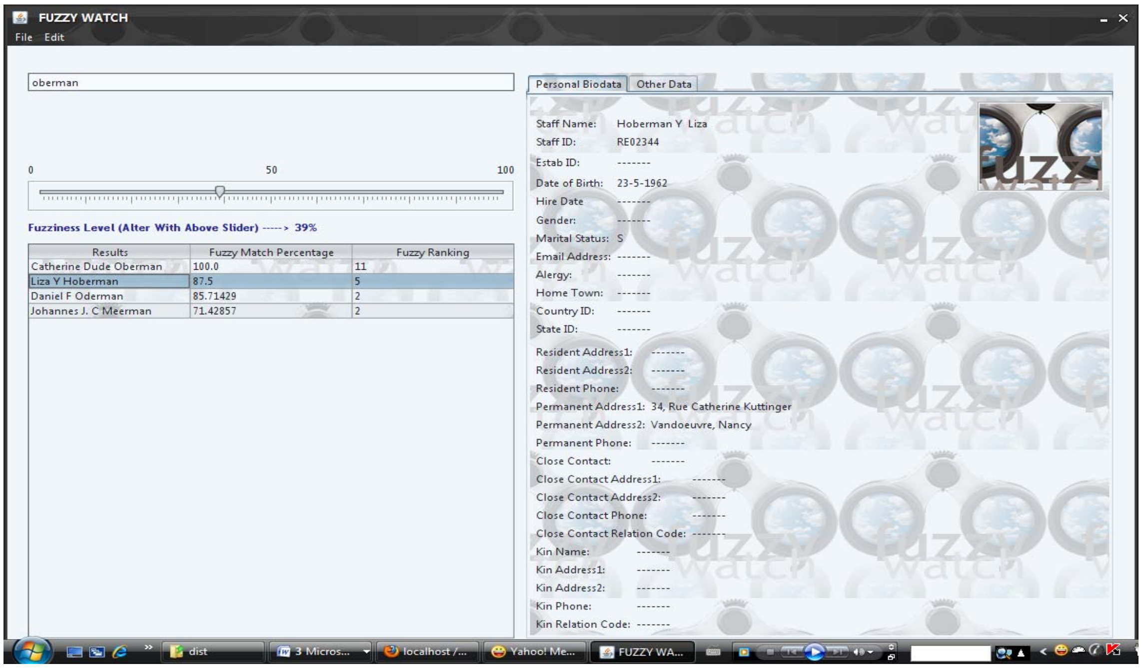


Figure 5.23: The importance of Fuzzy Slider

We mentioned earlier that, there are times when there appeared more result than can be easily perused. The user in such situation can reduce the fuzziness level and the resultant number of entries will be decreased accordingly. Looking at figure 5.23, we can see the effect of the fuzzy slider on the number of displayed result. Once the fuzziness level reaches 39%, the number of displayed result reduces to just only four. Another important thing is that, we can monitor the fuzzy ranking level to see the divergence between the original submitted strings to the database string. Our testing research informs us that most of the results after 30% fuzziness level are almost unreliable,

The last example we want to consider is the problem of accent and/or pronunciation. Imagine how an Englishman would write the name "Duffing" as pronounce by his French counterpart. The best he can either make out of it will be something like "Duffan" which far from the original and would results into no hit in most search operation.

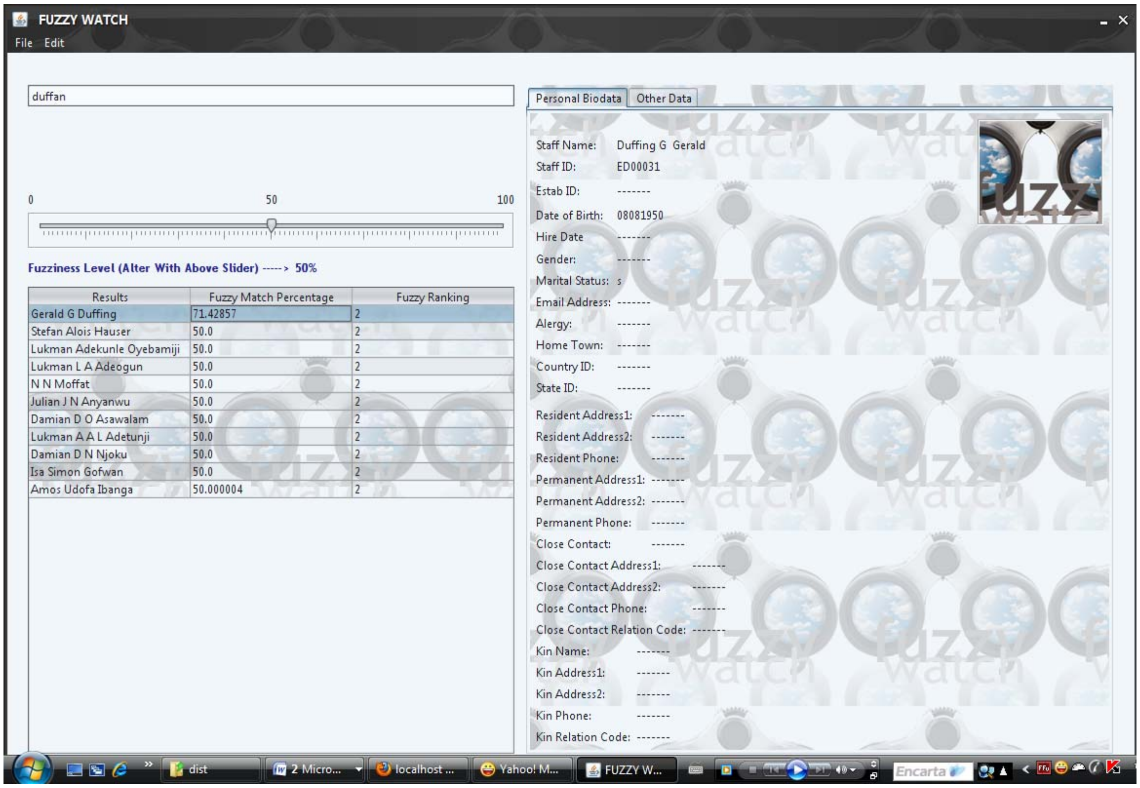


Figure 5.24: FuzzyMatch treatment of Accent/pronunciation in search

Figure 5.24 is the result of query string “Duffan” supplied to the database. Google demands if you mean “Duffin” and later went ahead and brought out results for “Duffan” which are not related to what the user intend. We can see from the figure that the fuzzy match return a 71.4% for the fuzziness and a ranking of 2. This does not disturb the users from having something close to their desire. Furthermore, in some cases, even after altering the fuzziness, the user may still be unsure about the result. The user can confirm by simply clicking on any result. This automatically fetches the details of the clicked record as well as a passport image if present. This portion of the system is implementation independent and has no relationship to the fuzzy search models. Confirmation can be made once the result is selected as it would be displayed on the right hand side known as the detailed result viewer space.

The taxonomy of dirty data depicts two broad categories leading to the dirtiness of the data. These are either missing data or not-missing data. It has been established by the common warehouse group (CMG) that enforcement or non-enforcement of null-entries is the main cause for missing data. Their treatment is therefore based on adequate monitoring measures. The non-missing counterpart is considered more difficult to handle as it has over thirty-three (33) possibilities for its occurrence. These non-missing can either be wrong data or not-wrong, again these are further broken down. These factors amongst other have contributed in no small mean to the concept known as "no data in the presence of data" i.e. user's query returning no-hit because of misspelling, transposition, different ordering, or ambiguous entries as in the case of dyslexia. To further demonstrate the effective of our tool, we make a comparison with existing tools and methods employed for information retrieval purposes

5.10 FuzzyMatch Operation with 'Not Missing but Wrong' & 'Not-Missing, Not-Wrong'

Earlier on, figure 5.9 presents hierarchical listing of the taxonomy of what was termed dirty data. Kim, et al., declared that the result of analyzing a database or warehouse of dirty data can be damaging and at best be unreliable. Broadly, dirty data include missing data, wrong data, and non-standard representation of the same data. Missing was categorized to result and be treated via the enforcement of 'null-not-allowed'. The missing case is not as dreadful as the resulting non-missing counterpart. If it is missing, then there is no entry which has been traced to null-factor. There are however around thirty-three (33) other possible scenario not associated to null-factor and which constitute "*not missing, but wrong data,…*" or "*not missing, not wrong, but unusable data …*". Several provisions have been made to remedy these occurrences and that is what we referred to earlier on as rat-race. It was also founded that 90% of needed data for organization's operation are resident in the warehouses or database but even with this percentage, no-hit still exist partly because of user's pattern of query or other issues. We demonstrate the operation of FuzzyMatch in handling the above mentioned cases in what follows below.

A typical string entered by a user can be ***'ONIEADF'*** instead of ***'ONIFADE'.*** This example is one of the cases resulting possibly from dyslexics or users unfamiliarity with such traditional names (e.g. Englishman's pronunciation of "Duffing" compared to his

French counterpart). It can be considered to be an extreme case of error from users' input, but the taxonomy of Kim, et al., clearly declared that such ambiguity is not just a possibility but was classified under *not-missing, not-wrong but ambiguous data* resident in the warehouse or database. The matching goes in three stages:

5.10.1 Stage1: Split Characters

These two strings to be compared are split into individual characters and stored concurrently because the user input represents a rearranged pattern of the database entry. Character splitting is a popular technique in information retrieval models and tools. However, their mode of implementation determines the robustness. Once a user input (query) is detected, the system performs automatic string partitioning and this is compared with the database entries. The operation on string is closely linked to the representation of documents and users information need. Since there are different categorizations of retrieval activities, it is important that the design should not only be flexible, but adaptive to accommodate inconsistency and ambiguity. In FuzzyMatch model, the string partitioning process is as shown in figure 5.25 below.

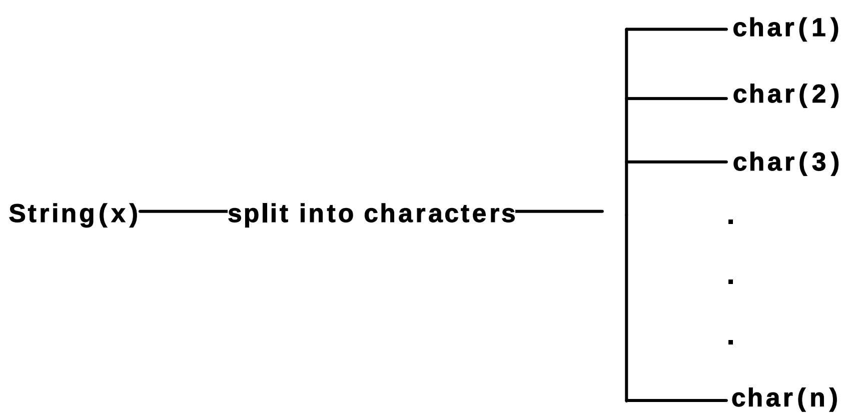


Figure 5.25: Strings' Splitting pattern in FuzzyMatch

5.10.2 Stage 2: Concurrent Strings Comparison and Buffer Operation

At this stage, the characters in the two strings are compared to check for similarity. The product of this stage is the *fuzzy match* value which can only be handled by the FuzzyMatch model. Once the user input has been tokenized, the comparison follows what is shown in figure 5.26. The partitioned string is compared with all entries in the database while fuzzy membership earlier composed assists in determining the level of belongingness of each string to the database content.

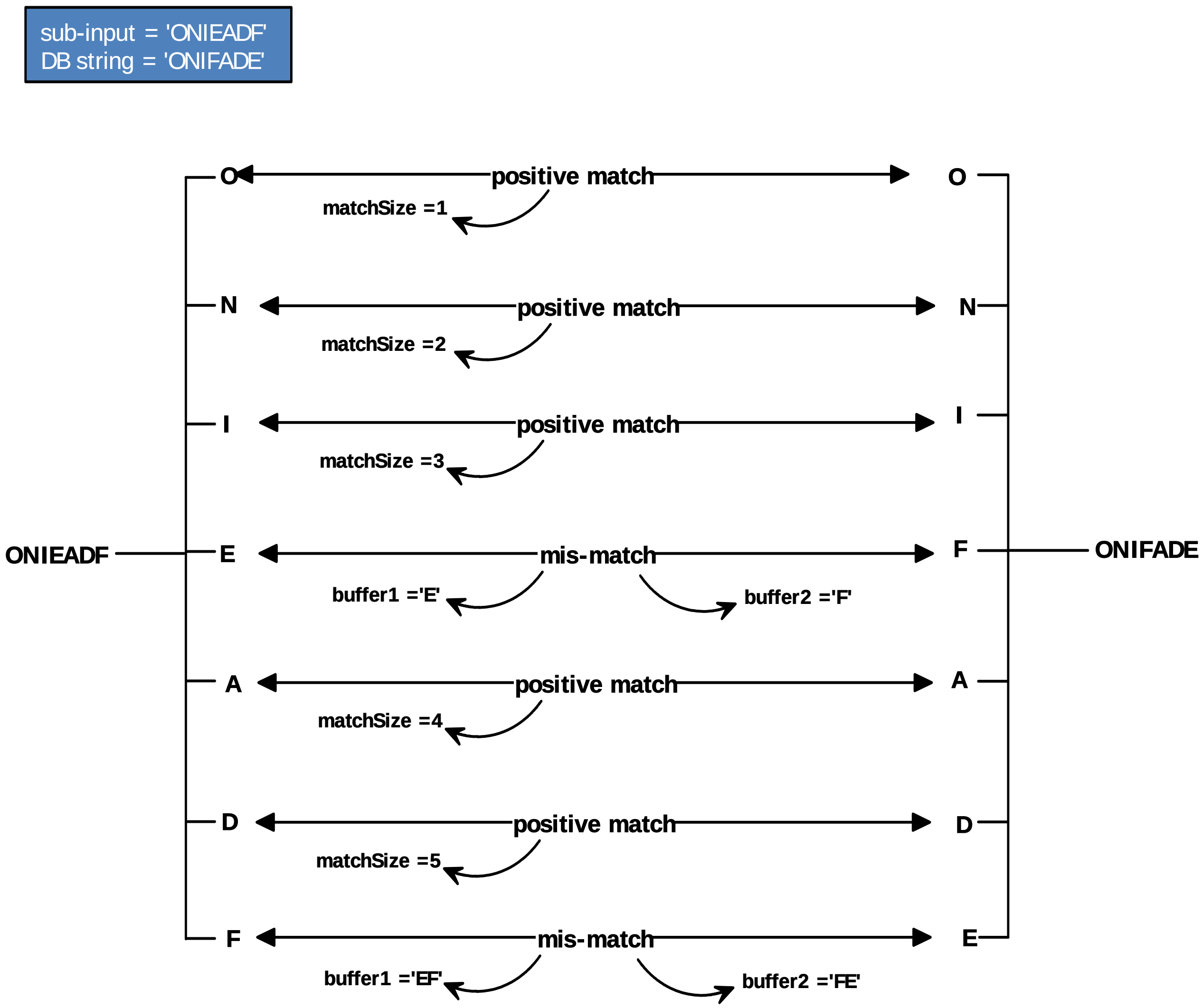


Figure 5.26: Comparing Strings

The operation of the buffering model (figure 5.16) is instrumental to the ability to formulate some of the database entries with the ambiguous user entry. The splitting is followed directly by the dynamic buffering which facilitates concurrent strings comparison and pattern generation from the user's entry to the database content. The next stage depicts this breakdown in a simpler manner.

