



AVERTISSEMENT

Ce document est le fruit d'un long travail approuvé par le jury de soutenance et mis à disposition de l'ensemble de la communauté universitaire élargie.

Il est soumis à la propriété intellectuelle de l'auteur. Ceci implique une obligation de citation et de référencement lors de l'utilisation de ce document.

D'autre part, toute contrefaçon, plagiat, reproduction illicite encourt une poursuite pénale.

Contact : ddoc-theses-contact@univ-lorraine.fr

LIENS

Code de la Propriété Intellectuelle. articles L 122. 4

Code de la Propriété Intellectuelle. articles L 335.2- L 335.10

http://www.cfcopies.com/V2/leg/leg_droi.php

<http://www.culture.gouv.fr/culture/infos-pratiques/droits/protection.htm>

Indexation de documents pédagogiques : fusionner les approches du Web Sémantique et du Web Participatif

THÈSE

présentée et soutenue publiquement le 29 Octobre 2009

pour l'obtention du

Doctorat de l'université Henri Poincaré – Nancy 1
(spécialité informatique)

par

Benjamin Huynh-Kim-Bang

Composition du jury

Rapporteurs : Éric Bruillard
Manuel Zacklad

Examineurs : Brigitte De la Passardière
Richard Hotte
Anne Boyer
Monique Grandbastien (directrice)

Mis en page avec la classe thloria.

Remerciements

À ma directrice, Monique Grandbastien, qui m'a guidé dans l'univers de la Recherche à l'aide de son expérience, de sa patience et de sa grande gentillesse.

Aux membres du jury, pour leurs judicieux commentaires et pour l'inspiration fournie par leurs pratiques de la Recherche.

À Eric Dané, doctorant en science de l'éducation, dont l'apport fut essentiel pour mes travaux. À nos longues discussions qui ont marqué ma vision de la Recherche.

Aux membres de l'équipe MOCAH, pour leur soutien jusqu'au dernier moment.

Aux enseignants et aux documentalistes qui ont pris le temps de me faire découvrir leurs métiers.

Aux membres du laboratoire LORIA et notamment aux doctorants, en compagnie de qui j'ai découvert le monde de la Recherche.

Durant ces trois dernières années, je suis passé par beaucoup de désillusions et de remises en question. Bien plus que des qualités personnelles, c'est le soutien de mes proches, amis et famille, qui m'a permis de ne pas me replier sur moi-même et de dépasser ces épreuves.

Pour toutes ces rencontres, pour tous ces encouragements, je vous suis profondément reconnaissant.

*« Si l'ordre est le plaisir de la raison, le désordre est le délice de l'imagination. »
Le soulier de satin – Paul Claudel*

Table des matières

Introduction	1
1 Contexte du travail : les EIAH	1
2 Intérêt du partage de documents pédagogiques	1
3 Approches classiques de l'indexation	2
3.1 Indexation automatique	2
3.2 Indexation par des humains	4
4 Nouvelles approches	5
4.1 Présentation du Web Sémantique et du Web Participatif	5
4.2 Oppositions apparentes de ces nouvelles approches	5
4.3 Fusionner le Sémantique et le Participatif	6
5 Objectifs de la thèse	7
6 Plan	8
7 Quelques points de vocabulaire	8
Chapitre 1 Contexte de l'indexation de documents pédagogiques numériques	11
1.1 Nouvelles possibilités numériques	11
1.2 Indexation de documents pédagogiques numériques	12
1.2.1 L'approche par Learning Object	12
1.2.2 Les standards du E-learning	14
1.3 Conclusion	18
Partie I Web Sémantique, ontologies et indexation pédagogique	21
Chapitre 2 État de l'art	23
2.1 Web Sémantique	23
2.1.1 Présentation générale	23
2.1.2 Les outils du Web Sémantique	27
2.2 Web Sémantique appliqué au domaine pédagogique	31

2.3	Conclusion	33
Chapitre 3 Projet LUISA		35
3.1	Contexte	35
3.1.1	Contraintes du projet	35
3.1.2	Description générale du projet	36
3.2	Travaux	38
3.2.1	Étude des besoins à l’université et choix d’un contexte d’expérimentation	39
3.2.2	Prototype LUISA : fonctionnalités et raisonnements sémantiques . . .	47
3.2.3	Ontologies développées	61
3.2.4	Expérimentations et discussion	72
Partie II Web Participatif, communautés et indexation pédagogique		75
Chapitre 4 État de l’art		77
4.1	Web Participatif	77
4.1.1	Présentation générale	78
4.1.2	Concepts sous-jacents aux sites participatifs	80
4.1.3	Blogs	80
4.1.4	Wikis	81
4.1.5	P2P	83
4.1.6	Réseaux sociaux	84
4.1.7	Plateformes de partage	85
4.1.8	Partage de signets	85
4.1.9	Bilan des fonctionnalités	86
4.2	Web Participatif appliqué au domaine pédagogique	86
4.2.1	Web Participatif côté apprenant : elearning 2.0	86
4.2.2	Wiki éducatif	88
4.2.3	Réseaux pair-à-pair pédagogiques	88
4.2.4	Partage de signets et tags pédagogiques	89
4.2.5	Incitations à participer dans une communauté éducative	91
4.2.6	Plateformes pédagogiques pour les enseignants	92
4.3	Conclusion	93
Chapitre 5 Projet Pépi		95
5.1	Contexte	96
5.1.1	Différentes communautés d’enseignants : exemple de Cartables.net . .	96

5.1.2	Étude de la communauté PGM (Professeurs en Génie Mécanique) . . .	98
5.2	Développements	100
5.2.1	Objectifs	100
5.2.2	Principes d’une plateforme participative de partage de ressources pé- dagogiques	101
5.2.3	Fonctionnalités soutenant la participation	101
5.2.4	Architecture technique	107
5.3	Expérimentations	108
5.3.1	Contact des utilisateurs	108
5.3.2	Amorçage	108
5.3.3	Participations	108
5.4	Analyses et discussions	109
5.4.1	Analyses de l’échec des expérimentations	109
5.4.2	Discussion des concepts et travaux similaires	111
5.5	Conclusions	114

Partie III Fusion du Web Sémantique et du Web Participatif, et indexation pédagogique **117**

Chapitre 6 État de l’art **119**

6.1	Fusion du Web Sémantique et du Web Participatif	119
6.1.1	Limites du Web Sémantique et du Web Participatif	119
6.1.2	Rapprocher traitement sémantique et participation des utilisateurs . .	121
6.1.3	Ontologies « d’objets sociaux »	122
6.1.4	Des tags aux concepts	125
6.2	Fusion du Web Sémantique et du Web Participatif appliquée au domaine pédagogique	133
6.2.1	Créer des cours décrits sémantiquement dans un Wiki : SweetWiki . .	133
6.2.2	Des tags Delicious aux champs LOM	135
6.2.3	Annoter collectivement des ressources pédagogiques avec des concepts de Wordnet	135
6.3	Conclusion	136

Chapitre 7 Projet SemanticScuttle **137**

7.1	Modèle et fonctionnalités	138
7.1.1	Objectif des travaux	138
7.1.2	Modèle d’Indexation Progressive et Multi-points de vue (IPM)	138

7.1.3	Fonctionnalités illustrant le modèle IPM	141
7.2	Développement	144
7.2.1	Développement basé sur Scuttle, un outil classique de partage de signets	144
7.2.2	Fonctionnalités de structuration progressive	149
7.2.3	Fonctionnalités de collaboration progressive	158
7.2.4	Autres fonctionnalités et caractéristiques	160
7.3	Terrain d'application	163
7.3.1	Sociopôle, des signets en sociologie	164
7.3.2	WIKINDX, tentative d'utilisation d'un outil d'indexation bibliographique	166
7.4	Expérimentations	168
7.4.1	Méthodologie des expérimentations	168
7.4.2	Résultats : Facilité d'ajout d'un signet	172
7.4.3	Résultats : Continuum de fonctionnalités structurantes	173
7.4.4	Résultats : Continuum de fonctionnalités collaboratives	177
7.4.5	Résultats : Mieux « profiter » des ressources	178
7.4.6	Bilan des expérimentations	180
7.5	Discussions et travaux proches	181
7.5.1	Rapprochement avec le domaine pédagogique	182
7.5.2	Spectre de l'intérêt des utilisateurs	184
7.5.3	Perspectives de recherche	186
7.5.4	Le Web Sémantique du W3C peut-il être social?	187
7.5.5	Améliorer l'indexation automatique à l'aide de contributions des utilisateurs	189
7.5.6	Chercher et expérimenter dans le domaine des EIAH	190

Conclusion générale et perspectives **193**

Chapitre 8 Apports principaux **195**

8.1	Rappel et évolution des objectifs	195
8.2	Nos propositions	195
8.2.1	Présentation du cadre : Articuler ensemble différents procédés de classification	195
8.2.2	Modèle IPM	197
8.2.3	Application SemanticScuttle	198
8.2.4	Leçons tirées des premières expériences	198
8.2.5	Originalité et limites	199

Chapitre 9 Perspectives	201
9.1 Dans quels contextes pédagogiques, le modèle IPM et l’outil SemanticScuttle pourraient-ils être les plus intéressants ?	201
9.2 Comment gérer une continuité d’expertise ?	201
9.3 Comment gérer les liens entre utilisateurs ?	202
9.4 Comment intégrer les analyses automatiques dans l’indexation ?	202
9.5 Comment gérer techniquement la complexité croissante des métadonnées ? .	202
Annexes	203
Annexe A Publications	205
Annexe B Détails des remue-méninges et des entrevues pour LUISA	207
Annexe C Référentiel officiel du C2I niveau 1	211
Annexe D Exemple de ressource C2I	215
Annexe E GCS adapté au contexte du C2I	217
Annexe F Ontologie des logiciels du C2I : ComputerLiteracy	221
Annexe G Détails des règles sémantiques dans LUISA	223
Annexe H Mail aux responsables de Cartables.net	227
Annexe I Scuttle, écran par écran	229
Annexe J Autres fonctionnalités de SemanticScuttle	235
Annexe K Détails du ”focus group” sur le Sociopôle	239
Annexe L Expérimentation avec utilisateurs du Mechanical Turk	245
Annexe M Nos contributions par questions	247
Glossaire	253
Bibliographie	255

Introduction

1 Contexte du travail : les EIAH

Le travail de recherche présenté dans ce document concerne le domaine des Environnements Informatiques pour l'Apprentissage Humain (EIAH).

Ce large domaine contient par exemple des travaux sur les plateformes de formation en ligne ou sur les tuteurs virtuels. En France, ce domaine de recherche est notamment visible à travers la revue STICEF¹, les conférences EIAH² et l'association ATIEF³.

Dans ce contexte, nos travaux portent sur la description et le partage de documents en général, et de documents pédagogiques en particulier.

2 Intérêt du partage de documents pédagogiques

Utilité des documents pédagogiques. Pour enseigner, les professeurs emploient des documents pédagogiques tels que les manuels scolaires, les programmes officiels, leurs cours des années précédentes ou ceux de collègues. Ils utilisent aussi des documents et outils propres à leurs disciplines comme des analyses d'œuvres en Français ou des logiciels informatiques en Mathématiques.

Diversité des contextes, diversité des documents. Lorsqu'ils enseignent, les professeurs se situent dans des contextes divers. Ces contextes sont caractérisés par la discipline et le niveau. Mais ils se caractérisent aussi par des paramètres moins précis tels que les aptitudes des élèves ou le style pédagogique de l'enseignant. La diversité de ces contextes ne permet généralement pas aux enseignants de réutiliser chaque année les mêmes documents, de manière identique. La diversité des documents pédagogiques est alors un atout si elle permet aux enseignants de trouver des documents proches de leur contexte et enrichissant leurs « trousseaux à outils » pédagogiques.

Diversité de documents favorisée par Internet. Les échanges sont un moyen de favoriser et de diffuser cette diversité de documents. Internet fut ainsi un vecteur essentiel dans cette diffusion. Ceci est illustré par les nombreux sites Web⁴ d'enseignants ou d'associations d'enseignants

¹Revue STICEF, Sciences et Technologies de l'Information et de la Communication pour l'Éducation et la Formation : <http://sticef.univ-lemans.fr/>

²Conférence EIAH 2009 : <http://eiah2009.univ-lemans.fr/>

³ATIEF, Association des Technologies de l'Information pour l'Éducation et la Formation : <http://www.inrp.fr/atief/>

⁴Voir par exemple le guide édité par l'association du Café Pédagogique recensant de très nombreux sites pédagogiques de qualité : <http://www.cafepedagogique.org/dossiers/indisp2007/>

proposant des ressources pédagogiques.

Diversité demandant des outils de recherche efficaces. Les professeurs sont donc intéressés par trouver des documents pédagogiques qui pourraient les aider dans leur tâche. Cependant, le temps des enseignants étant limité, les recherches doivent ramener des documents réutilisables rapidement. Les documents doivent ainsi pouvoir être recherchés selon leurs contextes d'utilisation.

3 Approches classiques de l'indexation

3.1 Indexation automatique

3.1.1 Présentation

Exemple de recherche. Prenons le cas d'un enseignant recherchant des documents pédagogiques sur Internet. S'il utilise un moteur de recherche généraliste comme Google⁵, il effectuera une requête ressemblant à « devoirs troisième mathématique symétrie ». Google lui retournera alors une liste de sites Web. Voici comment ce type de moteur fonctionne.

Fonctionnement de l'indexation automatique. Google est un moteur de recherche basé sur une indexation automatique des contenus.

L'indexation sur le Web est le processus pendant lequel les contenus des pages Web sont parcourus et analysés afin d'en extraire une organisation générale. Cette organisation est ensuite employée pour répondre aux requêtes des utilisateurs. Google réalise ce processus de manière automatique, à l'aide de machines qui parcourent régulièrement les pages du Web, et créent un index associant les pages aux termes issus de leurs contenus.

Pour répondre à la requête « devoirs troisième mathématique symétrie », un moteur à indexation automatique cherche donc dans son index les pages contenant ces quatre termes et les propose à l'utilisateur. Un problème se pose alors s'il existe un millier de pages répondant à cette contrainte. Sachant que généralement les utilisateurs ne lisent que la première page de résultat, le moteur doit ordonner ses réponses selon un ou plusieurs critères.

Popularité des pages. La solution de Google pour ce problème est l'emploi d'un algorithme, nommé PageRank⁶, calculant la popularité de chaque page Web en fonction des liens hypertextuels pointant dessus. De manière simplifiée, si beaucoup de sites sur le Web renvoient vers un site A alors il y a des chances que ce site A soit intéressant. Ainsi lorsque les machines de Google parcourent le Web pour indexer les contenus, elles comptent en même temps le nombre de liens partant et arrivant sur chaque page. Ceci permet de déterminer une popularité puis un ordre entre les pages.

Pour répondre à la requête « devoirs troisième mathématique symétrie », Google retournera les pages contenant les termes requis et classées selon la popularité calculée pour chacune d'elle.

⁵En France, Google représente 91% des recherches sur Internet : <http://www.xitimonitor.com/fr-fr/barometre-des-moteurs/barometre-des-moteurs-octobre-2008/index-1-1-6-150.html>

⁶Détail du brevet sur <http://www.google.com/patents/about?id=cJU1AAAEBAJ>

L'intérêt de ce procédé d'indexation est qu'il ne nécessite aucune compréhension par les machines du contenu des pages Web et qu'il peut s'effectuer sans intervention humaine. Il permet ainsi à Google d'indexer et d'ordonner des milliards de pages Web, avec une qualité de résultats ayant conduit à son succès.

3.1.2 Limites

Utilisation excessive des moteurs. La simplicité d'utilisation de Google et de ses concurrents leur ont permis d'être utilisés par une grande majorité des internautes. Ainsi « to google » est entré dans le dictionnaire⁷ anglais avec le sens de « chercher sur Internet ». Or cet usage-réflexe entraîne des conséquences négatives dont la principale est que beaucoup d'internautes ne savent pas qu'il existe d'autres moyens de recherche que les moteurs généralistes.

Or si ces derniers sont efficaces dans certaines requêtes comme trouver une recette de cuisine ou le site Web d'une entreprise, ces outils ont plus de difficultés avec des requêtes plus floues ou dépendant d'un contexte particulier, comme la recherche de documents pédagogiques.

Limites pour déterminer les concepts. Si l'utilisateur cherche « devoirs troisième mathématique symétrie », les moteurs à base d'indexation automatique vont chercher ces quatre termes dans leurs index. À la rigueur, ils peuvent déduire que « devoirs » et « devoir » sont deux formes du même terme et inclure les pages contenant « devoir » au singulier. Néanmoins, ces moteurs vont avoir des difficultés pour repérer les documents contenant non pas « devoirs » mais « contrôle », « examen » ou « D.S. » (Devoir Surveillé). Pourtant, ces documents auraient probablement intéressé l'enseignant effectuant la requête. Mais ces moteurs repèrent des termes et non des concepts sous-jacents.

Limites pour déterminer des critères subjectifs. Les moteurs de recherche basés sur l'indexation automatique vont aussi avoir des difficultés pour déterminer des caractéristiques implicites comme la qualité pédagogique d'un document. Ces moteurs comptent les liens entre pages mais ne jugent pas directement le contenu. S'il n'y a pas assez de liens entre des documents, ces moteurs seront incapables d'ordonner statistiquement les documents.

Limites pour élaborer des vues synthétiques. Enfin, ces moteurs étant encore très liés aux termes des documents, ils offrent peu d'aide à un utilisateur ne connaissant pas un domaine et ses mots clefs. Par exemple, un parent cherchant des exercices de mathématiques pour son enfant ne saura pas forcément avec quels termes précis chercher. De plus, ces moteurs ont encore des difficultés à synthétiser des ensembles de documents. Ils ne fournissent pas de listes de concepts principaux d'un domaine à l'usage des non-experts, pour faciliter leur exploration.

Dépasser ces difficultés nécessite alors de demander à des humains de réaliser l'indexation des documents.

⁷Verbe Google dans le dictionnaire : <http://www.merriam-webster.com/dictionary/google>

3.2 Indexation par des humains

3.2.1 Présentation

Indexation traditionnelle par des humains. L'exemple le plus courant d'indexation réalisée par des humains est celle des livres dans les bibliothèques. Pour fonctionner, des experts d'un domaine créent une « carte » des concepts d'un domaine. Puis d'autres experts, bibliothécaires et documentalistes, prennent chaque document qu'ils vont indexer selon ces concepts. On dit alors qu'ils associent aux documents des métadonnées (données sur les données).

La « carte » la plus connue est probablement la Classification Décimale de Dewey couvrant toutes les connaissances humaines et employée par des milliers de bibliothèques à travers le monde. Dans ces bibliothèques, les documents se rapportant au concept de « Mathématiques » sont toujours situés sur l'étagère 510.

Avantages. Contrairement à l'indexation automatique, l'indexation par les humains permet de quitter le niveau syntaxique des documents pour offrir des vues conceptuelles sur ces derniers. Par exemple, des experts vont pouvoir concevoir une « carte » globale d'un domaine, qui servira de référence aux nouveaux utilisateurs.

Les experts humains pourront aussi évaluer des aspects peu décelables par des machines, comme la qualité pédagogique d'un document.

3.2.2 Limites

La qualité permise par l'indexation humaine possède néanmoins des défauts qui se révèlent flagrants avec le nombre et la forme des documents présents sur le Web.

Trop de documents. En effet, l'indexation par des humains, et a fortiori par des experts, est coûteuse en temps et en moyens. Or le nombre de documents créés sur le Web chaque année, dépasse largement celui des livres publiés. Dans le domaine pédagogique, cela correspond au nombre de cours et exercices diffusés sur le Web par les enseignants en comparaison aux quelques manuels publiés par les éditeurs chaque année.

Des documents trop divers. De plus, les documents sur le Web prennent des formes diverses, dûes à des processus de diffusion moins stricts que ceux des livres. Lorsqu'un enseignant dépose un cours sur son site Web, il n'a aucune validation ou contrôle qualité à requérir, aucun ISBN ne lui est donné. Son document peut être d'excellente qualité ou encore à l'état de brouillon et d'ailleurs il peut du jour au lendemain mettre à jour le fichier sans prévenir personne. Ces formes numériques peu stables et très diverses complexifient encore la tâche d'indexation par les humains.

Par conséquent, l'indexation par des experts est un processus coûteux sur le Web, difficilement applicable à grande échelle.

Or cet équilibre entre indexation automatique et indexation par des humains est remis en question par de récentes évolutions des technologies et des usages du Web.

4 Nouvelles approches

4.1 Présentation du Web Sémantique et du Web Participatif

Le Web Sémantique. Le Web Sémantique est un courant technologique reposant sur des représentations informatisées des connaissances appelées « ontologies ». Généralement, ces représentations se font à l'aide de concepts et de relations entre concepts. Le but final est de permettre des raisonnements automatisés sur ces connaissances.

Par exemple, imaginons quelques experts en pédagogie qui se réunissent pour créer une ontologie du domaine pédagogique. Ils définissent trois concepts : « document », « devoir » et « contrôle ». Ils créent ensuite des relations entre ces concepts ; ils expriment que le concept de « document » est plus générique que les deux autres, et que les concepts de « devoir » et de « contrôle » sont similaires. Dans cette ontologie, ils formalisent une partie de leurs connaissances du domaine qui pourra alors être traitée par des machines.

Pour la recherche de documents, l'existence des ontologies permet d'indexer les documents non plus selon des mots-clés mais selon des concepts reliés entre eux. Le moteur de recherche pouvant s'appuyer sur l'ontologie précédente peut donc déduire que si un utilisateur recherche un « devoir », il peut être utile de lui fournir aussi les documents indexés par « contrôle ».

Grâce aux ontologies, les machines disposent ainsi d'une connaissance du domaine pouvant améliorer les résultats des recherches.

Le Web Participatif. Nous appelons « Web Participatif » la participation croissante des internautes dans la production et la gestion de contenus sur le Web. Ce terme recouvre en grande partie les expressions « Web 2.0 » ou « Web Social ». Contrairement au Web Sémantique lié à un changement technologique, le Web Participatif est principalement une évolution des usages liée à la démocratisation d'Internet. Cette participation accrue s'illustre par des sites comme l'encyclopédie collaborative Wikipédia ou encore les sites de partage de vidéos comme YouTube.

Dans le cadre de la recherche de documents, les applications du Web Participatif permettent souvent aux internautes d'exprimer leur point de vue, généralement à l'aide d'une succession de termes choisis spontanément et appelés « tags ». Par exemple, un professeur de mathématiques trouve sur Internet une animation vidéo expliquant de manière simple le théorème de Thalès. Il taggue cette ressource avec les termes « thalès quatrième mathématique ». Les tags offrent alors une navigation à travers des ensembles hétérogènes de contenus, sans nécessiter le coût d'intervention d'experts du domaine.

4.2 Oppositions apparentes de ces nouvelles approches

En permettant le traitement du sens par les machines et en faisant participer les utilisateurs à la description des contenus, les courants du Web Sémantique et du Web Participatif renouvellent les pratiques de gestion documentaire traditionnelle. Mais ces récentes approches semblent s'opposer sur plusieurs points.

Experts ou non-experts. Dans le Web Sémantique, les ontologies sont des représentations précises et complexes d'un domaine. Leur création est donc généralement effectuée ou au moins

validée par des experts du domaine. Par la suite, les autres utilisateurs devront se conformer à cette vision.

Au contraire, le Web Participatif se fonde sur des actions peu contraintes des utilisateurs, ayant généralement tous le même statut. Les représentations d'un domaine sont alors soit issues d'une collaboration entre utilisateurs, soit agrégées à partir de leurs actions individuelles e.g. les nuages de tags.

Structuration ou flexibilité. Les ontologies sont des structures de concepts et de relations qui doivent rester cohérentes pour permettre des raisonnements automatisés. Ceci implique un lourd processus de création et d'expertise. Par conséquent, lorsque le domaine représenté évolue, les ontologies ne répercutent pas immédiatement ces évolutions.

Au contraire, les tags du Web Participatif sont moins cohérents mais plus flexibles. Si un nouveau terme devient à la mode et apparaît dans des documents, il apparaîtra aussi rapidement parmi les tags les décrivant. Il fournira ainsi une entrée pour explorer ces documents.

Les techniques d'indexation sont souvent cloisonnées voire opposées, à l'image du Web Sémantique et du Web Participatif. Pourtant l'intérêt est grand à essayer de mutualiser leurs avantages et compenser leurs limites.

4.3 Fusionner le Sémantique et le Participatif

Différentes solutions explorent la complémentarité entre les approches du Web Sémantique et du Web Participatif. L'expression « Web 3.0 » ([Mika, 2007]) est ainsi de plus en plus employée pour décrire cette hybridation. Celle-ci peut alors s'effectuer en favorisant soit les échanges entre machines et les raisonnements automatisés, soit les négociations entre utilisateurs.

Favorisant les raisonnements automatisés. Ces solutions récupèrent les tags hétérogènes des utilisateurs puis tentent de les transformer en des concepts structurés, que les machines vont pouvoir plus facilement échanger et traiter. Ces solutions s'appuient généralement sur les technologies du Web Sémantique.

Mais ces solutions correspondent peu aux communautés d'enseignants que nous avons observées : d'une part parce qu'elles s'appliquent à de très grandes communautés pour effectuer des analyses statistiques, d'autre part parce qu'elles nécessitent ponctuellement un travail sur des ontologies, difficile pour des non-informaticiens.

Favorisant les négociations entre utilisateurs. Ces solutions emploient l'expression de « Web Socio-Sémantique » ([Zacklad, 2005]), pour insister sur le lien entre interactions inter-individuelles et sémantique. Ces solutions proposent ainsi des outils pour aider les utilisateurs à percevoir et négocier leurs différents points de vue, et faciliter par la suite les recherches. Cette approche semble adaptée aux communautés en ligne d'enseignants où les membres sont au même niveau, où aucune autorité ne peut imposer aux autres sa vision du domaine.

Néanmoins, en insistant sur les échanges humains, ces solutions délaissent les possibilités de raisonnements automatisés. Elles se coupent aussi des nombreux efforts de développement et outils issus du Web Sémantique.

En conclusion, le Web Sémantique et le Web Participatif offrent des possibilités prometteuses

pour améliorer la recherche de documents. Cependant, le rapprochement de ces techniques n'en est qu'à ses débuts.

Nos travaux explorent la complémentarité de ces approches pour le contexte pédagogique. A la différence des autres travaux rencontrés, notre objectif vise à permettre des raisonnements automatisés tout en soutenant les négociations entre utilisateurs.

5 Objectifs de la thèse

Pour appréhender la complémentarité du Web Sémantique et du Web Participatif, nos travaux s'intéressent d'abord à ces approches séparément avant d'explorer leur fusion. Dans ce mémoire, nous répondrons ainsi aux questions suivantes :

Questions centrales :

- Comment les techniques du Web Sémantique peuvent-elles améliorer le partage de documents pédagogiques ?
- Comment la tendance du Web Participatif peut-elle améliorer le partage de documents pédagogiques ?
- Comment les approches du Web Sémantique et du Web Participatif, qui semblent contradictoires dans leurs usages, peuvent-elles se compléter ?

Question méthodologique :

- Comment tester les approches liées au partage de documents pour obtenir des solutions validées expérimentalement et réutilisables ?

La première question concerne principalement les ontologies et les raisonnements automatisés proposés par le Web Sémantique. Ceci soulève les sous-questions suivantes : Quelles fonctionnalités permettent-ils autour de documents pédagogiques ? Dans quels contextes éducatifs peuvent-ils être implantés ? Quelles difficultés sont rencontrées pour les créer ?

La seconde question concerne l'indexation peu coûteuse et selon plusieurs points de vue proposée par le Web Participatif. Quelles fonctionnalités facilitent la participation ? Comment les enseignants partagent-ils dans les communautés en ligne ? A quoi ressemblerait un site participatif de partage dans le domaine pédagogique ?

La troisième question concerne les moyens de dépasser les oppositions du Web Sémantique et du Web Participatif ? Comment conserver leurs avantages respectifs ? Quels modèles et fonctionnalités permettent cette fusion ?

La dernière question a guidé nos choix de développements et de terrains applicatifs tout au long de nos travaux. Selon nous, un outil de partage de documents peut difficilement être validé autrement qu'en situation réelle, c'est à dire avec de vrais utilisateurs, avec des ressources réelles et sur une période suffisante. Comment alors trouver les conditions nécessaires à cette situation ? La réponse à la question méthodologique prendra la forme d'un court retour d'expériences sur nos expérimentations.

6 Plan

Ce mémoire comprend cette introduction, un chapitre sur le contexte puis trois grandes parties, et enfin une conclusion générale et des annexes.

Contexte de l'indexation pédagogique. Le premier chapitre situe succinctement le contexte de l'indexation de ressources pédagogiques numériques. Il présente la notion de Learning Object ainsi que les différents standards associés.

Ensuite, les trois parties correspondent à trois contextes différents de recherche, liés à trois développements distincts.

Travaux liés au Web Sémantique. La première partie porte sur l'indexation de documents pédagogiques à l'aide de technologies du Web Sémantique. Nos travaux s'inscrivaient dans un projet européen nommé LUISA et les documents indexés concernaient le C2I (Certificat Informatique et Internet).

Nos travaux ont consisté en la recherche d'un terrain d'implémentation des technologies du Web Sémantique au sein de notre université, puis dans l'élaboration des **spécifications d'un outil exploitant différentes ontologies pédagogiques**.

Travaux liés au Web Participatif. La seconde partie porte sur le partage de documents dans les communautés d'enseignants.

Nos travaux ont consisté en l'**analyse des fonctionnalités des applications du Web Participatif**, puis dans la proposition d'un **modèle d'outil de partage adapté à des communautés enseignantes**. Ce modèle fut en partie implémenté dans une application en ligne, Pépi, mais que nous n'avons pas pu expérimenter par manque d'utilisateurs. Nous détaillons alors les causes de cet échec expérimental.

Travaux liés à la fusion des Webs Sémantique et Participatif. La dernière partie porte sur les moyens de profiter à la fois de la participation des utilisateurs et d'obtenir des métadonnées structurées.

Nos travaux ont consisté en la conception d'un **modèle d'indexation (Modèle d'Indexation Progressive et Multi-points de vue)**. Ce modèle fut implémenté dans un outil, SemanticScuttle, mis en production et expérimenté pendant plusieurs mois auprès d'un groupe de documentalistes. L'**outil est diffusé** (téléchargé 150 fois par mois depuis un an) et réunit une petite communauté internationale d'utilisateurs.

7 Quelques points de vocabulaire

Travailler sur l'indexation de documents nous a enseigné que bien souvent les individus ne se comprennent pas car ils mettent des sens différents sous des étiquettes identiques, et qu'ils ne s'en aperçoivent pas. Avant de détailler plus en avant nos travaux, voici nos définitions pour quelques termes.

Indexation et annotation. Dans le contexte français, l'indexation est le processus qui associe des métadonnées à des documents pour permettre leur recherche ultérieure. L'annotation

correspond plus à l'ajout de commentaires à un document comme interprétation ou enrichissement du contenu. Néanmoins, ces deux termes peuvent porter à confusion car il nous semble qu'ils possèdent des sens différents en anglais où nous parlons par exemple d'« annotation tool » pour désigner un outil d'indexation. En général, nous nous efforcerons de nous conformer aux sens français sauf pour la traduction « d'Annotation Tool » qui restera « outil d'annotation ».

Échange, partage ou mutualisation de fichiers. De manière stricte, « l'échange de fichiers » devrait désigner un envoi réciproque de fichiers entre utilisateurs et le « partage » ou la « mutualisation de fichiers » devraient désigner un travail autour d'un même document. Néanmoins, nous emploierons indifféremment ces trois termes dans le sens général de la mise à disposition d'un fichier à d'autres utilisateurs.

Documents et ressources. Nous considérerons ces termes comme synonymes.

Communauté. Comme le montre Élise Garrot dans [Garrot, 2008], il existe de nombreuses variations sur le sens de « communauté ». Dans ce document, nous entendons « communauté » au sens de « communauté de pratique » défini par Étienne Wenger ([Wenger, 1998]) : « Les communautés de pratique sont des groupes d'individus qui partagent un intérêt ou une passion pour une activité qu'ils effectuent et qui apprennent comment faire mieux en interagissant régulièrement ». Nous essaierons de nous en tenir à ce sens.

Un glossaire plus complet est disponible à la fin de ce mémoire.

1

Contexte de l'indexation de documents pédagogiques numériques

Sommaire

1.1	Nouvelles possibilités numériques	11
1.2	Indexation de documents pédagogiques numériques	12
1.2.1	L'approche par Learning Object	12
1.2.2	Les standards du E-learning	14
1.3	Conclusion	18

La facilité de création et de diffusion des documents numériques complexifie leur classification. Mais les ordinateurs et leurs traitements numériques offrent aussi de nouvelles possibilités pour la classification de documents.

Ce chapitre expose succinctement les standards et les normes tentant de résoudre ce paradoxe dans le domaine pédagogique.

1.1 Nouvelles possibilités numériques

Une longue tradition. La gestion de documents est une discipline qui est probablement née il y a quelques milliers d'années, peu après l'apparition de l'écriture sur des supports pérennes et transportables. Cette discipline s'est développée avec les bibliothèques qui devaient stocker des milliers d'ouvrages et qui utilisèrent des catalogues pour les organiser. L'invention de l'imprimerie fit augmenter fortement le nombre d'ouvrages mais les principes de leur gestion restaient assez semblables.

Bouleversée par le numérique. En revanche, l'apparition des ordinateurs et des documents numériques, ces 50 dernières années, provoqua un bouleversement dans cette science de l'organisation. David Weinberger [Weinberger, 2007] explique ce changement majeur en considérant trois niveaux d'ordonnement du monde :

1. **Le niveau des objets physiques.** « Au premier niveau d'ordre, nous organisons les choses elles-mêmes. Nous mettons les couverts dans les tiroirs, les livres sur les étagères,

les photos dans les albums. »⁸

Le principal défaut de ce niveau est que les objets physiques prennent de la place. Pour les réarranger, il faut déplacer des masses ; pour les retrouver, il faut parcourir des distances.

2. **Le niveau des métadonnées physiques.** Ce sont les catalogues, annuaires et fiches papier. « Le catalogue sépare les informations sur les objets du premier niveau des objets eux-mêmes, listant les entrées de manière alphabétique par sujet. [...] Un code sur un objet du second niveau, par exemple une fiche de catalogue, pointe vers un emplacement physique » où un objet du premier ordre est rangé.

L'avantage du second niveau est que l'organisation et la recherche sont simplifiées. Une fiche de quelques centimètres carrés peut pointer vers des objets quelles que soient leurs dimensions. Plusieurs fiches peuvent aussi pointer vers le même objet permettant une organisation selon plusieurs critères. Dans une bibliothèque, cela correspondait à un tiroir avec les fiches des ouvrages classées selon les titres, et un autre tiroir avec les fiches classées selon les auteurs.

Le principal défaut du second niveau est que les fiches papier doivent être organisées correctement pour être utiles. Lorsque la quantité d'objets augmente et lorsque les classifications s'enrichissent, il devient obligatoire de faire appel à des experts pour organiser et tenir à jour les catalogues. La gestion de ces catalogues possède ainsi un coût et une complexité limitant la quantité de fiches sur chaque objet.

3. **Le niveau des contenus et métadonnées numériques.** Ce sont par exemple des photographies ou des livres numérisés accompagnés de leurs descriptions dans un ordinateur.

Le troisième niveau possède deux avantages principaux par rapport aux niveaux précédents. Tout d'abord l'information numérique prend une place minime comparée à son équivalent physique. Un livre pourrait posséder, sans grandes difficultés, une description (ou métadonnées) de la taille de son contenu tout entier. Ensuite, une information numérique peut facilement être recopiée ou être référencée de manière illimitée, offrant alors de nouvelles possibilités de classification.

Observons comment ces « nouveaux » contenus numériques et ces « nouvelles » possibilités de classification sont considérés dans le domaine pédagogique.

1.2 Indexation de documents pédagogiques numériques

1.2.1 L'approche par Learning Object

Coût élevé d'un cours. Créer un cours de qualité pour un enseignement traditionnel, face à une classe, prend du temps et requiert des compétences. Créer un cours pour un enseignement en ligne ajoute le coût élevé de réalisation de ressources numériques. En effet, cette réalisation nécessite des compétences pour maîtriser les outils d'édition car reproduire de manière statique

⁸Dans cette partie, les passages entre guillemets sont notre traduction de [Weinberger, 2007] p.17-23.

les contenus à l'écran apporte peu d'avantages et rend même la lecture plus ardue. Ainsi, l'affichage sur écran nécessite de construire des documents plus complexes et plus interactifs que des livres linéaires. Inévitablement le coût de réalisation augmente.

Partager les contenus. Pourtant, les contenus numériques offrent une économie d'échelle car les possibilités de réplique sont illimitées et peu coûteuses. Avec l'expansion du Web, l'idée a donc émergé de capitaliser le travail en échangeant le matériel pédagogique. Les contenus pédagogiques sont devenus alors des « Learning Objects » (LO), expression qui peut se traduire par « Objets d'Apprentissage ». (Néanmoins, pour plus de facilité, nous conserverons l'expression de « Learning Object » tout au long de ce document.)

Définition des Learning Objects. En 2002, l'institut IEEE⁹ donna une définition des « Learning Objects », en même temps qu'un moyen de les décrire à l'aide du standard LOM (Learning Object Metadata, détaillé en §1.2.2) :

« Learning Objects are defined here as any entity, digital or non-digital, which can be used, re-used or referenced during technology supported learning. Examples of technology-supported learning include computer-based training systems, interactive learning environments, intelligent computer-aided instruction systems, distance learning systems, and collaborative learning environments. Examples of Learning Objects include multimedia content, instructional content, learning objectives, instructional software and software tools, and persons, organizations, or events referenced during technology supported learning. »

Une des caractéristiques de cette définition est de pouvoir recouvrir quasiment toute chose existante. Documents aussi bien que logiciels, personnes ou événements.

Dans l'usage, nous avons trouvé très peu de travaux décrivant des individus ou des événements comme un Learning Object. L'usage majoritaire que nous avons aperçu considère principalement comme Learning Object des entités numériques, dans le but de les ré-utiliser.

Cet usage est illustré par des sites de partage de Learning Objects, appelés « Learning Object Repositories » (LORs) qui peut se traduire par « entrepôts d'objets d'apprentissage ». ARIADNE¹⁰ en Europe ou MERLOT¹¹ aux États-Unis en sont deux exemples majeurs.

Comme nous l'avons dit précédemment, le premier objectif était d'économiser des coûts de production en partageant les contenus. Le second objectif, à plus long terme, était de rendre possible l'assemblage de morceaux de cours pour créer des cours individualisés. La métaphore sous-jacente reprenait l'idée des briques Légo. On « construirait » des cours de la même manière que l'on construit des maisons en assemblant les petites briques en plastique.

Vision naïve et limites des LO. Si l'idée originale peut sembler simple et séduisante, elle cache des difficultés majeures liées à la notion de contexte. Pour fonctionner, l'approche des Learning Objects en tant que brique réutilisable nécessite :

⁹IEEE, Association professionnelle pour les avancées technologiques : www.ieee.org

¹⁰ARIADNE, Alliance of Remote Instructional Authoring and Distribution Networks for Europe : www.ariadne-eu.org

¹¹MERLOT, Multimedia Educational Resources for Learning and Online Teaching : www.merlot.org

- **Des contenus au maximum décontextualisés.** Ceci est indispensable pour facilement agréger ensemble deux contenus provenant de sources différentes. Néanmoins, un contenu se situe forcément dans un contexte précis, rien que par la langue qu'il emploie ou le public qu'il vise. D'autre part, certains auteurs comme Norm Friesen dans [Friesen, 2004b] montrent le paradoxe qu'il peut y avoir entre d'un côté la volonté d'une approche décontextualisée et d'un autre côté le travail didactique qui vise justement à contextualiser le contenu pour l'apprenant.
- **Des contenus dont le contexte est décrit au maximum.** Si une ressource ne peut totalement être décontextualisée, il faut alors décrire au mieux son contexte, afin de pouvoir la déplacer d'un contexte à un autre semblable. Or, comme nous l'avons vu précédemment, la définition de Learning Object couvre des entités qui peuvent être une image ou alors un cours d'un semestre entier. Cette échelle de granularité rend très difficile l'évaluation précise de la place de chaque contenu. De manière générale, l'hétérogénéité des contenus force à produire un grand nombre de métadonnées pour faciliter leur réutilisation. Or si ajouter un titre comme métadonnée est une action simple et neutre, produire des métadonnées pédagogiques de manière générique semble beaucoup plus ardu.

Une approche plus pragmatique des Learning Objects. Comme le conclut [Dehors, 2007], l'approche par briques des Learning Objects semble irréaliste dans sa vision originale car les problèmes de contexte et d'hétérogénéité ont été largement sous-estimés.

La vision originale des LOs s'échangeant à travers le monde comme des biens matériels quelconques, a été remplacée aujourd'hui par une approche plus pragmatique où les outils proposent des ensembles de LOs homogènes, aux formats semblables et issus d'un même contexte. Par exemple, dans [Buffa et al., 2005], plusieurs livres de mathématiques sont découpés en Learning Objects mais la recombinaison est limitée aux contenus provenant d'un seul ouvrage pour assurer la cohérence.

Dans les projets que nous allons présenter dans ce mémoire, nous avons toujours cherché à respecter cette vision pragmatique des Learning Objects et de leurs contextes d'utilisation.

1.2.2 Les standards du E-learning

Les Learning Objects sont intimement liés à la notion de standardisation : standardisation de leurs interfaces informatiques pour les rendre interopérables, standardisation de leurs descriptions pour faciliter leur recherche. Nous présentons dans cette section les standards les plus illustratifs de cette tendance, en nous basant notamment sur les synthèses de Sylvain Dehors [Dehors, 2007] et d'Olivier Catteau [Catteau, 2008].

Dublin Core : des métadonnées pour tout type de ressources. Le Dublin Core Metadata Element Set est un ensemble de 15 propriétés (identifiant, créateur, format...) employées pour décrire tout type de ressources.

Le nom « Dublin » est dû à l'origine du projet ayant eu lieu lors d'un atelier à Dublin dans l'Ohio en 1995. Le terme « core » (traduisible par « noyau ») rappelle que les propriétés ont été conçues pour être les plus génériques possibles, et utilisables pour décrire un large éventail de ressources.

Ce standard est géré par le DCMI¹² qui le fait évoluer régulièrement. Les dernières évolutions tendent à rapprocher ce standard du courant du Web Sémantique, que nous aborderons dans la première partie de ce mémoire.

Tous les éléments de Dublin Core sont optionnels et peuvent être répétés et présentés dans n'importe quel ordre. Des vocabulaires contrôlés sont suggérés mais en aucun cas imposés.

Comme l'exprime [Catteau, 2008], le Dublin Core, grâce à sa simplicité, est très utilisé. Sa simplicité devient alors une faiblesse car il ne permet pas de détailler suffisamment les caractéristiques pédagogiques des contenus.

Néanmoins, par son antériorité historique et par sa généricité, il a largement inspiré un autre schéma de métadonnées, spécifiques aux objets pédagogiques : LOM.

LOM : des métadonnées pour les ressources pédagogiques. Le schéma de métadonnées LOM a pour but de guider la description de Learning Objects pour en faciliter la recherche et la ré-utilisation.

LOM regroupe les caractéristiques des Learning Objects en neuf catégories :

1. **general** : permet de décrire les aspects généraux du Learning Object comme le titre, la langue ou la date de création ;
2. **lifecycle** : permet de décrire les évolutions du Learning Object de l'état brouillon à celui de révisé ;
3. **meta-metadata** : correspond aux informations sur les métadonnées, comme par exemple qui a ajouté quelles métadonnées et à quelle date ;
4. **technical** : correspond aux informations techniques de la ressource numérique, comme son URL si elle est sur le Web, ou alors les logiciels nécessaires pour la lire si c'est un fichier de données ;
5. **educational** : correspond aux informations pédagogiques telles que le type de ressources (exercices, cours, ...), le niveau d'interactivité ou de difficulté ;
6. **rights** : permet d'exprimer les différentes licences légales liées au Learning Object ainsi que les conditions de réutilisation ;
7. **relations** : permet de décrire des liens entre Learning Objects. Par exemple, permet d'exprimer qu'un Learning Object est requis ou référencé par tel autre ;
8. **annotation** : permet d'ajouter des commentaires sur les Learning Objects. Par exemple, cette caractéristique permet à un professeur de décrire son opinion sur une ressource suite à son utilisation en classe ;
9. **classification** : cette catégorie permet de décrire le Learning Object en fonction de classifications non présentes ailleurs dans LOM. Par exemple, un Learning Object peut être décrit selon une classification de disciplines ou de compétences propre à un domaine particulier ;

Plus de 80 caractéristiques (appelées « champs ») se répartissent dans ces neuf catégories. La spécification LOM précise la manière de remplir chaque champ, par exemple avec un texte libre ou un terme sélectionné dans un vocabulaire donné. Les vocabulaires proposés par LOM ne sont pas obligatoires mais fortement recommandés pour des raisons d'interopérabilité.

¹²Dublin Core Metadata Initiative : dublincore.org/

Malgré les opportunités offertes par LOM, plusieurs défauts rendent difficile son utilisation.

Par exemple, certaines critiques portent sur le champ permettant de décrire la granularité des ressources (Aggregation Level) et qui ne propose que quatre tailles différentes. Comme nous l'avons vu précédemment, la granularité est un problème crucial pour les Learning Objects et les quelques valeurs proposées par LOM ne semblent pas y répondre.

D'autres difficultés surgissent à cause d'ensembles de valeurs incohérentes. Par exemple, le type pédagogique des ressources peut être complété par un concept abstrait comme « exercice » aussi bien qu'un objet physique comme « diapositives ».

Ces problèmes apparaissent clairement dans l'étude [Friesen, 2004a] comparant les métadonnées de Learning Objects issus de plusieurs dépôts à travers le monde. Cette étude démontra ainsi les difficultés à utiliser le LOM original, notamment pour les aspects pédagogiques.

Néanmoins, les défauts des vocabulaires recommandés par LOM ne suppriment pas l'intérêt de la plupart des caractéristiques proposées. Nous verrons notamment dans le projet européen LUISA comment LOM fut enrichi à l'aide de vocabulaires spécifiques (issus d'une ontologie de compétences par exemple).

Profils d'application de LOM : adapter LOM pour des contextes précis. Afin de continuer à utiliser LOM, tout en réduisant son aspect trop générique, certaines initiatives ont mis au point des profils d'application de LOM.

« Un profil d'application est un ensemble d'éléments choisis parmi un ou plusieurs schémas de métadonnées et combinés dans un schéma composite. Son objet est d'adapter des schémas existants pour constituer un ensemble taillé à la mesure des exigences fonctionnelles d'une application particulière, tout en restant interopérable avec les schémas d'origine » ([Duval et al., 2002]).

Pour préciser encore le contexte, un profil d'application peut lui-même voir naître des sous-profils d'application. Par exemple, LOMFR¹³ est le profil d'application français de LOM et SupLOMFR¹⁴ est le profil d'application du LOMFR destiné aux institutions françaises de l'enseignement supérieur.

ISO MLR : Future norme pour les ressources pédagogiques. Lorsqu'un standard a suffisamment été diffusé et utilisé, lorsqu'il arrive à maturation, un processus de normalisation est enclenché. Au niveau mondial, ce processus est mené par l'Organisme de Normalisation International (ISO). L'ISO a mis en place un sous-comité technique dédié à l'apprentissage, l'éducation et l'enseignement. Celui-ci travaille à la mise en place de la future norme ISO/IEC 19788 MLR¹⁵ pour la description de ressources d'apprentissage. Adaptabilité culturelle, capacité d'extension, et interopérabilité sont au cœur des préoccupations du groupe de travail qui s'attache à ré-exploiter le Dublin Core et le LOM de manière à faciliter le passage à la future norme.

SCORM : Standard pour des paquets pédagogiques exécutables. L'initiative la plus importante en faveur d'une approche Légo des Learning Objects est la mise en place du standard

¹³Lom-fr : <http://www.lom-fr.fr>

¹⁴SupLOMFR : <https://suplomfr.supelec.fr/mediawiki/index.php/SupLOMFR>

¹⁵Metadata for Learning Resources

SCORM (Sharable Content Object Reference Model). Ceci est illustré notamment par deux aspects du standard, à savoir :

- une description précise du format d'un contenu pédagogique nommé CAM (Content Aggregation Model). Celui-ci définit les informations utilisées pour décrire les Learning Objects (à l'aide notamment de LOM) et leur agrégation dans des SCO (Shareable Course Objects). Il spécifie aussi comment les objets doivent être stockés physiquement dans des fichiers.
- un environnement d'exécution nommé RTE (Run Time Environment). Le standard décrit ainsi comment une plateforme pédagogique¹⁶ doit interpréter les métadonnées incluses dans les SCO. Les paquets de cours SCO sont « exécutés » sur les plateformes pédagogiques en ligne de la même manière qu'un programme informatique est exécuté sur un ordinateur.

Ce standard s'intéresse avant tout à l'interopérabilité technique entre Learning Objects numériques. Aujourd'hui, la plupart des plateformes pédagogiques en ligne sont conformes au standard SCORM. Néanmoins, il laisse de côté les aspects plus pédagogiques reposant sur les métadonnées LOM. Ainsi, des auteurs comme Norm Friesen ([Friesen, 2004b]) reprochent à cette vision d'être trop proche d'une vision industrielle ou militaire de l'enseignement (le consortium aérien AICC et le ministère américain de la défense soutiennent fortement ce standard) et d'inciter à réduire la diversité pédagogique des formes d'apprentissage.

IMS LD : formaliser les scénarios pédagogiques. L'objectif du standard IMS Learning Design (LD) est de représenter de manière formelle un scénario pédagogique¹⁷ à l'aide de concepts et relations prédéfinies.

Plusieurs avantages sont prévus : LD découpe un apprentissage en activités, ce qui semble une approche plus pertinente qu'une approche orientée par les contenus. LD serait un langage permettant aux praticiens de discuter de scénarios pédagogiques. L'aspect formel permettrait aux machines d'automatiser l'exécution de ces scénarios. Suivant l'approche par briques, les scénarios pédagogiques pourraient être découpés en morceaux, indépendants de l'activité complète, et être échangés et réutilisés comme des Learning Objects.

Les auteurs de LD ([Koper and Tattersall, 2005]) disent que ce standard est pédagogiquement neutre. Il devrait supporter des formes d'apprentissage aussi bien behavioriste que constructiviste.

Cependant la force apparente du langage doit être minorée, pour l'instant, par le manque de cas d'utilisation réelle, exploitant des traitements automatiques sur des scénarios spécifiés par LD.

Dans nos travaux, nous n'avons pas exploré plus en avant ce standard. De manière générale, nous avons plutôt travaillé avec la « matière » employée et échangée par la majorité des enseignants aujourd'hui, comme des documents de cours ou des feuilles d'exercices.

¹⁶Une plateforme pédagogique en ligne est une application sur le Web, où généralement les enseignants peuvent déposer des cours et proposer des activités pédagogiques aux apprenants qui s'y connectent. En anglais, ces plateformes sont appelées Learning (Content) Management System ou L(C)MS.

¹⁷IMS LD : <http://www.imsglobal.org/learningdesign/>

1.3 Conclusion

L'évolution des contenus vers des formats numériques a commencé à être prise en compte dans le domaine pédagogique comme le montre les standards présentés ci-dessus. Ces derniers ont notamment l'intérêt de préciser et de nommer des concepts pédagogiques souvent implicites.

Cependant ces standards rencontrent trois difficultés majeures.

La première difficulté est due au terrain sensible de l'éducation, où les contraintes pédagogiques induites par les formats sont très surveillées. De plus, les enjeux politiques et économiques sous-jacents ne facilitent pas les négociations au sein des processus de standardisation.

La seconde difficulté provient d'un manque important de maturité des standards. Si l'interopérabilité technique est un thème en cours de résolution, les avancées sur les descriptions pédagogiques (certains parlent d'interopérabilité sémantique) sont encore très faibles.

La troisième difficulté provient d'un manque de maturité dans l'application de ces standards. LOM est en définitive un standard laissant beaucoup de libertés dans son application. Or, son interprétation majoritaire a soutenu une vision naïve des contenus pédagogiques numériques sous la forme de briques Légo interchangeables. Cette vision nous semble plus soutenue par des informaticiens modélisateurs que par les retours des enseignants et praticiens de terrain.

Toutes ces difficultés montrent la nécessité de faire encore évoluer les standards. Les travaux que nous avons effectués dans le projet européen LUISA (§3) offrent par exemple des propositions d'enrichissements de LOM avec des ontologies de compétences ou de disciplines ou encore la prise en compte des avis des utilisateurs pour cerner des contextes favorables d'utilisation.

Pourtant, peut-être que les organismes de standardisation n'ont pas encore perçu totalement les conséquences de la numérisation des contenus et des échanges, à savoir la modification du rapport prescripteur-utilisateur. Ainsi, dans sa modélisation des trois niveaux d'ordonnement (physique, papier et numérique – voir §1.1), David Weinberger pense que le principal changement dû à la démocratisation du numérique est le rapport à l'autorité :

Les contenus et métadonnées numériques « ne font pas seulement sortir l'information de ses traditionnelles catégories organisationnelles mais suppriment aussi l'autorité implicite accordée par la publication dans le monde papier. [...] Nous avons des industries et des institutions entières bâties sur le fait que le niveau d'ordonnement lié au papier limite sévèrement la manière dont les choses sont organisées. Les musées, les cursus scolaires, les journaux, l'industrie du voyage et les programmes de télévision sont tous basés sur l'hypothèse que dans le monde du second niveau, nous avons besoin d'experts pour parcourir les informations, les idées et les connaissances et les ranger soigneusement. » Le troisième niveau « modifie la manière dont nous pensons que le monde s'organise et, peut-être plus important encore, qui a autorité pour nous le dire. »

Si les enseignants ne savent pas décrire un scénario pédagogique à l'aide d'IMS-LD, ceci ne les empêche nullement d'échanger des documents comme nous le verrons par exemple dans l'étude d'une communauté d'enseignants en génie mécanique pour le projet Pépi (§5). De plus, le développement des sites participatifs sur le Web et l'explosion des tags (mots-clefs choisis libre-

ment) pour décrire des contenus indiquent une évolution importante des processus d'indexation.

Les internautes en général, dont les enseignants et apprenants font partie, ne sont peut-être pas prêts à attendre que des instituts de standardisation leur dictent la manière de percevoir leurs documents et activités, ou le vocabulaire à employer pour les décrire. Quels que soient leurs niveaux (local, national ou international), les groupes de standardisation vont devoir prendre en compte ces nouvelles pratiques collectives. Le projet SemanticScuttle (§7) que nous présenterons à la fin de ce mémoire, tend à explorer cette nouvelle forme d'indexation des documents mêlant les descriptions d'une autorité régulatrice aux descriptions libres des utilisateurs.

Première partie

Web Sémantique, ontologies et indexation pédagogique

Dans cette partie, nous explorerons l'approche du Web Sémantique appliquée à l'indexation de documents pédagogiques.

Après un survol rapide des technologies sémantiques, nous étudierons une application concrète au contexte pédagogique, à travers le projet LUISA.

2

État de l'art

Sommaire

2.1	Web Sémantique	23
2.1.1	Présentation générale	23
2.1.2	Les outils du Web Sémantique	27
2.2	Web Sémantique appliqué au domaine pédagogique	31
2.3	Conclusion	33

2.1 Web Sémantique

2.1.1 Présentation générale

Le consortium W3C. D'un point de vue technique, le développement du Web est mené par un consortium international nommé le W3C (World Wide Web Consortium). Ce consortium regroupe des membres d'universités ou de grandes entreprises, discutant et orientant les évolutions techniques du Web. Ce consortium propose notamment des langages et des architectures informatiques. L'exemple le plus connu est probablement le langage HTML employé pour créer une page Web.

Ce consortium ne décide pas de toutes les évolutions techniques. Par exemple, des développements majeurs ont eu lieu en dehors de leurs standards, comme les plugins Flash développés par une seule entreprise. Néanmoins, le W3C est un organisme important et écouté.

Parmi ses membres, le plus connu est sûrement Tim Berners-Lee, inventeur du Web. En 2001, dans un article [Berners-Lee et al., 2001] fondateur, Tim Berners-Lee et ses collaborateurs présentent leur vision du futur Web, appelée « Web Sémantique »¹⁸, c'est-à-dire un Web dont le contenu est compréhensible par les machines.

Des pages Web pour les humains. Actuellement, une page Web est réalisée pour que le contenu soit compris par les humains alors que les machines n'en perçoivent que la forme. La Fig. 2.1 illustre cette différence. Un humain lit un texte qui évoque des concepts tandis qu'une machine perçoit juste des mots, successions de caractères alphabétiques séparées par des espaces.

¹⁸Dans ce mémoire, chaque fois que nous parlerons de Web Sémantique, nous ferons référence au Web Sémantique selon le W3C.

Une machine peut réaliser des traitements simples sur ces mots. Par exemple, reconnaître que les *mot2* et *mot10* sont identiques (i.e. « de »). Cette capacité suffit pour effectuer une indexation automatique, à la Google (§3.1).

Néanmoins, cette reconnaissance syntaxique est insuffisante quand un ensemble de mots désigne un objet du monde réel ou quand un mot peut posséder plusieurs sens. Par exemple, toujours dans la Fig. 2.1, un humain comprend aisément que les mots « Gérard » et « Dutilleul » désignent un être humain précis alors qu'une machine aura plus de difficultés si elle ne possède pas l'information que « Gérard » est un prénom français courant. De plus, le mot « cinquième », isolé, est ambigu et une machine peut difficilement savoir qu'il désigne un niveau scolaire.

En conclusion, une machine est limitée pour interpréter des contenus textuels. Il lui manque des connaissances extérieures aux contenus et des références non-ambiguës vers ces connaissances.

Leçon de mathématique		<i>mot1 mot2 mot3</i>
Par Gérard Dutilleul		<i>mot4 mot5 mot6</i>
Cinquième D		<i>mot7 mot8</i>
Théorème de Pythagore		<i>mot9 mot10 mot11</i>

FIG. 2.1 – Extrait de contenu pédagogique. Ce que lit un humain, à gauche, et ce que perçoit une machine, à droite.

Ontologies, classes et entités. Dans la vision du W3C, les traitements des machines peuvent être soutenus grâce à des représentations de connaissances informatisées : les ontologies.

Une ontologie est composée de classes et d'entités. Les classes sont des ensembles d'entités possédant des caractéristiques similaires. Une classe pour les machines est un peu l'équivalent d'un concept pour les humains.

Par exemple, dans la Fig. 2.1, la première ligne pourrait être interprétée comme le nom de deux entités « leçon » et « mathématique » appartenant respectivement à des classes « Ressources pédagogiques » et « Disciplines scolaires », appartenant elles-mêmes à une ontologie du domaine pédagogique.

La première caractéristique des ontologies est que les classes et les entités possèdent obligatoirement une dénomination pour les désigner (URI). Cette dénomination prend généralement la forme d'une adresse Web. Alors qu'en langage naturel, un mot comme « cinquième » peut désigner aussi bien une vitesse de voiture qu'un niveau scolaire, dans le Web Sémantique, chacun de ces sens correspondrait à une classe désignable par une dénomination/adresse. Par exemple : « <http://UnConstructeurAutomobile.com/ontoVoiture#cinquième> » ou « <http://MinistereEducation.org/ontoEducation#cinquième> »

Une seconde caractéristique des ontologies est que les classes et les entités peuvent posséder des relations entre elles et sur lesquelles les machines pourront effectuer des raisonnements automatisés. Par exemple, il pourrait exister une relation « enseigne » entre une classe « Humain » et une classe « Discipline scolaire ». Cette relation permettrait à une machine de répondre à « quelle est la discipline enseignée par Bernard Dutilleul ? ». Une machine peut aussi inférer de nouvelles connaissances, c'est-à-dire déduire par exemple que si « Bernard Dutilleul » enseigne, alors « Bernard Dutilleul » est un « Enseignant ». Par la suite, d'autres raisonnements pourront s'appuyer sur cette nouvelle connaissance et ainsi de suite. Les relations appartiennent aussi aux ontologies et possèdent une dénomination unique.

Le groupe de travail sur le Web Sémantique¹⁹ propose donc des pages Web employant des ontologies, des classes, des entités et des relations pour être comprises par des machines.

Reprenons l'exemple précédent, en considérant cette fois qu'un marquage sémantique a été effectué par l'auteur de la page. En notant les dénominations de manière simplifiée, nous obtenons une illustration de la compréhension des machines en Fig. 2.2. Dans cet exemple, il y a trois ontologies du domaine de l'éducation (*ontoEducation*), des fonctionnaires (*ontoFonctionnaire*) et du domaine mathématique (*ontoMathématique*). Grâce à un marquage sémantique, la machine peut comprendre que le document est de type « leçon » et concerne la discipline « mathématique » au sens de l'ontologie « *ontoEducation* ». L'auteur est un individu précis représenté dans l'ontologie des fonctionnaires. Le niveau scolaire du document est celui de la cinquième au sens de l'ontologie éducative. Enfin, la connaissance abordée est celle du « Théorème de Pythagore » définie dans une ontologie des mathématiques dans l'enseignement français.

Cet exemple illustre aussi le côté ouvert et distribué du Web Sémantique ; c'est-à-dire que n'importe qui a le droit d'exprimer sémantiquement de nouvelles connaissances, et ceci en employant différentes ontologies. Ainsi, l'auteur du document a employé trois ontologies pour exprimer que « Gérard Dutilleul » a réalisé un document en « Mathématique ».

Leçon de mathématique	<code><ontoEducation#typeDocument = ontoEducation#leçon></code>
	<code><ontoEducation#discipline = ontoEducation#mathématique></code>
Par Gérard Dutilleul	<code><ontoEducation#auteur = ontoFonctionnaire#GérardDutilleul></code>
Cinquième D	<code><ontoEducation#niveau = ontoEducation#cinquième></code>
Théorème de Pythagore	<code><ontoEducation#connaissance = ontoMathématique#ThéorèmePythagore></code>

FIG. 2.2 – Extrait de contenu pédagogique, dans l'esprit du Web Sémantique. Ce que lit un humain, à gauche, et ce que perçoit une machine, à droite, grâce aux ontologies.

Agenda de développement. Pour que le Web classique devienne le Web Sémantique, le W3C a proposé un agenda sous la forme de composants techniques à développer : le fameux « gâteau sémantique » (*semantic cake*) (Fig. 2.3). Chaque composant se sert des précédents pour fonctionner.

Les couches inférieures (URI, UNICODE, XML, RDF et RDF-S) ont été standardisées et sont largement employées sur le Web. Elles concernent des aspects syntaxiques, c'est-à-dire les formats à employer pour écrire, nommer et échanger des données de manière commune.

Actuellement, les efforts de développement portent sur les couches médianes (SPARQL, OWL et RIF/SWRL). Les difficultés portent notamment sur les ontologies coûteuses à concevoir et à rendre cohérentes entre elles.

Enfin, les couches supérieures à venir porteront sur des raisonnements améliorés. C'est-à-dire portant l'attention sur la validité des connaissances issues de leur prouvabilité mais aussi de la confiance dans leur origine. Par exemple, ces travaux servent à déterminer en quoi certaines connaissances sont vraies dans un lieu, à une époque ou pour certaines personnes, mais pas dans un autre contexte.

¹⁹Groupe du W3C sur le Web Sémantique, <http://www.w3.org/2001/sw/>

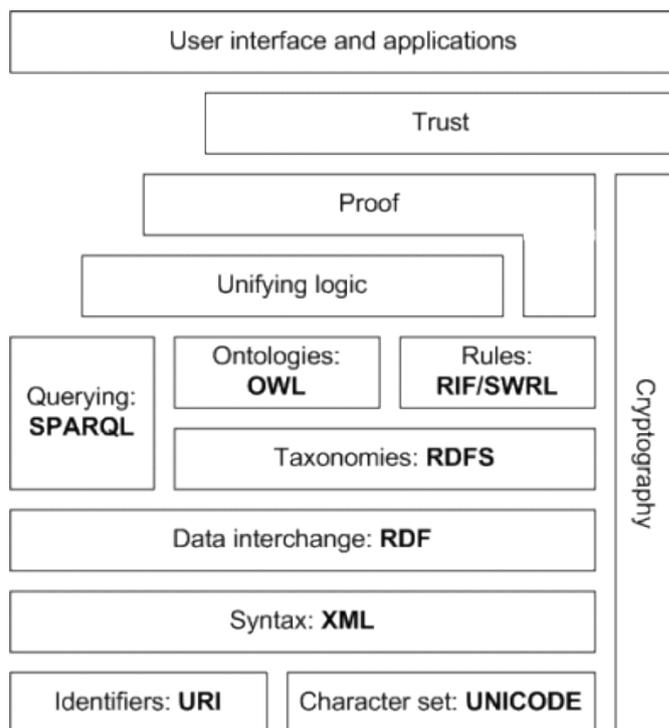


FIG. 2.3 – Le « gâteau sémantique » est une forme d’agenda de développement du Web Sémantique. La couche inférieure (de URI à RDF-S) est stabilisée tandis que les efforts actuels portent sur la couche médiane (SPARQL, OWL, RIF/SWRL).

RDF, OWL, WSML et SPARQL. Le « gâteau sémantique » présente les différentes technologies déployées pour atteindre le Web Sémantique. Dans nos travaux, nous nous intéresserons à RDF, OWL et SPARQL.

Le RDF permet de structurer de l’information sous la forme de triplets « sujet », « prédicat », « objet ». Par exemple, un document sémantique peut contenir les triplets « Bernard Dutilleul » « habite » « Paris », et « Bernard Dutilleul » « travaille à » « Cachan ». Chaque « sujet », « prédicat » et « objet » est désigné par une dénomination unique (URI). L’ensemble des triplets forme un graphe RDF.

Le langage OWL est une couche supérieure à RDF. Cela signifie qu’il réemploie le langage RDF mais en plus ajoute certaines fonctions. Ces fonctions permettent de structurer l’information selon des classes, des entités et des relations pour faciliter la modélisation d’un domaine. Ces fonctions offrent ainsi la possibilité de structurer l’information en ontologies, sur lesquelles vont pouvoir être plus facilement appliqués des raisonnements.

Dans le projet européen que nous présenterons par la suite (§ 3), le langage de modélisation choisi par les partenaires techniques n’est pas OWL mais WSML. Les différences entre ces deux langages sont faibles. Le choix porta sur WSML pour des raisons d’intégration avec des outils pré-existants.

SPARQL est un langage de requête sur les graphes RDF et donc sur les ontologies OWL. Ce langage est fondé sur l’union de motifs de graphes. C’est-à-dire que la requête est formée

d'un graphe RDF où certaines parties sont remplacées par des variables. Le langage répond à la requête en fournissant tous les graphes RDF correspondant. Par exemple, soit la requête « Bernard Dutilleul » « habite » « ?x » où « ?x » est la variable. Le système répondrait « Paris » si l'on suppose l'existence du triplet « Bernard Dutilleul » « habite » « Paris ».

2.1.2 Les outils du Web Sémantique

Cette section reprend en partie la synthèse issue de la thèse de Sylvain Dehors ([Dehors, 2007]).

Cinq catégories d'outils sont généralement employées dans les technologies du Web Sémantique ([Dehors, 2007]) :

1. des outils pour formaliser les connaissances dans les ontologies,
2. des outils pour exprimer ces connaissances dans les contenus,
3. des outils pour visualiser ces connaissances,
4. des outils pour rechercher parmi ces connaissances,
5. des outils pour naviguer parmi les connaissances et les ressources.

2.1.2.1 Éditeurs d'ontologies

Pour être échangées et manipulées, les connaissances doivent être formalisées grâce à des ontologies partagées. Les éditeurs d'ontologies en OWL, comme Protégé²⁰ ou SWOOP²¹ offrent des interfaces pour créer des ontologies puis générer les fichiers OWL.

Ces éditeurs permettent de naviguer dans la structure souvent à l'aide d'arbres hiérarchiques comme l'illustre la copie d'écran de l'éditeur Protégé 2.4. La structure hiérarchique des classes dans l'ontologie est visible à gauche. Les différentes caractéristiques des classes sélectionnées sont éditables sur la droite. Cette application peut être personnalisée au moyen de plug-ins pour créer des formulaires ou inclure par exemple des raisonneurs.

Un autre exemple est l'éditeur SWOOP de l'université du Maryland. Il offre les mêmes fonctionnalités de base mais est une application plus légère inspirée des interfaces Web (présence d'un bouton « retour », hyperliens). Ces deux outils requièrent une bonne connaissance du formalisme OWL et s'adressent plus particulièrement à des experts parfois appelés « ingénieurs de la connaissance » (knowledge engineers). Quand les ontologies doivent être construites directement par certaines catégories d'utilisateurs, comme les experts du domaine, des interfaces dédiées sont nécessaires.

Dans le projet européen LUISA, nous avons utilisé l'outil WSMO Studio²² qui implémente différents aspects de l'architecture WSMO. WSMO Studio est développé comme extension de l'environnement de programmation Eclipse²³ et propose des éléments d'interface similaires. L'éditeur d'ontologie inclus dans WSMO Studio enregistre les ontologies au format WSML, propre à l'architecture WSMO.

²⁰Éditeur d'ontologies Protégé, <http://protege.stanford.edu/>

²¹Éditeur d'ontologies SWOOP, <http://code.google.com/p/swoop/>

²²WSMO Studio, « open source Semantic Web Service and Semantic Business Process modelling environment » : www.wsmostudio.org

²³Eclipse : www.eclipse.org

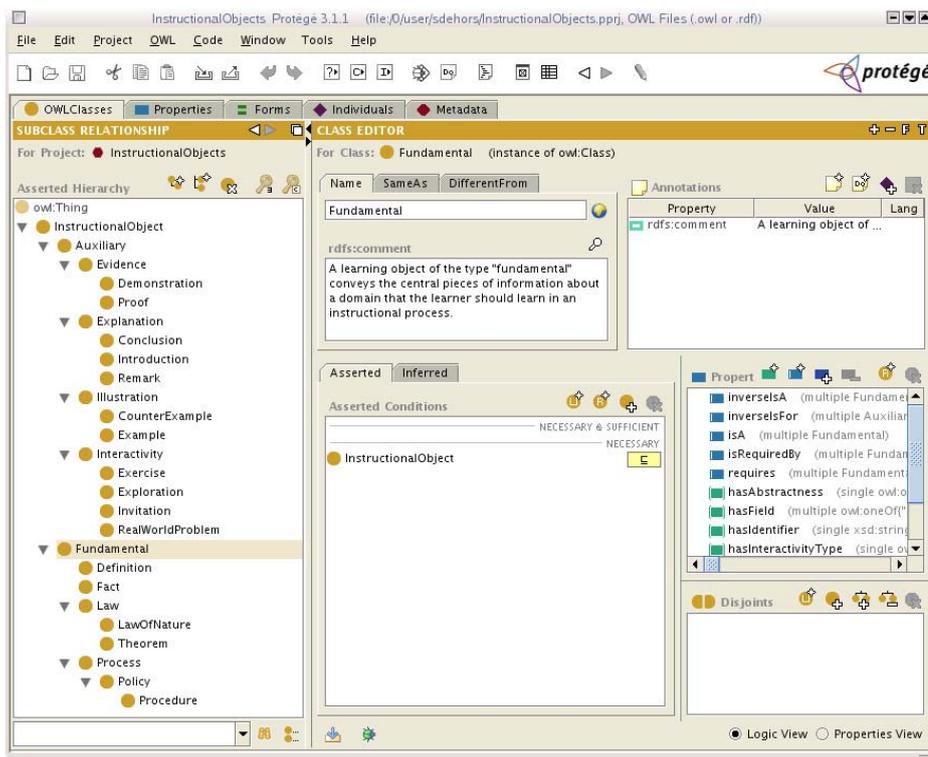


FIG. 2.4 – Copie d'écran de l'interface du logiciel Protégé

2.1.2.2 Outils d'annotation

Dans le Web Sémantique, toute ressource identifiée par une adresse unique (URI) peut être annotée. Afin d'effectuer l'annotation de ces ressources, il existe des outils spécifiques : les outils d'annotation. Une revue de ces outils pour le domaine du Knowledge management en général est proposé dans [Uren et al., 2006]. Sept points y sont identifiés :

1. L'éditeur doit fournir un format standard en entrée et en sortie.
2. Il doit être intégré dans l'environnement où les utilisateurs manipulent les documents et dans le processus collaboratif global.
3. Il doit supporter des références vers plusieurs ontologies.
4. Plusieurs formats de documents doivent être supportés.
5. L'évolution des documents doit être prise en compte.
6. L'annotation peut être stockée soit dans le document lui-même, soit dans un conteneur distinct.
7. Une aide à l'automatisation doit être fournie.

Dans le cadre du projet européen, nous avons employé l'outil d'annotation en ligne des partenaires, nommé SHAME dans sa version originale. Cet éditeur, assez récent, répond aux trois premiers points mais non aux quatre derniers. En revanche, la particularité de SHAME est d'essayer d'être générique. Cela signifie qu'un administrateur peut créer des interfaces d'annotation où il choisira quelles métadonnées seront à remplir (provenant de n'importe quel schéma) et sous quelle forme (texte libre, liste déroulante avec des valeurs du schéma, ...). L'outil est détaillé dans la section du projet LUISA (§3).

2.1.2.3 Visualisation

Les ontologies sont souvent représentées par des structures de graphes. Des outils comme IsaViz²⁴ ou des plugins pour Protégé, permettent de visualiser les graphes RDF.

WSMO Studio, employé dans le projet LUISA, nous a fourni aussi une visualisation des ontologies (Fig. 2.5).

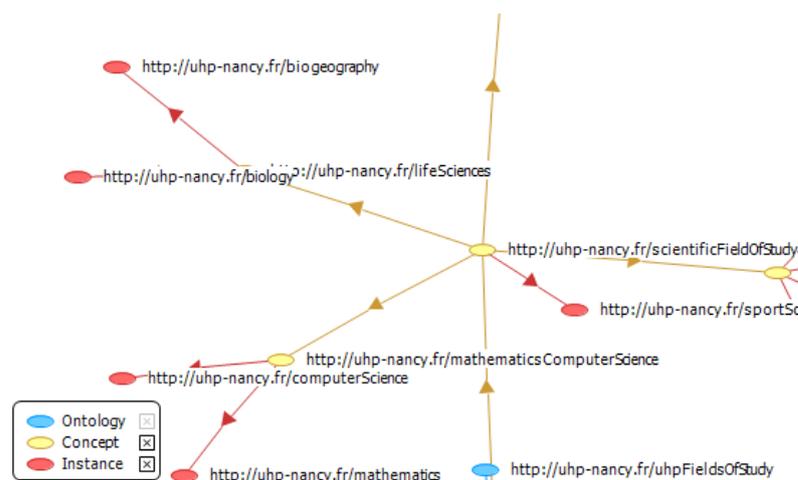


FIG. 2.5 – Visualisation de l'ontologie de discipline du projet LUISA dans WSMO Studio

Néanmoins, si ces représentations conviennent pour de petites ontologies, elles deviennent rapidement inutilisables quand le nombre de concepts et de relations augmente. De plus, le RDF ne spécifie pas d'emplacements visuels pour les nœuds ou les arcs. Les outils s'appuient alors sur des algorithmes automatiques pour le placement des nœuds et des arcs en deux dimensions. Or, comme expliqué dans la théorie des cartes conceptuelles [Novak and Canas, 2008], la disposition physique des concepts exprime une forme de sémantique.

Ces difficultés d'affichage des ontologies posent un problème important sachant que les ontologies ont pour objectif d'être partagées et donc comprises par plusieurs acteurs. Ces difficultés furent notamment abordées dans [Gandon, 2005] pour trouver automatiquement le meilleur affichage possible et avec le standard Fresnel²⁵ du W3C pour fournir un vocabulaire d'affichage pour RDF.

2.1.2.4 Moteurs de recherche

Les éditeurs permettent de formaliser des connaissances et les outils d'annotation permettent de relier les ressources à ces connaissances. Les moteurs de recherche sémantique vont pouvoir fouiller ces ressources à travers les connaissances qu'elles contiennent.

