

AMINE MESTIRI

**VERS UNE APPROCHE WEB SÉMANTIQUE DANS
LES APPLICATIONS DE GESTION DE
CONFÉRENCES**

Mémoire présenté
à la Faculté des études supérieures de l'Université Laval
dans le cadre du programme de maîtrise en informatique
pour l'obtention du grade de maître ès sciences (M.Sc.)

DÉPARTEMENT D'INFORMATIQUE ET DE GÉNIE LOGICIEL
FACULTÉ DES SCIENCES ET DE GÉNIE
UNIVERSITÉ LAVAL
QUÉBEC

2007

Résumé

Depuis sa création, il y a plusieurs années, par Tim Berners-Lee, le World Wide Web a révolutionné considérablement plusieurs domaines, notamment la société, l'économie et surtout la recherche et la manière dont celle-ci est conduite. Cependant, la majeure partie du Contenu du Web actuel a été conçue pour être lue et comprise par les êtres humains. Les applications et les agents logiciels n'avaient alors aucunes idées sur son sens réel. Le web sémantique, comme extension du Web actuel, vise à structurer son sens et permettre une meilleure collaboration homme-machine. On s'ouvre ainsi à de nouvelles possibilités d'automatisation sur le web et une gestion plus intelligente du contenu. Dans cette optique, plusieurs technologies clés comme XML, RDF et OWL ont été développées, puis standardisées par le World Wide Web Consortium (W3C).

Nous nous proposons d'étudier, dans le cadre de notre projet de maîtrise, l'approche Web sémantique dans les applications de gestion de conférences. Le projet SWOC (*Semantic Web Open Conference*) fait partie de cette réflexion. Il s'agit d'effectuer la ré-ingénierie d'une application en ligne de gestion des conférences, dans le but de la faire migrer vers la technologie du Web sémantique. Les ateliers de travail et les séminaires constituent, en effet, un partage de connaissances et d'idées entre des personnes d'une même communauté intellectuelle. Mettre leur gestion au cœur des innovations technologiques doit tenir compte de nombreuses contraintes. Or, au cours de notre recherche, nous avons relevé plusieurs limites dans les applications en ligne les plus utilisées, liées principalement à l'assignation des articles aux évaluateurs, l'aide à la saisie des données et la description des ressources.

L'idée qui se profile derrière ce projet est de cerner les plus-values du nouveau système et à en dégager les résultats d'assignation, afin de les comparer à ceux d'un système de base. Ce travail de recherche vise ainsi à analyser de plus près l'approche Web sémantique et son apport dans les applications de gestion de conférences, et d'en tirer des conclusions.

Avant-propos

Je tiens tout d'abord à remercier chaleureusement, mon directeur de recherche, M. Mamadou Tadiou Koné, pour son soutien constant et son dévouement pour son travail. Mr Koné a su répondre présent tout au long de ce travail. Je ne peux pas passer sous silence son inépuisable patience et sa générosité en partageant ses connaissances.

Je désire exprimer également ma sincère gratitude envers toutes les personnes qui ont contribué, de près ou de loin, à la réalisation de ce travail. J'adresse une reconnaissance particulière envers mes parents pour le support moral et les précieux conseils.

Je remercie aussi Messieurs Béchir Ktari et Danny Dubé pour avoir eu l'amabilité d'évaluer ce travail.

Finalement, je dédie ce mémoire à mon grand-père Hamed Mestiri. Puisse son âme reposer en paix.

Table des matières

Résumé.....	i
Avant-propos.....	ii
Table des matières.....	iii
Liste des figures.....	v
Chapitre 1: Introduction Générale.....	1
I Contexte de l'application.....	1
II Méthodologie de travail.....	7
III Structure du mémoire.....	7
Chapitre 2: État de l'art.....	9
I Introduction.....	9
II L'intérêt du Web Sémantique.....	10
III Web sémantique vs Intelligence artificielle.....	12
IV Architecture en couches du web sémantique.....	12
IV.1 La couche URI et Unicode.....	14
IV.2 La couche XML, NS, XML Schema.....	16
IV.3 La couche RDF et RDF Schema.....	19
IV.4 Couche ontologique.....	23
IV.5 Couche logique.....	29
IV.6 Couche Confiance et preuve.....	31
V Conclusion.....	32
Chapitre 3: Étude du domaine d'application.....	33
I Introduction.....	33
II Le processus de gestion des conférences.....	34
II.1 Première étape : Organisation de la conférence.....	34
II.2 Deuxième étape : Processus d'évaluation des articles :.....	36
III Étude du domaine : acteurs et cas d'utilisation.....	38
III.1 L'administrateur.....	39
III.2 Les auteurs.....	39
III.3 Les évaluateurs.....	41
III.4 Les membres du comité de programme.....	43
IV Présentation des applications les plus connues.....	46
V Conclusion.....	50
Chapitre 4: Architecture et solution proposée.....	51
I Introduction.....	51
II Problématique.....	51
III Ré-ingénierie de l'architecture.....	52
IV Annotation sémantique des ressources.....	57
IV.1 Aperçu.....	57
IV.2 Modèle de représentation des annotations sémantiques.....	58
IV.3 Automatisation des méthodes d'annotations.....	59
IV.4 Ingénierie de l'ontologie.....	64
IV.5 Moteur d'inférence.....	75
V Conclusion.....	83
Chapitre 5: Évaluation du projet.....	84
I Introduction.....	84
II Évaluation du système SWOC.....	84

II.1 Approche empirique :	85
II.2 Approche analytique :	85
III Apport de la technologie	86
III.1 Automatisation de la saisie d'information	86
III.2 Assignation des articles aux évaluateurs	91
IV Critiques sur le Web sémantique	98
V Conclusion	99
Chapitre 6: Conclusion générale	100
Bibliographie	102
Dictionnaire des acronymes	106

Liste des figures

Figure 1-1: Courbe d'évolution des serveurs sur le réseau Web (Source [NetCraft, 2005]).....	2
Figure 1-2: Résultat de recherche retourné par le moteur « Yahoo! », pour une requête donnée	3
Figure 1-3: Résultat de recherche retourné par le moteur « Google », pour une requête donnée	4
Figure 2-1: Architecture en couches du web sémantique (source: [Berners-Lee, 2001]).....	13
Figure 2-2: Évolution dans le temps des différentes couches du web sémantique.	14
Figure 2-3: Exemple d'utilisation des espaces de nommage.....	18
Figure 2-4: Représentation schématique d'un graphe RDF.....	21
Figure 2-5: Extrait de code RDF (source: http://websemantique.org/RDF).	22
Figure 2-6: Exemple d'une ontologie élémentaire du domaine.....	24
Figure 2-7: Différentes phases l'ingénierie ontologique.	26
Figure 2-8: Le cycle de vie d'une ontologie (source :[Fernandez et al, 1997]).....	27
Figure 2-9: Évolution des langages ontologiques.	28
Figure 2-10: Évolution des langages ontologiques.	29
Figure 2-11: Architecture des bases de connaissances.....	30
Figure 3-1: Procédure de sélection des membres du Comité de Programme.....	35
Figure 3-2: Procédure de sélection des évaluateurs	36
Figure 3-3: Procédure de traitement des articles reçus pour une conférence.....	38
Figure 3-4: Diagramme de cas d'utilisation liés à l'administrateur (Source [Cetic, 03]).	39
Figure 3-5: Diagramme de cas d'utilisation liés au chairman (Source [Cetic, 03]).	40
Figure 3-6: Diagramme de cas d'utilisation liés à l'évaluateur (Source [Cetic, 03])......	43
Figure 3-7: Diagramme de cas d'utilisation liés aux membres du comité de programme (Source [Cetic, 03]).	44
Figure 3-8: Diagramme d'états du traitement des articles reçus pour la conférence (Source [Celtic, 03])	45
Figure 3-9: Capture d'écran de la page principale de l'outil "OpenConf"	47
Figure 3-10: Capture d'écran de la page de configuration de l'outil "OpenConf"......	49
Figure 4-1: Représentation des différentes couches de l'architecture des applications en ligne.	52
Figure 4-2: Architecture en « client / Serveur » des applications en ligne.	53
Figure 4-3: Insertion du module « Web sémantique » dans l'architecture en couches des applications en ligne actuelles.....	54
Figure 4-4: représentation des relations existantes entre les différentes parties intervenantes de l'architecture du système SWOC.....	56
Figure 4-5: Annotation des ressources via l'outil KIM (extraite de [OntoText]).....	58
Figure 4-6: Processus d'annotation avec l'outil KIM.	61
Figure 4-7: Invitation des évaluateurs et création de leurs profils FOAF.....	62
Figure 4-8: Soumission et annotation semi-automatique d'un article avec le système SWOC..	63
Figure 4-9: représentation de l'ingénierie cyclique de l'ontologie du domaine du système SWOC (Source: [Yang, 2006]).	65
Figure 4-10: extrait de la conception de l'ontologie du système SWOC.	68
Figure 4-11: capture d'écran de la plateforme Protégé - Créations des concepts.....	70

Figure 4-12: Capture d'écran de la plateforme Protégé - Créations des relations entre ces concepts.....	71
Figure 4-13: Extrait du code OWL de l'ontologie du domaine du système SWOC.....	72
Figure 4-14: Organisation en packages de l'outil RAP	76
Figure 4-15: Processus de fonctionnement du module MemModel.	77
Figure 4-16: Extrait du code source du système SWOC - Inclusion de l'outil RAP.....	78
Figure 4-17: Extrait du code source du système SWOC - création d'un modèle de graphes RDF	79
Figure 4-18: interprétation du thème de l'article.	81
Figure 4-19: Assignation automatique de l'outil d'un article à quelques évaluateurs.	82
Figure 5-1: capture d'écran d'un formulaire de saisi d'un nouveau membre du CP (extrait du système SWOC).	87
Figure 5-2: Suite du formulaire de saisie d'un nouveau membre du CP (extrait du système SWOC)	88
Figure 5-3: capture d'écran d'un formulaire de soumission d'un document à la conférence (extrait du système SWOC).....	89
Figure 5-4: Suite du formulaire de soumission d'un document à la conférence (extrait du système SWOC).	90
Figure 5-5: Processus d'assignation des articles aux évaluateurs.....	91
Figure 5-6: Calcul de la moyenne de satisfaction de l'assignation des articles par l'outil OpenConf.	93
Figure 5-7: de la moyenne de satisfaction de l'assignation des articles par l'outil SWOC.	95
Figure 5-8: Courbes des moyennes de satisfaction de l'ensemble des articles pour les Outils OpenConf et SWOC.....	97

Chapitre 1: Introduction Générale

I Contexte de l'application

Depuis son apparition au début des années 90, la technologie Web a beaucoup gagné en popularité, en laissant un impact sans précédent sur la société et l'économie. Au fil des années, cette infrastructure est passée d'un simple réseau de communication et d'échange de données, à une gigantesque toile, considérée comme le plus grand répertoire d'information rassemblée et la plus grande source de partage de données.

On compte, en effet, aujourd'hui plus de 633.6 millions d'utilisateurs à travers le monde, selon les statistiques de 2003, calculées sur la base de l'année précédente [ClearLead, 2003]. Selon la même source, le nombre de documents en ligne est exprimé en termes de Billions. La société NetCraft, quant à elle, spécialisée dans les solutions d'analyses et de statistiques réseaux, a recensé en 2005, plus de 75 millions de sites sur l'ensemble de la Toile, dont 17,5 millions au cours des 12 derniers mois [NetCraft, 2005]. Le nombre de noms de domaine via le web a dépassé les 76 millions en 2005, comme l'indique la figure 1.1 [NetCraft, 2005].

Tous ces chiffres permettent de tracer l'évolution du web au cours de cette dernière décennie en donnant une idée sur l'évolution croissante des ressources utilisées à travers cette gigantesque toile. Une croissance due essentiellement à la présence de plus en plus de PME, qui ont compris les enjeux économiques de cette technologie.

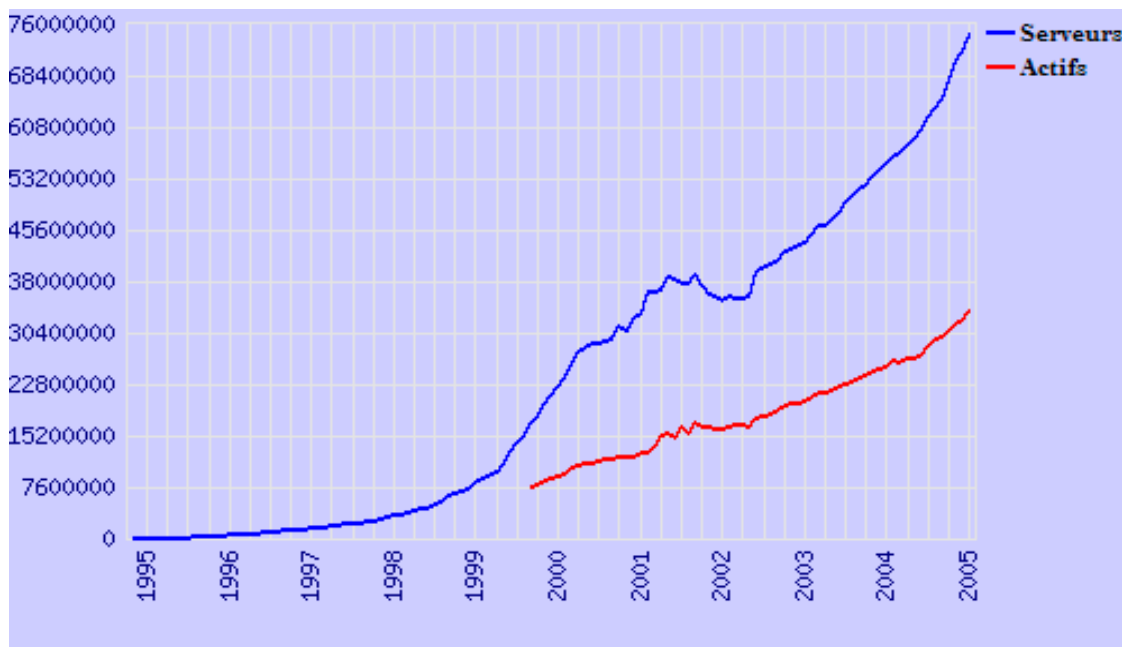


Figure 1-1: Courbe d'évolution des serveurs sur le réseau Web (Source [NetCraft, 2005]).

Cependant, malgré cette forte émergence qui ne cesse jour après jour d'améliorer nos façons de communiquer et de nous informer, le Web actuel présente encore certaines défaillances liées à la gestion des ressources et à l'automatisation de certaines tâches.

En effet, celui-ci est produit et conçu à l'origine pour être lu par des êtres humains, seuls aptes aujourd'hui à emprunter les autoroutes de l'information, de comprendre leurs sens et de décider de leurs pertinences. Ces informations ne sont pas dotées d'une sémantique interprétable par la machine, d'où la nécessité de l'intervention de l'homme dans la plupart des tâches. La problématique d'accès et de gestion des ressources, quant à elle, figure dans la plupart des ouvrages qui s'articulent autour des limites du web actuel. Elle est également pointée du doigt par différentes recherches, notamment celles sur le web sémantique [Bernes Lee *et al.*, 2001] et l'ingénierie des connaissances [Dieng *et al.*, 2000].

Par ailleurs, et tout au long de l'évolution du Web, l'attention a toujours été portée sur l'optimisation de la recherche d'information et de l'assignation automatique sous toutes ses formes. Ces points ont de lourdes conséquences sur les applications en ligne, notamment au niveau de la précision de la recherche, la pertinence des données retournées et la conformité des résultats d'assignation. On peut clairement remarquer les frustrations des utilisateurs, quant aux résultats insatisfaisants retournés par des moteurs tels que « Google » ou « Yahoo! »,

pourtant connus par la précision et la pertinence de leurs réponses aux requêtes, comparés à d'autres moteurs. Les résultats retournés sont souvent répartis sur plusieurs pages, où une information très pertinente peut se trouver à la fin, alors que les premières pages peuvent retourner des URL de sites qui n'ont rien à voir avec ce que cherche l'internaute. Les figures 1.2 et 1.3 montrent l'impertinence des recherches de ces deux moteurs. Il est clairement montré que les résultats de recherche ne correspondent absolument pas à ce que cherche l'internaute à travers sa requête.

The screenshot shows the Yahoo! search interface. At the top, the search bar contains the text "commission arbitrale de football en tunisie". Below the search bar, there are radio buttons for "the Web" (selected) and "just pages from Canada". The search results are displayed below, showing a list of six items. Each item includes a title, a brief description, and a URL. The results are as follows:

1. [Site Officiel de l'AFAF Moselle Est "Amicale Française des Arbitres de Football amicale Moselle Est" - Vers une ...](#) - Translate this page
... cadre juridique de la pratique **arbitrale**" et à répondre à ce qui apparaît comme " ... sécurité au plan social et fiscal tout **en** étant indemnisés au titre de cette ...
[arbitres-moselle-est.com/html/modules.php?name=News&file=article&...](#) - 37k - [Cached](#) - [More from this site](#)
2. [Voila.fr - Coupe du monde 2006 La Coupe du monde de A à Z](#) - Translate this page
... Angel Maria Villar Llona, le président de la **Commission arbitrale** de la Fifa. ... **football** mondial, les nations européennes ont repris leurs droits **en** Allemagne. ...
[sports.voila.fr/fr/.../200628/la-coupe-du-monde-de-a-a-z_100609.html](#) - 35k - [Cached](#) - [More from this site](#)
3. [Site Officiel de l'AFAF Moselle Est "Amicale Française des Arbitres de Football amicale Moselle Est" - Le Sénat a ...](#) - Translate this page
... proposition de loi prévoit la création d'une "**Commission** nationale consultative ... fonction **arbitrale** sous l'autorité de la fédération mais **en** toute indépendance ...
[arbitres-moselle-est.com/html/modules.php?name=News&file=article&...](#) - 51k - [Cached](#) - [More from this site](#)
4. [Nouvelles de Tunisie, Sports](#) - Translate this page
D'autre part, la paire **arbitrale** composée des Tunisiens Anis Béji et ... membre de la **commission de** contrôle antioppage lors du Mondial allemand **en** 2007. ...
[www.tunisie.com/sports/mondhand07.html](#) - 4k - [Cached](#) - [More from this site](#)
5. [FIFA.com Le site officiel de la Fédération Internationale de Football Association](#) - Translate this page
... CV plusieurs éditions de River Plate - Boca Juniors **en** Argentine, deux Copas ... La délégation **arbitrale de** la Conmebol comptera neuf représentants, si l'on ...
[fifa.com/fr/comp/index/0,2442,72548,00.html?comp=WYC&year=2003&...](#) - 54k - [Cached](#) - [More from this site](#)
6. [allAfrica.com: Cameroun: Au Tribunal - Les comptes de la Camair bloqués à](#)

Figure 1-2: Résultat de recherche retourné par le moteur « Yahoo! », pour une requête donnée

The image shows a Google search interface. At the top, the Google logo is on the left, and navigation links for 'Web', 'Images', 'Groupes', 'Actualités', 'Maps', and 'plus »' are on the right. A search bar contains the text 'commission arbitrale de football en tunisie'. To the right of the search bar is a 'Rechercher' button and links for 'Recherche avancée' and 'Préférences'. Below the search bar, it says 'Rechercher dans : Web Pages francophones Pages : Canada'. The search results are displayed under the heading 'Web' and show 'Résultats 1 - 10 sur un total d'environ 597 pour commission arbitrale de football en tunisie'. The first result is titled 'Tunisie - Sport - Trabelsi à Ajax jusqu'en 2006' and includes a snippet: 'Toute l'actualité de la Tunisie compilée par Bab-el-Web.com depuis plus de ... affirmant que la Commission arbitrale issue de la Fédération hollandaise qui ...'. The second result is titled 'Nouvel Obs - Coupe du monde 2006 La Coupe du monde de A à Z' and includes a snippet: '... répond Angel Maria Villar Llona, le président de la Commission arbitrale de la Fifa. ... Le genre de records dont le football se passerait volontiers. ...'. The third result is titled 'Actualité - Coupe du monde de Football 2006' and includes a snippet: '... "l'incident sera(it) étudié par la commission arbitrale qui se réunit de ... La Tunisie de Roger Lemerre doit gagner contre l'Ukraine si elle veut se ...'. The fourth result is titled 'La Tunisie organisera un séminaire international des arbitres de ...' and includes a snippet: '... de Football et de la commission des médias de l'Union arabe de football. ... la coopération entre le trio arbitral, les relations entre les arbitres et ...'. The fifth result is titled 'Sanctions définitives condamnant les incidents survenus lors du ...' and includes a snippet: 'La Fédération Turque de Football et les deux joueurs ayant déposé des demandes de reconsidération et d'atténuation des peines, la Commission de Recours de ...'. The sixth result is titled 'Voila.fr - Coupe du monde 2006 La Coupe du monde de A à Z' and includes a snippet: 'Bientôt les éliminatoires du Mondial 2006 de football sur Voila Sports! ... Angel Maria Villar Llona, le président de la Commission arbitrale de la Fifa. ...'. Each result includes a link to the source page, the page size, and links for 'En cache' and 'Pages similaires'.

Figure 1-3: Résultat de recherche retourné par le moteur « Google », pour une requête donnée

Le processus utilisé jusqu'à présent pour rechercher une information, est assez complexe, mais consiste à peu près à comparer la liste des mots saisis par l'utilisateur avec les mots clés de chaque site. C'est la définition même de « *la recherche par indexation* ». Cette méthode s'avère souvent très limitée, puisqu'il suffit juste que le mot recherché figure dans la liste des mots clés du site, au niveau de la balise « Méta » du code HTML du site web, pour que l'adresse URL de ce dernier soit retournée par le moteur. Par ailleurs, il faut avouer qu'il y a également des enjeux financiers très considérables, qui font que les propriétaires de ces moteurs ne cherchent pas nécessairement à améliorer leurs méthodes de recherche. Parmi ces

raisons, nous citons le fait que certaines entreprises payent des sommes importantes pour voir leurs sites affichés en tête des résultats retournés par le moteur. Cela constitue une véritable source de revenu pour les propriétaires des moteurs, certes, mais se fait, bien entendu, au dépend de la pertinence du classement des informations retournées.

Les assignations automatiques, au même titre que la recherche d'information, représentent l'un des handicaps de la technologie du web actuel, qu'on appelle syntaxique. Les évaluateurs, dans une conférence donnée, se voient assigner des articles qui, souvent, n'ont rien à voir avec leurs domaines de recherches. Il suffit que l'article aborde un thème qui s'articule autour de l'intelligence artificielle, par exemple, pour que le système trouve le premier correcteur du même domaine et le lui affecte. Cependant, l'intelligence artificielle est un vaste domaine. Être spécialiste dans les systèmes experts dans l'aide à la décision, ne veut pas dire nécessairement que l'on est expert en robotique.

Toutes ces faiblesses, en plus des besoins en croissance permanente sur le web, font que le web actuel, commence à perdre de sa crédibilité, pour devenir sujet à de nombreuses critiques. La gestion insuffisante des ressources, la nécessité de l'intervention humaine dans la plupart des tâches et la frustration des utilisateurs pour les résultats de recherches et d'assignations sont des facteurs qui nous poussent à penser à une nouvelle conception de l'architecture de cette toile étendue. Une nouvelle vision qui permettra d'assigner une sémantique aux ressources utilisées, de façon à permettre à des agents logiciels de les interpréter et d'effectuer les raisonnements nécessaires, de manière autonome, en se basant sur une ontologie du domaine. Nous reviendrons plus tard sur la définition de cette dernière. C'est dans cette optique qu'apparaît le web sémantique, comme extension du web actuel, [Berners-Lee, 2001], dans la mesure où il vient structurer son sens et permettre la gestion automatique et intelligente de son contenu. Les informations ne seront plus juste stockées, mais aussi comprises [Xavier Borderie, 2002].

Proclamée « *technologie du futur* », par son créateur, Tim Berners-Lee, cette nouvelle vision du web va permettre de s'ouvrir à de nouvelles possibilités d'automatisation dans le web et d'assurer une meilleure collaboration entre les humains et les machines [Berners-Lee, 2002]. Parmi les avantages qu'on peut compter pour cette nouvelle génération du web, nous citons la possibilité de réutilisation et de partage des données disponibles sur la toile du net [Berners-Lee, 2004]. Le web sémantique a fait aujourd'hui un grand pas vers l'avant, en voyant naître

beaucoup de standards proposés par le W3C, qu'est le consortium du World Wide Web [W3C].

Étant donné la contribution majeure qu'apporte cette nouvelle technologie à la gigantesque toile du web, et l'intérêt qu'elle joue par conséquent sur la société de l'information, en favorisant la gestion intelligente du contenu, nous avons pris l'initiative de l'étudier de plus près dans le cadre de notre projet de maîtrise en informatique. L'émergence de cette nouvelle technologie, définie par son caractère sémantique, donne l'opportunité de proposer une nouvelle approche de conception des applications web actuelles.

Par ailleurs, nous avons choisi les systèmes en ligne de gestion de conférences, comme domaine d'application. Il s'agit précisément d'effectuer la réingénierie d'une application existante, pour mettre en œuvre une nouvelle version, entièrement basée sur la technologie du web sémantique, afin de bénéficier des nombreux avantages qu'apporte cette dernière. Nous visons à travers ce projet, l'optimisation de la gestion des ressources utilisées (Documents, participants, évaluateurs, organisateurs, etc.) en permettant une recherche plus précise de l'information et une affectation d'articles plus ciblée plus conforme aux expertises des évaluateurs.