5.10.3 Stage 3: Unmatched Characters Comparison

The propensity of human to commit error willingly or otherwise is the basis for ambiguity and lack of desired results in our endeavours. However, this factor contributes in no small measure to risk in every facet of our life. We submitted the same input ***'ONIEADF'*** to Google search engine and the result is interesting. Before discussing the Google result, let us look at the string comparison vis-à-vis the buffering pattern employed by FuzzyMatch to resolve such ambiguity.

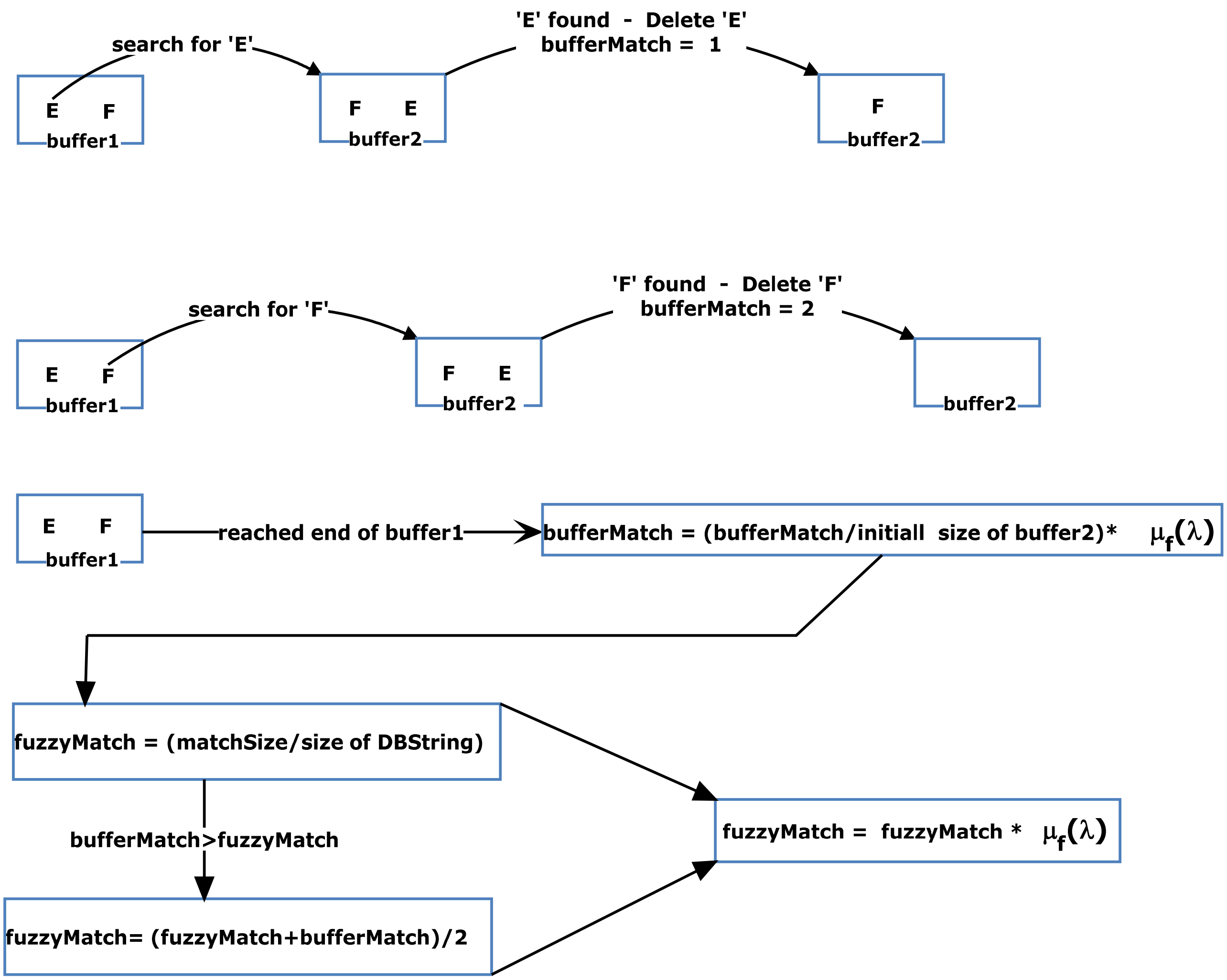


Figure 5.27: Strings Comparison and Buffering Operations

The model compares the two buffers containing the unmatched characters produced from stage 2 to check for similarity in their character content. The product of this stage is the buffer match value. Figure 5.27 attempts to explicitly capture the operations involved in string splitting, matching and the buffering pattern. Once a character presence can be established in the string, the buffer content continues to be manipulated dynamically until the last entry is considered in the string. This results into the fuzzy match which is the multiplicative effect of the buffer match and the level of belongingness. The fuzzy function employed helps to determine the level of fuzziness in the pattern of arrangement of the user's input and used same to assist in possible rearrangement.

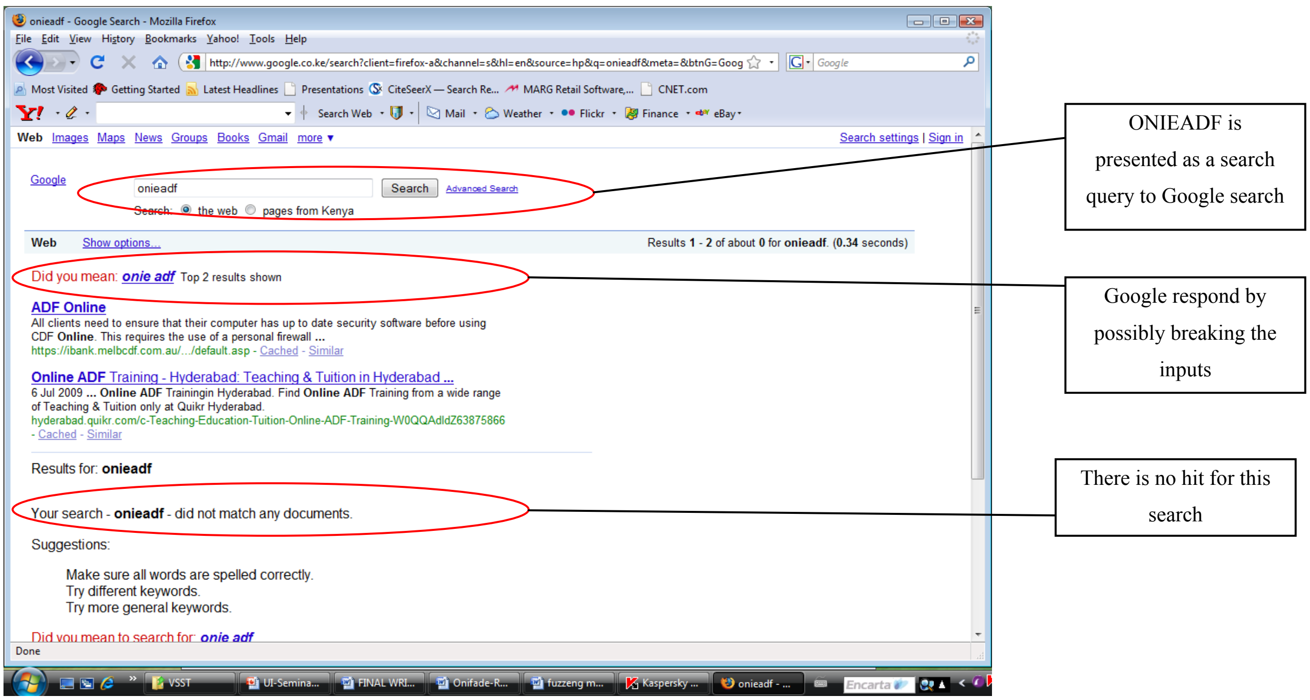


Figure 5.28: Google search result for *'ONIEADF'*

String partition in for resolving search queries is not uncommon phenomenon, in fact it is interesting to note from our review of search engines operations, more than 80 percent employ string partition. It is therefore not peculiar to the design of FuzzyMatch. Resulting back to our dyslexics query string example of ***'ONIEADF',*** we present the same input string to other known search engines and figure 5.28 captures the result from Google search engine.

Once the string is typed in by the user, the first thing Google did was to attempt to find a matching pattern in the supplied string. This led to the partitioning of the string to have two other substrings "***onie***" and "***adf***" which is an attempt to see if similar pattern could be established in the database entries. This we say is similar to the string partitioning of our proposed FuzzyMatch engine. The two substrings arrived at by Google depicts more than the partition visible on the surface. It informs us that presently *Google does not handle any ambiguous substring that is more than four letters,* again this is another walk mile away from the functionality of existing search engines. This is evident in the fact that the first three letters of the substring can directly match the desired result however, the supplied fourth character is actually the seventh in the original string which confuses the engine. "***Onie***" therefore has no representation in Google and it has to move on to the next substring for possible resolution. Since character's case (upper or lower) does not affect the result of search any longer, the second substring "***adf***" returns 2 hits which is not in any way close to the desire of the user. As if Google itself knows about the users' dissatisfaction, it declared that "*Your search – 'onieadf' did not match any documents*", and went ahead to give possible suggestions to resolve the problem.

The suggestions from Google are "make sure all words are spelled correctly" – implying that it is recognized that users' input can be wrong and the effect can be adverse to the point of not returning a tangible result. The second is "try different keywords" – this suggestion is only useful if the user knows that the query supplied is wrong. There are occasions when queries are not direct intention of the users, and this is almost always the case because queries are reflection of users' need not the exact. Thus, it might be impossible to supply other keywords on such occasion without deviating from the initial intention. Google also suggests that "try more general keywords" – again this is difficult when you are not very sure where the error comes from as will be shown in the implementation example we give in some later sections.

FuzzyMatch takes into consideration all the suggestions by Google and try to assist users even in the face of confusion by introducing user friendly options like the fuzzy slider, fuzzy ranking to accommodate various forms of ambiguity. FuzzyMatch unlike Google was able to resolve the dyslexics' entry of ***'ONIEADF'*** and assist the user resolve the ambiguous query with possible entries out of which the user can then select. It has been established that even though many people might not accurately recite their phone numbers or vehicle number, they can always recognize it when they see it. Thus given the user opportunity with the fuzzy slider and ranking to adjust entries improves the quality of search results. In the next section, we discuss the fuzzy ranking operation of FuzzyMatch

5.11 Recap

FuzzyMatch employs a much more balanced structure in the composition of the black-box components than any of the considered tools. The strength of Glimpse and Agrep amongst others lie in the character handling method. Similar principle is employed in *FuzzyMatch* to establish a match between user query and database entry not minding the level of misspelling and/or transposition (insertion, deletion, etc.). The tool is further enhanced by the enhanced fuzzy mechanism for both sorting and document ranking. The limitations of fuzzy AND was compensated for by the fuzzy OR whenever the need arise, thus the ranking employed has supposedly two levels as opposed to single level in Google.

FuzzyMatch will however benefit from a more rigorous test with increased volume of data to further determine the stream of character matching principle and the underlining graph theories which facilitates this operation. The focus presently is however more on the possibility of limiting the risk of non-retrieval in the presence of information. Uncertainty, ambiguity, and incompleteness which also plagued user's information needs generation was adequately tackled in this new tool. Risks accruable form query representation, interpretation, lexicographic, syntactic and phonological issues were considered and catered for in the design and development of *FuzzyMatch*. In all, while the development can still be enhanced with further additional functionalities, the intermediate results are encouraging, and will further found a place in modelling user information need alongside information retrieval process.

FuzzyMatch serves as a search engine that can operate on a database or warehouse to facilitate robust querying capable of dealing with the inadequacies earlier listed. It was developed from the string handling capabilities of Java and the ambiguous case handling ability of fuzzy inference system. The prototype was tested with sample database consisting over 5000 entries and various forms of misspelling, transposition, and dyslexic's entries were tested.

The results are promising, more importantly for the dyslexia and other ambiguous entries. The fuzzy ranking and fuzzy sorting operate concurrently to give a robust level of entries fuzziness which can be increased or decreased to accommodate or reduces the number of entries and the strictness of desired result. In all FuzzyMatch will probably be a search engine of the nearest future as the prototype continues to evolve.

“What is it that was, is now and will not be forgotten?”

Chapter 6
Conclusion, Recommendations and Perspectives

6.1 Conclusion

Economic Intelligence is presented in this research as the act of understanding the processes involved in the production of interpretable indicators for decision making, based on the information available to it. The leader (decision maker) is expected to identify the problem in terms of risk, challenge and threat, through the observation of certain parameters within the environment, and also inform of assumptions which demand/requires verification that can be achieved via the availability of information.