Une majorité des moteurs utilisent les ontologies en OWL et les annotations en RDF pour répondre aux requêtes exprimées en SPARQL par exemple :

²⁴IsaViz : un outil visuel d'édition de RDF, <http://www.w3.org/2001/11/IsaViz/>

²⁵Fresnel, « Display Vocabulary for RDF » : <http://www.w3.org/2005/04/fresnel\discretionary{-}{-}{-}info/>

- Ontobroker²⁶ devenu un produit commercial Ontoprise, et Kaon [Oberle et al., 2004].
- Le moteur de recherche sémantique Corese [Corby et al., 2004], fondé en interne sur des graphes conceptuels. Il fut intégré dans une plateforme Web complète nommée sewese (SEmantic WEb SErver).
- Le framework Jena²⁷ développé par HP
- Sesame²⁸
- Triple²⁹

Au delà des moteurs classiques, les moteurs sémantiques vont parfois pouvoir acquérir des connaissances qu'ils ne possèdent pas encore et qui ne sont pas explicitées. Par exemple, par un mécanisme d'inférence, si un moteur sait grâce à des annotations qu'une ressource est localisée à l'université de Nancy, et s'il sait que Nancy est une ville française grâce à une ontologie géographique, alors il pourra déduire/produire la connaissance que la ressource est française.

Dans le projet LUISA, le moteur de recherche sera un plugin à une plateforme pédagogique en ligne. La requête pour récupérer des ressources est envoyée et traitée par un ensemble de Web Services Sémantiques.

2.1.2.5 Navigateur sémantique

Un autre moyen d'exploiter les connaissances formalisées et les documents annotés est de suivre l'analogie avec le Web actuel et d'offrir des navigateurs sur les connaissances disponibles. Cela s'appelle la « navigation conceptuelle » (conceptual browsing).

La navigation conceptuelle présentée dans [Naeve, 2001] consiste dans la visualisation de graphes représentant la connaissance conceptuelle et donnant accès aux documents liés. Par exemple, la Fig. 2.6 montre l'interface du navigateur conceptuel Conzilla. Sur la partie de gauche, la représentation conceptuelle est visible employant un format proche de l'UML. Sur la droite, le contenu du document lié au concept est affiché.

Magpie [Dzbor et al., 2005] propose une autre forme de navigation. Magpie est un plugin pour le navigateur InternetExplorer. Au lieu de visualiser une structure de concepts, le navigateur se fonde sur une navigation hypertextuelle classique de pages Web améliorée par des annotations automatiques. Lors de la navigation sur le Web, le plugin surligne les termes des pages faisant référence à des concepts des ontologies que le plugin connaît. La Fig. 2.7 montre une page Web où les termes reliés à une ontologie météorologique sont surlignés. En cliquant avec le bouton droit sur ces termes, le plugin propose un menu contextuel permettant de naviguer parmi les concepts en relation.

Après ce survol du Web Sémantique et de ses outils, nous allons voir comment ces technologies s'appliquent au domaine pédagogique en général et à l'indexation de documents pédagogiques en particulier.

²⁶Ontobroker, <http://ontobroker.semanticweb.org/>

²⁷Jena, un framework pour le Web Sémantique en Java, <http://jena.sourceforge.net/>

²⁸Sesame, RDF Schema Querying and Storage : <http://www.openrdf.org/>

²⁹Triple, RDF query, inference, and transformation language for the Semantic Web : triple.semanticweb.org/

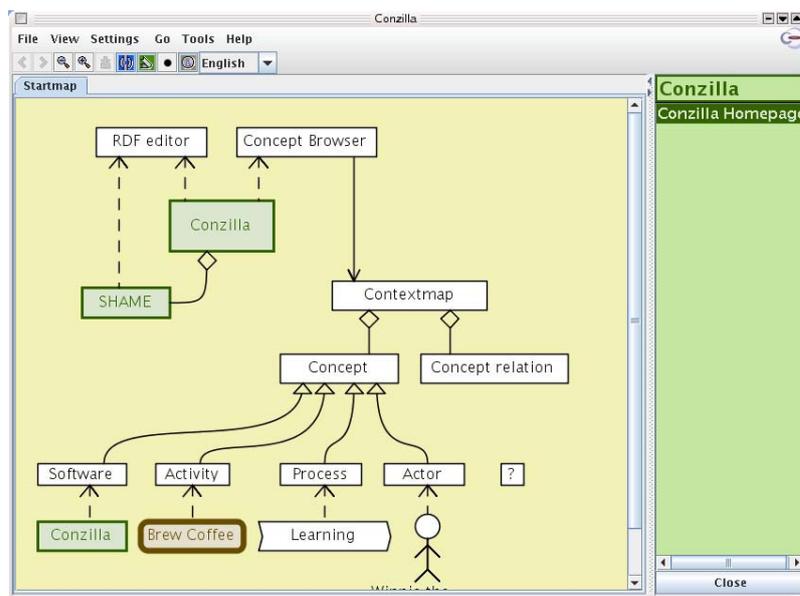


FIG. 2.6 – Copie d'écran de l'interface du logiciel Conzilla, navigateur conceptuel

2.2 Web Sémantique appliqué au domaine pédagogique

Le projet du Web Sémantique propose des solutions, dont beaucoup seraient intéressantes dans le domaine des EIAH.

Dépasser LOM. En 2001, [Stojanovic et al., 2001] propose d'employer le Web Sémantique pour effectuer des recherches et retrouver des ressources pédagogiques sur le Web. Mettre en oeuvre des ontologies résoudrait les problèmes de variabilité dus aux champs libres du standard LOM (§1.2.1). Le contexte d'apprentissage serait aussi plus précisément défini. La structure du matériel pédagogique peut être comprise par les machines si celle-ci fait référence à des ontologies. Fondés sur ces ontologies, la réutilisation de cours et leurs personnalisations sont présentées comme un apport direct du Web Sémantique.

Modéliser les théories « d'instructionnel design ». Une des ontologies les plus développées dans le domaine pédagogique est OMNIBUS ([Mizoguchi et al., 2007]) composée de plus de 1000 concepts autour des théories pédagogiques. Cette ontologie permet notamment de soutenir la création de scénarios pédagogiques pour la formation en ligne.

Agents artificiels. Selon [Devedzic, 2004], l'introduction de métadonnées sémantiques dans le matériel pédagogique permet de fournir un environnement interactif aux étudiants pour chercher du matériel et un cadre aux agents pédagogiques artificiels pour collaborer et produire du matériel de cours. L'auteur met en avant l'importance des standards et souligne les difficultés pour créer des ontologies puis pour annoter avec.

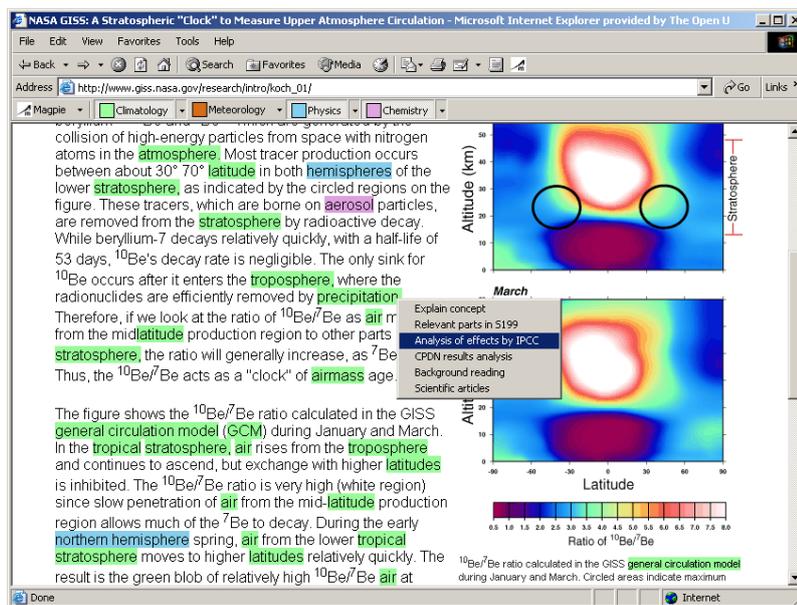


FIG. 2.7 – Copie d'écran illustrant le plugin Magpie : les termes de la page reliés à des concepts sont surlignés.

Web Services Sémantiques. [Aroyo and Dicheva, 2004] développent les questions d'interopérabilité. Les ontologies et les architectures orientées service sont vues comme les solutions permettant l'interopérabilité.

Différents points de vues sur les ressources. Une position originale est défendue dans [Naevé, 2001] et [Downes, 2004]. Les métadonnées sont vues comme des vues subjectives sur les ressources. L'aspect distribué du Web Sémantique permet alors aux systèmes de combiner différents points de vue et de fournir de meilleurs résultats de recherche qu'avec des schémas traditionnels de métadonnées. Dans ce courant, [McCalla, 2004] propose de récupérer ces métadonnées directement auprès des utilisateurs des LO, en enregistrant l'activité des membres des plateformes d'apprentissage.

Personnalisation de cours. Plusieurs travaux comme [Dehors, 2007] ou [Yessad et al., 2008] s'intéressent plus particulièrement à la génération automatique de cours ainsi qu'à leur personnalisation. Les ontologies permettent de décrire le domaine enseigné et les règles gérant le parcours d'apprentissage.

Sémantique pour les communautés. Le projet européen Palette³⁰ (février 2006-janvier 2009) vise à comprendre et à améliorer l'apprentissage dans les Communautés de Pratique. Pour atteindre ce but, des services, notamment sémantiques, furent développés et expérimentés avec une dizaine de Communautés de Pratique. Parmi eux, SweetWiki et BayFac.

SweetWiki est un wiki sémantique, fondé sur un moteur sémantique (CORESE) et des pages wiki au format RDF. Ce wiki est utilisé pour soutenir la création collaborative de cours. Nous

³⁰<http://palette.ercim.org/>

détaillons ce projet en §6.2.1.

BayFac est un outil de classification puis de recherche de ressources. Au début, la classification est effectuée manuellement à partir de concepts d'ontologies. Puis le système applique des algorithmes bayésiens qui en se basant sur les métadonnées appliquées par les premiers utilisateurs et en analysant le contenu de nouvelles ressources va automatiquement proposer des métadonnées. La recherche des ressources s'effectue à travers une présentation multi-facette des ressources. Les facettes sont issues des ontologies définies au départ.

2.3 Conclusion

Le Web Sémantique offre des possibilités de traitements automatisés de connaissances informatisées. En 2006, même si plusieurs articles prévoient les possibilités du Web Sémantique dans le domaine pédagogique, peu de projets exploraient concrètement cette perspective. C'est cette année-là que démarra le projet LUISA auquel nous avons participé.

Dans ce projet, nous verrons comment les technologies du Web Sémantique peuvent être intégrées au sein d'une université, et quels services pédagogiques les traitements automatisés peuvent fournir.

3

Projet LUISA

Sommaire

3.1	Contexte	35
3.1.1	Contraintes du projet	35
3.1.2	Description générale du projet	36
3.2	Travaux	38
3.2.1	Étude des besoins à l’université et choix d’un contexte d’expérimentation	39
3.2.2	Prototype LUISA : fonctionnalités et raisonnements sémantiques	47
3.2.3	Ontologies développées	61
3.2.4	Expérimentations et discussion	72

Premier des trois projets. Le premier des trois projets de ce mémoire, est le projet européen LUISA. Celui-ci aborde l’indexation de documents pédagogiques selon l’angle du Web Sémantique.

3.1 Contexte

3.1.1 Contraintes du projet

Contrainte d’un projet européen. Par rapport aux deux autres projets, le projet LUISA possède une certaine rigidité, laissant parfois peu de liberté dans nos travaux. Cette rigidité provient du format des projets européens de type STREP³¹ qui force à définir un calendrier et des objectifs de recherche très précis, bien avant le début officiel du projet. Par la suite, durant le projet, tout changement dans ces paramètres requiert une lourde tâche de justification.

Contrainte de la collaboration. De plus, contrairement aux deux autres projets où nous sommes seuls responsables de nos travaux, le projet européen nous a demandé de travailler avec des partenaires techniques aux visions parfois très différentes des nôtres. Bien sûr, ce travail à plusieurs mains est légitime et nécessaire. Néanmoins, cela explique parfois certaines motivations privilégiant le développement technique plutôt que l’intérêt pédagogique.

³¹STREP, Specific Targeted REsearch Project : http://cordis.europa.eu/fp6/instr_strp.htm

3.1.2 Description générale du projet

Web Service Sémantique. Les outils développés dans le projet LUISA emploient des Web Services Sémantiques pour échanger des documents.

Un Web Service est un serveur proposant une interface d'interrogation par d'autres machines, selon un protocole défini. Par exemple, quand un humain veut réserver des billets sur Internet, il va sur le site Web d'une agence de voyage puis rentre sa destination, ses dates prévues de voyage et un prix maximal. Mais l'agence de voyage peut offrir le même service d'interrogation à des machines, à travers un Web Service. Celui-ci répond à des requêtes contenant aussi les paramètres de destination, de dates et de prix. Ce Web Service peut être utile à des sites Web partenaires. Ces derniers pourront alors proposer aussi des réservations, en interrogeant directement le Web Service de l'agence de voyage.

Les Web Services consistent donc en des interfaces de communication entre machines sur le Web.

Un Web Service Sémantique est un Web Service employant des ontologies dans son protocole de communication. Il peut ainsi décrire selon les concepts d'ontologies quels types de paramètres il nécessite, et quels retours il fournit. Le Web Service Sémantique et la machine effectuant la requête possèdent donc une vue commune sur le monde à travers les ontologies.

Or qui dit ontologies, dit aussi raisonnements automatisés. La puissance d'un Web Service Sémantique réside dans la possibilité de recevoir une requête puis d'interroger d'autres Web Services Sémantiques pour y répondre. Par exemple, le Web Service Sémantique d'une agence de voyage reçoit une requête avec un prix maximal en dollar. Or ce Web Service fonctionne avec des euros. Il interroge donc un autre Web Service Sémantique qui effectue la conversion Dollar-Euros. Il récupère le prix en euros et peut lui-même renvoyer les propositions de voyage correspondantes.

Les Web Services Sémantiques offrent donc des possibilités d'interrogation et de raisonnements automatisés, de manière distribuée sur le Web.

Extrait du document officiel. Voici une description générale, extraite du document principal du projet LUISA [LUISA-Consortium, 2005]. Les phrases en gras seront expliquées par la suite.

« L'interopérabilité des ressources pédagogiques entre les plateformes d'apprentissage (LCMS) ou les dépôts de Learning Objects (LOR) s'est beaucoup améliorée grâce aux efforts de normalisation comme LOM, SCORM ou IMS-LD (voir 1.2.2). Pourtant cette interopérabilité n'est qu'un premier pas vers des niveaux plus élevés d'automatisation et des possibilités de délégation de tâches envers des agents et modules logiciels. Ces techniques avancées requièrent une sémantique plus riche que celle offerte actuellement par les spécifications de métadonnées pour ressources pédagogiques. **Les technologies du Web Sémantique et l'utilisation d'ontologies sont capables de fournir une sémantique calculable par des machines pour automatiser les tâches relatives aux Learning Objects comme la sélection et la composition. De plus, les Web Services Sémantiques (SWS) fournissent une architecture technique et les moyens de médiation rendant cette sémantique interopérable.**

En effet, les Web Services Sémantiques fournissent les représentations conceptuelles requises, ainsi que les capacités de traduction et d'intégration de divers sys-

tèmes, partageant le but commun de réutiliser des Learning Objects. Un moteur de Web Service Sémantique intégré aux technologies standardisées actuelles des plateformes d'apprentissage étendra les possibilités des apprenants, enseignants et concepteurs pédagogiques. Il fournira des outils de recherche sémantique capable de demander et retrouver des Learning Objects à tout fournisseur qui s'enregistre en tant que dépôt sémantique (LOR sémantique).

Cette vision se traduit par les objectifs principaux suivants :

1. Développer une architecture fondée sur des Web Services Sémantiques pour découvrir, sélectionner, négocier et composer des Learning Objects.
2. Développer des techniques d'annotations sémantiques pour les Learning Objects. Ils pourront ainsi être distribués efficacement par les Web Services Sémantiques.
3. Développer un outil d'annotation sémantique adapté à la description de Learning Objects.
4. Intégrer l'architecture à des plateformes pédagogiques LCMS.
5. Produire des cas d'utilisation. L'implémentation de l'architecture sera mis en pratique dans des cas réels, où des ressources existantes seront indexées selon des représentations ontologiques. **Les expérimentations se situeront dans le cadre universitaire** et dans celui de la formation professionnelle. »

Le terme de « composition » signifie assembler plusieurs Learning Objects pour en créer un nouveau. Par exemple, assembler plusieurs parties de cours pour en former un nouveau, selon certaines règles pédagogiques.

Le premier paragraphe que nous avons mis en évidence cite les deux aspects majeurs du projet à savoir décrire sémantiquement des contenus pédagogiques et exploiter leurs manipulations à travers des Web Services Sémantiques.

Notre position dans le projet. Le consortium LUISA se compose :

- de cinq partenaires techniques, dont le rôle est de développer l'architecture des Web Services Sémantiques ;
- et de deux partenaires d'expérimentation, l'Université Henri Poincaré et EADS.

Notre rôle consistait à trouver des situations pour expérimenter le système LUISA en milieu universitaire, à définir un prototype et les ontologies pour évaluer ce système, et finalement à expérimenter le prototype.

En 1.2.1, nous avons mentionné le risque de considérer un contenu pédagogique comme un objet pouvant être facilement décontextualisé puis réutilisé. Ce risque était important pour le projet LUISA, fondé sur des Web Services distribués. Notre travail a donc consisté à explorer les possibilités des outils sémantiques des partenaires techniques, tout en incitant à ne pas oublier le contexte d'apprentissage.

Fonctionnement global. La Fig. 3.1 donne une vue simplifiée du fonctionnement d'une requête sémantique dans l'architecture LUISA.

Considérons un utilisateur, enseignant ou apprenant, connecté à une plateforme d'apprentissage en ligne (point 1). Une de ses activités nécessite un contenu pédagogique. L'utilisateur recherche par exemple des exercices de mathématiques.

La plateforme envoie la requête, contenant des contraintes sémantiques, à la couche sémantique (2). Cette couche est le noyau de l'architecture LUISA. Cette couche va traiter la requête en s'appuyant sur différentes ontologies.

Une requête est ensuite envoyée à différents dépôts possédant chacun une interface de Web Service (3 et 3'). Ces différents dépôts représentent l'aspect distribué de l'architecture LUISA. Une fois les documents trouvés, les dépôts renvoient les documents vers l'utilisateur à travers la plateforme (4 et 4', 5 et 6).

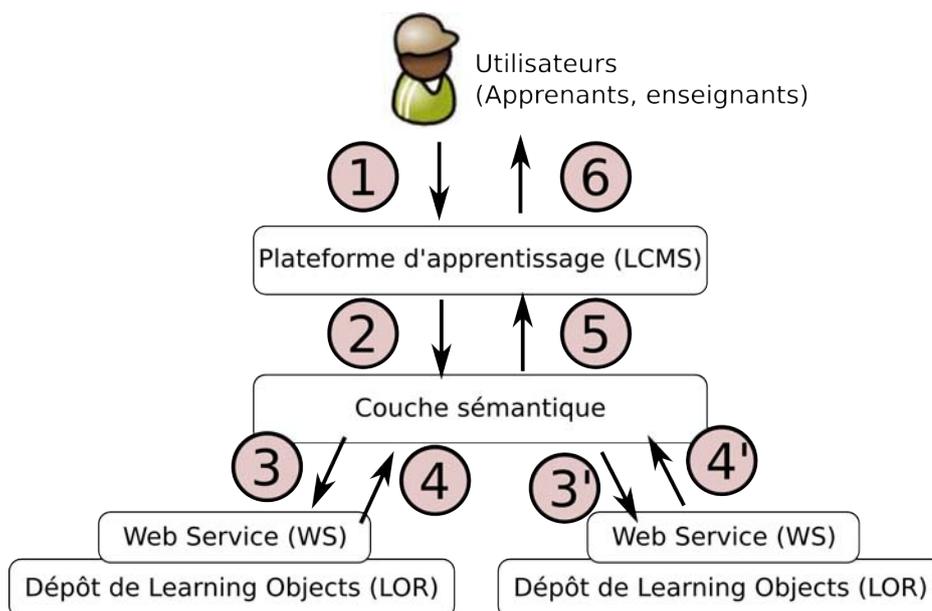


FIG. 3.1 – Fonctionnement d'une requête dans LUISA

3.2 Travaux

En travaillant sur le projet LUISA, nous avons pu explorer les possibilités du Web Sémantique pour l'indexation de documents pédagogiques.

Résumé de nos travaux.

Étude des besoins (§3.2.1) : notre premier travail a consisté à étudier, au sein de l'université Henri Poincaré, les besoins potentiels pouvant nécessiter et permettre l'emploi de technologies sémantiques.

Cette étude à base de séances de remue-méninges et d'entrevues a aussi permis de découvrir la perception des capacités du Web Sémantique par des non-informaticiens.

Cette étude s'est conclue par le choix du contexte du Certificat Informatique et Internet (C2I) et de la rédaction de cas d'utilisation s'y appliquant.

Conception de prototype (§3.2.2) : nous avons rédigé les spécifications d'un prototype permettant la recherche de documents du C2I et explorant les possibilités du Web Sémantique.

Les fonctionnalités du prototype permettent par exemple de décomposer une compétence en sous-compétences lors d'une requête, de prendre en compte l'opinion des utilisateurs dans le raisonnement sémantique ou d'expliquer le résultat de la requête aux apprenants.

Création des ontologies (§3.2.3) : les fonctionnalités du prototype exploitent quatre ontologies, que nous avons créées ou adaptées, liées au domaine du C2I ; la première est fondée sur LOM, les suivantes sont des ontologies de compétences, de disciplines et de logiciels.

Nous avons ensuite annoté 160 ressources selon ces ontologies (§3.2.3.6).

Expérimentation (§3.2.4) : l'expérimentation du prototype n'a pas pu se faire en situation réelle à cause de problèmes techniques de la version finale. Néanmoins, nous avons évalué les résultats du projet à l'aide de démonstrations puis d'entrevues avec différents membres de l'université.

Les expérimentations nous permettent de discuter les apports et limites du projet LUISA, et de certains aspects du Web Sémantique en général.

3.2.1 Étude des besoins à l'université et choix d'un contexte d'expérimentation

Malgré plusieurs années d'existence, les technologies du Web Sémantique sont peu présentes dans des applications à grande diffusion. Notamment, nous ne connaissons aucune réelle utilisation au sein d'universités françaises (hors expérimentation).

Avant toute réalisation technique, nos premiers travaux ont donc consisté à explorer la perception des membres de l'université sur les apports de cette technologie. Pour cela, nous avons réalisé des séances de remue-méninges (brainstormings) et effectué des entrevues en face à face (§3.2.1.1). Ceci nous a permis de déterminer les « espaces » potentiels où ces technologies pourraient s'implanter et répondre à des besoins exprimés (§3.2.1.2).

3.2.1.1 Remue-méninges et entrevues

L'objectif des remue-méninges et des entrevues était de comprendre les attentes des membres de l'université. Nous espérions récupérer :

- des témoignages sur l'expérience des membres par rapport à la manipulation des documents pédagogiques numériques à l'université,
- des besoins par rapport à ces documents,
- des idées de fonctionnalités à implémenter dans le projet LUISA,
- des contextes et des contacts pour expérimenter le projet LUISA.

Séances de remue-méninges. L'annexe B contient de manière plus détaillée la méthode et les résultats de ces séances. En voici les principaux points.

Nous avons organisé deux séances de remues-méninges réunissant au total 15 participants. Chaque séance durait deux heures.

Nous avons pu toucher des profils variés grâce à la présence de membres de services transversaux (bibliothécaires, ingénieurs e-learning) et d'enseignants de différentes disciplines.

La question de départ était : « En vous basant sur vos expériences professionnelles et personnelles, imaginez les services de demain autour des ressources numériques à l'université. » La

question posée fut volontairement vague, afin de laisser les participants exprimer librement leurs différents besoins en matière de documents.

Deux résultats principaux ressortirent de ces séances.

Le premier résultat provient des bibliothécaires. Selon leurs expériences, les étudiants ne savent plus effectuer des recherches documentaires correctement. En effet, les bibliothèques se vident car les étudiants effectuent la majorité de leurs recherches par le biais d'Internet. Ils essaient des requêtes sur les moteurs généralistes comme Google, et n'exploitent pas les bases de documentation présentes dans les bibliothèques universitaires. La conséquence est que les étudiants n'obtiennent souvent aucun résultat à leurs recherches et considèrent, à tort, que ces résultats n'existent pas.

Ce résultat montre la limite des moteurs généralistes et la **nécessité d'outils de recherche pour valoriser les documents pédagogiques de l'université**. Ces points vont dans le sens du projet LUISA qui cherche à améliorer les recherches des usagers en prenant en compte plus de sémantique.

Le second résultat est l'importance du contact humain autour des documents. Les utilisateurs ont parlé ainsi de « ressources vivantes », autour desquelles la présence des autres utilisateurs se faisait sentir, opposées à des « ressources mortes », simplement stockées. Ce besoin s'est notamment exprimé quand les participants ont décrit les plateformes de documents à l'aide des métaphores de la « pharmacie », où l'on tisse un lien avec le pharmacien et à qui on peut demander conseil, ou alors de la « cafétéria » où l'on peut discuter autour des tables avec convivialité.

Ce second résultat souligne un manque du projet LUISA, pour lequel aucun contact n'était à l'origine prévu autour des documents entre utilisateurs ou avec les auteurs. Ce retour des utilisateurs nous poussera à introduire cet **aspect plus « humain »** dans le prototype du projet LUISA, par exemple à travers la fonctionnalité prenant en compte l'opinion des utilisateurs dans les recherches.

Entrevues. L'annexe B contient de manière plus détaillée la méthode et les résultats des entrevues.

Les entrevues se déroulèrent généralement pendant une durée de deux heures. Elles ont eu lieu avec six enseignants de l'université, dont deux avaient participé à une séance de remue-méninges.

Là encore, les profils furent variés avec par exemple un enseignant de STAPS³² ou une responsable de la VAE³³.

Contrairement aux remue-méninges où la question de départ était très libre, les entrevues avaient pour objectif d'expliquer les possibilités des technologies du Web Sémantique et d'obtenir des cas d'utilisation possibles. Nous n'interrogeons pas les participants en tant que consommateurs des travaux issus du Web Sémantique, mais en tant que réalisateur potentiel d'une

³²STAPS, Sciences et Techniques des Activités Physiques et Sportives

³³VAE, Valorisation des Acquis de l'Expérience. La VAE est un service offert par l'université à des adultes pour transformer certaines compétences en diplômes universitaires. Cette « transformation » est validée à la suite de la préparation d'un dossier sur plusieurs mois.

application sémantique autour des documents pédagogiques.

Nous avons retiré trois résultats principaux de ces entretiens.

Le premier résultat est la **difficulté d'expliquer les possibilités du Web Sémantique à des enseignants** d'université. La connaissance du fonctionnement de ces technologies n'est pas utile pour en profiter comme internaute. En revanche, cette compréhension est vitale pour réaliser des outils adaptés aux documents pédagogiques et pour lesquels des enseignants devront forcément participer quant aux aspects pédagogiques. Pour des non-informaticiens ou des non-documentalistes, la notion d'ontologie est loin d'être intuitive. Réaliser des raisonnements les exploitant est encore plus abscons.

Le second résultat est la **difficulté pour les enseignants rencontrés de décrire précisément les aspects pédagogiques de contenus**. La responsable du service langue de l'université nous apprend par exemple qu'ils avaient voulu classer des contenus pédagogiques en anglais et que finalement un des critères les plus pertinents, pour lequel tout le monde était d'accord, avait été de classer par la longueur des textes.

Le troisième résultat est que les **cas d'utilisation obtenus ne concernaient généralement pas la recherche de documents pédagogiques**. La majorité des participants a proposé d'annoter sémantiquement les contenus pour les transformer en outils pédagogiques (e.g. questionnaires) pour les apprenants.

Discussion. Cette étude des besoins au sein de l'université est limitée principalement par deux défauts : le nombre restreint de réponses et la subjectivité des travaux.

Au total, une quinzaine de personnes participèrent. Ce petit nombre est loin d'être statistiquement représentatif de l'ensemble du personnel de l'université.

Néanmoins, la cohérence des réponses provenant d'horizon divers dans l'université est un bon critère pour dire que les résultats ne sont pas insignifiants.

De plus, les participants furent contactés car ils semblaient intéressés par le sujet des documents numériques pédagogiques (par exemple, car ils avaient déposé des documents sur la plateforme en ligne de l'université). Ils étaient tous volontaires, marquant encore leur intérêt pour ces sujets. Nous donnons donc d'autant plus de poids à leurs avis.

Le second défaut de ces travaux est forcément l'aspect subjectif que peuvent avoir les résultats. En menant les séances de remue-méninges ou les entrevues pour trouver des cas d'utilisation, nous avons forcément dû influencer un minimum les propos des participants.

Sans les prendre comme totalement précises, les réponses des enseignants donnent donc des indices sur leurs préoccupations et leurs idées sur l'indexation de documents.

Conclusion. Parmi les tendances esquissées, le premier point important est la difficulté pour les enseignants à comprendre les technologies du Web Sémantique et à décrire précisément les aspects pédagogiques de contenus. Pour le projet LUISA, cela signifie qu'il nous est difficile de demander à des enseignants de décrire les ressources. Pour les applications sémantiques en général, cela signifie que la participation d'enseignants dans la conception d'applications nécessitera

leur formation préalable.

Le second point important est la mention du contact humain à travers les ressources : contact avec les auteurs et contacts avec les autres utilisateurs. Les enseignants n'ont jamais mentionné des mécanismes d'assemblage de contenus ou de raisonnements intelligents pour les retrouver. En revanche, ils ont mentionné le besoin de discuter dessus avec d'autres utilisateurs. Ces aspects sont étonnamment proches des idées du Web Participatif (§ II). Nous verrons comment ils seront pris en compte dans le projet LUISA mais aussi dans les deux autres projets présentés dans ce mémoire.

Cette étude des besoins nous a permis de mieux comprendre les préoccupations des membres de l'université, par rapport aux documents pédagogiques. Parmi les besoins exprimés, le contexte répondant le mieux aux contraintes du projet LUISA concernait le Certificat Informatique et Internet.

3.2.1.2 Choix du contexte du C2I et cas d'utilisation

Lors des entrevues, un professeur de STAPS proposa d'appliquer le système LUISA aux documents du Certificat Informatique et Internet (C2I), qu'il enseignait.

Le C2I tient une position particulière dans le paysage éducatif français, par exemple de par son système de compétences. Or cette position correspond sur plusieurs points aux pré-requis des technologies mises en place dans LUISA.

Présentation officielle du C2I. Le C2I est un certificat des compétences informatiques mis en place par le ministère de l'Éducation Nationale. Le projet officiel du C2I est décrit en Fig. 3.2.

Intérêt du contexte du C2I. Nous avons choisi de travailler avec le contexte du C2I car :

- le C2I est un des rares modules de l'université fondé sur un **référentiel de compétences** à acquérir,
- étant obligatoire à l'université, ce certificat touche de **nombreux étudiants** quelle que soit leur filière,
- étant transversal à toutes les filières, l'enseignement s'adresse à des **étudiants aux niveaux disparates**. Or cette disparité nécessite des outils adaptés comme ceux que nous désirons expérimenter dans LUISA.
- touchant le domaine informatique, l'objet des contenus est **compréhensible par les partenaires techniques** dans LUISA,
- de **nombreuses ressources** sur le C2I sous licences libres sont accessibles sur le Web.

Ces aspects font du C2I le contexte idéal pour expérimenter les outils du projet LUISA, l'avantage principal étant l'approche par compétences.

Référentiel de compétences. Le ministère de l'enseignement supérieur et de la recherche propose un référentiel listant les compétences à acquérir pour obtenir le C2I.

Le référentiel national du C2I niveau 1 comprend deux parties : la première décrit les compétences générales et transversales visées par la certification, la seconde définit les différents domaines de compétences et les savoirs spécifiques et instrumentaux qui y sont associés. La Fig. 3.3 illustre les principaux domaines tandis que l'annexe C liste le référentiel complet.

Étudier dans la société de l'information

La rapide évolution des technologies de l'information et de la communication a engendré au cours de ces dernières années une progression notable des applications disponibles dans la vie courante et dans la vie professionnelle. Toute personne est aujourd'hui concernée par l'usage désormais banalisé d'outils informatiques.

Le Gouvernement a engagé un effort particulier pour favoriser la maîtrise de ces nouveaux outils de production, de transformation et de diffusion de l'information par l'ensemble de la société. L'éducation nationale contribue naturellement à ce projet gouvernemental d'une société de l'information pour tous qui nécessite un effort éducatif ambitieux. Son rôle est de dispenser à chaque citoyen la formation qui, à terme, le mettra à même de faire des technologies de l'information et de la communication une utilisation raisonnée, de percevoir les possibilités et les limites des traitements informatisés, de faire preuve d'esprit critique face aux résultats de ces traitements et d'identifier les contraintes juridiques et sociales dans lesquelles s'inscrivent ces utilisations. Dans l'enseignement scolaire cela se traduit notamment par les différents niveaux du brevet informatique et internet (B2I)®.

L'enseignement supérieur participe lui aussi largement à cet effort. Outre le développement du réseau RENATER, qui permet aux établissements d'enseignement supérieur et à leurs étudiants d'accéder à l'internet, les équipements informatiques se sont renforcés tandis que les ressources en ligne et les possibilités de formation à distance à travers les « campus numériques » commencent à se multiplier. Dans ce nouveau contexte, les finalités relatives aux technologies de l'information et de la communication sont doubles ; il s'agit :

- de permettre aux étudiants de maîtriser les compétences qui sont désormais indispensables à la poursuite d'études supérieures et d'être capables de faire évoluer ces compétences en fonction des développements technologiques ;
- de pouvoir établir qu'ils maîtrisent des compétences qui les aideront à s'insérer dans le monde des activités professionnelles à la fin de leur cursus.

Dans le but de développer, de renforcer et de valider la maîtrise des technologies de l'information et de la communication par les étudiants en formation dans les établissements d'enseignement supérieur, est donc institué un certificat informatique et internet (C2I)® comportant deux niveaux :

- un niveau 1 d'exigence applicable à tous les étudiants. Ce premier niveau sera exigible pour la titularisation des professeurs des écoles ;
- un niveau 2 faisant l'objet d'exigences plus élevées en fonction des orientations professionnelles des formations dispensées (à travers les enseignements de pré-professionnalisation et les filières professionnalisées). Son contenu sera précisé ultérieurement.

FIG. 3.2 – Présentation du C2I (Extrait du B.O. n° 19 du 9 mai 2002)

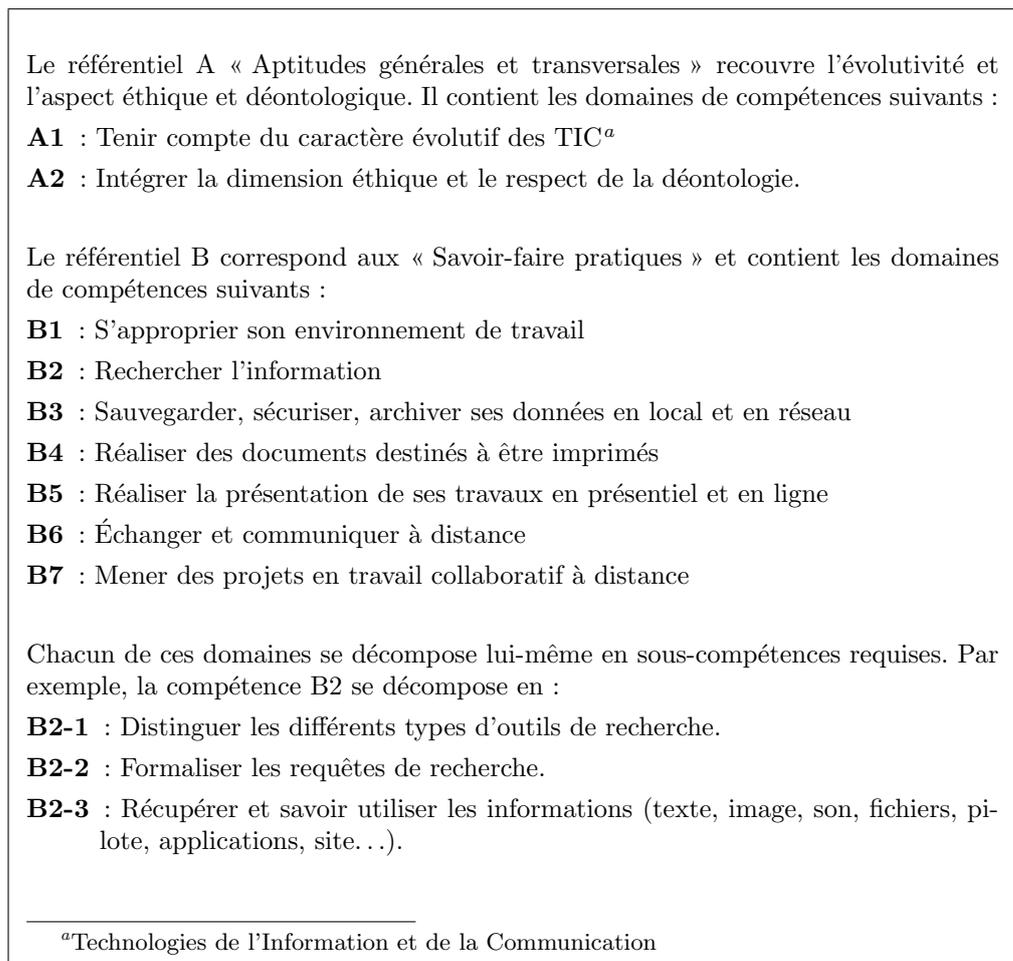


FIG. 3.3 – Extrait du référentiel de compétences du C2I

En listant les compétences à acquérir, le référentiel du C2I explicite des objectifs pédagogiques à atteindre. Cette explicitation est intéressante pour le projet LUISA car elle facilite le travail de formalisation de concepts pédagogiques dans des ontologies.

Les ressources du C2I. De nombreuses ressources ont été produites pour aider les professeurs et les étudiants à la préparation du C2I. D'un point de vue pédagogique, nous en distinguons quatre grands types :

- **Les cours** : leur rôle est d'expliquer les aspects théoriques aux étudiants. Ce sont souvent des présentations sous la forme d'un fichier PDF ou d'une page Web en HTML. Plus rarement, il s'agit de diaporamas accompagnés des commentaires audios de l'enseignant.
- **Les activités guidées** : leur rôle est d'aider les étudiants à développer des bases pratiques. L'activité consiste à un ensemble d'instructions, plus ou moins précises, que les apprenants doivent suivre avec un certain outil pour accomplir une tâche, telle qu'envoyer un email avec une pièce jointe.
- **Les activités non-guidées** : leur rôle est d'entraîner les étudiants pour l'évaluation. Une activité donne une tâche à accomplir mais ne donne pas d'indications pour l'effectuer.
- **Les tests** : Leur rôle est d'évaluer le niveau atteint par les étudiants. La plupart des tests sont des auto-tests, où l'apprenant doit répondre honnêtement à des questions (e.g. « Je

sais dézipper un fichier : oui/non »). Il existe aussi, mais moins fréquemment, des tests automatiques à choix multiples où l'étudiant est évalué de manière plus précise.

L'annexe D présente un exemple d'activité non-guidée, issu d'un module C2I donné à l'université Henri Poincaré.

Trouver des ressources sur le C2I. En 2006, sur le Web, plusieurs initiatives d'universités et d'enseignants mettaient à disposition des ressources pour le C2I. Les sites les plus fournis étaient le C2IMES³⁴ et PorTICES³⁵.

Ces sites proposaient généralement une recherche par compétences générales et parfois par mots-clefs à l'image du site officiel du C2I (Fig. 3.4).

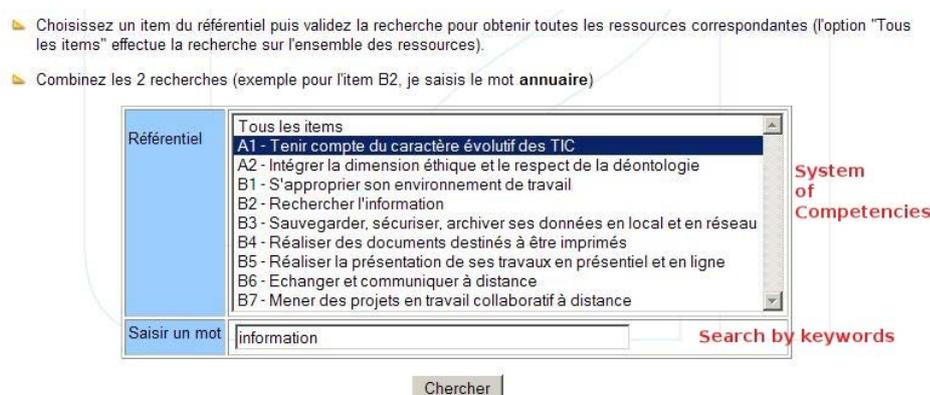


FIG. 3.4 – Moteur de recherche de ressources C2I du site officiel (en 2006)

Au sein de l'université Henri Poincaré, les documents préparant au C2I étaient déposés et gérés sur une plateforme d'apprentissage Moodle. Or si cette plateforme convient très bien pour créer des groupes d'étudiants ou paramétrer des forums, elle ne dispose pas d'une gestion évoluée pour organiser ou rechercher des contenus pédagogiques. Ces derniers sont généralement compartimentés par cours. L'écran 3.5 illustre l'organisation des ressources du C2I pour les étudiants de l'université Henri Poincaré.

Ainsi, la dispersion des ressources pour le C2I sur le Web et de manière générale, les outils basiques de recherche formaient un cadre idéal pour expérimenter de nouveaux moyens de recherches, comme ceux du projet LUISA, fondés sur les Services Web Sémantiques.

Besoins identifiés. La situation à l'université Henri Poincaré illustre deux grands besoins par rapport aux documents du C2I en particulier et aux documents pédagogiques en général.

Premièrement, l'université a **besoin d'outils pour gérer les productions numérisées de ses enseignants**. Sinon, elle perdra un grand nombre de ressources qui ne seront pas visibles en dehors des classes. Or les outils actuels, comme les plateformes d'apprentissage, ne sont pas prêts pour une gestion avancée des contenus pédagogiques dans une université et encore moins

³⁴C2IMES : <http://www.c2imes.org/>

³⁵PorTICES : <http://www.portices.fr>



FIG. 3.5 – Ecran Moodle d’organisation des cours. Les plateformes d’apprentissage ne sont généralement pas adaptées pour rechercher des documents pédagogiques, en dehors d’une recherche par cours.

entre plusieurs établissements.

Deuxièmement, il existe un besoin de fournir des **parcours d’apprentissage adaptés aux élèves** (par exemple, changeant de cursus). Ce besoin est dans le prolongement du processus de Bologne³⁶ qui tend à augmenter le nombre de ces étudiants « non-standard ». Dans le futur, il sera probablement nécessaire d’élargir l’utilisation des ressources en ligne pour mettre à jour ou approfondir une compétence à la demande d’un étudiant.

Or la situation de l’université Henri Poincaré ne semble pas être exceptionnelle. La plateforme d’apprentissage Moodle qu’elle utilise pour gérer les ressources pédagogiques numériques est aussi employée par la majorité des universités françaises.

La gestion des documents numériques pour le C2I révèle deux grands besoins, auxquels les technologies du Web Sémantique semblent pouvoir fournir des réponses. La section suivante expose les cas d’utilisation prévus, au sein du projet LUISA, pour répondre à ces besoins.

Cas d’utilisation. Une fois trouvé le contexte du C2I, nous avons proposé aux partenaires européens de travailler selon deux cas d’utilisation. Ces cas servent à définir le prototype implémentant le système LUISA.

Le cas d’utilisation principal se situe à l’université. Un enseignant prépare ses étudiants au C2I durant un module sur un semestre. Dans ce cours, le professeur alterne des phases de cours tous ensemble et des phases de travaux pratiques où les étudiants sont libres de travailler à leur

³⁶Le processus de Bologne vise à créer une aire européenne d’enseignement supérieur, dans laquelle les étudiants peuvent choisir parmi un large panel de cours et bénéficier de procédure souple de reconnaissance des diplômes : http://ec.europa.eu/education/policies/educ/bologna/bologna_en.html

rythme sur ordinateur.

Durant les phases libres, l'enseignant indique aux étudiants une page Web où se trouve le système LUISA. Cette page Web est intégrée dans la plateforme d'apprentissage Moodle. Les étudiants trouvent sur cette page un moteur de recherche qui propose des contenus (exercices, cours...) selon leurs profils. Les étudiants pourront y observer leur arbre de compétences, évoluant au fur et à mesure qu'ils accomplissent des exercices, ainsi que d'autres fonctionnalités permises par le système LUISA.

Le second cas se situe dans un cadre similaire mais pour les enseignants. Quand un professeur effectue une recherche de ressources pour le C2I sur la page du système LUISA, celui-ci propose des paquets de ressources, assemblées selon des critères pédagogiques. Ces paquets peuvent alors être fournis par l'enseignant aux étudiants, après d'éventuelles modifications par ses soins.

Voici maintenant le détail du prototype et de ses fonctionnalités.

3.2.2 Prototype LUISA : fonctionnalités et raisonnements sémantiques

Pour expérimenter les possibilités des technologies sémantiques, nous avons conçu un prototype implémentant six fonctionnalités fondées sur des ontologies.

3.2.2.1 Présentation générale du prototype

Entre technologies et besoins. Étant du côté utilisateur, nous avons été responsables des spécifications du prototype expérimenté en milieu universitaire. Cependant ces spécifications sont issues de régulières négociations avec les partenaires techniques. En effet, pendant tout le projet, nous avons dû chercher un équilibre entre expérimenter les technologies sémantiques des partenaires et répondre aux besoins pédagogiques des utilisateurs.

Pour présenter ce prototype, nous verrons les ontologies, les étapes de navigation et enfin les fonctionnalités.

Quatre ontologies. Les fonctionnalités sémantiques du prototype exploitent quatre ontologies :

- de documents pédagogiques (S-LOM), version sémantisée de LOM ;
- de compétences (GCS) reprenant le référentiel de compétences du C2I ;
- de discipline, représentant les disciplines de l'université Henri Poincaré ;
- de logiciels, représentant les différents logiciels employés dans la préparation au C2I.

Ontologies et fonctionnalités sémantiques sont un peu les deux facettes d'une même pièce, chacune aidant à la compréhension de l'autre. Dans ce mémoire, nous avons fait le choix de détailler les ontologies en §3.2.3 juste après cette présentation des fonctionnalités. Néanmoins, nous invitons les lecteurs à se référer aux ontologies auparavant s'ils en ressentent le besoin.

Navigation dans le prototype. Le prototype est intégré en tant que plugin à une plateforme d'apprentissage Moodle. L'idée est de faciliter dans une université la recherche de documents préparant au C2I (documents que nous avons indexés, voir §3.2.3.6).

Le prototype offre un parcours simple de navigation (Fig. 3.6). Les différentes étapes de la navigation sont :

1. **Identification** : Les utilisateurs s'identifient en entrant leur identifiant et mot de passe, probablement fournis par l'université. Cette identification, classique, permet surtout d'associer un profil à chaque utilisateur.
2. **Configuration** : La première fois qu'ils se connectent, les utilisateurs doivent configurer les préférences de leur profil. Celles-ci permettront au système d'adapter les résultats des recherches à chaque utilisateur. Le profil est exprimé en partie à l'aide des ontologies.
3. **Requête avec arbre de compétences** : La requête des utilisateurs s'effectue à travers un arbre de compétences du C2I, issu de l'ontologie de compétences. Cet arbre présente les compétences que l'utilisateur a déjà acquises et celles qu'il peut travailler pour atteindre le niveau du C2I. L'intérêt de cet arbre réside notamment dans son côté interactif, fondé sur la découverte de l'ontologie de compétences.
4. **Résultats** : Une fois la requête soumise, le système va rechercher les documents correspondant aux contraintes de l'utilisateur (compétences à travailler, préférences définies dans la configuration). C'est à cette étape que s'applique la majorité des raisonnements sémantiques permis par les ontologies. Le résultat correspond à une liste de documents, que l'utilisateur va pouvoir ou non sélectionner pour travailler avec.
5. **Plan de travail** : Les documents sélectionnés sont récapitulés sur cette page et forment un plan de travail pour l'apprenant. Ce plan est conservé et permet à l'utilisateur de quitter sa session puis d'y revenir plus tard en retrouvant le plan sauvegardé. Après avoir travaillé les documents du plan, le système propose à l'utilisateur de mettre à jour ses compétences. Si l'utilisateur accepte, lors d'une prochaine requête, l'arbre de compétences affichera le nouvel état des compétences de l'utilisateur.

En situation réelle, il faudrait une étape d'évaluation des compétences de l'étudiant. Néanmoins, dans le cadre de ce projet, nous n'avons pas pu nous intéresser à cette étape d'évaluation. Pour nos expérimentations, nous considérons donc qu'il suffit de demander à l'étudiant s'il désire mettre à jour ses compétences.

Fonctionnalités sémantiques. Autour de cette navigation viennent s'implémenter six fonctionnalités permises par l'emploi d'ontologies et enrichissant la recherche de documents :

- **Arbre de compétences** : permet aux utilisateurs de découvrir les compétences du C2I à travers un arbre de compétences interactif, puis d'y formuler leur requête (§3.2.2.2),
- **Recherche sémantique** : permet au système de fournir des résultats issus de raisonnements sémantiques fondés sur les différentes ontologies (§3.2.2.3),
- **Notation par les utilisateurs** : permet au système de retourner des ressources prenant en compte les notations des utilisateurs intégrées aux raisonnements sémantiques (§3.2.2.4),
- **Équivalence entre concepts** : permet au système de retrouver des concepts similaires (e.g. une compétence), dans des ontologies semblables provenant de différentes institutions. On parle de mapping entre concepts (§3.2.2.5),
- **Explication des résultats sémantiques** : permet aux utilisateurs de comprendre les raisonnements sémantiques effectués par le système pour fournir les résultats (§3.2.2.6),
- **Composition de paquets de ressources** : permet au système de proposer aux enseignants des paquets de ressources, construits automatiquement à partir de règles pédagogiques appliquées aux concepts des ontologies (§3.2.2.7).

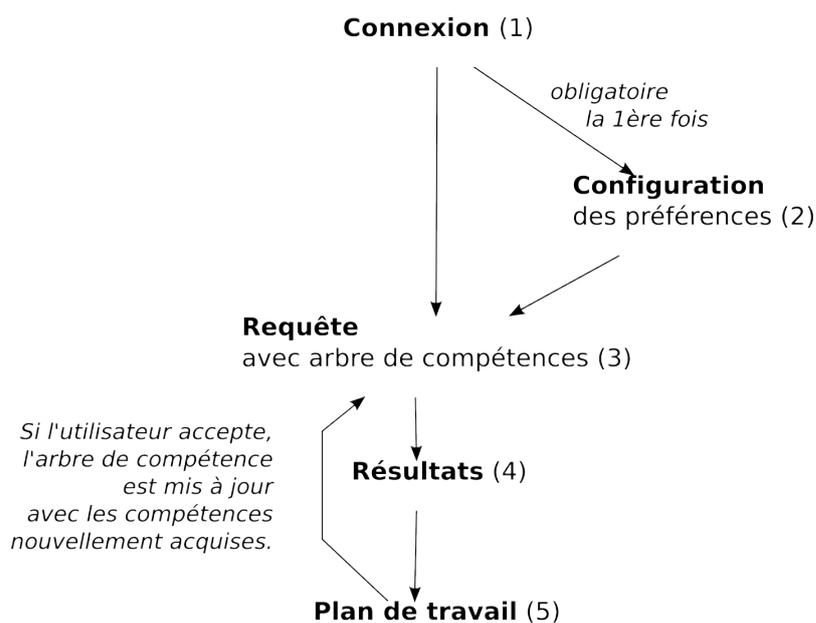


FIG. 3.6 – Les 5 étapes de navigation dans le prototype LUISA.

3.2.2.2 Arbre de compétences

Problème : les ontologies sont des représentations formelles d'un domaine censées être partagées par une communauté. Néanmoins, ces modèles sont souvent réalisés par des experts du domaine sur une période assez longue. Ceci a pour conséquence la complexité de ces représentations qui nécessite un travail sur leur présentation pour aider à leur compréhension par des utilisateurs non-experts.

Intérêt de ce problème : si représenter une hiérarchie de concepts est assez courant avec un arbre, représenter des relations transversales est tout de suite plus difficile s'il faut en même temps conserver une interface simple et non surchargée. Dans notre situation la question s'est posée pour aider les utilisateurs à comprendre l'ontologie de compétences, possédant des liens de dépendances entre compétences.

Solution proposée : au lieu de présenter l'arbre des compétences en entier, le système rend interactif l'arbre qui n'affichera pas immédiatement toutes les relations. L'idée est de laisser l'utilisateur « jouer » avec l'arbre pour se construire une représentation mentale de l'ontologie. Jouer consiste à développer des branches ou à sélectionner des nœuds, pour voir l'arbre évoluer en conséquence.

D'un point de vue pratique, nous avons fait le choix de centrer les requêtes de l'utilisateur sur les compétences. Or le référentiel du C2I contient une cinquantaine de sous-compétences que les apprenants doivent appréhender. De plus, ces compétences sont liées par des relations de dépendances transversales. Par exemple, la dernière compétence B7, sur le travail collaboratif, nécessite toutes les autres compétences (nous avons considéré qu'un apprenant ne pouvait pas travailler avec d'autres à travers les outils s'il ne maîtrisait pas déjà seul ces outils).

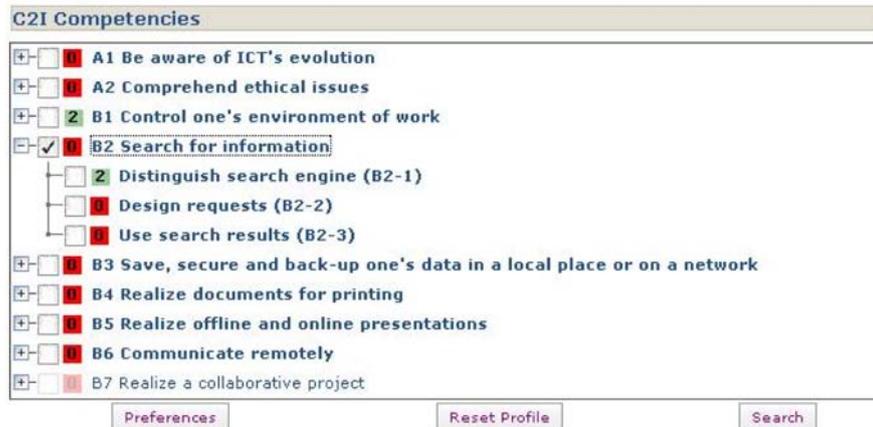


FIG. 3.7 – Copie d’écrans des étapes « Requête » (en haut) et « Résultats » (en bas). L’arbre de compétences affiche les compétences et leur niveau d’acquisition. L’utilisateur y sélectionne les compétences qu’il désire travailler.

L’interaction avec l’arbre est fondée sur les relations « details » et « requires » de l’ontologie de compétences (§3.2.3.3).

L’interface permet à l’utilisateur de sélectionner les compétences qu’il désire travailler. L’arbre des compétences est présenté de manière compacte avec seulement visibles les compétences supérieures (i.e. A1, A2 et de B1 à B7). La relation « details » est exploitée par le prototype car l’utilisateur peut déployer chaque compétence supérieure pour afficher les compétences incluses. Par exemple en cliquant sur B2, l’arbre affichera alors les compétences B2-1, B2-2 et B2-3. Ceci est la première interaction, classique pour représenter des hiérarchies. La Fig. 3.8 montre le détail de B2 après un clic dessus.

En revanche, l’arbre de compétences n’est pas qu’une vue générale sur le référentiel du C2I, c’est aussi une vue sur les compétences acquises et à acquérir par l’utilisateur. Dans la Fig. 3.8, l’échelle des niveaux de compétences apparaît sur le côté gauche de l’écran, allant de 0 à 3. Les niveaux apparaissent aussi sur l’arbre et représentent les niveaux de compétences de l’utilisateur. Il a par exemple acquis la compétence B1 dans sa globalité et la sous-compétence B2-1. Les



FIG. 3.8 – Arbre de compétences du C2I avec le détail de la compétence B2 et une compétence B7 inaccessible (Prototype UHP-LUISA)

cases à cocher, à gauche des compétences, servent à sélectionner les compétences que l'utilisateur désire travailler. Or, la compétence B7 apparaît en grisé et n'est pas sélectionnable. En effet, B7 nécessite de posséder toutes les autres compétences pour être ensuite atteignable. Cette compétence en grisé est donc une manière implicite de représenter la relation de dépendance entre B7 et les autres compétences.

L'utilisateur peut aussi sélectionner plusieurs compétences à travailler. L'arbre se met alors automatiquement à jour pour indiquer les nouvelles compétences atteignables. Par exemple, les compétences B2 à B6 dépendent directement de B1. Lorsque l'utilisateur arrive pour la première fois sur le prototype et ne possède aucune compétence, seuls A1, A2 et B1 sont sélectionnables, les autres compétences apparaissant en grisé. Or dès que l'utilisateur sélectionne B1 en cochant la croix à gauche de la compétence, les compétences B2 à B6 deviennent à leur tour sélectionnables, et l'utilisateur pourra demander en même temps à travailler B2 par exemple. L'arbre a ainsi réagi aux interactions de l'utilisateur pour indiquer, de manière plus légère qu'un texte, les dépendances entre compétences.

Conséquence de cette fonctionnalité : l'exemple ci-dessus n'est qu'une démonstration mineure des interactions possibles avec une ontologie. Néanmoins, alors que les ontologies et les référentiels sont des structures souvent complexes, proposer des interactions intéressantes, fondées sur les relations de l'ontologie, semble possible et utile à explorer pour aider les utilisateurs à appréhender un domaine et à y formuler par la suite des requêtes.

3.2.2.3 Recherche sémantique

Problème : le second problème est celui de la recherche de documents proprement dite en fonction de la requête et des préférences de l'utilisateur.

Intérêt de ce problème : pour rechercher des ressources, il existe deux grandes sortes de moteurs de recherche. Les moteurs généralistes comme Google se basent sur une indexation automatique des contenus mais n'offrent ni une vision, ni des caractéristiques pédagogiques pour leurs recherches. Les seconds moteurs sont spécialisés. Un des plus connus est celui de MERLOT qui offre des caractéristiques pédagogiques pour fouiller une base de documents indexés (généralement selon un modèle LOM). Pourtant ces moteurs spécialisés se fondent généralement sur une indexation syntaxique des contenus. Cela signifie que les champs contiennent des valeurs soit sous la forme d'un texte libre, soit sous la forme d'un terme issu d'une liste énumérée. Ces types de valeurs sont pratiques pour effectuer des filtrages simples : l'utilisateur demande des ressources pour travailler une compétence « Bureautique » (section « classification » de LOM), le système peut lui renvoyer toutes les ressources contenant « Bureautique » dans le champ correspondant. Éventuellement, le système peut élargir la requête en retournant aussi les ressources contenant la chaîne de caractère « Bureautique » dans les champs libres. Mais le « raisonnement » du système s'arrête là avec ce type de donnée. Si un document est indexé avec le terme « tableur » ou « Excel », le système ne pourra comprendre que le document traite de bureautique. Comment trouver alors des documents en fonction de termes employés dans la requête mais présents ni dans le document, ni dans ses métadonnées ?

Solution proposée : les technologies du Web Sémantique fournissent des réponses à ce problème en employant une modélisation du domaine avec des ontologies. En raisonnant sur ces ontologies, le système peut trouver des documents à partir de termes absents du document et de ses métadonnées. Le système peut aussi adapter le résultat au profil de l'utilisateur.

Dans le prototype LUISA, ces mécanismes ont été possibles grâce aux ontologies et grâce à l'indexation des ressources selon ces ontologies.

Dans ses préférences (étape 2 de la navigation, Fig. 3.9), l'utilisateur peut rentrer notamment :

- le système d'exploitation et les logiciels qu'il possède sur sa machine.
- la discipline qu'il étudie.

Ces préférences sont sauvegardées par le système en tant que référence aux concepts des ontologies de logiciel et de disciplines.

Au moment de la requête (étape 3 de la navigation), l'utilisateur sélectionne la ou les compétences qu'il désire travailler. Les compétences sélectionnables sont limitées par le système suite au traitement des dépendances entre compétences, comme nous l'avons vu précédemment.

Nous n'avons pas été intéressés par la recherche en texte libre qui peut être accomplie par exemple à l'aide de moteur spécialisé. Nous nous sommes concentrés sur les requêtes en fonction des compétences et affinées à l'aide des profils des utilisateurs.

A partir des préférences de l'utilisateur et de la requête en terme de compétences, le système va pouvoir appliquer les raisonnements suivants (grâce aux ontologies) :

- Décomposer une compétence en sous-compétences, pour trouver plus de documents répondant aux contraintes de l'utilisateur,
- Trouver éventuellement les logiciels correspondants à une compétence, pour aussi accumuler plus de documents répondant aux contraintes,



FIG. 3.9 – Ecran de configuration du profil utilisateur (Prototype LUISA)

- Restreindre les logiciels aux logiciels interopérables avec ceux de l'utilisateur.

Dans LUISA, les raisonnements s'effectuent à l'aide du framework IRS III³⁷ (Internet Reasoning Service). Nous avons décrit les raisonnements en langage naturel et les partenaires techniques les traduisirent en langage proche de la logique du premier ordre nécessaire à IRS III. Les raisonnements sémantiques sont détaillés en annexe G. Voici un aperçu de leur fonctionnement.

Le moteur de recherche sémantique LUISA fonctionne en deux étapes : il essaye de récupérer un maximum de documents puis pour chaque document est calculée une valeur correspondant à son adéquation avec chaque contrainte.

Décomposer une compétence en sous-compétences Pour trouver un maximum de documents, le système cherche tout d'abord tous les documents correspondant aux concepts de la requête, c'est à dire la ou les compétences recherchées. Si l'utilisateur veut travailler la compétence B2, le système va chercher tous les documents possédant dans le champ « Compétences visées » (targeted competencies) de leur description S-LOM, l'instance de la compétence B2. Puis, pour augmenter le nombre de documents et avoir ainsi plus de possibilités avant de restreindre le choix avec les contraintes, le système va chercher les documents indexés non pas avec la compétence B2 mais avec les sous-compétences B2-1, B2-2 et B2-3 trouvées grâce à la relation « details » de l'ontologie de compétences. On dit alors que la compétence B2 a été décomposée par le système en ses sous-compétences et ceci a permis de récupérer des documents dont les métadonnées ne répondaient pas à la requête et pourtant qui y répondaient.

³⁷IRS III : <http://technologies.kmi.open.ac.uk/irs/>

Trouver les logiciels correspondants à une compétence. Si trop peu de documents ont été trouvés, le système peut effectuer une sous-requête en cherchant les logiciels employés par la compétence cherchée puis trouver les documents employant ces logiciels (relation « about » de l'ontologie de compétence).

Par exemple, le système trouve trop peu de documents indexés par la compétence B2 « Effectuer une recherche sur Internet ». Grâce à la relation « about » le système déduit que la compétence traite d'outils nommé « moteurs de recherche » dans l'ontologie de logiciels. Puis que ce concept de « moteurs de recherche » est un concept général recouvrant des concepts tels que « Google » ou « Yahoo SearchEngine » (dans l'ontologie de logiciels). Grâce à ces connaissances, le système peut donc effectuer une nouvelle requête pour obtenir plus de documents en cherchant cette fois les documents indexés avec les termes ou instances « Google » ou « Yahoo SearchEngine ».

En effet, indexer en fonction d'aspects pédagogiques n'est pas une tâche aisée, comme nous l'avons mentionné auparavant. Ceci n'est pas une pratique habituelle des bibliothécaires qui ont l'habitude d'indexer en fonction de thématiques. Faire le lien entre les compétences et les logiciels est un moyen de passer d'un aspect pédagogique à un aspect plus thématique et de récupérer potentiellement des documents indexés selon ces aspects thématiques.

Logiciels interopérables avec ceux de l'utilisateur. Après la phase de récupération d'un grand nombre de ressources en fonction des compétences désirées, le système va essayer de les ordonner en fonction des préférences de l'utilisateur.

En effet, dans ses préférences, l'utilisateur définit le système d'exploitation ainsi que les logiciels installés sur sa machine. Grâce à la hiérarchie des logiciels offerte par l'ontologie des logiciels, l'utilisateur pourra rester à un niveau élevé (e.g. « j'ai un logiciel de traitement de texte ») ou alors être plus précis (e.g. « j'ai Microsoft Word 2007 d'installé »). Cette granularité dépendra du temps ou des connaissances qu'il possède pour remplir les champs demandés.

Au moment de l'ordonnement des ressources, le système calcule une valeur supérieure pour les documents portant sur des logiciels possédés par l'utilisateur. De plus, grâce à la relation d'interopérabilité (de l'ontologie de logiciels), le système peut à l'aide du système d'exploitation connaître les logiciels qui pourront être ou non installés. Par exemple, si l'utilisateur possède un système d'exploitation « Linux » et que le système a trouvé deux ressources sur le traitement de texte, mais l'une adaptée au logiciel « Word » et l'autre sur le logiciel « OpenOffice » alors le système avantagera la ressource portant sur « OpenOffice » car il déduira de l'ontologie de logiciels que « Word » n'est pas installable (interopérable) sur « Linux ».

3.2.2.4 Notation par les utilisateurs

Problème : Le Web Sémantique se base sur des ontologies, réalisées généralement par des experts d'un domaine de manière a priori, c'est à dire avant utilisation des ressources. Or il peut être utile de prendre en compte l'avis des utilisateurs pour ordonner les contenus selon les qualités perçues.

Intérêt de ce problème : L'avis des utilisateurs sur Internet fait plutôt partie de la tendance du Web Participatif (partie II). Mêler les approches sémantiques et participatives fait l'objet de

toute la troisième partie (partie III). Néanmoins, nous avons eu envie d'expérimenter au sein du projet LUISA comment les retours des utilisateurs pouvaient être intégrés aux raisonnements sémantiques, notamment pour aider à définir des contextes pour les ressources.

The screenshot shows the 'Selection phase' of the LUISA Moodle site. The user is logged in as 'Admin User'. The breadcrumb trail is 'LUISA > LUISA extension > Query phase > Selection phase'. The 'Your query:' section shows 'C2I Competencies: B1 Control one's environment of work' and '5 LO(s) found in workplan'. The 'Your selection:' section lists three resources with their CoFe ratings and checkboxes:

Resource ID	Resource Name	URL	CoFe Rating	Checkbox
1.	<< Moteurs de recherche et annuaires * >>	(http://www.uhp-nancy.fr/ontologies/LUISA#aa1881da151168c3e2c0d)	3 stars	<input type="checkbox"/>
2.	<< Systemes de gestion de fichiers (2) * >>	(http://www.uhp-nancy.fr/ontologies/LUISA#aa1881da151168c06f43d)	2 stars	<input type="checkbox"/>
3.	<< Systemes d'exploitation (2) Architecture de l'ordinateur * >>	(http://www.uhp-nancy.fr/ontologies/LUISA#aa1881da151168bf1a14f)	3 stars	<input type="checkbox"/>

Competencies for resource 1: k_distinguish_search_engine; b1; k_design_requests
Competencies for resource 2: b1; k_compress
Competencies for resource 3: b1

A button 'Work with these resources' is located at the bottom right of the selection area.

FIG. 3.10 – Ecran affichant les ressources proposées avec leur rating (étoiles) sur le côté (Prototype UHP-LUISA)

Solution proposée : Cette fonctionnalité consiste à laisser les utilisateurs noter les ressources suite à leur utilisation. Les notes sont alors agrégées par le système et leurs moyennes exploitées pour ordonner les résultats des prochaines requêtes. Cependant, contrairement à la même fonctionnalité dans un contexte du Web Participatif, dans notre proposition, les notes sont mises en rapport avec les profils des utilisateurs définis par des ontologies.

D'un point de vue pratique, le prototype laisse les utilisateurs noter les Learning Objects à l'étape 5 où est présenté le plan de travail. La note se situe entre 0 et 4 inclus. D'un point de vue ergonomique, la notation doit être la plus simple et rapide possible pour inciter les utilisateurs à l'effectuer. L'échelle de 0 à 4 s'inspire des interfaces des sites connus exploitant cette notation, comme Amazon.com pour la recommandation de livres. Un message à côté des étoiles explique qu'elles représentent le niveau de satisfaction : 0 signifiant que la ressource était totalement inadaptée ou inintéressante tandis que 4 qualifie une « bonne » ressource pour acquérir la compétence visée.

Au moment d'afficher les résultats à une requête (étape 4), le système ordonne les Learning Objects selon la moyenne des notes données par les utilisateurs aux profils similaires (Fig. 3.10).

Exemple. Les notes sont prises en compte selon les disciplines des utilisateurs (remplies dans les préférences). Considérons deux Learning Objects (intitulés LO1 et LO2) visant la même

compétence, et trois utilisateurs (nommés U1, U2 et U3). U1 et U2 sont en « Biologie » tandis qu’U3 étudie les « Mathématiques » (instances de l’ontologie de disciplines). Ils travaillent tous avec les ressources LO1 et LO2. La table suivante résume leurs notes entre 0 et 4 :

	Discipline	notation LO1	notation LO2
U1	Biologie	4	1
U2	Biologie	3	0
U3	Mathématiques	1	2
U1 et U2 (moyenne)	Biologie	3,5	0,5
U3 (moyenne)	Mathématiques	1	2

Ce tableau montre que les étudiants en Biologie préfèrent LO1 tandis que l’étudiant en Mathématiques préfère LO2. Considérons alors un dernier utilisateur U4. Si U4 étudie en Biologie, le système prendra en compte les notations effectuées par les utilisateurs de cette discipline. LO1 semble plus intéressant pour U4. Si U4 étudie les Mathématiques, la notation de U3 sera prise en compte où LO2 semble plus adapté.

Ainsi, grâce aux évaluations des utilisateurs et grâce aux profils renseignés selon l’ontologie de discipline, des contextes de disciplines peuvent être déterminés pour les ressources et améliorer la recherche de documents pédagogiques.

Suite de cette fonctionnalité : de manière plus large, cette fonctionnalité implémentée pour une simple ontologie de disciplines pourrait être employée avec des profils plus complets. Par exemple, au lieu de profils remplis par les utilisateurs, le système pourrait directement récupérer les données des dossiers de l’université³⁸ qui donnerait des critères par âge ou par options prises, et fournirait des « contextes » pédagogiques plus pertinents.

3.2.2.5 Équivalence entre concepts

Problème : Une ontologie est une représentation partagée par une communauté d’utilisateurs. Comment faire alors si deux communautés ont des représentations proches mais néanmoins différentes ? Par exemple, nous avons employé une ontologie définissant les compétences du C2I pour indexer les documents. Or d’autres organisations, sans rapport avec l’université ou le C2I, peuvent être intéressées par des ressources préparant aux compétences de la bureautique. Comment leur faciliter les recherches parmi les documents que l’université a indexés ?

Intérêt de ce problème : Ce problème est d’autant plus intéressant qu’il est fondamental pour le développement du Web Sémantique. Au sein même des réseaux d’universités, les organisations sont différentes. L’exemple de l’organisation des disciplines est frappant. Nous avons construit l’ontologie de discipline à partir du site Web de l’université. Or chaque université peut avoir une organisation différente, issue de passés différents. La difficulté est d’aider le système à reconnaître des concepts proches dans des ontologies différentes. Il existe des algorithmes pour essayer de reconnaître automatiquement les similarités dans les ontologies à partir des structures ou des termes employés ([Gasevic and Hatala, 2006], [Kalfoglou and Schorlemmer, 2003]), mais ces méthodes, dites d’alignement, sont encore peu précises pour l’instant.

³⁸En respectant bien sûr le chartes éthiques sur le recoupement de données.

Solution proposée : La proposition exploite l'architecture du projet LUISA fondée sur des Web Services Sémantiques. Les Web Services offrent une approche distribuée de la sémantique. Notre proposition consiste à créer un Web Service qui trouvera les équivalences éventuelles entre les concepts d'ontologies proches. Les équivalences sont trouvées manuellement par un expert analysant les deux ontologies.

Dans le prototype LUISA, la traduction fut appliquée pendant la phase de recherche de ressources pour découvrir le maximum de ressources. La traduction fut établie entre l'ontologie de compétences sur le C2I que nous avons développée, et l'ontologie de compétences développée par l'autre partenaire côté utilisateur du projet LUISA : EADS. Alors que nous développions des cas d'utilisation pour expérimenter l'architecture LUISA dans un contexte universitaire, EADS développait des cas d'utilisation pour la formation professionnelle dans le contexte de l'aéronautique. Ainsi EADS a développé une ontologie comportant plusieurs centaines de compétences employées dans leur domaine.

Après discussion avec EADS, nous avons trouvé qu'ils avaient eux aussi des compétences en bureautique, très proches de celles du C2I. Nous avons alors créé une correspondance (« mapping ») entre les concepts semblables. Ce mapping correspondait tout simplement à un tableau comme celui-là :

Compétences C2I	Compétences en Bureautique pour EADS
b4textprocessor	eadsCompetencyDefinition_2076
b4spreadsheet	eadsCompetencyDefinition_2077
...	...

Un Web Service Sémantique implémente ces relations de similarité et sert de dictionnaire pour d'autres Web Services ayant besoin de ces équivalences.

Faisabilité à grande échelle de ces traductions. La traduction manuelle de concepts est une tâche pouvant prendre beaucoup de temps et de moyens. Deux critères nous poussent pourtant à croire au possible développement de ces Web Services de « traduction ».

- Les ontologies sont des représentations assez stables. Par conséquent, une fois l'équivalence établie entre deux représentations, ces liens entre concepts ne seront pas à remettre à jour fréquemment.
- Si les services de traduction entre ontologies permettent à un moteur de recherche de découvrir de nombreuses ressources d'une autre base (aux concepts légèrement différents), ceci peut être économiquement très intéressant. Par conséquent, les Web Services Sémantiques de traduction entre ontologies pourraient être payants et faciliter ainsi leur diffusion.

Limites sémantiques à explorer. Dans notre contexte, nous avons juste fait le lien entre des concepts similaires. Dans un travail plus approfondi, nous aurions dû estimer si les niveaux pédagogiques demandés par le C2I universitaire et par EADS étaient semblables ou bien s'il fallait effectuer une équivalence de niveaux. Néanmoins, la traduction entre concepts peut être une approche utile à explorer dans le domaine pédagogique, tout en conservant en mémoire ses limites.

3.2.2.6 Explication des résultats sémantiques

Problème : les moteurs de recherches effectuent des processus de plus en plus complexes pour répondre aux requêtes des utilisateurs. Pourtant cette complexité est parfaitement dissimulée, donnant un semblant extérieur de simplicité. Le moteur Google symbolise parfaitement ces facteurs. Des études comme [Teevan et al., 2004] montrent que lorsque les utilisateurs ne trouvent pas les ressources répondant à leurs requêtes sur des moteurs généralistes, ils ont appris à exploiter les informations de la page de résultats pour reformuler leur requête et tomber, par essais successifs, sur les ressources souhaitées. Certains parlent alors d'une forme de communication évoluée entre le système « moteur de recherche » et l'utilisateur, chacun essayant de deviner le fonctionnement de l'autre pour adapter ses messages. Or pour aider l'utilisateur à comprendre les résultats du moteur de recherche, une perspective intéressante serait, non pas d'afficher la complexité brute sous-jacente mais de donner certains indices à l'utilisateur.

Intérêt de ce problème : les moteurs répandus, comme Google, sont des moteurs avant tout statistiques qui analysent des graphes de sites Web pour calculer des effets de popularité. Ces approches numériques semblent plus difficiles à « montrer »³⁹ que les approches des moteurs comme LUISA, fondés sur des représentations symboliques de la connaissance, les ontologies.

Solution proposée : Nous proposons d'afficher des extraits des raisonnements du moteur sémantique sur la page des résultats au dessus des documents. Nous avons prévu deux vues possibles : simplifiée (par défaut) et avancée.

L'interface de cette fonctionnalité n'a pu malheureusement être implémentée à temps dans le prototype. C'est pourquoi nous ne pourrions pas en montrer de copie d'écran.

La vue « Simplifiée » correspond à quelques lignes affichées en haut de page, au dessus des documents trouvés suite à la requête :

Explication des résultats : Objets impliqués par votre requête : <i>Compétence « B1 », Sous-compétence « Maîtriser la gestion de fichiers », Logiciel « Navigateur » ...</i> <u>Plus d'explications</u>

En cliquant sur « Plus d'explications », le rectangle s'agrandit et affiche des informations plus complètes sur la manière dont les résultats ont été trouvés. Pour le prototype LUISA, notre vision était assez basique avec uniquement l'affichage des logs de raisonnements effectués par le moteur sémantique.

Comme nous l'avons présenté plus haut (§3.2.2.3), le moteur applique un certain nombre de règles lorsqu'il recherche les documents. Les principales sont la décomposition de compétences, la traduction de concepts ou le passage d'une compétence à l'outil employé, afin d'augmenter le panel de ressources trouvées (avant ordonnancement par les contraintes).