La gestion des conférences est une tâche très lourde et délicate pour la gérer manuellement. Des outils de gestion et de suivi en ligne de ce processus sont disponibles actuellement sur le réseau du web, parmi lesquelles, nous citons **OpenConf** [OpenConf], **CyberChair** [CyberChair] et **ConfTool** [ConfTool]. Malgré l'importance de leur utilisation, ces outils présentent encore des limites. La réingénierie de l'une de ces applications vise surtout à faire face aux exigences des utilisateurs et répondre aux faiblesses des systèmes actuels. Au regard de l'architecture sous laquelle sera déployé le nouveau système, ce projet s'accorde tout à fait avec les enjeux du Web sémantique du W3C.

Depuis son apparition, le web sémantique s'est beaucoup intéressé au domaine médical et biologique, autour duquel s'articulaient les premières ontologies apparues. Ce n'est que par la suite qu'il a pris de l'envergure, pour s'ouvrir sur de nouveaux horizons en couvrir d'autres domaines. Cependant, aucun effort n'a encore été déployé sur les applications en ligne de gestion de conférences, qui constitue en soi, l'originalité de notre travail.

II Méthodologie de travail

Par souci d'une bonne organisation surtout de la continuité dans les tâches, nous avons jugé très utile de tracer, dès le départ, la méthodologie que nous suivrons pour ce projet. Celle-ci s'articule autour de deux technologies clés, à savoir, le web sémantique et les applications en ligne de gestion des conférences.

Les étapes à suivre sont les suivantes :

- **Ré-ingénierie de la conception et de l'architecture de l'application** libre « OpenConf » afin de lui intégrer la technologie du web sémantique ;
- **Ingénierie de l'ontologie du domaine**, ce qui sous-entend la conception et la création, ainsi que la gestion, de cette dernière.
- **Concevoir une annotation sémantique manuelle et/ou semi-automatique** complète et précise des ressources, de manière à faciliter leur interprétation et à garantir leur interopérabilité ;
- **Intégrer un moteur d'inférence** apte à interpréter les descriptions sémantiques des ressources et effectuer les raisonnements nécessaires sur l'ontologie du domaine.

III Structure du mémoire

Après avoir introduit le contexte du projet autour duquel s'articule ce mémoire, la motivation et la problématique, nous entamons la présentation du projet SWOC en tant que tel.

Le chapitre suivant portera sur la présentation de l'état de l'art. Tout en nous focalisant sur la technologie du web sémantique, nous introduirons plus en détail son architecture et son objectif, ainsi que l'état actuel des connaissances.

Dans le chapitre 3, nous discuterons du domaine d'application qu'est la gestion en ligne des conférences. Ce chapitre portera donc, sur la définition des acteurs et des cas d'utilisation d'un tel système. De plus, nous pointerons du doigt les faiblesses communes des systèmes actuels, afin de comprendre ce que nous attendons de l'intégration de la technologie du web sémantique.

Le chapitre 4, quant à lui, abordera l'aspect architectural de ce projet. L'intégration de la technologie du web sémantique doit passer par une révision complète de la conception du système et une réingénierie de son architecture. Celle-ci permettra l'annotation des ressources et le raisonnement basé sur les ontologies. Nous aborderons également les différentes phases du développement du nouveau système, tout en s'appuyant sur une présentation des interfaces et de quelques exemples du code.

L'évaluation de l'apport et de la contribution de la technologie du Web sémantique de façon générale, et dans le cadre de notre projet en particulier, sera abordée dans le chapitre 5. Une étude comparative sera alors menée sur les systèmes de base, qu'est le OpenConf, et celui que nous venons de développer, à savoir le SWOC. Les résultats seront interprétés et discutés dans ce même chapitre.

Nous concluons ce mémoire par une récapitulation des objectifs atteints ainsi que les perspectives d'avenir.

Chapitre 2: État de l'art

I Introduction

Cette dernière décennie a connu une évolution considérable de la gigantesque toile du web. Une évolution marquée par la croissante permanente des données et des ressources exploitées à travers cette gigantesque toile, ce qui rend très difficiles leurs localisations et leurs gestions, d'autant plus que le web actuel ne peut en aucun cas interpréter leurs sémantiques. La recherche d'une information, surtout avec des critères très précis et diversifiés, revient parfois à trouver une aiguille dans une botte de foin [Kun Yang, 2006]. Le web actuel serait incapable de donner suite à une requête telle que : « Quels sont les restaurants qui servent de la nourriture vietnamienne au Québec, qui est ouvert jusqu'à minuit et qui propose un spectacle en milieu de semaine ». C'est dans cette optique qu'est apparu le web sémantique comme une extension du web actuel pour structurer son contenu et lui donner un sens interprétable par la machine. Cette technologie, dont le but est d'améliorer la coopération homme/machine pour une gestion plus intelligente du contenu, permettra ainsi à des agents logiciels d'effectuer des tâches et des raisonnements logiques au nom des utilisateurs qui les leur ont délégués. On peut s'ouvrir ainsi à de nouvelles possibilités d'automatisation dans le web. [Koné]

En effet, et selon la vision de son créateur, Tim Berners-Lee, l'évolution du web est marquée par deux grandes étapes [Berners-Lee, 2001]. La première consiste simplement à relier les internautes entre eux pour partager des données et travailler en coopération. A cet effet, le HTML est largement suffisant, notamment à travers les liens hypertextes qui permettent de naviguer d'une ressource à l'autre. Toutefois, le défaut majeur du HTML est qu'il ne sépare pas le contenu de la présentation, ce qui pose un problème d'interopérabilité. Le second volet de cette évolution, quant à lui, a porté sur l'extension des fonctionnalités du Web, favorisant ainsi la coopération homme/machine. Pour cela, cette seconde génération de Web doit être apte à interpréter la sémantique des ressources utilisées. Ce principe repose déjà sur la séparation du contenu et de la présentation. Des standards mis au point par le W3C permettent déjà de structurer ce premier et de lui donner un sens, en lui affectant des méta-données. En se basant sur ces informations et les connaissances du monde réel, un système peut alors effectuer des inférences au nom des utilisateurs. Cette technologie permet d'assurer un plus grand

partage et une réutilisation des données, tout en favorisant la coopération homme/machine [Berners-Lee, 2001].

II L'intérêt du Web Sémantique

Le web sémantique apparaît comme un constat des insuffisances du web originel. La capacité à manipuler le contenu sur la base de sa sémantique est en soi une fin de la technologie du web sémantique. L'idée de cette technologie émergente est de trouver un moyen pour interpréter le sens des informations et gérer intelligemment le contenu du web. Elle permettrait de comprendre le sens des requêtes, la sémantique des ressources et de retourner ainsi des résultats plus pertinents.

Le contenu du web actuel, produit et conçu à l'origine pour être lu par des êtres humains, deviendra ainsi interprétable par des programmes qu'on appelle agents logiciels. Ces derniers pourront ainsi extraire de l'information, la manipuler, la traiter, la synthétiser et effectuer certaines opérations, faites jusqu'à présent à la main, et ce, aux noms d'utilisateurs qui leur auront délégué ces tâches. C'est pourquoi on dit que cette technologie est la clé pour s'ouvrir sur de nouvelles possibilités d'automatisation dans le web. Jusqu'à ce jour, le langage le plus utilisé pour la présentation des données est le HTML. Les balises de ce langage se limitent à décrire la manière dont le document doit être présenté. Le serveur web ne fait que les interpréter. La signification du contenu du document reste implicite et le document ne peut donc pas être manipulé sur base de cette signification. Le XML, qui est apparu comme une nouvelle technologie permettant de séparer le contenu de la présentation, ne représente en fin de compte qu'un simple format de message qui permet de structurer le contenu pour une meilleure opérabilité dans le partage et l'échange d'informations. Même s'il a résolu beaucoup de problèmes au niveau syntaxique, ce standard du web n'a toujours pas couvert les problèmes liés à la sémantique des ressources.

Tim Berners-Lee, l'inventeur du World Wide Web, était le premier à avoir compris l'importance du web sémantique. Il en a parlé en 1988, avant même l'apparition du web actuel, en le qualifiant de web de la prochaine génération. Cette technologie émergente n'est pas un substitut du web syntaxique, mais bien son extension, dans la mesure où elle vient donner un

sens à l'information, pour permettre la gestion automatique et intelligente de son contenu, ce qui aboutirait à une meilleure collaboration entre l'homme et la machine [Berners-Lee, 2001]. Le nom qu'elle porte est d'ailleurs très explicite, et ne fait guère référence à une définition ou un théorème mathématique, mais simplement à la manipulation des ressources du web sur la base de leur sémantique. Ce projet est développé et géré par le Consortium du World Wide Web, qu'est le W3C, et ce, en collaboration avec un grand nombre de chercheurs et de partenaires industriels. L'une des principales tâches de ce consortium est de définir des standards pour le web sémantique.

Le principe du web sémantique consiste brièvement à attacher des métadonnées à toutes les ressources disponibles sur le web. « Le Web sémantique consiste à attacher aux contenus Web des métadonnées — en vue notamment d'automatiser certaines tâches applicatives, telles que la syndication », expliquait Daniel Dardailler, directeur Europe du W3C [Dardailler, 2002]. Par ailleurs, il faut aussi définir des ontologies qui permettraient aux humains et aux machines de partager les connaissances du domaine et de collaborer ensemble. Nous les utilisons en général pour permettre aux machines de raisonner et d'interpréter les informations ainsi que d'améliorer la pertinence des recherches. Les agents auxquels les utilisateurs délèguent des tâches, devront communiquer entre eux et interpréter le contenu échangé de la même manière, c'est-à-dire en interprétant les termes décrivant le contenu de la même manière. D'où l'intérêt de cette ontologie. « Avec la notion de Web sémantique, vous définissez un espace virtuel où les hyperliens pointeraient non plus sur des documents (textes ou images), mais sur des concepts » explique Pierre Lévy — titulaire de la chaire de recherche en intelligence collective à l'université d'Ottawa (Canada), en ajoutant : « Le futur Web exprimera l'intelligence collective de l'humanité » [Lévy, 2003]. Cette nouvelle technologie, dite de l'avenir, permet d'autre part de faciliter le partage et la réutilisation des connaissances via les ontologies. Le principe consiste à définir une interprétation commune d'une partie du monde réel, et modéliser les concepts et des relations entre les concepts par des classes et des relations entre les classes. Une ontologie est en général gérée comme un wiki, dans le sens où chacun peut intervenir pour modifier certains sens ou ajouter des nouvelles définitions. L'ontologie doit donc être entretenue et mise à jour au fur à mesure de l'évolution du monde qu'elle modélise.

III Web sémantique vs Intelligence artificielle

De par sa gestion intelligente du contenu, en se basant sur les ontologies du domaine et les bases de connaissances, le web sémantique est perçu par certains chercheurs comme de l'intelligence artificielle appliquée au réseau. Le débat reste cependant ouvert car cette expression est interprétée d'une multitude de manières. Par ailleurs, il faut avouer que certaines définitions, en plus des techniques d'inférence utilisées telles que les chaînages avant et arrière, prêtent à confusion. Pourtant, il existe de nombreuses différences entre ces deux technologies de l'information.

En effet, si les systèmes de représentation des connaissances en intelligence artificielle sont centralisés et que les règles ne sont pas transférables, afin de ne pas risquer de changer leurs significations, il en est tout autrement avec le web sémantique qui est considéré comme versatile et pour lequel les chercheurs acceptent les paradoxes. Par ailleurs, l'architecture de ce dernier est très répartie.

L'idéal serait alors de qualifier le Web sémantique d'une nouvelle technologie qui se situe à mi-chemin entre l'informatique traditionnelle et l'intelligence artificielle, de façon à ce que les robots et les agents sachent lire et interpréter le sens de l'information¹.

IV Architecture en couches du web sémantique

L'évolution des travaux réalisés dans le cadre du web sémantique est marquée par différents niveaux de complexité. Si les plus élémentaires consistent en une simple affectation de métadonnées aux ressources utilisées, d'autres se basent sur des inférences très alambiquées et permettent l'exploitation de ressources hétérogènes [Fadhel, 2005]. Quel que soit le niveau de complexité, ces applications reposent toutes sur une architecture en couches commune exprimée par la figure 2.1, recommandée par le W3C. Il s'agit de la vision schématique du web sémantique proposée par son inventeur Tim Berners-Lee en 2001.

¹http://www.p2pfoundation.net/Web_s%C3%A9mantique

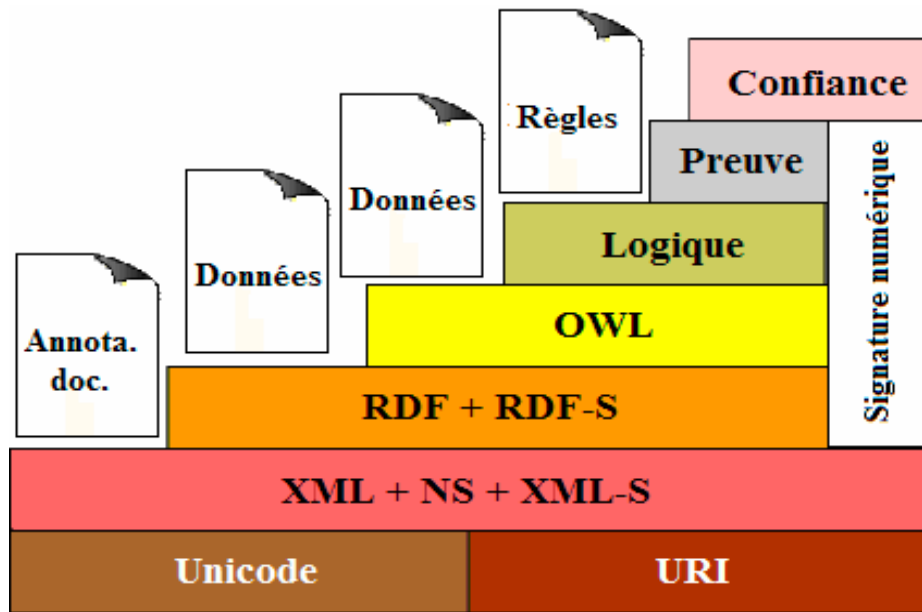


Figure 2-1: Architecture en couches du web sémantique (source: [Berners-Lee, 2001]).

Nous reviendrons un peu plus loin sur les détails de chacune des couches et leurs apports dans cette nouvelle vision de l'architecture du Web.

Étant donné le caractère très large et désordonné des données disponibles sur le réseau du net, on attend du web sémantique une meilleure structuration de ces dernières et une affectation optimale des métadonnées qui puisse simplifier leurs interprétations par les machines. Pour mener à bien ces objectifs, le W3C a mis au point, progressivement, une série de recommandations pour l'ensemble des couches de l'architecture. La figure 2.2 récapitule les standardisations faite par le W3C au sujet du Web sémantique, au cours du temps.

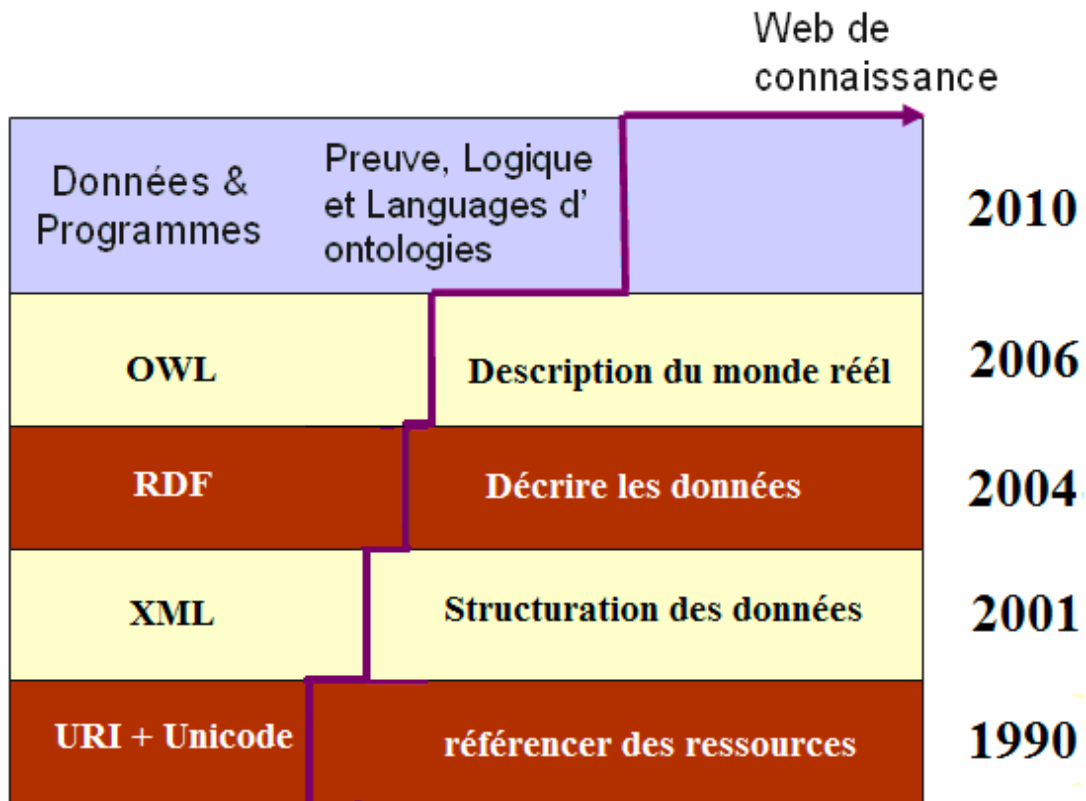


Figure 2-2: Évolution dans le temps des différentes couches du web sémantique.

Dans ce qui suit, nous présenterons plus en détail les différents niveaux de cette architecture.

IV.1 La couche URI et Unicode

L'URI (*Uniform Resource Identifier*) est justement un protocole simple et extensible pour identifier, d'une manière unique et uniforme, toute ressource sur le web. Il s'agit d'un aspect central de l'infrastructure, c'est pour cette raison que cet élément se trouve à la base de l'architecture en couches proposée.

La spécification de la syntaxe et de la sémantique de l'URI dérive des concepts introduits par l'initiative de globalisation de l'information dans le World Wide Web, qui date des années 1990. Elle faisait alors l'objet du protocole RFC1630. Son utilisation est décrite dans l'ouvrage

« *Universal Resource Identifiers in WWW* » et vise à réunir plusieurs recommandations concernant la localisation d'une ressource (URL) et son identification (URN). C'est pourquoi on dit, qu'il est le pilier du web syntaxique avant d'être celui du web sémantique. La notion d'URI a évidemment évolué depuis, pour être décrite par un nouveau standard du web qu'est le RFC2396.

URI est en fait un acronyme de *Uniform Resource Identifier*, qui veut dire en français Identificateur uniforme de ressources. D'une façon générale, le terme identificateur désigne une clé capable de référencer un objet ayant une identité. Dans le cas du web sémantique, l'URI est une séquence de caractères avec une syntaxe restreinte, qui permet d'identifier toute ressource utilisée dans le cadre d'une application web sémantique. On entend par ressource n'importe quel objet ayant une identité, telle qu'un document électronique, une page HTML, un fichier, une image, une vidéo, un service, une couleur, etc. Une ressource, n'est pas nécessairement un objet faisant partie du réseau, il peut s'agir d'un internaute, d'un livre, d'une corporation, etc. La notion d'uniformité, ou universalité, quant à elle, est d'autant plus importante que riche d'avantages. Elle permet d'abord à différents types d'identificateurs de ressources d'être utilisés dans le même contexte, même si le mécanisme qui leur permet d'y accéder est différent. Ensuite elle assure une interprétation sémantique uniforme des conventions syntaxiques communes pour les différents identificateurs. Par ailleurs, elle permet d'introduire de nouveaux types d'identificateurs de ressources sans interférer avec ceux existants. Enfin, cette uniformité permet la réutilisation de ces identificateurs dans différents contextes. Nous voyons là au moins 4 avantages de l'uniformisation des identificateurs des ressources.

Par ces propos, nous sous-entendons qu'il existe plusieurs types d'identificateurs de ressources. Un URI peut être classé, en effet, en 3 catégories, selon qu'il soit destiné à la localisation, au nommage ou au deux. Le terme URL (*Uniform Resource Locator*) désigne un sous-ensemble d'URI qui identifie les ressources via une représentation de leur mécanisme d'accès, plutôt que par le nom ou autre attribut de cette dernière, comme il en est le cas pour l'URN (*Uniform Resource Name*). L'URL et l'URN sont donc des cas particuliers d'URI. Voici quelques exemples d'URI :

- <ftp://ftp.ift.ulaval.ca/outils/openConf.rar>
Schéma FTP pour services File Transfer Protocol
- <http://www.ift.ulaval.ca/aminemestiri.html>
Schéma HTTP pour services Hypertext Transfer Protocol
- [Mailto : amine.mestiri.1@ulaval.ca](mailto:amine.mestiri.1@ulaval.ca)
Schéma d'adressage pour expédition de courriel
- <telnet://ift.ulaval.ca/>
Schéma pour services via le protocole TELNET

Par ailleurs, il est à noter que les données sont toujours encodées avec le jeu de caractères Unicode pour un maximum d'interopérabilité. C'est pourquoi cet élément figure dans cette couche de bas niveau, au même titre que l'URI.

IV.2 La couche XML, NS, XML Schema

À ce niveau d'architecture, nous ne sommes toujours pas au point d'affecter une sémantique à l'information, c'est-à-dire de la décrire et lui donner un sens. Il s'agit seulement d'une couche syntaxique, de bas niveaux, qui permet de structurer les données et organiser selon un format de message standard, et ce, grâce au langage de balisage extensible XML (*eXtensible Markup Language*). Ce dernier est donc est un sous-ensemble de SGML (*Standard Generalized Markup Language*), au même titre que le HTML (*Hyper Text Markup Language*), et en héritent par conséquent certaines de ses caractéristiques, notamment l'utilisation des chevrons (< >)¹. En d'autres termes, XML permet d'indiquer l'organisation logique de l'information d'un document, mais, a priori, ne permet pas d'en décrire le contenu². Structurer l'information, revient essentiellement à séparer la mise en forme et le contenu. Cette notion est au cœur de la définition de ce langage extensible.

¹<http://fr.wikipedia.org/wiki/XML>

²<http://lespetitescases.net/structurer-decrire-et-organiser-l-information-1>

Jusqu'à présent, dans le cadre du web syntaxique, le HTML est défini comme un standard pour éditer des pages web. Pourtant, celui-ci présente de nombreuses faiblesses, parmi lesquelles, on cite [Yang, 2006] :

- Le HTML définit le contenu et la présentation des données.
- Les données contenues dans un document HTML ne sont pas structurées.
- Les balises sont fixes, et ne permettent donc pas la personnalisation.

Tous ces inconvénients ont mis en question son utilité et surtout l'intérêt de développer un langage qui permet d'organiser l'information et de la définir selon un format avant de la décrire et lui donner un sens. D'où la définition du XML comme standard pour cette couche de l'architecture du web sémantique. En effet, même s'il joue le même rôle que le HTML, dans la mesure où il permet d'éditer le contenu, le XML, du fait de son extensibilité, est considéré comme un métalangage de balisage. Il permet aux utilisateurs de définir leurs propres balises, au lieu de traiter avec un ensemble de balises fixes, comme est le cas pour le HTML [Antonio, 2004]. Cela permet d'organiser et structurer l'information à la guise du développeur. Le XML est défini par le W3C comme un format de texte flexible, dont l'objectif initial est de faciliter l'échange automatisé du contenu entre des systèmes d'informations hétérogènes, notamment, sur Internet [W3C, 2004].

Avant de commencer à organiser les informations dans un document XML, il est impératif de définir la structure de ce dernier, afin de permettre notamment de vérifier sa validité. DTD (*Document Type Definition*) et XML-S (*XML Schema*) sont des langages de description de format de document XML. Cependant, on peut noter certaines différences entre eux. XML-S permet par exemple de définir des domaines de validité pour la valeur d'un champ, alors que cela n'est pas possible dans une DTD; en revanche, il ne permet pas de définir des entités. Par ailleurs, il est à noter qu'un fichier XML-S est lui-même un fichier XML¹. La norme XML n'impose pas l'utilisation d'une DTD pour un document XML, mais elle impose par contre le respect exact des règles de base de la norme XML, c'est pourquoi on fait la différence entre un document valide, pour un fichier XML qui comporte une DTD, et un document bien formé, pour un fichier XML qui respecte juste les règles de base du XML².

¹http://fr.wikipedia.org/wiki/XML_Schema

²<http://www.commentcamarche.net/xml/xmltdtd.php3>

Concrètement, un fichier XML peut faire appel à plus d'un document DTD ou XML-S. Sa structure peut donc être organisée selon plusieurs grammaires. Afin d'éviter les conflits de noms, dus au fait que ces documents sont en général développés indépendamment les uns des autres, le W3C a mis en place un nouveau standard, baptisé « Namespace », qui veut dire espace de nommage [Bray, 1998] [Walsh, 1998]. Selon la définition de ce consortium, les espaces de nommage fournissent un moyen simple pour qualifier les noms des éléments et des attributs dans le langage XML, en les associant avec des espaces de noms identifiés par URI.

La figure 2.3 est un exemple extrait des recommandations des espaces de nommage par le W3C, faisant office de type d'élément :

```
<x xmlns:edi= 'http://ecommerce.org/schema'>
  <!-- l'espace de nommage de l'élément « prix » est http://ecommerce.org/schema -->
  <edi:prix monnaie='Euro'>32.18</edi:prix>
</x>
```

Figure 2-3: Exemple d'utilisation des espaces de nommage

Notons que dans cet exemple, l'espace de nommage de l'élément « prix » est : <http://ecommerce.org/schema>.