Risk management is a popular branch of Knowledge Management where the quality of knowledge models is essentially employed for the insight and analysis in the domain of Economic Intelligence. Therein, division of labour has resulted into specialized classes of users dealing with information acquisition and its presentation, decision making and the acting agents. The concept of Actors, a kind of user has a broad definition in EI process: the set of users/actors include the decision maker, the watchers, coordinators and others yet without nomenclature. These actors interact to achieve the global objectives of the organization stemming from the identification of a decision problem, its interpretation to those involved in the search operations, search result treatment, presentation of search result and finally, taking inference from available data. While the above facilitates a robust process of decision making, their interaction sometimes subsumes risks emanating from non-quality data. In this regard, we proposed a risk factor model using the fishbone architecture to capture the intricacies and used this to develop the ontological based knowledge reconciliation model (KNOWREM)

KNOWREM is determined by a factor tagged Translation Credibility (TC). This factor attempts to measure the level of understanding derived during the process of knowledge reconciliation. It facilitates a robust manner via which the actors involved in decision problems resolution can rub minds and argue forth and back to arrive at logical conclusion devoid of avoidable misinterpretation and misconceptions. It neither eliminates human operation nor does it attempt to remove it, instead, it brings to focus intangible factors which could spell doom and at the same time avoidable. As an example, will it not have been better if Farouk Abdulmutallab has been denied boarding rather than the aborted suicide bombing of US flight 253. The fact remains that, it is not a factor of unavailability of information, but inadequate comprehension stemming from the manner by which available information can be properly garner, and also determine the relationship in extraneous queries which causes mismatch in the wrong spelling of Muthallab's name. This could have been averted if provisions were made in the system to determine the understanding of the possible consequences of allowing him on board.

We stressed that the importance of appropriate and adequate information toward decision making cannot be overemphasized. Considering the sheer volume of available information to information consumer and the rapidly advancing information management technologies facilitating corporate firms and organization to manage large and complex data sources aimed at exploring new opportunities towards strategic decision making. We identified a new concept known as "Information or Cognitive Overload". We proposed a model for what we tagged "Decisionability" that is the ability of a decision maker to adequate utilize available volume of information without necessarily impairing on the discharge of their duties. Using adequate and timely available information for delivering strategic decisions is the focus of economic intelligence, however with ever increasing volume of information made available to the decision maker, the risk of indecision, forgetfulness, amongst others result. We made bold to say that the consequent could be disastrous as the need arises for decision based on the interrelationship amongst the trio of decision situation, decision maker and the decision process in the face of information overload.

Human language is still the most effective tool to structure his intuition, experience and judgmental abilities towards employing these to model his environment. Unfortunately, ambiguity, vagueness, and imprecision can only be modeled with technology that has such provision. Complexity of problem definition derives from the differences in interpretation of each key actor in a given situation usually results into different level of comprehension of the event and to the explanations of the influences between events. Ontology potentially enables automated knowledge sharing and reuses among both human and computer agents; this is facilitated based on their ability to interweave human and machine understanding through formal and real-world semantics]. However, whenever ontology is faced with 'interpretation' there is always possibility of ambiguity which can highly impair the information search operation. In a bid to cub the risk of interpretation we proposed a new concept tagged "fuzzontology" to assist in knowledge reconciliation process towards delivering strategic decision based on available information.

The entire models proposed in this research were meant to handle one phase or the other during the process of decision making in economic intelligence process as depicted in the EI process stages. While information risk stems from the ability to adequate resolve and define decision problem, it has been found that inadequate or inappropriate information will also spell doom for any decision making. In other to guide against the risk of ambiguous, incomplete and uncertain information need representation, we developed a search engine code named FuzzyMatch. FuzzyMatch does not deviate from existing search tools rather; it understudied the problems associated with uncertainty, incompleteness and ambiguous representation and proffer solutions accordingly. The inclusion of two major components – character stream manipulation and fuzzy AND with OR resulted into a more flexible and enhanced retrieval activities not minding the level of incompleteness or ambiguity.

With this development, decision problem resolution and information retrieval process have better tools in what has been proposed in this research. Beginning with the determination of actors understanding via TC, watcher can go ahead to manipulate the signals with fuzzontological tool (same for the decision maker in defining decision problem) before finally deciding to commence on information search. Users now have a tool to deal with various forms of retrieval activities ranging from known items information search, conscious information need search, to confused information need search. In the next subsection, we exposed some of the present constraints, and future directions to the work

6.2 Recommendations

Our proposed models have been tested only as a prototype system. We want to recommend that testing of these models be performed in greater dimensions to facilitate a further development and enhancement of the models. Their adoption and usage will help in revealing the possible loop hole which might have been overlooked or unnoticed in the cause of this research.

6.3 Future Perspectives

It is noted that not all the thirty three object if search listed for review were adequately performed. This is sequel to the fact that some of the search services are for commercial purposes and thus privacy laws thus not permit access to the functionalities of their black-box to facilitate comprehensive comparison. However, with the ones considered, we are strong of the opinion that the performance of FuzzyMatch is quite impressive when compared to other existing search tools.

In the future, the following will be of interest:

- FuzzyMatch will benefit tremendously from cross platforms tests with a large volume of data. This is to determine the functionality and response time in the face of same constraint with tools like Google, Yahoo search and Bing. It will also be nice to include other tasks like crawling, filtering, mining, clustering amongst other to know how it would perform compared with others.

- The addition of another object of the taxonomy in the realm of reasoning with learning will improve the user adaptability of this development. It will be interesting to adequately employ user search pattern to assist in information search. This can also be used to determine the level of reasoning need for user's adaptability.

- Another future direction is the possibility of creating a standardized platform where users can manipulate both ontological definitions with fuzzy inference system. The manipulation presented is not flexible enough for easy development. It will be of great advantage if a tool can be developed to facilitate this without having to navigate between fuzzy and ontology differently. This is synonymous with what JMathlink is presently developing. Again, this can be improved upon in the immediate future.

References

Abate, M. L., Diegert, K. V., & Allen, H. W. (1998) "A Hierarchical Approach to Improving Data Quality", Data Quality, Vol. 4, no 1, pp. 365 – 369

Abraham, A. & Nath, B. (2000): "Hybrid intelligent Systems (HIS): A Review of a Decade of Research". www.citeseer.ist.psu.ed/523093.html.

Abraham, A.: "Neuro-Fuzzy Systems: State of the Art modeling Techniques". LNCS, accessed from: arxiv.org/pdf/cs.AI/10405011.pdf.

Abramova, N. A. & Kovriga, S. V., (2008) "Cognitive Approach to Decision-Making in Ill-Structured Situation Control and the Problem of Risk". Human System Interactions, 2008 Conference, 25 – 27 May, 2008, pp 485 – 490
ACM Computing Surveys, Vol. 30, No. 4, Dec. 1998, pp 528 - 552

Afolabi, B. and Odile Thiery: "A Model that considers users' behaviours in improving the responses of an information base". Laboratoire Lorrain de Recherche en Informatique et ses Applications (LORIA) Campus Sceintifique BP 239, 54506 Vandoeuvre-les-Nancy, France.

Afolabi, B. & Thiery, O: "Using Users' Expectation to Adapt Business Intelligence Systems" Laboratoire Lorrain de Recherche en Informatique et ses Applications (LORIA) Campus Sceintifique BP 239, 54506 Vandoeuvre-les-Nancy, France.

Agor, W. H. (1984) Intuitive management: Integrating left and right brain management skills.New York, NY: Prentice Hall Press.

Aguilar, J.: " A survey About Fuzzy Cognitive Maps Papers", Int'I Journal of Computational Cognition, Vol 3, no. 2, pp 27 – 33. www.yanGSKY.com/yangijcc.htm

Ajila, S. A., & Wu, Di,: "Empirical study of the effects of open source adoption on software development economics" In The Journal of Systems and Software 80 (2007) 1517–1529

Ajila, S. A., Bailetti, A.J., & Dumitrescu, R. T.: "Experience Report on Software product Line Evolution due to Market Reposition" In workshop QUTE_SWAP by ACM/SIGSOFT-FSE12, Nov. 5, 2004. Pp 25 – 33.

Alavi, M & Carlson, P.A. (1992): "A Review of MIS Research & Disciplinary Development" Journal of management Information System, 1992, vol 8, no 4, pp 45-62

Allan, G.M., Allan, N.D. Kadirkamanathan, V. & Fleming, P.J. (2007): "Risk Mining For Strategic Decision Making", K.M Wegrzyn-Wolska and P.S. Szazcepaniak (Eds): Adv In Intel Web, ASC 43, pp 21-28, 2007, Springer-Verlag Berlin Heideberg.

Alter, S. (2002). Information systems-Foundation of e-business. Prentice hall.

Andersen, J. A. (2000) 'Intuition in managers: Are intuitive managers more effective?', Journal of Managerial Psychology 15 (1): 46-67.

Armstrong, J. S. (1982): "The Value of Formal Planning for Strategic Decisions". Strategic Management Journal 3: 197 – 211. http://dx.doi.org/10.1002%2Fsmj.4250030303

Arthur B. Markman , Douglas , Douglas L. Medin , Julie Hwin , Daniet Kahneman , Richard, Thaler. "Decision Making". Downloaded from the following URL

http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.19.2228&rep=rep1&type=url&i=0
Baeza-Yates R, & Navarro G. "Fast Approximate String Matching in a Dictionary" Proc. SPIRE'98. IEEE CS Press. pp. 14–22. http://www.dcc.uchile.cl/~gnavarro/ps/spire98.2.ps.gz.
Ballou D. P., Wang R., Pazer H., & Tayi G. K. (1998), "Modeling Information Manufacturing Systems to Determine Information Product Quality", Management Science (44:4), Apr. 1998, Pp. 462 – 484.
Ballou, D. P. & Pazer, H. L. (1985) "Modeling data and process quality in Multi-input, multi-output information systems", Management Science 31 (2), 1985, Pp. 150 – 162.
Ballou, D.P., and Pazer, H.L., 1985, Modeling data and process quality in multi-input, multioutput information systems. *Management Science*. 31(2): 150-162.
Bar-Hillel, M. (1980). The base rate fallacy in probability judgments. *Acta Psychologica, 44*, 211-233.
Baron, J. (1994) "Non-consequentialist decisions". Behavioural and Brain Sciences, 17, 1 - 42
Baron, J., & Hershey, J. C. "Outcome bias in Decision Evaluation". Journal of Personality and Social Psychology. Apr. Vol. 54(4), 1988, pp 569 – 579
Barrette, R.: "Building a value driven organization – A whole system approach to cultural transformation", 2006. www.valuecentre.com/docs/fivemodels.pdf accessed on 06/10/08.
Barto, A.G. (1990). Connectionist learning for control: An overview. In W.T. Miller, III, R.S. Sutton, & P.J. Werbos (Eds.), *Neural networks for control*. Cambridge, MT: MIT Press
BBC (2003) Skin Cancer Web Sites 'Misleading', BBC News, http://news.bbc.co.uk/1/hi/health/1730436.stm.
Bednar, P. M. & Welch, C. (2008): "Bias, Misinformation and the Paradox of
Belew, R.K., Adaptative information retrieval, Proceedings of the 12th Annual International ACM/SIGIR Conference on Research and Development in information Retrieval, 1989, pp. 11–20, Cambridge, (MA)
Benbasat, I., & Taylor, R. N., (1982) "Behavioural Aspect of Information Processing for the Design of Management Information System," IEEE Transactions on Systems, Man, and Cybernetics, Vol. SMC-12, No. 4, July-Aug, 1982, pp 439 – 450.
Berger, J. (1990) "Statistical Decision Theory Foundation, Concepts and Methods". Springer Verlag
Besdok, E. & Civicioglu, P: "Adaptive Implicit-Camera Calibration in Photogrammetry using ANFIS". In B. Gabrys et al. (Eds) KES 2006, part I, LNAI, 4251, pp 606-613. Springer Verlage Berlin Heidelberb, 2006.
Best, J.B. (1989) "Cognitive Psychology" 4th Edn. U.S.A.: Allyn and Bacon
Bordogna G. & Pasi G., Controlling retrieval trough a user-adaptive representation of documents, International Journal of Approximate Reasoning, 12, 317–339, 1995.
Bordogna G. & Pasi G.: Modelling Vagueness in Information Retrieval, in Lectures in Information Retrieval, M. Agosti, F. Crestani and G. Pasi eds., Springer Verlag., 2001.
Boston.

Bouaka, N. & David, A. (2004): "A Proposal of a Decision Maker Problem for a Decision Maker Problem for a Better Understanding of information Needs." IEEE Explore, (2004), pp 551-552 http://ieexplore.ieee.org/iel5/9145/29024/01307879.pdf

Bouch, A. (2000) 'Quality is in the eye of the beholder: meeting users' requirements for internet quality of service', Proc. CHI 2000 Conference on Human Factors in Computing Systems (CHI 2000), The Hague, Netherlands, Pp. 297 – 304.

Bouyssou, D. (1984), Approches descriptives et constructives d'aide " la decision : fondements et comparaison, These de troisieme cycle, Universite de Paris - Dauphine, 270 pp.

Bouyssou, D. (1989) : "Modeling Inaccurate determination, Uncertainty, Imprecision using multiple criteria". In A.G. Lockett and G. Islei, editors, *Improving Decision Making in Organisations*, LNEMS 335, pages 78-87, 1989. Springer-Verlag, Berlin. Retrieved from http://citeseerx.ist.psu.edu/viewdoc/1243 on 01/10/2009

Bovee, et al., M. (2001): "A Conceptual Framework and Belief-Function approach to Accessing Overall Information Quality". Proc. Of 6th International Conference on Information Quality (ICIQ' 01), pp 311-324.

Brin, S., Page, L., Motwani, R., Winograd, T.: The PageRank Citation Ranking: BringingOrder to the Web, Technical report, Stanford University, 1998

Broder, A.: "A taxonomy of web search", In ACM SIGIR Forum, Vol. 36, No. 2. (2002), pp. 3-10, http://cs.wellesley.edu/~cs315/Papers/Broder_TaxonomyWebSearch.pdf,

Broglio, J., Callan, J.P., Croft, W.B., Nachbar, D.W., Document retrieval and routing using INQUERY system, Proceedings of the 3rd Retrieval Conference TREC,1995, pp. 29–38, Gaithersburg, Maryland.

Brown, S. M. (1997) "Preparing data for the data warehouse", Proceedings of the Conference on Information Quality, Cambridge, MA, 1997, Pp. 291 – 298.