Voici une vue possible des explications avancées :

³⁹Le moteur Wikia affichait les statistiques employées pour chaque requête mais le projet fut arrêté en mars 2009.

Explication des résultats :

Objets impliqués par votre requête : *Compétence « B1 », Compétence 2082 (EADS), Logiciel « Système d'exploitation » ...*

Explications avancées (logs)

1. Recherche selon la Compétence B1 dans l'ontologie de Compétence (30 LOs trouvés)
2. Traduction de concept dans ontologie EADS : Compétence 2082
3. Recherche selon la Compétence 2082 dans l'ontologie EADS (10 LOs trouvés)
4. Logiciel employé par compétence B1 : « Système d'exploitation » (ontologie de logiciels)
5. Recherche selon le Logiciel « Système d'exploitation » (15 LOs trouvés)
6. ...

Fermer

Par ce biais, les utilisateurs peuvent de temps en temps voir qu'aucune ressource n'a été trouvée et comprendre que les résultats sont limités par une option de leurs préférences.

Cette fonctionnalité semble un moyen utile pour inciter les utilisateurs à comprendre les représentations du domaine qu'emploie le système.

Le point critique de cette fonctionnalité est néanmoins le choix des informations à afficher aux utilisateurs. En effet, un code comme « Compétence 2082 » n'est pas très explicite. Les perspectives pourraient être de mettre un hyperlien sur le nom des concepts renvoyant vers de la documentation présentant succinctement les ontologies employées. Une autre approche pourrait consister à disposer des images ou des animations, présentant des extraits utiles de l'ontologie par rapport à la requête (e.g. montrant que la compétence « B1 » emploie des logiciels tels que « Système d'exploitation », « Gestionnaire de fichiers », etc.)

3.2.2.7 Composition de paquets de ressources

Problème : si de plus en plus de ressources sont diffusées sur le Web, trouver de « bonnes » ressources n'est pas suffisant. Dans une optique de « conception pédagogique », l'idéal est de trouver des ressources qui vont bien ensemble pour atteindre un objectif pédagogique. Est-il possible de formaliser des règles et des méthodes pédagogiques ([Reigeluth, 1999]) pour aider à la sélection de ressources pédagogiques puis à la création de paquets de ressources répondant à un objectif pédagogique ?

Intérêt de ce problème : des standards comme IMS-LD cherchent à décrire des scénarios pédagogiques. Des auteurs pensent que le chemin n'est pas loin pour y ajouter des méthodes de création automatique ([Sicilia, 2007]) de paquets.

Note : attention, le but final n'est pas de fournir des paquets pédagogiques, créés automatiquement, directement aux étudiants mais de les proposer aux concepteurs pédagogiques.

Solution proposée : suite à une recherche ayant fourni des ressources pour une ou des compétences données, le système propose aux enseignants quelques règles pédagogiques ainsi que des paquets formés automatiquement à partir des ressources trouvées précédemment et des règles paramétrées.

De manière pratique, suite à une recherche de ressources, le système propose aux enseignants un hyperlien « Laboratoire d’assemblage » (« Composition Lab »). En cliquant sur cet hyperlien, l’enseignant trouve une interface telle que celle montrée dans la Fig. 3.11.

The screenshot displays a web interface for configuring educational packages. It is divided into three main sections:

- Your selection:** Lists five selected resources with their details:
 - VR5 - Presentation du tableur OpenCalc: lecture, FrenchLanguage, Competencies: -
 - VR4 - exercices about using a spreadsheet tool (Excel 2003): exercise, FrenchLanguage, Competencies: -
 - VR1 - Exercices sur un tableur (OpenCalc): exercise, FrenchLanguage, Competencies: -
 - Function IF *: exercise, FrenchLanguage, Competencies: Basic Spreadsheet (B4-4)
 - VR7 - Test sur OpenCalc: exam, Language: I, Competencies: -
- Rules:** A configuration panel with the following settings:
 - Number Of Resources: min: 3, max: 3
 - Number Of Resources Lessons: 1
 - Number Of Resources Exercises: 2
 - Number Of Resources Exams: 0
 - Learning style: classic (selected), discovery
 - Consistency In Repository: yes, no (selected)
 - Consistency In Language: yes, no (selected)
 - Language: French (dropdown menu)
 - An 'Apply' button is located at the bottom of this section.
- Composed packages:** Shows two generated packages:
 - Package 0: VR4 - exercices about using a spreadsheet tool (Excel 2003), VR1 - Exercices sur un tableur (OpenCalc), VR5 - Presentation du tableur OpenCalc. Includes a 'Save' button.
 - Package 1: VR1 - Exercices sur un tableur (OpenCalc), Function IF *, VR5 - Presentation du tableur OpenCalc. Includes a 'Save' button.

FIG. 3.11 – Ecran affichant les règles permettant de composer des paquets pédagogiques (Prototype UHP-LUISA)

Le côté gauche montre les ressources préalablement sélectionnées dans les résultats de la recherche. Le côté droit montre les paramètres des règles régissant la composition des paquets et devant être complétés par l'utilisateur. En bas se trouvent les paquets proposés au format SCORM (voir §1.2.2).

Dans notre prototype, les règles et leurs paramètres sont expérimentaux, loin de correspondre aux besoins d'enseignants. Néanmoins elles appliquent tout de même des règles pédagogiques très simples.

Les quatre premiers paramètres à remplir correspondent à la composition désirée du paquet,

à savoir le nombre total de documents et parmi eux le nombre de documents de cours (lesson), d'entraînements (exercice) et d'évaluation (exam). Ces valeurs sont celles employées lors de l'indexation des documents du C2I et reprennent le vocabulaire du standard LOM.

Le paramètre « Learning Style » est une règle qui définit si les exercices sont situés après le cours (« classic ») ou avant le cours (« discovery »).

Si la cohérence dans le répertoire (« consistency in repository ») est activée, le système va former des paquets dont toutes les ressources proviennent du même entrepôt de Learning Objects (LOR). L'idée est que par exemple des ressources d'une même université ont plus de chance de posséder un contexte commun, plutôt qu'un paquet composé de ressources issues d'universités différentes. En revanche, un concepteur pédagogique peut préférer avoir des ressources provenant de sources différentes apportant une autre forme de richesse pédagogique.

Enfin, si la cohérence dans la langue (« consistency in language ») est activée, les paquets proposés seront exclusivement composés de ressources rédigées dans la même langue. Ceci dans le même objectif de cohérence que pour les répertoires.

3.2.2.8 Difficultés pour créer des raisonnements sémantiques

De notre expérience, nous tirons la difficulté de créer des raisonnements sémantiques qui permettront des fonctionnalités efficaces de manipulation des Learning Objects. L'élaboration des raisonnements sémantiques nous semble devoir s'effectuer de façon itérative car les raisonnements a priori pour simuler des procédés pédagogiques ont toutes les chances d'être simplistes et anecdotiques. Or dans le cas du projet LUISA, le temps entre l'idée d'un raisonnement sémantique et son implémentation était beaucoup trop long, et passait par trop d'intermédiaires, pour permettre des améliorations progressives.

3.2.3 Ontologies développées

Les ontologies sont les connaissances formalisées, à l'aide desquelles seront indexées les ressources pédagogiques. Elles seront aussi la base sur laquelle seront appliquées les raisonnements automatisés.

Dans le projet LUISA, afin de répondre aux besoins du contexte du C2I et afin d'explorer les possibilités des technologies du Web Sémantique, nous avons travaillé sur quatre ontologies. D'une part, nous avons repris et adapté deux ontologies, respectivement portant sur les documents pédagogiques (S-LOM) et sur les compétences (GCS). Elles ont été développées à l'origine par le partenaire Université de Alcalá. D'autre part, nous avons développé deux ontologies, de disciplines et de logiciels, pour le contexte de notre expérimentation, le C2I à l'université Henri Poincaré.

A la fin de cette section, nous détaillons la procédure d'indexation ayant employé ces ontologies.

3.2.3.1 Présentation générale des ontologies

Avant de détailler chaque ontologie, voici une présentation globale de leur fonctionnement et de leurs interrelations (Fig. 3.12).

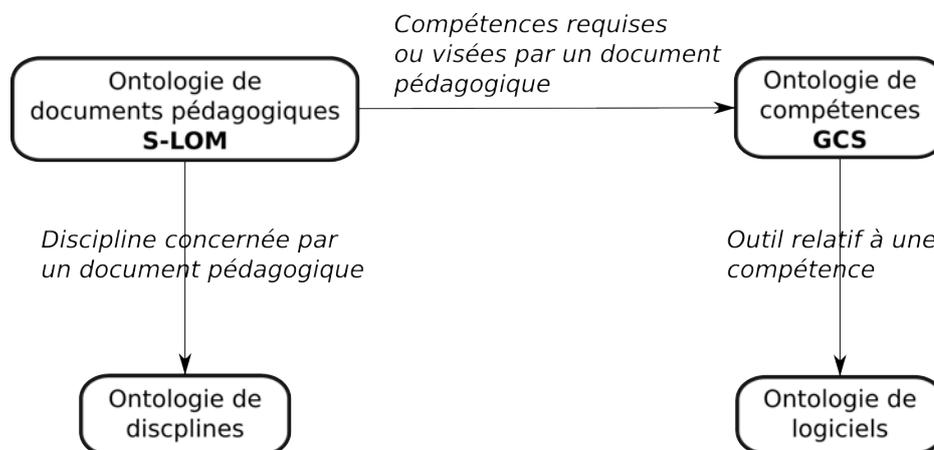


FIG. 3.12 – Relations entre les 4 ontologies du projet LUISA.

L'ontologie principale s'appelle S-LOM comme « Semantic LOM » et est une version sémantique de LOM, le standard pour décrire les documents pédagogiques (§ 1.2.2). Par exemple, au lieu de décrire une zone culturelle (champ « 1.6 Coverage » de LOM) avec simplement du texte libre, S-LOM permet de pointer vers des concepts pour définir une zone géographique ou une période temporelle.

La seconde ontologie d'importance est la General Competency Schema (GCS) qui offre un cadre conceptuel pour représenter les compétences.

Celles-ci peuvent être nécessaires à un emploi ou à un diplôme scolaire. GCS demande à être adapté pour chaque contexte. Une fois cette adaptation effectuée, GCS permet par exemple d'exprimer de manière formelle qu'un utilisateur « Georges » possède une compétence « Réaliser un diaporama » avec un niveau « Confirmé ».

S-LOM et GCS sont liées car pour décrire les compétences requises ou visées par un document pédagogique (catégorie « Classification » de LOM), S-LOM va permettre d'utiliser les compétences exprimées avec GCS.

Considérons qu'une documentaliste veuille décrire et indexer une série d'exercices entraînant à utiliser un outil de diaporama comme Powerpoint. Grâce à S-LOM et GCS, elle pourra exprimer que la compétence qui permet d'atteindre ces exercices est la compétence « Réaliser un diaporama » avec un niveau « Débutant ». Au moment d'une recherche, le système pourra en déduire automatiquement qu'il est inutile de proposer ces exercices à des utilisateurs, comme « Georges », possédant la compétence « Réaliser un diaporama » avec un niveau « Confirmé ».

L'ontologie de discipline, que nous avons développée, représente la structure des disciplines au sein de l'université Henri Poincaré. Pour décrire la discipline traitée par un contenu pédagogique (catégorie « Classification » de LOM), S-LOM utilisera les concepts issus de cette ontologie de discipline.

Enfin, l'ontologie de logiciels représente les logiciels employés dans la préparation au C2I (e.g. les traitements de texte ou un logiciel de mail). Cette ontologie est reliée à l'ontologie de compétences (GCS) car cette dernière permet de relier une compétence à un outil. Ceci sera

effectué en pointant encore vers l'ontologie de logiciel.

Par exemple, la compétence « Réaliser un diaporama » de GCS est liée au concept « Outils de diaporama » de l'ontologie de logiciels.

Voyons maintenant le détail de ces ontologies en commençant par l'ontologie S-LOM décrivant les documents pédagogiques.

3.2.3.2 Ontologie de Learning Objects : S-LOM

Mapping de LOM. S-LOM (Semantic LOM) est une version sémantique de LOM (§1.2.2). L'intention sous-jacente est donc aussi de décrire un Learning Object.

S-LOM résulte de l'expression du standard LOM en WSML (§2.1.1). On parle de « mapping ». Ce mapping fut réalisé par le partenaire Université d'Alcalá ([LUISA-Consortium, 2007]).

Le développement de cette ontologie respecta les principes suivants :

1. être compatible avec LOM dans le sens que tout champ LOM doit pouvoir être traduit dans l'ontologie en WSML,
2. conserver la terminologie LOM autant que possible,
3. fournir le plus de sémantique possible manipulable informatiquement.

Description. L'ontologie S-LOM est centrée autour du concept de « LearningObject » et chaque champ LOM correspond aux attributs d'une instance de ce concept.

Par exemple, le champ « Coverage » (traduisible par « Couverture ») de LOM devient dans S-LOM l'attribut « Coverage » du concept « LearningObject ». Dans LOM, ce champ est défini par :

« L'époque, la culture, l'emplacement géographique dans lesquels le Learning Object est employé. Ceci inclura typiquement la localisation spatiale (le nom d'un lieu ou les coordonnées géographiques), la période temporelle (nom d'une période, une date ou un écart entre deux dates) ou une juridiction (comme le nom d'une entité administrative). »

La valeur classique de LOM pour ce champ est une chaîne de caractère, c'est-à-dire du texte libre. S-LOM permet toujours cette description sous la forme d'un texte libre mais offre en plus la possibilité d'utiliser des représentations sémantiques issues de l'ontologie OpenCyc⁴⁰. S-LOM permet ainsi à un LearningObject de posséder un attribut nommé « Coverage » pointant vers les concepts OpenCyc de « GeographicalRegion » pour la localisation spatiale, « TimeInterval » pour la période temporelle et « AdministrativeUnit » pour l'unité administrative.

Alors qu'une chaîne de caractère ne permet que des traitements automatisés limités, la possibilité de pointer vers des concepts d'ontologies permet des traitements automatisés plus puissants. Par exemple, si un utilisateur recherche les ressources couvrant la France, il ne trouvera pas une ressource dont la Couverture mentionne dans une chaîne de caractères le terme « Région parisienne ». En revanche, si le champ « Couverture » pointe vers le concept de « Région parisienne » dans une ontologie géographique, un raisonnement automatisé peut retrouver sans

⁴⁰OpenCyc est une très grande ontologie généraliste (plusieurs milliers de concepts représentés) : www.cyc.com/cyc/opencyc

trop de difficultés que la région parisienne est située en France.

Lien avec les compétences. Dans LOM classique, le champ 9 « Classification » permet de pointer vers des valeurs externes à LOM par exemple pour les compétences. Dans S-LOM, les valeurs décrivant les compétences requises ou visées par un document pédagogique sont issues de l'ontologie GCS.

3.2.3.3 Ontologie de compétences : GCS

Sémantisation du modèle HrXML. GCS (General Competency Schema) est une ontologie du domaine des compétences, développée par l'université d'Alcala et détaillée pour la première fois dans [Sicilia, 2005].

GCS étend et sémantise le modèle HrXML (Human Resource XML) issu d'un consortium⁴¹ développant des spécifications XML liées aux ressources humaines dans les entreprises.

Vue globale. La Fig. 3.13 montre une vue globale du fonctionnement de GCS.

Sur la gauche, le concept de *Processor* représente une entité possédant des compétences. Généralement, ceci représente un individu mais rien n'empêche qu'une organisation ou un agent virtuel soit aussi un *Processor*. Pour simplifier, considérons ici l'individu Paul, instance de *Person* héritant de *Processor*.

Sur la droite, le concept de *CompetencyDefinition* représente une compétence de manière abstraite. Par exemple, considérons l'instance de *CompetencyDefinition* « rédiger un mémo ». C'est une compétence abstraite dans le sens où elle ne porte sur personne. C'est juste une définition.

Paul peut posséder la compétence « rédiger un mémo ». Dans ce cas, il existera une relation *isAbleToPerform* entre Paul et une compétence, cette fois concrète, « rédiger un mémo » instance de *Competency*.

La différence entre une compétence concrète (*Competency*) et une compétence abstraite (*CompetencyDefinition*) est qu'une compétence concrète est associée à un utilisateur et possède un niveau (*level*). Par exemple, Paul sait « rédiger un mémo » de manière experte qui correspondrait au *level* 3 d'une échelle définie par ailleurs.

La figure montre que les concepts *CompetencyDefinition* possèdent une relation *requires* pointant vers eux-mêmes. Cette relation représente la dépendance entre compétences. Par exemple, la *CompetencyDefinition* « rédiger un mémo » peut être nécessaire pour acquérir la *CompetencyDefinition* « diriger une équipe ». Dans ce cas il y aurait une relation « diriger une équipe » « requires » « rédiger un mémo ».

Il existe une seconde relation entre *CompetencyDefinition* mais qui n'apparaît pas sur la Figure, la relation *details*. Quand des *CompetencyDefinition* détaillent (*details*) une *CompetencyDefinition*, cela signifie que cette dernière se compose des premières. Cette relation sera bientôt illustrée dans notre adaptation de GCS pour le C2I.

⁴¹ Consortium HR-XML : <http://www.hr-xml.org>

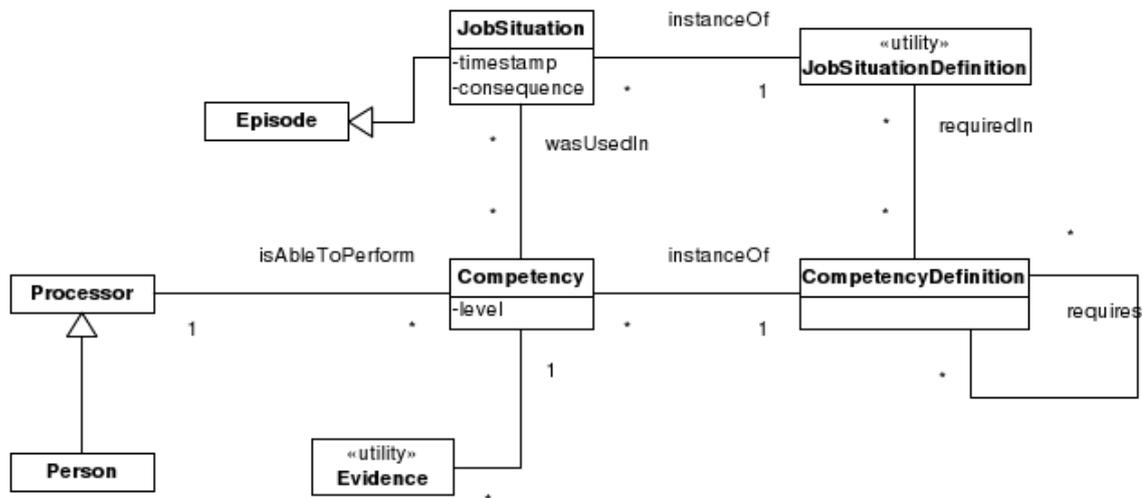


FIG. 3.13 – Ontologie de compétences, vue générale - [Sicilia, 2005]

Détail d'une compétence abstraite. Le schéma 3.14 présente une vue centrée sur le concept de compétence abstraite (*CompetencyDefinition*) qui se décompose en morceaux de compétences, *CompetencyElement*, de trois types : *KnowledgeElement*, *Attitude* et *Skills*.

Par exemple, « rédiger un mémo » pourrait se décomposer en « connaître les règles du mémo » et « être capable de rédiger un résumé de 3 lignes ». « connaître les règles du mémo » serait un élément de connaissance (*KnowledgeElement*) tandis que « être capable de rédiger un résumé de 3 lignes » serait un savoir-faire plus pratique (*Skills*). L'*Attitude* serait plutôt un savoir-être comme « être impartial » ou « être audacieux ». Nous n'avons pas exploré cet aspect dans LUISA.

La différence entre des *CompetencyDefinition* et des *CompetencyElement* est une question de granularité. Les concepteurs de l'ontologie considère que les *CompetencyDefinition* sont gérées par les ressources humaines d'une entreprise, tandis que les morceaux de compétences sont gérés par le formateur.

L'ontologie offre aussi un moyen de lier une connaissance (*KnowledgeElement*) à un concept d'une autre ontologie. Par exemple, la connaissance « connaître les règles du mémo » pourrait être liée (relation *about*) à un livre traitant de stylistique, instance dans une ontologie bibliographique.

Une relation de cette ontologie est floue à cause de deux relations portant le même nom (*requires*) alors qu'elles relient des concepts différents (entre *CompetencyDefinition*, et entre *CompetencyDefinition* et *CompetencyElement*). Cette incohérence, révélée durant le projet, ne put être corrigée avant la fin.

Adaptation. GCS considère les compétences indépendamment des contextes d'application. Au sein du projet LUISA, notre travail a donc consisté à spécialiser cette ontologie pour le domaine du C2I. Pour cela, nous sommes partis du référentiel de compétences du C2I (§3.2.1.2).

Le référentiel du C2I est constitué de compétences principales (e.g. « B2 : Recherche d'information ») et de compétences secondaires (e.g. « B2.2 : Concevoir une requête »). Toutes ces

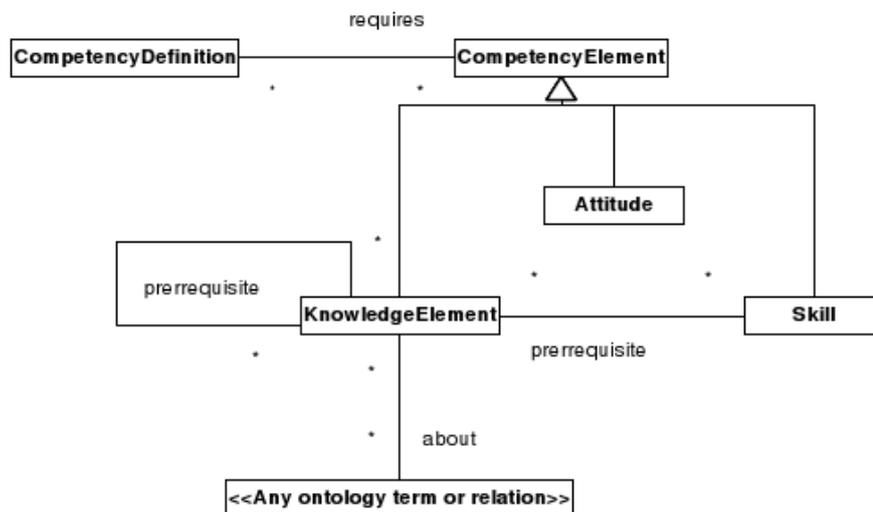


FIG. 3.14 – Ontologie de compétences, vue centrée sur les CompetencyDefinition - [Sicilia, 2005]

compétences sont abstraites dans le sens où elles ne sont liées à aucun individu en particulier, nous les avons donc représentées par des instances de *CompetencyDefinition*.

La différenciation entre compétences principales et secondaires se fait par la relation « details ». Par exemple, les compétences secondaires « B2.1 », « B2.2 » et « B2.3 » détaillent (*details*) la compétence principale « B2 ».

Certaines compétences requièrent d’en maîtriser d’autres auparavant. Cette relation de dépendance n’existe pas explicitement dans le référentiel du C2I mais nous avons pu l’extraire. Par exemple, la compétence « B1 : Maîtriser son environnement de travail » est une compétence de base nécessaire à toutes les autres compétences. Donc, dans notre adaptation de l’ontologie GCS, les compétences « B2 » à « B7 » requièrent « B1 ».

Quand une compétence est concrète (*competency*), elle est liée à un processeur (*processor*) et possède un niveau (attribut *level*).

Dans notre spécialisation de GCS, les *processor* sont les étudiants se préparant au C2I.

Les niveaux de compétences sont représentés par une échelle simple dans LUISA. Le but de cet échelle est, avant tout, de tester les possibilités des outils sémantiques pour raisonner sur des niveaux de compétences :

- Niveau 0 : l'utilisateur ne possède pas du tout la compétence donnée,
- Niveau 1 : l'utilisateur maîtrise en partie la compétence,
- Niveau 2 : l'utilisateur possède la compétence au niveau C2I,
- Niveau 3 : l'utilisateur maîtrise une compétence bien mieux que ce qui est requis pour le C2I.

Des vues plus détaillées sur les relations *details* et *requires*, ainsi qu’un extrait de code WSML se trouvent en annexe E.

3.2.3.4 Ontologie de disciplines

Une modeste ontologie. L'ontologie de discipline fait partie des deux ontologies, avec celle de logiciel, construites spécialement pour le C2I à l'université Henri Poincaré. Ceci signifie que la motivation sous-jacente n'est pas qu'elle soit diffusée au maximum et ré-employée, comme S-LOM ou GCS, mais juste qu'elle soit utile pour notre contexte expérimental.

Ré-utilisation d'OpenCyc. Cette ontologie représente les disciplines de l'université Henri Poincaré à l'aide du concept de *FieldOfStudy*. Ce concept reprend en fait le concept de *FieldOfStudy* d'une ontologie généraliste nommée OpenCyc contenant des milliers de concepts. Néanmoins, nous ne reprenons pas la structure des disciplines d'OpenCyc qui ne correspond pas exactement à celle de l'université Henri Poincaré.

Sous-disciplines. Basiquement, une discipline (*FieldOfStudy*) peut se décomposer en sous-disciplines. Par exemple, le domaine de la « Santé » contient ceux de « Médecine » et de « Pharmacie ». Dans l'ontologie, ceci est représentée par un concept *Health* qui est un sous-concept de *FieldOfStudy*, puis par les concepts *Medicine* et *Pharmacology* eux-mêmes sous-concepts de *Health*.

Relation « LinksTo ». De plus, certains liens existent entre disciplines en dehors de la notion de sous-disciplines. Par exemple, les sciences physiques font parfois appel à des théories mathématiques avancées et le cursus universitaire en sciences physiques contient des modules de mathématiques. Or, ceci n'est pas représentable par la relation de sous-disciplines.

Dans l'ontologie de discipline, nous avons donc créé la relation « linksTo » pour représenter ce lien entre disciplines. Cette relation est asymétrique, ce qui signifie que les sciences physiques nécessitent un apport des mathématiques, mais non l'inverse.

Structure générale. A partir des composantes de l'université Henri Poincaré⁴², nous avons réalisé la structure suivante :

⁴²Composantes de l'UHP : http://www.uhp-nancy.fr/universite/composantes_uhp

- FieldOfStudy
 - Mathematics-Computer Science
 - Mathematics
 - Computer Science
 - Physics and Chemistry sciences
 - Physics (linked to Mathematics)
 - Biophysics
 - Chemical Engineering
 - Health
 - Medicine (linked to Biology)
 - Pharmacology (linked to Biology)
 - Dentistry
 - Ergonomics
 - Life sciences
 - Biology
 - Biogeography
 - Sport sciences

3.2.3.5 Ontologie de logiciels

Comme le C2I porte sur le domaine de l'informatique et d'Internet, nous avons eu besoin de créer une ontologie des logiciels et matériels informatiques pour tester des raisonnements sémantiques autour des ressources du C2I.

Une ontologie nouvelle. Avant de construire une nouvelle ontologie, même modeste par sa taille, nous avons cherché à réutiliser des modèles déjà existant⁴³.

Pendant nous n'avons trouvé aucun modèle correspondant directement à nos besoins. Par exemple, il existe la « *Ontology of Computing Services* »⁴⁴. Mais si cette ontologie contient des concepts très précis, comme *ImageResolution*, elle ne contient pas les quelques propriétés de base que nous désirons comme le coût d'un logiciel ou l'interopérabilité avec les systèmes d'exploitation. Lier nos concepts à cette ontologie aurait demandé un travail important pour rester cohérente avec elle et finalement ne nous aurait fourni aucun avantage car nous n'avions pas besoin des relations ou propriétés qu'elle définit.

Par conséquent, nous avons développé les quelques concepts et relations dont nous avons besoin en nous affranchissant des ontologies pré-existantes, non adaptées à notre contexte. Ceci a donné naissance à l'ontologie appelée « *ComputerLiteracy* » d'une trentaine de concepts.

⁴³Méthodologie pour construire une ontologie : http://protege.stanford.edu/publications/ontology_development/ontology101-noy-mcguinness.html

⁴⁴Ontology of Computing Services, reliée à l'ontologie de haut niveau SUMO (www.ontologyportal.org/)

Hierarchie de logiciels. Cette ontologie a été réalisée manuellement⁴⁵ à partir des types de logiciel mentionnés dans le référentiel du C2I (Annexe C).

L'ontologie a une structure assez simple et de type hiérarchique (Fig. 3.15). Le concept racine est celui de `ComputerLiteracyItem` (« ComputerLiteracy » est le nom anglais de l'ontologie) dont hérite les concepts de « Software » et de « Hardware ».

Relations d'interopérabilité. Pour pouvoir expérimenter des raisonnements plus complexes sur l'ontologie, nous avons ajouté une relation « `interoperableWith` » reliant un Software à un autre concept de l'ontologie. Par exemple, le logiciel Microsoft Word XP est interopérable avec le système d'exploitation Windows XP mais non avec un système Linux.

Cette relation d'interopérabilité traverse l'ontologie et n'aurait pu être représentée dans une simple taxinomie hiérarchique. Cette relation a permis d'expérimenter certains raisonnements sémantiques dans LUISA mais pose en soi un problème car identifier toutes les interopérabilités possibles est une tâche lourde. Dans LUISA, cette relation ne concerne en définitive que quelques applications et systèmes d'exploitation.

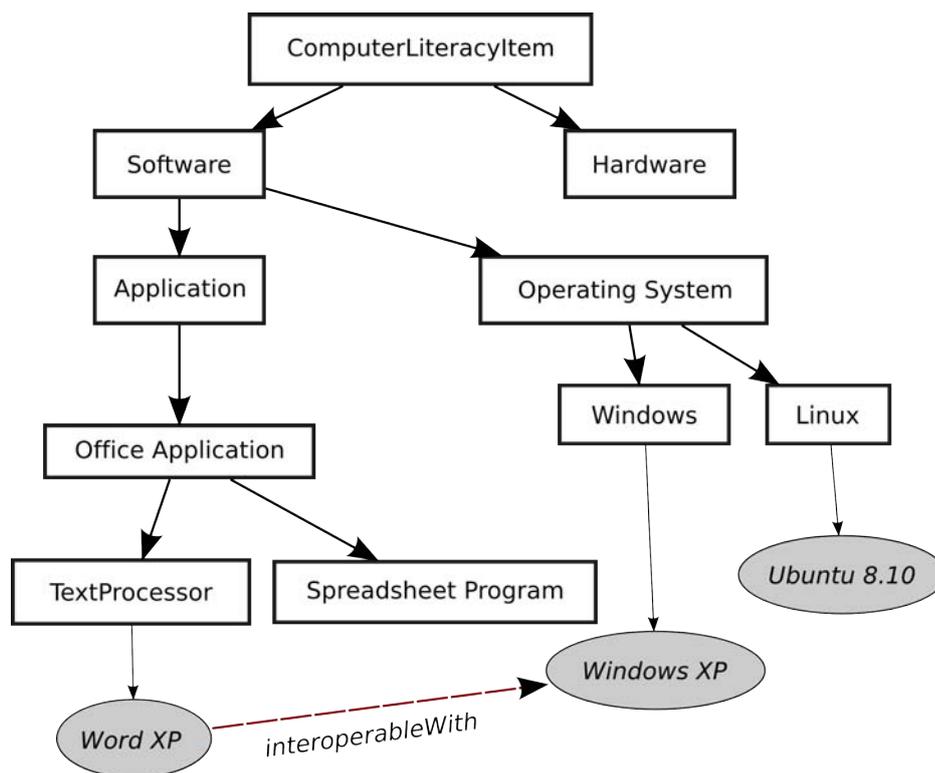


FIG. 3.15 – Extrait de l'ontologie de logiciel. Les traits épais sont des relations « `subConceptOf` », les traits fins des instanciations et le trait pointillé représente une relation d'interopérabilité, transversale à la hiérarchie.

Une vue globale des concepts et un extrait du code WSMML sont fournis en annexe F.

⁴⁵Nous avons découvert à la fin du projet LUISA une autre ontologie informatique, réalisée à partir d'analyses linguistiques automatiques : www.let.uu.nl/lt4e1/ (projet européen Lt4eL)

3.2.3.6 Indexation des ressources C2I à l'aide des ontologies

Modélisation et indexation en parallèle. Bien que nous présentons l'indexation après la construction des ontologies, ces deux processus furent effectués parallèlement. En effet, la construction des ontologies nécessitait de les tester en indexant quelques ressources, puis de modifier les ontologies, puis de ré-indexer quelques ressources, et ainsi de suite.

Provenance des ressources. Avec l'aide d'un stagiaire, nous avons indexé 160 ressources portant sur des sujets traités par le C2I. Ces ressources proviennent de :

- modules C2I de différents départements de l'université Henri Poincaré (Annexe D),
- ressources C2I du site C2IMES.org conseillé par le ministère,
- différents sites Web proposant des tutoriaux et des activités en ligne pour l'apprentissage de l'informatique en général.

Outil d'annotation sémantique : eLUISA. A partir des ontologies conçues et présentées dans la section précédente, nous avons travaillé avec les partenaires techniques afin de les intégrer au sein d'un outil d'annotation sémantique (§2.1.2.2) en ligne : eLUISA.

L'écran 3.16 montre un extrait du logiciel d'indexation dans sa première version. eLUISA est fondé sur l'outil SHAME⁴⁶ développé par l'université d'Uppsala en Suède. SHAME est un outil visant à être le plus générique possible grâce à des profils d'annotation. Un profil d'annotation est créé par un administrateur qui sélectionne des métadonnées dans des ontologies et qui choisit leur visuel dans l'outil d'annotation.

Par exemple, l'administrateur de eLUISA a créé un profil d'annotation permettant de décrire une ressource avec un champ « titre » qui est interprété comme un « titre » au sens de l'ontologie S-LOM (§3.2.3.2) et est représenté par un champ texte simple. Il a aussi ajouté un champ « SemanticTopic » qui pointe vers l'ontologie des logiciels (§3.2.3.5). Il a configuré ce champ pour qu'il soit une liste déroulante dont les valeurs sont issues des instances de l'ontologie (Fig. 3.16).

SHAME permet ainsi de créer de nouveaux profils d'annotation, à partir de « morceaux » d'ontologies différentes. Le but est de supporter des formes plus flexibles d'annotation qui pourraient être adaptées aux différentes utilisations, sans forcer par exemple à remplir tous les champs LOM, si les utilisateurs savent qu'ils ne seront pas tous exploitables plus tard.

Difficultés dans la création et manipulation des ontologies. Les ontologies forment le cœur des technologies du Web Sémantique. Or, si les possibilités de leurs représentations formelles sont séduisantes, leur création n'en reste pas moins un acte difficile dans un projet à plusieurs.

Tout d'abord, dans le projet, la création des ontologies avait été prévue bien avant la réalisation du prototype. Par conséquent, nous avons dû créer et adapter les ontologies sans pouvoir les expérimenter directement, c'est-à-dire sans pouvoir effectuer de requêtes sur les bases de documents pour tester les raisonnements. Ceci nous a semblé une grande difficulté des modèles a priori. Même si nous avons pu modifier un peu les ontologies par la suite, la **flexibilité des**

⁴⁶Site de SHAME : <http://kmr.nada.kth.se/shame/wiki/Overview/Main>

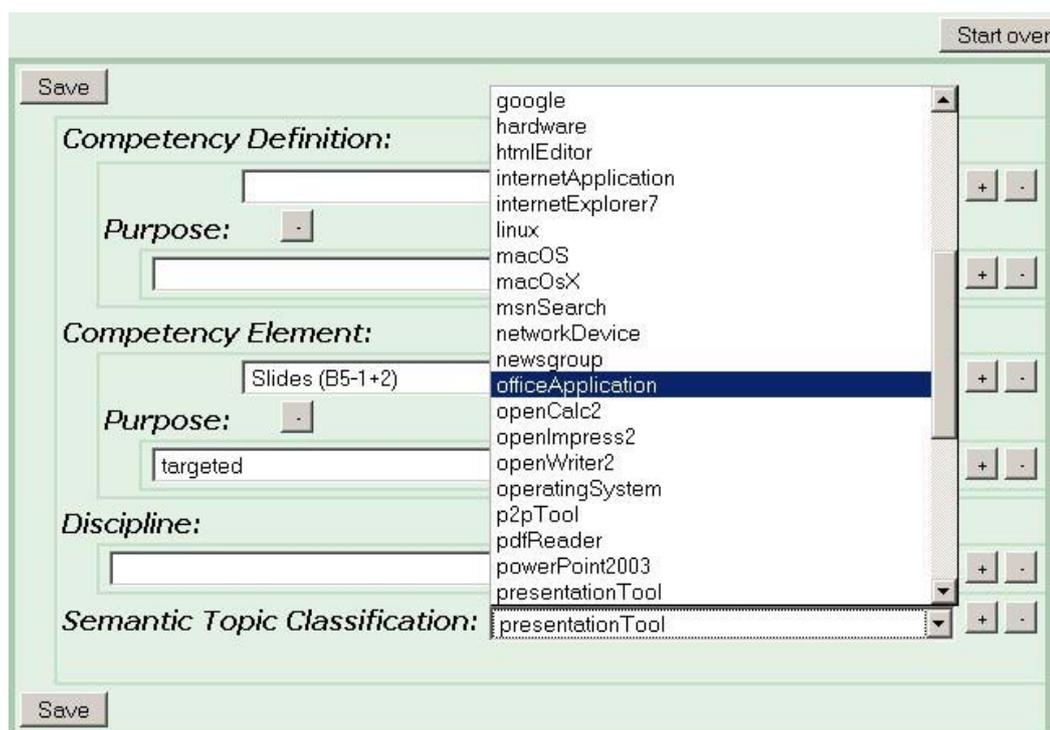


FIG. 3.16 – Ecran de l’outil d’annotation eLUIA. L’administrateur de l’outil a configuré un profil d’annotation définissant les champs, leurs valeurs possibles issues des ontologies, et leurs types d’affichage. Ainsi, le champ « SemanticTopic Classification » est une liste déroulante contenant les instances de l’ontologie de logiciel.

ontologies était réduite. Nous aurions aimé un processus de création des ontologies plus itératif.

Un autre problème rencontré fut la **difficulté de se comprendre autour des ontologies** avec les partenaires européens de LUISA. Malgré trente mois de travail en commun, nous ne sommes pas certains que tous les partenaires aient une vision identique des ontologies produites.

Ces incompréhensions peuvent résulter de problèmes inter-linguistiques, d’une mauvaise communication de la part des créateurs des ontologies ou de lacunes dans la gestion globale du projet. Néanmoins, ceci renforce l’idée que les structures formelles un peu évoluées sont rapidement des objets complexes ; elles sont rarement compréhensibles des humains à la simple vue d’un schéma.

Les difficultés que nous avons rencontrées dans la création des ontologies ne sont pas des cas isolés. Ce processus de création des ontologies est un point vital du Web Sémantique où des travaux nous semblent encore nécessaires, à l’image par exemple de [Karapiperis and Apostolou, 2006] ou [Braun et al., 2007] qui étudient la construction itérative et collaborative d’ontologies. Ces aspects importants seront ré-abordés plus tard avec l’approche du Web Participatif et de la fusion des Webs Sémantique et Participatif.

3.2.4 Expérimentations et discussion

Problèmes du prototype. En juin 2008, le prototype du projet LUISA est achevé. Néanmoins, ce prototype souffre de graves défauts :

- une instabilité forte, produite par des bugs et stoppant fréquemment le serveur exécutant le moteur LUISA,
- une lenteur importante lors des recherches sémantiques distribuées entre plusieurs Web Services Sémantiques, et empêchant des tests en situation réelle.

Ces difficultés ont empêché une évaluation satisfaisante du prototype et des concepts implémentés. Les évaluations furent avant tout les opinions et commentaires de membres de l’université suite à des démonstrations ou à des vidéos des prototypes LUISA.

Résultats. Voici un récapitulatif des résultats du projet LUISA dont le but était d’évaluer les possibilités des technologies du Web Sémantique pour la recherche de documents pédagogiques. Ces résultats sont issus principalement de notre expérience de conception (de la recherche d’un contexte d’expérimentation à l’observation des résultats implémentés, en passant par la conception des fonctionnalités) et des commentaires des membres de l’université lors des démonstrations du prototype final.

Les apports. Mesurer quantitativement l’apport d’une technologie est une tâche difficile, surtout sans avoir pu expérimenter d’outils en situation réelle. Nous ne pourrions pas en dire grand chose pour le projet LUISA.

D’un point de vue qualitatif, les différentes fonctionnalités conçues prouvent que les technologies du Web Sémantique offrent de nouvelles opportunités dans la manipulation de documents pédagogiques, aussi bien dans leur description que dans leur recherche.

Cependant, l’apport le plus important pour les utilisateurs potentiels n’est pas réellement un apport technique mais plutôt une approche. En effet, pour les enseignants et documentalistes de l’université Henri Poincaré, le point le plus positif mentionné lors des démonstrations de LUISA, fut **l’approche des documents par l’angle des compétences**, approche peu fréquente dans les universités, plus habituées aux approches thématiques et disciplinaires.

Cette approche par compétence n’est pas directement liée aux ontologies. Néanmoins, en y regardant de plus près, elle possède des points communs avec l’approche du Web Sémantique dans :

- l’explicitation de principes abstraits ;
- une approche dynamique des connaissances, considérées comme les jalons d’un chemin menant à des objectifs.

Pour l’instant, cette « nouvelle » manière de voir les ressources et les connaissances dans l’université est le principal apport des technologies du Web Sémantique, pour les utilisateurs. L’intérêt de cette approche est d’ailleurs renforcée par l’intérêt qu’y portent d’autres projets mêlant ontologies et compétences, comme le récent projet Intergéo portant sur le partage de constructions géométriques interactives ([Libbrecht et al., 2008]).

Les limites. Si les apports sont incertains, des difficultés sont clairement visibles et devront être résolues pour que les technologies du Web Sémantique puissent s'implémenter dans l'université.

Les **lenteurs des raisonnements sémantiques**, notamment lorsqu'ils sont distribués, ont alourdi les prototypes LUISA. Néanmoins, nous pouvons supposer que des projets moins expérimentaux et mieux optimisés pourront à moyen terme obtenir des résultats satisfaisants, au moins sur des ensembles limités de documents comme ceux produits par une université.

Les limites sur lesquelles nous voudrions insister sont celles de la conception des ontologies et des règles s'appliquant dessus. Les utilisateurs non-informaticiens avec qui nous avons travaillé dans l'étude des besoins ont eu beaucoup de mal à raisonner en terme d'ontologie. Nous mêmes avons eu des **difficultés à produire des ontologies et des règles sémantiques intéressantes**. Enfin, au sein du projet, entre membres tous informaticiens et pour beaucoup spécialistes des ontologies, les discussions autour des celles-ci furent difficiles, et leur manipulation fastidieuse.

Ces complications montrent les difficultés que ces objets soulèvent, notamment dans le milieu universitaire actuel.

Pourtant, sur de nombreux points, nous avons choisi des solutions simples voire parfois simplistes pour réussir à expérimenter les technologies du Web Sémantique :

- Nous avons employé le référentiel du C2I pour créer une ontologie de compétences alors que certaines compétences du C2I ne l'étaient pas vraiment,
- Nous avons créé des règles sémantiques un peu irréalistes sur les ontologies de logiciels,
- Nous avons créé des ontologies pour notre cadre limité et non dans un but de réutilisation.

Malgré tous ces efforts pour les rendre plus abordables et intégrables dans le monde réel, les ontologies pédagogiques sont restées lointaines et irréalistes.

C'est pourquoi les travaux sur la construction itérative et collaborative d'ontologie nous semblent un point majeur pour l'emploi réel d'ontologies pédagogiques.

Du Web Sémantique au Web Participatif. En réponse à cette difficulté de formaliser des ontologies, nos seconds travaux portent sur une forme d'indexation plus souple, réalisée par les utilisateurs de manière spontanée.

Deuxième partie

Web Participatif, communautés et indexation pédagogique

La première partie détaillait l'approche du Web Sémantique fondée généralement sur des ontologies réalisées par des experts. Cette approche permet d'améliorer certaines recherches mais pose des problèmes dans la création et l'évolution des représentations formelles partagées.

Cette seconde partie détaille l'approche du Web Participatif où les utilisateurs finaux des ressources participent à leur production et à leur organisation. La représentation formelle de quelques experts laisse place à des formes plus souples d'indexation des documents par des utilisateurs aux profils variés.

Par rapport à la partie précédente, le cadre de nos recherches évolue pour passer de celui d'un projet européen à celui d'une coopération informelle avec un doctorant en Sciences de l'Education, Eric Dané.

La participation des utilisateurs étant généralement liée à la notion de communauté, nous avons abordé avec ce doctorant une communauté d'enseignants en Génie Mécanique reliés par une liste de diffusion.

4

État de l'art

Sommaire

4.1 Web Participatif	77
4.1.1 Présentation générale	78
4.1.2 Concepts sous-jacents aux sites participatifs	80
4.1.3 Blogs	80
4.1.4 Wikis	81
4.1.5 P2P	83
4.1.6 Réseaux sociaux	84
4.1.7 Plateformes de partage	85
4.1.8 Partage de signets	85
4.1.9 Bilan des fonctionnalités	86
4.2 Web Participatif appliqué au domaine pédagogique	86
4.2.1 Web Participatif côté apprenant : elearning 2.0	86
4.2.2 Wiki éducatif	88
4.2.3 Réseaux pair-à-pair pédagogiques	88
4.2.4 Partage de signets et tags pédagogiques	89
4.2.5 Incitations à participer dans une communauté éducative	91
4.2.6 Plateformes pédagogiques pour les enseignants	92
4.3 Conclusion	93

La démocratisation des ordinateurs et le développement d'Internet ont favorisé l'émergence de nouveaux usages liés aux documents. Un des traits principaux de ces nouvelles pratiques porte sur la dimension sociale de la production et de l'organisation des documents, qui s'illustre dans des sites participatifs.

Dans cet état de l'art, nous parcourons les concepts et grands types d'outils du Web Participatif (« Web 2.0 »). Nous en tirons les principes pour le développement d'une plateforme participative de partage de documents pédagogiques.

4.1 Web Participatif

Contrairement au Web Sémantique promulgué par le W3C (§2.1.1), le « Web Participatif » recouvre un ensemble d'usages apparus de manière spontanée sur le Web. Au lieu d'étudier des standards comme pour le Web Sémantique, nous verrons plutôt les fonctionnalités implémentées sur les sites participatifs.

4.1.1 Présentation générale

Historique et définition Suite à la popularisation de l'accès à Internet, de nouveaux usages se sont développés au début des années 2000. Ceux-ci se fondent sur une participation massive des internautes à la production et l'organisation de ressources. Ce courant est illustré notamment par la popularité des wikis ou des blogs, des sites de partage de photos, de vidéos ou de signets.

Les premiers à remarquer cette évolution globale des usages furent Tim O'Reilly et ses associés qui proposèrent le terme de « Web 2.0 » ([O'Reilly, 2005]). Le suffixe « 2.0 » fait référence, comme pour un logiciel, à un changement de version important par rapport à l'usage traditionnel du Web.

Dans cet usage traditionnel, les contenus et services sur le Web reproduisaient des contenus et services du monde physique.

Puis, selon Tim O'Reilly, la bulle Internet explosa au début des années 2000 et le résultat fut l'élimination des nombreuses copies des modèles classiques pour laisser émerger les sites implémentant des modèles nouveaux d'échanges et de gestion des contenus.

La première définition du Web 2.0 de Tim O'Reilly et de ses associés prit la forme d'un tableau comparatif entre sites du Web 1.0 et sites du Web 2.0. La Fig. 4.1 en présente un extrait :

Web 1.0	Web 2.0
Britannica Online	Wikipédia
Répertoires (taxinomie)	Tags (« folksonomie »)
Sites personnels	Blogs
Mp3.com	Napster
DoubleClick	Google AdSense

FIG. 4.1 – Extrait de la comparaison Web 1.0 et Web 2.0 ([O'Reilly, 2005])

Britannica Online est la version en ligne de la fameuse encyclopédie Britannica, fondée sur l'écriture d'articles de références par des experts. Wikipédia est une encyclopédie participative en ligne apparue en 2003, où tout le monde peut contribuer. Contrairement à Britannica et grâce aux nombreuses contributions bénévoles, les contenus de Wikipédia sont gratuits et plus nombreux favorisant sa popularité sur le Web.

Au côté des structures en répertoires et des hiérarchies arborescentes pour classer des contenus par auteurs, sont apparus les tags, mots-clés attachés aux contenus par les utilisateurs.

Les sites personnels représentent les sites réalisés par des informaticiens tandis que 2003 annonçait le succès des blogs, ces sites que peuvent créer des internautes sans connaissance technique, grâce à des moteurs de blogs, comme Technorati ou Blogger.com, pour diffuser des contenus.

MP3.com était au début des années 2000 un magasin en ligne de vente de musique sous la forme de MP3. Napster était un des premiers réseaux pair-à-pair permettant aux utilisateurs d'échanger gratuitement entre eux des contenus comme des fichiers MP3.

Enfin, DoubleClick était une régie publicitaire reprenant le modèle classique de la publicité. Les agents de DoubleClick achètent des espaces de publicité sur les sites à hautes fréquentations puis les revendent aux entreprises désirant diffuser de la publicité. Google AdSense proposa une innovation importante en offrant un moyen automatique pour des créateurs de sites Web

de disposer des emplacements publicitaires à côté de leurs contenus, tandis que les annonceurs peuvent acheter automatiquement ces emplacements. Ce procédé sans intermédiaire a donné la possibilité à Google de le proposer à tous les sites, quelles qu'en soient leurs tailles. Ainsi, des sites de petite taille comme les blogs ont permis à leurs auteurs d'être rémunérés pour leurs contenus par la publicité.

Les exemples ci-dessus montrent la tendance générale liée à l'augmentation de la participation sur le Web. De nouvelles formes de gestion des contenus apparaissent à mesure que ces contenus sont :

- produits par des utilisateurs non forcément experts,
- plus nombreux,
- proposés gratuitement,
- échangés entre pairs, sans experts intermédiaires,
- organisés par les utilisateurs.

Ambiguïté d'une dénomination. Si l'expression « Web 2.0 » marque bien une évolution du Web, elle ne décrit aucune de ses nouvelles caractéristiques. De plus, les principales définitions se servant d'exemples pour expliquer le terme, beaucoup de sens peuvent en réalité se cacher derrière.

Dans ce mémoire, nous employons l'expression de « **Web Participatif** » pour signifier que la caractéristique essentielle qui nous intéresse est **la participation plus importante des internautes dans la production et l'organisation des contenus**. C'est cette participation qui renouvelle les formes classiques d'indexation, en fournissant des métadonnées moins chères et plus variées.

Nous n'avons pas employé l'expression « Web Social » car le terme « Social » insiste plutôt sur les relations entre individus. Pour tisser des relations par le Web, les gens doivent participer, mais ils peuvent être actifs sur le Web e.g. enregistrer leurs signets en ligne pour les sauvegarder, sans tisser de relations sociales. Le Web Social est donc plutôt un sous-ensemble du Web Participatif.

Effets de mode. L'expression « Web 2.0 » a fait beaucoup d'émules et désigne maintenant presque un courant « 2.0 ». Nous avons par exemple entendu parler de « Mode 2.0 » dans la presse féminine, de « Finance 2.0 », « d'Entreprise 2.0 » ou encore de « Bibliothèque 2.0 » parmi d'autres⁴⁷. Malgré toutes ces désignations, le suffixe 2.0 désignait généralement toujours le même concept, à savoir la participation accrue des utilisateurs finaux dans la production et l'organisation, en opposition avec le modèle classique distinguant les producteurs des consommateurs. Ce changement de rapports s'accompagne généralement d'une augmentation des liens directs entre utilisateurs, par exemple par Internet. Ainsi la « Mode 2.0 » signifie une mode créée par chacun et moins dictée par les marques ou les personnalités, la « Finance 2.0 » se rapporte à un site permettant le prêt direct d'individus à individus sans passer par une banque, « l'Entreprise 2.0 » étiquette un modèle d'entreprise où les salariés communiquent plus entre eux (notamment grâce aux outils du Web 2.0) au lieu des hiérarchies strictes, et enfin la « bibliothèque 2.0 » pointe l'évolution nécessaire des bibliothèques pour mettre en rapport les lecteurs entre eux, en leur permettant de commenter les ouvrages ou en les faisant se rencontrer.

⁴⁷eHub, blog fournissant une veille sur les dernières applications dites « Web 2.0 » : <http://www.emilychang.com/go/ehub/category/>

Ces exemples montrent que la notion de Web Participatif s'inscrit dans un courant majeur de la société, soutenu ou inspiré par les nouveaux usages d'Internet. Nos travaux s'intéressent à l'influence de ce courant sur le partage de documents pédagogiques.

Pour mieux cerner la tendance du Web Participatif et tenter de l'appliquer au contexte pédagogique, nous avons étudié les concepts sous-jacents et les fonctionnalités des sites participatifs.

4.1.2 Concepts sous-jacents aux sites participatifs

Nous appelons « sites participatifs », les sites Web dont une partie des contenus ou de leurs organisations exploite la participation des utilisateurs de ces mêmes contenus.

Nous avons repéré deux objectifs principaux nécessitant cette participation :

- **Produire et organiser plus de contenus, moins cher et différemment** : en faisant participer les internautes, les sites peuvent obtenir des contenus plus nombreux, aux coûts de production moindres mais aussi des contenus différents de ceux réalisés par quelques experts dans un domaine car produits par un panel plus large.
- **Faire passer du temps sur les pages Web** : en offrant aux utilisateurs des services pour produire des contenus, les sites possèdent des utilisateurs qui restent longtemps dessus et donc voient beaucoup de publicités, et offrent ainsi des revenus publicitaires plus importants. Néanmoins, du point de vue pédagogique, nous ne nous intéresserons pas davantage à cette motivation.

Nous classons les fonctionnalités et facteurs incitant à participer selon trois catégories :

- **Catégorie ergonomique** : le site permet d'effectuer des actions de production de manière simple et rapide.
- **Catégorie sociale** : le site permet aux utilisateurs de communiquer ou d'agir entre eux, ou au moins de voir les activités des autres. Ces possibilités permettent l'émergence de liens sociaux, favorisant ou motivant les contributions.
- **Catégorie pécuniaire** : le site permet aux utilisateurs d'être rémunérés pour leurs contributions.

Nous allons maintenant analyser les sites participatifs au regard des trois catégories présentées ci-dessus. Nous en déduisons des fonctionnalités et principes applicables au domaine pédagogique et regroupés en §4.1.9.

4.1.3 Blogs

A l'origine, les blogs désignent des sites offrant des contenus de manière chronologique. Selon nous, l'intérêt des blogs tient à la facilité de création permise par des outils comme Wordpress⁴⁸, Blogger⁴⁹ ou Technorati⁵⁰. Sur ces hébergeurs de blogs, les internautes peuvent créer un site personnel gratuitement et avec peu de connaissances techniques.

Catégorie ergonomique :

- **Facilité de création** : rien à installer pour commencer à ajouter des textes ou des images sur son espace blog. La création devient ainsi à portée de beaucoup d'internautes.

⁴⁸Wordpress : wordpress.org

⁴⁹Blogger : www.blogger.com

⁵⁰Technorati : technorati.com

- **Espaces configurables** : les hébergeurs de blogs comme Wordpress proposent de plus en plus de fonctionnalités permettant de personnaliser son espace. Cette personnalisation peut toucher la forme e.g. en changeant la couleur des pages, ou alors en ajoutant des modules complémentaires facilitant la navigation parmi le contenu du blog.

Catégorie sociale :

- **Commentaires sur les contenus** : la principale fonctionnalité diffusée par les blogs est sûrement la possibilité laissée aux lecteurs de commenter les contenus et éventuellement de débiter ainsi une conversation avec l’auteur. Aujourd’hui cette discussion autour des contenus sur Internet s’est répandue sur pratiquement toutes les pages Web.

Le concept de blogs cache différentes formes d’incitation à participer. Les hébergeurs de blogs incitent les internautes à créer des blogs et des contenus ensuite. Les bloggeurs incitent leurs lecteurs à participer en laissant des commentaires. Ces commentaires peuvent être en réponse une incitation à créer des contenus.

4.1.4 Wikis

Quoi de mieux pour définir un wiki que de reprendre la définition⁵¹. du wiki le plus connu, Wikipédia :

Un wiki est un système de gestion de contenu de site web rendant ses pages Web librement modifiables par tous les visiteurs y étant autorisés. Les wikis sont utilisés pour faciliter l’écriture collaborative de documents avec un minimum de contraintes. Ils ont été inventés en 1995 par Ward Cunningham, pour une section d’un site sur la programmation informatique qu’il a appelée WikiWikiWeb. Le mot « wiki » vient du redoublement hawaïen « wiki wiki », qui signifie « rapide ». Au milieu des années 2000, les wikis ont atteint un bon niveau de maturité ; ils sont depuis lors associés au Web 2.0. Créé en 2001, Wikipédia est devenu le site web écrit avec un wiki le plus visité.

Notre analyse des wikis portera principalement sur le site francophone de Wikipédia possédant plus de 750 000 articles, puis sur GoogleKnol, une approche plus récente des articles en ligne.

Pour Wikipédia, cette analyse se base notamment sur le dossier de l’INRP ([Endrizzi, 2006]).

Classement des fonctionnalités et principes selon les catégories ergonomiques et sociales :

Catégorie ergonomique :

- **Ajouter et modifier rapidement des contenus** : sur Wikipédia, chaque article possède en haut à droite un bouton « Editer » permettant à un lecteur de corriger ou d’améliorer le contenu. En un clic sur ce bouton, l’utilisateur voit le texte source de l’article. Après modification, en un second clic sur le bouton « Sauvegarder », la modification est prise en compte. Sauf cas particulier, ces modifications ne nécessitent pas de s’identifier.
- **Contenu sous licence libre** : Sur Wikipédia, tous les contenus sont légalement sous une licence libre définissant que le contenu peut être copié puis modifié à volonté sous la seule condition de mentionner la source. Cette licence semble indispensable pour l’appropriation des contenus par chacun et permettre une plus grande participation à leur édition.

⁵¹Définition d’un wiki sur Wikipédia : <http://fr.Wikipédia.org/w/index.php?title=Wiki&oldid=38200838>

Catégorie sociale :

- **Hiérarchie souple des rôles** : pour effectuer une modification normale, les utilisateurs n'ont pas besoin d'être identifiés et donc de se voir attribuer des autorisations particulières. Il existe néanmoins des rôles implicites et explicites sur Wikipédia. Chacun peut s'attribuer un rôle implicite comme se concentrer sur un thème précis pour y effectuer des modifications sur le long terme. Les rôles explicites permettent la gestion des conflits de manière la plus démocratique possible, c'est à dire selon les principes régissant l'encyclopédie et de manière transparente (voir fonctionnalité suivante). Ces rôles explicites sont principalement décidés par votes et dépendent notamment de l'implication des acteurs. Ainsi, les internautes ont plusieurs moyens de participer à Wikipédia et de trouver un rôle leur convenant. Ces possibilités semblent faciliter l'émergence d'une communauté Wikipédia et soutenir ainsi la participation.
- **Transparence des actions** : toute action (au delà de l'édition des pages) sur Wikipédia est enregistrée dans un fichier visible et recherchable par tous. Cette transparence va de pair avec l'aspect démocratique de Wikipédia, et cette qualité est probablement un facteur de participation à son organisation.
- **Profils des utilisateurs** : les utilisateurs ne sont pas que des identifiants mais chacun peut créer une page de profil pour s'y présenter ou pour expliquer en quoi il souhaite contribuer à Wikipédia. Cette fonctionnalité semble nécessaire au tissage de liens sociaux entre membres, qui peuvent ainsi mieux se connaître.
- **Espaces de discussion** : les membres de Wikipédia peuvent discuter de sujets généraux sur des chats ou des forums liés au site Web, et renforcer leurs expériences entre membres.
- **Espaces de négociation liés aux contenus** : chaque article possède un espace de discussion, aussi facilement éditable que l'article lui-même. Cet espace sert aux contributeurs à collaborer autour de l'article ou à discuter leurs différends.
- **Historique des modifications** : chaque article de Wikipédia est versionné. Ceci signifie que chaque modification est associée à une version et que l'ensemble de ces versions est affichable dans un historique de la page. En cas de conflit, il est aussi possible de revenir à une version antérieure de l'article. Cette historique semble un point fondamental pour la collaboration autour d'un objet commun.

Toutes ces fonctionnalités incitent aux échanges entre utilisateurs. Les premières fonctionnalités favorisent plutôt les liens entre membres et la formation d'une structure communautaire tandis que l'espace de négociation et l'historique sont liés plus directement au travail collaboratif autour d'un même objet.

Alors que Wikipédia est un exemple très connu de Wiki, nous voudrions présenter une initiative moins connue nommée GoogleKnol et qui s'intéresse aux articles d'experts.

GoogleKnol⁵² a été lancé officiellement en Juillet 2008. « Knol » signifie une unité de « knowledge » (connaissance) et désigne un article de référence sur un sujet. Bien que Wikipédia et GoogleKnol puissent sembler très proches (tout le monde peut écrire des articles), GoogleKnol se différencie en visant plus particulièrement la participation d'experts dans leurs domaines.

GoogleKnol possède de nombreuses caractéristiques ergonomiques pour faciliter la rédaction d'articles mais ces incitations à la participation d'experts sont les plus intéressantes :

⁵²GoogleKnol, par Google, « Put your name behind your content ; write a knol. » : knol.google.com

Catégorie sociale :

- **Reconnaissance des auteurs** : contrairement à Wikipédia, GoogleKnol autorise (voir incite) la création de plusieurs articles sur un même sujet. Les auteurs entrent alors en concurrence pour proposer les meilleurs articles. Par conséquent, alors que les auteurs de Wikipédia n'apparaissent pas sur la page d'un article, l'auteur d'un GoogleKnol est mis en avant et clairement identifié. GoogleKnol recommande aux auteurs de lister sur leur page-profil les facteurs établissant leur identité et leurs références dans le domaine de leurs articles. Par exemple, l'identité d'un auteur peut être validée par reconnaissance de la carte bancaire.
- **Récompenses symboliques** : pour valoriser les auteurs, le système calcule aussi des récompenses attribués aux auteurs possédant les articles les plus lus ou les mieux notés par les lecteurs. Ces récompenses symboliques prennent la forme de badges apparaissant par exemple à côté du nom de l'utilisateur.
- **Participation paramétrable** : alors que dans Wikipédia, tout auteur peut voir son travail effacé par d'autres, GoogleKnol permet à chaque auteur de configurer le niveau de participation qu'il souhaite pour son article. Ainsi un auteur peut demander à un autre auteur de faire une revue de son article, autoriser les lecteurs à modifier le document (comme sur Wikipédia) ou alors juste à ajouter des commentaires et des notes, qui apparaîtront sous le document. De plus, l'auteur peut paramétrer la licence légale du document sur un panel large allant d'une licence restrictive (aucun droit de copie) jusqu'aux licences libres autorisant diffusion et modification.

Catégorie pécuniaire :

- **Rémunération publicitaire** : un auteur peut recevoir, s'il le désire, une participation aux intérêts de la publicité générée par les pages qu'il a écrites et ainsi être incité à en écrire d'autres.

Contrairement à Wikipédia qui favorise une communauté d'auteurs, Google Knol recherche la participation d'experts pour produire des contenus. Pour cela, Google rétribue symboliquement ou de manière monétaire, et offre beaucoup de choix dans la configuration du contenu (licences légales et niveau de participation). Ces deux sites montrent ainsi un large choix de fonctionnalités participatives.

4.1.5 P2P

Le Pair-à-Pair (P2P) désigne les outils et les réseaux permettant des échanges de fichiers directement entre utilisateurs, et ne nécessitant pas de machine centrale. Cette pratique s'est notamment développée avec l'échange de fichiers musicaux au format MP3 ou de films au format DivX.

Napster fut le premier réseau P2P à succès. Suite à sa fermeture par la justice, d'autres protocoles et réseaux virent le jour, comme le protocole BitTorrent. Dans ce protocole les utilisateurs ne sont pas reliés un à un mais plusieurs à plusieurs. Ceci signifie que l'échange ne s'effectue pas simplement d'un utilisateur à un autre, mais qu'une centaine d'utilisateurs peuvent s'échanger un même fichier en même temps, chacun envoyant aux autres les bouts qu'il possède. Cette technique accélère fortement le téléchargement de gros fichiers.

Catégorie ergonomique :

- **Faciliter l'échange de gros fichiers** : en reliant directement les utilisateurs entre eux, ces protocoles multiplient la rapidité de téléchargement de gros fichiers. Beaucoup de ces protocoles sont conçus pour que plus un utilisateur envoie des données, plus il peut en recevoir.

4.1.6 Réseaux sociaux

Nous entendons par sites de réseaux sociaux les sites où les liens entre membres et la création de groupes d'utilisateurs (parfois appelés « communautés ») sont largement exploités. Il existe des réseaux généralistes comme Facebook⁵³, à l'usage spécialisé comme MySpace⁵⁴ pour la musique ou à visée professionnelle comme LinkedIn⁵⁵ ou Viadeo⁵⁶.

Alors qu'un outil comme Wikipédia permet la structuration d'une communauté autour d'objets communs (les articles encyclopédiques), un réseau comme Facebook permet l'apparition digitale de réseaux existants dans le monde physique. C'est à dire qu'un internaute pourra y retrouver ses amis. On pourrait considérer que le contenu créé par les utilisateurs sur un réseau social comme Facebook est l'ensemble des traces digitales de sa vie.

Catégorie sociale :

- **Retrouver ou inviter les amis** : les réseaux sociaux possèdent tous des fonctionnalités permettant de retrouver ses amis et connaissances sur le réseau social. En retrouvant ses amis, l'utilisateur recrée un cercle de confiance incitant à revenir sur le réseau social. Par exemple, de nombreux sites proposent de leur donner les mots de passe des comptes emails afin qu'ils y récupèrent la liste des connaissances issues du répertoire de contacts. Ils proposent aussi d'envoyer automatiquement des invitations aux contacts n'ayant pas encore de compte.
- **Commentaires sur toute action** : alors que les blogs inventaient les commentaires sur les articles, Facebook propose de commenter quasiment toute activité des amis. En effet, Facebook enregistre l'activité de tout utilisateur et permet à ses amis de la voir : ajout d'un module, changement d'une information sur la page profil, résultat à un jeu. Toutes ces activités sont susceptibles de lancer une discussion et ainsi de renforcer les liens entre les participants.

Catégorie pécuniaire :

- **Applications rémunérées** : les réseaux sociaux ne rémunèrent pas encore les membres mais des sites comme Facebook rémunèrent les développeurs d'applications pour le réseau. En effet, Facebook propose à ses membres de nombreuses applications exploitant le réseau social. Cela peut aller de jeux entre amis à une application de gestion de livres préférés que verront alors les amis, etc. Pour enrichir le choix dans les applications proposées, Facebook ouvre son interface de développement à tous les développeurs qui peuvent gratuitement créer une application sur le réseau et toucher les revenus issus de la publicité.

⁵³Facebook : www.facebook.com

⁵⁴MySpace : www.myspace.com

⁵⁵LinkedIn « strengthens and extends your existing network of trusted contacts » : www.linkedin.com

⁵⁶Viadeo : www.viadeo.com

4.1.7 Plateformes de partage

Nous entendons par plateformes de partage les sites permettant aux utilisateurs de déposer des contenus, généralement multimédias, tels que des photos ou des vidéos. Les sites les plus représentatifs sont YouTube ou Dailymotion pour les vidéos, Flickr pour les photos ou Scribd pour les documents textuels. YouTube sera le service qui illustrera majoritairement nos propos.

Catégorie ergonomique :

- **Facilité de dépôts** : YouTube permet de déposer des fichiers vidéos de plusieurs dizaines de Mo et notamment des vidéos en hautes définitions (HD) très lourdes. Si l'utilisateur voulait employer son propre serveur, il devrait dépenser des sommes assez importantes pour permettre la diffusion à plusieurs centaines d'utilisateurs alors que c'est gratuit même pour des milliers de spectateurs sur YouTube. YouTube propose enfin des services de dépôts directement à partir du téléphone mobile, multipliant les moyens de dépôts.
- **Intégration externe des documents** : les vidéos YouTube ou les photos de Flickr peuvent être intégrées et visualisées dans d'autres sites Web. Un utilisateur qui y dépose ses documents n'est ainsi pas captif de la plateforme et peut intégrer ses vidéos de manière assez transparente sur son propre site.
- **Services sur les documents** : en plus de permettre le dépôt de photographies, Flickr propose des outils pour modifier et éditer ces photos en ajoutant un cadre, du texte etc. De la même manière, YouTube propose des services avancés pour annoter les vidéos ou les sous-titrer. Le service de dépôt devient ainsi service d'édition, facilitant le passage du statut de déposant à celui d'utilisateur plus actif.

Catégorie pécuniaire :

- **Rémunération publicitaire** : YouTube expérimente depuis quelques mois des publicités qui seraient diffusées avant ou après certaines vidéos. Ces publicités, imitant le modèle de la télévision, serviraient à rémunérer les déposants.

Nous avons présenté les plateformes de partage, principalement liées au dépôt d'un contenu, qui sera ensuite éventuellement transformé (annotation des vidéos par exemple). Cependant, il existe de plus en plus de sites de partage où le contenu n'est pas déposé mais directement créé en ligne à l'aide des outils mis à disposition. Google Map illustre parfaitement cet aspect en proposant un service de cartographie avancée. Le service de base permet d'afficher la carte ou photographie d'une ville, ou de chercher son chemin. Néanmoins, le site explore de plus en plus les fonctionnalités permettant de créer des cartes annotées puis de les intégrer sur un autre site.

Ces sites incitent à la participation en offrant gratuitement des outils avancés qui permettent la manipulation de contenus généralistes (vidéos, photos) mais aussi des contenus plus spécialisés comme les cartes géographiques.

4.1.8 Partage de signets

Les signets (appelés aussi « favoris » ou « bookmarks ») sont des adresses Web (URL) associées à une description. A l'origine, les signets sont sauvegardés par les internautes sur leurs navigateurs Web pour retrouver rapidement un site.

Au lieu de les enregistrer sur son ordinateur, le partage de signets (« social bookmarking »

) consiste pour un internaute à enregistrer ses signets sur un site Web. Ses signets seront alors visibles par tous et permettront au site de présenter les signets les plus populaires par thèmes.

Des sites comme Delicious⁵⁷ ou MisterWong⁵⁸ permettent de sauvegarder et partager des signets organisés par mots-clefs spontanés, les tags.

Catégorie ergonomique :

- **Sauvegarder des contenus à distance** : dans un cadre solitaire, les utilisateurs pourraient très bien continuer à employer leurs signets au sein des navigateurs Web. Delicious permet de sauvegarder ses signets pour ne plus les perdre ou pour les retrouver même en changeant d'ordinateur.
- **Rapidité d'ajout** : un signet peut être ajouté en quelques clics grâce à l'installation d'une mini-extension dans le navigateur Web (bookmarklets).

Catégorie sociale :

- **Partage de signets** : Delicious propose deux moyens d'interagir avec les autres utilisateurs. Le premier moyen est de former des réseaux en désignant des utilisateurs comme des contacts. Une page spéciale permet alors d'observer les signets ajoutés par son réseau. Le second moyen est de pouvoir envoyer des ressources à ces contacts.

Nous traitons le partage de signets en dehors des autres contenus car ce type d'outil sera la base du projet de la troisième partie, nommé SemanticScuttle. §7.2.1 présente ainsi en détail un outil classique de partage de signets.

4.1.9 Bilan des fonctionnalités

Les sites étudiés ci-dessus permettent de lister un ensemble de fonctionnalités dessinant une plateforme participative de partage de contenu. Nous les résumons dans la Fig 4.2).

L'idée du projet Pépi, que nous présenterons en (§5.2.3), est d'appliquer ces facteurs et ces fonctionnalités au partage de documents pédagogiques et aux enseignants.

4.2 Web Participatif appliqué au domaine pédagogique

La tendance du Web Participatif a inspiré beaucoup de projets dans le domaine pédagogique et notamment par rapport au partage de documents. Voici quelques travaux et usages illustratifs.

4.2.1 Web Participatif côté apprenant : elearning 2.0

Dès 2005, peu après l'apparition de l'expression « Web 2.0 », Stephen Downes ([Downes, 2005]) fait le parallèle entre les mécanismes du Web Participatif et l'apprentissage. Il propose alors l'expression « E-learning 2.0 » pour décrire une forme d'apprentissage exploitant les outils du Web Participatif autour des apprenants.

Au lieu de proposer des connaissances statiques et enfermées, les outils proposent des connaissances inter-reliées, à l'image des blogs, et dépassant les limites d'une institution : « The e-learning application begins to look very much like a blogging tool. It represents one node in a

⁵⁷Delicious : delicious.com

⁵⁸Mister Wong : www.mister-wong.fr

Catégorie ergonomique :

- Rien à installer
- Espace personnalisable
- Ajouter et modifier rapidement des contenus
- Contenu sous licence libre (si collaboration)
- Facilité de dépôts de gros contenus
- Sauvegarder des contenus à distance
- Services sur les documents
- Intégration des documents sur des sites externes

Catégorie sociale :

- Hiérarchie souple des rôles (si structure communautaire émergente)
- Transparence des actions : visibles par tous
- Espaces de discussions
- Espaces de négociations liés aux contenus
- Historique des modifications (si collaboration)
- Profils des utilisateurs
- Reconnaissance des auteurs (si non collaboration)
- Récompenses symboliques des auteurs (si non collaboration)
- Participation paramétrable (si non collaboration)
- Retrouver ou inviter les amis
- Commentaires sur les contenus

Catégorie pécuniaire :

- Applications des développeurs rémunérées
- Rémunération publicitaire pour les auteurs

FIG. 4.2 – Récapitulatif des fonctionnalités soutenant la participation sur les sites du Web Participatif.

web of content, connected to other nodes and content creation services used by other students. It becomes, not an institutional or corporate application, but a personal learning center, where content is reused and remixed according to the student's own needs and interests. It becomes, indeed, not a single application, but a collection of interoperating applications—an environment rather than a system. »

Pour lui, les blogs et les wikis offrent de nouveaux espaces pour que les apprenants construisent la connaissance : « What happens when online learning software ceases to be a type of content-consumption tool, where learning is 'delivered,' and becomes more like a content-authoring tool, where learning is created ? »

Liens entre apprenants et constructions collectives offrent ainsi selon lui les possibilités pour une nouvelle forme d'apprentissage en ligne : « This approach to learning means that learning content is created and distributed in a very different manner. Rather than being composed,

organized and packaged, e-learning content is syndicated, much like a blog post or podcast. It is aggregated by students, using their own personal RSS reader or some similar application. From there, it is remixed and repurposed with the student's own individual application in mind, the finished product being fed forward to become fodder for some other student's reading and use. »

Cette vision décrit les opportunités pédagogiques liées aux nouvelles formes de contenus et d'outils sur le Web. Néanmoins, par rapport à nos travaux, cette vision s'intéresse plus aux apprenants qu'aux opportunités pour les enseignants.

4.2.2 Wiki éducatif

Parmi les projets de wikis éducatifs, le plus abouti et celui ayant la plus grande échelle est probablement le projet Connexion⁵⁹ où les enseignants peuvent rédiger des manuels de cours sous la forme wiki et avec une licence libre. Le but est ensuite de favoriser les traductions et les nouveaux manuels reprenant éventuellement des morceaux des anciens.

Wikiversity⁶⁰ est un autre projet international, soutenu par la fondation Wikimedia qui soutient déjà Wikipédia. Basé sur le même outil Wiki que Wikipédia⁶¹, Wikiversity a pour premier but la rédaction collaborative de cours en ligne. Wikiversity tente aussi de soutenir l'apprentissage par les interactions entre utilisateurs. L'idée serait éventuellement de voir se former des communautés d'apprentissage autour des cours. Cette logique suit celle de la communauté Wikipédia qui est décrite dans [Endrizzi, 2006] comme une organisation auto-apprenante, dans le sens où les débats entre les membres autour des articles sont un moyen de discuter les connaissances mais aussi les diffuser aux participants.

Les wikis sont un espace privilégié pour construire des articles de manière collaborative. Nos travaux ne porteront pas directement sur cet aspect car il nous semble aujourd'hui encore trop minoritaire. Nous nous sommes plutôt intéressés au partage de documents déjà existants.

4.2.3 Réseaux pair-à-pair pédagogiques

EduTella. EduTella⁶² est le principal projet ayant exploré l'échange de métadonnées pédagogiques sous la forme des réseaux pair-à-pair.