Jusqu'à ce niveau de l'architecture, le problème de l'interprétation de la sémantique de l'information par la machine n'est toujours pas résolu. Le XML a su donner une nouvelle dimension aux informations en les structurant à la guise du développeur, mais ne permet pas d'en décrire le contenu. Cela paraît plutôt logique, puisque la structuration et la description sont deux processus différents¹.

Dans ce qui suit, nous présentons les standards de représentation des connaissances qui enrichissent les informations apportées par XML, à savoir RDF/RDFS et OWL.

¹<http://lespetitescases.net/structurer-decrire-et-organiser-l-information-1>]

IV.3 La couche RDF et RDF Schema

Après avoir référencé les ressources avec le protocole URI et structuré les informations avec le XML, l'étape suivante consiste à les annoter, afin de les doter d'un sens interprétable par la machine. C'est justement le rôle de la couche RDF et RDF-S dans l'architecture du Web sémantique.

Selon le dictionnaire officiel québécois, le grand dictionnaire¹, le terme annotation est défini comme un commentaire ou une explication d'un document ou de son contenu, habituellement ajoutée en note. Il peut s'agir d'une explication, d'une notation, d'une note, d'une remarque, d'un renvoi ou d'un synonyme.

En informatique, ce terme fait référence à l'affectation de notes, qu'on appelle métadonnées, à un texte ou une partie de texte, et ce, d'une manière formelle². Dans ce sens, le W3C a développé plusieurs standards pour structurer ces informations, notamment le HTML, SGML et XML qui fût encore une fois recommandée pour la technologie du web sémantique. L'annotation sémantique se résume donc en la génération d'informations spécifiques, appelées métadonnées, qui sert à la description d'un document pour un accès et une gestion plus efficace des informations disponibles sur le Web [Yang, 2006]. Celles-ci peuvent prendre des formes graphiques ou textuelles. Il existe deux types de modèles d'annotations. Le premier consiste à créer une annotation séparée et détachée de la ressource. La seconde, quant à elle, suggère un schéma d'annotation encapsulé [Desmontils, 2002].

Le RDF (*Resource Description Framework*) est un modèle conceptuel, normalisé par le W3C, servant à encadrer la description de ressources, et ce, d'une façon simple et non ambiguë [Benayache, 2005]. Ses applications visent initialement le web sémantique, mais elles peuvent s'étendre plus largement à l'ingénierie des connaissances³. Ce modèle est associé à une syntaxe dont le but est de permettre à une communauté d'utilisateurs de partager les mêmes métadonnées pour des ressources partagées.

¹<http://www.granddictionnaire.com>

²<http://www.ontotext.com/kim/semanticannotation.html>

³<http://websemantique.org/RDF>

Le RDF repose sur un ensemble de vocabulaires pour écrire les métadonnées, dont les plus connus sont le DC (*Dublin Core*) et le FOAF (*Friend Of A Friend*). Dans le web sémantique, un vocabulaire (ou « vocabulaire contrôlé ») désigne un ensemble de termes utilisés pour décrire des ressources¹.

FOAF (*Friend Of A Friend*) est un projet créé par Dan Brickley et Libby Miller en 1998. Il s'agit d'un vocabulaire RDF permettant de décrire des personnes et les relations qu'elles entretiennent entre elles, et ce, d'une façon compréhensible par la machine². Tous les documents FOAF sont donc reliés, ce qui permet de faire du web un grand réseau de connaissances. Nous reviendrons plus tard, au niveau de l'architecture de notre système, sur la définition de ce langage. Le DC (*Dublin Core*), quant à lui, est un langage basé sur RDF et adapté à la description des documents ou ressources Internet. Il comprend officiellement 15 éléments de description, parmi lesquels nous citons : « *Description* », « *Subject* », « *Creator* », « *Date* », « *Language* », « *Title* »³. Les valeurs associées à ces propriétés peuvent être écrites en différentes langues, à la guise de l'utilisateur. Le Dublin Core est reconnu officiellement par le W3C et repose sur la norme ISO 23950.

RDF est très souple et propose des systèmes très efficaces pour lier et décrire des ressources, sous forme de graphes. Il repose le plus souvent sur le langage de balisage XML. C'est pourquoi on parle très souvent de RDF/XML [Beckett, 2003]. D'autres formalismes peuvent aussi être retenus, à l'instar du N3, du N-triples ou du turtle. Le choix de la syntaxe XML par le W3C tient plus à la convergence des technologies, car il rend possible la réutilisation par RDF des règles et des dialectes utilisés pour exprimer et traiter du XML, en particulier le principe des namespaces et le langage de feuilles de style XSL⁴.

La construction de base en RDF consiste en un triplet d'éléments (Ressource, Propriété, Valeur), qu'on appelle déclaration RDF. Par analogie, un triplet RDF est similaire à la déclaration < sujet — prédicat — objet > [Manola, 2002] :

Ressource (Sujet) : Cela peut être n'importe quel objet référencé par une URI, qu'il concerne le web (Page HTML, document PDF, fichier multimédia...), ou non (Personne, Région, Etc.).

¹<http://websemantique.org/Vocabulaire>

²<http://fr.wikipedia.org/wiki/FOAF>

³ http://fr.wikipedia.org/wiki/Dublin_Core

⁴<http://lespetitescases.net/structurer-decrire-et-organiser-l-information-2>

Propriété (prédicat) : Critère, caractéristique, attribut ou relation qui peut décrire la ressource (titre, couleur, taille, auteur, etc.). Une propriété ses valeurs permises et ses relations avec les autres propriétés.

Valeur (objet) : C'est la valeur qui sera affectée à la propriété de la ressource. Cette affectation peut être soumise à certaines restrictions.

Chaque élément de ce triplet peut être un URI [Berners-Lee & All, 1999], un littéral ou une variable. Une déclaration RDF est souvent schématisée par un graphe RDF, comme le montre la figure 2.4, où les sujets et les objets sont représentés par des ellipses, qu'on appelle nœuds, et le prédicat par une flèche qui va du sujet vers l'objet. Les rectangles, quant à eux, représentent des instances d'une ressource.

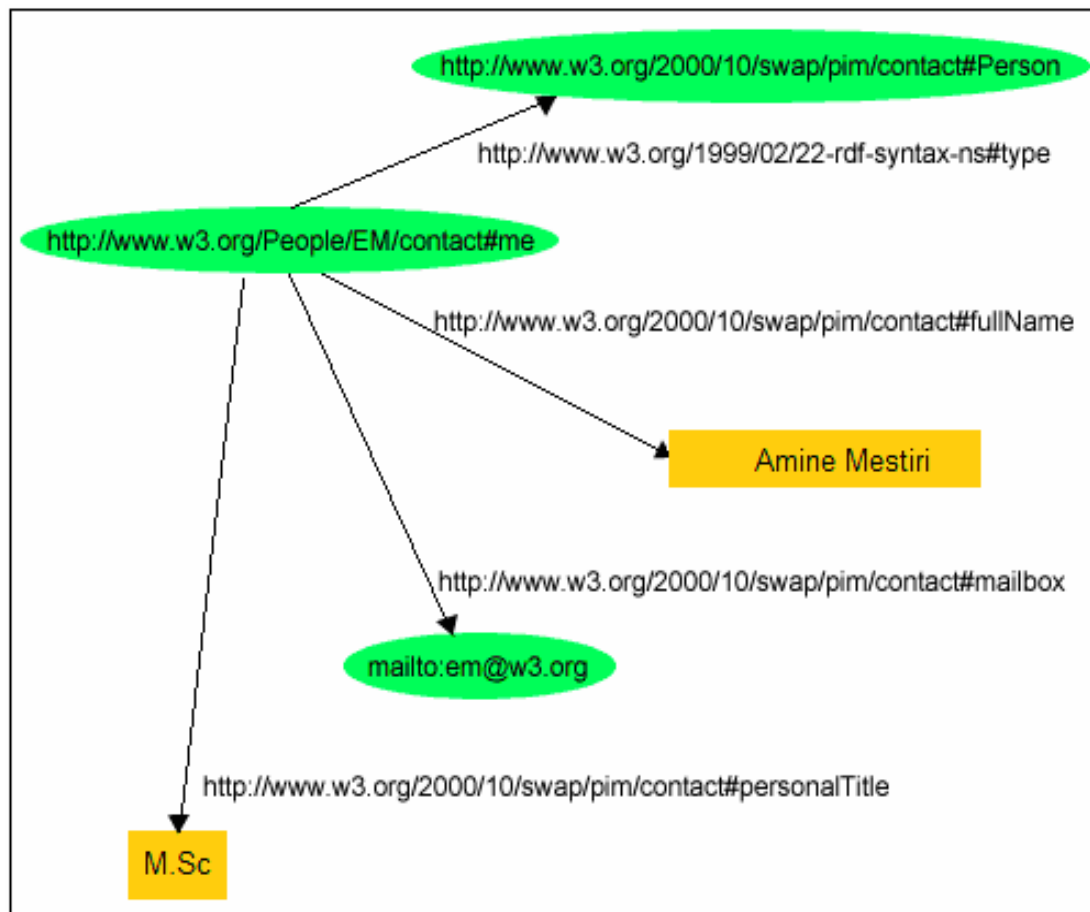


Figure 2-4: Représentation schématique d'un graphe RDF.

La figure 2.5 représente un bout de code de la description XML/RDF d'un document PDF, extrait du Wiki du web sémantique¹ :

```

conteneur RDF
<?xml version="1.0"?>
<rdf:RDF utilisation des espaces de noms rdf et dc
xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
xmlns:dc="http://purl.org/dc/elements/1.1/">
  <rdf:Description about précise la ressource à décrire
rdf:about="http://xml.coverpages.org/RadioIV-NewsML-en-
20020224.pdf"> valeur M. Onishi
propriété creator <dc:creator>M. Onishi</dc:creator>
  <dc:title>RadioTV-NewsML in Japan</dc:title>
  <dc:date>2002-02-21</dc:date>
  <dc:type>Text</dc:type>
  <dc:format>application/pdf</dc:format>
  </rdf:Description>
</rdf:RDF>

```

Figure 2-5: Extrait de code RDF (source: <http://websemantique.org/RDF>).

RDF-S (*RDF Schema*) est un vocabulaire de base pour décrire les déclarations RDF, au même titre que le XML-S pour le langage XML. Il ajoute à RDF la possibilité de définir des hiérarchies de classes et de définir les genres et les propriétés des ressources, d'assigner des contraintes spécifiques sur la nature des documents et de fournir des informations sur l'interprétation des déclarations RDF [Laublet *et al*, 2005]. Les schémas RDF permettent donc de garantir qu'un document RDF est sémantiquement consistant [Abdel, 2004].

¹<http://websemantique.org/RDF>

IV.4 Couche ontologique

Le terme ontologie est emprunté de la philosophie où il fait référence à la science qui « étudie l'être en tant qu'être ». Avec l'émergence de l'ingénierie des connaissances et du web sémantique, ce terme a pris une tout autre tournure pour désigner la problématique de représentation et de manipulation des connaissances dans un système informatique [Psyché *et al*, 2003]. Elles sont définies comme « *une spécification explicite d'une conceptualisation* » [Gruber, 1993]. En mettant l'accent sur l'importance du partage et la réutilisation des connaissances, cette définition a été reprise et modifiée : « *Une ontologie est une spécification formelle d'une conceptualisation partagée* » [Borst, 1997].

L'ontologie n'est en fin de compte qu'une modélisation du monde réel en concept et relation entre ces concepts. Les principales composantes qu'on peut distinguer sont donc les suivantes [Gomez-Pérez, 1999] :

- **Concept** : C'est la représentation abstraite des éléments du domaine. On peut également les appeler termes ou classes. Ces concepts peuvent être classés selon différents critères dans la taxonomie (niveau d'abstraction, atomicité, etc.) [Gomez-Pérez, 1999].
- **Relations** : Elles expriment les associations entre les différents concepts définis dans la taxonomie. Les différents types de relations qui peuvent exister sont : « *Spécialisation/Généralisation* », « *Agrégation ou Composition* », « *associé à* », « *composé de* », Etc.
- **Fonctions** : Il s'agit des relations particulières où un élément est défini par les n-1 autres éléments.
- **Axiomes** : Constituent des assertions considérées toujours comme vraies.
- **Instances** : Ce sont des exemples particuliers de concepts.

L'ontologie et l'ensemble des instances constituent la base de connaissance du système informatique. La figure 2.6 représente un exemple d'une ontologie élémentaire.

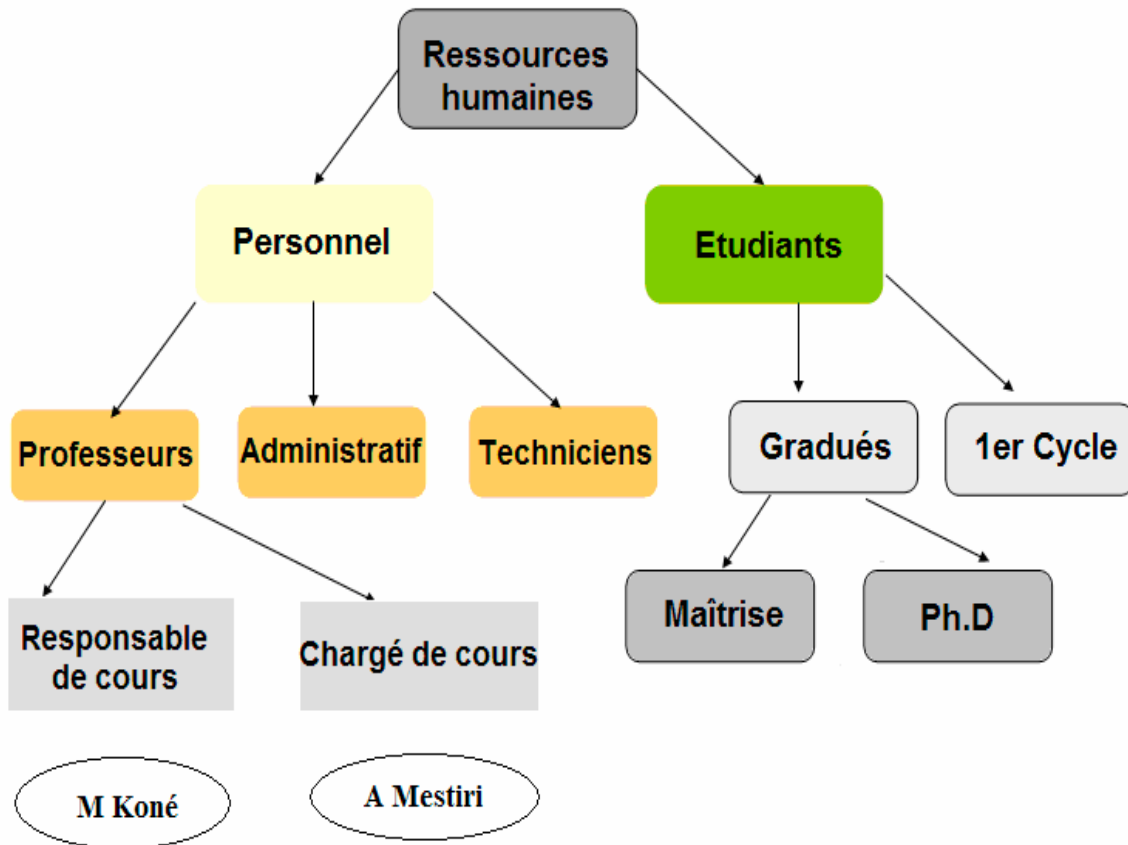


Figure 2-6: Exemple d'une ontologie élémentaire du domaine.

Même si le besoin de développer une ontologie est très varié et dépend du domaine d'application, nous pouvons facilement énumérer un certain nombre d'utilités, notamment :

- **La connaissance du domaine :** Les ontologies permettent la modélisation des connaissances dans un domaine particulier, dans lequel opère le système à développer.
- **La communication :** les ontologies assurent une communication fiable et hétérogène entre personnes et machines (agents logiciels ou organisations) [Uschold et al., 1996] du fait qu'elle permet de mettre en place un langage ou un vocabulaire conceptuel commun.
- **L'interopérabilité :** La représentation explicite des connaissances dans un domaine donné sous forme d'une ontologie, permet à son tour une plus grande réutilisation, un partage plus large et une interopérabilité plus étendue [Uschold et al., 1996].

- **L'aide à la spécification des systèmes :** La représentation conceptuelle des éléments du domaine, permet aux systèmes de réaliser des raisonnements logiques, qu'on appelle inférences, et de sortir avec des conclusions capables d'aider l'utilisateur ou le gestionnaire dans ses décisions.
- **L'indexation et la recherche d'information :** Dans le web sémantique, d'une façon générale, et dans notre application en particulier, les ontologies sont utilisées pour indexer et décrire les ressources utilisées. Cela permet une plus grande précision dans les résultats des recherches ou d'assignation des ressources.

Notons que l'apport des ontologies dans les systèmes informatiques basés sur la technologie sur web sémantique est très considérable et consiste surtout en la réutilisation et le partage des connaissances.

La construction d'une ontologie doit passer par différentes phases, à commencer par l'analyse des besoins, jusqu'à la validation et l'évaluation, et ce, comme le montre la figure 2.7.

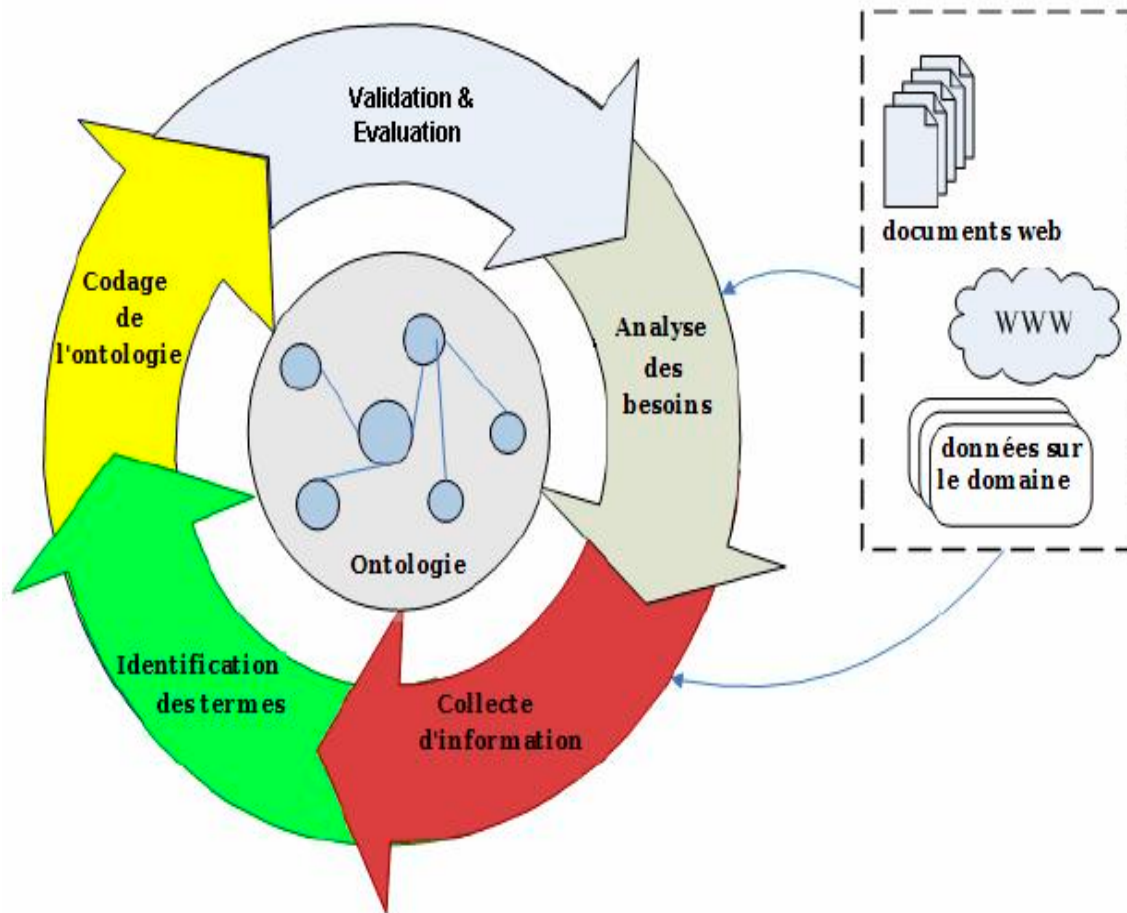


Figure 2-7: Différentes phases l'ingénierie ontologique.

Cette élaboration doit passer par une phase d'ingénierie qui détermine son cycle de vie. Étant donné que ces ontologies seront utilisées comme composantes de systèmes logiciels, leur développement doit s'appuyer sur les mêmes principes que ceux utilisés en génie logiciel [Benayache, 2005]. En effet, « *la vie d'une ontologie passe par les états suivants : spécification, conceptualisation, formalisation, intégration, implantation, et maintenance [...]. Ce cycle de vie permet l'inclusion, le déplacement où la modification de définitions n'importe quand durant le cycle de vie de l'ontologie. L'acquisition, la documentation et l'évaluation de connaissances sont des activités de support qui sont effectuées pendant la majorité de ces états.* » [Fernandez et al, 1997]. La figure 2.8 récapitule ces différentes activités.

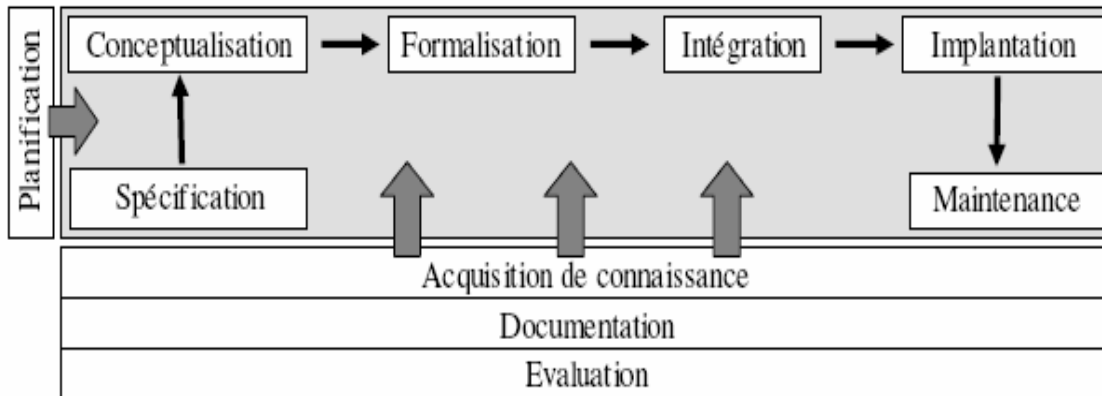


Figure 2-8:Le cycle de vie d'une ontologie (source :[Fernandez et al, 1997]).

Nous reviendrons plus en détail sur la méthodologie adoptée dans le cadre de notre application, tout au long du chapitre 4, qui traitera entre autres de l'architecture du système et de la conception de l'ontologie.

Parler de la couche ontologique, revient implicitement à parler des langages ontologiques, dont le plus connu est le OWL (*Ontology Web Language*). Il s'agit d'un vocabulaire XML basé sur le RDF [OWL-W3C 2004] qui permet d'élaborer des ontologies web structurées. Ce langage est recommandé par le W3C comme un standard pour le web sémantique depuis 2004 et constitue un pilier pour le web sémantique, selon Tim Berners-Lee [Berners-Lee, 2004]. Comme l'indique la figure 2.9, le OWL dérive de la fusion des deux langages OIL et DAML

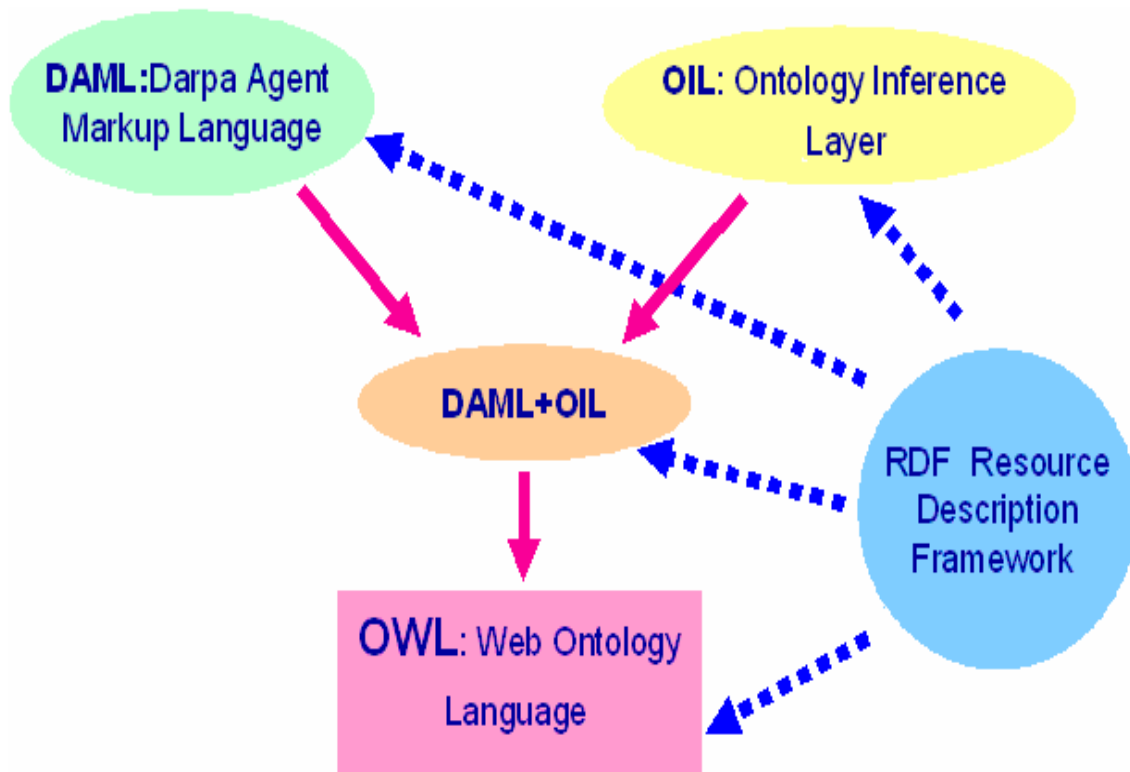


Figure 2-9: Évolution des langages ontologiques.