Buche, P., Dervin, C., Haemmerle, O. & Thomopoulos: "Fuzzy Querying of Incomplete, Imprecise & Heterogeneously Structured Data in the Relational Model using Ontologies & Rules". IEEE transaction on Fuzzy Systems, vol 13, no 3, June 2005, pp 373-383.

Buckland M. K. & Plaunt, C.: "On the construction of Selection Systems". In Library Hi Tech, 12(4), 1994.

Buckland, M. K. & Plaunt, C.: On the construction of Selection Systems. In *Library Hi Tech*, 12(4), 1994.

Bueno, D. & David, A. "METIORE: A Personalized Information Retrieval System". Proceedings of the 8th International Conference on user modeling, 2001, Pp. 168 – 177.

Bullen, G. & Sacks L. (2003): "Towards new Modes of Decision marking Complexity of Human factors". Version 1, Issues 1a, University college London, pp 1-5 29th August, 2003.

Burgess et al, M.S.E (2004): "Quality Measures and the Information Consumer". Proc. Of 9th International Conference on Information Quality (ICIQ'04).

Burgess et al., M. (2007): "Using Quality Criteria to Assist in Information Search". International Journal of Information Quality, vol 1, pp 83-99.

Burgess et al., M. (2007): "Using Quality Criteria to Assist in Information Search". International Journal of Information Quality, Vol. 1, Pp. 83 – 99.

Burgess. et al, M. (2002): "Establishing a Taxonomy of Quality for use in information Filtering". Proc. Of British National Conference on databases 2002, pp 103-113.

Caby, B.C., Pautke, R.W., and Redman, T.C., 1995, Strategies for improving data quality. *Data Quality*. 1(1): 4-12.

Callaos, N., & Callaos, B. (2002). Toward a systemic notion of information: Practical consequences. Informing Science: The International Journal of an Emerging Transdiscipline, 5(1), 1-11. Retrieved June 11, 2009, from http://inform.nu/Articles/Vol5/v5n1p001-011.pdf

Canfora, G. & Cerulo, L.:"A Taxonomy of Information Retrieval Models and Tools" Journal of Computing and Information Technology - CIT 12, 2004, 3, 175–194

Carroll, J. M., (1983) "Decision Support Risk Analysis, Computer and Security," vol. 2, issue 3, pp. 230 – 236, Nov, 1983

Celko, J. (1995) "Don't Warehouse Dirty Data". Datamation (Oct. 1995), pp 42 – 52.

Centre of Risk Management of Engineering Systems (CRMES), Ranking of Space shuttle FMEA/CIL items: the risk ranking and filtering (RRF) methods, University of Virginia, Charlottesville, 1991

Chaffey, D. & Wood, S. (2005). *Business information management: Improving performance using information systems.* Harlow, Prentice Hall, Pearson Education Limited.

Checkland, P. B. & Scholes, J. (1990). *Soft Systems Methodology in Action*. New York: John Wiley and Sons.

Chen, G. (1998): *Fuzzy Logic In Data Modeling: Semantics, Constraints, and Database Design*, (Kluwer Academic Publishers, Boston,1998

Chen, H., She, L.: Inductive query by examples IQBE: A machine learning approach, Proceedings of the 27th Annual International Conference on System Sciences, Information Sharing and Knowledge Discovery Track, 1994, Maui, Hawaii

Choo, C. W., Detlor, B., & Turnbull, D. . Information Seeking on the Web – An integrated model of browsing and searching. *Proceedings of the Annual Meeting of the American Society for Information Science (ASIS)*, 1999. Available at http://www.eric.ed.gov/ERICDocs/data/ericdocs2sql/content_storage_01/0000019b/80/16/15/6c.pdf

Chu, H. & Rosenthal, M. (1996). "Search engines for the World Wide Web: A comparative study and evaluation methodology," ASIS 1996 Annual Conference Proceedings, Baltimore, MD, October 19-24, 1996, 127-135. http://www.asis.org/annual-96/ElectronicProceedings/chu.html

Collins, J. C., & Porras, J. I.: 'Build to last: successful habits of visionary companies (New York: Harper business).

Coombs, C. H., (1987) "The Structure of Conflicts" American Psychologist, 42, 355 – 363

Cooper, W.S.: A definition of relevance for Information Retrieval. Information Storage and Retrieval, 7:19{37, 1971

Cooper, W.S.: Some inconsistencies and misnomers in probabilistic Information Retrieval. ACM Transactions on Information Systems, 13(1):100{111, 1995

CORAS A platform for risk analysis of security critical systems. IST-2000-25031, 2000. http://coras.sourceforge.net.

Cox, E.: "The Fuzzy system handbook". Second edition, 1999.

Crestani F. & Pasi G. eds., Soft Computing in Information Retrieval: Techniques and Applications, Physica Verlag, series Studies in Fuzziness, 2000.

Crestani F. & Pasi G., Soft Information Retrieval: Applications of Fuzzy Set Theory and Neural Networks, in: "Neuro-fuzzy Techniques for Intelligent Information Systems", N.Kasabov and Robert Kozma Editors, Physica-Verlag , Springer-Verlag Group , pp. 287–313, 1999.

Crestani, F. & Lalmas, M.: "Logic and Uncertainty in Information Retrieval" In M. Agosti, F. Crestani, and G. Pasi (Eds.): ESSIR 2000, LNCS 1980, pp. 179 - 206, 2000. Springer-Verlag Berlin Heidelberg 2000

Cykana, P., Paul, A., & Stern, M. (1996) "DOD Guidelines on Data Quality Management". Proc of the Int'l Conference in Information Quality (ICIQ'96), MIT, Boston, pp 154 – 171

Davenport, T. H. (1997). *Information ecology*. Oxford: Oxford University Press.

Davenport, T. H., & Prusak, L., (1998) "Working Knowledge: How organizations manage what they know". Boston – Harvard Business School Press

David, A. & Thiery, O.: 'Prise en Compte du Profil de I'Utilisateur dans un Systeme d'Information Strategique'. Laboratoire Lorrain de Recherche en Informatique et ses Applications (LORIA) Campus Sceintifique BP 239, 54506 Vandoeuvre-les-Nancy, France.

Davies, L. & Ledington, P. (1991). Information in action. London: MacMillan Press.

Dawes, R. M. (1998) "Rational Choice in an Uncertain World". Orlando, FL: Harcout Brace Jovanovich

Delone, W. H. & McLean, E. R. (1992) "Information systems success: the quest for the dependent variable", Information systems research 3(1), 1992, Pp. 60-95.

Delone, W.H., and McLean, E.R., 1992, Information systems success: The quest for the dependent variable, *Information Systems Research.* 3(1): 60-95.

Dervin B. (1983). An overview of Sense-Making Research: Concepts, Methods and Results to Date. Presented at the *International Communication Association Annual Meeting*. Dallas, May. Seattle: School of Communications, University of Washington.

Dissanayake, M. & Abourisk, S. M.: "Qualitative Simulation of Construction Performance using Fuzzy Cognitive Maps". In proc of the 2007 winter simulation conference, S.G. Henderson eds.

Dock, V.T., & Wetherbe, J.C. (1988). Computer Information systems for business. West Publishing.

Doernberg. D.L., and Ziegler, D.H., 1980, Due process versus data processing: An analysis of computerized criminal history information systems. *New York University Law.* 1: 55.

Dromey, R.G. (1996) 'Concerning the chimera , IEEE Software, Vol. 1, No. 1 pp. 33-43.

Duffing, G., David, A. & Thiery. O. (2005): Contribution dela Gestion du Risque a la Demarche d'Intelligence Economique. Fouille de Donnees Complexs, EGC' 05, 2005

Dumas, P., Boutin, E. Duvernay, & Gallezot, G.: "Is Communication Separable From Information?" First European communication conference, Amsterdam, 25-26 Novembre 2005 downloaded from http://hal.archives-ouvertes.fr/docs/00/06/27/07/PDF/sic_00001670.pdf. pp 1-18

Dunning, T., & Freedman, D.A. (*2008) Modeling section iffects. in Outhwaite, W. & Turner, S. (eds) Handbook of Social Science Methodology. London: Sage*

Dyke, N. V., (2006): "Levenshtein: Levenshtein Distance Metric in Scheme" http://www.neilvandyke.org/levenshtein-scheme/

Eaton, J. & Bawden, D. “What kind of resources is information: Int’l journal of information management. 11(2), 156 – 165, 1991.

Eden, Colin. “Cognitive Mapping and Problem Structuring for System dynamics model building”. In System Dynamic Review Vol. 10, No. 2-3, Summer-Fall 1994, John Wiley and Sons.

Edwards, W., Kiss, I., Majone, G. & Toda, M. (1984) What constitutes “a good decision”? Acta Psychologica, 56, 5 - 27

Eierman, M. A., Neiderman, F., & Adam, C. (1995) “ DSS Theory: A model of construct and relationships”, Decision Support Systems, vol. 14, 1995, pp. 1 -26

Einstein, A. (1909): “*The collected papers of Albert Einstein”. Vol. 2 : The Swiss years: writings, 1900-1909* (Princeton, NJ, 1989).

Eisenhardt, K. M. & Zbaracki, M. J. (1992) ‘Strategic decision making’, Strategic Management Journal 13 (Special Issue): 17-37.

Eisenhardt, K. M. (1989) ‘Making fast strategic decisions in high-velocity environments’, Academy of Management Journal 32 (3): 543-576.

Eisenhardt, K. M. (1999) ‘Strategy as strategic decision making’. Sloan Management Review, Spring: 65-72.

Elsbach, K. D. & Barr, P. S. (1999) ‘The effects of mood on individual’s use of structured decision protocols’, Organization Science 10 (2): 181-198.

Encarta English Dictionary (2007).

Engin, S., Kuvulmaz, N. J. & Omurlu, V.E.: “Modeling of a Coupled Industrial Tank System with ANFIS”. Journal of British Lib. Direct, ISSN 2970, pp 804-812, Springer-Verlag, Germany, 2004.

English, L. (1991): “Improving data warehouse and Business information quality. Methods for reducing cost and increasing profits. New York Wiley and Sons.

Eppeler, M. J. Helfert, M. & Gasser, Urs (2004): “Information Quality: Organizational, Technological, and Legal perspectives”, Studies in Communication Sciences 4/2 (2004), Pp. 1 – 16.

Epstein, S. (1998) ‘Emotions and psychology from the perspective of cognitive-experiential self-theory’, in W. F. Flack and J. D. Laird (eds) Emotions in psychopathology: Theory and research, Series in affective science, pp. 57-69. New York, NY: Oxford University Press.

Evans, M. H.: “Excellence in Financial Management - Course 12: Competitive Intelligence”. Pp. 1 – 35.

Even, A. & Shankaranarayanan, G. (2006): “Not All Data Are Created Equal: Inequality in Utility and Implication for Data management”. In Proceedings of the International Conference on Information Quality. ICIQ-06, Boston.

Even, A. & Shnakaranarayanan, G. (2005): “Value-Driven Data Quality Assessment In Proceedings of the International Conference on Information Quality”. ICIQ-05, Cambridge, M.A, Nov, 2005.

Even, A., & Shankaranarayanan, G. and Berger, P.D. (2005), “Profit Maximization with Data Management Systems” Proceedings of the Intl. Conference of Information Systems (ICIS), Dec, 2005, Las Vegas, NV.

Feldman, M.S., and March, J.G., 1981, Information in organizations as signal and symbol. *Administrative Science Quarterly*. 26: 171-186.

Felix, R.: "Real World Application of a Fuzzy Decsion Model Based on Relationship between Goals for Finance Services and Sequencing for Assembling of Cars". Journal of soft computing, Springer-Verlag, 2008, 12: 129 135.

Financial Accounting Standards Board (FASB). "Statement of Financial Accounting Concepts No. 2, Qualitative characteristics of Accounting Information", Original Pronouncements, Accounting Standards as of June 1, 199, Volume II (AICPA Pronouncements, FASB Interpretations, FASB Concepts Statements, FASB Technical Bulletins), Financial Accounting Standards Board, CN, USA, 1993.

Firquin, B.L (1992) "Quality is free… but how do you implement total quality and restructive information technology at the same time?. Proc. 20th (SIGUCCS) User Services Conference, New York, NY, USA, pp 93-96.

Fisher C. W., Chengalur-Smith I., & Ballou D. P. The Impact of Experience and Time on the Use of Data Quality Information in Decision Making, *Information Systems Research*/Vol. 14, No. 2, 2003.

Flake, G.W., Lawrence, S., Giles, C.L., Coetzee, F.M.: Self Organization and Identification of Web Communities, Journal of the IEEE Computer Society, 35(3), 2002, pp. 66–71

Floridi, L. (1999). *Philosophy and Computing, An Introduction*. London: Routledge, Taylor and Francis Group.

Fox, C., Levitin, A., and Redman, T., 1994, The notion of data and its quality dimensions. *Information Processing and Management.* 30(1):9-19.

Fox, R. (2004): "Moving from data to Information: OCLC system and series. Int'l digital library representatives, 20 (3), 96 – 101.

framework, Proceedings of the 17th ACM-SIGIR conference on Research and development in information retrieval 182–190.

Fuhr N, & Buckley C (1991) A probabilistic learning approach for document indexing, ACM Transactions on Information Systems 9:223–248

Gackowki, Z.J. (2007b). Informing Quality, 1 (2), 225-249.

Gackowski, J. (1986): Specific Sequence of computer information system components least constraining the designers' freedom in search of solutions ../../../../db/conf/acm/csc 1986.htm#Gackowski>:471

Gackowski, Z. J. (2004). Logical interdependence of data/information quality dimensions A purposedfocused view in IQ. Proceedings of the Ninth International Conference on Information Quality (ICIQ-04), November, 5-7, 2004, Cambridge, MA.

Gackowski, Z. J. (2005a). Information systems in business environment: A purposefouced view. Informing Science Journal, 8, 101-122. Retrieved from http://inform.nu/Articles/Vol8/v8p101-122Gack.pdf

Gackowski, Z. J. (2005b). Operations quality of information: Teleological operations research-based approach, call for discussion. Proceedings of the 10th Anniversary International Conference on Information Quality—ICIQ 05 November, 4-6, 2005, Massachusetts Institute of Technology (MIT), Cambridge, MA.