EduTella ([Nejdl et al., 2002]) est un réseau pair-à-pair pour partager des métadonnées sémantiques et non des fichiers. EduTella n'est pas un réseau unique mais permet plutôt à différents systèmes de former des réseaux pour échanger des métadonnées sémantiques selon des standards du Web Sémantique. Le noyau simplifié d'EduTella consiste en une librairie et en un langage de requête. La librairie permet de développer un fournisseur ou un client se connectant à un réseau EduTella. Le langage de requête (QEL) sert d'intermédiaire entre la représentation sémantique (RDF) employée sur les réseaux EduTella et les représentations internes à chaque fournisseur (en SQL par exemple si les métadonnées sont dans une base de données).

Nous n'avons pas trouvé de retours d'usages sur le projet EduTella.

⁵⁹Connexion : cnx.org

⁶⁰Wikiversity francophone : <http://fr.wikiversity.org>

⁶¹MediaWiki : www.mediawiki.org

⁶²Projet EduTella, P2P pédagogique : Issu de www.edutella.org/edutella/edutella.shtml

Échanges entre enseignants. Mis à part ce projet européen, une partie des enseignants se sont aussi appropriés les réseaux pair-à-pair pour y échanger les manuels scolaires et les carnets pédagogiques en fichiers PDF. N'ayant pas les autorisations des éditeurs, ces échanges se font dans l'illégalité mais ils prouvent que, les réseaux P2P généralistes ne servent pas qu'à l'échange de fichiers vidéos ou musicaux mais peuvent aussi servir aux échanges professionnels par exemple entre enseignants.

4.2.4 Partage de signets et tags pédagogiques

Cette section développe différentes expériences, en particulier, sur les tags appliqués aux documents pédagogiques et plus généralement pour des échanges avec des outils de partage de signets.

Altered Vista ou le partage de signets entre élèves. [Recker et al., 2000] détaille un travail sur le filtrage collaboratif de ressources pédagogiques par des apprenants en 2002, avant que Delicious ou le terme de « Web 2.0 » n'existent.

Cette approche « enables context-sensitive discovery and recommendation of learning objects. The discussion is based upon research in developing and evaluating a collaborative filtering system, which enables users to share ratings, opinions, and recommendations about resources on the Web. An additional benefit of this approach is that it also allows a user to locate other users that share similar interests for further communication and collaboration. »

« As we will describe, the approach supports metadata structures that incorporate what we call 'non-authoritative' data elements. This form of metadata attempts to capture the context of use and surrounding activities of the learning object. The data elements can also describe the community of users from which the learning object is derived. Moreover, any user (and not just the authorized cataloger) can contribute a metadata record. As a result, a particular learning resource may have multiple 'non-authoritative' metadata records, in addition to its 'authoritative' record. »

L'expérience conduite employa 15 élèves qui effectuèrent 172 revues sur 97 ressources. La Fig. 4.3 montre les 11 champs de metadonnées requises ainsi que leur format.

Dans ses conclusions, Mimi Recker estime l'intérêt du mécanisme pour aider les élèves à trouver des élèves aux intérêts communs et pour aider les enseignants à trouver des contenus de bonne qualité.

Néanmoins, l'auteur souligne aussi la difficulté à inciter les utilisateurs à décrire des contenus et avoue ne pas connaître de solution immédiate. Quelques années plus tard, le succès des sites de partage de signets montrèrent qu'un outil donnant aux utilisateurs le moyen de sauvegarder ses propres signets (pas seulement pédagogiques) et avec des mécanismes de description encore plus simples (comme les tags) pouvaient inciter cette participation. Les travaux de Mimi Recker étaient trop en avance pour voir l'émergence du Web participatif.

Annoter à partir de « social bookmarking tool ». Dans [Al-Khalifa and Davis, 2007], les auteurs expérimentent la description de contenus pédagogiques en informatique (des cours

Field	Description	Value's type
Web site title	The title of the site	Text box
Internet address	The URL of the site	Text box
Keywords	Keywords to classify resource	Multiple selection list
Added by	User name	Automatically generated
ADA Accessibility	Meets Disabilities Act design criteria	5-point Likert scale
Usability	How usable is the resource	5-point Likert scale
Authoritativeness	Authority base of document author	5-point Likert scale
Educational relevance	Educational relevance of the resource	5-point Likert scale
Description	Simple description of resource	Text box
Quality	The subjective quality of the resource	5-point Likert scale
Overall rating	Overall opinion	5-point Likert scale

FIG. 4.3 – Champs à remplir pour partager des signets pédagogiques ([Recker et al., 2000]), avant l'émergence du Web Participatif.

ou tutoriels sur le CSS⁶³) en exploitant les tags du site Delicious.

Si les tags sont riches, en tant que termes spontanés, ils sont parfois difficiles à interpréter par d'autres individus que l'auteur. Pour éviter ces problèmes, Al Khalifa et ses collègues éliminent simplement tous les termes ambigus (homonymes, termes avec des caractères non anglophones). Les pluriels sont automatiquement mis au singulier pour faciliter les regroupements.

Puis, Al Khalifa relie les tags restants aux concepts de trois ontologies : ontologie du Web Design, ontologie du domaine du CSS et ontologie des types de ressources techniques et pédagogiques. Ces ontologies lui permettent alors de remplir plus d'une dizaine de champs.

Al Khalifa finit en comparant les métadonnées obtenues à celles d'un expert seul. Ses conclusions sont que les métadonnées issues de Delicious couvrent un vocabulaire plus larges que celles fournies par un expert et peuvent faciliter ainsi les recherches.

Annotations de ressources pédagogiques Dans [Bateman et al., 2007], Scott Bateman explore aussi la description collective de documents pédagogiques mais axée principalement sur l'annotation des documents. Son système permet ainsi à des apprenants d'annoter les passages d'un cours numérique comme ils surligneraient un manuel papier (4.4). Ils peuvent ensuite ajouter un commentaire pour expliquer le surlignement ou tagguer le passage pour ensuite le retrouver. Les tags liés aux surlignements sont ensuite navigables par tous. Les documents peuvent aussi montrer l'ensemble des passages surlignés avec même les morceaux recouverts par le plus grand nombre de surlignement. Pour faciliter le démarrage du partage, les documents sont taggués à partir des mots-clefs repérés automatiquement dans les contenus.

Selon Scott Bateman, les intérêts d'un tel dispositif sont :

1. Les plateformes d'apprentissage manquent de supports pour auto-organiser les contenus pédagogiques.

⁶³Le CSS est un format pour décrire le visuel de pages Web, complémentaire du HTML.

2. Le tagging collectif a le potentiel d'enrichir les relations entre étudiants et de permettre la prise de conscience des autres autour des contenus pédagogiques.
3. Le tagging est une pratique réflexive, qui peut donner aux étudiants l'opportunité de résumer des idées tout en s'inspirant des autres étudiants (à travers leurs tags par exemple).
4. Les tags fournissent un aperçu de la compréhension et de l'activité des apprenants, ce qui peut être utile aux enseignants.

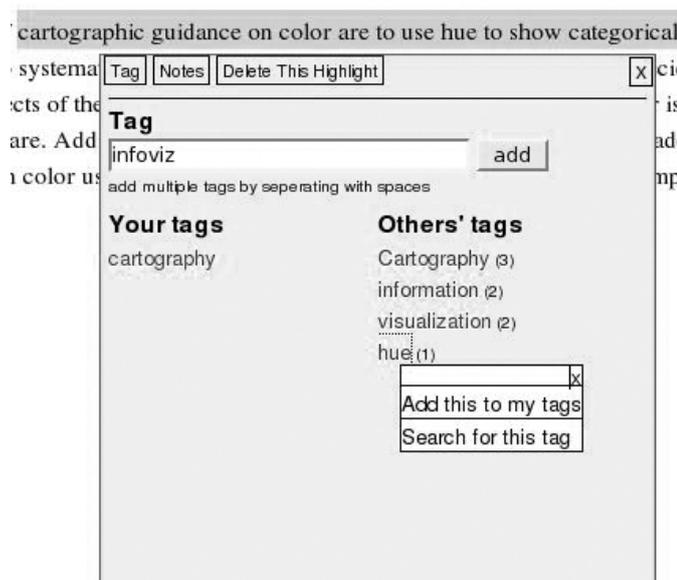


FIG. 4.4 – Ecran de OATS : L'extrait d'un document pédagogique peut être surligné (première ligne en haut) puis annoté et taggé de manière collective (fenêtre centrale).

4.2.5 Incitations à participer dans une communauté éducative

Soutenir la participation est une caractéristique majeure des sites participatifs.

Ainsi Julita Vassileva ([Cheng and Vassileva, 2006]) propose un système inspiré d'un site participatif d'information nommé Slashdot⁶⁴. Slashdot permet à n'importe qui de déposer des nouvelles liées à la culture informatique. La nouvelle est ensuite validée par des modérateurs et si cette validation est positive l'auteur de la nouvelle voit son « karma » augmenter. Les modérateurs sont choisis aléatoirement par le système, proportionnellement à leur karma. Par la suite, plus un auteur est réputé (karma élevé), plus ses nouvelles ont une chance d'être visible des lecteurs.

Le système de Julita Vassileva met au point un système de c-points, que l'on gagne en évaluant les ressources des autres utilisateurs. Plus on obtient de c-points, plus on pourra ensuite mettre en avant ses propres ressources. De plus une interface symbolique permet de mettre en avant des gold members, utilisateurs les plus réputés.

⁶⁴Slashdot : slashdot.org

4.2.6 Plateformes pédagogiques pour les enseignants

Nos travaux sont proches de ceux d'Elise Garrot [Garrot, 2008] portant sur les outils soutenant les communautés de pratique (CoP) en général, et plus particulièrement les communautés de tuteurs.

Dans un riche état de l'art, Élise Garrot classe ces outils en (pp. 71-82 de [Garrot, 2008]) :

- Support aux CoPs par des outils « classiques » (forums, listes de diffusion, ...)
- Universités et campus numérique : portails Web
- Hors des frontières des institutions
- Support aux CoPs par des outils « spécifiques »

Voici sa synthèse selon 8 critères :

- La technologie est indispensable à la CoP, c'est-à-dire qu'il s'agit d'une (1) communauté distribuée où les membres communiquent principalement à travers un environnement informatique.
- Le niveau des pratiques qui concernent les membres de la CoP : (2) au niveau local d'une institution voire d'un département d'enseignement ou (3) à un niveau global indépendant du contexte spécifique des pratiques de chacun.
- Les caractéristiques des outils informatiques supportant la CoP : (4) offrir un espace supportant l'identité communautaire, (5) favoriser l'entraide et l'échange entre pairs, (6) favoriser la collaboration et la réflexivité, (7) proposer des ressources contextualisées et (8) favoriser la recherche en rapport au contexte de la pratique d'un membre.

Systèmes et outils informatiques	1	2	3	4	5	6	7	8
Portails Web d'universités ([Sherer et al., 2007] ; [Lefoe et al., 2002])		X			X			
Environnements de campus à distance([Peyrelong and Follet, 2004] [Béziat and Caron, 2003])	X	X			X			
CoPs émergentes supportées par des blogs ([Pashnyak and Dennen, 2007]), listes de diffusion ([Daele, 2005])	X		X		X			
Tapped-In ([Schlager and Schank, 1997])	X		X		X	X		
CoPe_it! ([Karacapilidis and Tzagarakis, 2007])		X				X		
doceNet ([Brito-Mirian et al., 2006])		X				X	X	X
IFL (Inquiry Learning Forum) ([Barab et al., 2001])	X		X				X	X
Liste de diffusion @pretic ([Makni et al., 2008])	X		X		X			X

FIG. 4.5 – Extrait de la synthèse des environnements et outils supports aux CoPs en ligne d'enseignants et tuteurs ([Garrot, 2008])

Suite à cette synthèse (Fig. 4.5), Elise Garrot propose un outil basé sur une plateforme collaborative nommée Joomla! qui fournirait aux tuteurs un ensemble de fonctionnalités favorisant la participation et soutenant le partage de ressources.

Elle met notamment en avant le besoin d'offrir aux participants un niveau local (i.e. leurs propres écoles, départements d'enseignement ou institution) et global d'échanges (entre membres de différentes institutions).

Par rapport à ces nombreux projets, nous tirons les conclusions qu'aucun n'offre les caractéristiques d'une plateforme de partage participative telle que nous l'avons définie précédemment, et centrée sur les documents pédagogiques :

- Une plateforme globale mais où chacun peut créer les groupes qu'il désire.
- Centrée sur les documents déposés et inspirant éventuellement ensuite des rencontres autour : Inquiry Learning Forum est centré sur les discussions.
- Permettant une rapidité des dépôts.
- Offrant une possibilité d'exporter ses documents sur un autre site.
- Proposant un ensemble de services pour les documents pédagogiques : CoPe.it! propose bien un service pédagogique mais non lié aux documents.
- Etant indépendant des institutions : les portails Web sont liés à des universités.

4.3 Conclusion

De tous ces travaux nous avons conclu qu'il n'y avait pas de plateformes reprenant les principes du Web Participatif pour le partage de documents pédagogiques. Nous voulions réaliser une sorte de YouTube des documents pédagogiques.

De plus, contrairement à beaucoup de travaux, nous étions intéressés par les documents décrits et partagés par les enseignants (et non par les étudiants) et pour des contextes présentiels (et non pour la formation en ligne).

5

Projet Pépi

Sommaire

5.1	Contexte	96
5.1.1	Différentes communautés d’enseignants : exemple de Cartables.net	96
5.1.2	Étude de la communauté PGM (Professeurs en Génie Mécanique)	98
5.2	Développements	100
5.2.1	Objectifs	100
5.2.2	Principes d’une plateforme participative de partage de ressources pédagogiques	101
5.2.3	Fonctionnalités soutenant la participation	101
5.2.4	Architecture technique	107
5.3	Expérimentations	108
5.3.1	Contact des utilisateurs	108
5.3.2	Amorçage	108
5.3.3	Participations	108
5.4	Analyses et discussions	109
5.4.1	Analyses de l’échec des expérimentations	109
5.4.2	Discussion des concepts et travaux similaires	111
5.5	Conclusions	114

Pépi est un projet visant à réaliser une plateforme de partage de documents inspirée des applications du Web Participatif et spécialisée dans le domaine pédagogique.

« Pépi » est le diminutif de « Pépinière Pédagogique ». La pépinière fait référence à un lieu où poussent des plantes. Nous avons choisi ce nom suite aux remues-méninges effectués dans le projet LUISA (section 3.2.1.1). Dans ces remues-méninges, les utilisateurs ont établi à plusieurs reprises une distinction entre « ressources vivantes » et « ressources mortes ». De manière intuitive, ils ont caractérisé une « ressource vivante » comme une ressource où le travail humain apparaît en tant qu’auteurs ou qu’utilisateurs. Une « ressource vivante » est aussi une ressource qui évolue. Au contraire, une « ressource morte » est une ressource déposée dans un espace mais qui ne change plus. En référence à ces « ressources vivantes » que les membres de l’université semblaient attendre, nous avons décidé d’explorer les possibilités d’une plateforme où les ressources peuvent évoluer et où les utilisateurs sont conscients de la présence de l’activité des autres utilisateurs, favorisant éventuellement les contacts.

5.1 Contexte

5.1.1 Différentes communautés d'enseignants : exemple de Cartables.net

Expérimenter avec des communautés. Durant plusieurs mois, nous avons cherché à contacter des communautés pédagogiques afin de travailler avec elles sur la mise au point d'une plateforme de partage de documents. Nous avons alors contacté différentes communautés telles que WebLettres, Sesamath ou Cartables.net. Ces communautés parmi les plus actives sur Internet proposent généralement aux enseignants d'échanger des documents. Pour cela, chacune développe des outils ad hoc permettant le dépôt ou la recherche de ressources pédagogiques. Ces outils sont assez simples. Pour mieux comprendre le cas et les besoins de ces communautés, voici l'exemple de Cartables.net ayant mis l'accent sur un site à la conception graphique professionnelle et assez évolué techniquement.

Présentation générale. Nous avons contacté Cartables.net, un site s'adressant aux enseignants du primaire et proposant : des outils pour communiquer (listes de diffusion, forums et même un outil de chats), un répertoire de « liens pour la classe » (2000 liens annoncés) et un outil de partage de fichiers appelé « banque de fichiers ».

Le forum contient 4 320 inscrits et 277 590 messages. Le site est géré principalement par trois enseignants.

Motivation du partage de ressources pédagogiques. Pour nos travaux, nous nous sommes particulièrement intéressés aux fichiers et aux liens partagés.

La banque de fichiers, ses motivations et son fonctionnement sont présentés ci-dessous :

Ce qu'est la banque de fichiers :

- *Un espace ouvert de mutualisation de documents sans aucun a priori ou jugement de valeur, condition indispensable à nos yeux pour faire naître des échanges riches et variés.*
- *Nous avons choisi délibérément d'y accepter tous les documents postés à condition qu'ils respectent les droits d'auteurs et qu'ils correspondent à la description qui en est faite.*
- *Nous considérons que tout enseignant est apte à choisir ses outils, sa pratique, ses fichiers... que tout enseignant est capable de réflexion et d'esprit critique...*
- *Nous pensons qu'en étant le reflet de pratiques variées, la banque de fichiers peut aussi susciter des réflexions sur sa propre pratique pédagogique.*
- *En conséquence, la banque de fichiers ne peut être que ce que VOUS en faites. C'est aussi pour cette raison que nous avons mis à votre disposition un système d'évaluation des fichiers.*

Les points principaux que nous retenons de cette description sont la participation de tous les utilisateurs qui est mise en avant ainsi que leur égalité. Le site ne se donne pas d'autorité pour juger de la qualité des documents (autre que l'orthographe ou le droit d'auteur).

Interface de dépôt. Techniquement, l'utilisateur qui désire déposer une ressource sur le site peut soit envoyer un fichier, soit donner un lien hypertextuel pointant vers la ressource sur le Web. Un fichier ne peut excéder une taille de 500 Ko.

La ressource doit ensuite être décrite :

- en sélectionnant une catégorie parmi une cinquantaine proposée dans une liste mêlant thèmes, matières et manuels : Aide aux élèves en difficulté, Ateliers philo, Coopérative scolaire, Espace-Temps, Français - Généralités, Français - Lecture, Outils TICE, Recettes...
- en sélectionnant un type de documents parmi une douzaine : Activités-Exercices, Emplois du temps, Mémoires Professionnels, Textes Officiels...
- en remplissant un champ textuel libre de quelques lignes.

Recherche des fichiers. Les fichiers déposés peuvent être retrouvés en tapant des termes qui seront recherchés dans les descriptions attachées aux fichiers, et par filtrage en sélectionnant une catégorie ou un type de documents (par des listes identiques à celles proposées lors du dépôt).

Affichage des fichiers. La Fig. 5.1 montre comment sont affichés les résultats de la recherche.

eval-imparfait.zip		Eval imparfait				  604
 (1 vote) Évaluer ce fichier						
Français - Conjugaison	Activités/Exercices	Cycle 3	word	3 703	Sebb	16/03/09
affichage-temps.zip		Affichage présent indicatif, impératif imparfait, passé simple à agrandir en A3				  598
 (1 vote) Évaluer ce fichier						
Français - Conjugaison	Leçons/Aides-mémoire	Cycle 3	word	8 638	ValerieD	16/03/09

FIG. 5.1 – Présentation des résultats de la recherche de fichiers sur Cartables.net

En bas à gauche de chaque résultat, le lien « Évaluer ce fichier » permet aux utilisateurs d'évaluer une ressource en la commentant ou en la notant (Fig. 5.2). Ils peuvent aussi contacter éventuellement son auteur.

Évaluer le fichier "eval-imparfait.zip"

Soyez le plus objectif possible dans votre vote,
afin que le système d'évaluation garde tout son sens...
N'hésitez pas à laisser un commentaire justifiant votre notation.

1 2 3 4 5 6 7 8 9 10

Souhaitez-vous laisser un commentaire à l'auteur ?
Le propriétaire du fichier sera informé de votre adresse email
et pourra ainsi vous répondre si besoin.


VALIDER

FIG. 5.2 – Formulaire permettant d'évaluer le fichier d'un utilisateur, et possibilité d'interaction par email.

Limitation des dépôts. 14248 fichiers avaient été déposés jusqu'en mars 2009. Cette quantité importante a donc donné lieu à une politique de gestion présentée ainsi sur le site par les gestionnaires :

Le nombre de fichiers dans la banque de fichiers est très très important et donc difficilement gérable.

Il serait souhaitable, pour tous, avant d'envoyer un fichier, de réfléchir à ce qu'il peut réellement apporter à tous.

*Nous avons opté (après mûre réflexion et pour retarder la saturation de la banque et de ses gestionnaires) de **désactiver l'envoi de fichiers pendant le week-end** et de **limiter le nombre maximum de fichiers reçus à 10 par jour**. Cette décision n'est pas définitive, mais pour l'instant nous n'avons pas vraiment trouvé d'autre solution.*

Ainsi, malgré une description des fichiers par les déposants et malgré une vérification minimale des fichiers par les gestionnaires du site, ces derniers ont été obligés de limiter le nombre de fichiers déposés.

Conclusion. Parmi les sites d'enseignants que nous avons pu observer, plusieurs proposent de partager des documents pédagogiques (Weblettrés, Sésamath) mais Cartables.net était en 2007 celui ayant une des interfaces les plus évoluées de dépôts et de recherche. Pourtant, comme le montre la politique de limitation mise en place par les gestionnaires du site, le nombre de documents à gérer est trop important. Ceci est confirmé par les questions posées par mail à un gestionnaire du site (annexe H).

De ces cas, nous retenons l'idée du manque d'outils des sites d'association d'enseignants pour partager des fichiers pédagogiques, obligés de développer des solutions basiques et adhoc. De plus, dans l'esprit du Web Participatif et comme semblent l'indiquer les fonctionnalités de Cartables.net (descriptions par les auteurs, évaluation par les utilisateurs), les outils mis en place nécessitent une gestion des documents par les utilisateurs pour soulager les gestionnaires. Enfin, la possibilité de communiquer avec les auteurs d'un fichier, la mise en place de forums, chat' et listes de diffusion sur Cartables.net (mais aussi sur les autres sites d'association comme Sésamath) montrent le besoin de communiquer entre enseignants.

Dans ce cadre, une plateforme spécialisée dans le partage de fichiers pédagogiques et reprenant les fonctionnalités du Web Participatif nous semble adaptée aux besoins de ces communautés. En 2009, le site Cartables.net annonce qu'il va stopper son activité par manque de temps et d'énergie des gestionnaires. Des outils facilitant leurs tâches en reportant un peu de leur charge sur les utilisateurs semblent d'autant plus nécessaire.

5.1.2 Étude de la communauté PGM (Professeurs en Génie Mécanique)

Présentation de la communauté. Le contexte d'expérimentation que nous avons principalement choisi est celui d'une communauté de Professeurs de Génie Mécanique (PGM). Ce groupe est défini par l'usage d'une liste de diffusion nommée ListePGM.

ListePGM. Plus de 1500 professeurs de génie mécanique échangent depuis dix ans via la liste de diffusion ListePGM à raison de plus d'une centaine de messages par mois (hors vacances

d'été). ListePGM est une liste de diffusion liée aux groupes Yahoo! ⁶⁵. La configuration de la liste n'autorise pas les fichiers joints et n'autorise la diffusion des messages, qu'après modération de trois gestionnaires. Le gestionnaire principal est aussi le fondateur de la liste de diffusion.

Contenus des échanges. L'observation des messages de la liste montre que les enseignants échangent régulièrement des références vers des ressources techniques et pédagogiques en ligne. Ces ressources peuvent se trouver sur des sites institutionnels (e.g. Centre National de Ressources CNR-CMAO⁶⁶) ou sur des sites d'enseignants. En effet, un membre sur dix de cette communauté semble posséder un site Web⁶⁷.

Un mécanisme de synthèse publique a été mis en place sur la liste pour faciliter les échanges. Le principe est que lorsqu'un utilisateur a une question, les membres ayant une réponse la lui envoient non pas sur la liste en public mais par mail privé. Lorsque le membre a reçu suffisamment de réponses il envoie alors sur la liste une synthèse des réponses privées. Ce protocole permet d'éviter le nombre de messages envoyés sur la liste et d'obtenir des messages de synthèse riches.

Communauté de pratique. Les questions et les réponses échangées sur la liste nous permettent de dire que la ListePGM se rapproche d'une communauté de pratique [Wenger, 1998] et plus spécifiquement d'une communauté de pratique en ligne selon Elise Garrot [Garrot, 2008] :

Une Communauté de Pratique en ligne rassemble des membres de manière informelle dans un espace en ligne spécifique du fait qu'ils

- *ont des pratiques, des centres d'intérêts et des buts communs (e.g. partager des idées et des expériences, construire des outils communs, développer des relations entre pairs) ;*
- *échangent et s'entraident afin de développer leurs compétences et expertise et résoudre des problèmes de manière innovante ;*
- *développent une identité communautaire autour des connaissances partagées et des approches et pratiques communes établies ;*
- *et créent un répertoire partagé de ressources communes.*

Une communauté qui se structure. En tant qu'organisation humaine, cette communauté connaît plusieurs phénomènes de « structuration » quant aux rôles des membres, aux règles de modération et aux outils de partage. Le phénomène de structuration se produit par exemple avec la mise en place de règles de modération régissant la liste, devant être assez strictes pour éliminer les messages hors sujets et cependant assez souples pour encourager la participation et les échanges. L'organisation de la communauté s'est transformée, après dix ans de fonctionnement informel, en une association de loi 1901 (Union PGM, de 200 membres). A présent, le président-fondateur de la communauté, qui administre toujours la liste de diffusion, est entouré d'autres membres siégeant au conseil d'administration.

Un besoin d'outils adaptés. La constitution en association en 2008 s'est accompagnée d'une relance de la réflexion sur les outils de partage. En effet, la communauté se trouvait limitée dans ses moyens techniques de partage, une liste de diffusion étant peu pratique pour capitaliser les références vers les ressources en ligne. Les observations de la liste de diffusion, des entrevues

⁶⁵fr.groups.yahoo.com

⁶⁶SITE CNR-CMAO

⁶⁷Un annuaire de la communauté affiche 150 sites enseignants pour 1500 membres

et un sondage auxquels 160 membres ont répondu⁶⁸, révélèrent le besoin pour la communauté de regrouper et d'organiser les ressources éparpillées sur les nombreux sites institutionnels ou personnels. Les outils explorés furent un site Web classique.

Un site Web classique. Depuis plusieurs années, le fondateur de la liste de diffusion avait mis en place un site Web (ListePGM.fr) pour accompagner la liste. Ce site était maintenu principalement par lui et affichait de nombreuses informations utiles à la communauté comme les dernières nouvelles de la discipline, un annuaire de sites d'enseignants ainsi que des ressources sélectionnées (livres ou ressources en ligne).

Choix du contexte. Nous avons décidé de nous concentrer sur cette communauté pour plusieurs raisons :

- Nous pouvions travailler dessus avec un doctorant en sciences de l'éducation, Eric Dané.
- Les outils de la communauté montraient l'usage d'un échange de ressources.
- Les membres de la communauté produisaient et diffusaient de manière notable des ressources à travers les sites des enseignants.
- La liste de diffusion rencontrait des limites dans l'archivage et l'exploration des ressources.
- La structuration de la communauté en association officielle favorisait une phase de mutation rendant possible l'intégration de nos propositions.

5.2 Développements

5.2.1 Objectifs

Motivations. De l'observation des sites d'enseignants et des sites des communautés enseignantes, nous avons repéré le besoin d'outils pour faciliter la mise en ligne et l'organisation de ressources pédagogiques.

De l'étude des sites du Web Participatif, nous avons repéré les principes et les fonctionnalités soutenant la participation des utilisateurs, nécessaires au dépôt et à l'organisation des contenus.

Des séances de remues-ménages du projet LUISA, nous avons repéré le besoin de « ressources vivantes », c'est à dire de ressources où la présence humaine était plus apparente et caractérisée par son côté évolutif.

Objectifs. En accord avec les motivations ci-dessus, nous avons réalisé le modèle d'une plateforme soutenant le partage de documents pédagogiques. Notre principale source d'inspiration était le site YouTube pour le partage de vidéos. Nous voulions explorer les mêmes principes appliqués au contexte pédagogique.

Méthode de travail. Dans l'intention d'un développement itératif et participatif, nous avons désiré réaliser une plateforme aux fonctionnalités minimales (dépôts de fichiers) puis attirer des utilisateurs pour travailler avec eux sur les fonctionnalités plus avancées.

⁶⁸Sondage ListePGM, réalisé par Eric Dané : <http://festic.fr/pgm/questionnaire-0/resultats.php>

5.2.2 Principes d'une plateforme participative de partage de ressources pédagogiques

Les principes de l'outil que nous proposons s'inscrivent soit dans la tendance du Web Participatif reprenant les principes des sites Participatifs étudiés (4.1.9), soit plus particulièrement dans l'exploration des « ressources vivantes ».

Nous proposons donc les principes suivants pour une plateforme participative spécialisée dans le partage de ressources pédagogiques et adaptés aux sites enseignants :

- Faciliter le dépôt et l'organisation des ressources ;
- Permettre l'intégration des ressources dans d'autres sites.

Pour explorer les possibilités de « ressources vivantes », nous proposons les principes suivants :

- Contextualiser les ressources ;
- Faciliter les discussions entre enseignants autour des ressources ;
- Permettre de suivre l'évolution des ressources lors de leurs réutilisations.

5.2.3 Fonctionnalités soutenant la participation

Pépi⁶⁹ est un site Web proposant aux enseignants de déposer des documents pédagogiques et permettant leur recherche. Voici les fonctionnalités issues des principes présentés ci-dessus.

5.2.3.1 Faciliter le dépôt de ressources

Dépôt et description simplifiés des ressources. Le site international de ressources pédagogiques MERLOT proposait le dépôt en cinq pages. S'il n'est pas gênant de passer du temps à déposer un document lorsqu'on est documentaliste professionnel, cela devient dissuasif pour des individus dont ce n'est pas la tâche principale. De plus, décrire des documents de tailles importantes peut sembler plus envisageable que passer du temps sur des documents de taille plus réduites où la description prendrait presque autant de temps que la production.

Dans Pépi, une fois inscrit sur le site, cette tâche s'effectue en un minimum de champs comme le montre la Fig. 5.3 : un titre libre, une description libre, éventuellement un dossier-label (sur lesquels nous reviendrons plus tard) puis la ressource en elle-même (URL ou fichier).

Des ressources diffusables. Pour faciliter la réutilisation, tous les contenus déposés sont automatiquement sous licence libre de ré-utilisation et modification sous contrainte de mentionner l'auteur original et de ne pas exploiter le contenu de manière commerciale⁷⁰. Sur le même principe, les ressources ne peuvent être limitées pour un sous-groupe ou mise en privé. En effet d'après les témoignages recueillis lors des remues-ménages du projet LUISA, les participants mentionnaient que les différents niveaux de visibilité offerts par la plateforme Moodle (seulement pour un groupe ou un sous-groupe, etc) avaient tendance à compartimenter les ressources.

5.2.3.2 Permettre l'intégration des ressources sur d'autres sites

Exporter des contenus sur les sites des enseignants. Les enseignants de génie mécanique possédant un site Web y affichent généralement des ressources pédagogiques comme des cours ou des activités employées en classe. Or la maintenance de ces fichiers prend du temps pour les

⁶⁹Pépi est accessible : <http://pepi.fr>

⁷⁰Creative Commons

The screenshot shows the 'La Pépinière Pédagogique' website interface. At the top, there is a green banner with the text 'La Pépinière Pédagogique' and 'partager des documents, rencontrer d'autres enseignants'. Below the banner is a navigation menu with links: 'Accueil', 'Explorer', 'Rechercher', 'Discuter', 'A Propos', and 'Aide'. The main content area contains a form for adding or editing a resource. The form has the following fields and elements:

- Titre :** A text input field containing 'Démarche par l'analyse des contacts'.
- Description :** A large empty text area.
- Mettre dans un dossier :** A dropdown menu showing '-- Modélisation cinématique des assemblages'.
- Auteur :** A text input field containing 'evobg'.
- Url :** An empty text input field.
- ou Fichier (<10Mo) :** A text input field followed by a 'Browse...' button.

Below the form, there are three buttons: 'Sauvegarder', 'Supprimer', and 'Annuler'.

FIG. 5.3 – Ajouter ou éditer une ressource dans Pépi en un minimum de champs

organiser de manière claire, et de l'argent pour acheter de l'espace et y déposer des contenus faisant parfois entre 10 et 20 Mo comme les modèles en 3D employés en Génie Mécanique.

L'idée est de résoudre ces difficultés en permettant aux enseignants de déposer les contenus pédagogiques sur la plateforme Pépi puis d'intégrer ces contenus sur leurs pages Web en y collant un morceau de javascript de quelques lignes. Les contenus textuels apparaîtraient par exemple dans un petit encadré au format pdf, permettant de passer d'une page à l'autre.

Le site YouTube permet aux utilisateurs de déposer des vidéos de plusieurs dizaines de méga-octets tout en les intégrant ensuite sur leurs propres sites Web.

Recherche des contenus à partir des sites enseignants : API. Une API⁷¹ sur Internet est une sorte de protocole permettant à un site d'interroger un autre site, de recevoir les résultats et de les afficher.

Une telle API sur le site Pépi permettrait à un enseignant de disposer un moteur de recherche Pépi sur son site, qui interrogerait la base des ressources Pépi et afficherait les résultats sur le site de l'enseignant.

Par manque de temps, ces fonctionnalités ne purent être développées sur Pépi.

5.2.3.3 Contextualiser les ressources

Contexte humain. S'intéresser au contexte humain de la ressource signifie qu'au lieu de montrer des ressources isolées de leur auteur, celui-ci est mis en avant avec l'affichage de son identifiant mais aussi d'un lien permettant d'accéder à son profil.

⁷¹API, Application Programming Interface

Dossiers-labels. L'originalité de Pépi par rapport à d'autres sites de dépôts est de permettre à l'auteur d'organiser les ressources déposées dans une arborescence personnelle de dossiers (Fig. 5.4). Cette fonctionnalité reprend l'organisation des documents sur les sites des enseignants où la classification s'effectue par exemple par année scolaire, filière, cours, etc..

FIG. 5.4 – Les ressources d'un dossier apparaissent sur une page, avec leur description et les fichiers associés. En haut apparaît l'arborescence permettant de contextualiser les ressources.

Les avantages de cette organisation par dossiers est que lorsqu'un utilisateur retrouve une ressource, il peut voir aussi les autres ressources de l'auteur. Par exemple, une activité se trouvera en général dans le même dossier que le cours et l'examen qui l'accompagnent.

Ces dossiers-labels se distinguent des dossiers ou répertoires sur l'ordinateur par deux points :

- L'utilisateur peut ajouter une description aux dossiers-labels. Par exemple pour un dossier « Archives » : « Ce dossier contient les documents que je trouve utiles mais que je n'ai pas employé directement dans mes cours » .
- Les dossiers-labels sont employés lors des recherches comme base d'indexation. C'est l'aspect « label ».

En effet, une recherche effectuée parmi les ressources de Pépi explore les titres et les descriptions des ressources mais aussi l'arborescence des dossiers qui servent alors chacun de description. Exemple : voici l'extrait de l'arborescence de dossier d'un utilisateur :

- Mécanique du Solide Niveau BEP
 - Cinématique du solide
 - Statique du solide

Pépi déduit de cette arborescence que si un fichier est contenu dans le dossier « Statique du solide » alors il est décrit par les termes du dossier courant « Statique du solide » mais aussi

par les termes de tous les dossiers le contenant à savoir ici « Mécanique du solide niveau BEP » . Chacun des termes (i.e. « Mécanique », « solide », « BEP », « Statique » ...) sont autant de mots-clés qui permettront de retrouver la ressource ultérieurement, notamment par un autre utilisateur.

Nous n'avons pas implémenté directement de système de tags car les sites Webs des enseignants proposaient souvent des arborescences simples entre documents. Nous avons fait l'hypothèse que des tags ne permettraient pas de rendre compte de cette hiérarchie entre dossiers.

De plus, les dossiers-labels ont l'avantage des tags d'être choisis librement par les enseignants et d'être recherchables par l'application comme des tags. Ils offrent en outre la possibilité d'une arborescence des contenus (impossible avec des tags). Néanmoins, ils ont le défaut de ne pas permettre à un même utilisateur de posséder une ressource dans plusieurs dossiers.

Enfin, pour faciliter l'organisation des ressources, lorsqu'un utilisateur crée un compte, son espace contient automatiquement un premier dossier-label nommé selon l'année courante (e.g. nommé « Année scolaire 2009-2010 »).

5.2.3.4 Faciliter les discussions entre enseignants (notamment autour des ressources)

Comme nous l'avons vu dans l'étude des sites Participatifs, les possibilités d'interactions entre utilisateurs peuvent être une source de motivation pour participer.

Discussion à tous les niveaux. A l'image des commentaires sur les articles de blogs, Pépi implémente des zones de discussion pour chaque dossier-label et pour chaque ressource (Fig. 5.5).

Un forum général permet aussi aux utilisateurs de discuter. Il est accessible par le bandeau supérieur à l'onglet « Discuter » (Fig. 5.3).

Chaque communauté possède aussi un forum de discussion distinct pour ses membres.

Profil Une des particularités des sites participatifs est de généralement permettre aux utilisateurs d'éditer une page « profil » les présentant. Ces pages permettent de mettre une identité un peu plus précise sur un identifiant, éventuellement une photo, ou alors les motivations de l'utilisateur, par exemple sur les pages profils de Wikipédia.

Pépi permet ainsi aux utilisateurs de se décrire à l'aide d'un champ texte libre et d'un lien vers leur site personnel. D'un point de vue pédagogique, le profil affiche aussi la communauté à laquelle appartient l'utilisateur ainsi que les classes dans lesquelles il enseigne (Fig. 5.6).

Ces classes sont un moyen de permettre aux utilisateurs de trouver rapidement des collègues proches. Le problème est qu'il existe une très grande variété de classes et de filières différentes. Pour répondre à cette difficulté, le système définit une classe selon trois champs possibles, niveau, filière et discipline, qui peuvent être remplis de manière libre. En revanche, quand un utilisateur commence à taper les premiers caractères dans un champ, ce dernier propose toutes les valeurs déjà rentrées. Par exemple, la Fig. 5.7 affiche la discipline Électrotechnique, suite à l'entrée du premier caractère « e » par l'utilisateur.

Ressources réalisées par : laurentgranjon

Dans [ressources de laurentgranjon](#)

Cours de Mécanique du point matériel

Les deux chapitres de cours que vous trouverez ici correspondent au cours que j'ai enseigné de 1997 à 2001 en première année de l'ESISAR. Je n'ai jamais eu le temps de dactylographier la fin de ce cours (chap 3 à 10) et n'enseigne plus cette matière. Vous trouverez toutefois ci dessous la partie correspondant à la résolution d'équations différentielles et aux oscillations forcées.

Chapitre 1 : Introduction	Télécharger : mp_ch-1.pdf (0.01Mo)
Chapitre 2 : Cinématique	Télécharger : mp_ch2.pdf (0.64Mo)
Equations différentielles du second degré	Télécharger : equa_diff.pdf (0.09Mo)

Autres ressources de 'laurentgranjon'

1 sujet(s) dans ce forum - [Poster un sujet](#)

benjaminhkb : 2009-04-03 17:21:27	"Merci" Merci pour tous ces documents....	0 réponse(s)
---	---	------------------------------

FIG. 5.5 – Chaque dossier et chaque ressource peut être discuté à l'aide d'un mini-forum contextualisé situé en bas de page (un peu comme les commentaires sur les blogs)

Comme les ressources, les profils peuvent donner lieu à des discussions.

Communautés. Lors de son inscription, un utilisateur peut s'ajouter à une communauté. Ces communautés ont été définies par nos soins à partir de quelques communautés observées sur le Web ou rencontrées : professeurs des écoles (cartables.net), professeurs de mathématique (Sésamath), professeurs de génie mécanique ou documentalistes.

Une communauté permet de définir un contexte pour les enseignants. La communauté est affichée dans le profil. Chaque communauté possède aussi une page à part récapitulant les membres de la communauté, les dernières ressources déposées et les dernières discussions.

Navigation parmi les utilisateurs. L'utilité des profils et des communautés est renforcée par une option de navigation parmi les utilisateurs (Fig. 5.8) disponible dans l'onglet « Explorer » du bandeau principal visible en Fig. 5.3. La navigation peut ainsi être filtrée selon les classes ou selon les communautés existantes et affichées dans la partie supérieure de la fenêtre.

Login :	evobg
Site web :	http://evobg.free.fr/
Description et intérêts :	Enseignant en construction mécanique, je propose, ici, mes propres ressources pédagogiques aux niveaux BEP, BAC PRO et BAC STI. Je vais essayer d'organiser sur Pépi, les ressources provenant de : http://evobg.free.fr/
Classes :	
Communauté :	Génie Mécanique
Date de création :	2007-05-18 16:10:55
Date de mise à jour :	2007-06-25 22:42:32
Ressources :	<u>29 ressources dans 12 dossiers</u>

Laisser un message à evobg

0 sujet(s) dans ce forum - Poster un sujet
Aucun sujet pour le moment.

FIG. 5.6 – La page profil d’un utilisateur présente ses caractéristiques mais aussi une zone inférieure permettant de commencer une discussion avec lui (à l’image des « murs » de Facebook)

Classes actuelles (opt.) :	Niveau : <input type="text" value="Terminale"/>	Filière (opt.) : <input type="text"/>	Discipline (opt.) : <input type="text" value="E"/>
	<input type="button" value="Ajouter"/>		<input type="button" value="Electrotechnique"/>
	Ex : Terminale S Français Astuce : Taper les premières lettres fera apparaître une liste de suggestions.		

FIG. 5.7 – Les enseignants peuvent renseigner librement les classes dans lesquelles ils travaillent. Des suggestions sont proposées à partir des premiers caractères tapés.

5.2.3.5 Permettre de suivre l’évolution des ressources lors de leurs réutilisation

Ce point ne put être développé pour cause de manque d’utilisateurs. L’objectif, dès le début du projet, était de permettre de suivre l’évolution de chaque ressource, à l’image du suivi des modifications des articles de Wikipédia (historique des modifications). Néanmoins, une ressource pédagogique téléchargeable par des enseignants et modifiable à volonté sur leurs ordinateurs personnels ne permet pas le même suivi que celui des modifications d’un article édité en ligne.

Pour suivre ces évolutions, nous avons pensé nous inspirer des « réponses vidéos » de la plateforme YouTube. Sur cette plateforme, toute vidéo peut être commentée par un utilisateur à l’aide d’une autre vidéo qui sera par exemple une parodie de la première ou une vue complémentaire sur le sujet traité. En reprenant ce système de « réponses » volontaires, nous pensions laisser à un utilisateur la possibilité de copier un contenu d’un autre utilisateur sur son espace personnel puis de télécharger le fichier de la ressource, de le modifier comme bon lui semble en local, avant d’uploader le fichier modifié. En mettant en ligne le fichier modifié, l’utilisateur aurait désigné de quel fichier il se serait inspiré. Même si le mécanisme restait imprécis, ceci aurait permis, nous l’espérons, de conserver une trace des évolutions d’une ressource.

Liste des utilisateurs (Total : 32)

Filtrer par communautés :
 Toutes - **Génie Mécanique** - Primaire - Mathématiques - Documentation -

Filtrer par classes : (vos classes)
 Niveaux : **Tous** - Bac Pro - BEP - BTS - CAP - CM1 - CM2 - CP - L3 - Première - Terminale -
 Filières : **Toutes** - AVA - carrosserie - Sc. Éducation - STI -
 Disciplines : **Toutes** - AFSM - construction - Électrotechnique - sciences et techniques -

Login	Description	Ressources/dossiers	Classes	Création
benjaminhkb	Je suis le développeur principal de Pépi.	<u>0 res./0 dos.</u>		07-2007
jfraisse	Professeur de construction mécanique Lycée Raoul Dautry - Limoges	<u>2 res./0 dos.</u>		07-2007
cfresne	Enseignant en construction, Lycée de l'automobile et de la carrosserie - Niort	<u>0 res./1 dos.</u>	<ul style="list-style-type: none"> • BTS AVA - AFSM • Bac Pro Carrosserie - Construction • BEP carrosserie - construction • CAP carrosserie - construction 	06-2007
evobg	Enseignant en construction mécanique, je propose, ici, mes propres ressources pédagogiques aux niveaux BEP, BAC PRO et BAC STI. Je vais essayer d'organiser sur Pépi, les ressources provenant de : http://evobg.free.fr/ Site : http://evobg.free.fr/	<u>29 res./12 dos.</u>		05-2007

FIG. 5.8 – Il est possible de visualiser tous les utilisateurs inscrits avec la possibilité de filtrer par communautés, classes, filières et disciplines éventuelles

Pour l'auteur d'une ressource, un des principaux avantages de ce suivi des évolutions aurait été la satisfaction de voir son travail réutilisé. Pour l'application, ce suivi aurait éventuellement permis de mettre en avant les ressources ayant inspiré de nombreuses autres ressources, par un mécanisme basique de popularité. Enfin pour les chercheurs, ceci aurait pu laisser la chance de suivre les évolutions et les modifications d'un contenu pédagogique par plusieurs enseignants, dans plusieurs contextes.

5.2.4 Architecture technique

Pépi est une application Web développée à l'aide du langage PHP et d'une base de données MySQL. Afin d'accélérer le développement, nous avons employé un framework de développement PHP nommé Symfony⁷² qui aide à structurer une application PHP et offre des bibliothèques facilitant certains aspects du développement Web (par exemple, librairie AJAX pour les interfaces).

La base de données du projet contient 6 tables :

- « user » : pour la gestion des utilisateurs,
- « classroom » : pour la gestion des classes associées à chaque utilisateur,
- « comment » : pour la gestion des forums et tout ce qui est en rapport avec les discussions,

⁷²Site du framework Symfony : <http://www.symfony-project.org/>

- « community » : pour la gestion des communautés (Génie Mécanique, Primaire...),
- « res » et « contain » : pour la gestion des ressources et des dossiers les contenant.

Ceci correspond à la maintenance de plus de 900 fichiers sources occupant 3,6 Mo mais dont une partie fut produite par le framework Symfony.

5.3 Expérimentations

Malgré nos efforts pour faire participer des enseignants, nous n'avons pas réussi à provoquer l'intérêt nécessaire à des expérimentations satisfaisantes.

5.3.1 Contact des utilisateurs

Lors de nos travaux nous avons principalement contacté quatre groupes d'utilisateurs :

1. Nous avons contactés un responsable de Cartables.net, site réunissant des enseignants du primaire (section 5.1.1).
2. Nous avons contacté par mail un responsable de Sésamath, qui nous a renvoyé vers une liste de diffusion de sésamath spécialisé dans les TICE (Technologies de l'Information et de la Communication pour l'Enseignement).
3. Nous avons présenté la plateforme et ses principes pendant une heure devant deux classes d'élèves-documentalistes à l'IUFM⁷³ de Rouen.
4. Nous avons contacté à plusieurs reprises le responsable de ListePGM, par mail ou en face-à-face, ainsi que les membres de ListePGM, par plusieurs emails sur la liste de diffusion. Ceci devait être notre communauté principale.

5.3.2 Amorçage

Le démarrage est une phase importante du partage de documents, car une plateforme vide n'attirera pas d'utilisateurs, qui ne déposeront pas de documents et la plateforme restera vide.

Pour pallier ces difficultés, nous avons contacté trois professeurs de génie mécanique pour leur demander l'autorisation de déposer sur Pépi les contenus pédagogiques de leurs sites Web. Ces contenus ont permis de constituer une base de démarrage de 42 documents répartis dans 16 dossiers-labels et ont permis de tester la structuration en dossiers-labels.

Nous avons par ailleurs déposé 61 documents dans 18 dossiers pour le niveau primaire.

5.3.3 Participations

Suite à ces contacts entre mars 2007 et juin 2007, le résultat fut que nous n'avons pas pu créer de partenariat avec les communautés et que la participation sur la plateforme fut très faible.

D'un point de vue partenariat, aucune association (Cartables.net, Sésamath ou la communauté PGM) ne fut intéressée par notre proposition de collaboration où nous proposons de les aider à développer un outil participatif de partage de fichiers en échange de l'observation des

⁷³IUFM : Institut Universitaire de Formation des Maîtres

échanges.

D'un point de vue participation des utilisateurs, nos présentations et nos emails ont touché un certain nombre d'utilisateurs car il y eut :

- 4 inscrits en professeur des écoles,
- 4 inscrits en mathématiques,
- 4 inscrits en documentation,
- 2 inscrits en génie mécanique,

5.4 Analyses et discussions

5.4.1 Analyses de l'échec des expérimentations

Pas de participation sur Pépi. Le résultat principal issu de ces tentatives d'expérimentations est le non-attraire de Pépi pour les enseignants et pour les communautés enseignantes. Ce non-attraire s'illustre au niveau des enseignants par la non participation sur la plateforme, et au niveau des communautés par la non possibilité de tisser des partenariats dans le développement de la plateforme.

Plusieurs causes possibles. Cette non-participation, ironique pour une plateforme participative, peut s'expliquer par des difficultés rencontrées au niveau :

- de l'implémentation de plateforme ;
- des concepts sous-jacents ;
- de notre manière d'expérimenter ;
- de notre manière d'aborder la recherche en EIAH.

5.4.1.1 Analyse de l'implémentation.

Difficultés ergonomiques. Parmi les concepts facilitant la participation, un point important est la simplicité du système. Or les retours que nous avons eu des enseignants mentionnaient des difficultés pour comprendre l'interface des documents déposés.

Difficultés d'implémentation. Nous avons perdu beaucoup de temps à résoudre des problèmes dus au partage de documents en soi. Les enseignants en génie mécanique manipulent de lourds fichiers (entre 10 et 20Mo) que nous voulions pouvoir diffuser. Ce service aurait permis éventuellement d'attirer ces enseignants. Néanmoins ce genre de dépôts pose des difficultés techniques importantes. De plus, posséder des serveurs gérant les fichiers de tailles importantes a un coût non négligeable.

Partir de rien. La plus grande difficulté fut de partir de rien pour réaliser la plateforme. Nous pensions que la plateforme centrée sur les ressources ne cacherait pas de difficultés techniques trop importantes. Or ce fut une illusion et nous avons perdu beaucoup de temps dans le développement malgré le support du framework Symfony.

Concurrence des outils. Nous désirions partir d'un outil neuf pour implémenter à notre guise les concepts que nous voulions expérimenter. Nous avons découvert lors des expérimentations qu'il existe de nombreux outils de partage de documents. Par exemple, les élèves-documentalistes

de l'IUFM de Rouen étaient formés à employer BSCW⁷⁴. Cette plateforme était obligatoire durant l'année scolaire. Pépi, même avec une approche des documents différentes, était un doublon. Nous aurions dû donc réaliser un outil intégré à un autre.

5.4.1.2 Analyse des concepts sous-jacents

La non-participation des utilisateurs à Pépi nous semble plus remettre en question la manière dont les fonctionnalités étaient implémentées plutôt que les concepts sous-jacents.

Partager au niveau global. Une des caractéristiques de la plateforme est d'être disponible à un niveau global car elle se trouve sur le Web et n'importe qui peut s'y inscrire pour déposer des documents. Néanmoins Pépi permet d'appartenir par la suite à un niveau plus limité à l'aide des communautés. L'idée était de pouvoir accueillir n'importe quelle communauté enseignante.

A l'image des grandes plateformes participatives comme YouTube et Wikipédia, il semble intéressant de proposer un service à un niveau global, permettant ainsi de recueillir une participation plus importante, quitte à la contextualiser en sous-groupes par la suite. Néanmoins, à l'échelle de nos moyens, il fut peut-être illusoire de désirer répondre à des besoins trop larges sans cibler plus précisément une communauté.

Légitimité à recueillir des documents pédagogiques. Un point que nous avons découvert durant l'expérimentation est l'importance de l'adresse Internet qui valide les documents pédagogiques. Un document sur le site du ministère n'est pas perçu de la même manière que le même document sur le site d'un enseignant. De la même façon, télécharger des documents sur le site d'une association d'enseignants reconnue n'a pas le même sens que le même document sur une plateforme inconnue.

Ce problème fut confirmé lors d'un entretien avec le fondateur de Edumedia-share.

Edumedia⁷⁵ est une entreprise réalisant des animations multimédia pédagogiques (en Flash) dans le but d'être utilisées par des enseignants en classe.

Edumedia-share est un site de partage de contenus multimédia déposés par les enseignants. Ce site répond à de nombreux critères listés comme fonctionnalités participatives (4.1.9) : facilité de dépôts, facilité de description à l'aide de tags, commentaires... Edumedia a mis beaucoup d'énergie à créer un site au design agréable et pour lancer le partage, il a rémunéré des enseignants pour qu'ils y déposent des contenus de qualité.

Néanmoins, suite à une entrevue avec un des fondateurs d'Edumedia, nous avons appris que le site était pour eux un échec. Ils n'ont pas réussi à lancer la dynamique de partage. Plusieurs explications possibles ont été formulées :

- Déposer des contenus sur un site externe est problématique pour les auteurs.
- Déposer des contenus sur un site côtoyant une vitrine commerciale (Edumedia) n'a pas rassuré les enseignants.

⁷⁴Plateforme collaborative BSCW : <http://public.bscw.de/>

⁷⁵Edumedia-share, [sitedepartaged'imagesouschémasàviséepédagogiques:http://www.edumedia-share.com/](http://www.edumedia-share.com/)

5.4.1.3 Analyse de nos expérimentations

Développer participativement. Nous avons arrêté le projet Pépi prématurément car nous ne voulions pas développer une plateforme de partage sans implication d'utilisateurs dans le développement et dans l'utilisation.

Précipitation. A vouloir à tout pris accrocher l'intérêt des enseignants, il se peut que nous ayons précipité plusieurs étapes. Le manque de retours sur les usages en est une possible conséquence.

Manque de contacts. La plateforme se voulait être intégrable dans un maximum de communautés, or nous n'en avons contacté que quatre. Il aurait peut-être fallu essayer de toucher beaucoup plus de groupes pour avoir une chance de trouver des utilisateurs disponibles.

Difficulté avec la communauté principale. La communauté PGM devait être notre principal terrain d'expérimentation. Mais, malgré nos efforts pour tisser un lien avec celle-ci, nous n'avons pas réussi à gagner la confiance des responsables et par conséquent des utilisateurs. Ainsi, nous n'avons jamais réussi à faire mettre un lien vers notre outil à partir du site principal. Nous n'avons pas non plus réussi à y intéresser d'utilisateurs pour former un petit noyau d'utilisateurs impliqués.

5.4.1.4 Analyse de la recherche en EIAH (selon notre expérience)

Tension entre innovation et efficacité. La difficulté à créer une plateforme participative de partage provient, selon nous, des difficultés à créer à la fois un outil répondant aux besoins immédiats des utilisateurs pour les faire adhérer et en même temps à proposer des solutions nouvelles à explorer d'un point de vue Recherche.

Manque de légitimité des chercheurs. Notre volonté de soutien aux communautés ne fut jamais reconnue. Alors que les communautés enseignantes manquent de moyens et que nous proposons un soutien technique gratuit, nous avons reçu un accueil plutôt froid de la part des communautés que nous devons toujours relancer. L'exemple le plus flagrant fut avec la communauté ListePGM où le contexte semblait idéal pour nos contributions et où rarement nous fûmes interpellés. La raison principale de ces difficultés est, selon nous, le manque de légitimité de la communauté EIAH auprès des enseignants. Combien d'enseignants ont déjà entendu parler des EIAH? Au delà de notre travail, ceci nous semble un aspect préoccupant pour une communauté de Recherche âgée d'au moins une vingtaine d'année.

5.4.2 Discussion des concepts et travaux similaires

5.4.2.1 Caractéristiques d'une plateforme support à une communauté en ligne

[Garrot, 2008] s'intéresse aux moyens de soutenir le partage de documents et d'expériences entre tuteurs universitaires. Elise Garrot s'intéresse plus particulièrement aux tuteurs en tant que Communauté de Pratique en ligne (voir définition 5.1.2). Elle propose alors cinq grandes caractéristiques que doivent posséder les plateformes pour soutenir une Communauté de Pratique en ligne :

- Offrir un espace supportant l'identité communautaire
- Favoriser l'entraide et l'échange entre pairs

- Favoriser la collaboration et la réflexivité
- Proposer des ressources contextualisées
- Favoriser la recherche contextuelle

Nos travaux ont démarré non pas par l'approche des Communautés de Pratique mais par l'étude de la tendance du Web Participatif. Nous trouvons intéressant de comparer alors les fonctionnalités de Pépi aux catégories proposées par Elise Garrot. Pour chacune d'entre elle, nous rappellerons la définition d'Elise Garrot puis nous feront le lien avec Pépi.

Offrir un espace supportant l'identité communautaire « *De par l'entreprise commune comprise et continuellement renégociée entre les membres, ceux-ci développent un sentiment d'appartenance à la communauté et adhèrent à son identité. Au cours du temps, l'engagement au sein d'une pratique commune fait que les membres acquièrent une reconnaissance tout d'abord par leurs pairs, mais également au niveau institutionnel en définissant leur activité par des artefacts ou un vocabulaire adopté par tous. Il est alors nécessaire d'apporter une plateforme informatique supportant la construction et l'évolution de l'identité de la communauté par ses membres : l'engagement libre des personnes, la création de ressources communes et la mise en place d'un vocabulaire reflétant la spécificité de l'activité.* »

Dans Pépi, contrairement aux travaux d'Elise Garrot, la plateforme ne s'adresse pas à une seule communauté en particulier mais est un service ouvert à plusieurs communautés. En revanche, chaque communauté possède un espace distinct sur Pépi, avec la liste des membres, des ressources et des espaces de discussion distincts.

Comme dans les travaux d'Elise Garrot, la plateforme permet de s'inscrire librement comme membres et la visibilité des actions de tous favorise la perception de la participation des membres à la communauté.

Favoriser l'entraide et l'échange entre pairs « *La dimension d'engagement mutuel de la CoP suppose un rapport d'entraide entre les membres, nécessaire au partage de connaissances sur la pratique.* »

Dans ses travaux, Elise Garrot s'intéresse au partage en général allant du partage de documents au partage d'expériences. Dans Pépi, le but principal visait au partage de documents pédagogiques. Les discussions et les rencontres étaient prévues plutôt autour de ces documents échangés. Nous considérons que la liste de diffusion était un outil pratique et suffisant pour les discussions générales.

Favoriser la collaboration et la réflexivité « *Le développement professionnel repose en partie sur la réflexivité sur les pratiques. Dans le cas des enseignants, un bon enseignement n'est pas un acte répétitif mais incorpore les leçons apprises de l'expérience dans les pratiques d'enseignement pour les faire évoluer. C'est pourquoi nous pensons essentiel de leur fournir les moyens de témoigner de leurs expériences, même sans intention de les diffuser aux autres membres. [...]*

Les membres, en interagissant, sans forcément en avoir conscience d'ailleurs, peuvent construire un répertoire de ressources très riche en expériences, en bonnes pratiques, en documents, etc.,

qui constituent le répertoire partagé de la communauté. »

Contrairement à la plateforme d'Elise Garrot, Pépi ne permet pas explicitement de déposer un témoignage écrit et n'y incite pas. La réflexivité s'effectue dans l'ordonnancement de ses propres documents pédagogiques et dans la vision des documents pédagogiques des autres enseignants.

Proposer des ressources contextualisées « *Comme les ressources produites par la communauté ne prennent forme et n'ont de sens que dans un contexte qui leurs sont propres, l'enjeu majeur de nos travaux est de parvenir à « saisir » ce contexte pour le lier à l'artefact produit (message, document, discussion, etc.). »*

Dans Pépi, chaque ressources pédagogiques apparaît dans des dossiers-labels contenant d'autres ressources proches réalisées par le même auteur et donnant ainsi une première vision du contexte pédagogiques. De plus, les auteurs peuvent ajouter des descriptions aux dossiers pour préciser leur démarche pédagogique. Enfin, si les utilisateurs ont des questions ou des remarques sur les documents, ils peuvent les discuter publiquement dans les zones de discussion attachées à chaque document. Ces discussions ont pour but d'explicitier encore le contexte lié au fichier.

Favoriser la recherche contextuelle « *Le système informatique a pour rôle de donner facilement et rapidement accès aux ressources pertinentes, dans le contexte de la pratique de chaque membre. »*

Dans Pépi, le contexte d'un enseignant est principalement défini, dans son profil, par sa communauté et par les classes dans lesquelles il déclare enseigner. Le moteur de recherche de Pépi utilise ces éléments pour fournir des documents correspondant à ses recherches. Contrairement aux travaux d'Elise Garrot, Pépi ne propose pas de gestion de compétences plus précise. Selon l'approche observée dans les sites du Web Participatif, l'idée est que les utilisateurs ont du mal à remplir des profils détaillés selon des taxinomies précises.

Conclusion de la comparaison. Nos travaux et ceux d'Elise Garrot se rejoignent dans l'objectif d'outiller des communautés pédagogiques.

Néanmoins, Elise Garrot reprend le modèle et l'outil d'une plateforme collaborative (Joomla!). Ces plateformes sont globales et monoblocs dans le sens où elles tentent de supporter tous les échanges entre membres (partage de fichiers, discussions générales, partage d'expériences...). Ces plateformes demandent une implication notable des membres qui doivent décrire leurs compétences de manière fournie par exemple. L'intérêt de la plateforme provient avant tout des échanges directs avec les autres membres.

Dans l'optique des sites du Web Participatif, Pépi n'essaye pas de répondre à tous les besoins d'échange mais **se concentre sur le partage de fichiers**. L'idée est que l'outil puisse s'intégrer dans la galaxie formée par les sites des enseignants et les listes de diffusions par exemple. Pépi fournit un service de dépôts de documents mais peut s'intégrer aux sites des enseignants. Il favorise juste la manipulation des documents, leur organisation et leur recherche. Sur son site, l'enseignant peut continuer à publier les travaux de ces élèves ou différentes nouvelles sur sa discipline. De plus, les enseignants peuvent continuer à employer la liste de diffusion, pratique par sa simplicité, pour discuter de points généraux et pointant éventuellement vers les ressources

de Pépi.

Pépi nous semble correspondre à la tendance du Elearning 2.0, définie précédemment (4.2.1) où l'internaute n'est plus dépendant d'un outil et de son modèle mais crée sa propre expérience en agrégeant les services proposés sur Internet de plus en plus inter-reliés.

5.5 Conclusions

Ces travaux sur le Web Participatif ont permis d'explicitier des caractéristiques et fonctionnalités soutenant la participation au partage. Nous avons classé celles-ci en trois catégories : ergonomique, sociale et pécuniaire.

D'autre part, l'observation de communautés d'enseignants a justement révélé le manque d'outils informatiques génériques pour soutenir leurs échanges. Nous y avons vu une opportunité pour proposer une solution « utile », implémentant les caractéristiques du Web Participatif.

Pourtant, nos efforts furent limités par des difficultés techniques et « politiques ». Nous avons pris conscience de la difficulté à réaliser un outil de partage ergonomiquement simple. Nos seules compétences furent insuffisantes pour atteindre une qualité pouvant attirer les enseignants. Malgré des efforts répétés, nous n'avons pas réussi à « accrocher » une communauté ou au moins à réunir un noyau d'enseignants impliqués.

Avec du recul, que retenons-nous de cette « failure story » ?

Depuis nos travaux autour de Pépi, le site Cartables.net a annoncé sa fermeture par lassitude des trois fondateurs, et la communauté ListePGM n'a toujours pas trouvé vraiment d'outils pour dépasser la liste de diffusion. De notre point de vue, ces communautés manquent toujours d'outils pour se structurer et perdurer. Une plateforme de partage de documents pédagogiques telle que nous la proposons nous semble toujours d'actualité pour les soutenir.

Si nous devons reprendre aujourd'hui nos travaux, nous commencerions par impliquer des utilisateurs, point crucial à nos yeux. Peut-être irions-nous dans les IUFM pour trouver de jeunes enseignants en recherche de soutiens pour leur métier et probablement plus ouverts aux nouveaux usages des outils informatiques. D'un point de vue technique, nous essaierions d'**insérer l'outil de partage dans des réseaux sociaux** tels que Facebook. En plus de profiter de mécanismes de partage déjà intégrés, l'outil pourrait exploiter facilement les liens sociaux, par exemple entre élèves de classes d'IUFM. L'outil explorerait alors un aspect plus viral du partage, il favoriserait pleins de petits groupes de partage puis les agrégerait progressivement, plutôt qu'un partage tout de suite national sur un site classique (comme nous l'imaginions avec Pépi).

Vers une application plus « expérimentable » et socio-sémantique Une contrainte importante que nous avons fixé à nos travaux est de pouvoir expérimenter nos outils en situation réelle. Lorsque nous nous sommes aperçus que nous n'avions pas réussi à attirer des utilisateurs sur Pépi et que son développement était trop lourd, nous avons décidé de débiter un nouveau développement dans un contexte différent afin d'obtenir des résultats plus exploitables.

D'un point de vue implémentation, le nouveau développement ne se préoccupait plus du dépôt des fichiers pour se concentrer sur la description des ressources sous la forme de signets. De plus, au lieu de développer un outil à partir de rien, nous avons repris une application pré-existante et diffusée sur le Web.

D'un point de vue modèle, nous nous sommes intéressés davantage à la description collaborative des contenus, de manière de plus en plus structurée. Nous avons pour cela notamment transformé la notion de dossiers-labels en celle de « tags structurés », introduisant des liens d'inclusion et de synonymie entre les tags. Enfin, à travers cette structuration progressive et collaborative des descripteurs d'un domaine, nous avons pu explorer la complémentarité et la fusion du Web Sémantique et du Web Participatif.

Troisième partie

Fusion du Web Sémantique et du Web Participatif, et indexation pédagogique

Les approches du Web Sémantique et du Web Participatif peuvent sembler s’opposer dans leurs usages. Pourtant, dans cette partie, nous verrons différentes manières d’allier technologie sémantique et aspect social du Web.

Parmi les solutions possibles, nous proposons le modèle d’une Indexation Progressive et Multi-points de vue des documents (modèle IPM), et les spécifications de fonctionnalités mettant en oeuvre ce modèle.

Ce modèle et ces spécifications furent implémentés dans une application de partage de signets, nommé SemanticScuttle.

SemanticScuttle fut développé et expérimenté en collaboration avec des documentalistes en sociologie. Dans la discussion, nous verrons comment les résultats de cette expérimentation peuvent être généralisés au domaine pédagogique. Enfin, la relative bonne diffusion de l’outil nous permettra de discuter les apports d’un outil de recherche ré-utilisable.

6

État de l'art

Sommaire

6.1 Fusion du Web Sémantique et du Web Participatif	119
6.1.1 Limites du Web Sémantique et du Web Participatif	119
6.1.2 Rapprocher traitement sémantique et participation des utilisateurs	121
6.1.3 Ontologies « d'objets sociaux »	122
6.1.4 Des tags aux concepts	125
6.2 Fusion du Web Sémantique et du Web Participatif appliquée au domaine pédagogique	133
6.2.1 Créer des cours décrits sémantiquement dans un Wiki : SweetWiki	133
6.2.2 Des tags Delicious aux champs LOM	135
6.2.3 Annoter collectivement des ressources pédagogiques avec des concepts de Wordnet	135
6.3 Conclusion	136

6.1 Fusion du Web Sémantique et du Web Participatif

Dans un premier temps, nous détaillerons les limites liées aux approches distinctes du Web Sémantique et du Web Participatif. Puis nous présenterons les solutions explorant leur utilisation commune : solutions automatiques ou manuelles, fondées sur des ontologies pré-existantes ou émergentes.

6.1.1 Limites du Web Sémantique et du Web Participatif

Dans leurs usages majoritaires, les visions du Web Sémantique et du Web Participatif semblent s'opposer. Le Web Sémantique, dans la vision du W3C, se fonde sur des ontologies. Celles-ci sont des représentations formelles et structurées d'un domaine, généralement réalisées par quelques experts ce celui-ci. A l'opposé, le Web Participatif exploite la participation des utilisateurs pour indexer selon des termes spontanés, les tags.

Limites des ontologies (du W3C) Bien que le Web Sémantique soit actuellement un courant technologique majeur, par exemple en considérant le nombre de conférences scientifiques sur le sujet, des critiques sont formulées contre la vision du W3C.

Par exemple, Clay Shirky souligne les limites des ontologies d'experts et des raisonnements formels. (En France, Manuel Zacklad s'oppose aussi à la vision du Web Sémantique du W3C, nous reviendrons sur ce point lors de la discussion en §7.5.4)

Dans [Shirky, 2005], Shirky déclare que des catégories prédéfinies pour organiser des entités fonctionnent très bien si le domaine est :

- de taille réduite ;
- avec des catégories bien définies ;
- avec des entités stables ;
- avec un nombre restreint d'entités ;
- avec des limites claires.

De plus, le système d'organisation fonctionnera d'autant mieux que les participants :

- sont des catalogueurs experts ;
- respectent une autorité de jugement ;
- sont des utilisateurs avertis ;
- sont des utilisateurs coordonnés.

Ce système fonctionne dans des bibliothèques avec des bibliothécaires organisant les livres pour les abonnés ou chez les psychiatres états-uniens dont l'Association Américaine des Psychiatres (APA) fait autorité en fournissant un manuel définissant chaque maladie selon les symptômes.

En revanche, les systèmes fondés sur des représentations précises et prédéfinies d'un domaine fonctionnent d'autant moins bien que :

- le corpus est large ;
- il n'y a pas de catégories précises, formelles ;
- les entités sont instables ;
- de nouvelles entités apparaissent ;
- il n'y a pas de limites claires.

De même, ces systèmes d'organisation fonctionnent moins bien si les participants :

- sont des catalogueurs amateurs ;
- ne possèdent pas d'autorité à laquelle se référer ;
- sont des utilisateurs amateurs ;
- sont des utilisateurs non coordonnés.

Or toutes ces caractéristiques sont similaires à celles du Web actuel, hétérogène et décentralisé. Comme les experts ne peuvent ni lire dans les pensées des utilisateurs pour deviner leur manière d'exprimer une requête, ni prédire le futur avec certitude pour définir des modèles stables dans le temps, Shirky considère que les ontologies d'experts ont peu de chance de fonctionner dans le contexte du Web.

De plus, Clay Shirky pointe une seconde difficulté que rencontre le Web Sémantique quant aux mécanismes de déduction ([Shirky, 2003]). Les déductions se fondent sur des règles formelles qui sont rares dans le monde réel. Exemple, « Les gens qui vivent en France parlent français » est une règle simple que tout le monde pourrait accepter. Mais elle n'est pas universelle car il y a en France des expatriés et des immigrants qui ne parlent pas français. Les déductions sont des

processus dépendants du mécanisme de généralisation lié au contexte. Les déductions sont donc inutiles si les systèmes ne partagent pas d'abord un même contexte et une vue commune du monde. Or nous avons vu précédemment que partager une vue commune, comme des ontologies, est une tâche difficile à l'échelle du Web. Pour Clay Shirky, ceci rend improbable l'application automatisée de règles à l'échelle du Web.

En conclusion, ces critiques rappellent la nécessité d'un contexte réduit et bien défini pour profiter des ontologies d'experts. Pour le domaine ouvert du Web, une autre approche des ontologies semble nécessaire.

Limites des tags. Contrairement aux représentations prédéfinies d'un domaine, comme les ontologies, les tags fonctionnent d'autant mieux qu'il y a beaucoup de ressources et d'utilisateurs. Néanmoins, nous distinguons deux limites principales aux tags :

- Lorsque peu d'utilisateurs sont impliqués dans l'indexation des données, les tags sont peu efficaces. Les tags sont donc difficilement applicables dans les petites communautés ou les communautés qui émergent.
- Les tags ne permettent pas de structuration élaborée ([Guy and Tonkin, 2006] [Sen et al., 2006] [Passant, 2007]). Ils ne permettent qu'une structure à plat, laissant peu de moyens à l'utilisateur pour organiser ses tags les uns par rapport aux autres.
- L'effet des tags est diminué par des problèmes de typographie ou de synonymie produisant du « bruit » et du « silence » dans les résultats. En cherchant le tag « apple », un système classique fondé sur des tags renverrait aussi bien les ressources portant sur les pommes que sur l'entreprise du même nom. En cherchant « devoirs » au pluriel, ce système ne renverrait pas les ressources portant sur « devoir » au singulier.

En conclusion, la flexibilité des tags semble avoir forcément pour contrepartie une imprécision diminuant la qualité des résultats.

Récapitulatif des différences d'usage. Nous parlons d'usage pour distinguer le Web Sémantique tel qu'il est défini dans les documents du W3C, de sa mise en pratique. Ici, nous nous intéressons à l'usage majoritaire i.e. aux nombreux projets fondés sur des ontologies d'experts, comme le projet LUISA.

De notre point de vue, voici une comparaison des usages du Web Sémantique et du Web Participatif :

	Web Sémantique	Web Participatif
Origine	Technologie prescrite par W3C	Tendance émergente
Description	Ontologies par des experts	Tags par les utilisateurs
Indexeurs	Experts maîtrisant les ontologies	Utilisateurs aux profils variés
Échelle	Mieux à petites échelles	Mieux à grandes échelles
Avantages	Raisonnements automatisés	Indexation peu coûteuse et variée
Désavantages	Coûts élevés d'indexation	Exploitation difficile par les machines

6.1.2 Rapprocher traitement sémantique et participation des utilisateurs

Le rapprochement entre Web Sémantique et Web Participatif possède deux aspects principaux :

- **Ontologies « d'objets sociaux »** : Le premier aspect consiste à définir les ontologies dans lesquelles des entités seront représentées. Ces entités peuvent être directement sociales (relations interindividuelles) ou issues de constructions sociales (tags, vocabulaires. . .). Les raisonnements et les échanges sémantiques seront ensuite basés sur ces représentations.
- **Des tags aux concepts** : Le second aspect consiste à exploiter des phénomènes participatifs pour exprimer les données selon les représentations formelles définies précédemment. Ce processus de structuration permet de transformer des métadonnées hétérogènes et non structurées (généralement des tags) en des métadonnées plus cohérentes et structurées (les ontologies).

Plusieurs solutions existent pour rapprocher Web Sémantique et Web Participatif. Pour comprendre leurs différences, la prochaine section survolera quelques unes de ces ontologies d'objets sociaux. La section suivante détaillera les moyens de structurer les données hétérogènes selon les concepts de ces ontologies.

6.1.3 Ontologies « d'objets sociaux »

Voici quelques ontologies parmi les plus connues pour représenter des phénomènes sociaux et participatifs sur Internet.

6.1.3.1 Ontologie de relations sociales : FOAF

L'ontologie « sociale » la plus connue est probablement l'ontologie FOAF⁷⁶, Friend Of A Friend, permettant de représenter les liens sociaux entre les individus et par prolongement des réseaux sociaux.

Dans ce contexte, « social » s'entend dans le sens d'interactions entre plusieurs individus.

Cette ontologie permet à un utilisateur de créer un profil et d'y indiquer que tel autre utilisateur est un ami, un membre de sa famille, un collègue. . . Même si cette ontologie ne permet pas d'exprimer toute la subtilité des relations humaines (comment différencier les différents niveaux ou types d'amitiés ?), sa simplicité et son ancienneté (2000) en ont fait aujourd'hui une référence.

Un exemple de profil FOAF décrivant Anne et Bernard, sachant qu'Anne considère Bernard comme un ami et que Bernard possède un site Web à son nom :

```
<foaf:Person rdf:ID="anne" >
  <foaf:name>Anne Martin</foaf:name>
  <foaf:gender>Female</foaf:gender>
  <rel:friendOf rdf:resource="#bernard" />
</foaf:Person>

<foaf:Person rdf:ID="bernard">
  <foaf:name>Bernard Dupont</foaf:name>
  <foaf:gender>Male</foaf:gender>
  <foaf:homepage rdf:resource="http://www.bernarddupont.com"/>
</foaf:Person>
```

⁷⁶<http://www.foaf-project.org>

Plusieurs applications du projet FOAF facilitent aujourd'hui la création de profil FOAF et permettent des recherches parmi les relations entre profils. A terme, il devrait être de plus en plus simple de formuler des requêtes comme « quels sont les collègues de X, amis de Y ? ».

6.1.3.2 Ontologies de tags

Tag ontology. Pour représenter sémantiquement les tags, il existe des représentations comme la « Tag Ontology ». Cette ontologie ne représente pas le contenu des tags, qui peuvent être n'importe quel terme, mais leur fonctionnement. Elle définit principalement, qu'un « tag » est apposé par un « utilisateur » à un « moment donné » pour décrire une « ressource ».