Le OWL repose sur le RDF, dans l'architecture en couche proposée par Tim Berners-Lee, du fait qu'il vient lui apporter un niveau sémantique supplémentaire. En effet, ce langage ontologique a pour principale caractéristique de décrire des concepts et des relations entre ces concepts, afin de permettre un raisonnement logique au sein des systèmes d'information, qu'on appelle inférences. Dans le cadre de notre application, nous utiliserons les ontologies pour associer l'information d'une ressource (document soumis, évaluateur enregistré, etc.) à des structures de connaissances et à des règles d'inférence.

Selon le niveau d'expressivité qu'on veut exprimer, Le OWL peut être défini en trois sous langages complémentaires (voir figure 2.10) :

- **OWL Lite** : Il s'agit d'un sous-ensemble de OWL qui permet d'exprimer la classification, qu'on appelle taxonomie, et les relations simples entre les classes. Ce sous langage ne permet pas d'exprimer des contraintes complexes sur les classes ou sur les associations.
- **OWL DL** : Permet un meilleur niveau d'expressivité tout en maintenant la complétude (garantie de calculer toutes les conclusions) et la décidabilité (tous les calculs doivent être achevés en un temps fini) [Fadhel, 2005]. Ce sous-ensemble repose sur les caractéristiques de la logique descriptive pour inclure des propriétés utiles aux systèmes de raisonnement [Koné, 2006].
- **OWL Full** : Ce sous-ensemble offre un maximum d'expressivité, mais sans aucune garantie de calcul. OWL permet donc à une ontologie d'augmenter le sens du vocabulaire prédéfini [Fadhel, 2005].

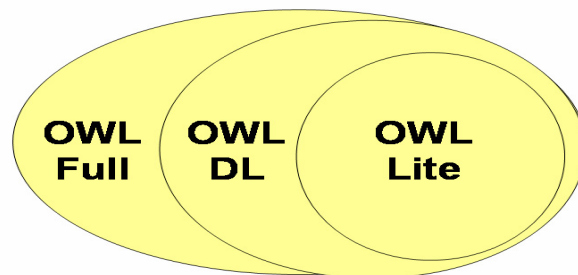


Figure 2-10: Évolution des langages ontologiques.

La figure 2.10 montre explicitement que toutes les ontologies OWL Lite sont perçues comme des OWL DL légalles. Sur la même logique, toutes les OWL DL sont considérées comme des ontologies OWL Full.

IV.5 Couche logique

La couche logique repose sur les langages ontologiques dans l'architecture recommandée par le W3C. En général, nous utilisons la couche logique pour exprimer les règles d'inférences.

Une large variété de logiques a été conçue jusqu'à présent. Étant le formalisme le mieux apprécié dans la représentation de la connaissance, la logique descriptive est celle qui est, généralement, la plus adoptée pour la représentation des règles d'inférences. La logique descriptive est définie comme étant « *une famille de formalismes de représentation de la connaissance basée sur la logique. Elle est conçue pour représenter et raisonner sur la connaissance d'un domaine d'application d'une manière structurée et bien comprise. Elle dérive des réseaux sémantiques* » [Sattler, 2003].

Il existe deux concepts de base dans les logiques descriptives, à savoir, les concepts (prédicats) et les rôles (relations binaires). Les notions de base à prendre en considération, sont la **Satisfiabilité** et la **subsumption**. Un concept « C » est dit satisfaisable quand son expression ne réfère pas à un ensemble vide. En d'autres termes s'il existe une interprétation possible pour ce concept. La Satisfiabilité est un cas particulier de la subsumption, qui n'est autre que la relation de généralisation/spécialisation. C'est pourquoi on parle d'une relation entre la Satisfiabilité et la subsumption).

La figure 2.11 représente l'architecture proposée pour l'inférence basée sur la logique descriptive.

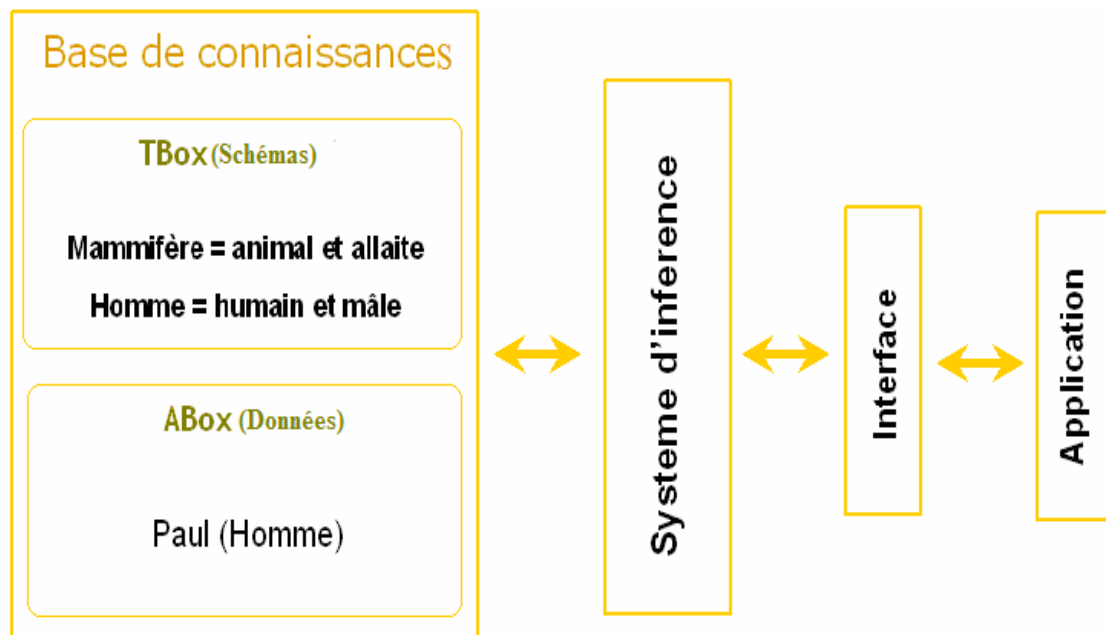


Figure 2-11: Architecture des bases de connaissances.

Les deux éléments principaux d'une telle base de connaissances, sont le **TBox** (*Taxonomy Box*) et le **ABox** (*Assertional Box*). Le premier représente la connaissance conceptuelle du domaine et englobe un ensemble de formules relatives aux informations terminologiques. Le second, quant à lui, définit les instances, c'est-à-dire l'ensemble de formules relatives aux informations sur les assertions. Celles-ci sont instables et dépendantes du domaine [Natalya, 2002]. Ainsi, toutes les informations connues sont alors modélisées comme un couple $\langle T, A \rangle$.

IV.6 Couche Confiance et preuve

Si les trois premières couches ont déjà été standardisées et recommandées par le W3C, dans le cadre de l'architecture du web sémantique, il en est tout autrement pour les couches de haut niveau, que très peu d'auteurs évoquent dans leurs littératures. En effet, les recherches menées jusqu'à présent, dans ce contexte, se sont plus concentrées sur le référencement des ressources, la structuration de l'information (XML), l'affectation des métadonnées (RDF), la modélisation du monde réel sous forme d'ontologies (OWL), en faisant abstraction de la sécurité et des preuves qui servent fondamentalement à justifier la pertinence de l'information.

La couche **Preuve** a pour but de prouver la pertinence de l'information retournée par les couches de plus bas niveau et des déductions obtenues à partir des inférences. Une des façons de le faire est de garder trace des sources d'information et des raisonnements effectués. Malheureusement, à l'heure où nous rédigeons ce mémoire, il n'existe encore aucun langage de preuves standardisé par le W3C. Un langage de preuve constitue un moyen simple pour prouver si une déclaration est juste ou pas. Une instance de ce dernier consiste en général en une liste de toutes les étapes d'inférence par lesquelles a transité l'information en question¹.

Le web est un environnement très ouvert et dynamique. De ce fait, toute personne est donc en mesure d'éditer et de publier des informations de façon très simple. La couche **Confiance**, dans l'architecture proposée par Tim Bernes-Lee, a pour objectif d'évaluer la fiabilité de l'information et des raisonnements. Cette couche repose sur les signatures numériques, le cryptage des données et sur la fiabilité des sources d'information (agents de

¹<http://infomesh.net/2001/swintro/>

confiances, certifications, etc.) [Yang, 2006]. Le web ne pourra atteindre son plein potentiel que si les utilisateurs aient confiance dans les transactions et la qualité de l'information fournie [Antoniou, 2004].

V Conclusion

Il va sans dire que l'apport du web sémantique est d'une importance capitale pour gérer intelligemment un contenu en croissance permanente. Le web sémantique est perçu depuis son apparition comme la technologie du futur [Berners-Lee, 2001]. Cela explique d'ailleurs l'intérêt majeur que manifeste le W3C à son égard, notamment à travers les nombreuses recherches et publications. Cependant, et malgré cette forte émergence qui a marqué la dernière décennie, cette technologie n'est toujours pas au point et son adoption par les industriels fait encore l'objet de nombreuses polémiques.

En effet, contrairement aux premières couches qui ont déjà fait l'objet de standardisation par ce consortium du web, il en est autrement pour celles de plus hauts niveaux, rarement évoquées dans les ouvrages. Seuls l'avenir et les recherches constantes sauront ouvrir les portes à cette nouvelle technologie et faire profiter pleinement les utilisateurs de son apport.

Chapitre 3: Étude du domaine d'application

I Introduction

Un système de gestion de conférence, est une application en ligne destinée à la gestion complète d'une conférence, à savoir : la création du site web officiel, l'évaluation des soumissions, l'inscription des candidats, l'indexation et le partage des résultats des soumissions. Notre projet porte sur la réalisation de la première application de gestion de conférences basée sur les technologies du web sémantique.

L'idée de ce projet consiste à faire une réingénierie d'un outil open source actuellement disponible sur le net, qu'est le « OpenConf », et adapter la plateforme de manière à ce qu'elle s'accorde tout à fait avec les enjeux du web sémantique. Le futur système pourra ainsi profiter pleinement de ses avantages afin de répondre aux faiblesses enregistrées par les applications actuelles. Un sens sera ainsi donné à toutes les ressources pour les rendre interprétables par la machine. Celle-ci pourra alors réaliser tous les raisonnements logiques et nécessaires pour automatiser les tâches et raffiner la recherche et l'assignation, et ce, en se basant sur l'ontologie du domaine.

Selon les contraintes initiales du projet, notre système n'apporte rien de nouveau, sur le plan fonctionnel, par rapport à ceux disponibles actuellement sur le marché. Nous avons même essayé de préserver la même ergonomie, de manière à ce que l'utilisateur ne se rende pas compte du changement. Cependant, c'est au niveau des besoins non fonctionnels que la différence se fera sentir considérablement. La refonte de cette application sous la nouvelle plateforme sémantique, vise une affectation plus soignée des articles aux évaluateurs conformément à leurs expertises. Nous ciblons également l'optimisation des résultats des recherches et l'automatisation de certaines tâches réalisées jusqu'à maintenant à la main.

À travers ce chapitre, nous visons à introduire tous les besoins fonctionnels que doit couvrir une application de gestion de conférences. Nous commencerons d'abord par présenter, en détail, tout le processus de gestion des conférences. Par la suite, nous effectuerons une

analyse complète des acteurs (utilisateurs potentiels), ainsi que des différents cas d'utilisations du système. Enfin, et avant de conclure cette section, nous présenterons les applications en ligne les plus connues pour la gestion de conférence, tout en justifiant le choix de l'outil OpenConf.

II Le processus de gestion des conférences

Le processus de gestion de conférences se déroule sur deux étapes majeures. La première consiste à organiser la conférence en constituant le comité d'organisation et en invitant des examinateurs à évaluer les articles qui seront soumis par différents auteurs. La seconde phase définit la procédure de traitement de tout article reçu, depuis sa soumission, jusqu'à l'avis de refus ou d'acceptation qui sera retourné à son auteur, en passant par les différentes étapes d'évaluation et de classement.

II.1 Première étape : Organisation de la conférence

Cette étape commence dès lors qu'une personne décide d'organiser une conférence sur un thème donné. Il en est donc le président, qu'on appelle « *Chair* ». Celui-ci va commencer par réunir un certain nombre de participants pour constituer le *comité de programme*, dont le rôle est de gérer le processus d'évaluation et d'organiser les réunions pour classer les articles évalués et décider de leurs sorts, comme représenté par la figure 3.1.

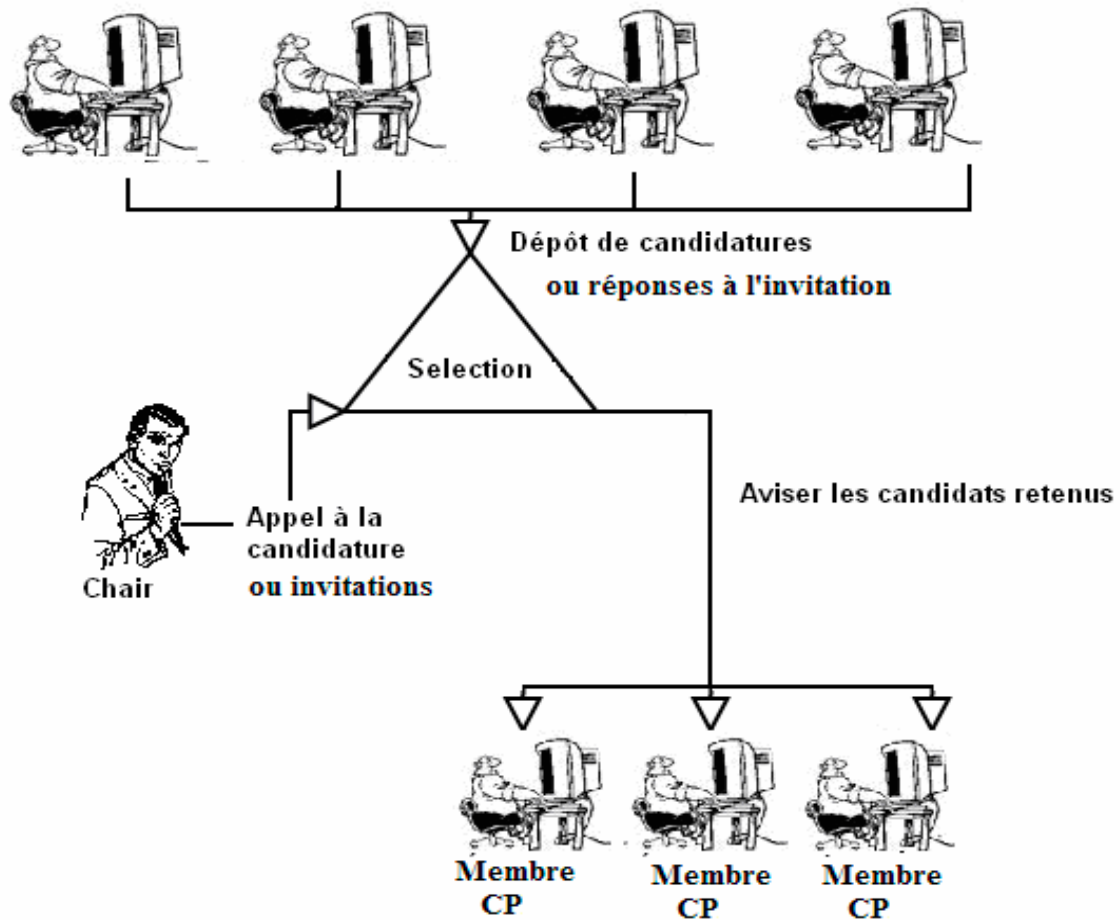


Figure 3-1: Procédure de sélection des membres du Comité de Programme.

Par la suite, et selon les cas, un appel de dépôt de candidature est lancé ou des invitations personnalisées sont envoyées à certains candidats pour constituer le comité d'évaluateurs (Figure 3.2). Il s'agit de chercheurs ou d'experts du domaine qui s'engagent à évaluer quelques articles qui leur seront assignés et de faire part de leurs appréciations au comité du programme. Nous aborderons plus en détail, dans la section suivante, le rôle des évaluateurs et la procédure qu'ils doivent suivre.

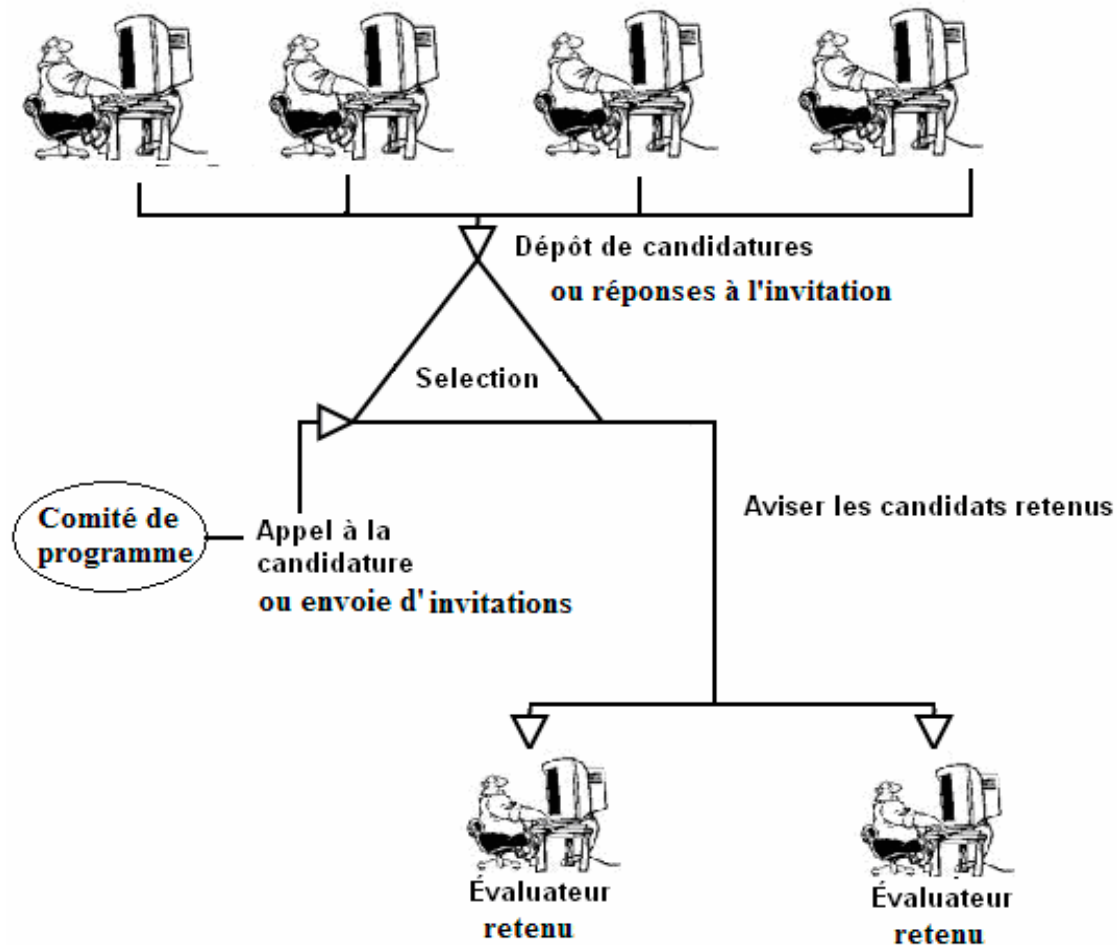


Figure 3-2: Procédure de sélection des évaluateurs

II.2 Deuxième étape : Processus d'évaluation des articles :

Une fois l'étape préliminaire achevée, un appel à la soumission est alors lancé au public. Tout en respectant l'échéance, chaque auteur est alors en mesure de se connecter sur le site de la conférence et de soumettre en ligne son article. Tous les documents reçus seront stockés sur le serveur. En s'inspirant de projet «Trial Solution» [Dehors Sylvain et al, 05], nous appellerons cet espace d'archivage « entrepôt virtuel ».

Une fois arrivée à l'échéance, la phase de soumission est close et les articles sont assignés aux évaluateurs conformément à leurs expertises. C'est notamment à ce niveau que nous devons intervenir pour améliorer les résultats et répondre aux frustrations des utilisateurs. Selon les paramètres de configuration, chaque évaluateur reçoit un nombre N d'articles, et

chaque article peut être évalué par un nombre 'M' d'évaluateurs, et ce, en tenant compte du thème et des mots clés de l'article, d'une part, et de l'expertise et l'expérience, des évaluateurs d'autre part. Il est à noter que l'affectation des articles doit se faire dans l'anonymat total, dans le sens où les évaluateurs ne connaissent pas l'auteur de l'article et vis et versa. L'appréciation étant faite, selon le barème initialement fixé dans la phase de configuration de l'application, il ne reste plus à l'évaluateur que de remettre le tout au comité de programme. Ce dernier se réunira afin de discuter et choisir les articles qui seront retenus pour la conférence.

Notons que dans beaucoup d'applications, le classement se fait encore de manière manuelle. Nous nous proposons, dans cette application, d'automatiser cette tâche par des inférences basées sur l'ontologie du domaine. Ceci fera l'objet du second volet qui sera traité par notre application.

Une fois que les décisions seront prises, il ne reste plus au comité de programme que d'aviser chaque auteur de la décision qui a été prise au sujet de son article. Trois cas sont envisageables, à savoir, une acceptation immédiate, un refus définitif ou un refus conditionnel. Dans ce dernier cas, l'auteur doit modifier ou corriger certaines parties en tenant compte des remarques des évaluateurs. Il sera, bien entendu, tenu de remettre la nouvelle version de son travail, dans un certain délai qui lui sera également communiqué.

La figure 3.3 récapitule schématiquement la procédure décrite ci-dessus, pour la soumission en ligne d'un ou plusieurs articles. Il est clair, selon cette figure, que la phase d'évaluation des articles est cyclique. Cela s'explique par le fait que tout article retourné à son auteur pour modification, peut être soumis de nouveau sur le serveur, selon le délai qui lui a été accordé. Si cela est le cas, le document doit absolument suivre le même processus de base, depuis son affectation à un évaluateur, jusqu'au compte rendu à l'auteur.

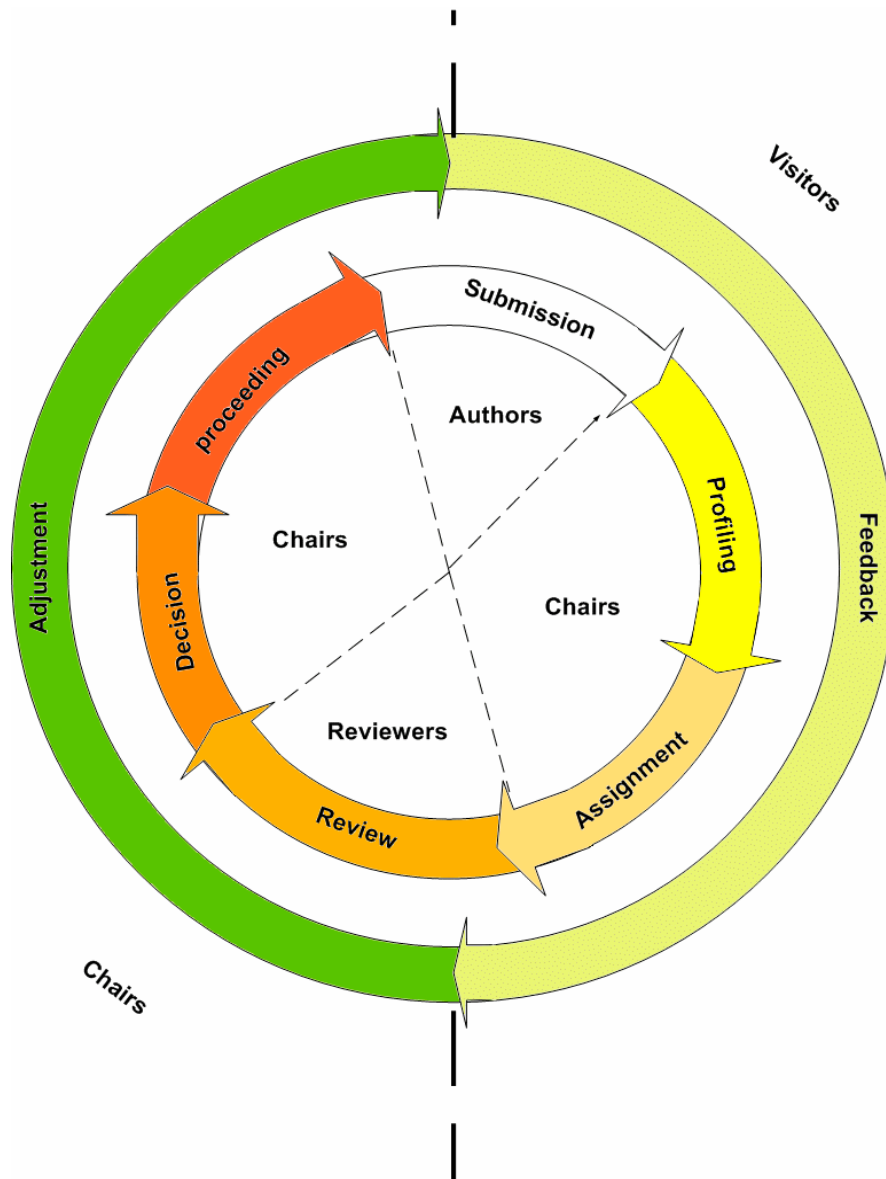


Figure 3-3: Procédure de traitement des articles reçus pour une conférence

III Étude du domaine : acteurs et cas d'utilisation

Dans cette section, nous présentons les utilisateurs potentiels du système, qu'on appelle acteurs, ainsi que les cas d'utilisation qui leur sont associés. En se basant sur plusieurs études menées sur ce sujet, notamment [Richard Van De Stadt] et [Cetic, 03], on peut clairement distinguer 4 principaux acteurs :

III.1 L'administrateur

Son rôle est de gérer les aspects techniques et informatiques de l'application (maintenance, correction de bugs, etc.). Parmi les tâches quotidiennes de l'administrateur, nous citons, la gestion de l'apparence du site, la définition des paramètres de configuration de l'application, et la gestion du compte du chairman, comme le schématise la figure 3.4.