Gackowski, Z. J. (2007a). A formal definition of operation quality of factors – A focus on data and information. International Journal of Information Quality, 1(2), 225-249.

Gackowski, Z.J. (2006a). Quality of Informing Credibility-Provisional model of functional dependencies. Proceedings of the 2006 Informing Science and Information

Technology, 3, 731-744. Retrieved from http://informaingscinece,org/proceeeding/InSITE2006?IISITZbig250.pdf

Gackowski, Z.J. (2007b). Informing as a discipline: An initial proposal. Proceeding of the Informing Science + IT Education Joint Conference InSITE 2007, University of Ljubljana, Slovenia, June 21 – 24, 2007. Retrieved from http://proceedings.Informingscience.org/InSITE2007/InSITE07187-208GACK346.pdf

Gackowski, Z.J. (2008). Informing for Operations: The First Principia http://proceedings . Informingscience.org/InSITE2008/IISITv5p679-704Gack490.pdf

Galhards,H. Florescu, D., Shasha, D., Simon, E & Saita, C-A, (2001) "Data Cleaning Language, Model, and Algorithms." In Proceedings of the Very Large Database Conference (VLDRB). 2001.

Gardyn, E. (1997) "A Data Quality Handbook for a Data Warehouse". Proc of the Int'l Conference in Information Quality (ICIQ'97), MIT, Boston, pp 267 – 290

Gary G. R., Gary R. G. P. E , & Frank C. (1999): "Application of Neuro-Fuzzy Systems to Behavioral Representation in Computer Generated Forces". Retrieved from http://www.hti.com/pdfs/neuro-fuzzy.pdf

Gary, S. et al., (2001) "Risk Management Guide for Information Technology Systems", Special pub. 800-300, National Institute Standards and Technology, 2001

Gatziu et al. (1999): "Design and management of Data Warehouse." Report on the DMDW 99 Workshop: http://sunsite.informatik.rwth-aachen.de/DMDW99/

Ge, Mouzhi & Markus Helfert (2006): " A Framework to Assess Decision Quality Using Information Quality Dimesions". In proceedings of the International Conference on Information Quality. ICIQ-06, Boston

Gilbert, D. (2006): "I'm Ok!, You are Biased" New York Times. Accessed from http://www.nytimes.com/2006/04/16/opinion/16gilbert.html? On 16/11/2009

Gillies, A.C. (1996) Software Quality-Theory and Management. International Thomson Publishing (1996)

Glover, E. J., Lawrence, S., Gordon M. D., Birmingham W. P., & Lee Giles C., Web Search – YourWay, Communications of the ACM, 1999.

Goodhue, D. L.: "Understanding user evaluations of information systems", Management Science 41(12), 1995, Pp. 1827 – 1844.

Goodhue, D.L., 1995, Understanding user evaluations of information systems. *Management Science*. 41(12): 1827-1844.

Goodman, S. K. (1993) 'Information needs for management decision-making', Records Management Quarterly 27(4): 12-23.

Google (2006) "Google search", http://www.google.com/.

Gordon, M.: Probabilistic and genetic algorithms for document retrieval, Communication of the ACM, 31(10), 1988, pp. 1208–1218.

Gordon, M.D.: User-based document clustering by redescribing subject descriptions with a genetic algorithm, Journal of the American Society for Information Science, 42(5), 1991, pp. 311–322.

Goria, S., Knauf, A., David, A. & Geffroy, P.: "The Economic Intelligence Process- a study according to the point of view of the information retrieval problem". Laboratoire Lorrain de Recherche en Informatique et ses Applications (LORIA) Campus Sceintifique BP 239, 54506 Vandoeuvre-les-Nancy, France.

Graham, L. & Metaxas, P. T. (2003) "Of course it's true; I saw it on the intermit!': critical thinking in the internet era', Communications of the ACM, Vol. 46, No. 5, Pp. 71 – 75.

Grefen, P.WP.J Apers, P.M.G: Integrity control in relation database systems. An overview. Data and knowledge Engineering Press, 10 (1993) 187-223

Grep. http: //www.gnu.org

Hall, H. (1994): "Information strategy and manufacturing Industry – case studies in Scottish textile industries. Int'l journal of information management, 14, 281 – 294.

Hammer, M. & Stanton, S. (1995)"The Reengineering Revolution". Haper-Colin, NY, 1995.

Harper, R.: (1999) "Information that counts: A sociological view of information navigation" in A. J. Munro, K, Hook and D. Benyon (eds), Social Navigation of Information Space, Springer, London, pp 81 – 89

Harper, S.C. (1990): "Intuition: What separate executives from manager from manager". W.H. Agor (Ed), Intuition on organization, 111- 124. Sage Publications, Newbury Park, Carlifonia

Harris, R., (1998)"Introduction to Decision Making", VirtualSalt – accessible on www.virtualsalt.com/crebook5.htm, July, 1998, pp 1 - 15

Hastie, R. (2001) "Problems for Judgment and Decision Making" Annual Review of Psychology, vol. 52, pp 653 – 683

Hayward, T. & Preston, J. (1998) 'Chaos theory, economics and information: The implications for strategic decision-making', Journal of Information Science 5 (3): 173-182.

Herrera – Viedman, E.: "Modeling the Retrieval Process of an Information Retrieval System using an Ordinal Fuzzy Linguistic Approach". Journal of the American Society for Information Science and Technology, 2001. Vol. 52: 460 – 475.

Hewett, R.: "Decision Making using incomplete Data", proc of 2004 IEEE int'l conference on systems, man & cybernetics. Pp 182-187.

Heydari, Z. F., Arabalibeik, F. H. & Parnianpour, M.: "Adaptive Neuro-Fuzzy Inference Systems for Classification of ACL-Ruptured Kness using Arthrometric Data". In Annals of Biomedical Engineering, 2008.

Hill, T & Westbrook, R. (1997): "SWOT Analysis – It's time for a Product Recall". Long Range Planning 30 (1): 46 – 52. http://dx.doi.org/10.1016%2FS0024-6301%2896%2900095-7

Hilton, R.W., 1979, The determinants of information system value: An illustrative analysis. *Journal of Accounting Research.* Autumn: 411-35.

Hocaoglu, F. O., Oysal, Y. & Kurban, M.: "Missing Wind Data Forecasting with Adaptive Neuro-Fuzzy inference System". Neural Computing & Applications. Springer Verlag, London, 2008.

Hong, T. & Lee, C.: "Learning Fuzzy Knowledge from training Examples" In ACM proc. Of 1998 on CIKM, Bethesda, MD, USA. Pp 161-166.

Hoppen, N & Meirelles, F. (2005): "Sisten de Informacia: un panorama da Pasquisa Cientifica Entre 1990-2003, Revista de Administration de Empresas 2005, vol 45, no 1, pp 24-35

Howell, D. C., (2009): "Treatment of Missing Data" http://www.uvm.edu/~dhowell/StatPages/More_Stuff/Missing_Data/Missing.html

http://choo.fis.utoronto.ca/fis/respub/aisis99/.

http://ieeexplore.ieee.org/ie15/9145/29024/01307879.pdf.
http://www.dcs.gla.ac.uk/Keith/Preface.html.
http://zing.ncsl.nist.gov/hfweb/proceedings/navarro-prieto/index.html.

Hunt, J. (2000) The anatomy of organisational change in the twenty-first century, paper

Inmon, W.H. 1996. Building the Data Warehouse. New York: John Wiley.

Inmon, W.H. 1999. Data Warehouse Performance. New York: John Wiley.

Isearch: http://www.cnidr.org/isearch/html

Isenberg, D. J. (1984) 'How senior managers think', Harvard Business Review Nov-Dec: 81- 86.

Ishikawa, K., (Lu. D. J. trans), 1985, What is Total Quality Control?, Prentice-Hall Inc., Englewood Cliffs, NJ.

Jang, J.S.R: "Adaptive-Network-Based Fuzzy Inference Systems" IEEE transaction on Systems and Cybernetics, 23(03): 665-685, May, 1993.

Jang, J. R. S (1995): "Inverse Learning for Fuzzy Controller Design". In proc of IEEE, pp 335-340, 1995.

Jang, J. R. S. & Sun, C.T.(1995): "Neuro-Fuzzy Modeling and Control". Proc of IEEE, March 1995, 83(3): 378-406.

Jang, J.S.R. (1996): "Input Selection for ANFIS Learning". Proc of IEEE on int'l conference on Fuzzy systems, New Orleans, Sept, 1996a. pp 1493-1499.

Jang, J.S.R: "Neuro-Fuzzy Modeling for Dynamic System Identification". Proc of IEEE on Artificial Intelligence. 1996(b), pp 320-325.

Janis, I. L. & Mann, L., (1977) "Decision Making: A Psychological Analysis of Conflict, Choice and Commitment." New York: Free Press

Jarke, M. & Vassiliou, Y. (1997): Data warehouse Quality: A Review of the DWQ project". Proc 2nd Conference on Information Quality, Cambridge 1997.

Jovanovic, B. B., Reljin, I.S. & Reljin, B. D.: "Modified ANFIS Arcitecture- Improving Efficiency of ANFIS Techniques". 7th IEEE seminar on NN application in electrical engineering, NEUREL 2004.pp215-220.

Jung, W. (2004): "A Review of Research: An Investigation of the Impact of Data Quality on Decision Performance". International Symposium on Information & Communication Technologies (ISITC'04), pp 166-171

Jungerman, H.: "One mode of decision making – no more, no less". Journal of Behavioural decision making, vol. 14, no 5, Pp.376 – 370, 2001. John Wiley and Sons Ltd.

Kacprzyk, J. & Zadronzny, S.: "Collective Choice Rules under Linguistic Preference: An example of Computing with Words/Perception Paradigm", in Information and Intelligent Systems. Proc of IEEE Explore, 2000, Pp. 786 – 791.

Kahn, B.K, Strong, D.M & Wang, R. Y. (2002): Information Quality Benchmark: Product & Services Performance". Communication of the ACM, April, 202, vol.45, no 4, pp 184-192.

Kahneman, D. & Tversky, A.: "Choice, Value, Frames, the Cambridge press, ISBN 0521621720.

Kahneman, D., & Tversky, A. (1982). Variants of uncertainty. *Cognition, 80,* 273-251.

Kahneman, D., & Tversky, A. (Eds.) (2000) "Choices, Values, and Frames." New York: Cambridge University Press.

Kahneman, D., Slovic, P., & Tversky, A. (1982) "Judgment under Uncertainty: Heuristics and Biases." New York: Cambridge University Press.

Kandel, A., Schneider, M. & G. Langholz: "The use of Fuzzy Logic for Management of Uncertainty in Intelligent Hybrid Systems" Proc of Uncertainty and intelligent Hybrid systems. Pp 569-587

Kang, B., Kim, D. & Lee1, S.: Semantic Indexing and Fuzzy Relevance Model in Information Retrieval. In Bo-Y. Kang et al.: Semantic Indexing and Fuzzy Relevance Model in Information Retrieval, Springer-Verlag Berlin Studies in Computational Intelligence (SCI) 2, 49–60 (2005)

Karen, G. & de Bruin, W. B.: "On the Assessment of Decision Quality: Consideration Regards Utility, Conflict and Accountability". In thinking psychological perspectives on reasoning and decision making. David Hardman and Laura Machi Ed., John Willey and Sons, Pp. 347 – 363, 2003.

Keeny, R. L., & Raiffa, H. (1976) "Decision with Multiple Objectives: Preference and Value Tradeoffs" New York: Wiley

Kelly, G. A. (1955) "The Psychology of Personal Constructs," New York, Norton 1955

Keren, G. (1991) "Additional tests of Utility Theory under unique and repeated conditions". Journal of Behavioural Decision Making, 4, 297 – 304

Kim, D.: "A Simulation Method for Cognitive Map". 1st Int'I conference on system thinking in management, 2002. pp294-299.

Kim, H.M., & Fox M.S (1994): Formal Models of Quality and ISO 9000 Compliance: an information systems approach. Proc. 48th Annual Quality Congress, Las Vegas, NV USA, (1994)

Kim, W. 1995. Modern Database Systems. ACM Press, 1995.

Kim, W. and Seo, J.Y. 1991. On classifying schematic and data heterogeneity in multidatabase systems. IEEE Computer, 24(12).

Kim, W., Chae, K.J., Cho, D.S., Choi, B.J., Kim, M., Lee, K.H., Lee, M.J., Lee, S.H., Park, S.S., and Yong, PIPER – SU1112 (AD) D3.7 Maximising Web Visibility

Kim, W., Choi, B-J., Homg, E,-K; Kim, S-K and Lee, D.(2003) "A Taxonomy of Dirty Data". Data Mining and Knowledge Discovery, 2003, Kluwer Academic Publishers. 7.2003. pp. 81-99

Kim,W., Choi, I.J., Gala, S., and Scheevel, M. 1993. On resolving schema heterogeneity in multidatabase systems. Distributed and Parallel Databases, 1(1):251–279.

Kimball R., Reeves L., Ross M., & Thornthwaite W. (2000), "The Data Warehouse Lifecycle Toolkit", Wiley Computer Publishing, New York, NY, 2000.

Kinney, (2000): "Information Quality Assurance and Internal Control For Management Decision making", McGraw-Hill Higher Education, Boston, Massachusetts, USA, 2000.

Knauf, A. & David, A. (2004): "The Role of the informediary in EI process". The World 8th Multi-conference on Sytemic, Cybernetics & Informatic- WMSCI'04.

Knox, T. K.: "An Investigation into the Information strategy formulation process within a higher education institution: A case study approach, Nottingham Buz. School. Nottingham, Nottingham Trent University: 256, 2007a.

Knox, T. K.: "The various and conflicting notions of Information". In Issues in information science and Information technology, vol. 4, Pp. 674 – 689, 2007b.

Koller, G. R., (2000) "Risk Assessment and Decision Making in Business and Industry: A practical guide" CRC press, LLC.