L'originalité de cette ontologie est qu'à l'image du phénomène spontané des tags, elle ne semble pas dirigée par une institution comme le W3C mais produite par les blogueurs. La référence la plus officielle se trouve donc à l'adresse d'un site personnel⁷⁷.

Ontologie de tags selon Gruber. Thomas Gruber propose dans [Gruber, 2007] les principes d'une autre ontologie de tags. Par rapport à la Tag Ontology, il mentionne notamment la nécessité d'enregistrer l'application (« source ») où le tag a été apposé, comme élément de contexte. Thomas Gruber propose aussi d'ajouter des tags négatifs c'est à dire pour un utilisateur d'affirmer qu'un tag n'est PAS associé à une ressource. Ce procédé serait utile pour éviter le phénomène de « mauvais » tags ajoutés par des spammeurs.

6.1.3.3 Ontologie de termes : SKOS

Parmi les ontologies les plus récentes se trouve SKOS visant à représenter sémantiquement des « concepts » et des « vocabulaires » de manière minimale. Nous considérons les vocabulaires en tant qu'objets sociaux, dans le sens d'objets fondamentaux issus des interactions sociales.

SKOS⁷⁸ propose une représentation formelle des concepts de thésaurus. Un concept peut avoir plusieurs noms (« labels ») et être relié à d'autres concepts.

Nous retiendrons de cette ontologie les trois relations principales entre concepts qu'elle offre :

Dans les systèmes d'organisation de la connaissance (Knowledge Organisation System), les relations sémantiques jouent un rôle crucial pour définir les concepts. La signification d'un concept n'est pas seulement défini par les termes en langage naturel de ses labels mais aussi par les liens vers les autres concepts du vocabulaire. A l'image des catégories fondamentales de relations qui sont employées dans les vocabulaires de thésaurus, SKOS fournit trois propriétés standards :

- « **skos : broader** » et « **skos : narrower** » permettent la représentation de liens hiérarchiques, comme la relation entre un genre et ses espèces plus spécifiques ou dépendants des interprétations, la relation entre un tout et ses parties ;
- « **skos : related** » permet la représentation de lien associatif (non-hiérarchique), comme la relation entre un type d'évènement et la catégorie des entités qui y

⁷⁷Tag ontology : <http://www.holygoat.co.uk/projects/tags/>

⁷⁸SKOS, Simple Knowledge Ontology System : <http://www.w3.org/2004/02/skos/>

participent habituellement. Une autre utilisation de « skos : related » se situe entre deux catégories où aucune n'est pas générale ou spécifique.⁷⁹

En mars 2009, la spécification SKOS était soumise à commentaires en tant que candidate pour devenir une recommandation du W3C.

6.1.3.4 Ontologie du Web Participatif : SIOC

L'ontologie SIOC (« Semantically-Interlinked Online Communities ») relie des individus, des productions sur le Web et des concepts pour décrire ces productions. En traitant ces trois aspects en parallèle, nous considérons cette ontologie comme la plus proche de représenter le Web Participatif.

Les « individus » appartiennent à des « communautés » qui peuvent être liées à plusieurs « sites Web ». Par exemple, nous présentions précédemment une communauté en ligne de professeurs en génie mécanique (§5.1.2). Cette communauté se retrouve autour d'une liste de diffusion, d'un site participatif et sur les sites « personnels » des enseignants. SIOC permet d'employer FOAF (§6.1.3.1) pour représenter les liens entre individus et organismes.

Pour SIOC, les « sites Web » sont considérés comme des « conteneurs » pouvant contenir des productions d'individus ou d'autres sous-conteneurs. Par exemple, un site Web peut contenir un « conteneur » permettant de déposer et partager des signets, et un forum considéré comme un « conteneur » de message de discussions.

Chaque production peut alors être décrite à l'aide de tags ou de concepts. Pour cela, SIOC permet d'employer les représentations de la Tag Ontology (§6.1.3.2) ou de SKOS (§6.1.3.3).

Le but de l'initiative SIOC est de permettre l'intégration entre les informations des communautés en ligne. L'ontologie fut employée dans différents outils participatifs permettant d'exporter les contenus au format SIOC grâce à des plugins⁸⁰. En devenant un moyen standard pour exprimer les contenus générés par les utilisateurs sur les sites participatifs, SIOC espère offrir de nouveaux usages autour des données des communautés en ligne et permettre le développement d'applications sémantiques exploitant les données du Web Participatif.

6.1.3.5 Conclusion

Les travaux ci-dessus répondent au besoin de formaliser les « objets » issus de l'activité des internautes, comme les relations sociales ou les vocabulaires employés.

Le point commun de ces ontologies est leur légèreté nécessaire pour représenter la diversité des activités des internautes. Par la suite, les membres des projets comme FOAF ou SIOC font l'hypothèse que ces représentations seront suffisantes pour échanger des concepts entre applications et effectuer des raisonnements automatisés utiles aux internautes.

Pourtant la jeunesse de ses projets ne nous permet pas de savoir si de réels échanges sémantiques seront possibles par ses ontologies et pour l'instant, nous n'avons pas rencontré de projets

⁷⁹Traduction de <http://www.w3.org/TR/2008/WD-skos-primer-20080829/#secsimple>

⁸⁰SIOC PHP Exporter API : <http://wiki.sioc-project.org/index.php/PHPEXportAPI>

appliquant des raisonnements évolués sur des données sociales en situations réelles.

Néanmoins, ces ontologies nous semblent permettre une amélioration des échanges de données simples entre applications, comme des termes ou des données basiques sur les utilisateurs. Ces échanges peuvent alors améliorer l'expérience des utilisateurs sur le Web en facilitant la gestion de données entre plusieurs sites. L'API OpenSocial⁸¹ illustre cette tendance en permettant le partage de profils d'utilisateurs entre applications. Ceci permet à un utilisateur de retrouver ses informations personnelles quel que soit le site de réseau social sur lequel il navigue.

6.1.4 Des tags aux concepts

Une fois les ontologies définies, comment exprimer les données hétérogènes du Web selon celles-ci ?

Certains objets sociaux comme les signets partagés ou les messages d'un forum sont bien définis pour être représentés formellement de manière automatique. Mais d'autres objets sociaux co-construits, dont les limites sont moins claires, sont plus difficiles à représenter sémantiquement. C'est le cas des concepts partagés entre plusieurs utilisateurs, auxquels nous nous intéressons.

Un « concept partagé » apparaît sous la forme d'un tag sur les sites participatifs et d'une instance, dans une ontologie comme SKOS, pour le Web Sémantique. Pour améliorer le partage de contenus à l'aide de concepts partagés, la question est de savoir comment passer des tags aux instances et concepts plus structurés du Web Sémantique.

Nous classons les différentes approches pour passer des tags spontanés et individuels aux concepts structurés et communs en quatre grandes catégories. Ces catégories dépendent des réponses à deux questions :

- le passage des tags aux concepts est-il effectué automatiquement ou manuellement par des utilisateurs ?
- les concepts pré-existent-ils indépendamment des tags ou en émergent-ils ?

Les réponses permettent de classer les approches les plus courantes dans le tableau suivant.

	Concepts indépendants des tags	Concepts issus des tags
Liens créés automatiquement	1	3
Liens créés par les utilisateurs	2	4

L'intégration du Web Sémantique et du Web Participatif ne peut se faire qu'en relâchant des spécificités d'au moins une des deux approches. Certains travaux préfèrent conserver le formalisme du Web Sémantique et explorent davantage les zones 1 et 2. D'autres travaux préfèrent conserver la flexibilité du Web Participatif et explorent les zones 3 et 4.

Pour notre part, nous partirons dans nos travaux d'un outil participatif de partage de signets. Nous explorerons le soutien aux utilisateurs pour les aider à enrichir et structurer progressivement leurs descriptions (zone 4).

⁸¹OpenSocial : <http://sites.google.com/a/opensocial.org/opensocial/>

6.1.4.1 Sémantisation automatique et concepts prédéfinis

Ces approches partent de schémas décrivant un domaine et tentent d'y relier les tags avec un traitement automatique.

Les schémas les plus courants sont des ontologies comme Wordnet ou DBpédia⁸², une ontologie reprenant les informations de Wikipédia.

Par exemple, [Gracia et al., 2006] propose de désambiguïser des mots-clefs en faisant référence à différentes ontologies. Les ontologies sont trouvées à l'aide d'un moteur de recherche d'ontologies⁸³. En se basant sur plusieurs ontologies, Gracia et al. utilisent un important ensemble de concepts possibles. La désambiguïstation est effectuée à l'aide de requêtes Google qui permettent de sélectionner le sens le plus plausible parmi plusieurs possibles. Pour cela, Gracia et al. utilisent la proximité sémantique entre deux termes, qu'ils définissent grâce au nombre de résultats fournis par Google et contenant les deux termes dans une même page.

Sofia Angeletou et al. ([Angeletou et al., 2007]) explicitent des relations entre tags. Les tags sont issus des sites participatifs Delicious et Flickr. Ces travaux ont notamment l'intérêt de distinguer différentes sortes de tags qui peuvent difficilement être retrouvées de manière simple dans les ontologies. C'est le cas par exemple des tags définissant une terminologie nouvelle ; les tags étant ajoutés continuellement, ils offrent une vue sur le vocabulaire courant employé par les utilisateurs. Ces termes récents mettent beaucoup plus de temps à apparaître dans les ontologies d'experts. C'est aussi le cas des tags issus d'un jargon ; le site Flickr permet le partage de photos. Il y apparaît de nombreux termes propres aux photographes (« canon », « d50 », « macro »). Ces vocabulaires spécifiques sont rarement abordés par les ontologies du Web Sémantique.

Néanmoins, en dehors de ces tags particuliers, les tags décrivant des noms communs peuvent généralement être retrouvés dans des ontologies. Sofia Angeletou et al. peuvent ainsi retrouver différents liens existant entre tags en employant les informations des ontologies.

Enfin, Maurizio Tesconi et al. proposent dans [Tesconi et al., 2008] un algorithme pour lier des tags de Delicious aux pages correspondantes de Wikipédia. L'association s'effectue en supposant que « plus un tag T est lié à une page Wikipédia, plus cette page contient de termes équivalents à des tags co-occurent à T dans Delicious. » Selon les auteurs, la méthode appliquée à Delicious permet de désambiguïser 90% des 3000 tags de 9 utilisateurs aux profils variés.

6.1.4.2 Sémantisation automatique et concepts émergents

Le principe de cette approche est d'explicitier le sens des tags et notamment leurs relations, en analysant de manière automatique leur organisation, celles des documents et celles des utilisateurs.

Mika propose ainsi dans [Mika, 2007] d'analyser les graphes sociaux, de tags et de documents. Mika propose le nom de « Web 3.0 » pour décrire ces approches mêlant technologies du « Web

⁸²DBpedia, « A community effort to extract structured information from Wikipedia and to make this information available on the Web » : dbpedia.org

⁸³Swoogle, moteur de recherche d'ontologies : <http://swoogle.umbc.edu/>

Sémantique » et tendance du « Web 2.0 ».

[Au et al., 2007] illustre cette approche tri-partite (tags, utilisateurs, documents) en désambiguïsant le tag « sf » dans Delicious. En effet, ce tag peut signifier aussi bien « Science Fiction » que « San Francisco » pour les utilisateurs. Pour cela, le système détermine les graphes des documents et des utilisateurs. Il y a un lien entre deux documents si un utilisateur les a taggués avec le même tag, et il existe un lien entre deux utilisateurs s'ils ont taggé un document avec le même tag. La Fig. 6.1 montre le graphe créé par 20000 utilisateurs ayant employé le tag « sf ». Les deux pôles de ce graphe montrent à quel point l'analyse des réseaux d'utilisateurs peut aider à déterminer puis éventuellement enrichir la sémantique des tags, en les situant dans des ensembles d'utilisateurs.

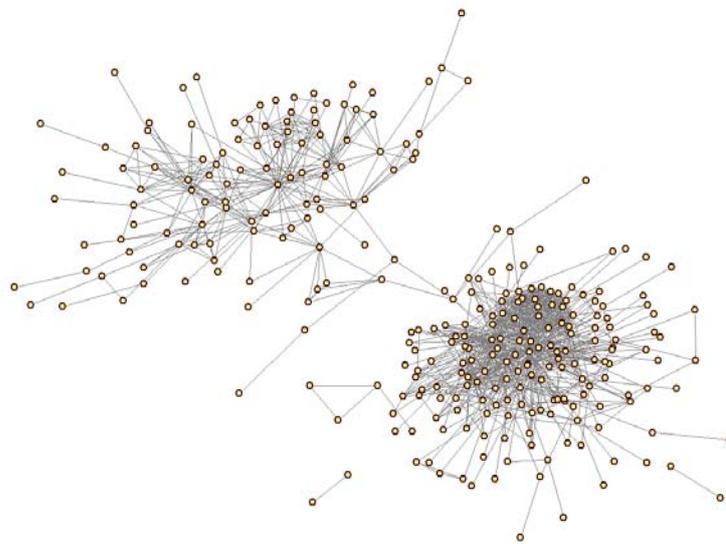


FIG. 6.1 – Graphe des utilisateurs de Delicious ayant employé le tag « sf ». Plus des utilisateurs ont de signets en commun et plus ils sont proches et reliés. La désambiguïsation de « sf » en « Science Fiction » ou « San Francisco » peut ainsi être soutenue par l'analyse des relations entre utilisateurs.

L'analyse automatique des tags se retrouve dans les GrowBags du projet FacetedDBLP [Diederich et al., 2007]. Ce projet explore les 500 000 mots clés associés aux 93 000 articles scientifiques enregistrés dans la base DBLP⁸⁴. Par analyse statistique, le système propose alors des taxinomies automatiques sur les mots-clés, navigables ensuite dans un navigateur (Fig. 6.2). Par exemple, si la majorité des occurrences du tag A se fait en co-occurrence avec le tag B, mais que le tag B est aussi souvent employé sans le tag A, alors un système peut faire l'hypothèse que le tag B subsume le tag A. Les relations de subsumptions sont représentées par des flèches entre les concepts ; l'épaisseur du trait qualifie l'assurance du système à proposer cette relation de subsumption avec des pointillés pour une confiance faible et avec un trait épais pour une confiance importante.

Cette approche émergente automatisée apparaît aussi dans [Simpson and Butler, 2008] qui parlent de « structures latentes » ou dans [Begelman et al., 2006] qui détaillent comment regrouper les tags de Delicious par des méthodes de clustering.

⁸⁴DBLP, Computer Science Bibliography : www.informatik.uni-trier.de/~ley/db/

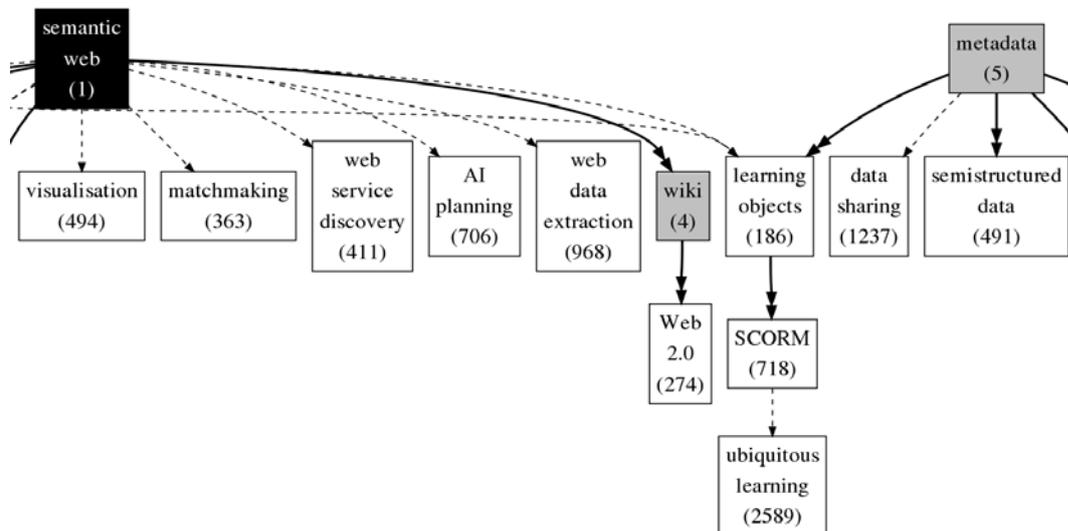


FIG. 6.2 – Hiérarchie de termes issues de l'analyse automatique des mots-clefs d'articles scientifiques dans DBLP ([Diederich et al., 2007])

6.1.4.3 Sémantisation manuelle et concepts prédéfinis

Cette approche fait reposer la « conceptualisation » sur les utilisateurs.

Création des liens sociaux de FOAF. De nombreux outils se sont développés pour aider les internautes à créer des profils FOAF. Par exemple, l'utilisateur remplit un formulaire en ligne en indiquant les pages Web des autres utilisateurs qu'il connaît et le système génère le profil FOAF au format RDF⁸⁵.

Relier manuellement des tags aux concepts. Pour les tags, des auteurs ont proposé des applications pour aider les utilisateurs à relier leurs tags à des concepts d'ontologie.

Une des applications les plus abouties est Faviki⁸⁶ permettant de tagguer des ressources mais seulement en employant le titre d'articles issus de Wikipédia.

[Marchetti et al., 2007] et le système MOAT d'Alexandre Passant [Passant and Laublet, 2008] explorent un système similaire mais en permettant de relier les tags à des concepts d'ontologies. La partie milieu-gauche de Fig. 6.3 montre les concepts possibles, définis par leur URL, pour le tag « Paris ».

Limites. Les tags sont fondés sur la simplicité et la rapidité de mise en oeuvre. Comme [Tesconi et al., 2008], nous pensons que ces approches manuelles cherchant à faire relier par les utilisateurs les tags à des concepts prédéfinis semblent limitées par le temps et l'énergie supplémentaires que doit fournir l'utilisateur pour choisir le bon concept, plutôt que de taper le premier terme lui venant à l'esprit.

⁸⁵Générateur de profil FOAF : <http://www.ldodds.com/foaf/foaf-a-matic>

⁸⁶Faviki : <http://www.faviki.com/pages/welcome/>

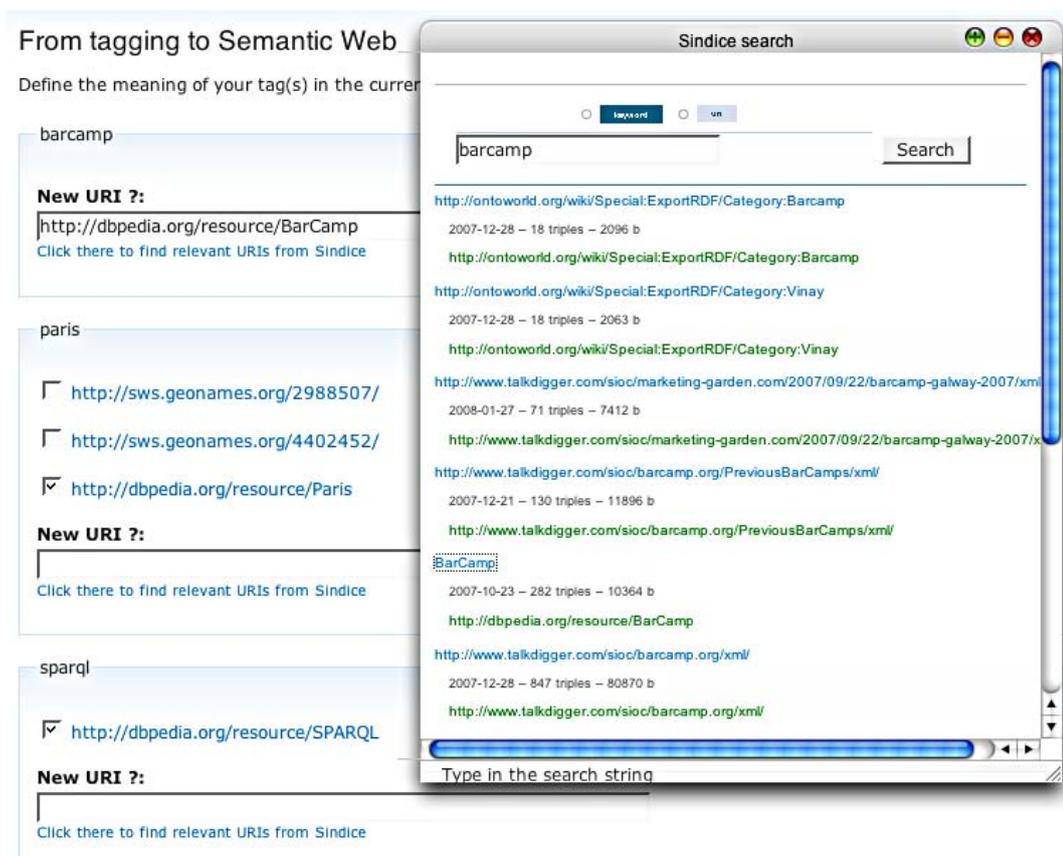


FIG. 6.3 – Relier les tags aux concepts dans MOAT, fenêtre de gauche.

6.1.4.4 Sémantisation manuelle et concepts émergents

Cette approche consiste à soutenir les utilisateurs dans la structuration de tags en concepts. Contrairement à l'approche précédente, les utilisateurs ne sont pas contraints d'employer une ontologie définie à l'avance.

Pour donner une vue plus large, nous traiterons dans cette section la construction manuelle d'ontologies en général (pas forcément issues des tags).

Techniques soutenant la structuration. Les tags ont la particularité d'être spontanés et donc de contenir des incohérences rendant difficile leur exploitation automatisée. Pour réduire ces incohérences, des outils en ligne comme Delicious Tag Cleaner⁸⁷ proposent par exemple des unifications entre tags. Par exemple, ce système repère qu'un utilisateur utilise le tag « article » et « articles » et propose de transformer automatiquement le tag au pluriel en singulier. Ces procédés ne produisent pas de concepts mais permettent de diminuer l'hétérogénéité des tags.

Lors de l'ajout d'un contenu, les outils participatifs recommandent aussi en général des tags basés sur les tags les plus populaires pour l'utilisateur ou pour la communauté en général. Des travaux comme [Sen et al., 2006] montrent l'efficacité des recommandations pour inciter un groupe d'utilisateurs à employer le même vocabulaire.

⁸⁷Delicious Tag Cleaner : <http://delicious.isnotworking.com>

Wikis sémantiques. Des wikis sémantiques permettent de créer des relations entre pages à la volée. C'est à dire que de la même manière que dans un wiki l'utilisateur modifie une page et ajoute du contenu librement, certains wikis sémantiques comme Semantic Mediawiki [Krötzsch et al., 2006] permettent de créer librement des relations sémantiques entre pages. [Buffa and Gandon, 2006] liste longuement les différents types de wikis sémantiques.

Alexandre Passant ([Passant, 2009]) propose avec SemSlate un processus complet d'indexation sémantique. Ce processus permet aux employés d'une entreprise de créer et maintenir une ontologie à l'aide d'un wiki sémantique (UfoWiki), puis d'indexer des contenus comme ceux de blogs à l'aide de tags, pointant vers les concepts du wiki sémantique.

Dans les wikis sémantiques, tous les utilisateurs travaillent sur les mêmes concepts associés généralement à une page du wiki. Ils travaillent donc sur la même représentation du domaine.

Nous n'avons pas réussi à trouver de références détaillant les structures sémantiques issues de ces outils ou sur les formes de collaboration qui s'y pratiquent.

De textes sources à l'édition fine de l'ontologie. L'outil ECCO ([Giboin et al., 2008]) offre un processus complet permettant à des « représentants des utilisateurs » de travailler avec des « ingénieurs de la connaissance » pour créer des ontologies.

Le cycle de conception comprend : 1) acquisition de textes sources ; 2) extraction de termes candidats et de leurs contextes dans les textes sources ; 3) élaboration collaborative d'un vocabulaire à partir des termes candidats, en distinguant notamment les concepts des relations ; 4) mise en place de la hiérarchie des termes du vocabulaire ; 5) édition fine de l'ontologie ; 6) ajout/édition de règles ; 7) tests de l'ontologie à l'aide de requêtes SPARQL (§2.1.1).

Les concepteurs de l'outil s'interrogent sur l'intérêt de donner de plus en plus de moyens aux utilisateurs (dans une optique Web Participative) et sur les difficultés que ceci peut engendrer. Leurs travaux en cours ne fournissent pas encore de retours d'usage.

Méthodologie itérative et recherche de consensus. [Karapiperis and Apostolou, 2006] propose une méthodologie de construction collaborative et itérative d'ontologies. Cette méthodologie emploie des groupes de 5 à 10 participants, aux profils variés. Les étapes consistent principalement pour le coordinateur à :

1. Expliquer le but de l'ontologie aux participants pour les motiver à participer ;
2. Discuter avec les participants pour déterminer les limites du domaine à définir et les requêtes auxquelles l'ontologie doit permettre de répondre ;
3. Construire une ontologie initiale à partir des informations recueillies ;
4. Soumettre l'ontologie initiale à une évaluation des participants. Une fiche (Fig. 6.4) guide l'évaluation en permettant de noter l'utilité de concepts, ou encore de proposer de nouveaux synonymes ou relations.
5. Comptabiliser les notations pour modifier l'ontologie et soumettre à un vote les ajouts. Reprendre les deux étapes précédentes jusqu'à un consensus complet entre les membres ;
6. Tester l'ontologie, notamment à partir des requêtes définies à l'étape 2.

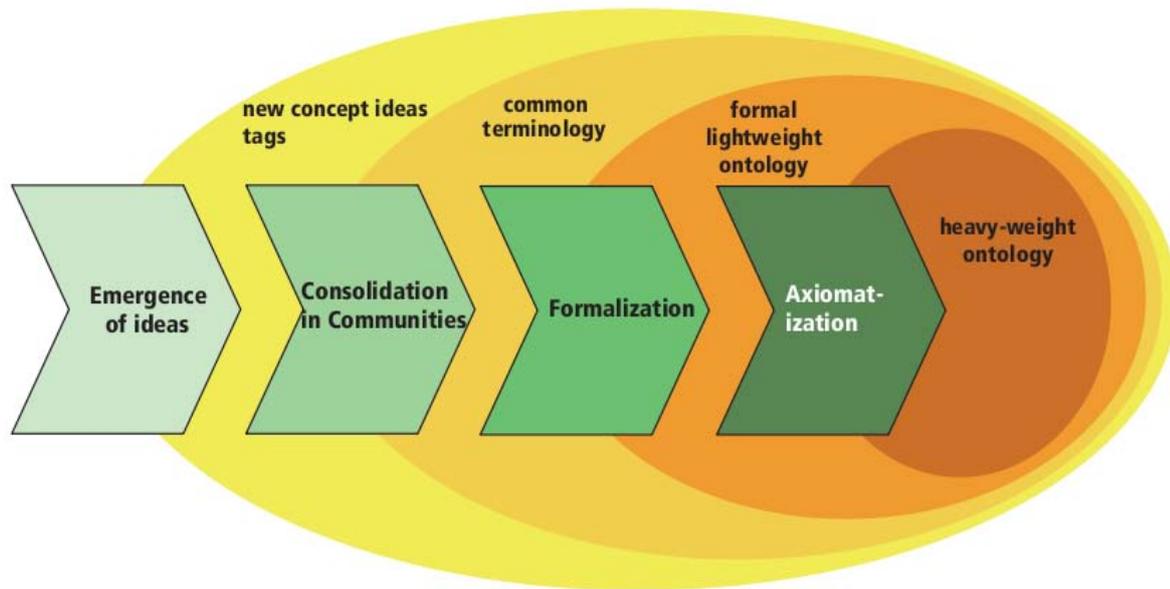


FIG. 6.5 – Processus de mûrissement des ontologies. Les tags sont progressivement consolidés, structurés puis formalisés ([Braun et al., 2007]).

Dans cet outil, les utilisateurs travaillent ensemble sur une même ontologie.

Web Socio-Sémantique (W2S). Le Web Socio-sémantique est une approche proposée par Manuel Zacklad et le laboratoire Tech-CICO. Dans [Zacklad, 2005], Manuel Zacklad explique que le Web Sémantique du W3C se fonde sur des ontologies limitant le sens afin de pouvoir y appliquer des raisonnements automatisés. Cette vision du Web Sémantique tend à un affaiblissement du sens. Il propose alors des « ontologies sémiotiques » où l'interprétation joue un rôle important (ouverture sémiotique).

Dans ce courant, [Cahier, 2005] propose d'améliorer le standard des TopicsMap pour définir des ontologies sémiotiques. Cet auteur propose alors des cartes HyperTopics. L'intérêt de cette approche consiste notamment en la prise en compte du contexte social des ontologies, défini au sein de communautés d'utilisateurs, et de la prise en compte du point de vue des utilisateurs.

Pour ces aspects nous nous inscrivons dans la vision du Web Socio-Sémantique et non dans la vision parfois universaliste des travaux du W3C. Néanmoins, la mise en oeuvre technique du Web Socio-Sémantique nous semble possible avec les outils de base proposés par le W3C.

Conclusion : coordonner différents points de vue. Parmi les techniques d'élaboration collaborative d'ontologies, la majorité force les utilisateurs à travailler sur une même représentation. Pour nous, dans le cadre décentralisé et hétérogène du Web, il nous semble primordial de conserver les différents points de vue des utilisateurs, d'essayer de les coordonner mais sans qu'une vision commune ne soit obligatoire (notamment pour faciliter l'utilisation des technologies sémantiques).

La coordination de différents points de vue pour construire une vision commune reste donc un problème peu abordé. La vision du Web Socio-Sémantique de Manuel Zacklad est l'une des seules approches explorant ce problème. De ce point de vue, nous pensons que nos travaux s'inscrivent donc dans ce courant.

6.1.4.5 Conclusion

Les techniques pour passer des tags aux concepts peuvent être organisées selon deux axes : techniques automatiques ou manuelles, visant des concepts prédéfinis ou émergents.

Parmi ces approches, nous avons décidé d'explorer l'approche manuelle vers des concepts émergents car :

- les techniques automatiques nécessitent de grandes quantités de données (milliers de participants) que ne peuvent pas produire les communautés enseignantes de taille beaucoup plus réduite.
- les techniques reliant les tags à des concepts prédéfinis perdent selon nous une caractéristique primordiale des tags à savoir de suivre l'apparition de nouveaux concepts dans les communautés.

Dans les motivations, nous sommes proches des travaux sur le Web Socio-Sémantique de Manuel Zacklad visant à soutenir les interactions entre utilisateurs. Néanmoins, dans les techniques, il nous semble possible d'employer certaines technologies sémantiques recommandées par le W3C. Ces dernières offrent notamment l'avantage de s'appuyer sur les développements d'une large communauté.

6.2 Fusion du Web Sémantique et du Web Participatif appliquée au domaine pédagogique

Voici quelques travaux spécifiques au domaine pédagogique tentant de relier sémantique et participatif.

6.2.1 Créer des cours décrits sémantiquement dans un Wiki : SweetWiki

Quand ils prennent de l'ampleur, les wikis rencontrent des difficultés dans l'organisation des données (structures anarchiques, nombre trop important de pages, chemin de navigation non mis à jour...).

SweetWiki ([Buffa and Gandon, 2006]) explore la conception d'un wiki fondé sur un serveur du Web Sémantique, c'est à dire l'emploi des technologies du Web Sémantique pour soutenir le cycle de vie des wikis. Le modèle des wikis a été décrit dans un cadre sémantique : une ontologie contient les concepts tel que les mots clefs des wikis (WikiWord), les pages, les liens entre pages, les auteurs, les dates de modification, les versions... Cette ontologie est ensuite exploitée par un moteur de recherche sémantique (Corese) installé sur le serveur. Ce moteur permet alors de répondre à des requêtes SPARQL pour la création d'index dynamiques ou la recommandation de liens « Pages similaires ».

En plus, SweetWiki intègre un éditeur standard en WYSIWYG⁸⁸ qui fut étendu pour supporter l'ajout d'annotations sémantiques comme des tags. En éditant une page, l'utilisateur peut la décrire en entrant librement des mots-clés dans un champ, où un mécanisme d'auto-complétion, basé sur des requêtes SPARQL, propose des termes issus des concepts aux labels compatibles. Le moteur montre aussi le nombre de pages déjà décrites par ces concepts. Ainsi les concepts sont collectés et employés comme des tags.

Pour maintenir l'ensemble des concepts, SweetWiki permet l'édition en ligne des ontologies et des annotations par des administrateurs. Ces derniers vérifient les tags ajoutés par les utilisateurs simples et peuvent les réorganiser en ajoutant des relations comme « subClassOf » ou « SeeAlso ». Les annotations des utilisateurs simples ne sont pas modifiées mais la navigation et la recherche fondées sur des requêtes sémantiques sont améliorées.

SweetWiki fut expérimenté pour créer un cours de Java entre plusieurs enseignants. Chaque page du cours contient des tags ajoutés par les enseignants (Fig. 6.6) et décrivant le contenu de la page. La copie d'écran montre une page sur l'héritage en Java et sur les concepts associés « subclass », « superclass »... Les administrateurs du Wiki maintiennent une ontologie du domaine, i.e. de Java, à laquelle ils vont raccrocher les tags des utilisateurs. Ainsi le tag « jbutton » est relié au concept « [button] » et celui de « javadoc » au concept de « [documentation] ».

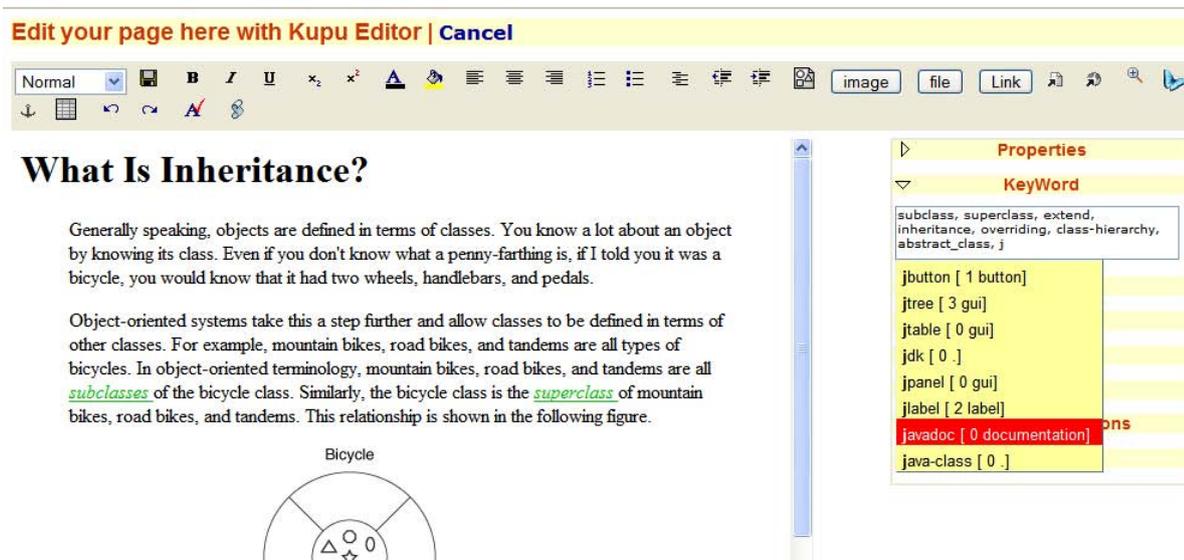


FIG. 6.6 – Ajout de tags lors de l'ajout d'une page dans SweetWiki (menu droit). Entre crochets apparaissent les concepts auxquels les administrateurs ont relié les tags des utilisateurs.

En reliant les tags aux concepts prédéfinis par des administrateurs, SweetWiki se situe dans la zone 2 – indexation manuelle et concepts prédéfinis – de notre organisation (voir 6.1.4). Les limites liées à cette approche résident dans la difficulté éventuelle à faire accepter aux enseignants-utilisateurs l'approche unique de l'ontologie sous-jacente, gérée par les administrateurs. Malheureusement nous avons trouvé peu d'analyses sur les usages de SweetWiki ; les analyses concernent la fréquentation globale de l'outil mais n'apprennent pas grand chose sur les contenus, tags et ontologies créés ou employés par les utilisateurs ([Ghali et al., 2009]).

⁸⁸WYSIWYG : éditeur permettant de modifier une page tout en voyant directement le résultat (« What You See Is What You Get »)

6.2.2 Des tags Delicious aux champs LOM

[Al-Khalifa and Davis, 2007] repèrent dans Delicious les signets portant sur les feuilles de style CSS. Le thème du design Web, auquel le CSS appartient, est très présent sur Delicious.

Puis, ils nettoient les tags pour ne conserver que les tags en anglais, sans variations orthographiques et sans ambiguïtés. Ils créent ensuite trois ontologies portant sur le domaine du design Web, sur les concepts des CSS et sur les types pédagogiques et techniques des ressources (e.g. éditeur CSS, exemple).

Enfin, ils relient les tags aux concepts et les concepts à certains champs de LOM ou propres au domaine : titre, description, mots-clefs, type pédagogique de la ressource, difficulté, niveau de recommandation des autres utilisateurs, concepts CSS abordés... Les tags sont reliés aux concepts des ontologies portant le même nom. Certaines règles sémantiques basiques sont aussi appliquées telle que « si une ressource est décrite par le tag 'font' (police de style) alors mettre le champ 'instructional level' à la valeur 'basique'. »

Ces travaux correspondent à l'approche 1, i.e. aux liens automatiques entre tags et concepts prédéfinis.

6.2.3 Annoter collectivement des ressources pédagogiques avec des concepts de Wordnet

Dans [Bateman et al., 2006], Scott Bateman et al. propose CommonFolks, une architecture permettant de décrire des ressources pédagogiques à l'aide des concepts de WordNet⁸⁹. Si un utilisateur ne trouve pas un concept il peut le rattacher à un autre concept à l'aide de la relation de subsomption.

Considérons une ressource pédagogique portant sur le XML. L'utilisateur veut dire que c'est un « tutoriel ». Ce concept n'existant pas dans WordNet, l'utilisateur doit l'ajouter. L'application lui propose de naviguer dans le graphe des concepts de WordNet pour trouver la meilleure position pour placer « tutoriel ». L'application peut proposer des recommandations de positionnement si d'autres utilisateurs ont déjà ajouté le concept de « tutoriel ». Par exemple, l'application va indiquer que le meilleur concept-parent est « leçon » car la majorité des autres utilisateurs l'ont choisi comme concept-parent. Ceci correspond à une sorte de consensus implicite parmi les utilisateurs. Si l'utilisateur est d'accord il choisira de situer là son concept de « tutoriel ».

Plus tard si l'utilisateur veut réemployer le concept de « tutoriel », il n'aura pas besoin de le ré-ajouter car le système lui proposera le concept qu'il a déjà défini. Tout nouvel ajout sera aussi vérifié, par le système, pour qu'il ne soit pas incohérent avec les concepts précédemment définis.

Ces travaux se situent entre les approches 2 et 4. C'est à dire que les utilisateurs relient les tags à des concepts pré-existants de WordNet. Néanmoins l'application permet aussi aux utilisateurs de placer leurs propres concepts en s'inspirant des vues des autres mais sans être contraints. Ces approches signifient aussi des limites dues à la charge cognitive pour placer un

⁸⁹Wordnet est une ontologie de concepts généraux

nouveau concept dans une ontologie, tout en conservant l'ensemble cohérent. Nous n'avons pas trouvé de retours d'usages sur l'emploi direct de cet outil.

6.3 Conclusion

Dans ce parcours des travaux existants, nous avons vu des méthodes automatiques et des méthodes manuelles pour relier les tags aux concepts, généralement prédéfinis.

Les méthodes automatiques sont d'autant plus efficaces qu'elles peuvent s'appuyer sur de grandes masses de données et un grand nombre d'utilisateurs. De plus, elles nécessitent généralement un nettoyage des tags, quitte à éliminer les tags ambigus ou n'ayant du sens que pour un sous-ensemble d'utilisateurs. Par conséquent, les concepts n'ont peut-être pas la richesse des points de vue offerts par les tags. Pour de petites communautés, de moins d'une centaine de participants, **il nous semble donc intéressant et utile d'explorer plutôt les méthodes manuelles.**

Le choix entre concepts prédéfinis ou concepts émergents renvoie au débat existant entre le Web Sémantique et le Web Socio-Sémantique. La difficulté des ontologies pré-définies par des experts nous semble être la compréhension par les utilisateurs. Ceci est d'autant plus vrai dans certains domaines, comme celui pédagogique, où les pratiques et les points de vues peuvent beaucoup varier d'un utilisateur à l'autre. Pour ces raisons, **il nous semble plus réaliste d'étudier l'émergence des concepts dans des phénomènes collaboratifs souples plutôt que leur pré-définition formelle par des experts.**

Néanmoins, les technologies du Web Sémantique version W3C ne nous semblent pas incompatibles avec ces approches souples de la sémantique. Par exemple, l'ontologie SIOC nous semble un cadre assez général pour représenter (ou au moins essayer de représenter) les phénomènes participatifs de descriptions et d'interactions autour de ressources.

Selon nous, la meilleure solution réside à l'intersection de tous ces aspects : experts définissant des représentations et utilisateurs exprimant leurs propres visions, visions hétérogènes et représentations structurées. . . Mais comment intégrer ces aspects dans un même outil d'indexation ?

Projet SemanticScuttle

Sommaire

7.1	Modèle et fonctionnalités	138
7.1.1	Objectif des travaux	138
7.1.2	Modèle d'Indexation Progressive et Multi-points de vue (IPM)	138
7.1.3	Fonctionnalités illustrant le modèle IPM	141
7.2	Développement	144
7.2.1	Développement basé sur Scuttle, un outil classique de partage de signets	144
7.2.2	Fonctionnalités de structuration progressive	149
7.2.3	Fonctionnalités de collaboration progressive	158
7.2.4	Autres fonctionnalités et caractéristiques	160
7.3	Terrain d'application	163
7.3.1	Sociopôle, des signets en sociologie	164
7.3.2	WIKINDEX, tentative d'utilisation d'un outil d'indexation bibliographique	166
7.4	Expérimentations	168
7.4.1	Méthodologie des expérimentations	168
7.4.2	Résultats : Facilité d'ajout d'un signet	172
7.4.3	Résultats : Continuum de fonctionnalités structurantes	173
7.4.4	Résultats : Continuum de fonctionnalités collaboratives	177
7.4.5	Résultats : Mieux « profiter » des ressources	178
7.4.6	Bilan des expérimentations	180
7.5	Discussions et travaux proches	181
7.5.1	Rapprochement avec le domaine pédagogique	182
7.5.2	Spectre de l'intérêt des utilisateurs	184
7.5.3	Perspectives de recherche	186
7.5.4	Le Web Sémantique du W3C peut-il être social ?	187
7.5.5	Améliorer l'indexation automatique à l'aide de contributions des utilisateurs	189
7.5.6	Chercher et expérimenter dans le domaine des EIAH	190

Le projet SemanticScuttle consiste en un modèle d'indexation prenant en compte différentes formes de métadonnées, d'un outil nommé SemanticScuttle et de son expérimentation, principalement avec des documentalistes en sociologie.

7.1 Modèle et fonctionnalités

Pour allier technologie sémantique et tendance participative, nous proposons le modèle d'Indexation Progressive et Multi-points de vue (modèle IPM). De plus, nous suggérons un ensemble de fonctionnalités le mettant en oeuvre.

7.1.1 Objectif des travaux

Point de vue des utilisateurs. Nos travaux visent à outiller des communautés en général et des communautés d'enseignants en particulier pour faciliter le partage de ressources entre les membres, au moyen de la construction d'une vue commune du domaine.

Le projet Pépi a révélé les difficultés techniques à réaliser un site de partage de fichiers pouvant inciter les enseignants à participer. L'implémentation du projet SemanticScuttle doit permettre une meilleure diffusion de l'outil et d'obtenir plus d'utilisateurs pour le tester.

Point de vue de la Recherche. Nos travaux visent à définir un modèle permettant d'exploiter la participation des utilisateurs pour finalement produire des représentations plus structurées, qui faciliteront le partage de ressources.

7.1.2 Modèle d'Indexation Progressive et Multi-points de vue (IPM)

Dans cette section, nous allons présenter un modèle que nous appelons « modèle d'Indexation Progressive et Multi-points de vue » ou modèle IPM.

Le modèle d'Indexation Progressive et Multi-points de vue (modèle IPM) vise à faciliter le partage de ressources entre des utilisateurs aux motivations et aux profils variés. Ce modèle propose les principes d'une structuration progressive des métadonnées et d'une collaboration progressive entre utilisateurs autour de cette tâche. (Fig. 7.1)

Après avoir rappelé la motivation de ce modèle, nous détaillerons ce que nous entendons par « progressivité » et par « multi-points de vue ». Ces concepts éclairent les deux principes que nous expliquerons ensuite : la « structuration progressive » puis la « collaboration progressive ».

Variété d'utilisateurs et de métadonnées. L'étude du Web Sémantique et du Web Participatif illustre la diversité des approches possibles de l'indexation : par des experts ou par les utilisateurs, avec des métadonnées structurées ou non-structurées, basées sur des modèles a priori ou émergents. Or chacune de ces approches possède des avantages qui sont abandonnés lorsqu'une autre approche est choisie.

A l'intérieur du courant explorant la fusion du Web Sémantique et du Web Participatif, nous nous sommes aperçus qu'il existait aussi une opposition : d'un côté certaines approches restreignent les différents points de vue des utilisateurs pour les contraindre à travailler sur une même ontologie, d'un autre côté, certaines approches délaissent les apports techniques du Web Sémantique et les raisonnements automatisés.

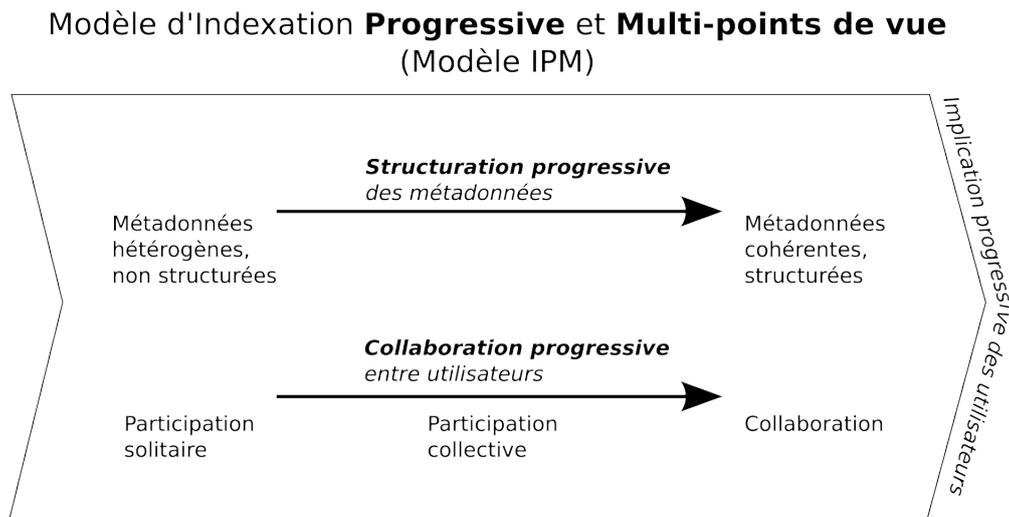


FIG. 7.1 – Illustration du modèle d'Indexation Progressive et Multi-points de vue

Notre contribution consiste à fournir un modèle et des fonctionnalités permettant l'exploitation commune des technologies du Web Sémantique et de la tendance du Web Participatif. Dans cette fusion, nous essayons de préserver les apports des raisonnements automatisés et des interactions entre utilisateurs.

Le nom du modèle que nous proposons comporte deux notions que nous allons maintenant expliquer : la progressivité et l'aspect multi-points de vue. Ces notions servent d'éclairages transversaux aux principes plus pratiques qui en découlent et que nous présenterons juste après.

Progressivité. La progressivité est liée à l'implication progressive des utilisateurs. Fournir des métadonnées de qualité, c'est-à-dire réfléchies et cohérentes, demande de l'énergie aux utilisateurs et donc une motivation importante. Ceci fonctionnera surtout sur les sites où les utilisateurs sont très impliqués. Mais la tendance du Web Participatif a montré que si l'on est moins intransigeant sur la qualité et la cohérence, il est possible de récupérer de nombreuses métadonnées fournies par tout type d'utilisateurs, aux motivations et implications variées.

L'indexation progressive consiste donc à faciliter la participation des utilisateurs à la hauteur de leur implication. Ceci implique des fonctionnalités soutenant plusieurs niveaux de participation et facilitant le passage d'un niveau à l'autre.

Multi-points de vue. L'aspect multi-points de vue se base sur le fait que plus un groupe d'individus est large, plus il y réside d'opinions, de visions du monde et de vocabulaires différents.

Il existe alors au moins deux solutions pour faciliter les échanges entre ces individus.

La première solution consiste à nier ou limiter ces différences. Ceci est effectué en contraignant les individus à avoir la même vision. Ou alors en affirmant que seuls les mots sont différents mais que derrière réside une vision commune du monde.

La seconde solution consiste à accepter les différentes visions et à les laisser s'exprimer. Ensuite, les échanges peuvent être facilités en laissant les visions s'influencer, ou en permettant

l'émergence de visions communes nouvelles.

Si la première solution est applicable en milieu contraint, elle semble plus illusoire en milieu ouvert où la participation est libre. La seconde solution semble alors plus adaptée comme le montre la tendance du Web Participatif.

En insistant sur la notion de points de vues, nous pensons nous inscrire dans le courant du Web Socio-Sémantique (§6.1.4.4) dont c'est aussi un aspect fondamental.

Nous allons voir maintenant comment la progressivité et l'aspect multi-points de vue éclairent transversalement les deux principes du modèle IPM.

Structuration progressive des métadonnées. Améliorer l'organisation de ressources passe par une amélioration des traitements par les machines permise par la structuration des métadonnées.

Dans la vision du Web Sémantique, les machines ont besoin d'entités formelles (concepts, relations. . .) pour comprendre le sens des contenus et pour raisonner dessus. Plus les métadonnées sont exprimées selon ces entités et plus ces métadonnées sont, ce que nous appelons, structurées.

Le problème est que la création de ces entités est une tâche lourde pour des humains quand le nombre de documents est important. En effet, l'indexation doit être précise et les nouvelles entités doivent rester cohérentes avec les anciennes.

Pour résoudre cette difficulté, nous proposons une structuration progressive des métadonnées.

Tout d'abord, l'outil d'indexation doit faciliter au maximum la participation des utilisateurs. Ceci signifie que les ressources et les métadonnées ajoutées par les utilisateurs seront probablement nombreuses et riches par leurs points de vues mais peu homogènes et faiblement structurées.

Ensuite, l'outil d'indexation doit permettre aux utilisateurs de structurer peu à peu ces métadonnées. Ceci passe par la création d'entités formelles et par l'expression des métadonnées selon ces entités.

Cette transformation des métadonnées ne pourra probablement pas se faire immédiatement. L'outil doit alors pouvoir fonctionner avec les métadonnées à la fois structurées et non-structurées.

Collaboration progressive. Améliorer l'organisation de ressources passe aussi par un rapprochement des utilisateurs permis par leur collaboration.

Dans la tendance du Web Participatif, les humains peuvent collaborer sur des objets communs, à l'image des articles de Wikipédia. Mais la collaboration pour créer ou organiser des ressources est généralement une tâche lourde. En effet, celle-ci implique une mise en commun des idées et des négociations entre utilisateurs. Or dans un système à participation libre, les utilisateurs ont des motivations diverses et des niveaux différents d'implications. Ils ne sont pas tous prêts à collaborer.

Pour résoudre cette difficulté, nous proposons, pour indexer les ressources, une collaboration progressive entre utilisateurs.

Nous distinguons ainsi trois niveaux de collaboration entre utilisateurs :

- L'utilisateur peut effectuer ses actions sans se préoccuper des autres ; par exemple, il peut employer l'outil d'indexation pour organiser des ressources en envisageant son seul usage. Ce niveau demande le moins d'implication. Ses métadonnées pourront être exploitées par d'autres utilisateurs mais lui n'a pas à s'en préoccuper. **C'est le niveau solitaire.**
- L'utilisateur peut effectuer ses actions en observant de temps en temps les actions des autres utilisateurs, en s'en inspirant parfois, mais en tout cas, en étant seul responsable de ses ressources et de leur organisation. Par exemple, l'outil d'indexation peut proposer des tags populaires à un utilisateur. Ce dernier pourra alors les ré-utiliser ou non. **C'est le niveau collectif.**
- Enfin, plusieurs utilisateurs peuvent travailler sur la même ressource. Si l'un d'eux veut la modifier, il doit alors négocier avec les autres utilisateurs. Cette participation demande une implication importante. **C'est le niveau collaboratif.**

Un outil soutenant une collaboration progressive est un outil permettant ces trois niveaux de collaboration. Ceci permet aux utilisateurs de participer à la hauteur de leur implication.

De plus, la progressivité doit permettre aux utilisateurs de passer facilement d'un niveau à un autre. Par exemple, si un utilisateur participe de loin depuis quelques mois, il peut soudain avoir un peu plus de temps pour participer davantage. Il va alors commencer à donner son point de vue sur certaines ressources et négocier autour de descriptions communes. Si soudain le temps lui manque, il pourra revenir à son utilisation solitaire de l'outil.

Remarque sur la flexibilité. Les principes ci-dessus appellent une remarque importante sur la cohabitation des formes d'indexation et la flexibilité de l'outil.

Le but des principes précédents est d'amener les utilisateurs à structurer leurs métadonnées et à travailler ensemble pour élaborer des visions communes. Néanmoins il est peu probable, même après un certain temps, que tous les utilisateurs structurent leurs métadonnées et qu'ils collaborent tous ensemble.

La progressivité et l'aspect multi-points de vue incitent alors à ne pas chercher d'absolu mais à laisser cohabiter des métadonnées différemment structurées, des utilisateurs aux participations différentes, des points de vues différents. La qualité de l'outil d'indexation réside dans la capacité à profiter de toutes ces formes d'indexation.

Cette cohabitation est alors un excellent moyen d'avoir un outil flexible, adaptable à différentes situations. Notre participation à plusieurs projets nous a montré la diversité des communautés, au sens large, dans leur composition ou dans leurs membres. La flexibilité est une qualité majeure pour permettre l'appropriation de l'outil par le plus grand nombre de ces communautés.

7.1.3 Fonctionnalités illustrant le modèle IPM

Le modèle IPM expose les principes généraux de structuration et de collaboration progressive. Voici les fonctionnalités que nous proposons pour illustrer la mise en application de ces principes.

Ces fonctionnalités ont été conçues pour un outil de partage de signets. Les lecteurs n'ayant jamais employé ce type d'outil peuvent trouver la présentation de l'un de ses outils en §7.2.1.

7.1.3.1 Fonctionnalités soutenant la structuration progressive

Voici une représentation de la progressivité de la structuration des métadonnées. En gras apparaissent les apports, dont nous détaillerons plus tard l'implémentation (§7.2). La lecture vers la droite va des métadonnées les plus simples vers les aspects les plus structurés.

— Structuration progressive —>

URL	Tags et description	Description individuelle de tags	Tags structurés et description balisée	Export sémantique
-----	---------------------	---	---	--------------------------

URL. A gauche, une ressource ne comporte aucune description. Elle est juste définie par une référence, généralement une URL sur le Web.

Tags et description. Puis, une case à droite, cette ressource peut être décrite à l'aide de tags et d'une description simple. Cette description est juste un texte libre de quelques lignes. Les outils de partage de signets emploient ce genre de métadonnées qui sont simples et rapides à ajouter.

Description individuelle de tags. La description individuelle de tags correspond à la possibilité d'ajouter une description à un tag. Un tag n'est généralement qu'un mot-clef e.g. « plateforme ». Nous proposons la fonctionnalité qu'un utilisateur puisse attacher des descriptions textuelles à ses tags. Par exemple, l'utilisateur peut décrire « plateforme » comme « Tag décrivant les sites parlant de plateforme d'apprentissage comme Moodle ».

D'un point de vue formel, le tag n'est pas véritablement plus explicite ou structuré. Néanmoins, un lien implicite est créé entre le tag et les termes de sa description (entre le tag « plateforme », et les termes « apprentissage » et « moodle »).

D'un point de vue cognitif, cette description peut permettre aux utilisateurs de réfléchir à leur activité d'indexation (méta-réflexion).

Cette fonctionnalité est déjà employée par le site Delicious. L'originalité de notre proposition provient de sa fonctionnalité-sœur permettant une description non pas individuelle mais collaborative des tags, présentées ultérieurement.

Tags structurés et description balisée. Les tags structurés sont peut-être notre contribution la plus originale. Ils consistent en la possibilité, pour les utilisateurs, d'ajouter des relations entre tags pour les structurer explicitement entre eux. Les types de relations possibles sont l'inclusion et la synonymie. Enfin, une caractéristique importante est que les relations peuvent être créées à la volée, en même temps que les tags.

Les descriptions balisées sont les descriptions des ressources, dans lesquels les utilisateurs peuvent créer des balises – e.g. [auteur][/auteur] – et les utiliser pour encadrer des mots de la description. Ces mots balisés pourront, par la suite, être exploités par des machines. Les mots balisés permettent de structurer souplement les descriptions.

Export sémantique. L'export sémantique est une manière d'exploiter le résultat des fonctionnalités précédentes pour produire une représentation formelle. Cet export exploite principa-

lement les relations entre tags structurés. Le format de sortie emploie les concepts de l'ontologie SIOC (§6.1.3.4).

7.1.3.2 Fonctionnalités soutenant la collaboration progressive

Voici une représentation de la progressivité de la collaboration entre utilisateurs. En gras apparaissent les apports, dont nous détaillerons plus tard l'implémentation. La lecture de gauche à droite suit la progression vers les fonctionnalités collaboratives :

— Collaboration progressive —>

Solitaire	Collectif		Collaboratif
Signets	Tags	Contacts, suggestion des tags des contacts, notes privées	Descriptions communes de tags, signets communs

Signets. Le premier usage des signets et des favoris est un usage solitaire. Ils n'étaient qu'une fonctionnalité pour enregistrer un site dans le navigateur Web. Les signets peuvent être employés de manière totalement isolée d'autres utilisateurs.

Tags. Les tags se situent entre un usage solitaire et collectif. En effet, un utilisateur peut employer des tags pour son seul usage, pour classer rapidement des ressources.

Néanmoins, les tags permettent en plus un usage collectif, notamment quand l'outil d'indexation suggère à un utilisateur les tags les plus populaires parmi les autres utilisateurs de l'outil.

Contacts et fonctionnalités relatives. De plus en plus d'outils du Web Participatif offrent la possibilité d'avoir des contacts, c'est-à-dire de créer un lien vers d'autres utilisateurs de l'outil. Par ce lien, l'outil sait que deux utilisateurs se connaissent et peut ainsi proposer des fonctionnalités nouvelles.

Par exemple, la fonctionnalité « Suggestion des tags des contacts » désigne une proposition effectuée par l'outil à un utilisateur en lui affichant les tags les plus employés par ses contacts. La suggestion de tags existe déjà dans plusieurs applications mais généralement la suggestion se base soit sur la seule activité de l'utilisateur, soit sur l'activité de l'ensemble du site. Nous proposons une suggestion employant les tags employés fréquemment par les contacts. Ceci permet de soutenir, au sein d'un outil de partage, les sous-groupes d'utilisateurs aux visions communes mais peut-être différentes de l'ensemble des utilisateurs.

La fonctionnalité « Notes privées » est une note textuelle ajoutée aux signets mais qui n'apparaît qu'à son auteur et à ses contacts. Les notes privées sont utiles pour soutenir les sous-groupes d'utilisateurs de l'outil, dont certaines remarques ne concernent pas les autres utilisateurs. Cette fonctionnalité sera illustrée dans l'expérimentation concernant des documentalistes.

Descriptions communes de tags, signets communs. L'usage collaboratif concerne les fonctionnalités où plusieurs utilisateurs modifient le même objet.

Les outils de partage de signets sont très généralement des outils collectifs, où chaque utilisateur ajoute ses signets et ses tags, mais non collaboratifs.

Pour soutenir la collaboration, nous proposons des espaces communs de descriptions, sous forme de wiki. Tout d'abord, des espaces communs pour décrire les tags. Ces espaces associés aux tags invitent à négocier une définition commune pour chacun d'entre eux. L'avantage de cette description commune est qu'elle peut servir de référence à la communauté mais que si un utilisateur a une définition différente, il peut ne pas se soucier de la définition commune.

Ensuite, nous proposons des espaces communs pour décrire les signets. C'est à dire que pour une ressource, l'outil peut proposer les descriptions et les tags fournis par chaque utilisateur (aspect collectif), mais aussi une description commune issue de négociations entre utilisateurs (aspect collaboratif).

Ces fonctionnalités seront détaillées plus amplement dans l'implémentation de l'outil SemanticScuttle. Avant, nous allons présenter un outil classique de partage de signet, Scuttle, sur lequel est basé SemanticScuttle.

7.2 Développement

Pour implémenter le modèle IPM, nous avons repris un outil nommé Scuttle offrant les fonctionnalités classiques du partage de signets.

Scuttle est alors devenu SemanticScuttle quand nous en avons complété les fonctionnalités pour obtenir une progressivité dans la structuration et la collaboration.

Enfin, nous présenterons quelques fonctionnalités ne s'inscrivant pas directement dans le modèle IPM mais, ayant été réclamées par les utilisateurs, illustrant l'usage général de l'outil.

7.2.1 Développement basé sur Scuttle, un outil classique de partage de signets

Le développement de SemanticScuttle fut basé sur une application libre de partage de signets, nommé Scuttle. Nous présenterons donc tout d'abord le fonctionnement de Scuttle, puis les concepts originaux de SemanticScuttle implémentant le modèle IPM.

Présentation de Scuttle. Scuttle est une application Web de partage de signets sous licence libre, imitant le fonctionnement de Delicious. Le projet Scuttle fut créé en mars 2005 sur Sourceforge⁹⁰ par Marcus Campbell (alias « sronide » sur Sourceforge). L'application est codée en PHP et s'appuie sur une base de données MySQL principalement (quatre tables).

Large diffusion de Scuttle. Cette application rencontra une large diffusion sur Internet qui peut se mesurer à plusieurs critères :

- au nombre important de traductions (anglais, chinois simplifié, danois, hollandais, français, allemand, japonais, lituanien, portugais...)
- au nombre de téléchargements (48000 téléchargements entre mars 2005 et mars 2009, soit en moyenne 1000 téléchargements par mois)
- aux applications Web basées dessus comme ScuttlePlus⁹¹ (offrant une interface avancée d'administration).

⁹⁰Sourceforge, site soutenant la création de logiciels libres : <http://sourceforge.net/projects/scuttle/>

⁹¹ScuttlePlus : <http://www.scuttleplus.org/>

- aux individus malveillants proposant des services pour spammer les sites employant Scuttle (comme le site Automatic Scuttle Submitter⁹² dont le slogan est « Drive Massive Traffic to Your Websites from Thousands of Scuttle Sites! »).

Dans le but d’obtenir un maximum d’utilisations de SemanticScuttle, s’appuyer sur une application répandue était un atout.

Fonctionnement général. Scuttle, comme Delicious, permet à des utilisateurs de s’inscrire puis d’enregistrer des signets sur le site. Chaque signet est décrit à l’aide d’un titre, d’une courte description textuelle et surtout de tags qui sont des termes choisis librement et séparés par des virgules.

Scuttle fonctionne selon deux types d’espaces, visibles de tous :

- **Des espaces individuels** où se retrouvent tous les signets et les tags par utilisateur. Chaque utilisateur est seul responsable de son espace. Cet espace est d’ailleurs à son nom avec une adresse ressemblant à « `www.example.com/bookmarks/bernard` ».
- **Un espace global** où sont agrégés automatiquement les signets et tags des espaces individuels. L’agrégation est effectuée généralement par popularité i.e. par ordre décroissant de nombre d’utilisation.

Quand il est connecté, un utilisateur peut retrouver tous ses signets dans son espace. Il peut parcourir ses propres signets ou ceux des autres utilisateurs grâce aux tags.

Si un utilisateur clique sur le tag « chimie », tous les signets décrits par « chimie » apparaissent. Comme les signets sont généralement décrits par plusieurs tags, le système peut afficher alors sur le côté les tags co-occurents à « chimie », c’est-à-dire apparaissant sur les signets décrits par « chimie ».

Fonctionnalités écran par écran. Scuttle illustre par ses fonctionnalités et son ergonomie générale de nombreux aspects vus dans les applications participatives : facilité d’utilisation, simplicité de l’interface, côté social mis en avant. Les Fig. 7.2 et 7.3 illustrent quelques points de Scuttle.

Remarque : Suite à des variations de traductions, nous parlerons toujours de « tags » mais la traduction française de Scuttle parlera généralement de « labels ».

Comme le développement de SemanticScuttle est basé sur celui de Scuttle, il nous semble important de détailler cette application écran par écran (Annexe I).

Modèle sous-jacent à Scuttle. La Fig. 7.4 est une représentation en UML du modèle sous-jacent à Scuttle. Ce modèle est le modèle classique des applications de partage de signets comme Delicious ; il nous servira ensuite à préciser par comparaison les nouveautés de SemanticScuttle par rapport à ces outils.

Dans Scuttle, une URL (en bas de la figure) est décrite par un ou plusieurs signets d’utilisateurs. Chaque signet possède un auteur (utilisateur), une description, et est associé à des tags. Un utilisateur peut avoir des contacts qui sont des utilisateurs dont il décide de suivre les

⁹²<http://www.scuttlesubmitter.com/>



FIG. 7.2 – Page d'accueil de Scuttle. La majorité des pages de Scuttle ont la même structure, simplifiant ainsi la navigation. Tout d'abord, la plus grande zone (1) affiche les signets les uns sous les autres. Chaque signet ne se compose que de quelques métadonnées : un titre, une description, une date d'ajout, un auteur et un ensemble de tags qui sont les termes en bleus séparés par des virgules. Ensuite, chaque page affiche sur le côté une zone de menu dynamique (2). Ces menus dépendent de la page visitée. Par exemple, la page d'accueil comporte un seul menu qui affiche les derniers tags employés par les utilisateurs. Enfin, toutes les pages possèdent un menu général dans le bandeau supérieur (3). Lorsque l'utilisateur est identifié, ce menu lui permet de naviguer parmi différents espaces personnels que nous verrons plus tard. Enfin parmi les autres fonctionnalités, nous pouvons voir l'icône des flux RSS (2) permettant de suivre les nouveaux signets pour chaque page.

scuttle bookmarks watchlist ajouter un signet quitter

ajouter un signet

Adresse ← Requis

Titre ← Requis

Description

Labels ← Séparés par des virgules

Vision ▼

Ajouter signet

1

Labels populaires

moteur recherche france usa

2

FIG. 7.3 – Page d’ajout d’un signet dans Scuttle. Dans Scuttle, tout est conçu pour accélérer l’ajout d’un nouveau signet. Cet écran montre le nombre restreint de champs (1). De plus, seul le titre et l’URL de la ressource sont obligatoires. Le menu « Vision » permet de déterminer si le signet sera visible uniquement par l’auteur, aussi par ses contacts (voir l’écran « watchlist » en Annexe I) ou par tous. La zone de tags permet d’ajouter les tags en les séparant par des virgules. Néanmoins une aide est fournie dans la partie inférieure (2) qui affiche les tags les plus fréquents. En cliquant sur un tag proposé, celui-ci est ajouté au signet.

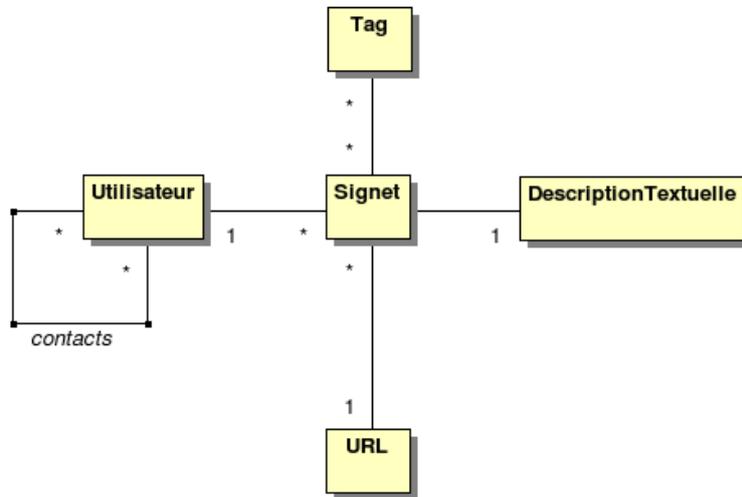


FIG. 7.4 – Modèle UML de Scuttle

derniers signets.

La Fig. 7.5 donne un exemple d’emploi de Scuttle pour deux ressources pédagogiques décrites par Anne (traits rouges) et Bernard (traits verts). Anne a indexé une seule ressource (signet 1) avec les tags « devoir » et « java » tandis que Bernard a indexé deux ressources (signets 2 et 3) avec les tags « java », « examen » et « exercice ». De plus, ils ont ajouté des descriptions textuelles pour les signets 1 et 2. Cette Fig. illustre notamment qu’une URL peut être décrite par plusieurs utilisateurs (signets 1 et 2) et qu’un tag permet de retrouver plusieurs signets (tag « java » attaché aux trois signets).

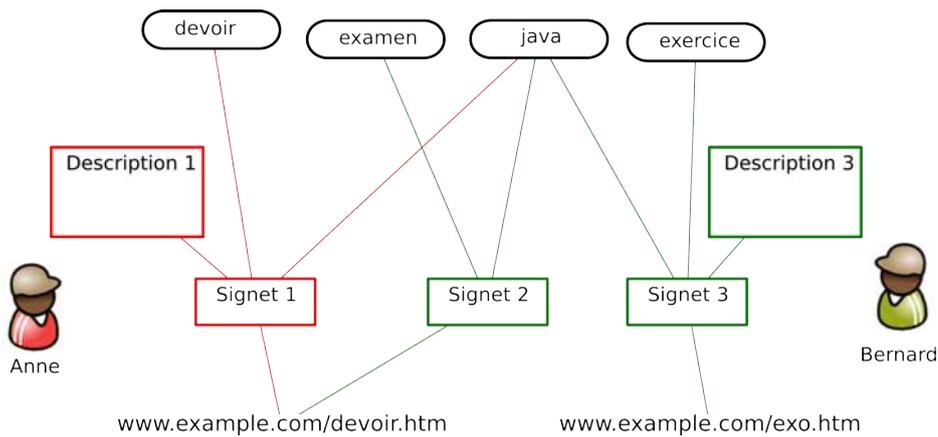


FIG. 7.5 – Fonctionnement de Scuttle avec deux ressources et deux utilisateurs.

De Scuttle à SemanticScuttle. Scuttle est un outil représentant parfaitement les outils classiques de « Social Bookmarking » dont il en possède les principales fonctionnalités. Néanmoins, comme les autres outils du même type, Scuttle ne propose pas de véritables moyens de

structurer les descriptions ni de fonctionnalités avancées pour collaborer.

C'est à partir de cet outil, avec ses avantages et ses limites, que nous avons développé et exploré le modèle IPM. La nouvelle version de Scuttle que nous maintenons s'appelle SemanticScuttle :

Projet SemanticScuttle : <http://sourceforge.net/projects/semanticscuttle/>
 Démo : <http://festic.fr/semanticscuttle/demo/fr/>

Techniquement, SemanticScuttle reprend la logique de Scuttle et l'ergonomie générale mais en y adaptant des fonctionnalités pour les rendre plus structurantes ou collaboratives. Voici une comparaison des fonctionnalités.

Scuttle	SemanticScuttle
Signets	
Individuels	Individuels ou commun
Description simple	Description simple et balisée
	Notes privées
	Signets certifiés
Export HTML	Export HTML et sémantique
Tags	
Tags	Tags structurés
	Descriptions individuelles et communes de tags
Contacts	
Surveiller les signets des contacts	Surveiller les signets des contacts Proposition des tags des contacts

Pour clarifier le discours, nous présenterons les apports de SemanticScuttle de la manière suivante :

- Fonctionnalités de structuration progressive :
 - tags structurés ;
 - descriptions balisées ;
 - export sémantique.
- Fonctionnalités de collaboration progressive :
 - notes privées ;
 - descriptions individuelles et communes de tags ;
 - signets communs.

Enfin nous présenterons quelques fonctionnalités n'entrant pas dans ces catégories : rôle d'administrateur et signets certifiés, widgets et moteur Google Custom Search.

7.2.2 Fonctionnalités de structuration progressive

Ces fonctionnalités permettent la structuration progressive des contenus, c'est à dire l'organisation des métadonnées d'un point de vue informatique.

7.2.2.1 Tags structurés

Comme Scuttle, SemanticScuttle permet d'ajouter des tags de manière libre et simple. Cependant il permet aussi, si l'utilisateur le désire, de les structurer. Ceci se produit quand l'utilisateur crée des liens entre les tags ; ce que nous appelons les « tags structurés ».

Le principe des tags structurés est que de la même manière qu'un utilisateur ajoute des tags librement et à la volée, il peut créer librement et à la volée des relations entre les tags au moment d'ajouter un tag à un signet.

SemanticScuttle exploite deux relations : inclusion et synonymie. Ces relations reprennent les relations de base définies par les systèmes de gestion de connaissances (voir SKOS 6.1.3.3). Si un tag A est inclus dans un tag B alors le système interprète que toutes les ressources taggées par A, le sont aussi par B. Si un tag A est synonyme d'un tag B alors le système interprète que toutes les ressources tagguées par A le sont aussi par B, et inversement. Ces relations permettent d'améliorer les recherches parmi les tags qui sont classiquement à plat.

L'ajout d'une relation entre tags se fait à la volée, au moment de l'ajout des tags. Exemple : considérons un utilisateur qui ajoute un signet pointant vers un tutoriel XML. Avec des tags classiques, il aurait ajouté dans une zone de tags quelque chose comme : « xml, tutorial, tutoriel ». Il pourrait ainsi retrouver le signet en tapant ces termes. Il met « tutorial » et « tutoriel » car ces deux termes sont pour lui quasiment équivalents et qu'il les emploie indistinctement. En mettant les deux tags dans la description du signet, il ne ratera pas celui-ci s'il cherche l'un ou l'autre des deux tags.

Avec les tags structurés, l'utilisateur peut dans la zone de tags mettre : « format>xml, tutorial=tutorial ».

Le caractère « > » indique une relation d'inclusion entre les tags « format » et « xml ». Cette relation est enregistrée par le système. Par la suite, chaque fois que l'utilisateur cherchera les signets portant sur un « format » informatique (en cherchant le tag « format »), le système renverra les signets taggués par « xml ». Plus tard, l'utilisateur pourrait ajouter une relation entre « format » et « html ». Par conséquent, en cherchant le tag « format », il trouvera les signets taggués par « xml » et par « html ». Le système construit ainsi peu à peu un graphe entre les tags.

Le caractère « = » indique au système de créer une relation de synonymie entre « tutorial » et « tutorial ». Ainsi chaque fois que l'utilisateur cherchera les signets taggués par « tutorial », le système lui fournira en plus les signets décrits par « tutorial ».

Si un utilisateur n'est pas satisfait d'une relation qu'il a précédemment ajoutée, il peut la modifier à travers une fenêtre de gestion des relations entre tags.

Les tags structurés ont l'avantage d'être robustes dans le sens où le système ne demande pas à être cohérent et ne peut pas bugger. Si l'utilisateur ajoute, par erreur, les relations cycliques « A>B>C>A » alors le système continuera tout de même à fournir des résultats. Il interprétera simplement que les tags A, B et C sont synonymes, c'est à dire qu'en cherchant les ressources taggées par le tag C, il renverra aussi les ressources taggées par A et par B. Cette robustesse s'oppose aux systèmes fondés sur des ontologies formelles devant généralement conserver une

cohérence stricte pour fonctionner. Cette robustesse est semblable aux systèmes de tags qui continuent à fonctionner même si un utilisateur pose un mauvais tag sur une ressource.

Le schéma UML 7.6 illustre les tags structurés. Un tag structuré n'est qu'un tag relié à d'autres tags pour un utilisateur donné.

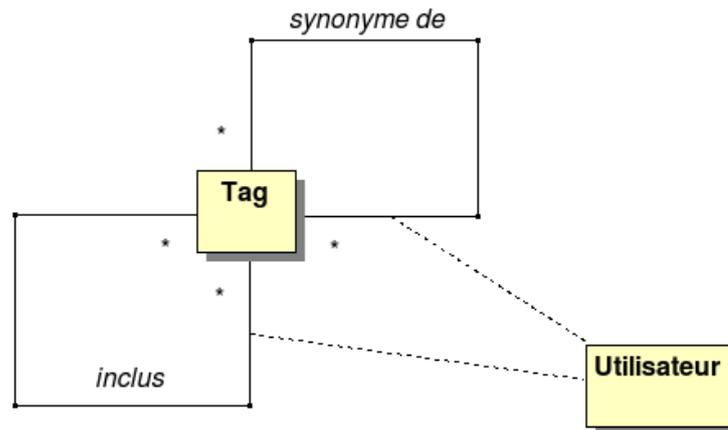


FIG. 7.6 – Représentation UML de la notion de tags structurés, qui sont des tags reliés par des liens (inclusion ou synonymie) propres à chaque utilisateur.