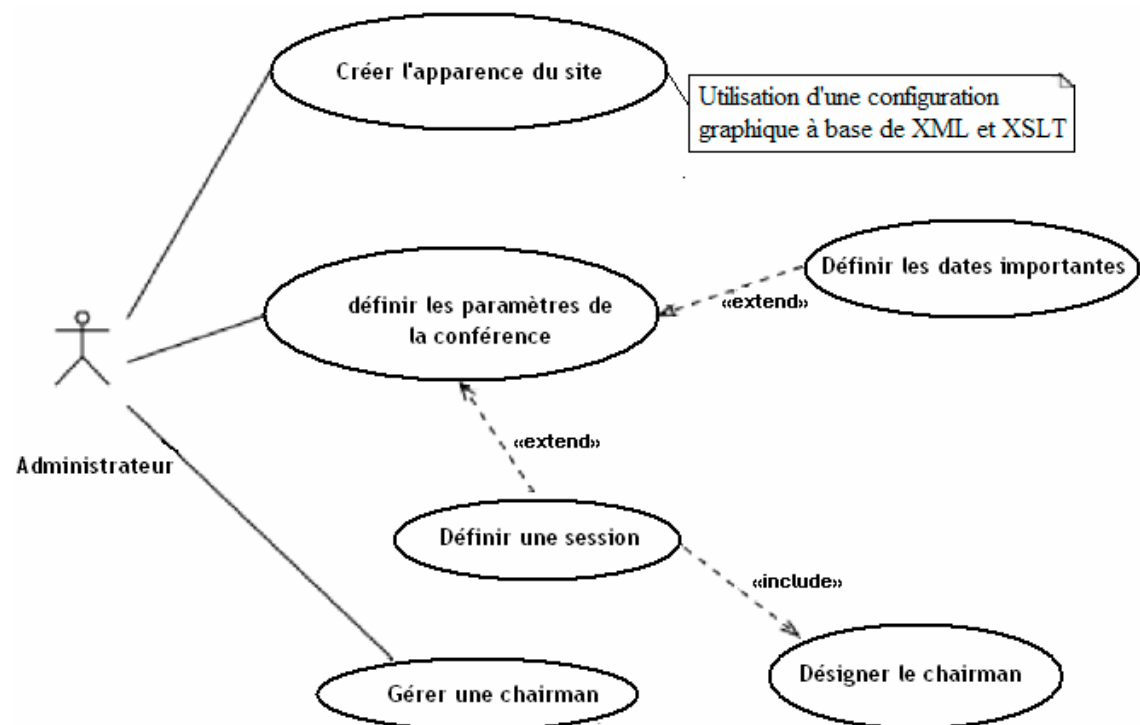


Figure 3-4: Diagramme de cas d'utilisation liés à l'administrateur (Source [Cetic, 03]).

III.2 Les auteurs

Il s'agit de toute personne ayant rédigé un article qui s'articule autour du thème principal de la conférence et qui souhaite le soumettre en ligne pour évaluation. Cette soumission est précédée par une phase d'identification, où l'auteur doit remplir un formulaire pour fournir un certain nombre d'informations, notamment, son nom, un mot de passe, un bref résumé (*Abstract*) et les différents thèmes qu'il aborde.

Une fois que ces informations ont été correctement saisies, le système donne la main à l'utilisateur pour sélectionner le document en question à partir de son arborescence locale. Un identifiant unique, dit « ID », est automatiquement attribué à cette soumission. Celui-ci est indispensable, au même titre que le mot de passe de l'utilisateur, pour le suivi de son article ou toute modification ultérieure. La figure 3.5 est un diagramme qui récapitule les différents cas d'utilisation associés à l'auteur.

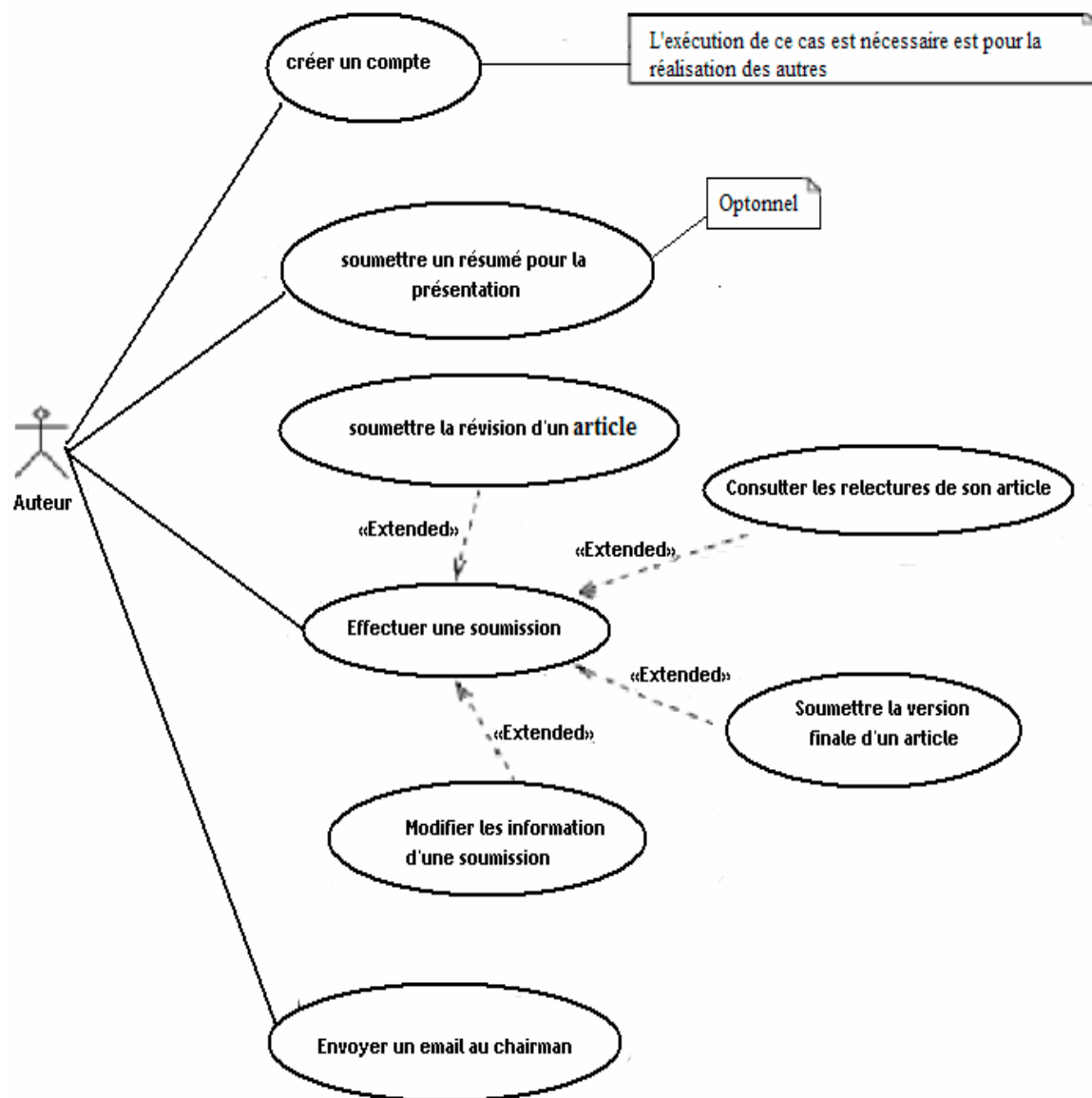


Figure 3-5: Diagramme de cas d'utilisation liés au chairman (Source [Cetic, 03]).

III.3 Les évaluateurs

Un évaluateur est une personne ayant développé une certaine expertise dans un ou plusieurs thèmes évoqués par la conférence. Celle-ci est alors invitée à partager ses connaissances, en évaluant les articles soumis par les auteurs, afin de décider de leur pertinence.

Un système de gestion de conférences doit être en mesure d'assigner des articles aux évaluateurs conformément à leurs expertises. Idéalement, le système devrait aussi tenir compte de leurs expériences dans le domaine. Toutefois, un évaluateur peut sélectionner lui-même les articles qu'il souhaite évaluer. Dans le cas échéant, la méthode la plus utilisée est la suivante :

- L'évaluateur accède à la liste des thèmes de la conférence.
- En cliquant sur un thème, le système affiche la liste des articles qui s'articulent autour de celui-ci.
- L'examineur coche les cases correspondantes aux articles qu'il souhaite évaluer.
- Pour chaque article choisi, L'évaluateur doit spécifier le niveau de priorité (*Low – Medium – High*).
- En tenant compte de ses préférences et des paramètres de configuration (Nombre d'article par évaluateur, Nombre d'évaluation pour chaque article, etc.), le système affecte à chaque évaluateur un certain nombre d'ouvrages parmi ceux qu'il a sélectionnés.
- Après la lecture de chaque article, l'évaluateur doit remplir une fiche d'évaluation pour noter ses impressions. Cette dernière peut prendre plusieurs formats (Note sur la présentation, note sur le contenu, ...). Selon le modèle « MAKE CHAMPIONS EXPLICIT » de Nierstrasz [Nierstrasz, 2000], la classification suivante est utilisée suite à l'évaluation :
 - « *I'll Champion this paper rat the PC meeting* »: C'est à dire que l'évaluateur accepte et soutient cet article.

- « *I can accept this paper, but I'll not champion it* »: C'est à dire que l'évaluateur accepte éventuellement cet article, mais ne le supporte pas devant le comité. Il peut être sujet à modifications, auquel cas il sera retourné à son auteur pour le corriger.
- « *This paper should be rejected, though I will not fight strongly against it* »: Cela veut dire que l'examineur rejette cet article, à priori, mais qu'il pourrait être accepté ultérieurement.
- « *Serious Problems. I'll argue to reject this paper* »: L'article est catégoriquement refusé par l'évaluateur.

De plus, les évaluateurs doivent indiquer leurs compétences et leurs expertises dans le sujet évoqué par l'article. Ceci doit être fait pour les articles qui ne correspondent pas à l'expertise de l'évaluateur, afin de juger de l'utilité de faire évaluer l'article par quelqu'un d'autre.

Il est très important de noter que la décision finale revient toujours au comité de programme, et ce, quel que soient les impressions de l'évaluateur sur un article.

La figure 3.6 est un diagramme qui récapitule les différents cas d'utilisation associés à un auteur qui souhaite soumettre son article pour une conférence.

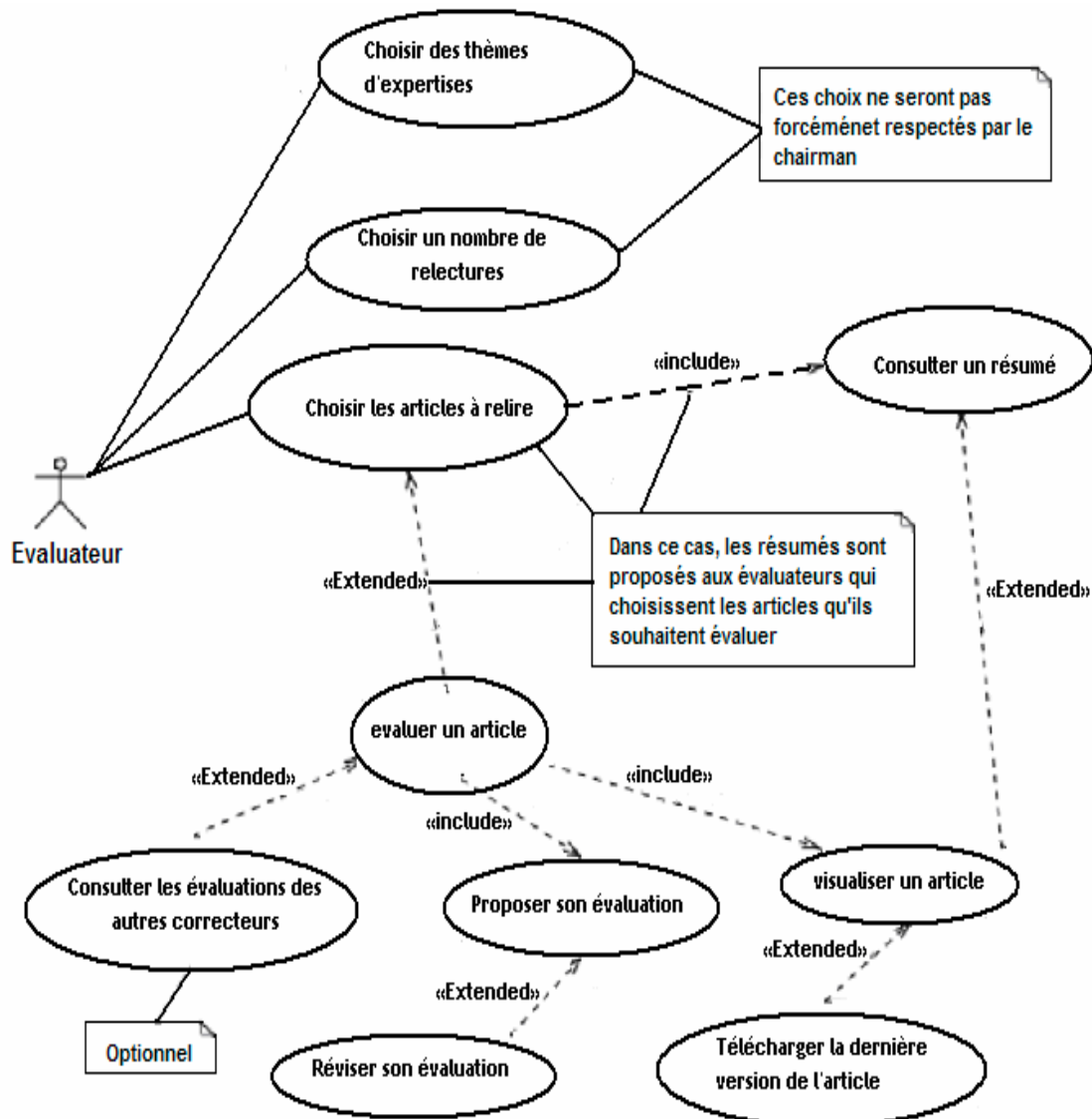


Figure 3-6: Diagramme de cas d'utilisation liés à l'évaluateur (Source [Cetic, 03]).

III.4 Les membres du comité de programme

Dès lors qu'une application de gestion de conférence telle que « **Cyber Chair** » ou autre est utilisée, le rôle de ces membres est nettement minimisé. À la base, leurs tâches consistent à assurer une assignation conforme et ciblée des articles aux différents évaluateurs, et à organiser des réunions pour décider du sort de chaque article, en se basant sur les appréciations des évaluateurs. Le comité du programme (**CP**) a ainsi pour rôle principal d'assurer le contrôle et la coordination du processus d'évaluation. C'est les membres du comité de programme qui votent aussi pour le meilleur article.

La figure 3.7 est un diagramme qui récapitule les différents cas d'utilisation associés à l'auteur.

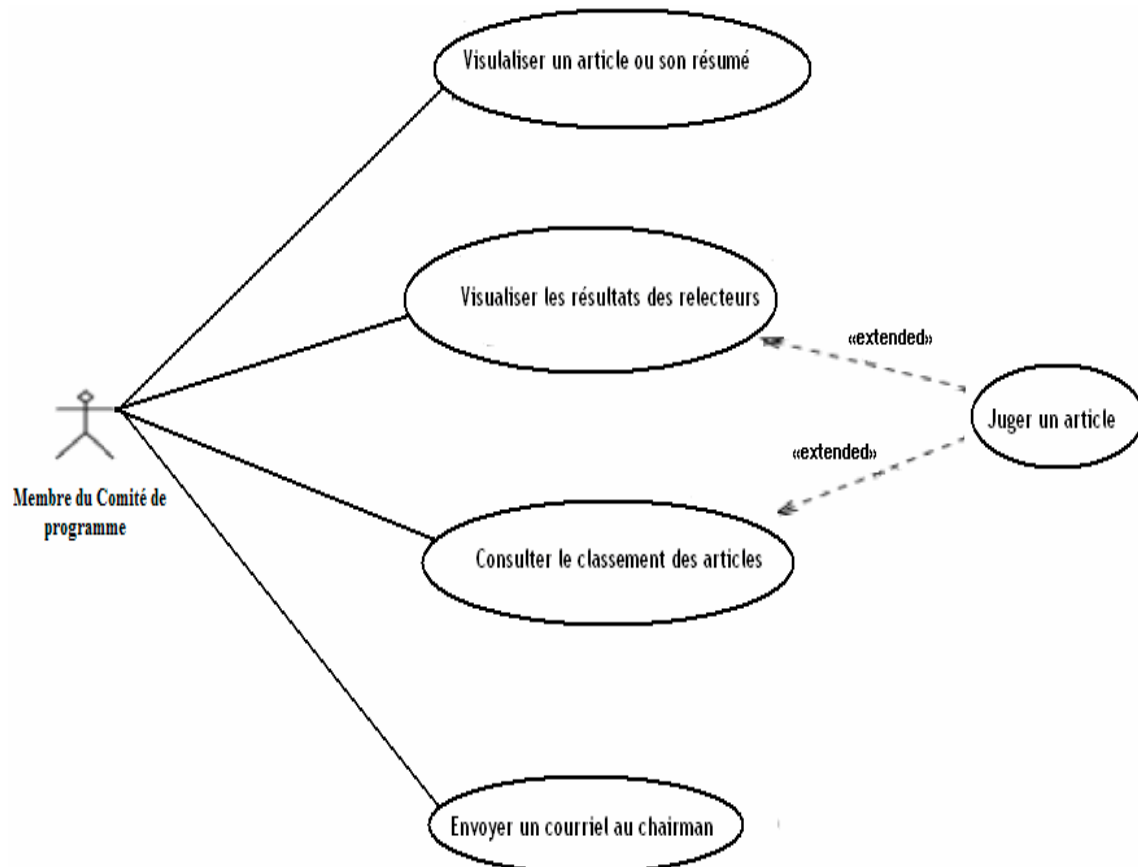


Figure 3-7: Diagramme de cas d'utilisation liés aux membres du comité de programme (Source [Cetic, 03]).

À la lumière de ce qui a été présenté comme acteurs et cas d'utilisations qui leur sont associés, il est fort à constater que la plus part de la recherche s'articule autour du problème majeur de l'assignation des articles reçus, aux différents évaluateurs. La figure 3.8 est un diagramme d'états qui récapitule les différentes phases de ce processus

Cependant, l'assignation n'est pas le seul problème pris en considération par la refonte de cette application. En effet, on parle également de recherches raffinées des ressources et de l'automatisation de certaines autres tâches. Le futur système donnera la main à tout utilisateur de spécifier l'URL de son profil FOAF, s'il y a lieu, afin de lui éviter de saisir toutes les données. L'auteur se voit aussi offrir le moyen de spécifier l'URL de la description RDF de son article, afin ne pas avoir à re-saisir toutes ces données sur le système

IV Présentation des applications les plus connues

Le monde du management, sous tous ses angles, est le principal bénéficiaire de l'avancée technologique et informatique. Nous nous intéressons à un cas particulier de gestion, dans le cadre de ce projet, celui de la gestion et l'organisation de conférences. Cette dernière est une tâche très lourde et délicate pour la gérer manuellement. Un grand nombre d'applications en ligne ont fait leurs apparitions sur le marché. Malgré les différents avantages que présente chacune d'elles, elles représentent toutes approximativement les mêmes faiblesses:

- Les résultats frustrants des assignations.
- La nécessité de l'intervention du comité de programme pour classer les articles
- La nécessité de saisir toutes les données dans les formulaires, même si ces dernières sont disponibles quelque part dans le web.

Un très grand nombre d'outils en ligne de gestion de conférences sont disponibles actuellement sur le net, parmi lesquels nous citons, CyberChair [CyberChair], ConfTool [ConfTool] et OpenConf [OpenConf]. C'est sur ce dernier que nous effectuons une réingénierie, pour introduire une plateforme de web sémantique. La figure 3.9 représente la page principale de cette application.

Canadian Semantic Web Working Symposium 2007	
OpenConf Conference Management System	
OpenConf Home	Email Chair
<p>Authors:</p> <ul style="list-style-type: none"> • Submit a Paper • Edit Submission • Re-Upload Paper • View Paper <p>Review & Program Committees:</p> <ul style="list-style-type: none"> • Sign In • Sign up keycode required: <input type="text"/> <input type="button" value="Enter"/> <p>Program Chair:</p> <ul style="list-style-type: none"> • Sign In 	
<p>Powered by OpenConf Copyright ©2002-2006 Zakon Group LLC</p>	

Figure 3-9: Capture d'écran de la page principale de l'outil "OpenConf"

OpenConf est d'abord, un outil conçu et produit sous **PHP/MySQL** par le « *Zakon Group* » en 2003 [zakongroup]. Il s'agit donc d'une application en ligne, déployée sous une architecture simple à 2 tiers, plus connue sous le nom de Client/serveur. Nous reviendrons plus en détail, dans le chapitre suivant, sur l'aspect architectural.

Le choix de cet outil, dans notre projet de ré-ingénierie est fondé sur les nombreux avantages qu'il présente, à savoir :

- **Facilité d'installation** : il suffit de suivre les trois ou quatre points décrits dans la documentation disponible en ligne [OpenConf].
- **Disponibilité du code** : Du fait qu'il s'agit d'une application Open source, le code est disponible et bien commenté, ce qui rend la migration vers la plateforme web sémantique beaucoup plus simple.
- **Popularité** : Depuis sa création en 2003, OpenConf a été utilisé dans plus de 600 conférences. Rien que pour l'année 2007, qui est encore à son tout début au moment où nous rédigeons ce mémoire, cet outil est utilisé dans plus de 131 conférences dans le monde [OpenConf].
- **Personnalisation** : La conception du front-end est très simple et efficace. OpenConf permet également au chairman de configurer cet outil à sa guise. La figure 3.10 représente quelques paramètres parmi tant d'autres, que le chairman peut modifier. OpenConf est conçu de façon à permettre un processus de soumission et d'évaluation itératif, mais surtout très simple

Change Config	
Conference Full Name:	<input type="text" value="Canadian Semantic Web Working Symposium 2007"/> (\$OC_confNameFull) <i>Conference name used on Web pages and in email messages.</i>
Conference Short Name:	<input type="text" value="CSWWS07"/> (\$OC_confName) <i>Conference abbreviated name, primarily used in email subject lines.</i>
Chair Email:	<input type="text" value="amine.mestiri.1@ulaval.ca"/> (\$OC_pccemail) <i>Used as the general contact email and in case of errors or other follow-up.</i>
Confirm Email:	<input type="text" value="amine.mestiri.1@ulaval.ca"/> (\$OC_confirmmail) <i>Receives a copy of confirmation emails sent to authors and reviewers (e.g., submissions, edits, signups); see Email Notification below.</i>
Conference URL:	<input type="text" value="http://www.CSWWS07.ca"/> (\$OC_confURL) <i>Conference Web page.</i>
Reviewer Keycode:	<input type="text" value="evaluateur"/> (\$OC_keycode_reviewer) <i>Keycode for signing up as a reviewer committee member.</i>
Program Cmt. Keycode:	<input type="text" value="membre"/> (\$OC_keycode_program) <i>Keycode for signing up as a program committee member.</i>
Paper Formats:	<input checked="" type="checkbox"/> Adobe PDF <input checked="" type="checkbox"/> Microsoft Word <input type="checkbox"/> Corel WordPerfect <input checked="" type="checkbox"/> HTML <input type="checkbox"/> PostScript <input checked="" type="checkbox"/> RTF <input checked="" type="checkbox"/> Text <i>Select at least one paper format to accept.</i> (\$OC_extar)

Figure 3-10: Capture d'écran de la page de configuration de l'outil "OpenConf".