Kotz, D., & Gray, R.S.(1999) Mobile Agents and Future of Internet. ACM Operating Systems Review, 33: 3(1999)7-13

Kraft D., Bordogna G., & Pasi G.: Fuzzy Set Techniques in Information Retrieval, in: "Fuzzy Sets in Approximate Reasoning and Information Systems", J. C. Bezdek, D. Dubois and Prade eds, volume of the series "The Handbooks of Fuzzy Sets Series", Kluwer Academic Publishers, pp. 469–510, 1999.

Kraft, D. H., Pasi, G. & Bordogna, G.: "Vagueness and uncertainty in information retrieval: how can fuzzy sets help?" Proc. of the 2006 international workshop on Research issues in digital libraries, 2006. http://doi.acm.org/10.1145/1364742.1364746

Kroenke, D. M. (2007). *Using MIS*. Upper Saddle River, New Jersey: Pearson, Prentice Hall.

Kruse, R. Borgelt, C. & Nauck, D.: "Problems & prospects in Fuzzy Data Analysis". In B. Azvine et al.(Eds): intelligent systems and soft computing, LNAI 1804,pp 95- 109, 2000.

Kruse, R., Borgelt, C. & Nauck, D. "Fuzzy Data Analysis: Challenges and Perspectives". In proc of 1999 IEEE int'l conference on fuzzy systems, August 25-29, Seoul. Pp 1211-1216.

Kruskal, Joseph, & Mark Liberman. [1983] . The Symmetric Time-Warping Problem: From Continuous to Discrete. In *Time Warps, String Edits and Macromolecules: The Theory and Practice of Sequence Comparison*, ed. David Sankoff and Joseph Kruskal. 125–161. Reprint, with a foreword by John Nerbonne, Stanford, CA: CSLI Publications.

Kühne, T. (2005): What is a model. In Dagstuhl seminar proc. 04101 on Model-Driven Language Engineering. http://drops.dagstuhl.de/opus/volltexte/2005/23

Langefors, B. (1966): Essays on Infology - Summing up the planning for the future. Lund: Studentlitteratur

Langefors, B. (1966): Theoretical analysis of information systems. Lund: Studentlitteratur.

Langley, A., Mintzberg, H., Pitcher, P., Posada, E. & Saint-Macary, J. (1995) 'Opening up decision making: The view from the black stool', Organization Science 6 (3): 260-279.

Lawry, J. (2001): "A Methodology for Computing with Words". Int'l journal of Approximate Reasoning 28 (2001), 51-89. www.elsevier.com/locate/ijar.

Lee J.H.: (1994) On the evaluation of Boolean operators in the extended Boolean retrieval

Lee, H. M., (1996) "Group Decision making using Fuzzy set theory for evaluating the rate aggressive risk in software development," Fuzzy Sets and Systems, vol. 80, issue 3, pp 261 – 271, June, 1996

Lee, Y. W., Strong, D. M., Kahn, B. K. & Wang, R. Y. (2002): AIMQ: A Methodology for Information quality assessment. Elsevier Information and Management 40 (2002) 133 – 146.

Lerner, J.S. & Tetlock, P.E. (1999) "Accounting for the Effect of Accountability". Psychological review, 125, 255 – 275

Lesser, V., Horling, B., Raja, A, Zhnag X., & Wagner, T. (2000): Resource Bound Searches on an Information market place. IEEE Internet Computing 4:2 (2000) 49-58

Levine, P. & Pomerol, J. C. (1989) "System Interactifs d'Aide a la Decision et System Expert," Hermes, Paris

LEXA. http: //nora.hd.uib.no/lexainf.html

Lima, L.F.R. Macada, A.C.G. & Vargas, L.M (2006):"Research Into Information Quality: Study of the State-of-Art in IQ and its Consolidation." In Proceedings of the International Conference on Information Quality. ICIQ-06, Boston

Lin, C., Chen, K. & He, Y.: "A Hybrid Method to Construct Fuzzy Cognitive Maps". In proc of 5th WSEAS int'l conf on information security & privacy, Venice, Italy, 2006. pp118-123.

Lin, X., Soergel, D., Marchionini, G.: A self-organizing semantic map for information retrieval, Proceedings of the 14th Annual Int. ACM SIGIR Conference on Research and Development in Information Retrieval, 1991, pp. 262–269, Chicago (IL)

Lipshitz, R. (1989) "Either a medal or corporal": The Effects of success and failure on the evaluation of decision making and decision maker. Organization Behaviour and Human Decision Processes, 44, 380 - 395

Lipshitz, R. (1995) "Judgment by Outcome ": Why is it interesting? A reply to Hershey & Baron: "Judgment by Outcome: When is it justified?" Organization Behaviour and Human Decision Processes, 62, 123 - 126

Liu, S. (2007): "Computing with Words using Fuzzy Logic: Possibilities for application in automatic text summarization". In O. Castillo, et al., (Eds): Theor. Adv. & Appl. of Fuzzy Logic, ASC 42, pp 151 – 160, Springer Verlag, Berlin Heidelberg.

Lokuge, P. & Alahakoon, D.: "Hybrid BDI Agents with ANFIS in Identifying Goal Success factors in a Container Terminal Applications". Proceeding of the 2004 ACM int'l conference on Autonomous Agent and Multiagents Systems, AAMAS'04, July 19 – 23, 2004, New York, USA.

Long, J. A & Seko, C.E. (2002) "A New Method for Database Data Quality Evaluation at the Canadian Institute for Health Information (CIHI)." Proc of the 7th Int'l Conference in Information Quality (ICIQ'02), MIT, Boston, pp 238 – 250

Lu, Y., Yeh, F., Li, C. & Wu, M.: "Study of using ANFIS to the prediction in the Bore-Expanding Process". Int'I Journal of Adv. Manufacturer Technology (2005) 26:544-551. Springer- Verlag, London.

Lueg, C. (2001). Information, Knowledge and Networked minds. Journal of Knowledge Management, 5(2), 151 – 159.

Lui, L. & Chi, L.N. (2002): "Evolution Data Quality: A Theory Specific View". In Proceedings of the Seventh International Conference on Information Quality. ICIQ-02, pp 229-304.

Machlup, F. (1983): "Semantics quirks in studies of Information: In F. Machlup and U. Mansfield – the study of information: Interdisciplinary messages, New York, Willey and Sons 1983.

Macleod, K.J., Robertson, W.: A neural algorithm for document clustering, Information Processing & Management, 27(4), 1991, pp. 337–346

Maes, P. (1994): Agents that Reduce Work And Information Overload. Communication of the ACM, 37:7 (1994) 31-40

Maes, P. Guttman, R.H., & Moukas, A.G.(1999): Agents that Buy and Sell: Transforming Commerce as we know it. Communication of the ACM, 42: (1999) 81-91

Makela, J.M. "The Impact of Spatial Quality on Company's decision making". http://www.tkk.fi/Units/Cartography/research/impactspatial.html

Malaga, R.A. (2005). Information systems Technology. Person Prentice Hall

Management. 13(2): 79-83.

Maqsood, T., Finegan, A. D., & Walker, D. H. T: "Biases and Heuristics in Judgment and Decision Making: The Dark Side of Tacit Knowledge". In issues in Informing Science and Information Technology, pp 295 – 301, 2004.

Mar, J. & Lin, F.J.: "An ANFIS Controller for the Car-following Collision Prevention System". IEEE transaction on vehicular technology. voI. 5, no 4, July 2001. pp 1106-1113.

Marchand, D. A. (2000). Competing with information: A manager's guide to creating business value with information content. Chichester: John Wiley and Sons.

Marlin, B.: "Missing Data problems & Collaborative Filtering". http://www.cs.toronto.edu/~marlin/teaching/2535S08-lecture-slides-6up.pdf

Martinez, L. (2007): "Sensory Evaluation Base on Linguistic Decision Analysis". Int'l journal of Approximate reasoning 44 (2007), Pp. 148 – 164.

Martre, H. (1994) :, "Intelligence Economique et strategie des enterprises", Rapport du Commissariat General au Plan, Paris, La Documentation Francaise, 1994, Pp. 17, 18.

Mateou, N.H & Andreou, A.S.(2005): "Tree Strctured Multi-Layered FCM for Modeling Large Scale Complex Problems." Proc of the 2005 int'I conf on intelligent agents, web technology & internet commerce (CIMCA-IAWTIC'05).

Mateou, N.H., Andreou , A.S. & Stylianou, C.(2006): "Evolutionary multi-layered FCMs- A Hybrid System Design to Handle Large Scale, Complex Real World Problems". In proc of 2006 IEEE explore, pp 1663-1668.

Mateou, N.H., Mosieos, M. & Andreou, A.S.(2005): "Multi-Objective Evolutionary Fuzzy Cognitive Maps for Decision Support". In proc 2005 IEEE, explore pp 824-830.

Matsumura, A. & Shouraboura, N. (1996) "Competing with Quality Information" Proc of the Int'l Conference in Information Quality (ICIQ'96), MIT, Boston, pp 72 – 86

May, R., (2006): "Cognitive Bias in Strategic Decision Making" http://www.businesspundit.com/cognitive-bias-in-strategic-decision-making/

McCord, M. & de Neufville R. (1983), Fundamental deficiency of expected utility analysis, in S. French et al (Eds.), *Multiobjective Decision Making*, pp. 279-305, Academic Press, London.

McCord, M. & de Neufville, R. (1983), Fundamental deficiency of expected utility analysis, in S. French et al (Eds.), *Multiobjective Decision Making*, pp. 279-305, Academic Press, London.
McGraw-Hill, 1983

Mende, Jens (2005): "The Poverty of Empiricim". In Informing Science Journal, Volume 8, 2005, pp 189-210.

Merriam-Webster's Collegiate Dictionary. (1999). Springfield, Mass: Merriam-Webster Inc. Third Edition.

Meta Group – Data warehouse Score Group, 1999.

Miao, Y. Lui, Z. Li, S. & C.K. Siew.: '' Dynamic cognitive Networks : An extension of Fuzzy Cognitive Maps" Accessed from ieeexplore.ieee.org/ie15/6582/17565/00809764. pdf?arnumber=809764.

Mike, (2009): "The problem of dirty data" http://www.articlesbase.com/print/1111299 Accessed on 05/11/2009
Miller, F.J., 2002,1 = 0 (Information has no meaning). *Information Research.* 8(1).
Miller, H. (1996) "The Multiple Dimension of Information Quality". Information System Management, vol. 13, no 2, pp. 79 – 82.
Miller, H., 1996, The multiple dimensions of information quality, *Information Systems*
Mingers, J. (1997). The Nature of Information and Its Relationships to Meaning. In R.L. Winder, S. K. Probert and I. A. Besson, (eds), *Philosophical Aspects of Information Systems*. London: Taylor & Francis, pp. 73-84.
Mintzberg, H. (1990): "The design school, Reconsidering the basic premises of strategic management". Strategic Management Journal, 11, 171 – 195.
Mintzberg, H. (1994): "The Rise and Fall of Strategic planning". The Free Press: New York.
Mintzberg, H., Raisinghani, D. & Theoret, A., (1994): "The structure of 'unstructured' decision processes". Administrative Science Quarterly, 21, 246 – 275.
Mitra, S. & Hayashi, Y.: "Neuro-Fuzzy Rule Generation: Survey in Soft Computing Framework". IEEE transaction on neural networks, vol. 11, no 3, pp 748-768, May, 2000.
Mizzaro, S., A cognitive analysis of information retrieval, Proceedings of CoLIS2, 1996, pp. 233–250, Copenhagen Denmark.
Mizzaro, S., How many relevancies in information retrieval?, Interacting with Computers, 10(3), 1998, pp. 305–322
Moens, M.F (2000) Automatic Indexing and Abstracting of Document Texts, Kluwer Academic Publishers
Mon, Y.: "Airbag Controller Design by Adaptive-network- based Fuzzy inference System (ANFIS)". Fuzzy Sets and Systems 158 (2007) 2706-2714, Elsevier publications.
Morris, C, 1946, *Signs, Language and Behavior,* Prentice Hall Braziller, New York.
Mousavi, S.J., Ponnambalam, K. & Karrray, F.: "Inferring Operating Rules for Reservoir Operations using Fuzzy Regression and ANFIS". Elsevier pub. Science direct on fuzzy sets & Systems, 158(2007), pp 1064-1082.
Muramatu, J. & Pratt, W.: Transparent queries: Investigating Users' Mental Models of Search Engines. Proceedings of SIGIR 2001.
Naeza-Yates, R., & Riebeiro-Neto, B.: Modern Information Retrieval, Addison Wesley, NewYork, 1999.
Nake, F., 2001, Data, information, and knowledge, in: *Organizational Semiotics: Evolving a science of information systems,* K.Liu, R.J.Clarke, P.B.Anderson, R.K.Stamper, and E.Abou-Zeid, ed., Kluwer Academic Publishers, Boston, pp. 41-50.
Nauck, D.(1997): "Neuro-Fuzzy Systems: Review and Prospects:. Proc of 5th E.U congress on intelligent Techniques & Soft Computing (EUFIT'97), Aachen, Sept. 8-11, 1997, pp1044-1053.
Naumann, F. (2001) 'From databases to information systems information quality makes the difference', Proc. Int. Conf. on Information Quality (ICIQ 2001), Boston, MA, USA, Pp. 244 – 260.
Navarro G (2001). "A guided tour to approximate string matching". *ACM Computing Surveys* **33** (1): 31–88. doi:10.1145/375360.375365.

http://www.egeen.ee/u/vilo/edu/20023/Tekstialgoritmid_I/Articles/Approximate/Navarro_Review_on_Approximate_Matching_p31-navarro.pdf.

Navarro, G., Ricardo Baeza-Yates, E. Sutinen and J. Tarhio (2001). "Indexing Methods for Approximate String Matching". *IEEE Data Engineering Bulletin* 24 (4): 19–27. http://www.dcc.uchile.cl/~gnavarro/ps/deb01.pdf.

Navarro-Prieto, R., Scaife, M., & Rogers, Y.: Cognitive Strategies in Web Searching. Proceedings of the 5th Conference on Human Factors & the Web, 1999. Available at http://zing.ncsl.nist.gov/hfweb/proceedings/navarro-prieto/index.html

Nayak, P.C., Sudheer, K. P D., Rangan, M. & Ramasatri, K.S.: "A Neuro-Fuzzy Computing Technique for Modeling Hydrological Time Series". Journal of hydrology, Elsevier pub. Science direct, 2004. pp 52-66.