La Fig. 7.7 propose un exemple pour deux ressources et deux utilisateurs.

De cet exemple découle les vues suivantes pour chacun des utilisateurs. Pour Anne :

- activité
 - devoir, examen
 - exercice

Pour Bernard :

- langage
 - php
 - java

Si l'utilisateur désire effectuer une modification de ses tags structurés, il peut le faire à l'aide d'une interface spécifique (7.8) lui permettant d'ajouter ou d'effacer une relation (« > » ou « = ») entre deux tags.

Agrégation des relations. L'avantage des relations telles que nous les avons définies est qu'elles peuvent ensuite être agrégées, comme des tags, sans trop de difficultés. Par exemple, le système pourrait calculer une vue d'ensemble des tags structurés en affichant tous les tags et toutes les relations entre eux. Exemple : si « A>B » pour un utilisateur et « A>C » pour un autre utilisateur, alors le système peut déduire globalement que A inclus B et C.

Cette arborescence agrégée peut ensuite être améliorée en fonction du nombre de participants qui ont créé une même relation c'est-à-dire pointant d'un même tag pour aller vers le même autre. Certaines relations entre tags pourront être plus ou moins importantes et mises en

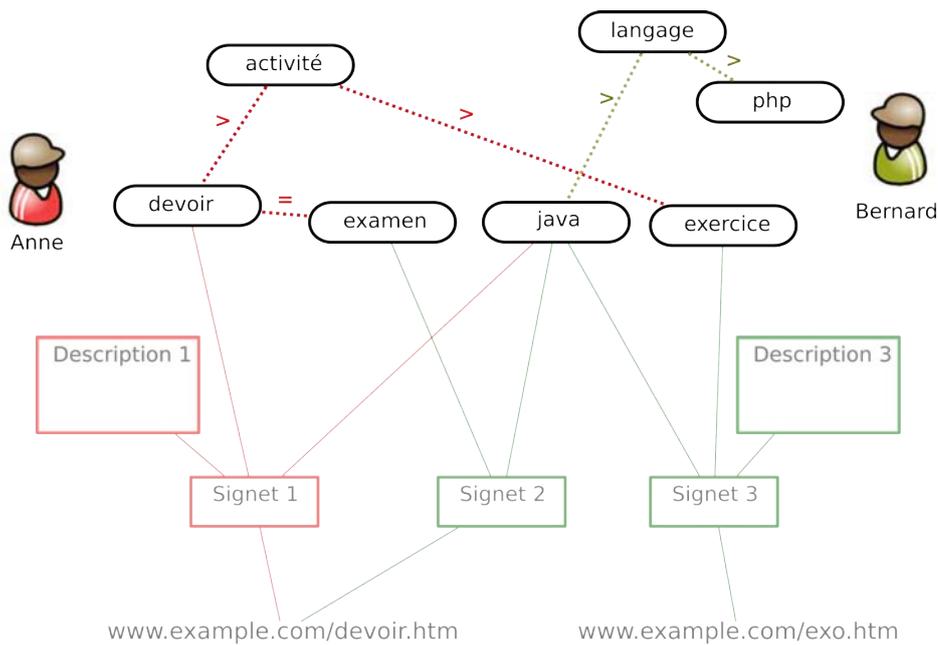


FIG. 7.7 – Exemple de tags structurés (boîtes noires en haut) ajoutés par deux utilisateurs sur trois signets.

ajout d'un lien entre tags:

Créer un nouveau lien

>

Note : inclure un tag dans le tag 'menu' (e.g. menu>countries) affiche ce tag dans la boîte de menu

Liens existants : langage > java langage > php

FIG. 7.8 – Interface de gestion des tags structurés

avant selon le nombre de contributeurs.

Ces vues agrégées ne sont pas totalement cohérentes, à l'image des tags qui ne sont pas des ensembles parfaitement cohérents. Néanmoins, ces agrégations permettent d'obtenir une vue globale sur un domaine, utile pour le parcours des ressources.

Tags « menu ». Les tags structurés sont un premier moyen d'organiser les tags. SemanticScuttle offre un second moyen permettant de mettre en avant certains tags par rapport à d'autres. Ce sont les tags menu.

Lorsqu'un tag est déclaré comme tag menu, il apparaîtra dans une boîte menu sur la page de l'utilisateur. Dans Scuttle, quand Anne visitait l'espace de Bernard, elle voyait sur la gauche les signets de Bernard et sur la droite les tags les plus fréquemment employés par Bernard. Dans

SemanticScuttle, elle voit aussi sur la droite un menu formé des tags menu de Bernard et des sous-tags inclus dans les tags menu (7.9).

Pour conserver la notion de rapidité des tags, les tags menu peuvent être ajoutés à la volée. Il suffit pour un utilisateur d'inclure le tag dans le tag spécial nommé « menu ». Chaque utilisateur peut ainsi définir son propre menu de tags.

Lorsque SemanticScuttle possède des administrateurs, ces derniers peuvent aussi définir des tags menu, qui pourront alors apparaître sur la page principale du site. Cette vision permet à un visiteur ne connaissant pas le domaine ou le site Web, d'appréhender rapidement un angle intéressant sur les contenus.

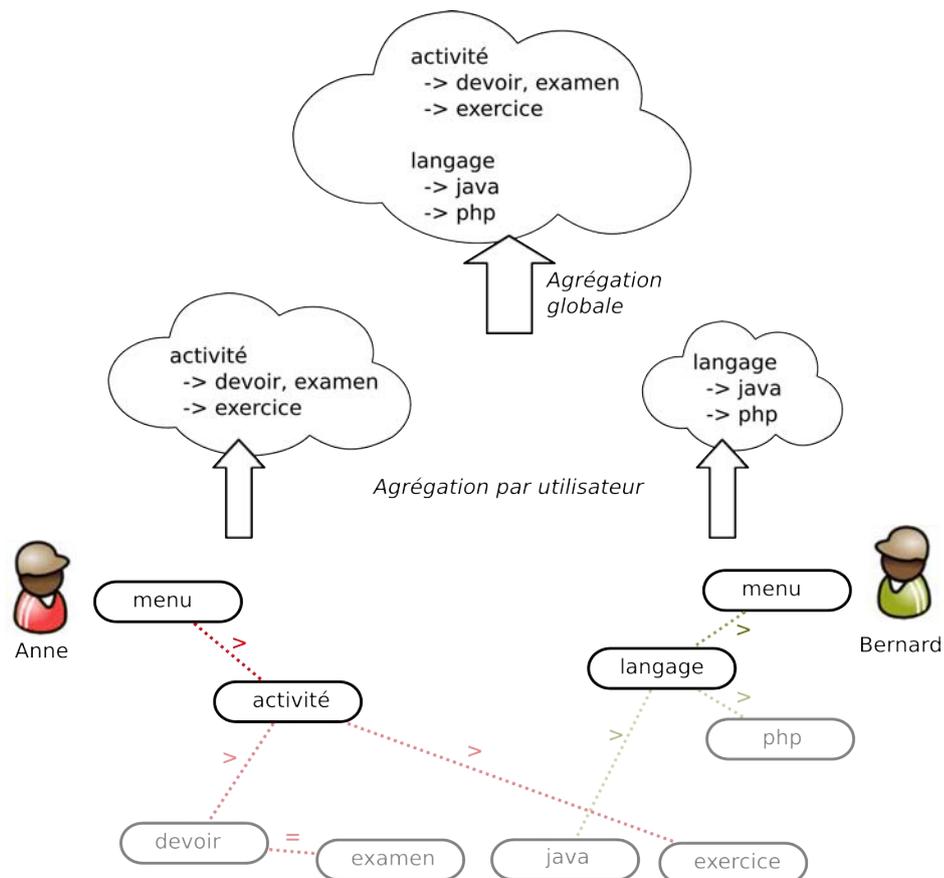


FIG. 7.9 – Tags menu et mécanisme d'agrégation au niveau de chaque utilisateur puis au niveau global.

Le bandeau latéral affichera une vue agrégée comme sur la Fig. 7.10.

Que se passe-t-il si... Nous avons développé et exploré les possibilités des tags structurés pour quelques utilisateurs. Que se passe-t-il si un nombre important d'utilisateurs crée des tags structurés ? Qu'affichera le système en cas d'incohérence ?

Le premier problème qui peut surgir est celui des cycles entre les tags, c'est-à-dire si un tag A inclut le tag B, qui inclut le tag A. Au moment d'une recherche, le système ne bugue pas



FIG. 7.10 – Agrégation des tags (cette version de l’interface ne montre pas les synonymes)

mais il va juste chercher les ressources associées à A et renverra aussi les ressources associées à B, et réciproquement. Lors de l’affichage des tags structurés, le système commence toujours par afficher les tags menu et ne ré-affiche pas un tag déjà affiché dans l’arborescence. Ainsi, si dans le cycle $A > B > A$, le tag A est un tag menu, alors le système va afficher une arborescence commençant par A, puis le tag inclut B mais il ne va pas ré-afficher le tag A qui apparaît déjà dans l’arborescence. Ainsi, les cycles ne posent pas de problème lors de l’affichage arborescent.

Le second problème concerne la compression des tags structurés, c’est-à-dire si le système possède une centaine de tags menu structurés, lesquels afficher prioritairement.

Plusieurs critères sont possibles pour sélectionner les tags à afficher. En voici deux :

- Par popularité. Comme pour des tags classiques, le système peut sélectionner à l’affichage les tags qui ont été choisis comme tag menu par le plus grand nombre d’utilisateurs.
- Par quantité de tags descendants. Le système sélectionne les tags menu dont le nombre de tags descendants (inclus, ou inclus dans des tags inclus, etc) est le plus grand.

Ces critères peuvent ensuite être combinés entre eux ou alors combinés avec les critères des tags classiques comme le nombre de ressources décrites par ces tags ou la date de dernière utilisation.

7.2.2.2 Descriptions balisées

Un des avantages de SemanticScuttle est le petit nombre de champs à remplir pour ajouter un nouveau signet (URL du site, titre, description libre et tags). Mais en travaillant avec les documentalistes (§7.3), nous nous sommes aperçus qu’elles avaient parfois des champs supplémentaires – dans leur précédent outil – qu’elles voulaient employer. Par exemple, beaucoup de leurs signets possèdent un ou plusieurs éditeurs définis par une dénomination et une ville (e.g. « Bibliothèque nationale de France, Paris »). Les documentalistes ne voulaient pas perdre ces informations structurées.

SemanticScuttle propose alors un système de balises libres que les utilisateurs peuvent créer à l’intérieur du champ Description (voir point 1 sur 7.11). Ces balises s’écrivent entre crochets sous la forme « [nomBalise]valeurBalise[/nomBalise] ». Pour un éditeur, ceci donne par exemple « [Editeur]Bibliothèque nationale de France, Paris[Editeur] ». Les utilisateurs peuvent inventer les balises qu’ils veulent sans contraintes.

Lors de l’affichage d’un signet, SemanticScuttle interprète automatiquement un champ balisé pour qu’il apparaisse sous la forme plus agréable de « nomBalise » valeurBalise ». Exemple pour

l'éditeur : « Editeur » Bibliothèque nationale de France, Paris ».

Toujours dans un esprit de flexibilité, les utilisateurs sont libres de créer des balises ou de ne pas en utiliser du tout. Néanmoins, SemanticScuttle permet d'inciter à employer certaines balises qui apparaissent à droite du champ. Sur la Fig. 7.11, ceci correspond à « Auteur Editeur FormatTechnique » dans l'encadré 1. En cliquant sur l'un de ces termes, l'application crée automatiquement la balise correspondante dans le champ Description. (E.g. en cliquant sur « Auteur », l'application fait apparaître « [Auteur][/Auteur] » dans le champ Description).

L'avantage des descriptions balisées est que les utilisateurs peuvent structurer leurs descriptions s'ils le désirent. Éventuellement, lors de l'export des données, ces champs pourront être repérés et traités automatiquement par une autre application.

The screenshot shows the 'editer le signet' interface with the following elements:

- Adresse:**
- Titre:**
- Description:**
 - ← Requis
 - ← Requis
 - ← Vous pouvez utiliser des balises pour délimiter des attributs. Par exemple : [publisher]blah[/publisher]
 - Balises suggérées : [Auteur](#) [Editeur](#) [FormatTechnique](#)
- Note privée:** (2)
- Tags:**
 - ← Visible uniquement par vous et vos contacts.
 - ← Séparés par des virgules
 - note : utiliser ">" pour inclure un tag dans un autre. ex: europe>france>paris
 - note : utiliser "=" pour rendre deux tags synonymes ex: europeu
- Accès:** Publique
- Buttons: Enregistrer les modifications, Annuler, Supprimer le signet, **éditer la description** (4)

FIG. 7.11 – Ajouter un signet sur SemanticScuttle avec description balisée (1), note privée (2), tags structurés (3) et éventuellement une description commune (4).

7.2.2.3 Export sémantique (SIOC)

SIOC, présentée en §6.1.3.4, est une ontologie permettant de représenter sémantiquement les sites participatifs à travers notamment les participants, les contenus ajoutés et les tags.

SemanticScuttle s'inscrivant parfaitement dans le domaine d'application de SIOC, nous avons réalisé un plugin exportant le contenu d'un site SemanticScuttle au format SIOC. Selon ce schéma, les utilisateurs sont représentés à l'aide de l'ontologie de relations sociales – FOAF – et les différents termes entre utilisateurs à l'aide de l'ontologie de termes – SKOS.

Puis cet export permet à un moteur sémantique externe de « raisonner » sur les ressources d'une application SemanticScuttle, dans l'optique du Web Sémantique vu par le W3C. Ceci est le dernier maillon permettant d'aboutir à une représentation sémantique selon le W3C, des données hétérogènes de Scuttle.

Voici un extrait de l'export SIOC pour l'exemple précédent avec deux utilisateurs et trois signets :

```
<rdf:RDF>

<sioc:Site rdf:about="http://www.site.com/">
<rdf:label>SemanticScuttle</rdf:label>
</sioc:Site>

<sioc:User rdf:about="http://www.site.com/profile/bernard">
<sioc:name>bernard</sioc:name>
<sioc:member_of rdf:resource="http://www.site.com/">
</sioc:User>

<sioc:User rdf:about="http://www.site.com/profile/anne">
<sioc:name>anne</sioc:name>
<sioc:member_of rdf:resource="http://www.site.com/">
</sioc:User>

<bm:Bookmark rdf:about="http://www.site.com/history/9d5347e3e0dc85f6bb96dee2cf85caa1">
<dc:title>Exercices</dc:title>
<dc:description>Description 3</dc:description>
<bm:recalls rdf:resource="http://www.example.com/exo.htm"/>
<sioc:owner_of rdf:resource="http://www.site.com/profile/bernard"/>

<sioc:topic>
<skos:concept rdf:about="http://www.site.com/bookmarks/bernard/java"/>
</sioc:topic>

<sioc:topic>
<skos:concept rdf:about="http://www.site.com/bookmarks/bernard/exercice"/>
</sioc:topic>
</bm:Bookmark>

...

<skos:concept rdf:about="http://www.site.com/bookmarks/bernard/langage"/>
<skos:narrower>
< dct:creator>
<sioc:User rdf:about="http://www.site.com/profile/bernard">
</dct:creator>
<skos:concept rdf:about="http://www.site.com/bookmarks/bernard/php"/>
</skos:narrower>
<skos:narrower>
<skos:concept rdf:about="http://www.site.com/bookmarks/bernard/java"/>
< dct:creator>
<sioc:User rdf:about="http://www.site.com/profile/bernard">
</dct:creator>
</skos:narrower>
</skos:concept>

...

</rdf:RDF>
```

L'extrait ci-dessus représente l'export SIOC dans son état actuel qui n'implémente pas encore complètement la sémantique de SemanticScuttle, notamment pour les liens entre concepts SKOS.

De ce fait, cet export peut être interrogé par d'autres applications sémantiques. Par exemple, le site Sparkler⁹³ permet d'effectuer une requête SPARQL sur n'importe quelle application SemanticScuttle en indiquant la bonne URL :

```
PREFIX dc: <http://purl.org/dc/elements/1.1/>
PREFIX sioc: <http://rdfs.org/sioc/ns#>
PREFIX skos: <http://www.w3.org/2004/02/skos/core#>

SELECT ?n ?u
FROM <http://www.site.com/api/export_sioc.php>
WHERE
{
  ?b sioc:topic ?n .
  ?b sioc:owner_of ?u .
  ?b dc:title ?title
  FILTER (regex(?title, "java")) .
}
```

La requête ci-dessus recherche ainsi les tags (et leurs auteurs) attachés à des signets dont le titre contient le terme « java ».

En conclusion, l'export en SIOC permet de représenter selon un formalisme partageable les utilisateurs, les signets et descriptions de SemanticScuttle. Les relations entre tags peuvent y être inscrits en reprenant les relations de SKOS même si certains points ne sont pas encore résolus comme :

- **la transitivité de l'inclusion** que considère SemanticScuttle mais qui n'est pas directement exprimable avec SKOS. SKOS ne permet d'exprimer la transitivité qu'au moment de la recherche (« broaderTransitive » ou « narrowerTransitive ») ;
- **le point de vue sur les concepts.** Les spécifications de SKOS mentionnent bien la possibilité de regrouper les concepts dans différents thésaurus (« scheme »), de préciser qui a créé un concept (« dc :creator ») ou la possibilité de relier des concepts entre eux (« skos :exactMatch »). Pourtant ces possibilités ne sont pas encore établies par des usages et nous manquons de recul sur leur utilisation. Comme le précisait le guide d'utilisation de SKOS « @@Note : les sémantiques de skos :exactMatch, skos :closeMatch, skos :broadMatch, skos :narrowMatch et skos :relatedMatch sont risquées et peuvent être changées dans des versions futures de SKOS.@@ »⁹⁴

Selon nous, ces difficultés dûes notamment à la jeunesse de SKOS pourront être corrigées ultérieurement.

Néanmoins, la formalisation des termes employés par les internautes ne signifie pas une compréhension totale des échanges au niveau sémantique par les machines. Selon nous, la structuration des contenus doit aussi être accompagnée d'une collaboration progressive entre les utilisateurs pour qu'ils tendent vers des visions proches de leur domaine, indépendamment de la compréhension du domaine par les machines.

⁹³Requêtes SPARQL en ligne : <http://www.sparql.org/sparql.html>

⁹⁴Traduction du guide utilisateur de SKOS version d'Octobre 2008 : <http://www.w3.org/TR/2008/WD\discretionary{-}{-}skos\discretionary{-}{-}primer\discretionary{-}{-}20080829/#secmapping>

7.2.3 Fonctionnalités de collaboration progressive

Les fonctionnalités suivantes permettent la collaboration autour de l'organisation des données. Elles facilitent la création de vues communes entre les utilisateurs.

7.2.3.1 Notes privées visibles par les contacts

Les documentalistes avec lesquelles nous avons collaboré, ont exprimé le besoin d'ajouter des notes aux signets mais qui n'apparaîtraient pas pour les visiteurs simples. Ces notes contiennent les remarques propres à la description des signets, e.g. indiquant un signet à mettre à jour, ou servant à se souvenir d'une question relative au signet (Schéma UML en Fig. 7.12).

Ce champ Note privé apparaît sur la Fig. 7.11 dans l'encadré 2. Lors de l'affichage des signets, le contenu de ce champ n'apparaît qu'au propriétaire du signets et à ses contacts.

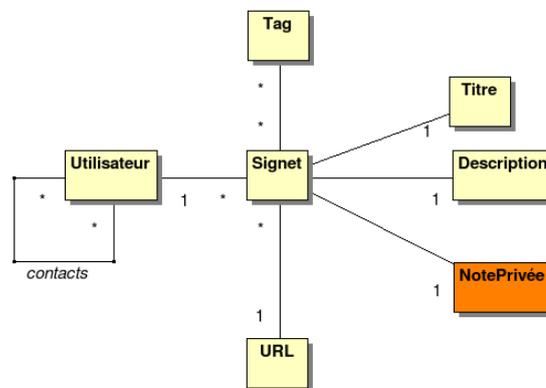


FIG. 7.12 – Les notes privées sur les signets ne sont visibles que des auteurs et de leurs contacts.

La Fig. 7.13 illustre une note privée laissée par Bernard sur le premier signet. Lorsque les signets sont affichés, les notes apparaissent en bas à droite des signets, en majuscule.



FIG. 7.13 – Le premier signet de Bernard possède une note privée (« Je ne suis pas sûr... »), visible par Anne si elle est dans ses contacts.

7.2.3.2 Descriptions de tags : individuelles et communes

Pour les outils de partage, nous définissons trois formes de participation : solitaire, collective et collaborative (voir modèle IPM, 7.1.2).

SemanticScuttle explore le passage de descriptions collectives, propres à un utilisateur mais visibles par les autres, à des descriptions collaboratives, communes à tous les utilisateurs. Ceci s'effectue à l'aide des espaces de descriptions des tags.

En effet, dans SemanticScuttle, un utilisateur peut ajouter une définition à chacun des tags. Il peut le faire dans un espace individuel et dans ce cas, il est le seul à pouvoir éditer la description ou alors il peut le faire dans un espace commun (format wiki) et alors tout le monde pourra modifier cette description. L'idée sous-jacente est de laisser les participants collaborer progressivement, s'ils le désirent, à une vue commune sur un tag.

Les descriptions ne sont pas employées par le système informatique pour rechercher des signets, elles ne servent qu'aux participants humains dans le but de s'organiser. Cette fonctionnalité s'inscrit dans la vision du Web Socio-Sémantique (§6.1.4.4) où les ontologies ont pour but de favoriser les interactions entre humains.

Le schéma 7.14 détaille ces descriptions.

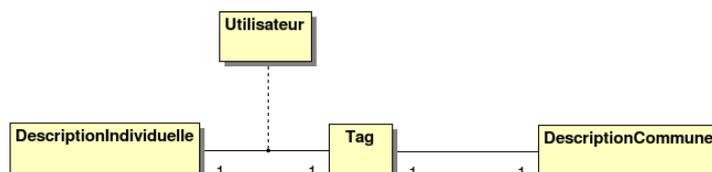


FIG. 7.14 – Les tags peuvent posséder une description individuelle, liée à un seul utilisateur, ou bien commune i.e. éditable par tous sous la forme d'un wiki.

La Fig. 7.15 illustre la description commune pour le tag « java ».

Pour l'instant, le seul moyen offert par SemanticScuttle pour négocier est un espace wiki. Dans le futur, l'idéal serait d'ajouter un forum intégré à l'outil ou bien un espace de t'chat pour communiquer de manière synchrone.

7.2.3.3 Signet commun (ou description commune de signet)

SemanticScuttle implémente aussi un signet commun pour décrire une URL (schéma 7.16). Pour ne pas complexifier l'implémentation, nous n'avons pas exploré des tags communs associés aux signets communs ; un signet commun est donc aussi parfois appelé « description commune de signets ».

Sur la Fig. 7.11, le point 4 montre un lien permettant d'éditer une description commune d'une URL. Comme sur un espace Wiki, chaque ajout d'un utilisateur modifie les travaux de l'utilisateur précédent cependant toutes les versions sont conservées en base de données. La Fig. 7.17 montre l'édition et la Fig. 7.18 la visualisation.

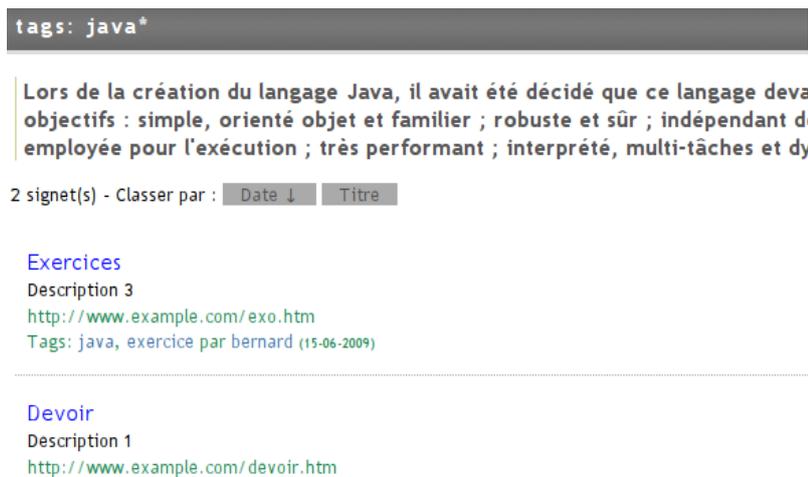


FIG. 7.15 – La description commune du tag « java » apparaît au dessus des signets taggués « java ». (Un lien permet d'éditer cette description mais n'apparaît pas sur la figure.)

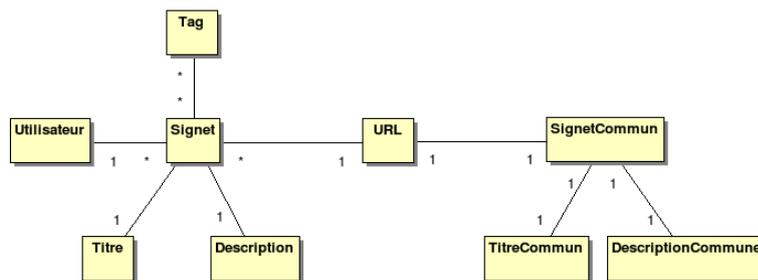


FIG. 7.16 – Chaque référence (URL) peut être décrite à l'aide de signets individuels, propres aux utilisateurs, ou à l'aide d'un signet commun (espace wiki).

7.2.4 Autres fonctionnalités et caractéristiques

Le développement de SemanticScuttle a été effectué en prenant en compte autant que possible les besoins des utilisateurs. Voici les principaux ajouts réclamés par ces derniers et illustrant l'usage général de l'outil.

7.2.4.1 Mode avec ou sans administrateurs

Dans Scuttle, tous les utilisateurs inscrits étaient au même niveau ; il existait un seul « rôle ». SemanticScuttle offre un niveau supérieur avec les administrateurs (appelés aussi « utilisateurs certifiés »).

Du point de vue des actions possibles, les administrateurs peuvent afficher des tags comme principaux ou éditer les tags des autres utilisateurs. Les tags principaux sont des tags apparaissant sur la page principale de l'outil ; c'est l'équivalent des tags "menu" mais sans être agrégés.

editer la description commune du signet: <http://www.example.com/devoir.htm>

Titre

Description

Description commune de devoir...

Description collaborative : ces champs peuvent être vus et modifiés par tous les utilisateurs.

Dernière modification : 2009-06-15 12:53:45, bernard

FIG. 7.17 – Fenêtre d’édition d’un signet commun. Comme sur un wiki, toutes les versions sont sauvegardées (Bernard fut le dernier éditeur).

Dans notre vision, l’édition des tags des utilisateurs simples ne doit être employée qu’en cas de spam. Il semble sinon préférable de leur laisser une assez grande autonomie.

Du point de vue affichage, les signets des administrateurs possèdent des indices graphiques (étoile et bandeau de couleur jaune) pour les différencier des contenus non certifiés. De plus, l’espace de chaque administrateur comporte un message « Cette page est gérée par un utilisateur certifié. »

7.2.4.2 Identification des ressources certifiées

Les utilisateurs certifiés (appelés aussi « administrateurs ») se distinguent par deux points particuliers des autres utilisateurs.

Tout d’abord leurs signets apparaissent avec une couleur distincte facilement identifiable sur le site. Ainsi si un simple utilisateur effectue une recherche, il voit rapidement les signets certifiés de ceux ajoutés par n’importe quel utilisateur inscrit sur le site. Ceci peut jouer dans la confiance donnée aux différents signets.

Ensuite, la page principale peut afficher les tags structurés des utilisateurs certifiés. Ainsi, ces utilisateurs certifiés peuvent créer une arborescence pour parcourir les signets du site et proposer cette arborescence aux internautes arrivant sur le site. Néanmoins les internautes ne seront pas obligés d’employer cette arborescence.

7.2.4.3 Accès externes aux signets : Widget

Comme nous le mentionnions dans les caractéristiques du Web Participatif, les utilisateurs ont d’autant plus intérêt à ajouter des contenus à un site, s’ils peuvent ensuite afficher ces contenus sur d’autres sites leur appartenant.

Les widgets sont de mini-plugins qui se développent sur le Web et permettent de personnaliser facilement une page Web. NetVibes⁹⁵ et IGoogle⁹⁶ sont les deux principaux sites permettant

⁹⁵NetVibes, « Leading personal start page to manage your digital life » : www.netvibes.com

⁹⁶IGoogle : www.google.fr/ig



FIG. 7.18 – Visualisation du signet commun et des versions individuelles pour l’URL `http://www.example.com/devoir.htm`

à des internautes de créer leurs pages personnelles en accumulant différents widgets, plus ou moins interactifs. Les widgets peuvent par exemple afficher la météo, les dernières nouvelles issues d’un journal en ligne ou être de simples jeux.

L’explosion du développement de widgets semble indiquer que les widgets, services pour les internautes, vont aussi devenir de plus en plus des services pour les webmestres qui pourront créer des sites Webs en proposant différents services provenant d’autres sites.

Pour expérimenter cet aspect, nous avons donc créé un widget exportant les signets d’une application SemanticScuttle vers des pages NetVibes ou IGoogle (Fig. 7.19). Le widget se configure en donnant l’adresse du site employant SemanticScuttle et son nom. En ajoutant un membre et des tags, l’utilisateur du widget peut restreindre l’affichage aux signets d’un certain membre du site, ou aux signets décrits par certains tags, ou en effectuant une combinaison des deux (Fig. 7.20).

7.2.4.4 Intégration du moteur de recherche Google Custom Search

La société Google propose un service nommé Google Custom Search⁹⁷ qui permet d’effectuer une recherche avec les technologies Google non pas sur tout le Web mais uniquement sur un ensemble de pages sélectionnées par les utilisateurs. Ce service est intégrable à d’autres sites Web à l’aide d’une API proposée par Google.

SemanticScuttle exploite cette API et offre ainsi un autre mode de recherche sur les signets. En effet, lorsqu’un utilisateur emploie le moteur de recherche interne à SemanticScuttle pour chercher par exemple « français », SemanticScuttle recherche la chaîne de caractère « français »

⁹⁷Google Custom Search : <http://www.google.com/coop/cse/>



FIG. 7.19 – Sur NetVibes, les utilisateurs peuvent configurer leurs pages Web avec des « widgets ». Le widget de gauche affiche les signets d'un site employant SemanticScuttle.

dans les titres des signets, leurs descriptions libres et les tags associés, c'est à dire dans les métadonnées ajoutées par l'auteur du signet. En effectuant la recherche sur le moteur Google Custom Search, celui cherchera la chaîne « français » (ainsi que ses possibles dérivés syntaxiques comme « française ») dans le contenu même des pages (et non dans leurs métadonnées).

Ce mécanisme permet d'appliquer la puissance d'un moteur généraliste aux ressources plébiscitées par les utilisateurs.

Les autres fonctionnalités de SemanticScuttle sont décrites en annexe J.

7.3 Terrain d'application

Changement de terrain Dans la partie précédente, nous évoquions la difficulté à travailler avec une communauté de professeurs en génie mécanique. Par conséquent, nous avons cherché un nouveau terrain d'expérimentation. Nous l'avons trouvé chez les documentalistes de l'université Paris V. Celles-ci⁹⁸ ne partagent pas des documents pédagogiques mais des sites Web portant sur le domaine de la sociologie. Néanmoins, la structure de la communauté et la manière de partager nous permettent de voir des ressemblances entre les deux terrains. Ces points communs et ces différences sont discutés en 7.5.1.1.

⁹⁸Nous emploierons le féminin pour désigner le milieu des documentalistes en raison de la forte proportion de la gente féminine en ses rangs et en espérant ne froisser personne.



FIG. 7.20 – Configuration d'un widget pour SemanticScuttle. Il est possible de définir les signets à afficher e.g. selon l'auteur (« user ») ou/et les tags attachés.

Présentation du nouveau contexte Les documentalistes de Paris V maintiennent depuis 1998 un site Web répertoriant des liens vers des ressources portant sur le domaine de la sociologie. Ce site s'appelle le « Sociopôle » (§7.3.1) et était principalement maintenu par une seule personne. En 2008, la décision fut prise d'ouvrir la gestion des liens à plusieurs collaboratrices et d'en profiter pour moderniser le site Web qui n'était qu'un ensemble de pages Web statiques. Les documentalistes essayèrent alors un outil bibliographique en ligne nommé Wikindx (§7.3.2), mais celui-ci ne rencontra pas l'adhésion des premières collaboratrices à cause d'une inadéquation au contexte du partage de signets (§7.3.2.2). C'est pourquoi finalement, les responsables du Sociopôle ont été intéressées par nos travaux, et pour collaborer autour d'un outil plus léger et mieux adapté à leurs besoins.

7.3.1 Sociopôle, des signets en sociologie

Les documentalistes sont responsables d'un site nommé le Sociopôle. Nous parlerons de Sociopôle original pour parler de la première version du site, que notre contribution viendra remplacer plus tard.

Objectif du Sociopôle. « L'objectif de Sociopôle est de rassembler les ressources électroniques pertinentes en sociologie pour la communauté des sociologues internautes, chercheurs, enseignants et doctorants, tant au niveau national qu'international. Notre approche ne vise pas à l'exhaustivité mais plutôt à une sélection qualitative des contenus présents sur la Toile, en tenant compte de la diversité des pensées sociologiques. »⁹⁹

Responsables et contexte. De 1998 à 2007, une seule personne en lien avec l'IRESO¹⁰⁰ développa le Sociopôle.

En 2008, le Sociopôle reçut un financement visant à développer une activité collaborative au sein d'un réseau thématique de professionnels de l'information. Les documentalistes décidèrent donc d'essayer différents outils collaboratifs pour gérer les contenus du Sociopôle.

Organisation du Sociopôle original. Chaque ressource du Sociopôle original était composée d'une URL, d'un titre, d'un éditeur et d'une description textuelle.

⁹⁹Site du Sociopôle : <http://www.sociopole.cnrs.fr/>

¹⁰⁰Institut de Recherches sur les Sociétés Contemporaines : www.iresco.fr/

Le fond pouvait être exploré grâce à un moteur de recherche (l'outil Ht ://Dig¹⁰¹) qui indexait l'ensemble des pages. Néanmoins l'organisation du site favorisait le parcours selon des catégories hiérarchiques réalisées par la responsable originale du site. Voici quelques détails sur l'organisation originale des 1900 ressources afin de mieux comprendre la vision de départ des documentalistes.

Deux sections principales mettaient en avant la localisation des pages Web répertoriées :

Sociopôle France : Ressources situées en France

Sociopôle francophone : Ressources francophones, situées dans le monde entier

La section francophone proposait une page listant les zones géographiques et les pays qu'elles contiennent :

[Sociopôle francophone]

Orientations

AFRIQUE DU NORD Algérie Maroc Mauritanie Tunisie	AFRIQUE SUBSAHARIENNE Bénin Burkina Faso Cameroun Côte d'Ivoire Mali Niger Sénégal Togo
AMÉRIQUE DU NORD Canada	ASIE SUD-EST ET OCÉAN INDIEN Cambodge Vietnam Comores La Réunion Madagascar
EUROPE Allemagne Belgique France Italie Luxembourg République Tchèque Russie Suisse	PROCHE-ORIENT Égypte Liban Turquie Yémen
Francophonie	

FIG. 7.21 – Page redirigeant vers les zones et pays francophones

Pour chaque pays, les ressources sont listées selon des catégories telles que Écoles doctorales, Unités de recherche, Revues scientifiques, Associations...

Par exemple, la Fig. 7.22 montre les ressources issues d'organismes au Viêt-Nam et trois catégories employées :

Participation. La Fig. 7.21 montre un lien nommée « Orientations » ouvrant une fenêtre pop-up contenant des explications sur l'organisation du fond. Il y est notamment demandé de participer en écrivant un mail à la responsable du site. Par ce biais, la documentaliste travaillant

¹⁰¹Moteur d'indexation et de recherche pour les petits domaines : www.htdig.org/

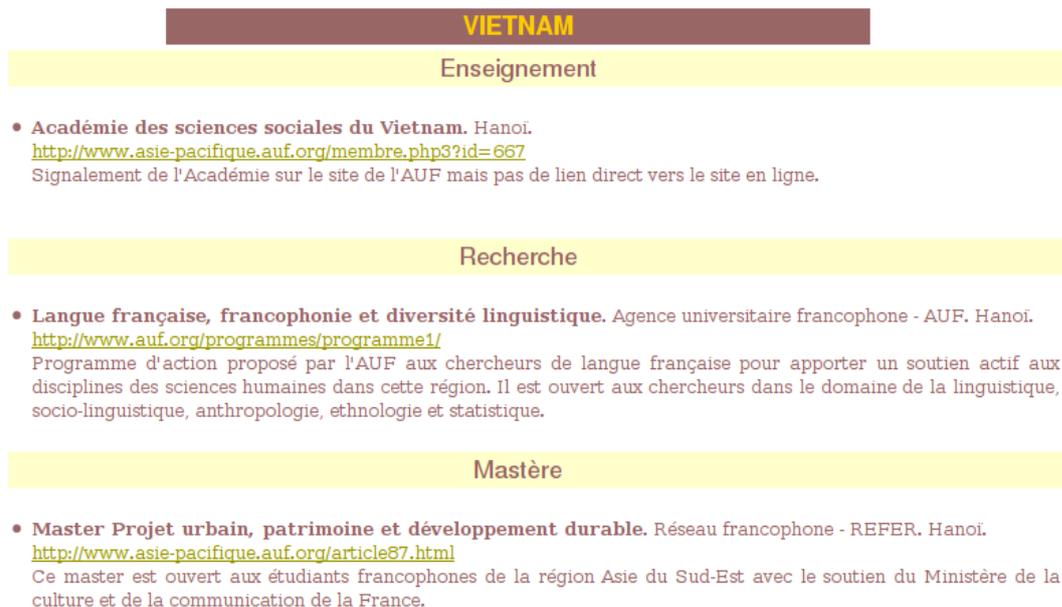


FIG. 7.22 – Ressources situées au Viêt-Nam selon trois catégories

sur le Sociopôle a reçu des demandes de modification ou d'ajout, généralement de chercheurs souhaitant corriger les informations les concernant.

Conclusion. Les documentalistes possédaient dans le Sociopôle original aux environs de 2000 ressources mais sans réels mécanismes de gestion. La modernisation du Sociopôle avait pour but de faciliter cette gestion grâce à :

- **la collaboration entre documentalistes** : i.e. la gestion des ressources par une dizaine de documentalistes ;
- **la participation des visiteurs** : i.e. que les visiteurs puissent facilement proposer des contenus ou des descriptions.

Néanmoins, le nouvel outil devait respecter les choix éditoriaux du Sociopôle original à savoir :

- **la qualité des ressources** : i.e. la mise en avant des contenus de qualité plutôt que le choix de l'exhaustivité ;
- **un guidage parmi les ressources** : i.e. la possibilité d'une navigation selon les catégories et les zones géographiques définies par les documentalistes.

7.3.2 WIKINDX, tentative d'utilisation d'un outil d'indexation bibliographique

Pour moderniser le Sociopôle, les documentalistes expérimentèrent tout d'abord un outil d'indexation bibliographique WIKINDX. Voici une présentation de l'outil, qui illustre les pratiques traditionnelles d'indexation des documentalistes.

web_site: Test site

Authors

Firstname Initials (e.g. M.N. or Th.V.) Prefix Last name A selection from the pre-existing names will be overridden by a surname in the input boxes

 *

URL: *

(http://)

Date accessed

Day: * Month: * Year: * (e.g. 1999)

Publication year:

Year of original publication

Abstract:

Note:

DOI:

ID no. (ISBN etc.):

Peer reviewed:

Categories

Each resource must belong to at least one category:

General Arts Computing Demographics Environment Film Music/Sound

(Multiples can be chosen)

Keywords:

 3-D image registration 3-DIARY/FLUORENE'S ... ANCEMENT; OLIGOMER Acoustic sound Acoustics Action Potentials; ... ognition; Fixation

(Multiples can be chosen)

You can select keywords relating to the resource from the list and/or enter them as a comma-separated list

(comma-separated list)

Add this resource to user bibliographies:

 My Bibliography Lectures d'Alain Sem antic Web References Zwischen Begehren ... chung - zur 222 Just a test

(Multiples can be chosen)

FIG. 7.23 – Troisième page permettant d'ajouter une page Web dans WIKINDX

7.3.2.1 Présentation de WIKINDX

« WIKINDX est un système sous licence libre de gestion de bibliographie et de citations/-notes et d'écriture d'articles. Il est conçu pour une utilisation par un seul utilisateur ou pour un usage collaboratif sur Internet. »¹⁰²

WIKINDX est un outil très complet. Il permet de gérer 39 types de ressources différentes, des auteurs, des éditeurs, des collections, des années de publications, des catégories, des mots-clés et des tags. WIKINDX est ainsi utilisé dans plusieurs universités à travers le monde.

En 2008, tous les contenus du Sociopôle furent intégrés dans WIKINDX. Puis l'outil fut testé par les documentalistes du groupe de travail. Or le test ne fut pas concluant car ces documentalistes n'apprécièrent pas les interfaces pour des raisons de complexité et de manque de lisibilité des données.

La Fig. 7.23 illustre cette complexité. C'est la troisième page de champs à remplir pour ajouter une ressource sur le Web dans WIKINDX. Le nombre de pages à parcourir et de champs à remplir fut une des sources principales de réticence des documentalistes à employer cet outil.

¹⁰²<http://wikindx.sourceforge.net/>

7.3.2.2 Limites de WINKINDX

A partir des remarques des documentalistes et de l'analyse ergonomique de WIKINDX, nous concluons les points suivants :

- Si l'exhaustivité de WIKINDX est peut-être pertinente pour des livres et des articles de recherche, elle semble moins adéquate pour des pages Webs qui sont des ressources beaucoup moins stables et de plus petites granularités.
- Si la navigation très complète (par catégories, mots-clefs, champs éditeurs, collection, etc) est peut-être pertinente dans certains contextes, cette diversité semble être une source possible de complexité pour les utilisateurs et les administrateurs.

En conclusion, WIKINDX est un outil trop strict et à l'emploi trop fastidieux pour ajouter des signets.

Vers un outil plus adapté au partage de signets Suite aux difficultés à faire adopter WIKINDX, les responsables du Sociopôle se sont tournées vers un outil plus léger et plus adapté au partage de signets.

7.4 Expérimentations

Expérimentations réelles. Si les projets LUISA (chapitre 3) et Pépi (chapitre 5) rencontrèrent de nombreuses difficultés dans la phase d'expérimentation, le projet SemanticScuttle permit un niveau beaucoup plus correct d'expérimentation, notamment grâce à la motivation des documentalistes et grâce à la stabilité et diffusion de l'outil. SemanticScuttle et le modèle purent ainsi être améliorés de manière régulière sur plusieurs mois.

7.4.1 Méthodologie des expérimentations

7.4.1.1 Objectifs généraux des expérimentations

Les expérimentations ont deux buts :

Améliorer l'outil : Améliorer l'outil en corrigeant les bugs, les défauts ergonomiques et en ajoutant les fonctionnalités nécessaires aux utilisateurs et non prévues au début.

Améliorer le modèle sous-jacent : Observer l'usage des fonctionnalités nouvelles issues du modèle initial (§7.1.2), vérifier ou réfuter leur validité et en tirer des conclusions pour améliorer le modèle sous-jacent à l'outil.

Dans le contexte d'une application participative, ces deux aspects sont directement corrélés sachant qu'en améliorant l'outil, la participation des utilisateurs augmente et que celle-ci est nécessaire pour valider le modèle sous-jacent.

Dans le cadre d'un développement participatif, il est difficile de distinguer réellement phases de développement et phases d'expérimentation car très tôt l'outil fut mis en production puis adapté suite aux interactions avec les documentalistes. L'expérimentation prend donc souvent moins la forme d'étapes ponctuelles de tests que de régulières modifications en accord avec les utilisateurs. Cela signifie que développement, recueil de données et analyse furent souvent effectués en parallèle.

7.4.1.2 Rappel des questions de Recherche et des sources d'observations

Questions de recherche. L'objectif final des expérimentations est de vérifier à quel point le modèle IPM permet d'améliorer l'offre et la recherche de ressources dans une communauté.

- Pour cela, nous observerons à quel point SemanticScuttle permet aux utilisateurs de :
- facilement participer ?
 - progressivement structurer les ressources ?
 - progressivement collaborer autour de l'organisation des ressources ?
 - **finale**ment mieux « profiter » des ressources ?

7.4.1.3 Plan des expérimentations

Pour obtenir un maximum de retours, nous avons cherché à expérimenter l'outil dans différents contextes. Voici un récapitulatif chronologique :

Sociopôle bêta : Installation de SemanticScuttle en tant que Sociopôle version bêta. Travail avec le groupe de documentalistes responsables du site (novembre 2008-février 2009).

Sociopôle : Site du Sociopôle mis en production. Observation de l'activité des visiteurs du Sociopôle (avril 2009-juin 2009). Poursuite du travail avec le groupe de documentalistes.

Focus group : Expérimentation du Sociopôle avec un groupe d'utilisateurs potentiels (7 mai 2009).

Mechanical Turk : Expérimentation de SemanticScuttle avec une dizaine d'internautes (Mechanical Turk), et des ressources pédagogiques.

7.4.1.4 Expérimentation du Sociopôle avec le groupe de travail des documentalistes : utilisation par des experts

Entre novembre 2008 et janvier 2009, nous avons travaillé avec deux documentalistes responsables du Sociopôle pour migrer les contenus du site original et pour adapter SemanticScuttle à leurs besoins.

La migration fut effectuée à partir de la base de données WIKINDEX. Néanmoins, plus qu'un simple transfert de base à base, le travail a consisté à repenser avec les documentalistes l'organisation des signets pour passer d'une hiérarchie à une classification à l'aide de tags structurés.

L'adaptation de SemanticScuttle consista en l'ajout de fonctionnalités comme le rôle d'administrateur, les descriptions balisées ou les notes privées. Ces fonctionnalités sont aussi issues d'un travail de longues discussions pour ne pas perdre le principe de simplicité propre aux applications participatives.

Par exemple, les documentalistes voulaient ajouter plusieurs rôles pour administrer le site mais nous avons insisté pour n'en ajouter qu'un, au moins au début, pour que les utilisateurs possèdent des droits équivalents. Notre hypothèse est que des droits identiques favorisent le partage et la collaboration entre utilisateurs.

De même, alors qu'au début les documentalistes voulaient ajouter de nombreux champs supplémentaires, nous avons créé la notion de descriptions balisées pour éviter l'ajout de champs superflus et conserver la simplicité initiale de l'interface.

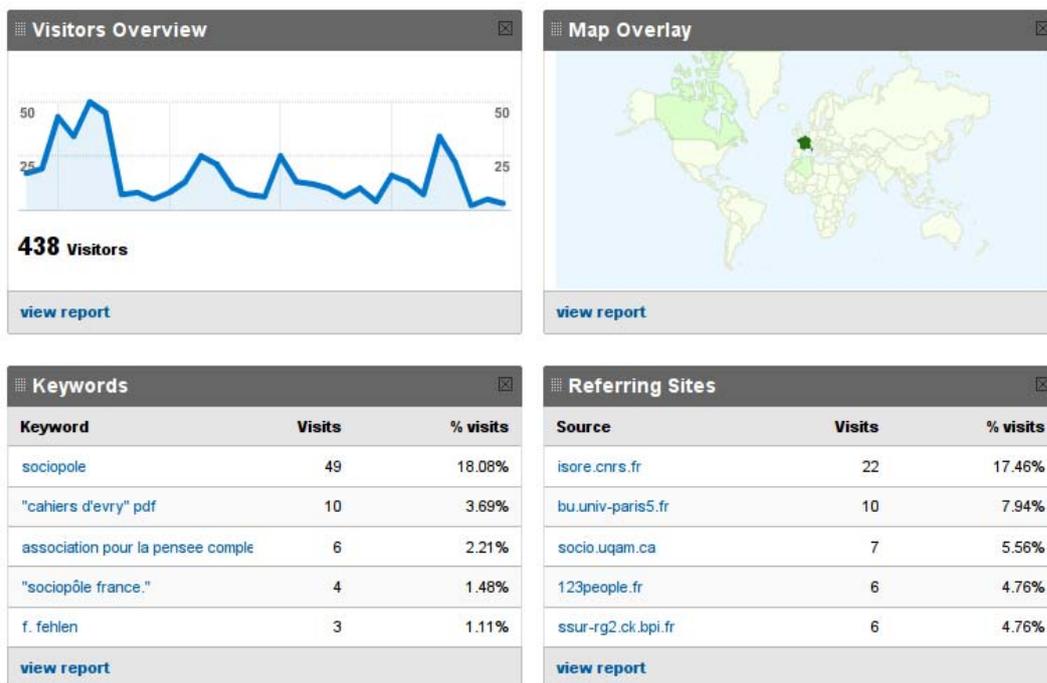


FIG. 7.24 – Ecran GoogleAnalytics affichant le résumé des statistiques du Sociopôle (évolution des visites, provenance géographique, mots-clefs cherchés, sites référents)

Entre février et mars 2009, le site du Sociopôle fut expérimenté par toutes les documentalistes du groupe de travail (8 personnes) pour ajouter, corriger des signets et de manière générale pour s'approprier l'outil.

Enfin à partir d'avril 2009, l'information concernant la nouvelle version du site fut diffusée.

En juin 2009, le site comportait 1950 signets publics pour 8 documentalistes du groupe de travail et 3 utilisateurs externes.

Pendant toutes ces périodes, nous avons recueilli les demandes et questions des documentalistes. Nous avons aussi observé l'usage du site à l'aide de l'outil Google Analytics.

Statistiques du Sociopôle avec Google Analytics. Le suivi de l'activité sur le site du Sociopôle fut effectué avec l'outil Google Analytics (Fig. 7.24). A l'aide d'un petit script exécuté lors de l'affichage de chaque page Web, Google Analytics permet d'obtenir des résultats extrêmement détaillés sur les usages d'un site Web. Les données peuvent être croisées entre autres selon le profil des visiteurs (ville de connexion, navigateur, langue...) ou des types de navigation (durée de visites, nombre de clics, pages vues, site précédemment visité...).

7.4.1.5 Expérimentation du Sociopôle avec un groupe d'utilisateurs potentiels : « focus group »

L'expression « Focus Group » peut couvrir différents sens, selon qu'elle est employée par des sociologues, des ergonomes ou des publicitaires¹⁰³.

Pour notre part, nous entendons par ce terme, une réunion avec un groupe d'utilisateurs aux profils variés autour d'un « produit » nouveau. Pour les concepteurs du produit, le but est de récupérer les premières impressions des utilisateurs sur le produit : ce qu'ils apprécient, ce qu'ils ne comprennent pas, ce qu'ils craignent, etc. En fonction de ces réactions, les concepteurs peuvent alors améliorer le produit en améliorant des fonctionnalités, en adaptant la présentation ou en modifiant le public visé par exemple.

Par rapport à une entrevue individuelle, l'expérimentation en groupe a pour objectif d'atteindre rapidement plusieurs publics et de faciliter la verbalisation des utilisateurs par des discussions entre eux. En contrepartie, le risque est que certains utilisateurs imposent leurs opinions aux autres. C'est à l'animateur de prévenir ce genre de comportements et de permettre à chacun de s'exprimer pour enrichir les échanges.

Participants. Le « focus group » a réuni 9 utilisateurs, contactés par les documentalistes du Sociopôle et sélectionnés pour la diversité des profils :

- 3 chercheurs en sciences humaines et sociales (SHS)
- 1 post-doctorante en SHS
- 2 doctorants en SHS
- 1 ingénieur pédagogique
- 2 conservateurs de bibliothèque SHS

La session de focus-group s'est déroulée sur deux heures : une heure d'activités individuelles sur machine en répondant à des questions sur le Sociopôle, puis une heure de discussion générale, reprenant les questions en ayant pour but de développer les réponses des utilisateurs.

Le protocole et les résultats détaillés du focus group se trouvent en annexe K.

7.4.1.6 Expérimentation de SemanticScuttle avec des utilisateurs sur Internet : Mechanical Turk

Cette expérimentation ne concerne pas le Sociopôle mais SemanticScuttle pour des ressources pédagogiques.

L'entreprise Amazon propose un service nommé « Mechanical Turk » permettant d'employer rapidement de nombreux utilisateurs pour des tâches très simples en ligne.

Le nom de « turc mécanique » provient d'un canular ayant été réalisé au XVII^e siècle et dans lequel un homme disait avoir inventé un automate capable de jouer aux échecs. Cet automate prenait la forme d'un joueur turc. L'automate parcourut l'Europe et remporta bien des victoires jusqu'au jour où l'on s'aperçut que l'automate contenait en réalité un joueur humain

¹⁰³Une introduction à la méthodologie des « focus group » : http://www.lergonome.org/so/u.htm#/index.php?option=com_content&task=view&id=27&Itemid=29

qui animait la machine de l'intérieur. Depuis, l'expression d'automate turc désigne un système où derrière une intelligence artificielle se cache en réalité une intelligence humaine.

Le service d'Amazon reprend ce concept en proposant une place de marché où des utilisateurs proposent des tâches très rapides et simples mais qu'ils doivent réaliser de grande quantité de fois. De par leur simplicité, ces tâches sont rémunérées une vingtaine de cents (\$) par exemple. D'un autre côté, d'autres utilisateurs, les « turcs », effectuent ces tâches. Des études¹⁰⁴ montrent que les « turcs » sont généralement des internautes aux États-Unis qui effectuent ces tâches pour se distraire. Ils utilisent les rémunérations pour s'acheter des produits sur le site Amazon.

L'existence de ce service, pas très éloigné de la notion de Web Participatif, nous semble un point important pour résoudre la question de l'expérimentation d'outil de partage comme SemanticScuttle. En effet, les « turcs » sont des utilisateurs participants mais qui passent le minimum de temps sur la tâche, se rapprochant selon nous d'un utilisateur en situation réelle. En revanche, les actions de nombreux « turcs » permettent d'évaluer des phénomènes de masse plus facilement qu'en laboratoire où les expérimentations sont généralement limitées.

Dans ce contexte, nous avons proposé dix tâches où des utilisateurs anglophones devaient chercher dix ressources pédagogiques sur l'informatique et les ajouter à une plateforme SemanticScuttle. Cette expérimentation avait pour but :

- d'explorer les possibilités des services « Mechanical Turk » pour étudier les outils de partage de masse ;
- d'observer les structures de tags émergeant pour 100 ressources ajoutées.

L'installation de SemanticScuttle expérimentée avec des signets pédagogiques par des membres du Mechanical Turk est visible sur :
<http://festic.fr/semanticscuttle/demo/mturk/populartags>

Suite à l'expérimentation, il y eut 32 utilisateurs inscrits, sur lesquels 16 ont ajouté 122 signets, 668 tags et 81 relations entre tags. La différence entre le nombre de participants et les dix utilisateurs requis réside du fait que certains participants ont abandonné la tâche en cours d'expérience, après n'avoir ajouté que quelques ressources.

L'énoncé de la tâche se trouvent en annexe L.

7.4.2 Résultats : Facilité d'ajout d'un signet

Le défi de SemanticScuttle est de permettre la description de ressources au plus grand nombre. Pour cela, ajouter un signet doit être le plus simple possible.

Résultat : L'ajout de signets est « plutôt simple ».

Source : focus group

¹⁰⁴« Mechanical Turk : the demographics » : <http://behind-the-enemy-lines.blogspot.com/2008/03/mechanical-turk-demographics.html>

Sur 9 utilisateurs du focus group, 6 ont trouvé l'ajout d'un signet simple ou normal, et 3 l'ont trouvé plutôt difficile. Parmi ces 3 réponses, il y avait une participante peu habituée aux outils informatiques dont la réponse ne surprend pas. Il y avait aussi un utilisateur avancé qui a réussi à ajouter des signets mais qui expliqua sa réponse dans le sens où il aurait aimé des fonctionnalités très avancées comme un glisser-déposer (drag'n drop) à partir de la barre d'adresse du navigateur.

Nous concluons que l'ajout a posé peu de problèmes aux utilisateurs moyens qui ont réussi la tâche. Ce résultat est confirmé par le fait que la majorité de ces utilisateurs moyens a déclaré ne jamais avoir ajouté de contenus sur un site Participatif. Ils n'étaient donc pas particulièrement habitués à cette tâche et ont cependant réussi à créer un compte et à l'effectuer sans difficultés.

7.4.3 Résultats : Continuum de fonctionnalités structurantes

Le continuum de fonctionnalités structurantes se compose :

- des tags « classiques » sur les ressources ;
- des tags structurés ;
- des descriptions balisées ;
- de l'export sémantique.

7.4.3.1 Tags classiques

Résultat : Un usage classique des tags « classiques ».

Sources : Utilisateurs du Mechanical Turk, et Sociopôle

L'interface des tags « classiques » de SemanticScuttle reprend le mécanisme de Scuttle qui copiait lui-même le mécanisme éprouvé du site Delicious.

Mechanical Turk. Pour 100 ressources pédagogiques, le nuage de tags créé par les utilisateurs du Mechanical Turk correspond tout à fait au domaine au vu tags les plus fréquents : « computer science », « tool » et « tutoring » (Fig. 7.25).

Sociopôle : nécessité de tags spontanés. Les documentalistes avec lesquelles nous avons collaboré avaient l'habitude de suivre des lignes de conduite et des vocabulaires contrôlés lors de l'indexation de documents.

L'objectif de l'emploi de SemanticScuttle était de les ouvrir à une indexation plus spontanée. Alors qu'au début le vocabulaire employé fut celui des catégories du Sociopôle original, les documentalistes ont rapidement eu besoin d'ajouter des termes non prévus. Néanmoins ce mouvement spontané fut stoppé par la crainte de mélanger ces tags non contrôlés avec ceux négociés auparavant. Nous leur avons alors proposé de préfixer leurs tags non négociés d'un caractère particulier.

Deux documentalistes ont alors employé le « _ » et ajouté des termes comme « _anthropologie sociale, _art extra-européen, _économie sociale, _féminisme ». L'intérêt de ces tags est de décrire des thématiques invisibles dans les catégories initiales du Sociopôle. Pour l'instant, ces tags ne décrivent qu'un signet chacun. Néanmoins plus tard, si leur usage se répand, les documentalistes



FIG. 7.25 – Derniers tags ajoutés par les utilisateurs du Mechanical Turk.

pourront éventuellement lors de réunions discuter ces tags en tant que propositions pour qu'ils deviennent des tags au sens négociés entre elles.

7.4.3.2 Ajout de tags structurés

Les tags structurés sont une des principales originalités de SemanticScuttle.

Ils peuvent être ajoutés de deux manières : « à la volée » au moment de l'ajout d'un signet, ou plus tard, par une page de gestion des tags structurés où l'utilisateur peut créer ou effacer une à une les relations entre tags.

Usage réfléchi des tags structurés. L'observation de l'activité des documentalistes a révélé que la création de liens entre tags ne s'est pas effectuée de manière spontanée.

Pour les tags classiques, les documentalistes faisaient référence au vocabulaire du Sociopôle original. De même, pour les relations entre tags, nous les avons vu discuter lors des réunions pour savoir quelles relations ajouter et quelles nouveautés les nouvelles relations pouvaient apporter. Ce rapport aux tags est loin de celui que nous avons prévu, c'est à dire de tags ajoutés rapidement et de manière individuelle, puis agrégés ou discutés. En conservant la vision historique de leur discipline, les documentalistes discutent avant de créer le tag ou la relation.

En observant l'usage de SemanticScuttle sur des initiatives indépendantes de nos incitations, nous remarquons fréquemment un usage mono-utilisateur. C'est-à-dire un utilisateur qui installe une plateforme SemanticScuttle pour y organiser majoritairement ses propres signets. Par

exemple sur le site *Projet Civ*¹⁰⁵, un enseignant emploie *SemanticScuttle* pour ses élèves et incite ses collègues à participer. Il reste cependant le contributeur majoritaire et a organisé de manière réfléchie ses tags structurés qui deviennent véritablement un menu proposé par l'administrateur et non pas une structuration par tous.

Cet usage fait aussi que les tags structurés modifient l'emploi des tags classiques. Ainsi, nous remarquons sur l'annuaire de *Projet Civ* que l'auteur ne tagge parfois ses ressources qu'avec un unique tag faisant partie d'une arborescence de tags structurés. Cet usage est différent des tags classiques dont trois ou quatre sont généralement employés pour chaque ressource.

Usage spontané possible. L'expérimentation menée avec les utilisateurs du *Mechanical Turk* illustre toutefois la possibilité d'un usage spontané. Sur les dix utilisateurs, tous ont réussi à créer des tags structurés. Pour rappel, leur niveau de compétences en informatique déclaré était élevé pour trois, moyen pour six et faible pour un.

Nous avons juste remarqué une erreur de l'un des participants ayant pour quelques ressources mis une URL dans les tags structurés. Par exemple, il a écrit dans le champ des tags « `http://www.dataminingresources.info > data mining` » comme s'il voulait non pas structurer des tags entre eux mais des tags et des sites.

Compréhension des notions de « tag » et « tag structuré » De manière plus large, cette erreur renvoie à une difficulté rencontrée aussi avec les documentalistes ou lors du focus group sur le sens de « tag ». En français, certains sites emploient la traduction de « label » ou « d'étiquette ». Pour notre part, nous avons expliqué un tag en parlant de « mot-clefs ». Néanmoins, la notion de « tags structurés » ajoute une couche de complexité aux tags qui en font un nouveau concept, demandant à être expliqué aux utilisateurs. Le principal avantage est qu'un utilisateur n'est pas obligé d'employer de tags structurés, s'il ne comprend pas leur usage.

7.4.3.3 Ajout de descriptions balisées

Les descriptions balisées ont rapidement été employées par les documentalistes pour ajouter des « auteurs » ou des « éditeurs ». Il est à noter que toutes les documentalistes n'emploient pas les raccourcis proposés. Certaines écrivent les balises à la main et ceci provoque des différences dans l'écriture des noms des balises. Certaines écrivent « [éditeur] » d'autres « [editeur] » au lieu d'« [Editeur] ». Néanmoins ces variations semblent mineures et ne pas empêcher par la suite un traitement automatique pour récupérer le contenu des différentes balises.

Usage en tant que balise intégrée au contexte. Une documentaliste ne faisant pas partie du groupe de travail a ajouté un signet sur le *Sociopôle* avec un usage original des descriptions balisées. En effet, alors que nous envisagions une balise comme devant apparaître de manière extérieure à la description, cette documentaliste a inclus la balise à l'intérieur de manière naturelle, comme un hyperlien HTML. C'est-à-dire qu'au lieu d'écrire la description suivante :

Cet annuaire est une publication annuelle qui regroupe principalement les comptes rendus des conférences des enseignants-chercheurs. . .
[Editeur]École pratique des hautes études[Editeur]

¹⁰⁵ Annuaire du site *Projet Civ* : <http://www.projet-civ.com/annuaire/>

Elle a écrit :

Cet annuaire de l'[Editeur]École pratique des hautes études[Editeur] est une publication annuelle qui regroupe principalement les comptes rendus des conférences des enseignants-chercheurs...

Comme expliqué dans la présentation de la fonctionnalité, les balises sont traitées pour apparaître sous la forme "balise » contenu de la balise". La description de la documentaliste apparaît donc de manière correcte à l'affichage du signet :

Cet annuaire de l'Editeur » École pratique des hautes études est une publication annuelle qui regroupe principalement les comptes rendus des conférences des enseignants-chercheurs...

Cette documentaliste a employé la balise d'une manière proche des conseils du W3C incitant à ajouter du code RDF dans les pages Web classiques (plus précisément le code XHTML) pour faciliter leur traitement sémantique ultérieur¹⁰⁶. L'exemple employé par le W3C est d'écrire :

Le contenu de ce site est sous licence Creative Commons .

où le terme « license » est un mot-clef repérable par les machines.

L'usage spontané de cette documentaliste est un indice de l'utilité des descriptions balisées. Cependant, d'autres utilisations et d'autres observations seront nécessaires pour découvrir le potentiel de ces balises incluses dans les descriptions.

7.4.3.4 Export sémantique (SIOC)

SemanticScuttle permet d'exporter les données (utilisateurs, signets, tags structurés) sous un format sémantique conforme aux recommandations du W3C. Toutefois deux points viennent limiter ces résultats.

Pas de besoins exprimés chez les utilisateurs. Les données sémantiques ne servent à rien sans applications pour les exploiter, et les applications sémantiques ont peu d'intérêts sans les données sémantiques. Dans notre contexte, même si nous avons fourni la possibilité d'un export sémantique, les utilisateurs n'ont actuellement pas d'outil leur permettant de les exploiter.

Par exemple, les documentalistes nous ont plutôt demandé un export au format OAI-PMH (voir les fonctionnalités prévues en annexe J) commun aux bibliothèques françaises.

Cependant, la diffusion croissante des outils et sites exploitant les formats sémantiques du W3C, e.g. DBpedia¹⁰⁷, montre que ces interfaces seront probablement de plus en plus exploités par des outils de recherche globaux.

D'une ontologie légère à des ontologies plus lourdes. En exportant au format SIOC, SemanticScuttle propose une ontologie très légère où les tags sont transformés en des termes inter-reliés avec des relations basiques (inclusion ou synonymie). Les raisonnements qui peuvent s'appliquer dessus restent donc aussi basiques.

¹⁰⁶<http://www.w3.org/TR/xhtml1-rdfa-primer/>

¹⁰⁷DBpedia reprend sous un format sémantique les données de Wikipédia : dbpedia.org

Au contraire, les ontologies de LUISA emploient des relations plus contextuelles comme la relation d'interopérabilité entre un logiciel et un système d'exploitation, permettant des raisonnements plus spécifiques comme « fournir tel document à cet utilisateur si le document utilise un logiciel interopérable avec le système d'exploitation de l'utilisateur ».

SemanticScuttle soutient ainsi la création d'une sorte de thésaurus informatisé au sein d'une communauté. Pour passer de ce thésaurus à une ontologie permettant des raisonnements sémantiques plus sophistiqués, il faudrait pouvoir relier les tags structurés de SemanticScuttle à des concepts extérieurs à la plateforme.

Les travaux d'Alexandre Passant sur MOAT ([Passant and Laublet, 2008]) explorent par exemple ce passage. Cependant, compte-tenu de la difficulté à obtenir une vue partagée dans un petit groupe d'utilisateurs, il nous semble pour l'instant prématuré de chercher à relier cette vue à des visions plus élargies.

7.4.4 Résultats : Continuum de fonctionnalités collaboratives

Le continuum de fonctionnalités collaboratives est composé :

- de la page des contacts (usage collectif)
- des notes privées (usage collectif)
- des descriptions communes des signets (usage collaboratif)
- des descriptions communes de tags (usage collaboratif)

7.4.4.1 Page des contacts

La page des contacts permet d'ajouter des utilisateurs à ses contacts et de voir sur la page leurs derniers ajouts de signets.

Techniquement, ajouter des contacts ne semble pas trop difficile et deux documentalistes du groupe ont effectué cette opération.

Lors des réunions, les documentalistes ont trouvé utile cette page de contacts pour voir les travaux des autres et éventuellement travailler à plusieurs. En effet, l'organisation initiale des documentalistes consistait à se répartir l'indexation des signets selon des catégories. Néanmoins, certaines catégories aux limites floues demandent de travailler à plusieurs sur les mêmes sites et donc de surveiller les ajouts des collègues.

7.4.4.2 Notes privées

Les notes privées permettent de commenter un signet uniquement pour soi-même et ses contacts.

Cette fonctionnalité fut réclamée par les documentalistes pour soutenir leur travail d'indexation professionnelle. Lors de l'ajout des signets à partir de l'ancien outil (WIKINDEX), ce champ fut automatiquement rempli avec l'ancien champ « note ».

Les documentalistes du groupe de travail ont employé cette fonctionnalité qui ne semble pas poser de difficulté. Le contenu des notes permet notamment de mémoriser des questions sur les signets. Ces questions semblent s'adresser soit à tous les utilisateurs :

Dans quelle(s) catégorie(s) mettre ce signet ? Quels autres tags ?

Est-ce que ce genre d'information concerne le sociopôle ? En tout cas, il est sûr qu'elle intéresse les 3e cycles. . .

soit viser un utilisateur précis :

Sandra, que penses-tu de ce signet pour le Canada ? Cécile

Ainsi les notes semblent bien répondre à une communication entre utilisateurs, au moins experts sur SemanticScuttle. Nous avons des données insuffisantes pour savoir si des utilisateurs moins impliqués emploieraient cette fonctionnalité avec leur réseau de contacts.

7.4.4.3 Description commune des signets

SemanticScuttle permet la description libre d'un signet. Cette description appartient à un unique auteur. Néanmoins, nous avons ajouté une fonctionnalité permettant à plusieurs utilisateurs de travailler sur la même description d'un signet (titre et description) sous la forme d'un espace commun.

Nous n'avons pas observé de description commune des signets. Lorsque nous avons présenté la fonctionnalité aux documentalistes, elles n'ont pas marqué d'intérêts. En revanche, elles ont exprimé plutôt le besoin de discuter autour des signets.

Par conséquent, dans une prochaine version, nous pensons transformer éventuellement l'espace de description commun (style wiki) en un espace de discussion (style forum), à l'image de la page de discussion associée à chaque article Wikipédia.

7.4.4.4 Descriptions communes/individuelles de tags

Dans SemanticScuttle, les utilisateurs peuvent ajouter des descriptions sur les tags. Quand un visiteur affichera les signets associés à ce tag, la définition apparaîtra au-dessus des signets.

Les documentalistes ont demandé à ce que l'espace commun ne soit modifiable que par les administrateurs du site, ainsi avons nous ajouté cette option.

Durant nos observations, nous avons remarqué que cet espace était très utile à la collaboration entre documentalistes. En effet, celles-ci travaillent avec des tags qu'elles considèrent comme des tags conventionnels qui ne doivent être employés que dans un seul sens. Or pour appréhender ce sens, les définitions leur permettent de se mettre d'accord sur ce sens. La Fig. 7.26 illustre la description du tag « annuaires et signets ».

En revanche, il n'y a eu aucune utilisation des descriptions individuelles des tags. Nous n'avons jamais insisté sur cette fonctionnalité auprès des documentalistes et elles ne semblent pas en avoir eu besoin. Ceci tend à montrer que les descriptions de tags sont particulièrement utiles dans un cadre collaboratif plutôt que pour une participation solitaire au site.

7.4.5 Résultats : Mieux « profiter » des ressources

Au final, en tant qu'outil de gestion de signets, SemanticScuttle a pour but :

- d'améliorer la recherche de signets ;

tags: annuaires et signets*

Cette catégorie recense des sites et pages web d'annuaires et de liens/signets en sociologie et sciences sociales. S'y ajoutent des sites que l'on peut qualifier d'ouverts au dialogue citoyen. Sites d'établissements publics ou d'universités ou encore d'associations, ces signets ont pour objet d'ouvrir sur l'environnement de la sociologie.

38 signet(s) - Classer par :

 **Rezo.net : le portail des copains** ☆
 Développé en 1999, le portail évolue vers une présentation d'informations extraites de manière automatique de diverses sources journalistiques. Ces informations sont classées par grands thèmes. Parmi les thèmes suivis, le travail, la précarité, le racisme, les services

FIG. 7.26 – Description du tag principal « annuaires et signets », affiché au dessus des signets

– d'améliorer l'offre de signets.

Malgré les expérimentations, nous n'avons que des indices sur les possibilités de SemanticScuttle et donc sur le modèle IPM.

7.4.5.1 Recherche simple et avancée

Lors du focus group, une activité (Annexe K) demandait aux utilisateurs de trouver certaines ressources parmi les signets. Ceci pouvait être effectué en navigant parmi les signets ou en employant le moteur de recherche interne.

Par la suite, dans le questionnaire, 5 utilisateurs ont décrit la recherche comme « simple » ou « normal » contre 3 qui l'ont trouvé « plutôt difficile ».

Mais, lors de la discussion, il a pu ressortir une distinction entre recherche simple et recherche avancée. En effet, parmi les 3 utilisateurs ayant trouvé la recherche « plutôt difficile » se situent 2 utilisateurs aux compétences élevées en informatique. Ils ont précisé alors la difficulté à effectuer des recherches avancées comme croiser les signets décrits par plusieurs tags ou effectuer des recherches mêlant à la fois tags et moteur interne.

Nous concluons de ces remarques que les recherches simples sont accessibles à quasiment tous les utilisateurs mais que les recherches avancées doivent être améliorées. Dans un premier temps ceci peut être effectué en améliorant l'ergonomie de l'application car des fonctionnalités de recherches avancées n'ont pas été trouvées par les utilisateurs, bien qu'elles existent.

7.4.5.2 Améliorer l'offre

L'offre de ressources est améliorée dans le sens où le travail des responsables du site est enrichi des apports d'utilisateurs moins réguliers.

Jusqu'à maintenant, une seule ressource fut ajoutée spontanément par un utilisateur externe au groupe de travail. Ce signet, ajouté par une documentaliste, est pertinent par rapport au Sociopôle. L'analyse des trois tags ajoutés montrent que deux reprennent les tags structurés des

documentalistes tandis que le troisième est un tag nouveau dans le Sociopôle qui concerne un thème de Recherche. Cet ajout spontané ne démontre rien statistiquement mais il est un premier indice sur le potentiel participatif permis par l'outil.

7.4.5.3 Identification des ressources certifiées

Lorsqu'un administrateur ajoute un signet, celui-ci est automatiquement marqué sur le site par un bandeau jaune et une étoile, afin de le distinguer des contenus ajoutés par les simples utilisateurs. Ces indices visuels étaient expliqués sur la page principale et par une infobulle apparaissant lorsque le pointeur survolait l'étoile.

Lors du focus group, 7 utilisateurs sur 9 ont découvert et retenu la signification de ces indices visuels. Nous fûmes surpris car nous pensions que ces indices ne seraient pas remarqués ou alors qu'ils ne seraient pas compris sachant que tous les signets étaient certifiés au moment du focus group.

Lorsque SemanticScuttle est configuré en mode « avec administrateurs », il est intéressant de trouver un équilibre entre la participation de tous et celles des administrateurs. En effet, si les ressources des administrateurs sont seules mises en avant alors ceci peut décourager les utilisateurs simples à participer en particulier si leurs contributions sont dissimulées.

La bonne identification des ressources certifiées tend à montrer que SemanticScuttle peut mêler ressources certifiées et non-certifiées sans trop désorienter les utilisateurs.

7.4.6 Bilan des expérimentations

Les expérimentations ne nous permettent pas de tirer de conclusions définitives sur les usages de SemanticScuttle, ni sur la validité du modèle IPM. Néanmoins, les expérimentations nous ont permis de recueillir de nombreux indices.

Les résultats concernent deux publics différents :

- public expert et impliqué des documentalistes ;
- public novice des simples utilisateurs.

7.4.6.1 Public expert

Bénéfice d'un continuum de fonctionnalités. Le modèle IPM incite aux développements de fonctionnalités plus ou moins structurantes, à l'image des tags classiques et structurés. Or ces différents choix de structuration semblent être un atout important dans l'appropriation de l'outil.

Par exemple, d'un côté, l'organisation générale du Sociopôle original a pu être intégré dans la version sous SemanticScuttle en conservant la hiérarchie des catégories. Nous avons vu les documentalistes reproduire les mêmes procédures de négociation avec les relations entre tags, qu'elles employaient pour l'indexation classique. Pourtant, d'un autre côté, les documentalistes ont pu aussi explorer l'usage des tags de manière spontanée; c'est ce qui s'est produit avec l'emploi du préfixe « _ ». Le continuum de fonctionnalités semble faciliter ainsi le changement des pratiques des documentalistes.

Diffusion naturelle de l'outil. L'outil SemanticScuttle a plu aux documentalistes qui y ont vu un outil léger et flexible d'indexation de signets. Cet intérêt est indiqué notamment par le poster sur SemanticScuttle¹⁰⁸ qu'elles ont réalisé puis présenté lors de rencontres de professionnels de l'information. Cet intérêt est aussi indiqué par les projets qu'elles prévoient, pour installer l'outil dans d'autres domaines que la sociologie.

7.4.6.2 Public simple

Utilisabilité des fonctionnalités Participatives. Le focus group ou l'expérimentation avec les utilisateurs du Mechanical Turk montrent que les utilisateurs aux compétences moyennes en informatique n'ont, en majorité, techniquement pas de difficultés particulières pour créer des comptes, ajouter des signets, des tags ou des tags structurés. De plus, avec la diffusion croissante des outils participatifs en ligne, il semble probable que ces tâches assez simples le deviennent de plus en plus.