Par ailleurs, il serait dommage et très subjectif de se baser sur l'outil OpenConf, dans le cadre de notre mémoire, sans en citer les inconvénients.

Le problème majeur de cette application en ligne consiste en fait, en l'absence d'une section « Posters » dans le processus d'évaluation, ce qui n'est pas très pratique pour les distinguer des articles normaux¹. Aussi, le processus d'assignation automatique des articles ne

¹<http://wiki.loria.fr/wiki/OpenConf>

prend pas en considération les membres du comité de programme, qui peuvent être évaluateurs en même temps. Dans un tel cas, il faut avoir alors recours à une assignation manuelle.

D'autres problèmes de second ordre viennent s'ajouter à la liste des faiblesses de cet outil, notamment, le fait qu'il ne soit disponible qu'en anglais. C'est au prix d'un certain effort qu'il a été traduit pour la conférence *TALN/RECITAL 2005*¹, puisque les textes sont directement écrits dans le code PHP. La notation, quant à elle, est aussi limitée à quelques critères, alors que d'autres modes de visualisation s'avèrent bien utiles pour toutes les tâches courantes².

V Conclusion

La réingénierie que nous envisageons de réaliser, doit s'effectuer sur la base de l'outil **OpenConf**, en répondant aux mêmes besoins et en essayant, dans la limite du possible, de préserver la même ergonomie.

Comme nous l'avons mentionné précédemment, ce projet devra s'accorder tout à fait avec les enjeux du Web sémantique en répondant conformément aux standards du W3C, et ce, notamment au regard de l'architecture sous laquelle il sera déployé. A cet effet nous aborderons l'aspect architectural tout au long du chapitre qui suit.

Cette nouvelle plateforme devra répondre aux faiblesses enregistrées par les systèmes actuels, notamment l'assignation des articles aux évaluateurs, la saisie automatique de toutes les informations sur un évaluateur ou un article par simple spécification de l'URL de sa description RDF, et l'optimisation du classement des articles.

¹<http://m43.limsi.fr/site/talnRecital05/>

²<http://wiki.loria.fr/wiki/OpenConf>

Chapitre 4: Architecture et solution proposée

I Introduction

Après avoir introduit la description de la technologie du Web sémantique et son contexte général, nous allons entamer la phase de développement tout au long de ce chapitre. Nous commencerons tout d'abord par rappeler la problématique du projet et les faiblesses retenues par notre étude préliminaire sur les applications de gestion de conférences actuelles. Nous aborderons, par la suite, l'aspect architectural du système, pour présenter la plateforme proposée qui nous a permis de faire migrer l'application existante sous le Web actuel vers la technologie du web sémantique. Le troisième volet de ce chapitre portera, quant à lui, sur la structuration des données et l'affectation de métadonnées aux différentes ressources. Enfin, nous présenterons en détail l'ontologie sur laquelle repose notre système. Celle-ci constitue sa base de connaissances qui lui permettra de réaliser toutes les inférences nécessaires.

II Problématique

Les conférences, les ateliers de travail et les séminaires constituent un partage de connaissances et d'idées entre des personnes d'une même communauté intellectuelle. Mettre leur gestion au cœur des innovations technologiques doit tenir compte des nombreuses contraintes que nous avons présentées dans le chapitre précédent.

Le projet **SWOC** (*Semantic Web Open Conference*), faisant l'objet de ce mémoire, fait partie de cette réflexion en proposant une nouvelle architecture basée sur la technologie du Web sémantique. Il est très important de mettre l'accent sur la problématique et de rappeler que l'idée de ce projet est d'analyser l'approche Web sémantique dans les applications en ligne de gestion de conférence. Celle-ci s'effectuera à travers la ré-ingénierie d'une application en ligne existante, pour la faire migrer de l'environnement du Web actuel vers la nouvelle génération afin de comparer les résultats et tirer une conclusion quant à l'efficacité de la

technologie que nous étudions. Nous nous proposons ainsi de présenter des solutions innovatrices qui puissent répondre aux faiblesses des outils utilisés actuellement sur le marché (recherche d'articles ou de personnes, assignation des articles souvent peu conforme aux expertises des évaluateurs, etc.). La structuration des informations, l'annotation des ressources, la modélisation des connaissances sous forme d'ontologies et l'intégration d'un moteur d'inférences sont tous des éléments destinés à assurer un meilleur partage de l'information, une plus large réutilisation des données et une gestion plus intelligente du contenu. Chacun d'eux fait l'objet d'une couche de l'architecture du web sémantique [Berners-Lee, 2001].

III Ré-ingénierie de l'architecture

Les applications en ligne actuelles de gestion de conférences sont toutes basées sur une architecture en couches où l'on retrouve les trois niveaux d'abstraction suivants : Données, Traitement et Présentation. La première couche concerne le stockage des données pertinentes à l'application. La seconde, quant à elle, traite les requêtes du client en interrogeant les supports de stockage, en effectuant les différents calculs et en retournant les résultats à la troisième couche. Cette dernière gère alors leur affichage. La figure 4.1 représente schématiquement la superposition de ces couches.

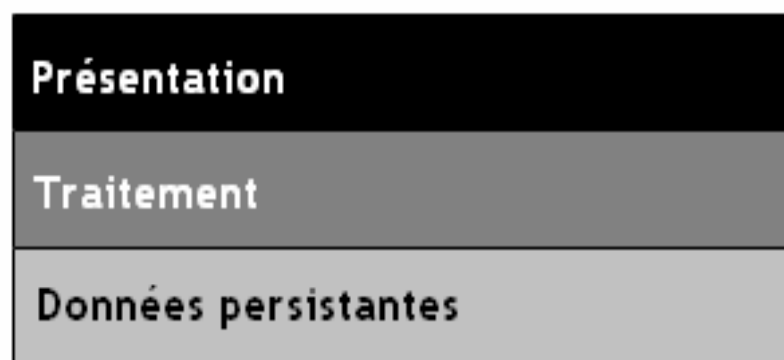


Figure 4-1: Représentation des différentes couches de l'architecture des applications en ligne.

Une telle architecture repose en général sur une topologie à deux tiers, qu'on appelle « **Client - Serveur** », représentée par la figure 4.2.



Figure 4-2: Architecture en « client / Serveur » des applications en ligne.

On retrouve dans cette architecture autant d'intervenants que de niveaux d'abstraction. Ces derniers se partagent, en effet, les tâches comme suit : le serveur de la base de données permet le stockage des informations pertinentes et persistantes pour le système. Le serveur Web, quant à lui, en plus de traiter les requêtes HTTP en provenance du client, gère un moteur PHP lui permettant d'interpréter les scripts PHP attachés à ces dernières.

Dans le cadre de ce projet, nous voulons effectuer une réingénierie de cette architecture, en vue d'intégrer la technologie du Web sémantique, dont les objectifs ont été clairement expliqués au cours des chapitres précédents. À cet effet, nous nous proposons d'insérer un **module web sémantique** entre la couche de données pertinentes et celle du traitement. La figure 4.3 reprend schématiquement cette architecture. Dans le cadre de son projet de conception d'une plateforme de commerce électronique basée sur la technologie du web sémantique, Kun Yang a baptisé cette couche **SVM** (*Semantic Virtual Machine*) [Yang, 2006].

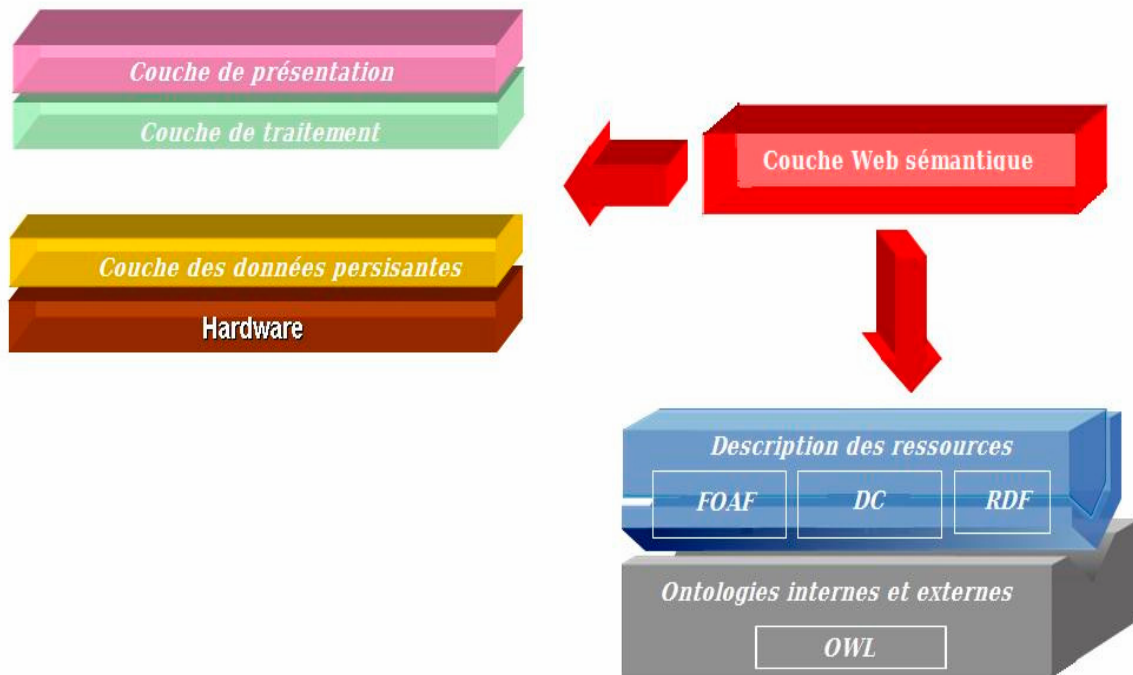


Figure 4-3: Insertion du module « Web sémantique » dans l'architecture en couches des applications en ligne actuelles.

Malgré la présence des annotations sémantiques pour décrire les ressources et les ontologies pour représenter les connaissances du monde réel, il doit toujours exister une part d'interaction entre le système et les différents supports de stockage classiques, tels que les fichiers et les bases de données. Le système peut, en effet, recourir aux données pertinentes stockées dans les différents supports gérés par la couche du plus bas niveau de l'architecture. Aussi, ce dernier se voit souvent programmer pour y stocker certaines de ses données, malgré la présence des couches de données de la plateforme du Web sémantique, C'est la raison pour laquelle la couche web sémantique repose directement sur celle des données.

Par ailleurs, après avoir réalisé les raisonnements logiques nécessaires, deux cas de figure se présentent au module web sémantique. Le premier consiste à communiquer directement ses résultats à la couche de présentation, qui gère alors leur affichage et leur présentation à l'utilisateur. Dans ce sens, le langage le plus utilisé est alors le **XSL** (*eXtensible Stylesheet Language*), étant donné qu'il dérive du XML et que ce dernier est le format utilisé dans la technologie du Web sémantique pour structurer les données. Recommandé par le W3C

depuis 1999, le langage XSL a pour rôle de récupérer les informations à partir de structures XML et de s'occuper de leurs présentations¹. C'est d'ailleurs pourquoi on qualifie très souvent ce dernier de complément de XML [Van Lancker, 2001]. Rappelons, au passage, que si le HTML a été conçu pour afficher l'information via des requêtes HTTP, le XML, quant à lui, n'a pour but que de structurer l'information, et repose donc sur le principe de la séparation entre le contenu et la présentation

Le second cas de figure, quant à lui, consiste à récupérer les résultats d'inférence du module Web sémantique et d'effectuer des traitements supplémentaires avant de les faire passer à la couche supérieure, et ce, avec un langage serveur. C'est d'ailleurs à cette technique que nous avons fait appel dans le cas de notre application, et c'est le langage PHP² qui sera utilisé à cet effet. Le HTML est dans ce cas largement suffisant pour afficher ses données.

Dans le cas de notre application, la couche de traitement joue toujours le rôle d'intermédiaire entre le moteur d'inférence et la couche de présentation, conformément au schéma de la figure précédente. La figure 4.4 est une représentation schématique des relations qui existent entre les différentes parties intervenantes de notre architecture.

¹<http://www.w3.org/TR/1999/REC-xslt-19991116>

²<http://www.php.net/>

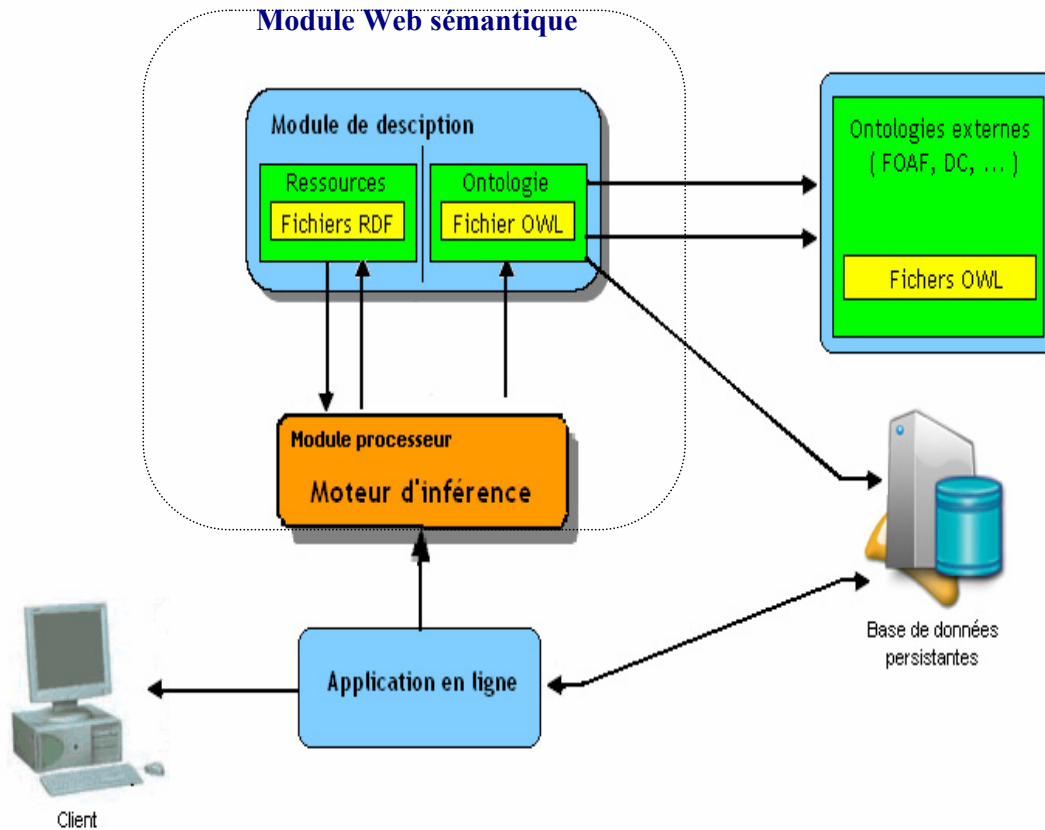


Figure 4-4: représentation des relations existantes entre les différentes parties intervenantes de l'architecture du système SWOC.

Cette architecture divise la couche Web Sémantique en deux modules, à savoir, le module de description et le module processeur. Le premier permet d'annoter sémantiquement les ressources pour les doter d'un sens afin de les rendre interprétables par la machine. Le second, quant à lui, comprend un moteur d'inférences pour effectuer les raisonnements logiques nécessaires, en se basant sur les connaissances du domaine et les descriptions des ressources à travers les métadonnées qui leurs ont été affectées.

À la lumière de ce qui a été présenté, nous pouvons clairement constater que cette nouvelle architecture permet à la couche traitement d'opérer toujours d'une façon classique en jouant le rôle d'intermédiaire entre les données persistantes et leurs présentations, et de faire recours aux techniques du web sémantique pour une meilleure structuration des données et une gestion plus intelligente du contenu. Celle-ci reste donc conforme aux dires de Tim Berners-

Lee qui a affirmé que le web sémantique n'est pas un substitut du web actuel, mais un complément qui vise à structurer l'information et lui donner en sens en vue de favoriser la coopération Personnes-machines [Berners-Lee, 2001].

IV Annotation sémantique des ressources

IV.1 Aperçu

Dans l'approche que nous développons, et qui s'aligne tout à fait avec les enjeux du Web sémantique, l'annotation consiste à assigner des descriptions sémantiques aux différentes ressources utilisées selon des schémas prédéfinis. Celles-ci permettent de les rendre interprétables par la machine et de s'ouvrir ainsi sur de nouvelles possibilités d'automatisation dans le Web. Le langage le plus utilisé à cet effet est le RDF, vu que ce dernier a été défini comme un standard par le consortium du Web, Le W3C. Rappelons que ce langage repose lui-même sur le XML comme nous l'avons expliqué dans le second chapitre. L'annotation sémantique est aussi définie dans ce contexte comme un processus d'instanciation des données à partir d'une ontologie du domaine [Tanior, 2006].

Il existe de plus en plus d'outils qui assistent l'utilisateur dans l'annotation de ses documents ou ressources, notamment l'outil **KIM** (*Knowledge and Information Management*)¹ développé par le laboratoire Ontotext en 2003 [OntoText]. Ce dernier propose un schéma explicite pour expliquer l'annotation sémantique, représenté par la figure 4.5.

¹<http://www.ontotext.com/kim/>

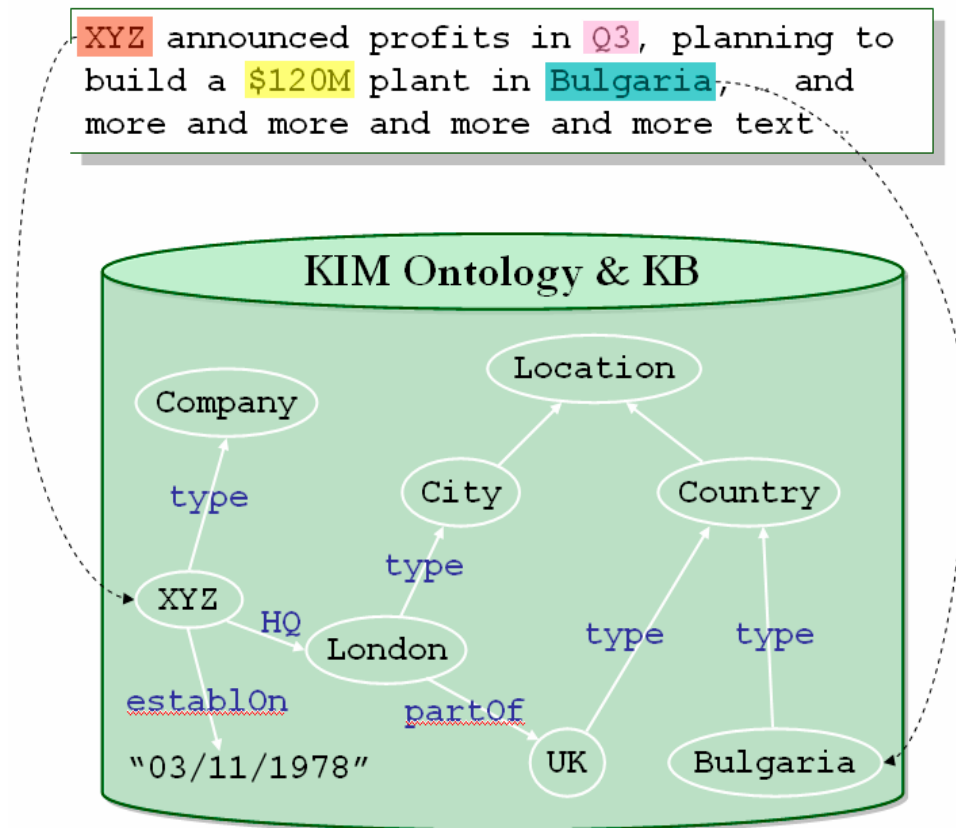


Figure 4-5: Annotation des ressources via l'outil KIM (extraite de [OntoText]).

IV.2 Modèle de représentation des annotations sémantiques

Le principe d'annotation revient à affecter des métadonnées aux différentes ressources utilisées, c'est-à-dire des données sur les données, afin de les définir, les décrire et les localiser [Guenther, 2001]. Cette structure permet ainsi une gestion plus intelligente et plus précise de l'information. Dans le cadre du Web sémantique, l'objectif des métadonnées est de rendre les différentes données interprétables par des agents logiciels et de favoriser ainsi la coopération personne-machine [W3C Technology, 2001].

Notons que dans tous les cas, l'annotation d'une ressource repose sur deux composantes de bases, à savoir :

- Le langage de représentation de ces métadonnées, dont le plus couramment utilisée est le RDF du fait de sa recommandation par le W3C.

- L'ontologie du domaine qui permet d'identifier chaque terme de ce langage. Cette dernière est décrite avec le langage OWL.

Selon une analyse menée par Fensel [Fensel, 2002] au sujet des ontologies, des langages de représentation des connaissances et des formats, il a été conclu que le langage RDF est le mieux adapté à l'annotation des ressources et le partage des connaissances pour la communauté du web sémantique. Celui-ci facilite considérablement la réutilisation de l'ontologie.

Le RDF, comme n'importe quel autre langage de représentation des connaissances, repose sur une ontologie Web qui définit les entités (Classes), les attributs (propriétés) et les relations entre ces classes. Cette ontologie permet une gestion plus efficace et un partage plus vaste et une meilleure réutilisation des connaissances [Yang, 2006]. Le langage OWL est le plus utilisé à cet effet. Ce dernier a été défini comme standard par le W3C en 2004 [W3C Recommandation, 2004]. Actuellement de plus en plus d'ontologies sont construites avec OWL.

IV.3 Automatisation des méthodes d'annotations

Nous pouvons citer plusieurs processus d'annotation des ressources web, selon le degré de leur automatisation. Les 3 approches principales sont les suivantes : Manuelle, semi-automatique et automatique.

IV.3.1 L'approche manuelle

C'est la toute première approche apparue dans les années 90, et qui servait à annoter une page web. Le langage le plus connu fut le HTML-A [Decker, 1999] qui était comme une extension du HTML. Par ailleurs, la définition des mots clés dans la balise <Meta> du langage HTML, est en soi une forme d'annotation manuelle d'une page web. Cette technique, dite d'indexation, permet au moteur de recherche de comparer les mots clés saisis par l'internaute à ceux définis dans cette balise pour savoir si la page en question répond à sa requête ou pas. Ce type

d'annotation est en général de nature informelle et repose sur le langage naturel, ce qui constitue une faiblesse au niveau de la recherche et de l'assignation.

IV.3.2 L'approche automatique et semi-automatique

Contrairement à la première méthode, celles-ci sont plutôt d'ordre formel. En général elles reposent sur un langage guidé par une ontologie ou une base de connaissances. C'est l'approche qui a été adoptée dans le cadre de notre projet **SWOC**, et ce, en se basant sur le langage RDF. Pour créer un nouvel utilisateur dans le système, le chairman crée un compte en remplissant un formulaire d'une façon très classique. Le système génère automatiquement un profil FOAF (Friend Of A Friend), qu'il enrichit en se basant sur ses connaissances. Il en est de même pour les articles soumis, qui seront décrits avec le langage DC. Nous aborderons ces deux langages dans la section suivante. Il existe de nombreux outils qui annotent automatiquement des textes ou une partie d'un texte, parmi lesquels on peut citer **KIM** (*Knowledge and Information Management*). Celui-ci a été élaboré par le laboratoire OntoText, et permet de reconnaître certains termes dans un texte et de les enrichir avec ses propres connaissances. Par exemple, quand KIM retrouve le terme « Québec », il lui affecte des métadonnées sur sa localisation géographique, sa population, sa densité et autre.

OntoText définit son outil KIM comme étant une plateforme qui fournit une infrastructure de services pour automatiser l'annotation sémantique, l'indexation et la récupération d'un contenu structuré ou semi-structuré. La figure 4.6 récapitule schématiquement l'outil KIM.

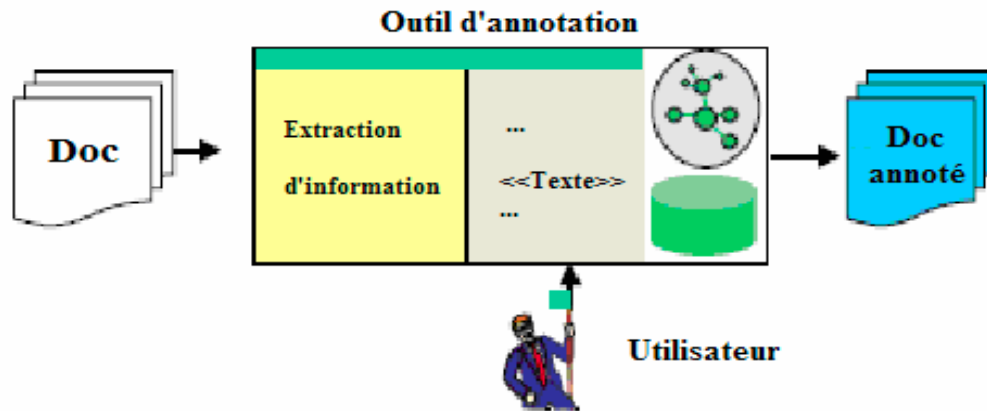


Figure 4-6: Processus d'annotation avec l'outil KIM.

KIM peut s'avérer très pratique pour enrichir l'annotation des articles reçus pour la conférence, et reconnaître facilement les thèmes qu'ils évoquent, voire même en créer un résumé.