Nayer, M. "Information Systems management", 10(2), Spring 199, 51-58.

Neill, S. D. (1992). *Dilemmas in the Study of Information: Exploring the Boundaries of Information Science*. New York: Greenwood Press.

Neutrality". In Informing Science – Int'l journal of emerging transdiscipline. Vol. 11, Pp

Nick Craswell, David Hawking & Stephen Robertson (2001). Effective site finding using link anchor information. SIGIR'01. Available at http://pigfish.vic.cmis.csiro.au/~nickc//pubs/sigir01.pdf

Niskanen,V. (1989). Introduction to imprecise reasoning. *Uncertainty, decision making and knowledge engineering,* editor Keskinen,K.. Publications of the Society for Artificial Intelligence in Finland, no 1. Helsinki (in Finnish).

Niskanen,V., (1998) What is fuzzy reasoning? *"Possibilities of fuzzy logic",*

Nourrisier, P., Thiery, O. David, A. & Micheal Ignatio: "Dynamic System of Management of Information Adapted to the Portal of Company". Laboratoire Lorrain deRecherche en Informatique et ses Applications (LORIA) Campus Scientifique BP 239, 54506 Vandoeuvre-lesNancy, France

Numann, F. & Rolker, C. (2000): "Assessment methods for incorporating Quality Criteria." Proceeding of 2000 conference on Information Quality (IQ-2000), Cambridge, MA, USA, pp 148-162.

Numann, F. (2002): "Quality- Driven Query Answering for Integrated Information Systems". Springer Veriag, Germany, 2002.

Nurnberge, A. & Kruse, R.: "Neuro-Fuzzy Techniques under MATLAB/SIMULINK Applied to a Real Plant". Proc of IEEE 1998, pp 572-576.

Nutt, P. C. (1999) 'Surprising but true: Half the decisions in organizations fail', Academy of Management Executive 13 (4): 75-89.

O'Brien, J. A. (2003). Introduction to Information Systems. (Eleventh Edition), Mc Graw-Hill/Irwin.

O'Brien, J. A. (2004). Management information systems (Sixth Edition), Mc Graw-Hill/Irwin

OCP. http: //info.ox.ac.uk/ctitext/resguide/resourceso125.html

O'Leary, D. E. & Selfridge, P. (2000). Knowledge management for best practice. Communications of the ACM, 43(11), pp: 12-24.

Oliveira, P., Rodrigues, F. & Henriques, P. (2005): "A Formal Definition of Data Quality Problems". In Proceedings of the Tenth International Conference on Information Quality ICIQ-05.

Onifade O.F.W & OJesanmi O.A (2005): “Agent-Based Stock Monitoring and Security Systems (ASTOM)”, Journal of Research in Physical Science, Volume 1, Number 1, 2005. pp 56-60

Onifade O.F.W. , Thiery O., Osifisan., A.O. & Duffing G. (2008b): “ Ontological Framework for Minimizing the Risk of Non Quality Data during Knowledge Reconciliation in Economic Intelligence”, in proc of the 13th Int’l Conference in Information Quality (ICIQ’08), MIT, Boston, pp 296 – 309.

Onifade O.F.W. , Thiery O., Osifisan., A.O. & Duffing G. (2010): “FUZZONTOLOGY: Resolving Information Mining Ambiguity in Economic Intelligent Process”. In the proc. Of International Conference on Information Systems, Technology and Management (ICISTM), Bangkok, Thailand, March 11-13, 2010 (To Appear)

Onifade, O. F. W., “Cognitive Based Risk Factor Model for Strategic Decision Making in Economic Intelligence Process”. GDR-IE Workshop, June 16 and 17, 2008 http//s244543015.onlinehome.fr/ciworldwide/wp-content/uploads/2008/06/nancy_onifadeofw.pdf.

Oxford English Dictionary (OED) (1989).

Page, L., Brin, S., Motwani, R. & Winograd, T. (1998) ‘The PageRank citation ranking bring order to the web’ Stanford Digital Library Technologies Project, Vol. 5, p. 26

Pasi G., Modelling Users’ Preferences in Systems for Information Access, International Journal of Intelligent Systems, Vol. 18, pp. 793–808, 2003.

Pasi, G.: “Fuzzy Sets in Information Retrieval: State of the Art and Research Trends”. In H. Bustince, et al., (eds.), *Fuzzy Sets and Their Extensions: Representation, Aggregation and Models,* Springer, pp 517 – 535, 2008

Passam, H. C., Tocatlidou, A., Mahaman, B. D. & Sideridis, A. B., (2003): “Methods for Decision making with Insufficient Knowledge in Agriculture”. In proc of EFITA 2003 Conference, 5 – 9 July, Debrecen, Hungary. Pp 727 – 731

Peguiron, F. & Kislin, P.” “Activity-Based Classification of University Actors for the Construction of a Domain Oriented Data Warehouse.” Laboratoire Lorrain de Recherche en Informatique et ses Applications (LORIA) Campus Sceintifique BP 239, 54506 Vandoeuvre-les-Nancy, France.

Pelta, D.A., A. Sancho-Royo, C. Cruz y J.L. Verdegay(2006): Using memory and fuzzy rules in a co-operative multi-thread strategy for optimization. Information Science, 176 (2006), 1849 – 1868

Pešl, J. & Hřebíček, J. (2003): “Soft Systems Methodology Applied to Environmental Modeling”. Environmental Informatics Archives, Volume 1 (2003), 261-266

PIPER – SU1112 (AD) D12.1 Study of Web-based Dissemination http://piper.ntua.gr/reports/searcheng/doc0014.htm (Accessed on 20/11/2009)

Pipino, L.L., Lee Y. W. & Wang, R.Y. (2002): “Data Quality Assessment”. The Communication of the ACM, April 2002, vol. 45, no 4e, pp 211-218.

Piratla, M.N. & Jayasumana, A. P.: ‘A Neural Network based Real-Time Gaze tracker”. Journal of Network and computer Applications, 25, pp 179-196, July 2002.

Pomerol, J. C. (1997) “Artificial Intelligence and Human Decision making” European Journal of Operation Research, 99 (1997), 3 – 25

Pomerol, J. C. (1997) “Cognition and Decision about some recent results in Neurobiology”, proceedings ISDSS `97, Lausanne, 1997, pp 115 – 125

Pomerol, J. C. (1998) “Scenario Development and Practical Decision Making under Uncertainty: Application to Requirements Engineering” 3 (1998) 174 – 181
Pomerol, J-C. (2001) “Scenario Development and Practical Decision Making under Uncertainty”. Decision Support Systems 31(2001), 197 – 204.
Pomerol, J-C., Roy, B., Rosenthal-Sabroux., & Saad, A. (1995) “An ‘intelligent’ DSS for the multicriteria evaluation of Railway timetable” Foundation of Computing and Decision Sciences 20(3) 219 – 238.
Raisinghani, V.T. (1999): “Cleaning Methods in Data Warehouse”. http://citeseer.ist.psu.edu/raisinghani99cleaning.html
Raneses, A., Deyong, C. & Glasstein, N. (2006): “Modeling Operational Risk In Data Quality”. In Proceedings of the International Conference on Information Quality. ICIQ-06, Boston.
Redman, T.C. (1998): “The Impact of poor Data Quality on the Typical Enterprise”. Communications of the ACM, Volume 41, Issue 2 (Feb. 1998), Pages: 79-82
relevant? ...probably. A survey of probabilistic models in information retrieval.
Revelli, C., “Intelligence strategique sur Internet”, Paris, Dunod, 1998, Pp. 18, 19.
Riichiro Mizoguchi: http://www.ei.sanken.osaka-u.ac.jp/pub/miz/Part1-pdf2.pdf
Robert, C. P. (1994). The Bayesian choice: a decision theoretic motivation. New York: Springer-Verlag.
Robertson, S.E., Sparck Jones, K.: Relevance weighting of search terms, Journal of the American Society for Information Sciences, 27(3), 1976, pp. 129–146
Robins, D.: Interactive Information Retrieval: Context and Basic Notions, Information Science, 3(2), 2000, pp. 57–61.
Rocchio, J.J., Relevance Feedback in Information Retrieval, Prentice Hall, 1971.
Rolker C. & Kramer, R. “Quality of Servcie transferred to Information Retrieved An Adaptive Information Retrieved System” Proc. 8th Int. Conf. Information Knowledge management, Kansas City, Mo USA (1999)
Roy B. & D. Bouyssou (1986), Comparison of two decision-aid models applied to a nuclear power plant citing example, *EJOR*, Vol. 25, pp. 200-215.
Roy, B. & Vincke, P. H. (1984). “Relational System of Preference with one or more pseudo-criteria: some new concept and result”. Management Science, 30(11), pp 1323 – 1335
Roy, B. (1988a), Main sources of inaccurate determination, uncertainty and imprecision in decision models, in B.R. Munier and M.K. Shakun (Eds.), *Compromise, negotiation and group decision*, pp. 43-62, Reidel, Dordrecht.
Rutkowska, D., Nowicki, R. & Hayashi, Y.: “Parallel Processing by Implication-Based Nero-Fuzzy Systems”. In R. Wyrzkowski et al. (Eds) PPAM 2001, LNCS 2328, pp599-607, 2002. Springer Verlag Berlin Heidelberb, 2002.
Salton G, Fox, E. A, & Wu H.: Extended Boolean information retrieval, Communications of ACM, December, 1983, vol. 26, issue 11, pp 1022 - 1036
Salton G., Automatic Text Processing - The Transformation, Analysis and Retrieval of Information by Computer, Addison Wesley Publishing Company, 1989.
Salton G.,& McGill M.J.: Introduction to modern information retrieval. New York, NY:
Saracevic, T., RELEVANCE: A Review of and a Framework for the thinking of the notion in information science, Journal of the American Society for Information Science, 26(6), 1975, pp. 321–343.

Sattler, K.& Schallehn, E., "A Data Preparation Framework based on a Multidatabase Language." In Proceedings of International Database Engineering and Applications Symposium (IDEAS 2001), Grenoble, France. IEE Computer Society. 2001. pp. 219-228.

Schneider, M., Schnaider E, Kandel, A & Chew, G.: "Constructing Fuzzy Cognitive Maps". Proc of Int'I conference on fuzzy systems, 1995, pp 2281-2288.

Schoemaker, P. H. J. (1982) "The Expected Utility Model: Its variant, purposes, evidence, and limitation." Journal of Economic Literature, 20, 259 – 563

Schutz, A. (1967). Phenomenology of the social world. Northwestern University Press.

Search Engine Watch http://www.searchenginewatch.com

Search Engines". Proceedings of SIGIR 2001.

Sebastiani, F., On the Role of Logic in Information Retrieval, Information Processing & Management, 34(1), 1998, pp. 1–18.

Selberg, E. & Etzioni, O.: The Meta Crawler Architecture for Resource Aggregation on the Web. IEEE Expert, 12:1 (1997) 8-14

Sellers PH (1980). "The Theory and Computation of Evolutionary Distances: Pattern Recognition". *Journal of Algorithms* **1**: 359–73. doi:10.1016/0196-6774(80)90016

Setlak, G. : "The Fuzzy-Neuro Classifier for Decision support". Int'l Journal of Information Theories & Application. Vol 15, pp 21-26, 2008.

Shaarawy, M., Belal, M. & ElGindy, E.: "A Neuro-Fuzzy Model for Motion Cognition". In 6th IEEE/ACIS Int'l Conference on Computer and Information Science (ICIS'07). Pp. 406 – 411.

Shafer, G. (1976). A mathematical theory of evidence. Princeton, NJ: Princeton University Press.

Shankaranarayanan G., Ziad M., & Wang R. Y. (2003), "Managing Data Quality in Dynamic Decision Making Environments: An Information Product Approach", Journal of Database Management (14:4), Oct – Dec 2003, Pp. 14 – 32.

Shannon, C. F., & Weaver, W. (1949) "A Mathematical Model of Communication." Urbana, IL: University of Illinois Press

Shapiro C., & Varian H. R. (1999) "Information Rules", Harvard Business School Press, Cambridge, MA.

Sidhom, S., Thiery, O. & David, A.: "Human Resources Consultancy (HR) in France: on the tendency of web services innovation to e-HR". Labortoire Lorrain de Recherche en Informatique et ses Applications (LORIA) Campus Sceintifique BP 239, 54506 Vandoeuvre-les-Nancy, France.

Silver, M. S., (1990)"Decision Support Systems: Directed and Non-directed change", Information Systems Research, Vol. 1, No 1, 1990, pp 47 – 70

Simier et al., P. "Historical Review of the Evolution of Information Systems" Laboratoire Lorrain de Recherche en Informatique et ses Applications (LORIA) Campus Sceintifique BP 239, 54506 Vandoeuvre-les-Nancy, France.

Simon, H. A. (1987) 'Making management decisions: The role of intuition and emotion', Academy of Management Executive 1 (1): 57-64.

Sinclair, M, & Ashkanasy, M. N. (2005): "Intuition: Myth or a Decision-Making Tool? Journal of Management Learning (2005) 36 (3): 353 – 370.