Intérêt des fonctionnalités Participatives. La très faible participation sur le site du Sociopôle révèle la nécessité d'adapter les contenus aux visiteurs.

Des fonctionnalités participatives sont inutiles si les visiteurs ne voient pas d'intérêt à participer. Les discussions du focus group montrent le besoin pour les documentalistes de plus prendre en compte le point de vue des utilisateurs dans l'indexation des contenus. L'écoute du public peut alors s'effectuer de manière externe ou interne à l'outil.

De manière externe, les documentalistes vont parcourir les pages Web des chercheurs pour récupérer leurs liens et leurs vocabulaires. De plus, elles contacteront l'association française de sociologie pour tisser des liens avec des chercheurs et les inciter à décrire certains signets sur le Sociopôle.

De manière interne, l'écoute du public est possible car chaque utilisateur peut avoir son propre modèle du domaine. Dans SemanticScuttle, ceci apparaît par exemple à travers les fonctionnalités de description commune de tags. Cette notion de point de vue, induite par le modèle IPM, nous semble primordiale pour faciliter l'apparition puis éventuellement la discussion entre plusieurs points de vue. Cette caractéristique devra néanmoins être confirmée lors des travaux futurs avec les documentalistes.

7.5 Discussions et travaux proches

Cette section discute les limites des résultats obtenus dans le projet SemanticScuttle. Néanmoins, nous élargirons aussi la discussion en comparant ces résultats aux travaux menés dans les deux projets Luisa et Pépi, ainsi qu'avec les outils apparaissant aujourd'hui sur le Web et les travaux proches du domaine.

¹⁰⁸Poster intitulé « SemanticScuttle : Pratiques collaboratives et interactivité autour d'une base de signets en sociologie (Sociopôle) » présenté aux journées RPIST 2009 : <http://rpist.inist.fr/>

7.5.1 Rapprochement avec le domaine pédagogique

Alors que nous avons principalement expérimenté SemanticScuttle sur des ressources de sociologie avec une communauté de documentalistes, quels résultats pouvons-nous généraliser au domaine pédagogique ?

7.5.1.1 Communautés de documentalistes VS communautés d'enseignants

Au début du projet SemanticScuttle, nous avons essayé de développer cet outil en partenariat avec la communauté des enseignants de Génie Mécanique. Néanmoins, nous n'avons pas été soutenus par les responsables de la communauté et nous n'avons pas eu de participation suffisante autour du projet.

C'est pourquoi lorsque nous avons été contactés par les documentalistes, nous avons choisi de concentrer nos efforts sur leurs besoins.

Quels sont les différences et les points communs entre les deux contextes ?

Les communautés enseignantes comme ListePGM regroupent des individus partageant des documents de manière libre, en dehors du cadre de leur institution professionnelle. Les documentalistes ordonnent des ressources dans le cadre de leur fonction et possèdent des compétences liées à cette tâche. Elles organisent les ressources pour d'autres utilisateurs, consommateurs des ressources. Ces points semblent totalement séparer les deux groupes.

Pourtant, les communautés d'enseignants en ligne, comme ListePGM, possèdent rarement une participation homogène. Nous observons plutôt un noyau de participants réguliers, puis des participants ponctuels et finalement un grand ensemble d'utilisateurs passifs. Sur la liste de diffusion ListePGM, ceci correspond à quelques membres actifs et motivés, face à un grand nombre d'inscrits n'ayant jamais posté.

D'un autre côté, les sites réalisés par les documentalistes ne sont pas totalement fermés aux contributions des utilisateurs, même s'ils n'incitent pas toujours à la participation. Par exemple, l'ancien site du Sociopôle permettait d'envoyer un e-mail pour ajouter des ressources et la responsable du Sociopôle a reçu plusieurs fois des ajouts ou des demandes de corrections.

Par conséquent, les contextes de ListePGM et du Sociopôle dessinent un même modèle où un noyau d'utilisateurs spécialisés enrichit et organise un ensemble de ressources à l'usage d'un plus grand nombre d'utilisateurs aux participations plus sporadiques. Ce modèle tend à la structuration des contenus mais employant la participation des utilisateurs.

Les contextes tendent même à se rapprocher avec d'un côté des enseignants qui cherchent à mieux structurer les ressources, et du coup qui ont besoin de s'organiser collectivement (création d'une association, recherche d'outils plus structurants mais moins accessibles) et d'un autre côté, les documentalistes qui ne peuvent efficacement organiser les ressources pour des chercheurs sans leur participation.

A ce modèle de « communauté progressivement structurante », le modèle de SemanticScuttle

The screenshot shows the homepage of the 'Projet-civ' collaborative directory. At the top, there is a logo with a globe and the text 'Projet-civ Annuaire collaboratif'. To the right, there are links for 'mots-clés populaires à propos' and 'se connecter s'enregistrer'. A search bar with the text 'Chercher dans' is also present. Below the header, a dark bar contains the text 'tous les signets'. A central box contains a welcome message: 'Bienvenue dans l'annuaire collaboratif de [Projet-civ.com](#). Connectez-vous pour ajouter des signets. Cliquez sur les flèches (▶) dans le menu pour afficher les sous-menus ou bien cliquez sur le mot-clé pour afficher les signets.' Below this, there are three articles listed with their titles, URLs, dates, and authors. On the right side, there is a 'mots-clés menu' section with categories like '1er cycle', 'Pédagogie', '2e cycle', 'Sites généraux', and 'Actualités', each with sub-items. At the bottom of the menu, there is a link for 'Mots-clés populaires →'.

FIG. 7.27 – Page d'accueil et tags structurés de l'annuaire Projet Civ

offre une outillage possible, notamment sous la forme des principes de fonctionnalités progressivement structurantes et progressivement collaboratives.

7.5.1.2 Usages spontanés en milieu pédagogique

Si nous n'avons pas réussi à intéresser les enseignants de Génie Mécanique, nous nous sommes aperçus que d'autres enseignants étaient intéressés par l'outil sans que nous ayons cherché à les convaincre. La Fig. 7.27 illustre le cas d'un enseignant québécois et du site [Projet-civ](#)¹⁰⁹ qui présente un « annuaire collaboratif de liens pour l'enseignement du domaine de l'univers social. ».

En juin 2009, ce site contient plus de 900 signets. Trois enseignants sont inscrits sur le site dont le fondateur qui a ajouté la majorité des signets.

Pour l'instant, le nombre réduit d'inscrits ne permet pas de valider l'aspect collaboratif du site. En revanche, les contenus et les tags structurés démontrent la possibilité d'employer l'outil avec une visée pédagogique.

¹⁰⁹ Annuaire du site [Projet Civ](http://www.projet-civ.com/annuaire/) : <http://www.projet-civ.com/annuaire/>

7.5.1.3 Revenir aux compétences : SemanticScuttle pour le C2I

Le contexte principal d'expérimentation de SemanticScuttle fut le domaine de la sociologie. Cependant, il est intéressant de considérer l'emploi d'une plateforme SemanticScuttle dans le contexte du projet LUISA. Ce contexte consistait en l'indexation de documents pédagogiques selon les compétences informatiques du C2I.

Dans LUISA, nous avons par exemple travaillé avec une ontologie de compétences fondée sur le référentiel de compétences du C2I. Ce référentiel, établi par quelques experts, était pratique pour réaliser ensuite des raisonnements sémantiques. Néanmoins, comme nous l'avons justifié dans un contexte ouvert, les ontologies peuvent difficilement être développées par des experts et ré-employées dans tous les contextes de la même manière.

C'est pourquoi pour faciliter l'implémentation des ontologies en milieu pédagogique, nous avons exploré dans LUISA l'alignement entre plusieurs ontologies similaires (voir §3.2.2.5). L'idée était d'assouplir la vision monolithique d'une unique ontologie de compétence.

Cette souplesse face aux ontologies apparaît aussi dans le modèle SemSLATES d'Alexandre Passant. Ce modèle propose une chaîne d'outils sémantiques, partant de la création collaborative d'une ontologie à travers un wiki sémantique jusqu'à l'annotation de ressources en ligne (e.g. billets de blogs) en employant les concepts développés précédemment. Dans SemSLATES, la création de l'ontologie concerne un plus grand nombre d'individus mais ces derniers, en définitive, doivent toujours travailler autour de la même ontologie du domaine. Appliqué aux compétences du C2I, un wiki sémantique permettrait une représentation peut-être commune à un plus grand nombre d'utilisateurs, que le référentiel officiel. Il est intéressant de noter que SemSLATES fut développé dans le contexte des entreprises où l'organisation peut inciter plus ou moins fortement les employés à se conformer à une même vision.

Nous n'avons pas pu expérimenter SemanticScuttle pour des documents concernant le C2I cependant nous pouvons remarquer que la plus grande différence se situe au niveau des fonctionnalités qui permettent l'expression de différents points de vue. Différents enseignants ou universités pourraient avoir différentes organisations des compétences en informatique. Ceci semble une obligation dans le domaine pédagogique, milieu moins contraint que celui des entreprises.

7.5.2 Spectre de l'intérêt des utilisateurs

Durant nos trois années de travaux, nous avons observé divers stades de vie d'outils de partage.

Cartables.net est un site pour les enseignants du primaire, où ces derniers peuvent déposer leurs fichiers, relus par quelques modérateurs pour éviter les fautes d'orthographe. Bientôt, le site va fermer car les trois modérateurs principaux sont las de maintenir le site. Le site ne s'arrête pas par manque d'échanges mais à cause de difficultés de maintenance.

ListePGM.org est à l'origine une liste de diffusion. En 2007, lors de la formation en association, les membres ont employé, en plus, une plateforme collaborative Joomla ! pour communiquer et échanger des ressources, ainsi qu'une plateforme nommée Windchill initialement prévue pour la conception collaborative en mécanique. En Juin 2009, l'association s'est recentrée sur la liste

adossée à un site avec un simple forum pour discuter.

Quand nous avons présenté Pépi ou SemanticScuttle à la communauté ListePGM, ces outils ne les ont pas intéressés. Mais les documentalistes de Sociologie furent intéressées par ce même SemanticScuttle pour leur travail.

Ces différents cas soulèvent la question de l'adéquation entre les outils et les communautés qui partagent. Sachant que tous les outils ne sont pas adaptés à toutes les communautés, quels paramètres déterminent cette adéquation ?

A partir des sites observés et de l'expérience issue des projets SemanticScuttle et Pépi, nous proposons un « spectre participatif » propre à chaque outil et permettant de mieux comprendre ses emplois possibles, ainsi que les contextes et communautés éventuels où il pourrait être implanté. Ce spectre n'a pas de valeur numériquement parlant mais aurait un intérêt symbolique pour classer les outils de partage.

Le « spectre participatif » est un graphe ayant en abscisse la motivation des utilisateurs mesurés selon leurs actions qu'ils ont la motivation d'effectuer sur le site, et en ordonnée le pourcentage d'utilisateurs prêts à effectuer ces actions.

La Fig.7.28 illustre une représentation possible des spectres participatifs pour les communautés ListePGM et celles du Sociopôle.

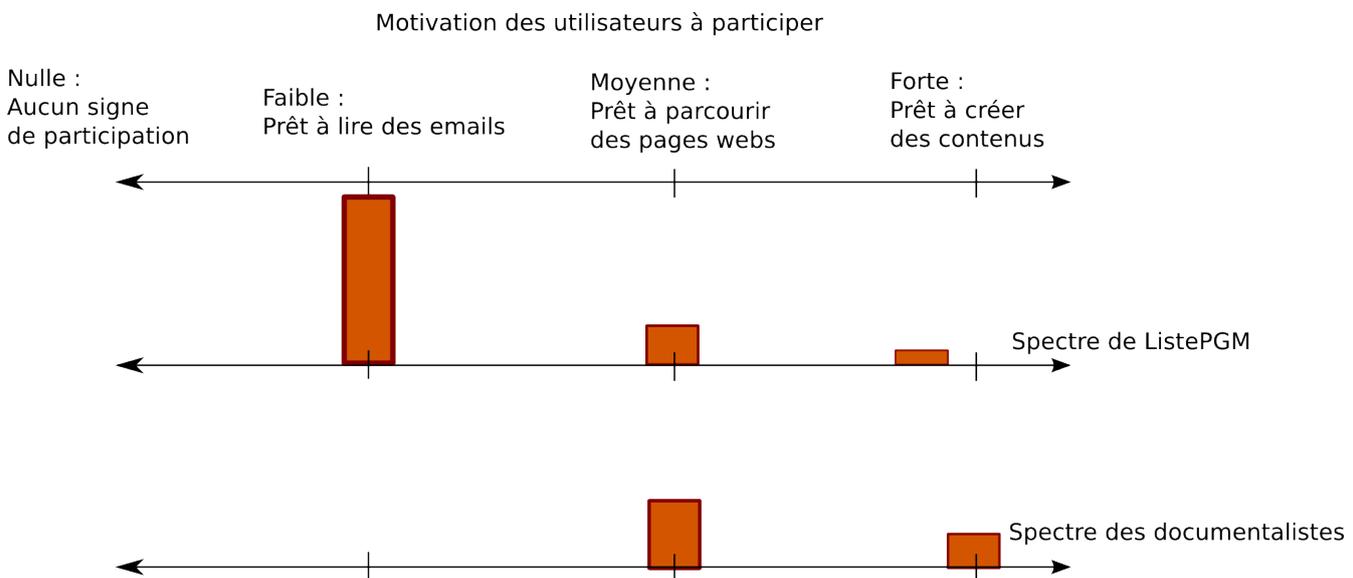


FIG. 7.28 – Spectre de motivation des participants de communautés

Ainsi, le site Cartables.net possède une organisation (outil + gestion) dont les caractéristiques ne correspondent plus à la communauté des utilisateurs. En arrêtant leur activité, les gestionnaires arrêtent les échanges entre membres sur le site.

Une liste de diffusion est un outil très simple d'utilisation de par son intégration aux emails. Cette simplicité a probablement soutenu la popularité de la ListePGM et d'autres listes d'ensei-

gnants regroupant parfois des dizaines de milliers d'abonnés. En revanche, le passage à des outils demandant une plus grande implication des membres (aller sur une page Web, s'identifier...) n'est pas évident. Ceci peut expliquer les difficultés des différents outils collaboratifs expérimentés par la communauté ListePGM.

Pour les caractériser, un spectre participatif pourrait être symbolisé sur les outils participatifs. Ceci pourrait préciser les communautés visées comme « Convient de préférence aux communautés avec un noyau d'utilisateurs réguliers ».

Par exemple, l'intérêt spontané des documentalistes pour l'outil SemanticScuttle et les fonctionnalités qu'elles ont demandé d'ajouter semble prouver l'utilité d'outil de partage sur le modèle de SemanticScuttle pour des communautés institutionnelles, contenant un noyau de responsables et prêt à laisser participer d'autres membres.

7.5.3 Perspectives de recherche

7.5.3.1 Agrégation et représentations avancées des tags structurés

Dans SemanticScuttle, nous avons proposé un moyen de relier les tags par des liens explicites d'inclusion et de synonymie. Nous avons exploré une forme d'agrégation simplifiée où une hiérarchie de tags est créée à partir des relations.

Les tags structurés offrent néanmoins des formes plus avancées d'agrégation, où par exemple une relation entre deux tags est créée pour toute la communauté quand plus de X% des membres ont créé cette relation.

Néanmoins, les contextes dans lesquels nous avons travaillé, ne nous ont pas offert l'occasion d'explorer ces possibilités de par le nombre restreint d'utilisateurs. Dans une optique pratique, nous avons préféré concentrer notre énergie sur les fonctionnalités proches des contextes abordés.

7.5.3.2 Réseau de contacts, suggestions et confiance

Dans SemanticScuttle, les utilisateurs peuvent créer une liste de contacts, dont ils pourront suivre les derniers ajouts de signets. De plus, lorsqu'un administrateur ajoute un nouveau signet, SemanticScuttle lui suggère, en plus de ses tags, les tags des autres administrateurs.

Un ajout simple serait de mêler ces deux aspects en proposant à un utilisateur, les tags populaires de son réseau de contacts, voire de son réseau élargi intégrant les contacts de ses contacts.

Le succès des interfaces de réseaux sociaux, comme Facebook ou OpenSocial, indique une tendance majeure à développer des fonctionnalités prenant en compte les aspects sociaux. La proposition de tags est un des aspects les plus simples. D'autres aspects peuvent inclure la suggestion de contacts ayant une grande part de signets ou tags en commun, par exemple.

Aux suggestions se superpose la notion de confiance sur laquelle nous n'avons pas réellement eu l'occasion de travailler. Dans SemanticScuttle, la confiance est principalement prise en compte par le statut d'administrateur dont les contributions possèdent un bandeau jaune pour les signaler. Néanmoins, dans des contextes où le nombre d'utilisateurs est plus important, il devient intéressant et possible de calculer la confiance d'un utilisateur en fonction de son réseau de

contacts. Par exemple, si un utilisateur est suivi par cinq autres, on peut supposer qu'il possède la confiance de cinq utilisateurs qui s'intéressent à ses signets.

7.5.3.3 Fusionner les techniques émergentes et automatisées

Dans 6.1.4, nous présentons quatre ensembles de techniques pour rapprocher les tags des concepts, fondés selon l'automatisation (ou non) de la tâche, et selon l'émergence (ou non) de l'ontologie. SemanticScuttle explore les techniques manuelles visant des ontologies émergentes.

Une perspective intéressante serait de fusionner les différentes techniques au sein d'un même outil. Alors que SemanticScuttle explore la coordination entre plusieurs utilisateurs humains, de nouvelles perspectives seraient l'ajout d'agents automatiques dans le système. Par exemple, un utilisateur spécial de SemanticScuttle serait en réalité un outil dont les tags structurés seraient une ontologie issue du traitement automatique des descriptions des signets du site¹¹⁰.

Cette approche est notamment possible car SemanticScuttle permet à différentes ontologies d'exister sur une même site.

7.5.4 Le Web Sémantique du W3C peut-il être social ?

Nos travaux pour les EIAH s'inscrivent en partie dans le cadre plus général de la communauté de l'Ingénierie des Connaissances.

Or dans la communauté Ingénierie des Connaissances française, il semble exister plusieurs visions quant aux possibilités du Web Sémantique défini par le W3C d'intégrer les interactions humaines.

Une vision est notamment défendue par Fabien Gandon qui soutient que le « Web Sémantique n'est pas antisocial » ([Gandon, 2006]), c'est-à-dire que le Web Sémantique n'empêche pas la création des applications du Web Participatif :

- Dans la définition du Web Sémantique([Berners-Lee et al., 2001]), « l'information se voit associée à un sens bien défini améliorant la capacité des ordinateurs et des hommes à travailler en coopération ». « Le scénario motivant du web sémantique utilisé dans sa définition même, le place au cœur d'une problématique sociale ».
- Mais cette vision peut être déformée par des usages particuliers, comme l'intérêt important à l'encontre du Web Sémantique provenant des chercheurs issus des langages de représentation des connaissances formelles.
- « Le Web Sémantique est avant tout une vision et ne doit pas être confondu avec les formalismes qu'il mobilise. » « Cette vision est celle de l'intégration de données à l'échelle du web, celle de la connexion des données à leur définition et leur contexte, pour inférer, pour permettre une collaboration effective et une réutilisation à différentes échelles (personne, groupe, organisation, communauté, etc.), pour réduire les coûts technologiques et sociaux de l'intégration effective des données en réseau, pour ajouter des données manipulables par les machines là où il n'y avait que des données manipulables par les humains, le tout en reposant sur des standards de représentation ouverts assurant la flexibilité des inscriptions »

¹¹⁰Techniques proposées par exemple dans le Workshop « The People's Web meets NLP : Collaboratively Constructed Semantic Resources » de ACL/IJCNLP-2009 : <http://www.ukp.tu-darmstadt.de/acl-ijcnlp-2009-workshop/>

- « Nous ne pensons pas que réaliser une application du web sémantique signifie implanter une solution uniquement avec les outils du web sémantique [...] Le seul effort qui lui est demandé est celui de faire le seul pas vers l'interopérabilité qui ne peut être fait à sa place : rendre explicites ses structures de données et la conceptualisation sur laquelle elle se fonde. C'est le défi des ontologies mais à l'échelle du web. »

Une autre interprétation du Web Sémantique semble s'y opposer et est défendue notamment par Manuel Zacklad. Ce dernier propose un modèle du Web Socio Sémantique (W2S), notamment soutenu par un protocole appelé HyperTopic étendant la norme des TopicMaps. Selon son interprétation :

- le W3C soutient une vision du Web Sémantique dont l'essentiel des gains promis par les WS provient d'une remontée en formalisation. Ceci peut transparaître dans le cake sémantique où la preuve et la vérité apparaissent au sommet (Fig. 2.3).
- Cette vision a guidé la quasi-totalité des publications et financements du Web Sémantique.
- « Dans le Web Socio Sémantique, le social et le sémantique se co-construisent mutuellement dans chaque transaction (formalité sémiotique ouverte). Dans le Web Sémantique Computationnel la sémantique est ancrée dans des situations transactionnelles standardisées et doit permettre de reproduire cette standardisation (fermeture sémiotique) ».
- « Dans le WS formel, l'interopérabilité doit être comprise au sens syntaxique ou au sens de la sémantique formelle, projection dans des modèles isomorphes permettant d'aboutir à des « valeurs de vérité » cohérentes. De manière radicalement distincte, dans le W2S, l'interopérabilité renvoie à la problématique de l'intercompréhension qui englobe l'ensemble des questions liées aux dimensions culturelles et linguistiques pour l'établissement d'un accord entre des participants.
- Les ontologies formelles du Web Sémantique du W3C ferment les situations transactionnelles tandis que les ontologies sémiotiques les ouvrent ».
- Les ontologies sémiotiques s'appuient sur trois dimensions de la sémantique (selon Mounin)
 - Définitionnelle : relation de type logique entre des termes (synonymie, hyperonymie) –
 - Contextuelle : signification ancrée dans un texte et un intertexte –
 - Situationnelle : signification ancrée dans des pratiques sociales basées sur des milliers d'expériences d'usage des termes en situation »

Enfin selon Manuel Zacklad, les deux approches ont des objectifs différents quant aux vocations des modèles. Pour le Web Sémantique du W3C, l'objectif est calculatoire et nécessite avant tout l'interopérabilité. Pour le Web Socio-Sémantique, l'objectif est communicationnel et nécessite l'intercompréhension.

Dans ce débat complexe, nous voyons deux niveaux au concept de « social ». Dans un sens « social » fait référence à l'implication d'utilisateurs dans les algorithmes et dans la modélisation. « Social » signifie que des utilisateurs ajoutent, organisent ou échangent des contenus. Selon nous, c'est le sens de « social » dans la vision du W3C. A nos yeux, Manuel Zacklad défend un niveau plus riche sémantiquement où « social » signifie des interactions non prédéfinies entre les utilisateurs et où un sens nouveau émerge par co-construction.

Or SemanticScuttle s'inscrit dans ces deux visions.

Nous retenons que le Web Sémantique selon Fabien Gandon permet la manipulation des machines sur le sens par l'intermédiaire de l'interopérabilité des systèmes à l'aide d'ontologies. Cet

aspect est implémenté dans SemanticScuttle à l'aide de l'export sémantique alliant les ontologies FOAF, SKOS et SIOC.

Selon Manuel Zacklad, le Web Socio-Sémantique soutient la richesse des interactions humaines, qui définissent et se définissent par le sens des objets qu'elles manipulent. Manuel Zacklad cite trois dimensions de la sémantique (définitionnelle, contextuelle et situationnelle) qui se retrouvent en partie, selon nous, dans SemanticScuttle à travers les liens d'inclusion et de synonymie entre tags (définitionnelles), les signets associés aux tags (contextuel) et le soutien à la communication dans la communauté à travers les espaces de type wiki (situationnel).

7.5.5 Améliorer l'indexation automatique à l'aide de contributions des utilisateurs

Google est le moteur de recherche le plus employé sur Internet et, malgré une diversification des services proposées par l'entreprise (mails, traitement de texte en ligne...), la majorité des dépenses concerne l'amélioration du moteur de recherche. Voici quelques travaux sur les innovations industrielles actuelles des moteurs généralistes à l'aide des contributions des utilisateurs. Comme pour les travaux de Recherche présentés précédemment, nous avons trouvé peu d'indices sur l'usage des outils suivants.

Google Custom Search. Google Custom Search (GCS) est un moteur de recherche Google mais que les utilisateurs peuvent configurer pour qu'il ne fournisse que des résultats parmi une liste de pages sélectionnées. Nous utilisons ce moteur dans SemanticScuttle sur les pages référencées par des signets.

Mais Google Custom Search permet aussi aux utilisateurs d'ajouter des tags aux pages sélectionnées transformant le moteur en un outil de partage de signets collaboratifs. Néanmoins, GCS exploite peu ces tags et n'offre par exemple pas de vues par nuage de tags.

D'autres outils comme HeyStaks¹¹¹ demandent à l'utilisateur d'ajouter une extension au navigateur qui permettra par la suite d'enregistrer des résultats de requêtes sur Internet. Ces résultats pourront alors être organisés par tags et partagés avec d'autres utilisateurs.

Google Search Wiki. En 2009, Google introduisit une nouvelle interface au moteur de recherche Google, permettant à un utilisateur ayant un compte Google, de personnaliser ses résultats. Lors de l'affichage des résultats à une requête, l'utilisateur peut indiquer qu'un site est particulièrement pertinent pour lui et doit apparaître en première position dans les résultats. Il peut aussi dire qu'un site n'est absolument pas intéressant pour une requête et ne pas le faire apparaître dans les résultats. Enfin, il peut ajouter un commentaire sous un résultat. Ce commentaire n'est pour l'instant visible que de lui-même mais Google indique si d'autres utilisateurs ont commenté le même site et l'on peut supputer que Google exploitera prochainement ces commentaires, par exemple en affichant les commentaires d'un ami¹¹².

¹¹¹HeyStacks.com

¹¹²Google Contact permet de catégoriser ses contacts lorsqu'on possède un compte Google : <http://www.google.com/contacts>

7.5.6 Chercher et expérimenter dans le domaine des EIAH

Durant ces trois années de recherche, nous avons tenu à travailler de manière proche du terrain et des utilisateurs. Les raisons sont que ceci permet :

- d’avoir des résultats utilisables à plus ou moins courts termes sur le terrain ;
- d’avoir des résultats validés en situation réelle et non juste en laboratoire ;
- de diffuser finalement les résultats directement auprès des communautés intéressées.

Nous entendons par « résultats » aussi bien les concepts théoriques que les outils les implémentant et les retours d’expériences.

Cette proximité du terrain et cette volonté de rapidité peuvent sembler en opposition avec les principes de la Recherche. Néanmoins cette rapidité n’est qu’à l’image de la rapidité de développement des outils sur le Web et de l’évolution parallèle des usages. Les blogs ou les wikis ne se sont pas diffusés sur le Web suite à des articles de Recherche mais par appropriation progressive par les utilisateurs. La notion de tags n’est pas un concept de recherche mais fut popularisé par des sites comme Delicious ou Flickr. Il nous semble qu’aujourd’hui, dans le contexte des EIAH, le meilleur moyen de diffuser un concept est de l’implémenter dans un outil puis de le diffuser. Pour que cette diffusion s’effectue, l’outil doit avoir été conçu avec les utilisateurs pour les satisfaire et qu’ils en fassent la publicité par la suite. Pour que l’outil satisfasse les utilisateurs, il doit correspondre à un ensemble de besoins basiques et avancés. Or nous ne pouvons fournir un outil avec des fonctionnalités avancées si les fonctionnalités de base sont insuffisantes. Ces arguments justifient le temps de développement important passé sur Pépi puis SemanticScuttle.

Mais cette proximité du terrain est ardue car :

- En tant que chercheurs en informatique, nous devons publier des articles pour les autres chercheurs en apportant des « nouveautés » dans les modèles. A nos yeux, cette pression à la nouveauté empêche de passer du temps sur les fonctionnalités de base qui rendront l’outil utilisable. Cette pression pousse à la complexité des outils entrant en opposition avec la facilité nécessaire à l’appropriation par les utilisateurs.
- Face aux enseignants sur le terrain, nous sommes parfois ressentis comme une gêne plutôt que comme une aide. Ceci est peut-être dû au manque de solutions concrètes, réutilisables que les travaux dans les EIAH pourraient s’attribuer.

En définitive, nous concluons à un certain succès dans la diffusion sur le terrain de l’outil SemanticScuttle, et des concepts qui le supporte :

- Un outil libre avec 150 téléchargements par mois et de régulières questions et suggestions soumises par les utilisateurs ;
- Des traductions envoyées par des contributeurs en québécois, allemand et japonais ;
- Une documentation d’une vingtaine de pages soumise par un utilisateur allemand ;
- Un réseau de documentalistes proposant l’outil à d’autres réseaux de documentalistes ;
- Des propositions pour expérimenter l’outil au sein des universités de Nancy.

Peut-être que dans un ou deux ans, SemanticScuttle aura totalement été oublié. Néanmoins, nous pensons qu’un temps important passé au développement a fourni un maximum de chances à l’outil et aux concepts pour se répandre. Mais cet aspect nous a contraint à certains choix peu valorisés par la Recherche :

- Reprendre un logiciel libre pour récupérer une communauté d’utilisateurs. Cet avantage a

- le défaut de devoir être contraint aussi à reprendre l'architecture initiale ;
- Développer en PHP car les serveurs les plus accessibles (coûts, popularité) sur Internet sont en PHP, alors qu'une majorité de travaux de Recherche se font en Java, probablement pour des raisons de facilité d'implémentation ;
 - Choisir de travailler sur le partage de documents entre enseignants du primaire et secondaire, pour l'enseignement présentiel, car ceci concerne la très grande majorité des enseignants, alors que de nombreux travaux en EIAH portent sur des formations à distance et sur l'enseignement supérieur.

Conclusion générale et perspectives

Apports principaux

8.1 Rappel et évolution des objectifs

Durant nos travaux, nous avons exploré le partage de documents pédagogiques dans différents projets, et plus particulièrement l'aide que les techniques du Web sémantique et du Web participatif peuvent apporter pour susciter et faciliter ce partage. Concernant les techniques du Web sémantique, nous avons complètement spécifié une étude de cas dans un contexte universitaire et développé les ontologies et règles permettant l'indexation et la recherche de documents fondés sur les descriptions sémantiques des domaines concernés. Concernant les techniques du Web Participatif, nous avons d'abord constaté leur quasi absence dans les sites de ressources utilisés par les enseignants, puis proposé un outil Pépi qui n'a pu être expérimenté. Ensuite, nous avons comparé dans un chapitre d'état de l'art les fonctionnalités proposées par chacune de ces approches, leurs avantages et leurs limites, ainsi que divers systèmes ayant déjà implanté des approches d'indexation de ressources par tags et par concepts.

Ces travaux nous ont conduit à faire des propositions qui ne concernent pas spécifiquement le domaine des documents de formation, mais qui sont une contribution plus générale à l'analyse des processus de partage et de gestion de documents sur le Web.

En conclusion de ce mémoire, la partie suivante résume donc l'essentiel de nos résultats sur **l'indexation et la recherche de documents, impliquant des experts – experts du thème des documents et experts de l'indexation – et des néophytes**. (Voir annexe M pour un récapitulatif de nos contributions en fonction des questions de départ.)

8.2 Nos propositions

8.2.1 Présentation du cadre : Articuler ensemble différents procédés de classification

Différents procédés de classification des documents. Un procédé de classification peut être schématisé en deux étapes : une étape de *modélisation* du domaine et une étape d'*indexation* des documents selon cette modélisation.

La modélisation du domaine peut-être *a priori* (avant l'usage des documents) ou *a posteriori* (après la lecture des documents). Dans ce second cas, la modélisation s'effectue en même temps que l'indexation (e.g. les tags).

Enfin, les actions de modélisation ou d'indexation peuvent être effectuées par des *machines*, des

Modélisation	A priori	Machines	1	2 <i>onto TAL</i>	3
		Experts	4	5 <i>bibliothèque onto experts</i>	6
		Utilisateurs	7	8	9
	A posteriori	(même que indexation)	10 <i>PageRank</i>	11	12 <i>tags</i>
		Machines	Experts	Utilisateurs	
Indexation					

FIG. 8.1 – Procédés de classification des documents selon la manière de modéliser et d’indexer *experts* humains ou des *utilisateurs* variés (pas forcément experts).

La manière de modéliser et d’indexer permet de classer les procédés de classification des documents selon douze cas (Fig. 8.1).

Par exemple, le procédé traditionnel des bibliothèques où des experts réalisent une « carte » du domaine puis des experts de l’indexation rangent les ouvrages se situe en case 5 du tableau. Lorsque des experts se réunissent pour créer une ontologie, le procédé se situe aussi en 5 car généralement ce sont aussi des experts qui vont indexer selon l’ontologie.

Lorsque des ontologies sont créées à partir de l’analyse automatique des documents (technique du TAL), la classification se situe plutôt en 2.

Les tags se situent plutôt en case 12 lorsqu’ils sont apposés au moment de la lecture du document par les utilisateurs.

Le fonctionnement initial de Google fondé sur l’analyse automatique des hyperliens entre documents (PageRank) se situe en 10.

Pourquoi choisir un procédé plutôt qu’un autre ? Au niveau de la modélisation, les modèles a priori sont généralement plus stables car les auteurs ont eu du temps pour les rendre cohérents. Cette stabilité permet alors d’y appliquer des raisonnements automatisés (cf. Web Sémantique). Néanmoins, si le domaine évolue les modèles a priori ne représentent plus la réalité des documents et des connaissances des utilisateurs. Sur ces aspects, la modélisation a posteriori peut devenir plus utile.

Au niveau des catégories d’« acteurs » (machines, experts ou utilisateurs) impliquées, chacune possède des avantages et des inconvénients plus ou moins adaptés à un contexte de classification :

Machines : Avantages : peuvent traiter et indexer un grand nombre de documents rapidement, offrent des analyses statistiques. Inconvénients : difficultés à repérer des concepts précis et pertinents dans les documents.

Experts : Avantages : possèdent une expérience stabilisée d’un domaine, peuvent fournir des modèles précis (e.g. ontologies) permettant éventuellement des raisonnements automatisés. Inconvénients : experts rares et coûteux, difficultés à connaître les conceptions et le vocabulaire des utilisateurs non-experts.

Utilisateurs : Avantages : potentiellement nombreux, peu coûteux, connaissent implicitement leurs besoins et leurs conceptions du domaine. Inconvénients : éventuelles incohérences des conceptions entre plusieurs utilisateurs, rarement disponibles pour effectuer des tâches longues et fastidieuses.

A noter que les cases vides de la Fig. 8.1 représentent éventuellement de nouvelles solutions possibles d'organisation. Par exemple, la case 11 (indexation a posteriori par des experts) correspondrait à un site Web où des experts indexent les contenus avec des tags. L'indexation serait plus souple qu'avec un modèle a priori tout en profitant d'une qualité d'expertise supérieure à l'indexation par de simples utilisateurs.

Nécessité d'articuler ensemble des procédés de classification. Pour dépasser les limites de chaque catégorie d'« acteurs », les procédés d'organisation des documents doivent de plus en plus mêler des techniques et des acteurs différents. Par exemple, les moteurs de recherche largement automatisés comme Google font aussi intervenir des testeurs humains rémunérés pour évaluer les résultats des requêtes. Leurs évaluations sont alors prises en compte pour améliorer le moteur (mêlant les cases 10 et 11 de la table 1). Des projets de recherche (e.g. [Passant, 2009]) explorent la création d'ontologies par les utilisateurs finaux et non par des experts (mêlant les cases 6 et 9).

Notre apport : l'indexation a posteriori par des experts et des utilisateurs. Dans ce contexte, nos travaux portent sur la fusion de différents procédés et en particulier sur l'indexation a posteriori par des experts et des utilisateurs (cases 11 et 12). L'objectif principal est de profiter de la diversité des points de vues des utilisateurs, et de leur aide à l'indexation tout en bénéficiant également de l'apport de personnes expertes (sur le thème des documents ou sur l'indexation i.e. documentalistes) pour rendre l'ensemble cohérent. L'objectif secondaire est de réussir à formaliser une partie de la classification afin d'y appliquer des raisonnements automatisés (cf. Web Sémantique).

Pour atteindre ces objectifs, nous avons dû dépasser trois sous-objectifs :

- Sous-objectif 1 : faciliter la participation des utilisateurs ;
- Sous-objectif 2 : faciliter la collaboration entre utilisateurs et avec les experts ;
- Sous-objectif 3 : trouver un moyen de formaliser une partie de la classification.

8.2.2 Modèle IPM

Notre solution consiste en un modèle, IPM (Fig. 8.2), et son implémentation dans une application de partage de signets, SemanticScuttle.

Le modèle d'Indexation Progressive et Multipoints de vues (IPM) est basé sur l'implication progressive des utilisateurs, c'est à dire qu'il leur propose une continuité de fonctionnalités allant des plus simples et rapides aux plus longues et complexes. Cette continuité leur permet de commencer à participer avec des efforts minimaux (Sous-objectif 1). Puis si l'utilisateur le désire et s'il en a le temps, il peut s'impliquer davantage en prenant progressivement en compte l'activité des autres utilisateurs (jusqu'à collaborer, Sous-objectif 2) ou en structurant davantage les métadonnées des documents (jusqu'à presque les formaliser, Sous-objectif 3).

Cette progressivité de l'implication nécessite que différents points de vue puissent coexister dans le système d'indexation. Tout d'abord, ceci donne une facilité aux utilisateurs qui n'ont pas

besoin de comprendre un autre point de vue avant de participer. Deuxièmement, cet aspect multipoints de vue permet la cohabitation entre des utilisateurs aux expériences différentes, ou entre utilisateurs et experts.

Modèle d'Indexation **Progressive** et **Multi-points de vue** (Modèle IPM)

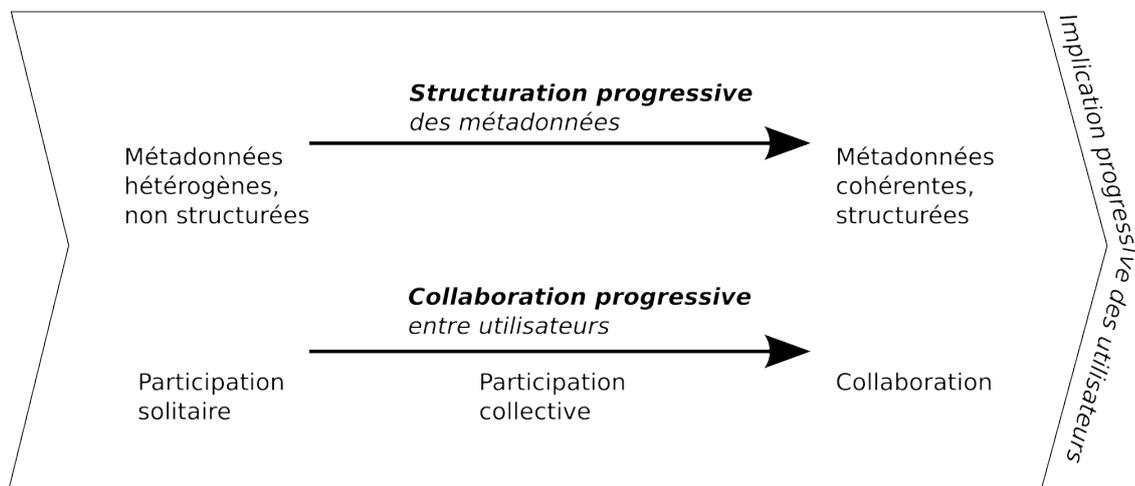


FIG. 8.2 – Modèle IPM

8.2.3 Application SemanticScuttle

Le modèle IPM fut implanté dans une application de partage de signets nommée SemanticScuttle. Comme les applications traditionnelles de ce type, SemanticScuttle permet aux utilisateurs de tagguer des ressources en ligne, mais SemanticScuttle offre de nouvelles possibilités de structuration des métadonnées et de collaboration entre utilisateurs.

La structuration progressive des métadonnées est permise car un participant peut décrire une ressource avec l'URL et un titre, y ajouter une description et des tags, y ajouter des « tags structurés » (reliés entre eux par des relations d'inclusion et de synonymie). L'utilisateur peut ensuite ajouter une description aux tags et enfin obtenir un export au format sémantique (ontologie SIOC) de la classification des documents.

La collaboration progressive entre utilisateurs est permise car un participant peut travailler en ajoutant des signets dans son propre espace (participation « solitaire »), puis il peut s'inspirer de l'activité des autres en observant leurs tags, ou en recevant des suggestions automatiques issues des tags de ses contacts (participation « collective »), enfin il peut travailler avec les autres utilisateurs sur des descriptions communes dans des espaces wiki (participation « collaborative »).

8.2.4 Leçons tirées des premières expériences

L'expérimentation de SemanticScuttle, et donc du modèle IPM, s'est déroulée selon deux angles : sur le Web et avec une communauté de documentalistes en sociologie.

Tout d'abord, SemanticScuttle a été mis à disposition sur le Web sous licence libre. Les téléchargements (en moyenne 150 tél./mois sur une année), les contributions (traductions, bouts de code, documentations) et les messages réguliers des utilisateurs montrent l'intérêt d'un outil

de partage de signets apportant plus de collaboration et plus de structuration. SemanticScuttle est généralement employé soit dans un usage mono-utilisateur (organiser et mettre à disposition ses liens préférés sur son site personnel), soit dans de petits groupes (entreprises en intranet, groupe d'enseignants¹¹³ ou associations¹¹⁴).

Ensuite, le développement de SemanticScuttle s'est déroulé en collaboration avec des documentalistes en sociologie, pour qui l'outil fut mis en production. Dans un premier temps, ces documentalistes furent particulièrement intéressées par la légèreté d'emploi des tags, la structuration permise par les tags structurés, ainsi que la collaboration entre elles. Dans un second temps, elles furent intéressées par les moyens de faciliter la participation des visiteurs. L'emploi de l'outil révéla alors le décalage entre les documentalistes et le public visé. Il semble difficile de faire participer les visiteurs sur un site aux contenus déjà structurés et employant un vocabulaire qu'ils n'utilisent pas, même si d'un autre côté les fonctionnalités les incitent à participer.

8.2.5 Originalité et limites

Au niveau du modèle, l'originalité de notre apport consiste en un procédé d'indexation visant à soutenir à la fois la participation d'experts et d'utilisateurs aux profils variés, puis à articuler leurs métadonnées pour tendre vers une vue partagée du domaine (collaboration) et des métadonnées plus structurées (structuration utile aux traitements par des machines cf. Web Sémantique).

Au niveau applicatif, l'originalité de notre apport consiste en un outil utilisable par de petites communautés et diffusé dans différents pays.

La principale limite à notre apport est le manque d'observations de l'outil dans différents contextes, autres que celui initial.

¹¹³Enseignants québécois : <http://www.projet-civ.com/annuaire/>

¹¹⁴Association artistique : <http://labomedia.lobotomie.org/scuttle/populartags.php/>

Perspectives

9.1 Dans quels contextes pédagogiques, le modèle IPM et l’outil SemanticScuttle pourraient-ils être les plus intéressants ?

Au début du développement de SemanticScuttle, alors qu’il n’était qu’un outil classique de partage de signets, nous l’avons expérimenté dans une communauté des enseignants de génie mécanique qui échangeaient autour d’une liste de diffusion. Mais, cette communauté ne fut pas intéressée.

Toutefois, le modèle IPM visant à faire cohabiter experts et non-experts, il serait intéressant d’observer son implémentation dans d’autres contextes pédagogiques que l’échange entre enseignants mais visant **le partage vers les apprenants, par définition non-experts**. Par exemple, nous pensons à la communauté Sésamath¹¹⁵ pour mettre à disposition des ressources pédagogiques directement aux élèves. Ceci pourrait aussi être fait dans l’enseignement supérieur avec les UNT dont la mission consiste à « favoriser la valorisation, la production et la diffusion de ressources pédagogiques numériques validées produites par les établissements d’enseignement supérieur »¹¹⁶. Dans tous ces cas, le modèle IPM pourrait avoir un intérêt pour améliorer l’indexation, qui, réalisée exclusivement par des experts, ne correspond pas toujours aux connaissances des apprenants.

9.2 Comment gérer une continuité d’expertise ?

Le modèle IPM propose une continuité de fonctionnalités. Cette continuité semble alors peu à peu transformer la notion d’experts car il n’y a plus de réelle distinction entre experts avec beaucoup de droits d’un côté, et simples utilisateurs de l’autre. Il y a un continu d’utilisateurs s’impliquant plus ou moins. De plus, trop mettre en valeur les contenus des experts par rapport aux contenus des autres utilisateurs est risqué, car ces derniers peuvent se demander à quoi sert leur participation si elle est dissimulée sur le site. Dans SemanticScuttle, les contenus des experts sont mêlés aux contenus des utilisateurs, avec un bandeau de couleur pour les distinguer. Ce problème soulève par exemple la question de la nécessité de tags propres aux experts, distincts des tags des utilisateurs et uniquement modifiables par des experts. Pour répondre à une requête,

¹¹⁵Communauté d’enseignants en mathématiques développant de nombreux projets, notamment autour des logiciels libres : <http://www.sesamath.net/>

¹¹⁶Présentation des Universités Numériques Thématiques (UNT) : <http://www.educnet.education.fr/superieur/unt>

comment prendre en compte les métadonnées des experts et celles des utilisateurs ? Doivent-ils indexer les mêmes aspects des documents ?

9.3 Comment gérer les liens entre utilisateurs ?

SemanticScuttle permet à un utilisateur d'ajouter un autre utilisateur dans ses contacts. Il peut ainsi facilement observer sur une page l'activité de ses contacts. Au moment de l'ajout de métadonnées, le système propose aussi les tags des contacts. Ces liens entre utilisateurs peuvent potentiellement créer des groupes d'utilisateurs au sein de l'ensemble des utilisateurs. Comment doivent être pris en compte ces sous-groupes ? Faut-il les privilégier au risque de décomposer l'ensemble ? Pour un utilisateur, faut-il privilégier les suggestions issues des experts ou des contacts ?

9.4 Comment intégrer les analyses automatiques dans l'indexation ?

Par rapport à la classification des procédés d'indexation (8.1), SemanticScuttle explore un procédé permettant de recouvrir les cases 11 et 12 (indexation a posteriori par experts et utilisateurs). Néanmoins, il nous semble intéressant d'explorer la fusion avec les analyses automatiques par des machines (case 10). Par exemple, un utilisateur de SemanticScuttle pourrait être un agent virtuel proposant une sélection et une organisation de documents fondées sur le traitement linguistique automatique des contenus (e.g. [Zouaq and Nkambou, 2008]). Il serait alors intéressant d'observer comment les contributions de cet agent virtuel d'indexation seraient prises en compte par les experts et les utilisateurs. Faudrait-il alors cacher ou souligner le côté virtuel de cet utilisateur particulier ? Pourrait-on créer plusieurs utilisateurs virtuels chacun employant une technique d'analyse automatique différente ?

9.5 Comment gérer techniquement la complexité croissante des métadonnées ?

SemanticScuttle est basé sur un schéma de base de données ad hoc. Les perspectives précédentes montrent que les métadonnées vont avoir tendance à se complexifier à mesure que différents points de vues seront possibles, que différents réseaux d'utilisateurs émergeront et que le système devra fournir des analyses prenant en compte tous ces facteurs. D'un point de vue technique, l'emploi d'ontologies et de moteurs sémantiques génériques et rodés semblent nécessaires, de manière native pour le système, pour faciliter la gestion de cette complexité croissante.

Annexes

A

Publications

Nos travaux ont donné lieu à 10 publications.

Chapitre d'ouvrage :

Knowledge Framework supporting semantic search of learning resources Grandbastien M., Huynh-Kim-Bang B., Monceaux A. Dans Metadata and Semantics, MA. SICILIA & M.D. LYTRAS (Ed.) (2008) p. 259-268 - <http://hal.archives-ouvertes.fr/hal-00365884/fr/>

Conférences internationales :

Merging semantic and participative approaches for organising teachers' documents Huynh-Kim-Bang B., Dané E., Grandbastien M. Dans Proceedings of ED-Media 08 - ED-MEDIA 08 - World Conference on Educational Multimedia, Hypermedia & Telecommunications, Vienna : Autriche (2008) - <http://hal.archives-ouvertes.fr/hal-00365714/fr/>

Conférences nationales :

Social bookmarking et tags structurés Huynh-Kim-Bang B., Dané E. Dans 19èmes Journées Francophones d'Ingénierie des Connaissances - IC 2008 - Ingénierie des Connaissances - IC 2008, Nancy : France (2008) - <http://hal.archives-ouvertes.fr/hal-00359418/fr/>

Les ontologies du prototype LUISA, une architecture fondée sur des web services sémantiques pour les ressources de formation Grandbastien M., Huynh-Kim-Bang B., Monceaux A. Dans Actes de la conférence Ingénierie des Connaissances 2008 (IC'08) - Ingénierie des Connaissances (IC), Nancy : France (2008) - <http://hal.archives-ouvertes.fr/hal-00365871/fr/>

Une recherche de ressources guidée par une ontologie de compétences Grandbastien M., Huynh-Kim-Bang B. Dans Actes de TICE 2008 - TICE 2008, Paris : France (2008) - <http://hal.archives-ouvertes.fr/hal-00365877/fr/>

Indexation collaborative à base de tags structurables à l'usage des communautés d'enseignants Huynh-Kim-Bang B., Dané E. Rencontre Jeunes Chercheurs en EIAH 2008 (rjc-EIAH 2008), Lille : France (2008) - <http://hal.archives-ouvertes.fr/hal-00365706/fr/>

Ateliers :

Social bookmarking tool based on structurable tags for communities of teachers Huynh-Kim-Bang B., Dané E. Young Researchers Track in International Conference on Intelligent Tutoring System 2008 (YRT-ITS'08), Montréal : Canada (2008) - <http://hal.archives-ouvertes.fr/hal-00365717/fr/>

Ontologies for enabling Learning Objects Metadata Semantic Interoperability : A case study

Grandbastien M., Huynh-Kim-Bang B. SWEL Workshop in ITS 2008, Montreal : Canada (2008)
- <http://hal.archives-ouvertes.fr/hal-00365882/fr/>

The LUISA framework for enabling semantic search of learning resources Grandbastien M.,
Huynh-Kim-Bang B., Pariente-Lobo T., Sicilia M.-A. LORNET Workshop in ITS 2008, Mon-
treal : Canada (2008) - <http://hal.archives-ouvertes.fr/hal-00365881/fr/>

Poster :

Quelles contributions du Web 2.0 au partage de documents pédagogiques ? Huynh-Kim-Bang
B. 3ième conférence en Environnement Informatique pour l'Apprentissage Humain - EIAH 2007,
Lausanne : Suisse (2007) - <http://hal.archives-ouvertes.fr/hal-00359399/fr/>

B

Détails des remue-méninges et des entrevues pour LUISA

Séances de remue-méninges

Ces séances ont eu lieu le 30 mars et le 6 avril 2006, réunissant 12 puis 3 personnes.

Méthode

Les participants faisaient partie des services de documentation de l'université, du service E-learning, du service des langues ou étaient professeurs aux départements Biologie, Sciences de la Matière, Médecine, Informatique de l'université.

Après les présentations, les participants effectuèrent une séquence de remue-méninges de 5 minutes portant sur une question humoristique pour "débloquer l'imagination". Puis, pendant le reste du temps, la séquence réelle porta sur la question "Basée sur vos expériences professionnelles et personnelles, imaginez les services de demain autour des ressources numériques à l'université."

Les participants se mirent par groupe de trois ou quatre pour trouver des idées, puis les réponses furent mises en commun. Un dernier temps servi à regrouper les idées par thèmes.

Les résultats présentés ci-dessous résultent de notre post-analyse et furent envoyés aux participants.

Résultats des séances de remue-méninges

Sentiment et interrogation majeurs du groupe :

- Besoin fort mais difficilement expliquable d'avoir des "ressources vivantes".
- Il manque un rapport humain autour des ressources numériques pour les faire vivre.
- Qu'apporterait à l'utilisateur cette présence humaine ?
- Comment se manifesterait-elle ?

Citation de l'expérience des bibliothécaires :

"Aujourd'hui, les gens abandonnent les bibliothèques et ne s'adressent plus à nous, spécialistes de la recherche. Ils préfèrent aller sur Google et croient maîtriser leurs

recherches en tapant deux mots. Alors qu'en fait, ils savent de moins en moins bien chercher."

- Que gagnent les gens en employant Google? Que perdent-ils en ne faisant pas appel aux bibliothécaires?
- Peut-on intégrer l'apport humain et l'apport de la machine?

2 types de ressources :

- la ressource "morte", simplement déposée et mise à disposition de certains et donc cachées à d'autres
- la ressource "vivante", matière et catalyseur, source et résultat d'un travail humain et entre humains

La ressource "vivante" ne peut être extraite de son environnement. C'est-à-dire de son contexte humain. Les caractéristiques de cet éco-système apparaissent à travers plusieurs métaphores et idées fortes :

Métaphore de la cafétaria :

- Besoin d'un lieu de rencontre convivial.
- Où rien n'est caché : égalité.
- Aussi bien lieu de travail que lieu de détente : informalité.
- Lieu simple de partage, comme autour d'une table : facilité.
- Lieu de rencontre non segmenté (hasard et richesse des rencontres).

Métaphore du marché (où l'on fait ses courses) :

- Liberté de déplacement : se balader.
- Où l'on trouve de tout : diversité.
- Où l'on peut demander conseil sur un produit. En discuter.
- Où l'on peut tâter, sentir le produit. Voir son vécu et usage.
- Une recherche "naturelle".

Métaphore du pharmacien :

- Le spécialiste ouvert sur la rue : Lorsqu'on veut un renseignement médical, on ne va pas voir le médecin mais le pharmacien. Sentiment d'accessibilité et de proximité.

Outil d'élaboration de ressources conservant la trace de l'homme :

- Permettant la construction collaborative .
- Permettant le suivi de l'évolution en fonction de l'apport de chacun.
- Capitalisation des échanges.

Vers une nouvelle forme de pédagogie liée à la distance et aux ressources numériques :

- Comment accompagner les apprenants avec les ressources? Quel degré d'autonomie lui laisser?
- Comment aider le professeur dans sa nouvelle facette de guide parmi les ressources?
- Comment les ressources peuvent-elles motiver les apprenants?

Entrevues

Les entrevues se déroulèrent, juin 2006, généralement pendant une durée de deux heures. Elles ont eu lieu avec six enseignants de l'université, dont deux avaient participé à une séance de remue-méninges.

Méthode

Nous nous sommes aperçus que la manière de présenter les technologies employées par le projet LUISA avait une influence non nulle sur les entrevues. La plupart des personnes rencontrées étant totalement nophyte quant à ces technologies, notre but n'était pas de les leur présenter en détail mais juste de les sensibiliser aux apports et au fonctionnement général.

Nous avons présenté les technologies des Services Web Sémantiques en tant que modèles riches d'un domaine (les ontologies), appliqués pour décrire des documents (annotation) et pouvant être traité informatiquement pour améliorer leur manipulation (recherche, modification, ...) grâce à des raisonnements avancées (règles d'inférence).

Nous expliquions ensuite par l'exemple à quoi pouvait ressembler une ontologie et à quoi pouvaient ressembler les raisonnements dessus. Nous illustrions notamment notre discours à l'aide de l'exemple très connu (dans le domaine des SWS) de l'agence de voyage.

”Un utilisateur désire partir en vacance au soleil pour quinze jours durant le mois d'avril et pour une somme maximale de X euros. Il se rend sur le site Web d'une agence de voyage et formule sa requête sur l'interface. Au lieu que le site web ne traite directement sa demande avec une base de donnée classique, celui-ci va découper la requête en plusieurs sous-requêtes qui seront envoyées à des Services Web appropriés. Par exemple, il va demander à un service web géographique quels sont les pays ensoleillés proche du pays de l'internaute. Considérons que le Service Web renvoie une liste de pays avec la Tunisie en première position. Le site Web va alors interroger un Service Web touristique en Tunisie pour demander une liste d'hôtels disponible en Avril pour un prix inférieur à X euros. Le Service Web tunisien va alors demander à un Service Web de conversion quel est l'équivalence actuel entre des euros et des dinars (monnaie tunisienne). Enfin il renverra une liste d'hôtels, répondant aux critères, au site Web touristique qui l'affichera à l'internaute.” Cette exemple illustre l'aspect distribué des Web Services. Pour communiquer entre eux les Services Web ont besoin de partager des portions commune de vision du monde qui sont les ontologies monétaire (pour la conversion), touristique (pour les hôtels) et géographique (pour situer les pays ensoleillés les plus proches). Enfin, le déroulement général illustre les raisonnements basiques que peuvent appliquer les Services Web en découpant une requête en sous-requêtes par exemple.

L'exemple ci-dessus ne porte pas directement sur de la recherche de documents. Nous attendions des personnes interrogées en quoi les technologies présentées pourraient s'implanter et être intéressante dans un contexte d'apprentissage et universitaire.

Résultats des entrevues : projets proposés par les enseignants

Voici les objets pédagogiques proposés par les personnes de l'université interrogées.

Travailler collaborativement autour de documents de VAE La VAE est la Validation des Acquis de l'Expérience. Ceci correspond au service de l'université permettant de transformer une expérience professionnelle en un diplôme universitaire. Cette validation s'effectue sur une période d'à peu près neuf mois et prend la forme d'un "dossier d'expression" réalisé par le professionnel avec l'aide d'un accompagnateur. Or de nombreux postulants n'amènent pas le dossier à terme par manque de motivation ou de suivi. Le service de VAE travaille donc sur une mise en communauté des postulants pour qu'ils progressent ensemble sur l'accomplissement de leur projet. Les technologies sémantiques serviraient à organiser les documents et à soutenir le processus collaboratif pour la communauté.

Indexer des lames virtuelles Les lames virtuelles sont la numérisation de lames observées au microscope. Les lames virtuelles sont des images riches dans lesquelles on peut naviguer, zoomer et elles possèdent un fort intérêt pédagogique. Ce projet était proposé par le responsable de la microscopie électronique à Nancy.

Indexer des exercices d'anglais La responsable du service des Langues de l'UHP nous a proposé d'indexer les exercices et supports en langue anglaise.

Exerciseur pour diapositives de cours Un professeur nous a proposé d'employer les technologies sémantiques pour réaliser un exerciseur à partir de l'annotation des diapositives de cours selon des concepts et des termes. Selon lui, chaque diapositive de cours possède une ou plusieurs idées fortes liées à un vocabulaire précis. Il désirait présenter aléatoirement les diapositives aux étudiants et les interroger sur les idées fortes en les laissant remplir un champ libre pour chaque diapositive. L'objectif était ensuite de faire analyser automatiquement les champs libres pour repérer les termes et les concepts-clefs. En comparant les concepts trouvés et l'annotation des diapositives par l'enseignant, l'objectif était d'aider l'enseignant à évaluer la compréhension du cours par les élèves.

Exerciseur pour auto-formation au C2I Un professeur de STAPS propose de décrire les exercices du C2I pour réaliser ensuite un exerciseur d'auto-évaluation des étudiants (voir contexte ci-dessous).

C

Référentiel officiel du C2I niveau 1

Le référentiel national de compétences du C2i niveau 1 comprend 44 compétences réparties en 9 domaines. Il fut créé en 2005 et modifié légèrement en 2008¹¹⁷.

Domaine A1 : Tenir compte du caractère évolutif des TIC Aptitudes générales :

- Être conscient de l'évolution constante des TIC et de la déontologie qui doit leur être associée et être capable d'en tenir compte dans le cadre des apprentissages.
- Prendre conscience des nécessaires actualisations du référentiel du C2i® niveau 1.
- Travailler dans un esprit d'ouverture et d'adaptabilité (adaptabilité aux différents environnements de travail, échanges).
- Tenir compte des problèmes de compatibilité, de format de fichier, de norme et procédure de compression et d'échange.

Domaine A2 : Intégrer la dimension éthique et le respect de la déontologie

- Respecter les droits fondamentaux de l'homme, les normes internationales et les lois qui en découlent.
- Maîtriser son identité numérique.
- Sécuriser les informations sensibles - personnelles et professionnelles - contre les intrusions frauduleuses, les disparitions, les destructions volontaires ou involontaires.
- Assurer la protection de la confidentialité.
- Faire preuve d'esprit critique et être capable d'argumenter sur la validité des sources d'information.
- Mettre en œuvre des règles de bons comportements, de politesse et de civilité.
- Comprendre et s'approprier des chartes de comportement.

Domaine B1 : S'approprier son environnement de travail

- Organiser et personnaliser son bureau de travail.
- Être capable, constamment, de retrouver ses données.
- Structurer et gérer une arborescence de fichiers.
- Utiliser les outils adaptés (savoir choisir le logiciel qui convient aux objectifs poursuivis).
- Maintenir (mise à jour, nettoyage, défragmentation, ...).
- Organiser les liens (favoris-signets) dans des dossiers.
- Se connecter aux différents types de réseaux (filaire et sans fil).

¹¹⁷<http://www2.c2i.education.fr/sections/c2i1/referentiel>

Domaine B2 : Rechercher l'information

- Distinguer les différents types d'outils de recherche.
- Formaliser les requêtes de recherche.
- Récupérer et savoir utiliser les informations (texte, image, son, fichiers, pilote, applications, site, ...).

Domaine B3 : Sauvegarder, sécuriser, archiver ses données en local et en réseau

- Rechercher un fichier (par nom, par date, par texte, ...).
- Assurer la protection contre les virus.
- Protéger ses fichiers et ses dossiers (en lecture/écriture).
- Assurer une sauvegarde (sur le réseau, support externe, ...).
- Compresser et décompresser un fichier ou un ensemble de fichiers/dossiers.
- Récupérer et transférer des données sur et à partir de terminaux mobiles.

Domaine B4 : Réaliser des documents destinés à être imprimés

- Réaliser des documents courts (CV, lettre, ...).
- Élaborer un document complexe et structuré (compte rendu, rapport, mémoire, bibliographie, ...).
- Maîtriser les fonctionnalités nécessaires à la structuration de documents complexes (notes de bas de pages, sommaire, index, styles, ...).
- Intégrer les informations (images, fichiers, graphiques, ...).
- Traiter des données chiffrées dans un tableau (formules arithmétiques et fonctions simples comme la somme et la moyenne, notion et usage de la référence absolue), les présenter sous forme de tableau (mise en forme dont format de nombre et bordures) et sous forme graphique (graphique simple intégrant une ou plusieurs séries).
- Créer des schémas (formes géométriques avec texte, traits, flèches et connecteurs, disposition en profondeur, groupes d'objets, export sous forme d'image).

Domaine B5 : Réaliser la présentation de ses travaux en présentiel et en ligne

- Communiquer le résultat de ses travaux en s'appuyant sur un outil de présentation assistée par ordinateur.
- Adapter des documents initialement destinés à être imprimés pour une présentation sur écran.
- Réaliser des documents hypermédias intégrant textes, sons, images fixes et animées, liens internes et externes.

Domaine B6 : Échanger et communiquer à distance Utiliser à bon escient et selon les règles d'usage :

- le courrier électronique (en-têtes, taille et format des fichiers, organisation des dossiers, filtrage) ;
- les listes de diffusion (s'inscrire, se désabonner) ;
- les forums de discussion (modéré, non modéré) ;
- le dialogue en temps réel ;
- les terminaux mobiles.

Domaine B7 : Mener des projets en travail collaboratif à distance Travailler dans un environnement de travail collaboratif :

-
- utiliser les outils d'un espace de travail collaboratif (plateforme, bureau ou campus virtuel, ...);
 - élaborer en commun un document de travail (assurer le suivi des corrections, ajouts et suppressions);
 - gérer différentes versions d'un même document.

D

Exemple de ressource C2I

Voici un exemple de documents autour des compétences du C2I, provenant d'un département de l'université Henri Poincaré.

Examen de bureautique.

24 mai 2006
durée 1h
(tous documents de TD autorisés)

L'objectif de cet exercice est de réaliser une page web (html) présentant un album (cd) d'un artiste. L'exemple choisi est l'album "Bad Cowboy" de "Axel Bauer", vous pouvez choisir un autre album d'un autre artiste. Le résultat est présenté dans le document "Bad Cowboy.pdf" (ne tenez pas compte des bordures blanches qui apparaissent à gauche et en bas de ce document).

Quelques informations pratiques :

- La table fait 500px de large.
- La 1^{ère} colonne fait 150px de large.
- Le font d'écran vous est fourni, il est contenu dans le document "background_textures_013.jpg".
- La jaquette de l'album ne vous est pas fournie. Vous devez la trouver via une recherche sur Internet.
- Le nom de l'artiste et la jaquette sont des liens qui mènent au site officiel de l'artiste "<http://www.axelbauer.com/>".

Le résultat de votre travail doit être une page html nommée "nom_prénom.html" (remplacez nom et prénom par les vôtres :-). Cette page est à envoyer à la fin de l'heure par e-mail (courriel) à l'adresse "XXXXXXXXXX@loria.fr".

Bon travail!

FIG. D.1 – Enoncé d'examen autour des sites en ligne (compétence B5 du C2I)



Artiste/Groupe :	Axel Bauer
Année :	2006
Album :	Bad Cowboy
Titres :	<ol style="list-style-type: none">1. Celle qui m'arrive2. Fais comme si3. Tu me tues4. Si je pouvais5. Jam 1 - Part 16. Bad Cowboy7. Ma liberté8. Tout reste encore à faire9. Alligator10. Aveugle et sourd11. Tout va bien12. Jam 1 - Part 213. Shine14. Un place au soleil15. Ma liberté - Part 216. Jam 217. Merci

FIG. D.2 – Ressource jointe à l'énoncé de l'examen

E

GCS adapté au contexte du C2I

Vue générale avec des schémas

Voici une représentation des instances de *CompetencyDefinition* (CD) ainsi que leurs relations *details* et *requires*.

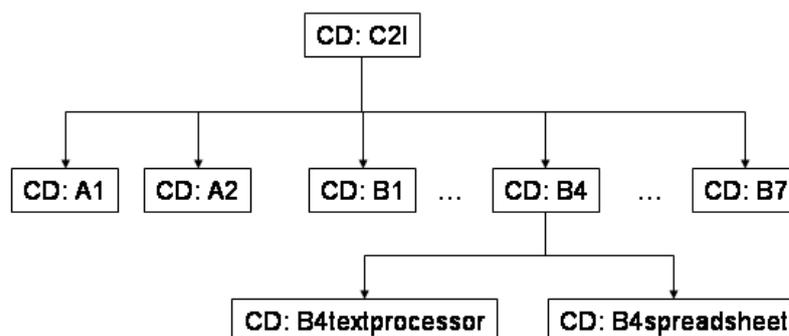


FIG. E.1 – Vue des *CompetencyDefinitions* selon la relation *details*

Une *CompetencyDefinition* peut se décomposer en différents *CompetencyElementDefinition* comme les *KnowledgeElementDefinition*. La Figure E.3 montre comment une *CompetencyDefinition* est lié à des *KnowledgeElementDefinition* qui sont reliés à des instances de l'ontologie des logiciels.

Extrait du code WSMML de GCS adapté au C2I

Code WSMML illustrant la compétence "B2" détaillée par ses compétences secondaires ("k_distinguish_search", "k_design_requests", "k_use_search_results"), ainsi que la compétence "B1" dont elle dépend (relation *requires*).

```
ontology _"http://www.uhp-nancy.fr/ontologies/LUISA#uhpGCS"
```

```
importsOntology
  _"http://www.cc.uah.es/ie/ont/gcs#gcs-ont"
```

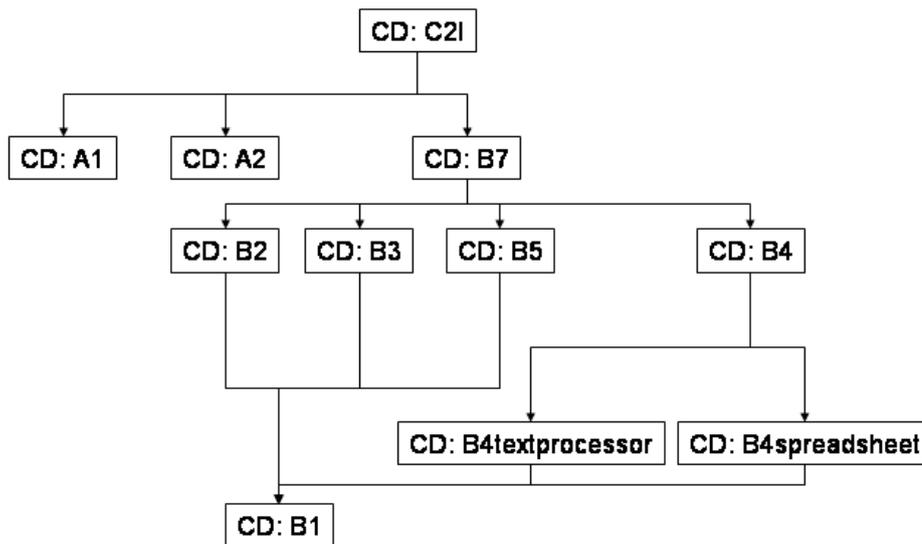


FIG. E.2 – Vue des *CompetencyDefinitions* selon la relation *requires*

```

concept uhpKnowledgeElementDefinition subConceptOf
  _"http://www.cc.uah.es/ie/ont/gcs#knowledgeElementDefinition"
  about impliesType (1 *) computerLiteracyItem
...

instance b1 memberOf _"http://www.cc.uah.es/ie/ont/gcs#competencyDefinition"
  nonFunctionalProperties
    rdfs#label hasValue "b1"
  endNonFunctionalProperties
  completelyDefined hasValue "_boolean(\"true\")"
  requires hasValue {k_desktop, k_filetree, k_OS_maintenance,
k_select_software, k_bookmarks, k_connect_to_network}

instance b2 memberOf _"http://www.cc.uah.es/ie/ont/gcs#competencyDefinition"
  nonFunctionalProperties
    rdfs#label hasValue "b2"
  endNonFunctionalProperties
  completelyDefined hasValue "_boolean(\"true\")"
  requires hasValue {k_distinguish_search_engine, k_design_requests,
k_use_search_results, b1 }
...

instance k_distinguish_search_engine memberOf knowledgeElementDefinition
  nonFunctionalProperties
    rdfs#label hasValue "k_distinguish_search_engine"
  endNonFunctionalProperties
  about hasValue _searchEngine
  
```

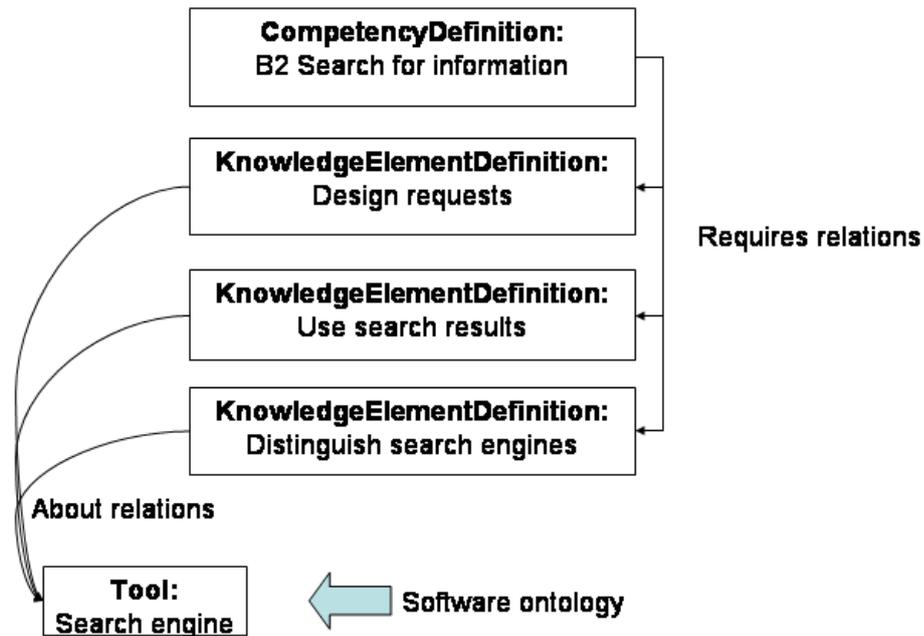


FIG. E.3 – Vue des *CompetencyDefinitions* selon la relation « requires » - niveau *KnowledgeElements*

```
instance k_design_requests  memberOf knowledgeElementDefinition
  nonFunctionalProperties
    rdfs#label hasValue "k_design_requests"
  endNonFunctionalProperties
  about hasValue _searchEngine

instance k_use_search_results  memberOf knowledgeElementDefinition
  nonFunctionalProperties
    rdfs#label hasValue "k_use_search_results"
  endNonFunctionalProperties
  about hasValue _searchEngine

...
```


F

Ontologie des logiciels du C2I : ComputerLiteracy

Vue générale

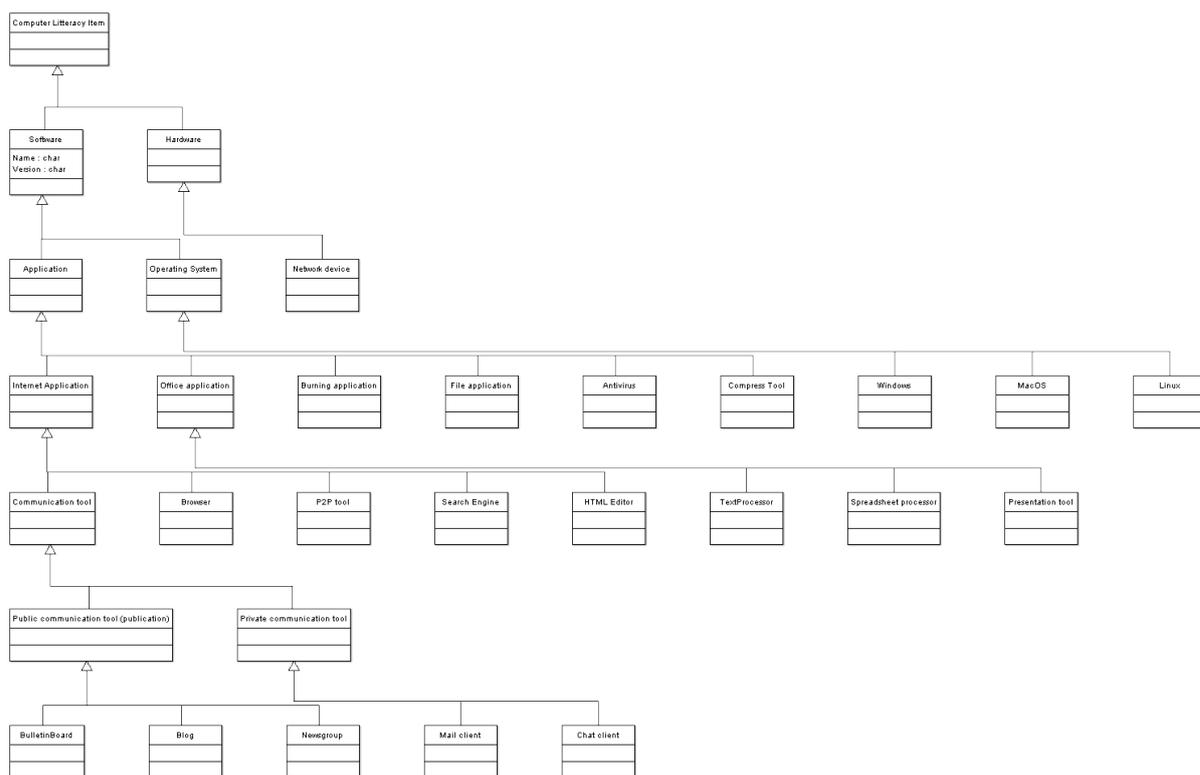


FIG. F.1 – Représentation des concepts de l'ontologie de logiciels (*ComputerLiteracy*) employée dans LUISA

Code WSML

```
ontology _"http://www.uhp-nancy.fr/ontologies/LUISA#uhpComputerLiteracy"
```

```
    nonFunctionalProperties
      wsmostudio#version hasValue "0.6.0"
    endNonFunctionalProperties

concept computerLiteracyItem

concept hardware subConceptOf computerLiteracyItem

concept networkDevice subConceptOf hardware

concept software subConceptOf computerLiteracyItem
  cost impliesType (0 1) _string
  interoperableWith impliesType computerLiteracyItem

concept operatingSystem subConceptOf software

concept windows subConceptOf operatingSystem

concept macOS subConceptOf operatingSystem

concept linux subConceptOf operatingSystem

...

instance windows98 memberOf windows
  nonFunctionalProperties
    rdfs#label hasValue "windows98"
  endNonFunctionalProperties
  cost hasValue "notfree"

instance windowsMillenium memberOf windows
  nonFunctionalProperties
    rdfs#label hasValue "windowsMillenium"
  endNonFunctionalProperties
  cost hasValue "notfree"

...
```

G

Détails des règles sémantiques dans LUISA

Voici un extrait des raisonnements du projet LUISA, employés lors de la recherche de documents par le prototype. Nous les avons spécifiés en langage naturel puis les partenaires techniques les ont transformé en règles formelles appliquées par le moteur d'inférence IRS-III. L'extrait est en anglais.