C'est l'approche d'annotation semi-automatique qui sera adoptée dans le cadre de notre application de gestion des conférences, nous avons deux types de ressources à annoter. En effet, nous avons, d'une part, les ressources humaines, à savoir les évaluateurs, auteurs, membres du comité ou autres. C'est le langage FOAF (Friend Of A Friend) qui sera utilisé à cet effet [FOAF]. Comme nous l'avons expliqué dans le deuxième chapitre, ce vocabulaire repose sur le langage RDF et permet de décrire le profil d'une personne (ses préférences, son expertise, ses expériences, etc.). La figure 4.7 représente le processus d'invitation d'évaluateurs et la création de leurs profils FOAF.

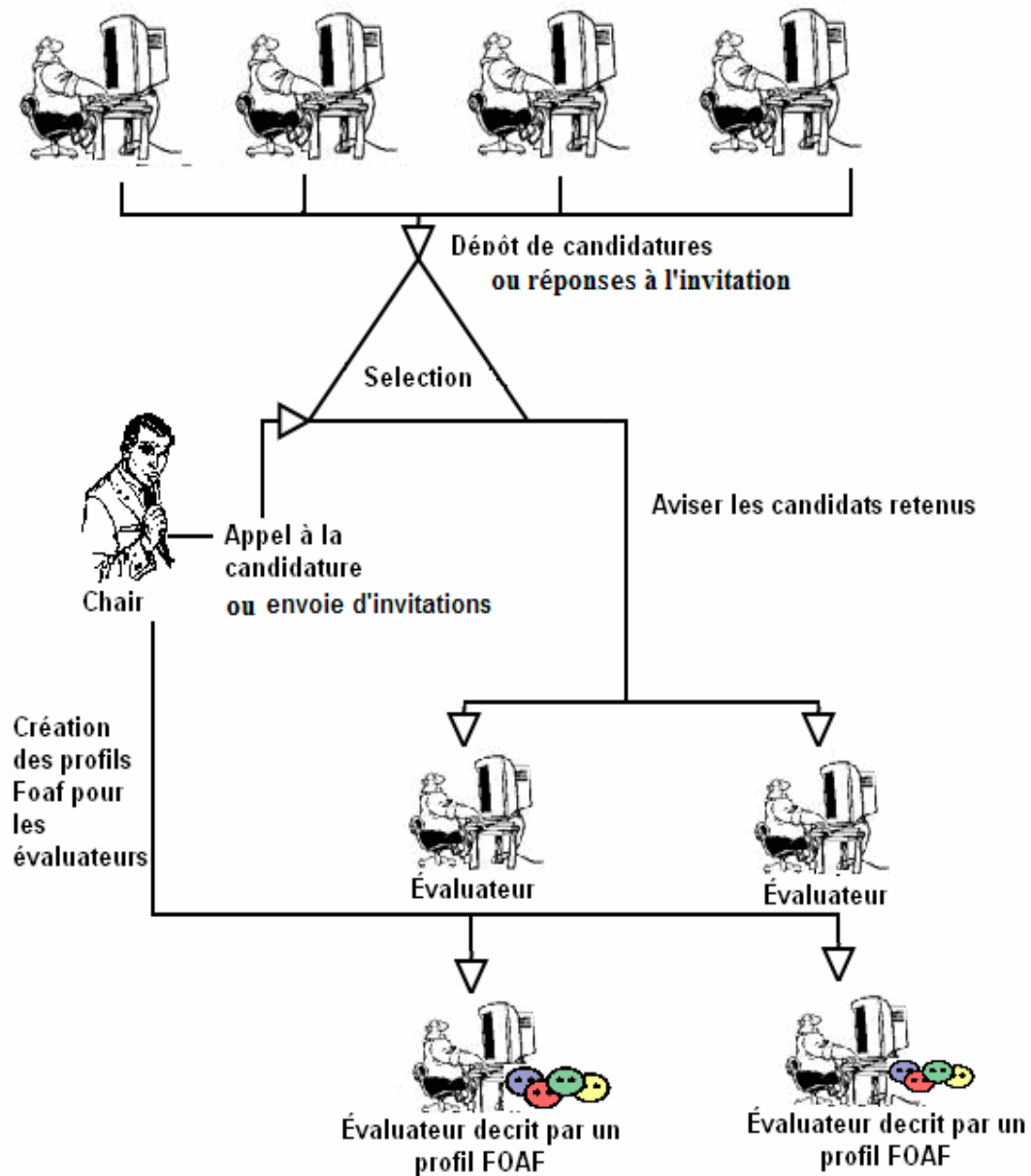


Figure 4-7: Invitation des évaluateurs et création de leurs profils FOAF.

D'autre part, nous avons à annoter aussi les ouvrages soumis par les auteurs dans le cadre d'une conférence. Il peut s'agir d'articles, de posters, ou autre. Ces derniers seront annotées en utilisant le vocabulaire DC (*Dublin Core*) [Dublin Core], qui repose aussi sur le langage RDF. La figure 4.8 explique le processus de soumission et d'annotation semi-automatique avec l'outil SWOC.

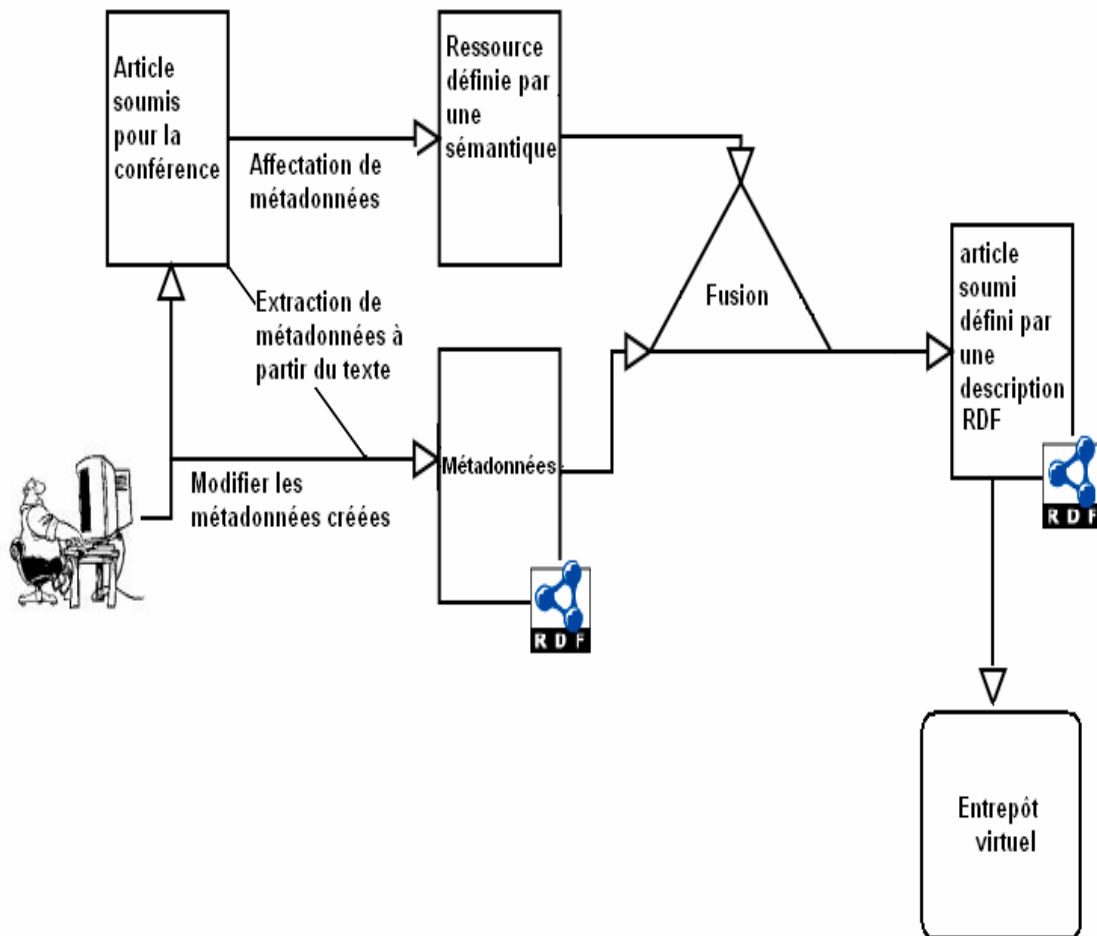


Figure 4-8: Soumission et annotation semi-automatique d'un article avec le système SWOC.

Dans tous les cas, c'est la méthode d'annotation semi-automatique qui sera adoptée. L'utilisateur ne fait que remplir les formulaires relatifs à son inscription, et le système génère automatiquement un profil FOAF. Le système peut enrichir cette description en se basant sur ses propres connaissances modélisées sous forme d'ontologie du domaine. Cette dernière fera l'objet de la section suivante.

IV.4 Ingénierie de l'ontologie

Le terme ontologie est emprunté de la philosophie, où il signifie « La science de l'être », et ce, depuis environ le XIXe siècle. Avec l'évolution de l'intelligence artificielle et l'émergence de l'ingénierie des connaissances, ce terme a pris une toute autre dimension [BENAYACHE, 2005]. Plusieurs définitions ont donc été proposées par différents auteurs, notamment, [Neeches et al., 1991], [Gruber, 1993] et [Borst, 1993]. En se basant sur leurs ouvrages, telle est la meilleure définition que nous pouvons tirer : « Une ontologie est une spécification explicite, et souvent formelle, d'une conceptualisation partagée ».

Il est à noter que le succès du web sémantique tient par-dessus tout à la prolifération des ontologies, qui requièrent une rapidité et une simplicité dans l'ingénierie d'une part, et de s'en tenir uniquement aux connaissances pertinentes d'autre part [Macdche, 2001]. L'ingénierie ontologique est l'ensemble des travaux qui s'articulent autour des différentes phases du cycle de vie d'une ontologie, à savoir analyse, conception, construction, validation, Etc. Suivre ces différentes étapes, simplifie considérablement l'élaboration de l'ontologie au développeur. La figure 4.9 récapitule graphiquement une des méthodes de l'ingénierie des ontologies, dite méta-modèle [Yang, 2006]. Cette forme cyclique de développement de l'ontologie a été adoptée pour la construction de l'ontologie du domaine du système SWOC.

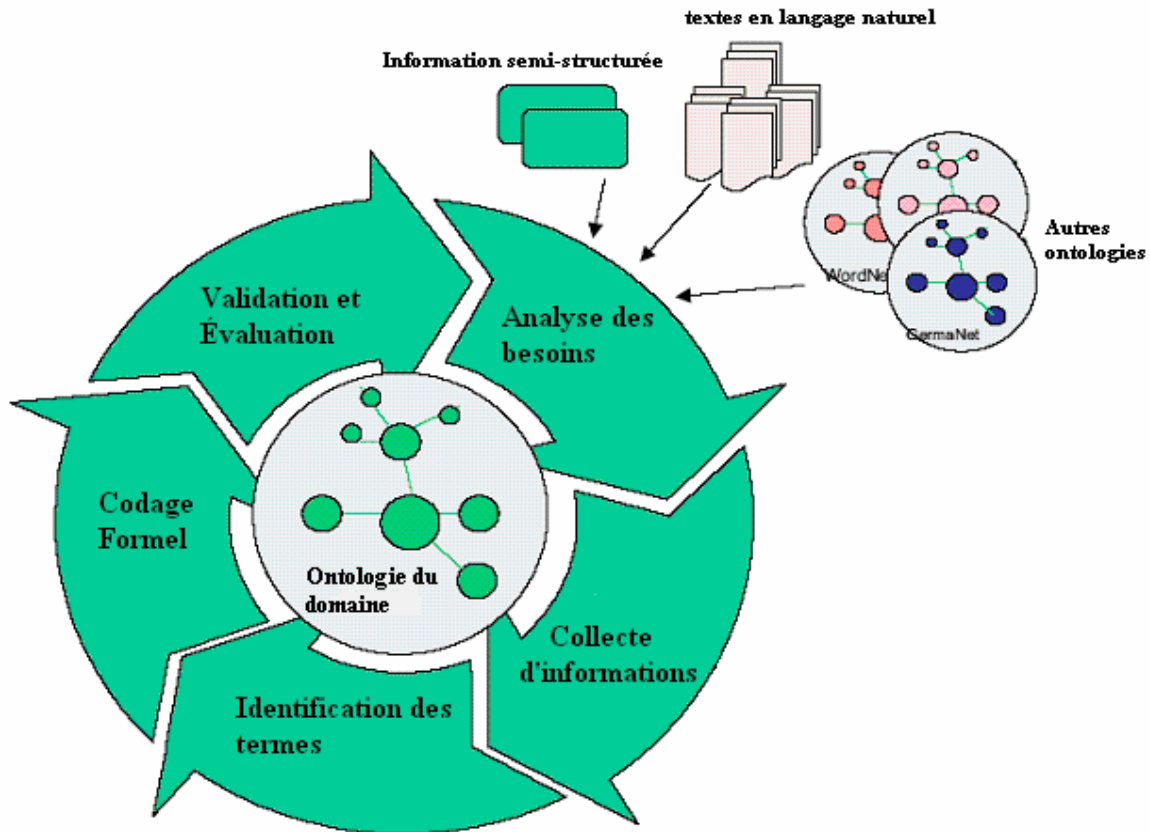


Figure 4-9: représentation de l'ingénierie cyclique de l'ontologie du domaine du système SWOC (Source: [Yang, 2006]).

Le méta-modèle est une méthode d'ingénierie d'ontologie qui a été conçue et adoptée en se référant aux différentes approches proposées par [Uschold, 1996], [Fernandez, 1997] et [Kietz, 2000], et en considérant les méthodes classiques du génie logiciel et de l'ingénierie des connaissances. On peut clairement déduire, à travers cela, qu'il n'existe pas une méthode standardisée pour le développement des ontologies, mais bien une variété de façons de faire, selon le domaine d'application. Dans tous les cas, certains critères présentés par [Gruber, 1995] doivent être pris en considération, notamment, la clarté, l'extensibilité et la cohérence.

Au cours de ce chapitre, nous présenterons les différentes phases de l'ingénierie de l'ontologie dans le cadre du développement du système SWOC. Celles-ci ont été préalablement citées, et sont très semblables à celle du cycle de développement dans le domaine du génie logiciel.

IV.4.1 Phase d'analyse des besoins

L'activité de cette phase se résume en l'identification des objectifs, des limites et des contraintes, ainsi qu'à l'étude du domaine d'application autour duquel s'articule le système en cours d'élaboration. Cette étape est pour l'ingénierie de l'ontologie ce qu'est l'analyse des besoins pour le génie logiciel et se doit de donner une réponse aux différentes questions qui tournent autour du domaine d'application, de l'objet, de l'objectif, de la durée de vie et de la langue utilisée pour cette ontologie.

Par ailleurs, un rapport récapitulatif, dit rapport d'analyse, sera élaboré à la fin de cette phase. Celui-ci servira d'appui pour la conception de l'ontologie au cours de la phase suivante.

IV.4.2 Phase de conception

Cette phase fait suite à celle de l'analyse des besoins et consiste à concevoir l'ontologie qui servira comme base de raisonnement du système en cours d'élaboration. Cette tâche n'est pas du seul rôle du développeur. En effet, il est extrêmement recommandé de faire intervenir des experts du domaine pour garantir la validité et la complétude de l'ontologie en question. La phase de conception se déroule en général sur deux étapes à savoir, la collecte d'informations et l'identification des termes.

A Collecte d'information

Il s'agit de collecter le maximum d'information sur le domaine d'application du système en cours d'élaboration. Ces données sont alors appelées *connaissances du domaine*. Cette tâche est souvent guidée par un expert du domaine, des documentations ou d'autres ontologies comme ce fut le cas pour l'application SWOC. En effet, la conception de l'ontologie du domaine du système est largement inspirée de l'ébauche proposée par l'ESWC2006¹. L'organisation de cette conférence est une initiative vers gestion basée sur la technologie du Web sémantique. Elle fut la première à élaborer une ontologie du domaine et à affecter des descriptions RDF (FOAF, DC, Etc.) à ses ressources, même si le système ne s'en

¹<http://www.eswc2006.org/>

est pas réellement servi pour l'affectation des articles ou quelconque autre tâche en rapport avec la gestion. En nous inspirant de cette description formelle du domaine d'application et en l'enrichissant de nouvelles connaissances, nous avons mis au cœur de ce projet, les notions de partage et de réutilisation des données. La plus value que nous avons apportée à cette ontologie consiste surtout à définir en détail un des domaines autour duquel pourrait s'organiser une conférence. Nous avons choisi à cet effet le domaine du web sémantique, du fait que nous ayons longtemps travaillé dessus, et que par conséquent nous sommes aptes à définir sa portée et ses extensions.

Il faut noter, par ailleurs que l'on fait souvent appel aux techniques de génie logiciel pour assurer cette tâche, notamment, le Brain Storm, les interviews, les documentations formelles ou informelles et autres [Yang, 2006].

B Identification des termes

Le développeur, en collaboration avec un expert du domaine, entame cette tâche aussitôt que les informations sur le domaine d'application seront réunies. Cette tâche consiste plus exactement à retracer une abstraction du monde réel, sous forme de concepts et de relations entre ces concepts. Pour des fins de simplification et de clarté, il est vivement recommandé de représenter cette abstraction en langage naturel, avant de commencer à la coder de façon formelle. La figure 4.10 est extraite du document de conception de l'ontologie développée dans le cadre du projet SWOC.

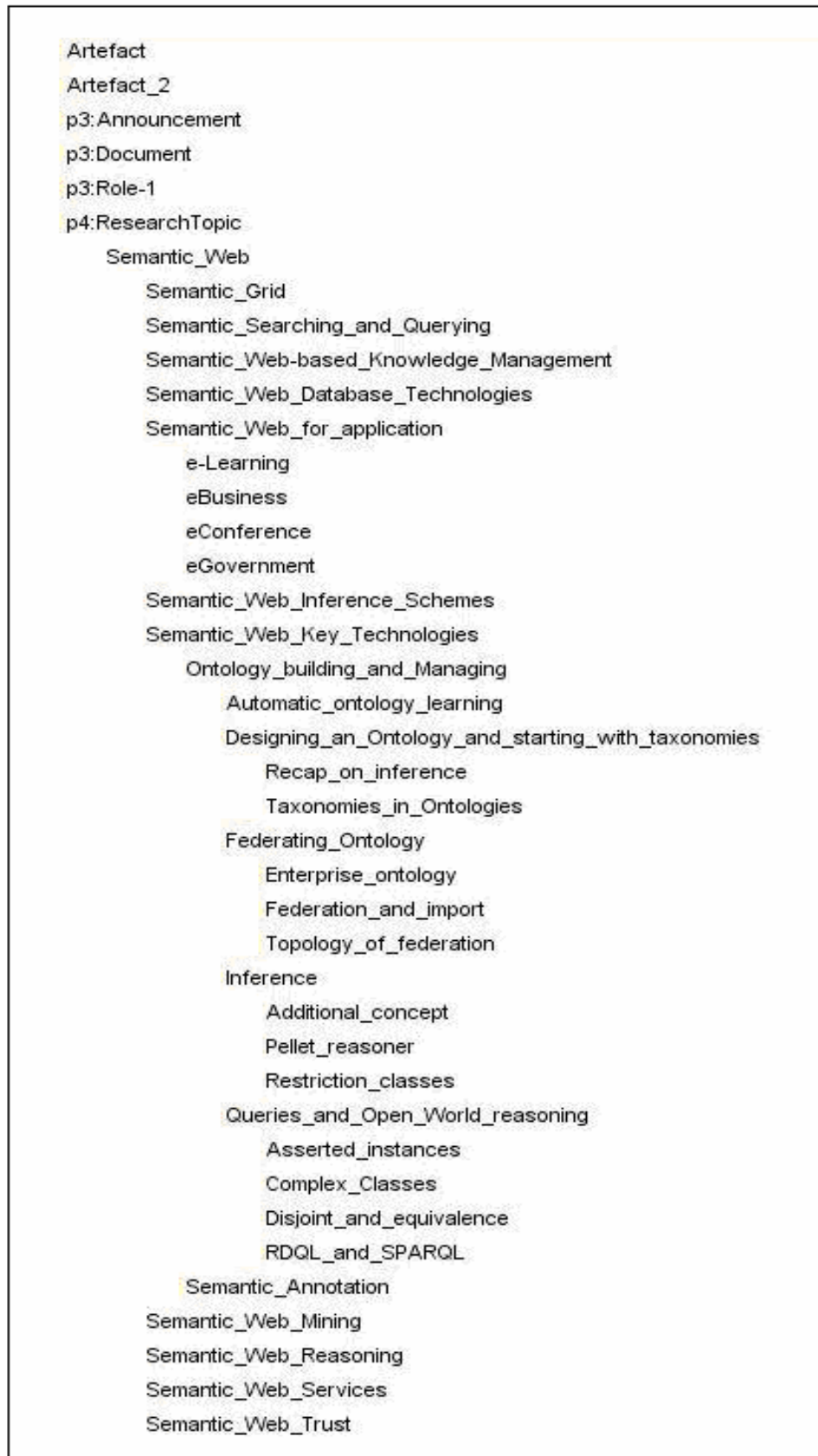


Figure 4-10: extrait de la conception de l'ontologie du système SWOC.

Il est très important de noter qu'au terme de cette phase de l'ingénierie de l'ontologie, un artefact doit être livré. Il s'agit du document de conception de l'ontologie, qui servira d'appui au développeur pour le codage de l'ontologie dans un langage formel.

IV.4.3 Phase de développement

Dans la phase précédente (Conception) les connaissances du monde réel étant représentées en concepts et relations entre ces concepts, il ne reste plus à présent qu'à les coder d'une façon formelle. Le langage utilisé à cet effet est l'OWL, du fait qu'il est recommandé par le W3C, comme nous l'avons précisé dans le second chapitre. Une fois le développement réalisé. Le code source résultant de cette phase doit être bien entendu enregistré dans un document qui portera l'extension « .OWL ». Celui-ci servira de référence pour le partage et la réutilisation des connaissances sur le monde réel, d'une part, et comme base de raisonnement pour les moteurs d'inférence du système SWOC d'autre part.

Il existe une multitude d'outils pour le développement d'ontologies qui diffèrent selon leurs faiblesses et leurs points forts. Ces derniers sont appelés dans le jargon informatique EDI (*Environnements de développement intégrés*). L'encyclopédie Wikipédia les définit comme étant « *des programmes regroupant un éditeur de texte, un compilateur, des outils automatiques de fabrications et souvent un débogueur* ». En d'autres termes, il s'agit d'un environnement permettant d'assister le développeur dans sa tâche de programmation ou de génération du code source.

Parmi les EDI les plus utilisés pour l'élaboration des ontologies, nous citons : Chimaera [CHIMAERA], DAG-Edit [DAG-EDIT], OilEd [OILED], Protégé [PROTEGE]. C'est d'ailleurs ce dernier qui a été adopté pour la construction de l'ontologie du système SWOC. Ce choix est basé sur une étude comparative réalisée par l'université de Linköping en suède [LAMBRIX, 2003], selon la méthode REAL [Löwgren, 1993].

Selon le site officiel de l'outil [PROTÉGÉ], Protégé est défini comme étant un éditeur d'ontologie libre et gratuit. Il est développé sous Java et se distingue par son caractère extensible et flexible. De nombreux modules peuvent être téléchargés et intégrés d'une façon

très simple, afin d'enrichir ses fonctionnalités. Les ontologies élaborées à travers cette plateforme peuvent être enregistrées sous une variété de formats, notamment RDF(S), OWL, XML-S.

La figure 4.11 représente une capture d'écran de la plateforme « Protégé », pendant que le développeur définit les concepts.

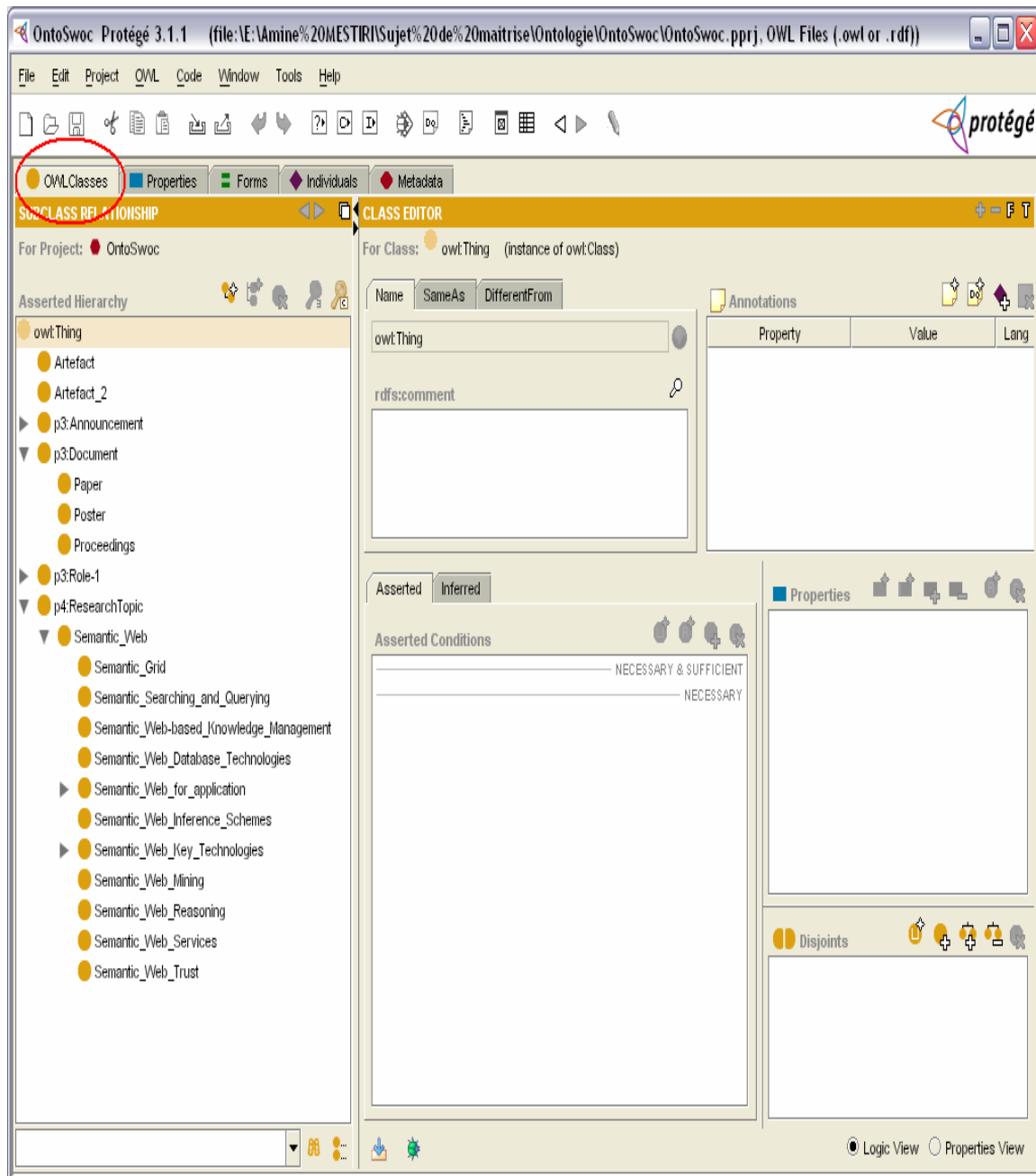


Figure 4-11: capture d'écran de la plateforme Protégé - Créations des concepts

La figure 4.12 représente une capture d'écran de la plateforme « Protégé », pendant que le développeur identifie les relations entre les concepts qu'il a définis.