Smart, K. L., (2002) "Assessing Quality Documents," ACM Journal of Computer Documentation, vol. 26, no. 3, 2002, pp 130 – 140

Smeaton, A.F. Progress in the application of Natural Language Processing to Information Retrieval tasks. The Computer Journal, 35(3):268-278, 1992.
Smith, L.C., & Warner, A.J.: A taxonomy of representation in information retrieval design, Journal of Information Science, 8, 1984, pp. 113–121.
Spinellis, D. (2003) 'The decay and failures of web references', Communications of the ACM, Vol. 46, No. 1, Pp. 71-77.
Srivastava, R.P & Mock, T.J Belief Functions in Business Decisions, Physica Verlag Heidelberg, Springer – Verlag Company, 2002.
Srivastava, R.P. & Mock, T.J. "Introduction to Belief Functions," In Belief Functions in Business Decisions, edited by R.P Srivastava and T. Mock, Physica –Verlag, Heidelberg, Springer-Verlag Company, 2002, 1-16.
Stach, W., Kurgan, L., Pedrycz, W. & Reformat, M.: "Evolutionary Development of Fuzzy Cognitive Maps". In proc of 2003 IEEE Int'I Conference on Fuzzy systems, pp 619-624.
Stachowiak, H. (1973): Allegemeine Modeltheorie. In Spinger-Verlag, Wien and New York, 1973.
Stephenson, B. Y., (1985)"Management by Information", Information Strategy: The Executive Journal, Vol. 1, No 4, 1985 summer. Pp 26 – 32.
Sterritt, R. (2000): "Soft Computing and Fault Management". Workshop on Soft Computing, University of Ulster with Nortel Networks, 21st September 2000, pp 1 – 14
Strong, D. M., Lee, Y. W. & Wang. R. Y. (1997): "Data Quality in context", Communications of ACM, 40(5): 103 – 110, May, 1997.
Stylios, C.D. & Groumps, P.P.: " Modeling Complex Systems using Fuzzy Cognitive Maps" IEEE trans. On Systems, Man & Cybernetics-part A: system & humans, vol.34, no 1, pp 155-162, 2004.
Stylios, C.D., Georgopoulos, V.C., & Groumpos, P.P: "The use Fuzzy Cognitive Maps in Modeling Systems. Accessed from www.med.ee.nd/MEDs-1997/papers/067.pdf
Su, X., Bolzoni, D. & van Eck, P. (2007):" Understanding and specifying Information Security Needs to Support the Delivery of High Quality Security Services". The International Conference on Emerging Security Information Systems & Technologiees. (SECUREWARE-2007). Pp 107-114.
Susan & Guijun Wang (1996). "Information Fusion with ProFusion," Webnet 96 Conference,San Francisco, CA October 15-19, 1996. http://www.csbs.utsa.edu:80/info/webnet96/html/155.htm
Sveiby, K. E. (1997). The new organizational wealth: Managing and measuring knowledge-base assets. San Francisco: Berret-Koehler
Takagi, H.: "Application of Neural Networks and Fuzzy Logic to Consumer products". Proc of 1992 int'l conference on industrial electronics, control, instrumentation & automation. Pp 1629-1633.
Tang Kai Yin, C. S. George Lee, (1995) Fuzzy Model-Reference Adaptive Control, IEEE Transactions on Systems, Man, and Cybernetics, vol. 25, no. 12, December 1995.
Tanino, T. (1984). "Fuzzy Preference Ordering in Group Decision Making", fuzzy sets and systems, vol. 12, pp117 – 131
Tayi, G. K., & Ballou (1998) "Examining Data Quality". In communications of the ACM, 41, 2, Feb. 1998

Tejay, G., Dhillon, G., & Chin, A. G.: “Data Quality Dimension for Information system security: A theoretical exposition” – Invited paper. In Security management, Integrity & Internal Information Control in Information Systems, Springer Books, Boston, pp 21 – 39, 2006.

Tetlock, P.E. (1991) “An Alternative Metaphor in the study of judgment and choice: people as politicians”, Theory and Psychology, 1, 451 – 475

The MathsWorks: “Fuzzy Logic Toolbox 2 user’s guide. www.mathsworks.com 2008.

The Spider’s Apprentice http://www.monash.com/spidap.html

Thiery, Odile et David, Amos “Modelisation de l’utilisateur, systemes d’informations strategiques et Intelligence Economique”, La Lettre de l’Adeli, (Eleventh Edition), Mc Graw – Hill/Irwin.

Todd, P. & Benbasat, I., (2000) “The Impact of Information Technology on Decision Making: A Cognitive Perspective”, In Framing the Domains of IT Management, Pinnaflex Educational Resources, Inc., R. W. Zmud (ed.), 2000, pp 1 – 14.

Tomaiuolo, N. G. & Packer, J. G. (1996). "An analysis of Internet search engines: assessment of over 200 search queries." Computers in Libraries. v16 n6 (June 1996), p58 (5). The list of queries used is in: Quantitative Analysis of Five WWW "Search Engines".

Tomer, J. F. (1996) ‘Discussion of modes of economizing’, Journal of Economic Behavior and Organization 29: 261-262.

Tuomi, I., (1999)“Data Is More than Knowledge: Implication of the Reversed Knowledge Hierarchy for Knowledge Management and Organizational Memory”, Journal of Management Information Systems, winter, Vol. 16, No. 3. 1999, pp. 103 – 117

Turtle, H., Croft. W.B.: Inference networks for document retrieval, Proceedings of the13th Annual Int. ACM SIGIR Conference on Research and Development in Information Retrieval, 1990, pp. 1–24, Brussels, (Belgium).

Tversky, A. & Kahneman, D. (1973) “Availability: A heuristic for judging frequency and probability”. Cognitive Psychology, 5, 207 – 232

Tversky, A. & Kahneman, D. (1983) ‘Extensional versus intuitive reasoning: The conjunction fallacy in probability judgment’, Psychological Review 90: 293-315.

Tversky, A. & Simonson, I. (1993) “Context-dependent preferences”. Management Science 39, 1179 – 1189

Tversky, A., (1977) “On the elicitation of Preferences: Descriptive and Prescriptive Considerations” in Bell, D. Kenney, R., & Raiffa, H., eds. “Conflicting Objectives in Decisions”, Willey, New, York.

Vachtsevanos, G.J., Kim, S.S., Echauzak, J.R., & Ramani, V.K.: “Neuro-Fuzzy Approach to Decision making – in Application to Check Authorization from incomplete Information”. Journal of Intelligent & Fuzzy Systems, IOS press, Vol. 6, n2/1998, pp 259-278

van Rijsbergen, C.J., Information Retrieval, Butterworth’s, London, 1979. Available at

Vassiliadis et al., P. (2000): “Towards Quality Oriented Data Warehouse Usage and Evolution”. Towards quality-oriented data warehouse usage and evolution. Information Systems 25(2), 2000.

Vassiliadis, P. (2000): “Data Warehouse Modeling and Quality Issues”. These de la National Technical University of Athens, 2000.

Vaughan, F. E. (1979) Awakening intuition. New York, NY: Doubleday.

Viera, J., Dias, F.M & Moba, A.: “Artificial Neural Networks & Neuro-Fuzzy System for Modeling & controlling RTS”. Elservier publications in Engineering Applications of Artificial intellience 17 (2004), pp 265-273.

Vincke P., Multicriteria Decision Aid, John Wiley & Sons, 1992.

Virrantaus, K., (2003): “Analysis of the uncertainty and imprecision of the source data sets for a military terrain analysis application”. In proc of Int’l Symposium on Spatial Data Quality (ISSDQ 2003), pp 1 -10

Voorhees, E.M & Harman, D.K (2000). *Overview of the Eigth Text Retrieval Conference* (TREC-8). In: Information Technology: The Eigth Text Retrieval Conference (TREC-8). NIST SP 500-246, pp.1-23, GPO: Washington, D.C

Wagner, R. A. & M. J. Fisher. (1974). The string-to-string correction problem. Journal of the Association for Computing Machinery, 21(1):168-173, January 1974

Wakami, N.: Fuzzy Control and Neural Networks: Application for Consumer Products”. In fuzzy Logic & Fuzzy Control. Springer Berlin pub. Vol. 833/1994, pp 97-105.

Wall, A. (2007): “History of Search Engines: from 1945 to Google 2007” http://www.searchenginehistory.com/

Wally, S. & Baum, R. J. (1994) ‘Personal and structural determinants of the pace of strategic decision making’, Academy of Management Journal 37 (4): 932-956.

Wand, Y. & Wang, R.Y. “Anchoring Data Quality Dimension in Ontological Foundations”. Communications of the ACM, November 1996, 39(11), 86-95.

Wand, Y. & Weber, R.: “On Deep Structure of Information System”. Journal of Information Systems (1995) 5, Pp. 203 – 223.

Wand, Y. & Weber, R.: “On the Ontological Expressiveness of Information Systems Analysis and Design Grammars”. Journal of Information Systems (1993) 3, Pp. 217 – 237.

Wang R. Y. (1998), “A Product Perspective on Total Quality Management”, Communications of the ACM (41:2), Feb. 1998, Pp. 58 – 65.

Wang, H., Johnson, T. R. & Zhang, J. (1998): “UEcho: A Model of Uncertainty Management in Human Abductive Reasoning”. In proc. Of 20th Annual Meeting of the Cognitive Science Society, 1998, Hillsdale, NJ

Wang, P., Chao, K., Huang, C., Lo, C. & Hu, C.: “A Fuzzy Decision Model of Risk Assessment through Fuzzy Preference Relation with Users’ Confidence Interval”. Proc of 20th Int’l conference on advance Information Network and Application, (AINA’06), IEEE Society, 2006, Pp. 889 – 893.

Wang, P.P. (Ed): “Computing with Words”. Willey Series on Intelligent Systems, John Willey and Sons, 2001.

Wang, R. & D. M. Strong: “Beyond Data Accuracy: What Data Quality Means to Data Consumers”. Journal of Management and Information Systems, Springer 1996, Vol. 12, no 4, Pp. 5 – 34.

Wang, R. & Strong, D. M.: “Beyond Data Accuracy: What Data Quality Means to Data Consumers”. Journal of Management and Information System, Springer 1996, Vol. 12, No. 4, Pp. 5 – 34.

Wang, R. Y., Madnick, H. B. and Stuart, E. (1993): “Data quality requirement analysis and modeling. In 9th int’l conference on Data Engineering, 1993.

Wang, R.Y & Strong D.M. Journal of Management Information Systems, Spring 1996, Vol. 124, 5-34.

Wang, R.Y., Reddy, M.p., & Kon,(1995) H.B Decision Support Systems, 13, 349-372.
Wang, R.Y., Ziad, M., and Lee, Y.W., 2001, *Data Quality,* Kluwer Academic Pubhshers,
Wang, Y & Wang, Y (2004) "On Cognitive Models of the Brain", IEEE Transactions on System, Man, and Cybernetics, vol. 34,
Wang, Y., Wang, Y., Patel, S. & Patel, D. (2004) "A Layered Reference Model of the Brain", IEEE Transactions on Systems, Man and Cybernetics, vol. 34
West L. A. Jr. (2000), "Private Markets for Public Goods: Pricing Strategies of Online Database Vendors", Journal of Management Information Systems (17:1), 2000, Pp. 59 – 84.
Whatis.com (2005) "Taxonomy". http://searchciomidmarket.techtarget.com/sDefinition/0.sid83 gci31416,00.html
Widyantoro, D. H. & Yen, J. (2001) "Using Fuzzy Ontology for Query Refinement in a Personalized Abstract Search Engine" IFSA World Congress and 20th NAFIPS International Conference, 2001. Pp 610 – 615.
Wikipedia (2009): "SWOT Analysis". Accessed from http://en.wikipedia.org/wiki/SWOT_analysis on 11/11/2009
Wikipedia, (2009) "List of cognitive biases " http://en.wikipedia.org/wiki/List_of_cognitive_biases
Williams, N., Ivins W.K. & Burgess, M.S.E (2006): "Data Quality & Agile Method: A BT Perspective." In Proceedings of the Eleventh International Conference on Information Quality. ICIQ-06.
Wilmes, J. (2009): "Defeating Dirty Data – address the mess to improve your appearance in customers' eye". http://www.destinationcrm.com/articles/PrintArticle.aspx? Accessed on 05/11/2009
Wilson, R.A & Keil, F.C. (2001) "The MIT Encyclopedia of the Cognitive Sciences", MIT press
Wilson, T. D. (1986). Information needs in social services. Research highlights in Social Work, 13, 12 – 24. retrieved from http://information.net/tdw/pub1/paper/infoneed85.html
Wong, K. W., Gedeon T. D. & Koczy, L. T.: "Efficient Fuzzy Cognitive Modeling for Unstructured Information". 2006 IEEE Int'l conference on Fuzzy systems, Vancouver, BC, Canada, July 16-21, 2006. Pp. 358 – 363.
Xing, Z. Y., Zhang, L Jai & Hu, W.: "Design of Interpretable and Accurate Fuzzy Model from Data". In L. Wang & Jin, Y. (Eds.) FSKD 2005, LNAI 3613, pp 69-78, 2005. Springer-Verlag, Berlin 2005.
Yakov, Y. H (1998) "Risk Modeling, Assessment and Management," John Willey, 1998
Yingxu Wang, Dong Lui & Guenther Ruhe (2004): "Formal Description of the Cognitive Process of Decision Making". Proceedings of the third IEEE International Conference on Cognitive Informatics (ICCI'04)
Zadeh L. A., The concept of a linguistic variable and its application to approximate reasoning, parts I, II, Information Science, 8, pp. 199–249, pp. 301–357, 1975.
Zadeh, L. A.: Fuzzy Logic= Computing with words (CW) "IEEE transactions on fuzzy systems, vol 4, no 2, May 1996, pp 103-111.
Zadeh, L.: "Fuzzy Logic and the Calculus of Fuzzy if-Then Rules". OMRON Technics, vol 31, no 4, pp 316-320.1991.

Zadeh, L.A. (1965) "Fuzzy Sets," Information and Control, June 1965, pp. 338-353 8, p.p.338-353, 1965.

Zadeh, L.A. (1994). Soft Computing and Fuzzy Logic. IEEE Software 11, 6, 48-56.

Zadeh, L.A: “Fuzzy Logic, Neural Networks & Soft Computing”. In communication of the ACM, March 1994, vol 37. no 2.pp 77-84.

Zhang, W. & Chen, S. “A Iogical Architecture for Cognitive Maps”, In 2nd IEEE Int”I Conference on Neural Networks , vol. 1,pp 231 -238, 1998.

Zhang, X. Y., Jin & Guo, G.: “ANFIS Applied to a Ship Autopilot Design”. Proc of the 5th conference on machine learning and cybernetics, Dalian, 13-16 August, 2006. pp 2233-2236.

Zhang, X., Jin, Y. & Guo, G.: “ANFIS Applied to Ship Autopilot Design”. Proc of 5th int’l conference on machine Learning & cybermatics, 2006. pp 2233-226.

Zmud, R. (1978) “Concepts, theories and techniques: an empirical investigation of the dimensionality of the concept of information,” Decision Science 9(2), 1978, Pp. 187 – 195