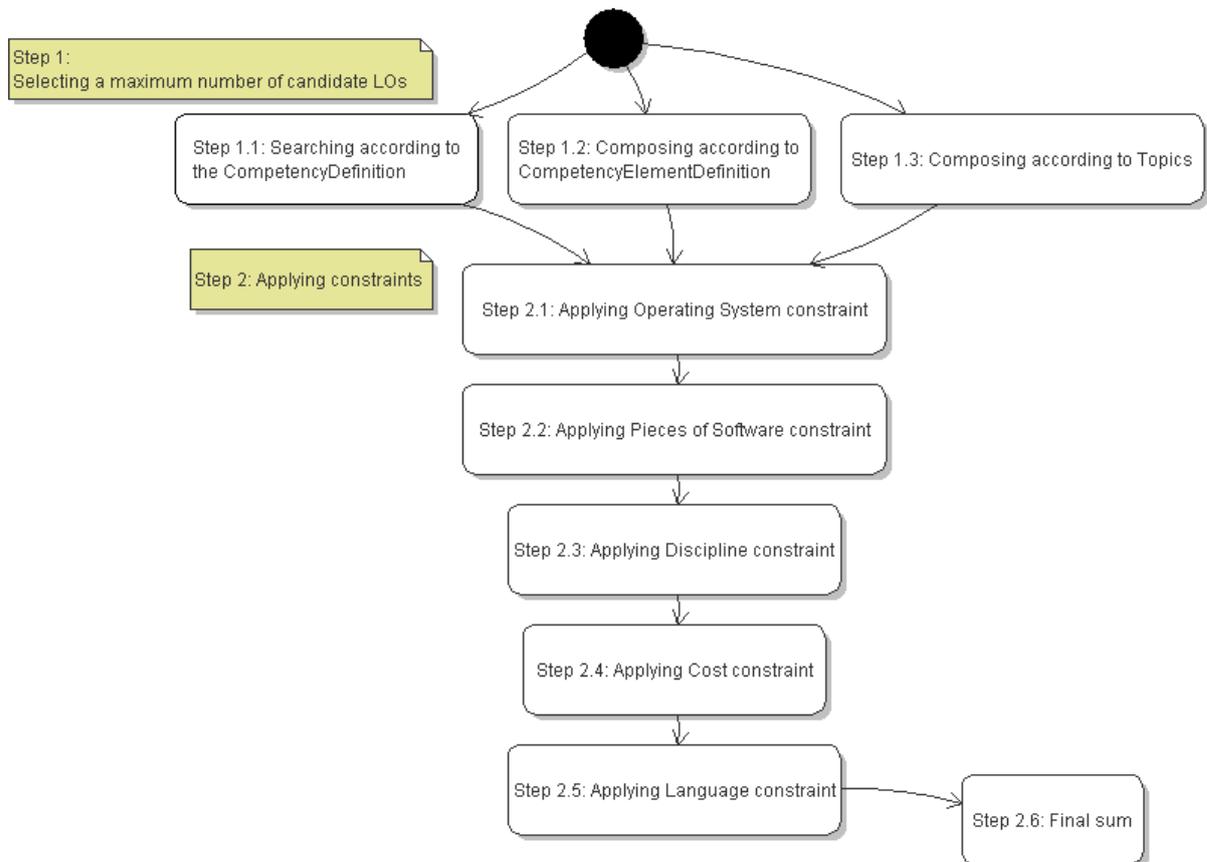


FIG. G.1 – Diagramme d'état montrant les différentes étapes pendant la recherche des documents (Prototype LUISA)

The searching process is composed of steps (see Fig. G.1). The first step aims at providing a wide range of LOs (by simple selection in the LOR and by LO composition). The second step and its sub-steps (from 2.1 to 2.5) apply constraints. The goal of these sub-steps is to value the LOs according to these constraints. At the end, values are summed by LOs (step 2.6) and the LOs with highest values are proposed to the user. All the values are between 0 and 1 but are weighted with coefficients.

Example : To illustrate the value system, we can imagine a very simplified example (see Fig. G). A user in Chemistry looks for a LO in French. All the searched LOs will be compared to this query. Let's imagine a LO in French but oriented to Biophysics (it means the spreadsheet exercises takes place in a biophysics context). How will this LO be valued ?

	Query	LO
Discipline (step 2.3)	Chemistry	Biophysics
Language (step 2.5)	French	French

FIG. G.2 – Simplified example of searching process

We will just observe the step 2.3, step 2.5 and step 2.6 (final sum). The language comparison is the easiest one. The LO has the same language as the query so it is valued 1 at the step 2.5. But Biophysics is not the requested Chemistry. At this step 2.3, a non-semantic system would have valued 0 the LO because of this difference of the field of study¹¹⁸. However, in the Discipline ontology, Biophysics and Chemistry are a specialization of the same above concept. The system can infer that these two disciplines are close to each other at the university. So for the step/constraint "FieldOfStudy", this LO is not valued 1 because it doesn't exactly fit the query ; it is not valued 0 because the system infers a proximity ; so the LO will be valued 0.5 for example. The final sum (step 2.6) is a weighted average of the value of each step (see Fig. G). In this simplified example, we can forget the weights. The final value of the LO according to the query will be : $(1 + 0.5)/2 = 0.75$

	Query	LO	Weights	Values
Discipline (step 2.3)	Chemistry	Biophysics	1	0.5
Language (step 2.5)	French	French	1	1
Final sum (step 2.6)				0.75

FIG. G.3 – Simplified example of valuation

By modifying the weights, it is possible to give more importance to a constraint. The weights configuration results in different query resolvers easily manageable.

Now we will explain the role of each step. The values at each step correspond to the standard query resolver. Other values would result in other searching behaviours and other query resolvers.

Note : one important benefit of this steps process is to make easier the adding of new constraints.

¹¹⁸We consider "Field of study" as a synonym of "Discipline".

Step 1

The sub-steps of step 1 try to provide the greatest number of LOs according to the query.

Warning : in the basic prototype, rules are simplified in order to experiment basic behaviors. Thus steps 1.1, 1.2 and 1.3 are in parallel (as shown in the steps process diagram) elsewhere the dependencies would have much more confused the process.

Step 1.1

The system looks basically for LOs according to the targeted CompetencyDefinitions selected by the user. The LOs issued from this step are valued 1 at this step.

Step 1.2

The system tries to compose few LOs according to CompetencyElementDefinitions. For example, B5 competency is detailed by B5.1, B5.2 and B5.3 according to the GCS ontology. If the user looks for B5, at this step, the system will group LOs targeting B5.1, LOs targeting B5.2 and one targeting B5.3. The LOs issued from this step are valued 0.7 at this step. Note : this step is not dependent from the Step 1.1 because it doesn't look LOs according to the same values. Warning : this tentative composition tests the capacity of the LUISA system to compose LOs using semantic reasoning based on competencies ontologies.

Step 1.3

The system tries to compose few LOs according to Topics. For example, B5 competency is about presentation tool and HTML editor. If the user looks for B5, at this step, the system will group LOs whose topic is presentation tool and LOs whose topic is HTML editor. The LOs issued from this step are valued 0.3 at this step. Note : this step is not dependent from the Step 1.1 because it doesn't look LOs according to the same values. Warning : this tentative composition tests the capacity of the LUISA system to compose LOs.

Step 2

The sub-steps of step 2 apply constraints to the LOs selected in step 1.

Step 2.1

The operating system (OS) of the user is compared to the LO's technical requirements. The LO is valued 1 to this step if they are the same. Else the LO is valued 0 at this step.

Step 2.2

The pieces of Software of the user are compared to the LO's technical requirements. The LO is valued 1 to this step if all the LO's technical requirements are in the user profile. Else the LO is valued 0 at this step.

Step 2.3

The discipline of the user is compared to the LO's discipline. The LO is valued 1 to this step if the user's discipline is the same as the LO's discipline. The LO is valued 0.75 if the user's discipline is linked to the LO's discipline. The LO is valued 0.5 if the user's discipline and the LO's discipline are generalized by the same super-concept. Else the LO is valued 0 at this step.

Step 2.4

The cost accepted by the university (university's profile) is compared to the LO's cost. The LO

is valued 1 to this step if the LO's costs are inferior to the accepted cost. Else the LO is valued 0 at this step.

Step 2.5

The language accepted by the university (university's profile) is compared to the LO's language. The LO is valued 1 to this step if they are the same. Else the LO is valued 0 at this step.

Step 2.6

The weighted average of the steps' values is calculated. The LOs are sorted according to their final value. Some constraints can be mandatory. If they are not checked by a LO, this LO is removed from the list. This can be done by configuring weights and thresholds.

At the end of this searching and sorting process, all the LOs are valued. The system proposes the list of the LOs to the user in the result page. The best valued LO is in the proposition box at the top of the list.

H

Mail aux responsables de Cartables.net

Extrait d'une discussion par e-mail avec une responsable du site cartables.net, site de partage entre enseignants du primaire, en juillet 2007.

En dehors des questions de poids des fichiers, quelles limites rencontrez-vous ?

Les limites de notre temps disponible pour vérifier que les fichiers sont conformes à ce qui est annoncé et qu'ils ne violent pas les droits en vigueur. C'est ce temps qui nous manque et c'est la source de nos plus grosses difficultés.

Avez-vous la participation que vous désirez ?

Pour l'envoi de fichiers, oui, aucun problème, on en a même trop puisqu'on a été obligé de fixer un quota par jour.

Pour la vérification, ça tourne à peu près, on a des membres du site qui nous aide de manière plus ou moins assidue mais dans l'ensemble ça ne va pas trop mal.

La recherche de documents fonctionne-t-elle bien pour les usagers ?

A priori oui, on a assez rarement de demandes à ce niveau-là, même si le nombre de fichiers est très important et que du coup, ça complique un peu les choses.

Les utilisateurs déposent-ils les documents que vous espérez ?

Il y a du bon et du moins bon, mais comme dès le départ on est parti du principe que ceux qui envoyaient étaient assez grands (et aussi compétents que nous) pour juger de ce qui pouvait être intéressant ou non, on n'intervient pas dans ce domaine. (mais bon, certains fichiers ne présentent pas un intérêt majeur c'est certain)

I

Scuttle, écran par écran



FIG. I.1 – Présentation des signets correspondant à un label dans Scuttle. Cette page correspond à la recherche des signets décrits par le tag (ou label) "usa". Nous pouvons voir que les menus sur le côté (1) se sont modifiés pour afficher les tags en relation et les tags populaires. Les tags en relation décrivent les ressources déjà décrites par "usa". Le "+" à gauche du tag "moteur" permet d'ajouter ce tag à la requête qui deviendra "usa+moteur" et qui affichera les signets comportant ces deux tags à la fois. Ceci permet de raffiner la navigation sur un ensemble restreint de signets. Les tags populaires correspondent aux tags les plus employés par l'utilisateur courant. Le menu supérieur (2) affiche les espaces personnels de l'utilisateur qui s'est identifié avec son mot de passe. L'espace "bookmarks" affiche tous les signets de l'utilisateur. L'espace "watchlist" correspond aux contacts de l'utilisateur et "ajouter un signet" permet à l'utilisateur d'ajouter un nouveau signet.

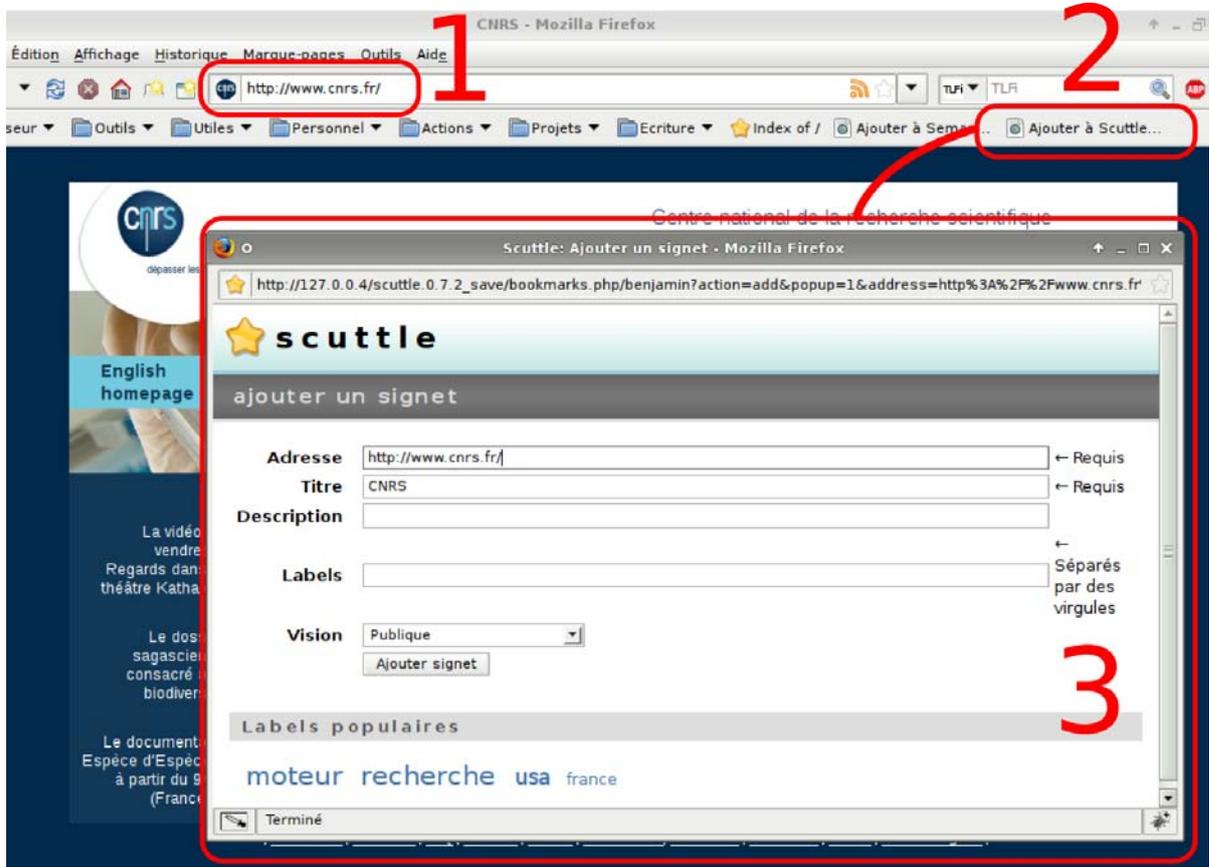


FIG. I.2 – Page illustrant l’emploi des bookmarklets dans Scuttle. Pour ajouter un signet encore plus vite, Scuttle propose d’ajouter des ”bookmarklets” à son navigateur. Un bookmarklet est un lien qui s’ajoute par un déplacer-déposer (drag’n drop) dans ses favoris. Imaginons qu’un utilisateur découvre le site du CNRS (1). Pour l’ajouter à ses signets sur Scuttle, il clique sur le bookmarklet (2) qui ouvre une fenêtre Scuttle en pop-up. Le ”bookmarklet” a automatiquement rempli les deux champs obligatoires en reprenant l’URL et le titre de la page courante. L’utilisateur n’a plus qu’à ajouter une description et des tags s’il le désire et à valider. La fenêtre pop-up se referme automatiquement permettant à l’utilisateur de poursuivre sa navigation sur la page du CNRS.



FIG. I.3 – "Watchlist" ou page des contacts dans Scuttle. Lorsqu'un utilisateur est identifié, Scuttle lui permet de suivre l'activité d'autres utilisateurs. Pour cela, il faut ajouter ces utilisateur à sa "watchlist" que l'on pourrait traduire par "liste des utilisateurs que j'observe". Ceci correspond plus ou moins sur d'autres sites participatifs à "mes amis", "mes contacts", "mon réseau" ou "ma communauté". La page "watchlist" correspond à un espace de signets qui ont pour point commun d'être dans mes observés. L'espace central (1) affiche pêle-mêle les signets de mes contacts tandis que le menu latéral (2) liste mes contacts.



scuttle

[à propos](#)

[s'enregistrer](#)

[se connecter](#)

profil: benjamin

Nom d'utilisateur	benjamin
Nom	Benjamin HKB
Page personnelle	benjamin.hbk.fr
Membre depuis	13 February 2009
Description	Doctorant en plein expérimentation.
Watched By	monique

FIG. I.4 – Page présentant le profil d'un utilisateur dans Scuttle. Les sites participatifs incitent les utilisateurs à participer en produisant généralement des contenus qui correspondent à des signets et des tags dans Scuttle. Pour les inciter, les utilisateurs sont identifiés par un surnom unique qui s'affiche sous chacun de leurs signets. Ce surnom est relié à une page "profil" visible de tous et où l'utilisateur peut se présenter succinctement. Cette page profil laisse la possibilité aux utilisateurs de mieux se connaître, d'avoir éventuellement plus confiance dans les autres utilisateurs et peut favoriser le partage.

J

Autres fonctionnalités de SemanticScuttle

Autres fonctionnalités ajoutées

Réaliser une application libre dans le but qu'elle soit diffusée, signifie qu'il faut forcément en développer des aspects utiles aux utilisateurs même s'ils n'ont pas d'intérêt direct pour nos recherches. Voici les principales fonctionnalités ajoutées, généralement à la demande des utilisateurs et parfois avec leur aide (i.e. avec leur bout de codes proposé). Cette liste a pour but d'illustrer le travail nécessaire pour passer d'un outil de recherche à emploi expérimental à un outil diffusable à plus large échelle.

Plugin recherche dans le navigateur (opensearch)

Comme nous l'avons vu précédemment avec les "widgets", les contenus des applications s'exportent de plus en plus en dehors des sites. En plus des widgets, SemanticScuttle permet d'installer dans son navigateur un plugin basé sur le format OpenSearch¹¹⁹ qui permet de partager des résultats de recherche entre applications. Ainsi de n'importe quelle page sur Internet, l'utilisateur peut effectuer de son navigateur une recherche dans les signets de son site supporté par SemanticScuttle.

¹¹⁹OpenSearch : <http://www.opensearch.org>

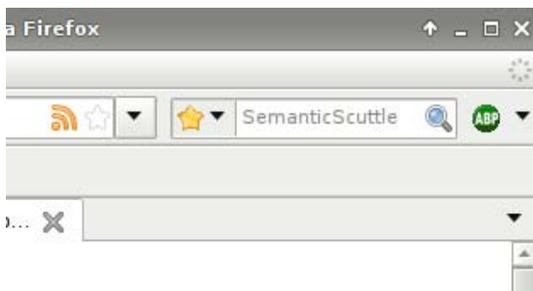


FIG. J.1 – Extrait du navigateur présentant un champ de recherche parmi les signets d'une plateforme SemanticScuttle



FIG. J.2 – Élément de l’interface affichant les dernières recherches effectuées sur la plateforme. Les nombres entre parenthèses correspondent au nombre de signets trouvés.



FIG. J.3 – Affichage d’un signet avec sur le côté gauche un aperçu du site.

Dernières recherches

Deux aspects nous ont poussé à enregistrer et afficher les recherches des utilisateurs sur SemanticScuttle.

Le premier aspect est le besoin des administrateurs de connaître les usages de la plateforme pour éventuellement améliorer les contenus et leur organisation.

Le second aspect consiste à soutenir la participation des utilisateurs en donnant des indices de la présence d’autres utilisateurs.

Pour répondre à ces deux besoins, la page principale de SemanticScuttle affiche une ”boîte” contenant les trois dernières recherches effectuées sur SemanticScuttle par l’ensemble des utilisateurs (Fig. J.2). A côté des termes recherchés apparaissent le nombre de signets trouvés. Les requêtes sans aucun résultat n’apparaissent pas.

Boîte ”nouveaux utilisateurs”

Comme pour les dernières recherches, SemanticScuttle affiche dans une boîte latérale les derniers utilisateurs inscrits sur la plateforme sur la page d’accueil du site. Un visiteur peut alors rapidement accéder au profil ou aux signets des nouveaux arrivants.

Image des sites à côté des signets

Une fonctionnalité largement réclamée par les utilisateurs est l’aperçu des sites apparaissant à côté de chaque signet. Cet aperçu est récupéré auprès de bases d’aperçus (”thumbnails”) gratuites sur internet¹²⁰. L’écran J.3 illustre l’affichage d’un signet et de son aperçu.

¹²⁰Nous employons par exemple le service d’ArtViper : <http://www.artviper.net/website-thumbnails.php>

Vérification des signets valides

SemanticScuttle est spécialisé dans l'indexation de sites Web dont la caractéristique est d'évoluer rapidement. Les sites de signets peuvent alors rapidement devenir des bases de liens morts, peu utiles aux visiteurs. Pour aider la base à rester à jour, nous avons ajouté à SemanticScuttle une fonctionnalité permettant d'envoyer une requête à tous les sites Web pour vérifier que les pages référencées existent encore. SemanticScuttle affiche ensuite un bilan des pages ont l'hébergeur a envoyé une erreur (généralement une erreur 404, page inexistante). Plus tard, il est prévu qu'en plus d'afficher un bilan des pages Web désuètes, le système mette automatiquement les signets concernés en visibilité privée, c'est-à-dire que seul son auteur puisse le voir.

Export pour tableurs (format CSV)

Collaborer avec le groupe de documentalistes sur SemanticScuttle a permis de faire surgir un certain nombre de besoins propres aux utilisateurs très impliqués sur un site. Par exemple, les documentalistes ont besoin de gérer des centaines de signets qu'ils doivent pouvoir traiter en grande quantité.

Pour faciliter ces traitements, SemanticScuttle permet un export des données d'un utilisateur au format CVS, c'est-à-dire sous la forme d'un tableau où les colonnes sont séparées par des ";" ou ";" et les lignes par des retours à la ligne.

L'avantage du format CSV est qu'il est importable dans tous les logiciels de tableurs comme Excel ou OpenCalc. Les documentalistes peuvent alors à partir du tableur imprimer les centaines de signets ou alors filtrer et trier les colonnes selon les tags, effectuer différentes statistiques.

Fonctionnalités prévues

La mise en place d'une plateforme participative chez des documentalistes renouvelle l'approche de leur travail. C'est pourquoi notre travail avec elles se prolongera probablement après la fin de la thèse pour continuer à améliorer l'outil et pour les aider à adapter leur manière de travailler. Voici quelques une des tâches prévues.

Export OAI-PMH

La principale demande de fonctionnalité à venir concerne l'export des signets au format OAI-PMH¹²¹ (Open Archives Initiative's Protocol for Metadata Harvesting). Ce format est un standard employé par de plus en plus de bibliothèques pour permettre à des services externes de récupérer leurs contenus (moissonnage) et valoriser ainsi leur travail.

Croisement entre tags et moteurs internes

Lors du focus group, les utilisateurs aux compétences informatiques élevées ont exprimé le besoin de pouvoir effectuer une recherche à l'aide des tags puis d'effectuer parmi les signets trouvés une recherche à l'aide du moteur (ou inversement).

Or permettre d'allier des modes de recherche complémentaires paraît un point important pour exploiter des métadonnées riches (provenant de différents contributeurs et plus ou moins structurées). Nous prévoyons d'implémenter ces recherches croisées prochainement dans SemanticScuttle.

¹²¹Présentation d'OAI-PMH : <http://www.culture.gouv.fr/culture/d11/OAI-PMH.htm>

K

Détails du "focus group" sur le Sociopôle

Contexte

Le "focus group" eut lieu le 7 mai 2009, pendant deux heures, dans les locaux de l'université Paris V, où travaillent les documentalistes responsables du Sociopôle.

Participants

Le tableau suivant récapitule le profil des participants, obtenus par les questionnaires remplis lors du "focus group". Sauf l'ingénieur pédagogique, ils sont tous liés au domaine des sciences humaines et sociales.

Les compétences informatiques possibles étaient : faibles, moyennes et levées. Les participants devaient dire s'ils connaissaient "le Web 2.0 ou communautaire" ("Non, jamais entendu parlé"; "Oui mais jamais utilisé"; "Oui et j'utilise les sites suivants..."). Enfin une dernière question demandait s'ils connaissaient l'ancienne version du sociopôle (§7.3.1).

Participant	Age	Profession	Comp. info.	Web 2.0	Ancien Sociopôle
1	32	Post-doctorante	Faibles	Oui, jamais utilisé	Non
2	52	Chercheur	Moyennes	Non	Non
3	43	Chercheur	Elevées	Oui, jamais utilisé	Non
4	26	Conservateur	Elevées	Oui	Oui
5	46	Ingé. pédagogique	Elevées	Oui	Non
6	64	Chercheur	Faibles	Non	Non
7	24	Doctorant	Moyennes	Oui, jamais utilisé	Non
8	30	Doctorant	Moyennes	Non	Oui
9	35	Conservateur	Moyennes	Oui	Oui

Mise en place

La séance du "focus group" était découpée en deux phases d'une heure :

- Phase d'activités individuelles sur machine et de réponses à un questionnaire
- Phase de discussion en groupe

Afin de ne pas influencer les participants, ces derniers savaient juste en arrivant qu'ils allaient expérimenter un site de ressources en sociologie. Nous ne leur donnèrent aucune autre explications avant la fin de la séance. C'est à dire que la découverte du site fut effectuée sans formation préalable.

Activités et questionnaire (1 heure)

Activités

L'énoncé de l'activité fut préparée avec les documentalistes responsables du Sociopôle afin notamment de définir les recherches à effectuer sur le site. Chaque participant est seul devant une machine. Voici l'énoncé :

Vous allez arriver sur le site pour y effectuer certaines actions (recherche, ...).

Merci de suivre dans l'ordre les consignes.

1/ Veuillez vous rendre sur le Sociopôle à l'adresse suivante : <http://www.sociopole.cnrs.fr>

2/ Prenez 1 ou 2 minutes pour découvrir le site.

3/ Effectuez les recherches suivantes en restant sur le Sociopôle :

- R1 : Combien trouvez-vous de sites de revues belges ?
- R2 : Combien trouvez-vous de sites dont l'auteur est Edgard Morin ?
- R3 : Combien trouvez-vous de sites dans le domaine de l'épistémologie ?
- R4 : Combien trouvez-vous de sites d'écoles doctorales à l'Université Paris Descartes ?
- R5 : Combien trouvez-vous de sites de bibliothèques dans la région du Proche-Orient ?
- R6 : Combien trouvez-vous d'universités éditrices ?

4/ Créer un compte sur le site.

5/ Dans une autre fenêtre de votre navigateur Web, rendez-vous à l'adresse : www.senspublic.org

6/ Ajouter ce site aux ressources du Sociopôle.

Le but de ces manipulations n'était pas de récupérer des réponses exactes aux questions mais de faire explorer le site par les participants, afin d'enrichir par la suite la discussion de groupe.

Questionnaire

Nous avons conseillé aux utilisateurs de prendre quinze minutes pour répondre au questionnaire suivant, sans reparcourir le site. L'idée était de récupérer les éléments qu'ils avaient compris de leur navigation préalable. Voici ces questions (hormis celles servant à décrire leur identité et détaillée dans K) :

Repérage des informations (sans revenir sur le site svp)

1. A quoi sert ce site ?
2. Qui réalise ce site ?
3. Qui peut ajouter et organiser des ressources sur ce site ? Qu'en pensez-vous ?
4. Combien de ressources se trouvent approximativement sur le Sociopôle ?
5. Sur l'interface du Sociopôle, que signifie la couleur jaune ou l' « étoile » ?

Vos impressions

1. Le contenu de ce site vous est-il utile ? Si oui, pour quelles tâches ?

-
2. Trouver des contenus sur le site, est-ce : Très simple, simple, normal, difficile, très difficile
 3. Ajouter des contenus sur le site, est-ce : Très simple, simple, normal, difficile, très difficile
 4. Selon vous, quels sont les aspects les plus intéressants sur le site ?
 5. Quels sont les aspects les plus gênants ?
 6. Que verriez vous d'utile sur une prochaine version du site ?

Discussion de groupe (1 heure)

Après la phase individuelle, nous avons animé une discussion de groupe. Celle-ci fut basée les questions remplies précédemment. L'objectif étant de faire expliciter et approfondir les propos des participants.

La discussion fut riche et tous les participants nous ont semblé s'exprimer.

Résultats

Voici les résultats détaillés par question du formulaire plus les impressions générales émergent de la discussion. Le questionnaire contenait des questions intéressants aussi bien le Sociopôle en tant que sites de ressources en sociologie que SemanticScuttle en tant qu'outil. Nous nous intéresserons davantage aux réponses portant sur SemanticScuttle.

Qui peut ajouter et organiser des ressources sur ce site ? Qu'en pensez-vous ? 8 participants sur 9 ont compris que le site permettait à n'importe quel internaute de créer un compte et d'ajouter des ressources. Ceci est probablement dû au fait que l'activité demandait d'ajouter un signet. Seule le participant 6, âgé de 64 ans et peu habitué aux outils informatiques n'a pas réussi à ajouter un signet et n'a pas su répondre à la question.

La fonctionnalité participative fut en revanche mis en doute par les participants qui craignent le "bruit" provoqué par les participations multiples. La participation des internautes sur le Sociopôle, même encadrée par des utilisateurs certifiés, est synonyme de dérives et de baisses de la qualité du site.

Sur l'interface du Sociopôle, que signifie la couleur jaune ou l' « étoile » ? Le but de cette question était de vérifier si les utilisateurs distinguaient correctement les ressources certifiées par les administrateurs du site (repérables par la bandeau jaune et une étoile), de celles ajoutées et décrites par n'importe quel utilisateur.

7 utilisateurs sur 9 ont compris le sens du bandeau jaune. Le participant 6 qui a eu des difficultés durant l'activité n'a pas su répondre. Ce résultat semble prouver que des utilisateurs moyens arrivent facilement à distinguer les ressources certifiées des autres.

Le contenu de ce site vous est-il utile ? Si oui, pour quelles tâches ? Cs questions ont eu plusieurs ensemble de réponses selon notamment les professions des participants.

Les chercheurs ont trouvé que l'organisation du domaine ne correspondait absolument pas à leur activité habituel. En effet, les signets sont décrits à l'aide de grandes catégories (sites traitant de sociologues, sites de formations universitaires en sociologie, sites de laboratoires) alors que les chercheurs ont besoin de signets organisés par thèmes de sociologie. Les chercheurs ont aussi tenu un discours paradoxal en disant qu'ils préféreraient les moteurs de recherche style Google

pour l'efficacité des recherches, tout en disant qu'ils voudraient bien des sites pour naviguer parmi les ressources à plat. Or les tags de SemanticScuttle offrent justement cette navigation à plat. Ce discours paradoxal se résoud selon nous par le fait que le Sociopôle offre bien une navigation par tags mais que ceci étant, au moment du focus group, uniquement organisés par des documentalistes, ils ne collent pas aux centres d'intérêt des chercheurs.

Les conservateurs ont trouvé un intérêt au site pour les personnes faisant de la veille et pour trouver des ressources étrangères (le Sociopôle couvre les sites francophones et souligne constamment le pays ou le continent d'origine).

Un utilisateur a trouvé un intérêt au site pour des renseignements pratiques. Par exemple, pour trouver des universités. Cet aspect faisait effectivement partie des objectifs du Sociopôle.

Certains utilisateurs ont mentionné l'intérêt par rapport à Google d'organiser le domaine.

Le participant 7, jeune doctorant habitué aux sites comme YouTube, a trouvé un intérêt pour trouver et partager des sites sur le domaine qui l'intéresse.

En conclusion, le Sociopôle dans l'état lors du focus group ne semble pas coller aux besoins des participants, notamment de par les contenus et leur organisation. Le focus group a révélé un décalage important entre les visions des documentalistes et des chercheurs notamment. Pour gagner en intérêt, le site doit alors se concentrer sur des publics et des usages. La participation des usagers semble envisagé par les utilisateurs habitués aux sites Participatifs (généralement les populations plus jeunes). L'organisation des contenus est intéressante simplement si elle apporte une valeur ajoutée à Google. Par exemple, pour des utilisateurs novices dans le domaine ou alors si le site regroupe les signets par thèmes précisément utiles aux visiteurs. Cette précision nécessite alors forcément la participation des visiteurs car les documentalistes n'ont pas les capacités à connaître tous les sous-domaines.

Trouver des contenus sur le site, est-ce : Très simple, simple, normal, difficile, très difficile 8 utilisateurs ont répondu : 1 Simple, 4 Normal, 1 Normal-Difficile, 2 Difficile. Voici la répartition des réponses en fonction des compétences en informatique déclarées :

	Simple	Normal	Plutôt Difficile
Faibles	×	×	1
Moyennes	×	4	×
Elevées	1	×	2

Le participant 6, peu habitués aux outils informatiques, a de manière presque prévisible trouvé plutôt difficile la recherche. Plus surprenantes sont les réponses des participants 3 et 5 aux compétences élevées ayant aussi trouvé les recherches difficiles. Lors de la discussion générale, ils ont alors exprimé les difficultés à effectuer des recherches avancées. Nous faisons donc l'hypothèse que leurs compétences élevées leur ont permis de percevoir les limites de leurs recherches : notamment la difficulté à sélectionner des tags correspondant à plusieurs tags, ou à croiser recherche par tags et par le moteur plein-texte.

Nous retenons néanmoins une majorité d'utilisateurs trouvant la recherche simple ou normale plutôt que difficile (5 contre 3).

Ajouter des contenus sur le site, est-ce : Très simple, simple, normal, difficile, très difficile Durant l'activité, les participants ont du "créer un compte" puis ajouter un signet. La difficulté résidait notamment dans le terme "créer un compte" de l'énoncé qui n'apparaissait pas sur l'interface. Les participants devaient comprendre que ceci correspondait au lien "S'enregistrer", écrit en haut à droite de tous les écrans.

Les participants ont trouvé la création de compte et l'ajout de signets sur le site : 1 Très Simple, 4 Simple, 1 Normal, 1 Normal-Difficile et 2 Difficile.

En agrégeant, l'ajout fut considéré :

- Plutôt Simple par 5 participants
- Normal par 1 participant
- Plutôt Difficile par 3 participants

Les réponses "plutôt difficile" contiennent encore le participant 6 peu habitué à l'outil informatique et un participant avancé qui a pu ajouter un signet mais aurait apprécié une fonctionnalité encore plus simple intégré au navigateur.

En conclusion, la création d'un compte et l'ajout d'un signet est plutôt considéré de manière simple et fut accompli sans aides par 7 participants sur 9. En revanche, seulement 4 utilisateurs sur ces 7 ont décrit le signet avec des tags. La consigne ne donnant aucune obligation de descriptions, nous ne savons pas si les utilisateurs ayant omis les tags le firent volontairement ou s'ils ne savaient pas comment effectuer cette action.

Selon vous, quels sont les aspects les plus intéressants sur le site ? 6 participants mentionnent l'intérêt de partager des contenus. En revanche, ce partage est envisagé différemment selon les compétences informatiques des participants. Ceux peu habitués au Web Participatif parle de "partage d'informations par des internautes", les participants aux compétences plus élevées sont plus précis. Ils parlent de "partage entre chercheurs" et de partage "sur un thème".

3 participants mentionnent l'intérêt d'avoir une participation communautaire intégrée à des ressources validées. Un conservateur de bibliothèque repère l'intérêt d'avoir de nombreuses ressources mais contrôlées.

1 conservateur, aux compétences en informatique élevées, parle de l'intérêt de raffiner des recherches en croisant les tags.

En conclusion, l'aspect de partage offert par le site est bien repéré par la majorité participants. En revanche, la mise en pratique réelle de ce partage (en petites communautés, sur des thèmes précis, avec certification de certaines ressources) ne semble être repéré que par les participants habitués aux outils informatiques ou aux problématiques de documentation (conservateurs).

Quels sont les aspects les plus gênants ? Les points gênants peuvent être regroupés en trois catégories : difficulté d'orientation, difficulté de recherches avancées, remise en cause de la participation des utilisateurs.

Quelques participants ont mentionné plusieurs difficultés ergonomiques dans la navigation globale du site. Ils auraient aimé une aide en ligne et des termes parfois plus précis pour les menus généraux (créer un signet, s'enregistrer).

5 participants sur 9 mentionnent la difficulté à effectuer des recherches avancées comme croiser les tags ou comprendre les menus latéraux qui sont contextuels à la recherche.

1 chercheur mentionne que les signets ne pointent pas, selon lui, vers des sources d'informations scientifiques. Les chercheurs en général mentionne le risque de laisser les internautes participer à l'ajout de ressources.

En conclusion, les remarques ergonomiques nous ont permis d'améliorer la navigation générale dans le site. Nous avons cherché à faciliter la navigation au moyen non pas d'une aide en ligne mais de messages contextuels pour expliquer les composants de l'interface. Les composants permettant des recherches avancées furent aussi améliorés pour les rendre plus accessibles. Selon nous, le risque lié à la participation de tous est limité par la veille des documentalistes. Ces

dernières peuvent éliminer rapidement les contenus sans aucun rapport. Elles peuvent de plus inciter les chercheurs à plus participer et du coup améliorer la qualité globale des contenus.

Que verriez vous d'utile sur une prochaine version du site ? Les principales améliorations demandées par les participants concernent la valoriation et l'enrichissement des thèmes :

- En approfondissement en sous-thèmes. Par exemple : inclure dans le tag "master", les sous-tags "pro" et "recherche".
- En ajoutant des terminologies officielles du domaine.
- En ajoutant des thèmes proches employés par les chercheurs.

Un chercheur propose aussi que les signets portent davantage sur des ressources bibliographiques, pour être plus utile aux chercheurs.

Synthèse Techniquement, l'outil permet sans difficultés des recherches simples et l'ajout de signets.

Les recherches avancées (comme croiser des tags) demandent à être améliorées ergonomiquement. Des messages d'aides contextuels (tooltip) nous semblent notamment pertinents.

Le plus gros travail à court terme semble être l'amélioration des structures et des tags des documentalistes pour plus coller aux termes et structures des publics visés (particulièrement les chercheurs).

Justement, l'aspect participatif de SemanticScuttle semble pertinent pour effectuer le rapprochement entre les descriptions et contenus des documentalistes avec celles des chercheurs. Néanmoins, cette participation externe demande une veille des documentalistes pour éviter les contenus hors de propos.

L

Expérimentation avec utilisateurs du Mechanical Turk

Le site Mechanical Turk est une place de marché en ligne où d'un côté des visiteurs proposent des tâches sur le Web, rapides à effectuer, et d'un autre des visiteurs les accomplissent pour quelques cents la tâche. Pour tester SemanticScuttle, nous avons ainsi proposé 10 tâches répondant à l'énoncé suivant :

Description de la tâche

Experiment a social bookmarking tool and tag 10 pedagogical resources

We would like to experiment a social bookmarking tool (called SemanticScuttle) and its possibilities to tag and describe pedagogical resources. Summary : You will create an account, search for 10 resources and add them to the social bookmarking tool.

Attention : Tasks will be rejected if you add less than 10 resources or if they are not about teaching or learning.

Your profile

1. What is your level with a computer ?

Low - Average - High

Instructions

2. Go to the social bookmarking tool experiment : <http://festic.fr/semanticscuttle/demo/mturk/>

3. "Register" with the top-right link. (You are logged and can see your possible actions at the top of the screen).

What is your chosen nickname? ____

4. Click on the upper link "add a bookmark". (This is the page used to add and describe the resources into the tool. At the bottom of the page, you have "bookmarklets (pop-up)" that you can drag'n drop into your browser to speed up your actions. When clicking on a bookmarklet,

you can directly add and describe a website.)

5. In a new window of your computer, open your favorite search engine (e.g. Google).

6. In this new window, search a pedagogical resource about Computer Science in English.

7. Add the resource to the bookmarking tool. The most important field is the tags' one. The tool allows to add links between tags. For example, if you write "europe>france" then the tool will add a link of inclusion between "europe" and "france"; later, when looking for "europe", the tool will also return resources described with "france". Please create links between tags because we observe the emerging structures of tags.

8. Do again, steps 6 and 7, X 9 to obtain finally 10 resources into the bookmarking tool. ATTENTION : Please add various pedagogical resources (from different websites, different sizes, different types, ...).

9. Go to your "tags" page (upper link) and manage your linked tags (right menu) to obtain a consistent structure of linked tags.

Your feelings

8. According to you, what is the major point we need to improve into the tool? ____

Thank you for your participation ;o)

For your information, on your "tag" page (upper-link) you can see all your tags and "links" between tags. On the home page, you can see the "linked tags" from every users.

Please provide any comments you may have below, we appreciate your input! ____

M

Nos contributions par questions

Questions de départ

Nos travaux avaient pour but d'explorer et de répondre aux questions suivantes :

Questions centrales :

- Comment les techniques du Web Sémantique peuvent-elles améliorer le partage de documents pédagogiques ?
- Comment la tendance du Web Participatif peut-elle améliorer le partage de documents pédagogiques ?
- Comment les approches du Web Sémantique et du Web Participatif, qui semblent contradictoires dans leurs usages, peuvent-elles se compléter ?

Question méthodologique :

- Comment tester les approches liées au partage de documents pour obtenir des solutions validées expérimentalement et réutilisables ?

Voici un récapitulatif de nos réponses.

Comment les techniques du Web Sémantique peuvent-elles améliorer le partage de documents pédagogiques ?

Apports

A travers le projet européen de Recherche & Développement LUISA, nos travaux illustrent les possibilités offertes par des ontologies pour la recherche de ressources pédagogiques.

Fonctionnalités sémantiques. Parmi les utilisations directes des ontologies, les prototypes, dont **nous avons réalisé les spécifications**, permettent notamment :

- de décomposer les requêtes en compétences et sous-compétences pour améliorer la recherche de ressources ;
- de trouver des compétences similaires dans différentes ontologies issues de différents contextes ;

- de proposer des paquets de ressources composées selon des règles pédagogiques simples.

Fonctionnalités sémantiques impliquant les humains. Parmi des utilisations plus originales des ontologies, les prototypes permettent :

- aux apprenants d’interagir avec une représentation de l’ontologie de compétences dans le but de soutenir leur compréhension de l’ontologie ;
- aux apprenants de noter les ressources puis d’intégrer ces notes aux raisonnements sémantiques ;
- de visualiser les concepts employés lors des raisonnements sémantiques effectués par le moteur de recherche, dans le but de soutenir la compréhension du domaine par les utilisateurs.

Limites

Ces apports sont néanmoins limités par différentes difficultés apparues lors du projet.

Lenteur des raisonnements distribués. D’un point de vue technique et dans un cadre distribué, des raisonnements sémantiques provoquent de nombreuses interactions entre Web Services et du coup entraînent des temps de latence importants lors des recherches. Ces temps correspondent à une quarantaine de secondes pour nos prototypes non optimisés mais appliquant des raisonnements assez simples.

Manipulation et appréhension difficiles. D’un point de vue humain, les ontologies sont des objets complexes à manipuler et à appréhender.

La difficulté de manipulation est apparue lorsque nous avons par exemple demandé des évolutions des ontologies aux partenaires. Les modifications des concepts ou des raisonnements sont laborieuses dès que plusieurs individus issus de contextes différents sont impliqués.

La difficulté d’appréhension est apparue lorsque nous avons recherché un contexte d’expérimentation au sein de notre université (**entretiens et séances de remues-méninges**). Les enseignants, non informaticiens, ont eu du mal à comprendre la notion d’ontologies et encore plus à formaliser un tant soit peu des raisonnements pédagogiques que pourraient appliquer des systèmes automatisés. Pour être diffusés dans le contexte des enseignants, les ontologies pédagogiques, en tant que visions partagées d’un domaine, demanderont probablement un effort important de communication et de formation.

Comment la tendance du Web Participatif peut-elle améliorer le partage de documents pédagogiques ?

Apports

Fonctionnalités participatives. Notre étude des sites du Web Participatif à succès, nous a permis de dégager un ensemble de fonctionnalités pouvant soutenir la participation au partage. Ces fonctionnalités se classent selon trois catégories : ergonomique, sociale et pécunière.

Proposition d'un modèle de YouTube pédagogique. Sur les principes précédents et suite à l'étude des sites de différentes communautés d'enseignants, nous avons **proposé le modèle** d'un site libérant ces communautés de la lourde gestion des ressources en ligne. Sur le modèle d'un YouTube pédagogique, ce site viserait à faciliter le dépôt de ressources pédagogiques. Il proposerait aussi différents services pour les enseignants comme suivre l'évolution d'une ressource pédagogique, rechercher selon des critères pédagogiques ou afficher les ressources sur les sites des communautés enseignantes.

Limites

Difficultés « politiques ». Partager des ressources ludiques n'est pas similaire à partager des ressources professionnelles. Les enseignants attachent de l'importance aux sites où ils déposent leurs ressources. De plus, la répartition du pouvoir au sein des communautés est interdépendant des outils qu'elles emploient. Ces critères font que proposer un outil de gestion de ressources pédagogiques ne peut se faire que difficilement sans une certaine légitimité vis-à-vis des enseignants et des communautés. Or cette légitimité accompagnant l'outil est une contrainte forte, requérant une légitimité préalable ou alors beaucoup de moyens pour être créé à partir de rien.

Difficultés techniques. Bien que non suffisant, l'aspect ergonomique est nécessaire pour rendre un site facile d'utilisation et donc permettre la participation du plus grand nombre. Si l'on ne possède pas un vivier d'utilisateurs « captifs » ou très impliqués, il faut donc développer une application ergonomique requérant des moyens importants, ou alors basée sur une application pré-existante déjà rodée.

Comment les approches du Web Sémantique et du Web Participatif, qui semblent contradictoires dans leurs usages, peuvent-elles se compléter ?

Apports

Modèle IPM. Suite à l'étude des différentes solutions reliant sémantique et participatif, nous avons proposé un **modèle d'Indexation Progressive et Multi-Points de vue**. Ce modèle d'application vise à faciliter la participation des utilisateurs tout en soutenant une structuration progressive des métadonnées et une collaboration progressive entre utilisateurs.

Implémentation dans un outil diffusé. Nous avons prouvé la faisabilité technique du modèle IPM en l'implémentant dans un outils de partage de signets. Cet outil est librement disponible¹²² et a été **téléchargé plus de 150 fois par mois depuis un an** et une petite communauté internationale s'est formée soumettant régulièrement questions, réponses et bouts de code.

Expérimentation en situation réelle. Nous avons installé l'outil pour des documentalistes que nous avons ensuite **accompagnés dans la modification de leurs pratiques** durant un

¹²²Projet SemanticScuttle : <http://sourceforge.net/projects/semanticscuttle/> et démo : <http://festic.fr/semanticscuttle/demo/fr>

semestre. Ces travaux confirment le besoin d'outils mêlant participation d'experts et de non-experts, mêlant structuration et flexibilité. L'outil que nous avons développé est aujourd'hui recommandé par les documentalistes à leurs collègues.

Limites

Justifier la participation ? L'expérimentation avec les documentalistes nous a montré qu'un site participatif a peu d'intérêt si les visiteurs ne se reconnaissent pas dans les contenus et s'ils ne voient pas d'intérêt à participer.

Comme nous allons l'expérimenter dans des **travaux futurs avec les documentalistes**, les solutions possibles sont une meilleure détermination du public visé par l'outil, et des collaborations ponctuels avec certains utilisateurs pour ajouter des ressources et amorcer la participation.

Généraliser aux enseignants ? Notre expérimentation principale a porté sur des documentalistes ; nos résultats sont donc incomplets sur la possibilité d'un outil pour les enseignants.

Néanmoins, certaines observations montrent un rapprochement entre documentalistes et enseignants. D'un côté, les communautés d'enseignants semblent avoir du mal à employer de manière pérenne des outils de partage structurés ; une solution serait le soutien de spécialistes de la documentation. D'un autre côté, les documentalistes ressentent le besoin de faire participer les visiteurs pour enrichir l'organisation des ressources. Entre ces deux besoins, le modèle d'Indexation Progressive et Multi-points de vue offre une solution flexible et cohérente.

Comment tester les approches liées au partage de documents pour obtenir des solutions validées expérimentalement et réutilisables ?

Proximité du terrain. Durant notre thèse, nous avons passé beaucoup de temps à chercher des terrains pour expérimenter. Trouver des utilisateurs intéressés nous semble indispensable pour réaliser des outils qui auront une chance d'être expérimentés en situation réelle par la suite.

Dans notre thèse, nous avons appliqué ce principe en stoppant le projet Pépi et le travail peu fructueux avec les enseignants de génie mécanique. Cette décision fut difficile mais fut ensuite récompensée par l'expérimentation de SemanticScuttle sur plusieurs mois.

Outils en tant que véhicules des idées. Les usages sur Internet évoluent à grande vitesse, à l'image des réseaux sociaux ou du microblogging qui n'existaient pas il y a trois ans. Les travaux de recherche permettent de mieux comprendre ces usages. Néanmoins comment faire pour que ces connaissances puissent un jour être utiles aux utilisateurs ?

De notre point de vue, l'outil n'est pas juste un moyen de « recueillir » des connaissances mais aussi de soutenir de nouvelles pratiques. Durant la conception, la participation des utilisateurs ne permet pas juste d'avoir un outil adapté à leurs besoins mais les soutient aussi dans une réflexion plus profonde sur les concepts sous-jacents. Durant la diffusion, étape souvent oubliée, les utilisateurs s'approprient l'outil et ses concepts sous-jacents en fonction de leurs contextes. De plus, entre utilisateurs, l'outil sert d'objet-support pour discuter, réfléchir et éventuellement diffuser de nouvelles pratiques, presque indépendamment du chercheur. Plus qu'une conception

participative, nous entrevoyons là peut-être une forme de Recherche-Action¹²³, nécessaire quand les usages dans un domaine évoluent très vite.

Dans notre thèse, nous avons appliqué cette approche dans le projet SemanticScuttle.

D'un côté, l'outil a été développé avec des documentalistes. Il fut mis très tôt en ligne¹²⁴ pour permettre des commentaires suite à un usage en situation réelle. Il fut aussi le sujet de plusieurs réunions entre documentalistes indépendamment de notre présence. Enfin, celles-ci communiquent autour de l'outil avec leurs collègues, par exemple en réalisant un poster pour une rencontre nationale dans leur domaine.

D'un autre côté, l'outil est un projet libre et diffusé, téléchargé plus de 150 fois par mois en moyenne depuis un an. En employant le logiciel, les utilisateurs manipulent ainsi le modèle sous-jacent. En complément aux publications scientifiques diffusant les idées sur des durées moyennes, l'outil diffuse les idées de manière « plus rapide » et directement auprès des utilisateurs mais sous une forme moins explicite.

¹²³Introduction à la Recherche-Action : <http://recherche-action.fr/pages/portail/introduction-a-la-recherche-action.php>

¹²⁴Sociopôle, ressources francophones en sociologie : <http://www.sociopole.cnrs.fr/>

Glossaire

Glossaire de quelques termes employés dans ce mémoire.

C2I : Le C2I (Certificat Informatique et Internet) est un certificat obligatoire au niveau Licence dans toutes les universités françaises. Le contexte de ce certificat, et les documents qui l’y préparent, servent de terrain d’expérimentation au projet LUISA (§ 3).

FOAF : FOAF (Friend Of A Friend) est une ontologie permettant notamment de décrire les liens sociaux entre des individus : Jean est l’ami de Véronique, qui est la collègue de Pierre, qui est de la même famille que Bernard.

LCMS : Les LCMS (Learning Content Management System) sont des plateformes d’enseignement et d’apprentissage en ligne. Les enseignants peuvent y créer des groupes d’apprenants et y déposer des documents de cours. Il peuvent aussi y configurer des outils tels que des forums, des chats, des QCM pour faire travailler les apprenants. Un des LCMS les plus connus s’appellent Moodle. Il fut employé dans le projet LUISA (§ 3) pour y intégrer un moteur de recherche à base de Web Services Sémantiques.

LOM : LOM (Learning Object Metadata) est un schéma de métadonnées pour les ressources pédagogiques. Ce standard de l’organisation IEEE propose des champs et des valeurs pour décrire des contenus pédagogiques. L’objectif est de les décrire de manière identique pour faciliter par la suite leur recherche. Dans le projet LUISA (§ 3), l’ontologie S-LOM est basé sur LOM.

Moodle : Voir LCMS.

SIOC : SIOC (Semantically-Interlinked Online Communities) est une ontologie permettant de représenter le contenu des sites participatifs : communautés, utilisateurs, messages, forums, tags... SIOC est présentée en §6.1.3.4 et employée dans le projet SemanticScuttle pour exporter le contenu sous forme sémantique (§7.2.2.3).

SKOS : SKOS (Simple Knowledge Ontology System) est une ontologie permettant de représenter des systèmes de connaissances comme des thésaurus : termes, concepts...SKOS est présentée en §6.1.3.3 et employée dans le projet SemanticScuttle pour exporter le contenu sous forme sémantique (§7.2.2.3).

Bibliographie

Les numéros apparaissant à la fin des références correspondent aux pages de ce mémoire où ces références sont citées.

- [Al-Khalifa and Davis, 2007] Al-Khalifa, H. S. and Davis, H. C. (2007). Replacing the monolithic LOM : a folksonomic approach. In *Proceedings of ICALT 2007*, pages 665–669, Nigata, Japan. 89, 135
- [Angeletou et al., 2007] Angeletou, S., Sabou, M., Specia, L., and Motta, E. (2007). Bridging the gap between folksonomies and the semantic web : An experience report. In *ESWC workshop. Bridging the Gap between Semantic Web and Web 2.0*. 126
- [Aroyo and Dicheva, 2004] Aroyo, L. and Dicheva, D. (2004). The new challenges for e-learning : The educational semantic web. *Educational Technology & Society*, 7(4) :59–69. 32
- [Au et al., 2007] Au, Gibbins, N., and Hadbolt, N. (2007). Understanding the semantics of ambiguous tags in folksonomies. In *The International Workshop on Emergent Semantics and Ontology Evolution (ESOE2007) at ISWC/ASWC 2007*. 127
- [Barab et al., 2001] Barab, S., MaKinster, J., Moore, J., Cunningham, D., and Team, I. D. (2001). Designing and building an online community : The struggle to support sociability in the inquiry learning forum. *Educational Technology Research and Development*, 49(4) :71–96. 92
- [Bateman et al., 2006] Bateman, S., Brooks, C., and McCalla, G. (2006). Collaborative tagging approaches for ontological metadata in adaptive ELearning systems. In *Proceedings of SW-EL'06@AH'06*, pages 3–12, Dublin, Ireland. 135
- [Bateman et al., 2007] Bateman, S., Brooks, C., McCalla, G., and Brusilovsky, P. (2007). Applying collaborative tagging to E-Learning. In *Proceedings of the Workshop on Tagging and Metadata for Social Information Organization, held in conjunction with the 16th International World Wide Web Conference*, Banff, Canada. 90
- [Begelman et al., 2006] Begelman, G., Keller, P., and Smadja, F. (2006). Automated tag clustering : Improving search and exploration in the tag space. In *World Wide Web Conference 2006*, Edinburgh, Scotland. 127
- [Berners-Lee et al., 2001] Berners-Lee, T., Hendler, J., and Lassila, O. (2001). The semantic web. *Scientific American Magazine*. 23, 187
- [Braun et al., 2007] Braun, S., Schmidt, A., Walter, A., Nagypal, G., and Zacharias, V. (2007). Ontology maturing : a collaborative web 2.0 approach to ontology engineering. In *Workshop on Social and Collaborative Construction of Structured Knowledge (CKC), 16th International World Wide Web Conference (WWW 2007)*, Banff, Alberata, Canada. 71, 131, 132

- [Brito-Mirian et al., 2006] Brito-Mirian, C., da Nobrega, G., and de Oliveira, K. (2006). Integrating instructional material and teaching experience into a teachers collaborative learning environment. In *First European Conference on Technology Enhanced Learning (EC-TEL 2006)*, pages 458–463, Crete, Greece. 92
- [Buffa et al., 2005] Buffa, M., Dehors, S., Faron-Zucker, C., and Sander, P. (2005). Towards a corporate semantic web approach in designing learning systems : Review of the TRIAL SOLUTION project. In *Proceedings of International Workshop on Applications of Semantic Web Technologies for E-Learning, AIED'2005*, Amsterdam. 14
- [Buffa and Gandon, 2006] Buffa, M. and Gandon, F. (2006). SweetWiki : semantic web enabled technologies in wiki. In *Proceedings of the 2006 international symposium on Wikis*, pages 69–78, Odense, Denmark. ACM. 130, 133
- [Béziat and Caron, 2003] Béziat, J. and Caron, C. (2003). Le campus numérique FORSE et ses tuteurs. In *2ème colloque de Guéret*. http://www.pedagogies.net/Colloque03/Colloque03_Programme/Programme/Resume/At16.html. 92
- [Cahier, 2005] Cahier, J. (2005). *Ontologies sémiotiques pour le Web socio sémantique*. PhD thesis, Université de Technologie de Troyes. 132
- [Catteau, 2008] Catteau, O. (2008). *Le cycle de vie de l'objet pédagogique et de ses métadonnées*. PhD thesis, Université de Toulouse. 14, 15
- [Cheng and Vassileva, 2006] Cheng, R. and Vassileva, J. (2006). Design and evaluation of an adaptive incentive mechanism for sustained educational online communities. *User Modeling and User-Adapted Interaction*, 16(3) :321–348. 91
- [Corby et al., 2004] Corby, O., Dieng, R., and Faron-Zucker, C. (2004). Querying the semantic web with corese search engine. In *Proceedings of ECAI 2004*, Valencia. IOS Press. 30
- [Daele, 2005] Daele, A. (2005). Développement professionnel des enseignants dans un contexte de participation à une communauté virtuelle : une étude exploratoire. In *Symposium SYMFONIC (SYMposium, FORMation et Nouveaux Instruments de Communication)*. 92
- [Dehors, 2007] Dehors, S. (2007). *Exploiting Semantic Web and Knowledge Management Technologies for E-learning*. PhD thesis, Université de Nice Sophia Antipolis. 14, 27, 32
- [Devedzic, 2004] Devedzic, V. (2004). Education and the semantic web. *International Journal of Artificial Intelligence in Education*, 14(2) :165–191. 31
- [Diederich et al., 2007] Diederich, J., Balke, W., and Thaden, U. (2007). Demonstrating the semantic growbag : automatically creating topic facets for faceteddblp. In *Proceedings of the 7th ACM/IEEE-CS joint conference on Digital libraries*, pages 505–505, Vancouver, BC, Canada. ACM. 127, 128
- [Downes, 2004] Downes, S. (2004). Resource profiles. *Journal of Interactive Media in Education*, 2004(5). 32
- [Downes, 2005] Downes, S. (2005). E-learning 2.0. *eLearn Magazine*, 2005(10). <http://elearn-mag.org/subpage.cfm?section=articles&article=29-1>. 86
- [Duval et al., 2002] Duval, E., Hodgins, W., Sutton, S., and Weibel, S. L. (2002). Metadata principles and practicalities. *D-Lib Magazine*, 8(4). 16
- [Dzbor et al., 2005] Dzbor, M., Motta, E., and Stutt, A. (2005). Achieving higher-level learning through adaptable semantic web applications. *International Journal of Knowledge and Learning*, 1(1/2) :25 – 43. 30

-
- [Endrizzi, 2006] Endrizzi, L. (2006). L'édition de référence libre et collaborative : le cas de wikipedia. Technical report, INRP. <http://www.inrp.fr/vst/Dossiers/Wikipedia/sommaire.htm>. 81, 88
- [Friesen, 2004a] Friesen, N. (2004a). International LOM survey : Report. Technical report, ISO/IEC JTC1/SC36. http://mdlet.jtc1sc36.org/doc/SC36_WG4_N0109.pdf. 16
- [Friesen, 2004b] Friesen, N. (2004b). Three objections to learning objects and e-learning standards. In *Online Education Using Learning Objects*, pages 53–63. McGReal, R., London, routledge falmer edition. 14, 17
- [Gandon, 2005] Gandon, F. (2005). Generating surrogates to make the semantic web intelligible to end-users. In *Proceedings of the IEEE/WIC/ACM International Conference on Web Intelligence 2005*, pages 352–358. 29
- [Gandon, 2006] Gandon, F. (2006). Le web sémantique n'est pas antisocial. In *Proceedings Ingénierie des Connaissances*, pages 131–140, Nantes. 187
- [Garrot, 2008] Garrot, E. (2008). *Plate-forme support à l'Interconnexion de Communautés de Pratique (ICP). Application au tutorat avec TE-Cap*. PhD thesis, INSA de Lyon. 9, 92, 99, 111
- [Gasevic and Hatala, 2006] Gasevic, D. and Hatala, M. (2006). Ontology mappings to improve learning resource search. *British journal of educational technology*, 37(3) :375–389. 56
- [Ghali et al., 2009] Ghali, A. E., Tifous, A., Buffa, M., Giboin, A., and Dieng-Kuntz, R. (2009). Using a semantic wiki in communities of practice. In *Proceedings of the 2nd International Workshop on Building Technology Enhanced Learning solutions for Communities of Practice*, Crete, Greece. 134
- [Giboin et al., 2008] Giboin, A., Priscille, D., and Gandon, F. (2008). Ingénierie ontologique participative : essai de mise en oeuvre avec l'éditeur collaboratif ECCO. In *Atelier IC2.0 joint à IC2008, 19èmes Journées Francophones d'Ingénierie des Connaissances*, Nancy. 130
- [Gracia et al., 2006] Gracia, J., Trillo, R., Espinoza, M., and Mena, E. (2006). Querying the web : a multiontology disambiguation method. In *Proceedings of the 6th international conference on Web engineering*, pages 241–248, Palo Alto, California, USA. ACM. 126
- [Gruber, 2007] Gruber, T. (2007). Ontology of folksonomy : A mash-up of apples and oranges. *International Journal on Semantic Web & Information Systems*, 3(2). 123
- [Guy and Tonkin, 2006] Guy, M. and Tonkin, E. (2006). Folksonomies : Tidying up tags? *D-Lib Magazine*, 12(1). 121
- [Kalfoglou and Schorlemmer, 2003] Kalfoglou, Y. and Schorlemmer, M. (2003). Ontology mapping : the state of the art. *Knowl. Eng. Rev.*, 18(1) :1–31. 56
- [Karacapilidis and Tzagarakis, 2007] Karacapilidis, N. and Tzagarakis, M. (2007). Web-based collaboration and decision making support : A multi-disciplinary approach. *International Journal of Web-Based Learning and Teaching Technologies*, 2(4) :12–23. 92
- [Karapiperis and Apostolou, 2006] Karapiperis, S. and Apostolou, D. (2006). Consensus building in collaborative ontology engineering processes. *Journal of Universal Knowledge Management*, 1(3) :199–216. 71, 130, 131
- [Koper and Tattersall, 2005] Koper, R. and Tattersall, C. (2005). *Learning Design : A Handbook on Modelling and Delivering Networked Education and Training*. Springer-Verlag New York, Inc. 17

- [Krötzsch et al., 2006] Krötzsch, M., Vrandečić, D., and Völkel, M. (2006). Semantic MediaWiki. In *The Semantic Web - ISWC 2006*, volume 4273/2006 of *Lecture Notes in Computer Science*, pages 935–942. Springer, Berlin / Heidelberg. 130
- [Lefoe et al., 2002] Lefoe, G., Hedberg, J., and Gunn, C. (2002). The changing role of tutors : Forming a community of practice in a distributed learning environment. In *Proceedings of International Conference on Computers in Education (ICCE'02)*, pages 729–733. 92
- [Libbrecht et al., 2008] Libbrecht, P., Desmoulins, C., Mercat, C., Laborde, C., Dietrich, M., and Hendriks, M. (2008). Cross-Curriculum search for intergeo. In *Intelligent Computer Mathematics*, volume 5144/2008 of *Lecture Notes in Computer Science*, pages 520–535. Springer-Verlag, Berlin / Heidelberg. 72
- [LUISA-Consortium, 2005] LUISA-Consortium (2005). Description of work for LUISA project. Technical report, LUISA-Consortium. 36
- [LUISA-Consortium, 2007] LUISA-Consortium (2007). A LOM-based ontology of learning objects in WSML. Technical report, LUISA-Consortium. 63
- [Makni et al., 2008] Makni, B., Khelif, K., Dieng-Kuntz, R., and Cherfi, H. (2008). Utilisation du web sémantique pour la gestion d'une liste de diffusion d'une CoP. In *8èmes journées francophones "Extraction et Gestion des Connaissances" (EGC'2008)*, INRIA Sophia Antipolis. 92
- [Marchetti et al., 2007] Marchetti, A., Tesconi, M., and Ronzano, F. (2007). SemKey : a semantic collaborative tagging system. In *WWW07 Workshop, Tagging and Metadata for Social Information Organization*. 128
- [McCalla, 2004] McCalla, G. (2004). The ecological approach to the design of E-Learning environments : Purpose-based capture and use of information about learners. *Journal of Interactive Media in Education*, 2004(7). 32
- [Mika, 2007] Mika, P. (2007). *Social networks and the Semantic Web*. Springer, New York. 6, 126
- [Mizoguchi et al., 2007] Mizoguchi, R., Hayashi, Y., and Bourdeau, J. (2007). Theory-Aware and Standards-Compliant authoring system. In *Fifth international Workshop of Ontologies and Semantic Web Services for E-Learning (SWEL 2007)*, USA. 31
- [Naeve, 2001] Naeve, A. (2001). The concept browser a new form of knowledge management tool. In *Proceedings of the 2 nd European Web-based Learning Environments Conference (WBLE)*, Lund, Sweden. 30, 32
- [Nejdl et al., 2002] Nejdl, W., Wolf, B., Qu, C., Decker, S., Sintek, M., Naeve, A., Nilsson, M., Palmér, M., and Risch, T. (2002). EDUTELLA : a P2P networking infrastructure based on RDF. In *Proceedings of WWW2002*, Honolulu, Hawaii, USA. 88
- [Novak and Canas, 2008] Novak, J. and Canas, A. (2008). The theory underlying concept maps and how to construct them. Technical Report IHMC CmapTools 2006-01 Rev 01-2008, Florida Institute for Human and Machine Cognition. 29
- [Oberle et al., 2004] Oberle, D., Volz, R., Motik, B., and Staab, S. (2004). An extensible ontology software environment. In Staab, S. and Studer, R., editors, *Handbook on Ontologies*, International Handbooks on Information Systems, pages 311–333. Springer. 30
- [O'Reilly, 2005] O'Reilly, T. (2005). What is web 2.0. <http://www.oreilly.com/pub/a/oreilly/tim/news/2005/09/30/what-is-web-20.html?page=5>. 78

-
- [Pashnyak and Dennen, 2007] Pashnyak, T. and Dennen, V. (2007). What and why do classroom teachers blog? In *IADIS Web Based Communities Conference*, pages 172–178, Salamanca, Spain. 92
- [Passant, 2007] Passant, A. (2007). Using ontologies to strengthen folksonomies and enrich information retrieval in weblogs : Theoretical background and corporate use-case. In *International Conference on Weblogs and Social Media*, Boulder, Colorado, USA. 121
- [Passant, 2009] Passant, A. (2009). *Technologies du Web Sémantique pour l'Entreprise 2.0*. PhD thesis, Université Paris-IV Sorbonne. 130, 197
- [Passant and Laublet, 2008] Passant, A. and Laublet, P. (2008). Meaning of a tag : A collaborative approach to bridge the gap between tagging and linked data. In *Proc. of The Linked Data on the Web Workshop of the World Wide Web Conference 08*, Beijing, China. 128, 177
- [Peyrelong and Follet, 2004] Peyrelong, M. and Follet, M. (2004). Le tuteur, entre écrit et chuchotements. In *Conférence TICE Méditerranée*, Nice. 92
- [Recker et al., 2000] Recker, M. M., Walker, A., and Wiley, D. A. (2000). Collaboratively filtering learning objects. In *The Instructional Use of Learning Objects : Online Version*. D. A. Wiley. recker2000cfl. 89, 90
- [Reigeluth, 1999] Reigeluth, C. M. (1999). *Instructional-Design Theories and Models : A New Paradigm of Instructional Theory, Vol. 2*. Lawrence Erlbaum, 1 edition. 59
- [Schlager and Schank, 1997] Schlager, M. and Schank, P. (1997). TAPPED IN : a new online teacher community concept for the next generation of internet technology. In *Second International Conference on Computer Support for Collaborative Learning*, pages 231–240, Toronto, Canada. 92
- [Schuler and Namioka, 1993] Schuler, D. and Namioka, A. (1993). *Participatory design*. Lawrence Erlbaum Associates.
- [Sen et al., 2006] Sen, S., Lam, S. K., Rashid, A. M., Cosley, D., Frankowski, D., Osterhouse, J., Harper, F. M., and Riedl, J. (2006). tagging, communities, vocabulary, evolution. In *Proceedings of the 2006 20th anniversary conference on Computer supported cooperative work*, pages 181–190, Banff, Alberta, Canada. ACM. 121, 129
- [Sherer et al., 2007] Sherer, P., Shea, T., and Kristensen, E. (2007). Online communities of practice : A catalyst for faculty development. *Innovative Higher Education*, 27(3) :183–194. 92
- [Shirky, 2003] Shirky, C. (2003). The semantic web, syllogism, and worldview. http://www.shirky.com/writings/semantic_syllogism.html. 120
- [Shirky, 2005] Shirky, C. (2005). Ontology is overrated : Categories, links, and tags. http://www.shirky.com/writings/ontology_overrated.html. 120
- [Sicilia, 2005] Sicilia, M. (2005). Ontology-based competency management : Infrastructures for the knowledge-intensive learning organization. In *Intelligent learning infrastructures in knowledge intensive organizations : A semantic web perspective*, pages 302–324. IGI Publishing ; 64, 65, 66
- [Sicilia, 2007] Sicilia, M. (2007). On the general structure of ontologies of instructional models. In *Post-Proceedings del IV Simposio Pluridisciplinar sobre Diseño, Evaluación y Desarrollo de Contenidos Educativos Reutilizables*, Bilbao, Espagne. 59

- [Simpson and Butler, 2008] Simpson, E. and Butler, M. H. (2008). Analyzing communal tag relationships for enhanced navigation and user modeling. In *Collaborative and Social Information Retrieval and Access : Techniques for Improved User Modeling*, page 390. M. Chevalier, C. Julien, C.Soule-Dupuy, information science reference edition. 127
- [Stojanovic et al., 2001] Stojanovic, L., Staab, S., and Studer, R. (2001). Elearning based on the semantic web. In *WebNet2001 - World Conference on the WWW and Internet*, pages 23—27. 31
- [Teevan et al., 2004] Teevan, J., Alvarado, C., Ackerman, M. S., and Karger, D. R. (2004). The perfect search engine is not enough : a study of orienteering behavior in directed search. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 415–422, Vienna, Austria. ACM. 58
- [Tesconi et al., 2008] Tesconi, M., Ronzano, F., Marchetti, A., and Minutoli, S. (2008). Semantify delicio.us : Automatically turn your tags into senses. In *Proceedings of First Workshop on Social Data on the Web (SDoW2008)*, Karlsruhe, Allemagne. 126, 128
- [Uren et al., 2006] Uren, V., Cimiano, P., Iria, J., Handschuh, S., Vargas-Vera, M., Motta, E., and Ciravegna, F. (2006). Semantic annotation for knowledge management : Requirements and a survey of the state of the art. *Web Semantics : Science, Services and Agents on the World Wide Web*, 4(1) :28, 14. 28
- [Weinberger, 2007] Weinberger, D. (2007). *Everything Is Miscellaneous : The Power of the New Digital Disorder*. Times Books. 11, 12
- [Wenger, 1998] Wenger, E. (1998). Communities of practice : Learning as a social system. *The Systems Thinker*, 9(5). 9, 99
- [Yessad et al., 2008] Yessad, A., Faron-Zucker, C., Dieng-Kuntz, R., and Laskri, M. (2008). Adaptive course generation in semantic web context. In *Proceedings of World Conference on Educational Multimedia, Hypermedia and Telecommunications*, Chesapeake, VA. 32
- [Zacklad, 2005] Zacklad, M. (2005). Vers le web socio sémantique : introduction aux ontologies sémiotiques. In *Actes IC'2005*, Nice. 6, 132
- [Zouaq and Nkambou, 2008] Zouaq, A. and Nkambou, R. (2008). Building domain ontologies from text for educational purposes. *Learning Technologies, IEEE Transactions on*, 1(1) :49–62. 202

Abstract

Current ways for Web indexing are not sufficient for learning resources. Indeed, automatic indexing, e.g. Google, can hardly raise above the syntactical level of contents while indexing by human experts implies high costs.

But, recent approaches like Semantic Web and Participative Web (Web 2.0) offer promising solutions.

A first part of our works concerns the study of Semantic Web applied to learning resources. We explore possibilities of reasonings applied to educational ontologies.

A second part is about the study of functionalities on participative websites making easier the adding of content and metadata by visitors. Then we propose a model of participative website adapted to communities of teachers.

Nevertheless, Semantic Web and Participative Web are often opposed. Formal ontologies generally produced by experts are opposed to heterogeneous tags added by numerous users with various profiles.

In a third part, we propose a model melting Semantic and Participative. The goal of this model is to help developing applications for sharing resources into communities of practice. It is based on a Progressive and Multi-points of view Indexing in which :

1. users progressively structure metadata, to finally allow semantic reasonings by computers ;
2. users progressively collaborate, to finally allow a shared vision of the domain by humans.

This model is implemented into a social bookmarking tool, called SemanticScuttle, offering original features like tags structured by relations of inclusion and synonymy, or wiki spaces to describe tags. The tool was developed and tested with librarians in sociology during several months.

Finally, our works allow us to formulate hypotheses about a social and technical model supporting sharing between teachers. They also contribute to models melting different indexing solutions : automatic or by humans, including experts or simple users, based on structured models (e.g. ontologies) or on flexible metadata (e.g. tags).

Keywords: Indexing (documentation), Pedagogy – Internet resources, Web 2.0, Online communities, Semantic Web, Ontologies (computer science).

Résumé

Les techniques actuelles d'indexation sur le Web ne sont pas satisfaisantes pour les ressources pédagogiques. En effet, l'indexation automatique, e.g. Google, peut difficilement dépasser le niveau syntaxique des contenus tandis que l'indexation par des documentalistes est coûteuse en main d'oeuvre.

Or de récentes approches telles que les techniques du Web Sémantique ou la tendance du Web Participatif (Web 2.0) offrent des solutions prometteuses.

Une première partie de nos travaux porte ainsi sur l'étude du Web Sémantique appliqué aux ressources pédagogiques. Nous y explorons les possibilités de raisonnements automatisés appliqués à des ontologies pédagogiques.

Une seconde partie porte sur l'étude des fonctionnalités des sites participatifs, facilitant l'ajout de contenus et de métadonnées par les visiteurs. Nous proposons alors un modèle de site participatif adapté aux communautés d'enseignants.

Néanmoins, Web Sémantique et Web Participatif sont souvent opposés. Aux ontologies formelles généralement réalisées par quelques experts s'opposent les tags hétérogènes ajoutés par de nombreux utilisateurs aux profils variés.

Dans une troisième partie, nous proposons donc un modèle fusionnant Sémantique et Participatif. Ce modèle vise à développer des applications de partage de ressources, principalement pour des communautés de pratique. Il se fonde sur une Indexation Progressive et Multi-points de vue (modèle IPM) consistant à :

1. permettre aux utilisateurs de structurer progressivement les métadonnées, pour finalement favoriser des raisonnements sémantiques par les machines ;
2. permettre aux utilisateurs de collaborer progressivement, pour finalement favoriser une vision partagée du domaine par les humains.

Ce modèle est implémenté dans un outil de partage de signets, nommé SemanticScuttle, proposant des fonctionnalités originales comme des tags structurés par des relations d'inclusion et de synonymie, ou des espaces wiki pour décrire des tags. L'outil a été développé et expérimenté avec des documentalistes en sociologie sur plusieurs mois. Diffusé, l'outil est employé par des utilisateurs dans plusieurs pays.

Finalement, nos travaux nous permettent de formuler des hypothèses sur un modèle socio-technique soutenant le partage entre enseignants. Ils sont aussi une contribution à des modèles intégrant différentes formes d'indexation : automatique et par des humains, faisant intervenir des experts et les utilisateurs, fondée sur des modèles structurés (e.g. ontologies) et sur des métadonnées flexibles (e.g. tags).

Mots-clés: Indexation (documentation), Pédagogie – Ressources internet, Web 2.0, Communautés virtuelles, Web sémantique, Ontologies (informatique).