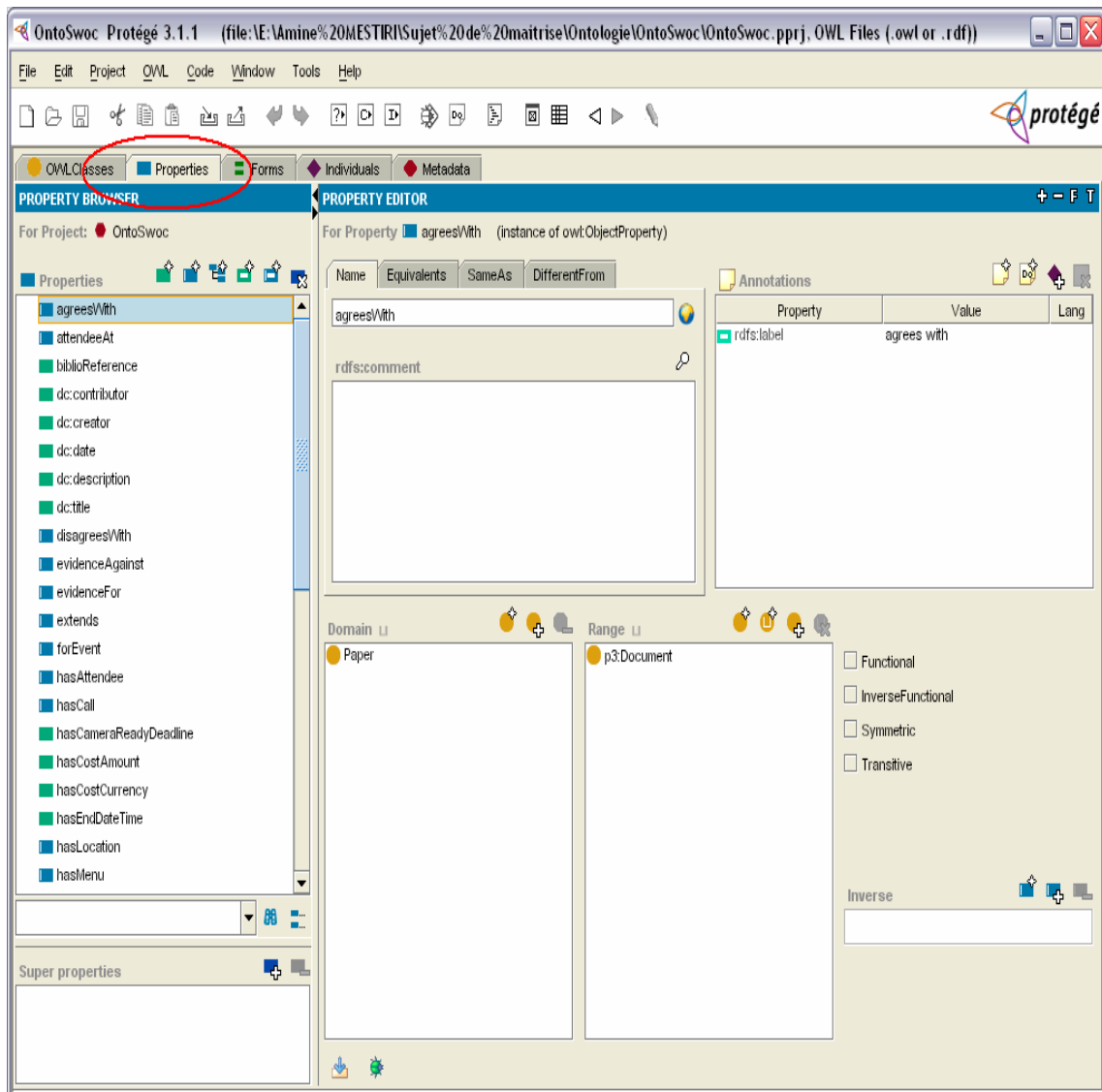


Figure 4-12: Capture d'écran de la plateforme Protégé - Créations des relations entre ces concepts

La figure 4.13 est un extrait du code OWL généré automatiquement à la suite de ce qui a été défini via la plateforme Protégé. On peut clairement voir sur cette image, la modélisation du monde réel en concepts (*OWL:Class*) et relations entre ces concepts (*OWL:ObjectProperty*).

Il est très important de rappeler que ce code doit être sauvegardé dans un fichier portant l'extension « OWL », afin que le système puisse le reconnaître et comprendre qu'il s'agit de la modélisation du domaine d'application.

```

<owl:Class rdf:ID="organisingCommitteeMember">
  <rdfs:subClassof rdf:resource="#Role"/>
  <rdfs:label>Organising Committee Member</rdfs:label>
</owl:Class>

<owl:Class rdf:ID="ProgrammeCommitteeMember">
  <rdfs:subClassof rdf:resource="#Role"/>
  <rdfs:label>Programme Committee Member</rdfs:label>
</owl:Class>

<owl:Class rdf:ID="SessionChair">
  <rdfs:subClassof rdf:resource="#Role"/>
  <rdfs:label>Session Chair</rdfs:label>
</owl:Class>

<!-- Artefacts, Events, Places expected to use hasPart/isPartof relations -->
<owl:ObjectProperty rdf:ID="hasPart">
  <rdfs:domain rdf:resource="http://www.w3.org/2002/07/owl#Thing"/>
  <rdfs:range rdf:resource="http://www.w3.org/2002/07/owl#Thing"/>
  <rdfs:label>has part</rdfs:label>
</owl:ObjectProperty>

<owl:ObjectProperty rdf:ID="isPartof">
  <rdfs:domain rdf:resource="http://www.w3.org/2002/07/owl#Thing"/>
  <rdfs:range rdf:resource="http://www.w3.org/2002/07/owl#Thing"/>
  <rdfs:label>is part of</rdfs:label>
</owl:ObjectProperty>

<!-- Artefacts, AcademicEvents, Calls expected to use hasTopic/isTopicof relations -->
<owl:ObjectProperty rdf:ID="isTopicof">
  <rdfs:domain rdf:resource="http://swrc.ontoware.org/ontology#ResearchTopic"/>
  <rdfs:label>is topic of</rdfs:label>
  <!-- range left blank -->
  <!-- inverse of #hasTopic -->
</owl:ObjectProperty>

<owl:ObjectProperty rdf:ID="hasTopic">
  <!-- domain left blank -->
  <rdfs:range rdf:resource="http://swrc.ontoware.org/ontology#ResearchTopic"/>
  <rdfs:label>has topic</rdfs:label>
  <!-- inverse of #isTopicof -->
</owl:ObjectProperty>

```

Figure 4-13: Extrait du code OWL de l'ontologie du domaine du système SWOC

Jusqu'à présent, nous n'avons fait qu'alimenter le sous-module description par des informations sur les ressources, sous forme d'annotations [Yang 2006]. Au terme de ce qui a été réalisé jusqu'ici, on peut déjà parler d'une première ébauche d'une application en ligne de gestion de conférence, basée sur la technologie du web sémantique. Le système peut créer maintenant et stocker les articles soumis, leurs descriptions DC et les profils FOAF des différents acteurs du système sur le serveur. On appelle souvent cet espace entrepôt virtuel. Tout ce travail, facilite la réutilisation et le partage des connaissances et des informations liées aux ressources, mais ne permet pas au système pour autant d'effectuer des inférences et de gérer plus intelligemment le contenu. Il manque en effet une composante capitale dans cette architecture qui jouera le moteur d'inférence. Il s'agit bien entendu du sous-module processeur. Ce dernier est guidé par l'ontologie du domaine que nous avons construite, et qui modélise d'une façon abstraite toutes les connaissances du système.

IV.4.4 Phase de validation et évaluation de l'ontologie

La validation de l'ontologie, de façon générale, doit se faire au regard des tâches qui lui sont attribuées, dans la mesure où cette étape de l'ingénierie ontologique ne peut se concevoir que par l'usage [Bourigault, 2001]. Cependant, et comme il est absurde de demander aux utilisateurs de valider un à un tous les termes définis dans cette ontologie, nous avons jugé plus judicieux d'effectuer une série de tests. Il s'agit en effet de vérifier pour chaque test, si les termes retournés comme résultats répondent à la requête de l'utilisateur ou non.

Par ailleurs, [Yang, 2006] propose une autre technique d'évaluation de l'ontologie du domaine. Celle-ci est plus conforme à la méthode de méta-modèle utilisée pour l'élaboration de l'ontologie et se distingue des méthodes classiques par des critères spécifiques d'évaluation. Il est à rappeler que cet auteur a lui-même adopté la même méthodologie dans l'élaboration de l'ontologie du domaine pour son système de recommandation de cartes de crédit, basé sur la technologie du web sémantique. Parmi ces critères, nous citons :

- ***Déterminer les limites du domaine d'application*** : il est très difficile de cerner les limites du domaine et d'affirmer le niveau d'abstraction que doit couvrir cette ontologie. L'essentiel est que cette ontologie puisse être utilisée par tous les utilisateurs et qu'elle définisse suffisamment de concepts et de relations entre ces concepts afin de limiter la dépendance vis à vis de l'expert du domaine.

- ***Souligner l'importance de l'analyse des besoins*** :

Il est très important de vérifier si l'ontologie satisfait tous les besoins cités dans le rapport d'analyse. Si tel est le cas, l'ontologie répond alors aux attentes des usagers, sinon il faudrait revoir la conception.

- ***Être attentif à l'évolution de l'ontologie*** :

Que ce soit en génie logiciel, ou en ingénierie des connaissances, l'évolution du système est un facteur d'une importance capitale à étudier de près. Il en est autant pour l'ontologie dans le cadre de l'ingénierie ontologique. Du fait de sa forme cyclique, la méthodologie que nous avons adoptée pour l'élaboration de l'ontologie ne laisse d'autre choix à l'ontologie que d'évoluer graduellement du plus approximatif au plus adapté.

- ***Assurer la mise à jour de la documentation***:

La documentation doit évoluer dans le même sens et à la même vitesse que l'élaboration de l'ontologie. Il s'agit de donner des comptes-rendus tout au long des différentes phases. Le rapport d'analyse des besoins, le rapport de conception et le rapport de développement font tous partie de cette documentation.

- ***Assurer le partage et la réutilisation des connaissances***

Il est à rappeler que le principal objectif de l'ontologie est de décrire un domaine de façon abstraite et formelle en concepts et relations entre ces concepts, afin de permettre un meilleur partage et une plus large réutilisation des connaissances. Il s'agit de deux idées clés dans la philosophie de la mise en place de l'ontologie. Au cours de la validation, le développeur doit donc s'assurer que son ontologie couvre tout le domaine, dans la mesure du possible, afin de ne plus dépendre d'un expert du domaine.

IV.5 Moteur d'inférence

Tout le travail que nous avons réalisé jusqu'à maintenant, consiste à annoter les ressources utilisées dans le cadre de notre application, d'une part, et à représenter les connaissances sous forme d'une ontologie du domaine, d'autre part. Cela est loin de pouvoir permettre au système de réaliser les raisonnements logiques nécessaires. L'intégration d'un moteur d'inférence s'impose donc, conformément à l'architecture proposée par la figure 4.4 au début du chapitre. Il existe une large variété de moteurs sur le Web tels que Jena [Jena], Racer [Racer], Rap [RAP]. Ceux-ci se différencient en général par leurs forces et leurs faiblesses, mais surtout par le langage sous lequel ils ont été développés.

IV.5.1 L'outil RAP

Dans le cadre du projet de développement de l'application **SWOC**, C'est le moteur d'inférence **RAP** (*RDF API For PHP*) qui a été adopté et qui sera intégré en tant que module à part entière dans le système. Avant de définir ce dernier, il est très important de rappeler le contexte dans lequel il a été choisi. En effet, à partir du moment où nous avons pris la décision de migrer d'une application existante sur le web classique vers la nouvelle technologie du web sémantique, au lieu de développer un nouveau système, nous nous sommes posés comme restriction de travailler avec le langage PHP, étant donné que toutes les applications les plus connues sont toutes développées avec ce dernier. Or, seul le moteur RAP est adapté à ce langage de programmation Web. RAP est par ailleurs défini comme étant un package de classes, permettant de compiler, chercher et manipuler des modèles RDF [RAP]. Il s'agit d'un outil conçu pour le web sémantique, à l'université de Freie, à Berlin, en 2002. Depuis ce jour, il a bénéficié de nombreuses contributions internes et externes au niveau de la programmation. RAP est maintenant un outil open source. Il peut être utilisé par quiconque sous les termes de la licence GNU Lesser General Public License (LGPL).

Au niveau conceptuel, RAP se présente sous forme d'un ensemble de classes, organisé autour de deux APIs de base, à savoir **Model API** et **ResModel API** [RAP]. La figure 4.14 illustre l'organisation de ses différents modules.

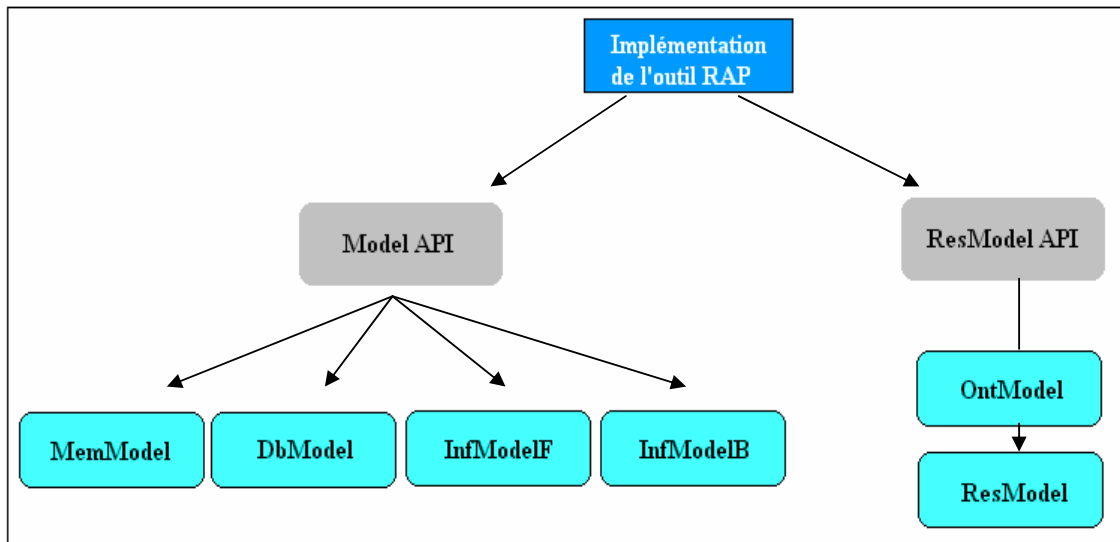


Figure 4-14: Organisation en packages de l'outil RAP

Si le ResModel API est plus adapté à la manipulation des graphes représentant des ressources, il en est autrement pour le MemModel qui permet de manipuler des graphes contenant un ensemble de « statements ». Un statement est une déclaration, appelée aussi triplet, RDF, conformément à ce qui a été présenté dans le second chapitre du présent document. Il se compose en général de trois éléments, à savoir :

- Le sujet qu'est la ressource à décrire.
- Le prédicat, qu'est la propriété qui étiquette l'arc (ou la déclaration).
- L'objet, qui peut être un littéral ou une ressource pointée par l'arc.

Un fichier RDF/XML représente en général l'annotation d'une ou plusieurs ressources web. Afin de pouvoir récupérer les descriptions contenues dans un tel fichier, le moteur d'inférence doit reproduire son contenu dans un modèle qu'il aura construit. Ce dernier est appelé arbre de déclarations. Le principe se résume donc en trois étapes :

- Créer un modèle vide :
- Récupérer le code RDF
- Structurer le modèle selon l'organisation du RDF

Il est très important de signaler, que dans le cadre de notre projet SWOC, c'est le « Model API » que nous utiliserons aussi bien pour récupérer de l'information d'un fichier

RDF que pour annoter une ressource. C'est plus particulièrement le sous-module **MemModel** qui sera adopté. Celui-ci stocke toutes les déclarations au niveau de la mémoire, contrairement au **DbModel** qui effectue un enregistrement sur un support de stockage. La structure de données utilisée à cet effet est un tableau dynamique. Chaque fois qu'un triplet vient s'ajouter au modèle, l'index doit être mis à jour pour accélérer le temps de réponse aux requêtes [Rap], et ce, conformément à la figure 4.15.

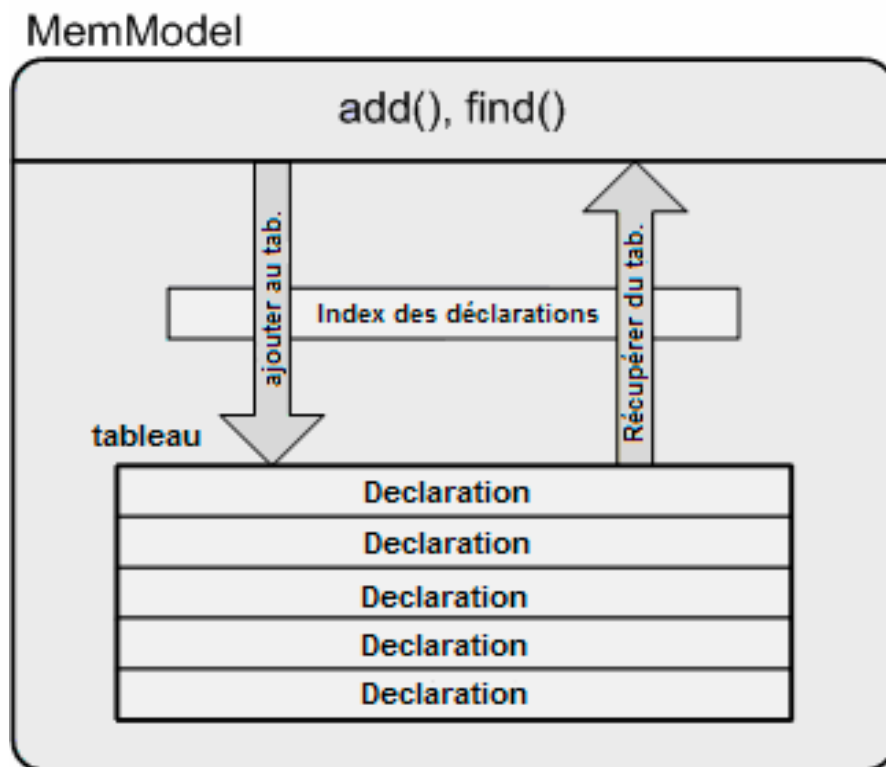


Figure 4-15: Processus de fonctionnement du module MemModel.

Comparé à tous les modèles des différents APIs, Le MemModel est le plus efficace en termes de temps de réponse.

IV.5.2 Intégration de RAP dans l'application SWOC

Après avoir fixé le choix du moteur d'inférence qui sera déployé pour effectuer les raisonnements logiques nécessaires, ce dernier doit être intégré en tant que module dans le système, conformément à l'architecture prévue. Il pourra ainsi être apte à interpréter les annotations des ressources utilisées par l'application SWOC, et raisonner en se basant sur l'ontologie que nous venons de construire dans la section précédente. La figure 4.16 est un extrait du code source de l'application et montre comment le développeur doit faire appel au package RAP pour pouvoir manipuler ses classes et les différentes fonctionnalités qu'elles proposent.

```
require_once "../include.php";
require_once "../$OC_statusCfgFile";
//creating and saving the RDF annotation file
define("RDFAPI_INCLUDE_DIR", "C:/Program Files/wamp/www/openconf/rdfapi-php/api/");
include(RDFAPI_INCLUDE_DIR . "RdfAPI.php");
include(RDFAPI_INCLUDE_DIR."vocabulary/foaf.php");
include(RDFAPI_INCLUDE_DIR."vocabulary/foaf_C.php");
//include(RDFAPI_INCLUDE_DIR."sparql\SparqlParser.php");
```

Figure 4-16: Extrait du code source du système SWOC - Inclusion de l'outil RAP.

La figure 4.17, quant à elle, représente un bout de code relatif à l'annotation d'une ressource humaine, via l'application SWOC. On peut clairement voir les principales étapes citées dans la sous-section à savoir : la création des triplets, la création d'un modèle, et l'insertion de ces déclarations dans ce dernier.

```

/creating and saving the FOAF profile

//Création d'une ressource
$newReviewer = new Resource("$rid");
//Création des statements
$statement0 = new Statement($newReviewer, FOAF::PERSON(), new Literal("$rid"));
$statement1 = new Statement($newReviewer, FOAF::FAMILY_NAME(), new Literal ($_POST["name_last"]));
$statement2 = new Statement($newReviewer, FOAF::FIRST_NAME(), new Literal ($_POST["name_first"]));
$statement3 = new Statement($newReviewer, FOAF::ORGANIZATION(), new Literal ($_POST["org"]));
$statement4 = new Statement($newReviewer, FOAF::BASED_NEAR(), new Literal ($_POST["country"]));
$statement5 = new Statement($newReviewer, FOAF::MBOX(), new Literal ($_POST["email"]));
$statement6 = new Statement($newReviewer, FOAF::PHONE(), new Resource($_POST["phone"]));
$statement7 = new Statement($newReviewer, FOAF::HOMEPAGE(), new Resource($_POST["url"]));

//Creation du model et ajout des statements à celui-ci
$model = ModelFactory::getDefaultModel();
$model->add($statement0);
$model->add($statement1);
$model->add($statement2);
$model->add($statement3);
$model->add($statement4);
$model->add($statement5);
$model->add($statement6);
$model->add($statement7);

```

Figure 4-17: Extrait du code source du système SWOC - création d'un modèle de graphes RDF.

Un moteur d'inférence n'est pas qu'un simple outil de création et d'extraction de triplets RDF. C'est avant tout une interface permettant d'interpréter les ressources en se basant sur leurs descriptions sémantiques, d'une part, et l'ontologie du domaine, représentative des connaissances du domaine, d'autre part. D'ailleurs techniquement, il se présente comme étant une API (*Application Programming interface*) qui comprend un ensemble de classes permettant d'assurer cette tâche.

Il est cependant fort à constater que malgré les points forts de cet outil, celui-ci ne prend en charge le langage OWL que de façon partielle, ce qui constitue en soi une grande faiblesse et un sérieux obstacle pour développer le module d'assignation des articles aux différents évaluateurs. Il suffit juste de se référer aux différents forums et ouvrages sur l'outil RAP pour constater qu'il s'agit véritablement d'une question de faisabilité, due à la limite de l'outil RAP plus qu'autre chose. Dans ce sens, nous étions bloqués à plus d'une reprise dans le développement de l'application. Cela nous a souvent amené à apporter certains changements au niveau de son code source, dans la mesure du possible, afin de corriger certaines défaillances.

L'utilisation du moteur d'inférence RAP, malgré ses faiblesses, revient au fait qu'il soit le seul outil adapté au langage de programmation PHP. Or, à partir du moment où nous avons décidé de partir d'une application existante pour la faire migrer vers une plateforme Web sémantique, nous n'avons d'autre choix que d'utiliser ce langage, puisque toutes les applications existantes jusqu'à ce jour sont développées sous PHP. Faire migrer une application vers une nouvelle plateforme et changer en plus de langage de programmation, revient en fin de compte à développer un nouveau système, ce qui ne constitue pas l'objectif de ce projet. Nous nous sommes donc fixé comme restriction de travailler sous ce langage serveur.

Par ailleurs, et par souci de réutilisation, nous avons décidé de faire en sorte que le module d'assignation automatique des articles soit construit de façon indépendante du reste du système, afin de faciliter son intégration dans n'importe quelle autre application existante, y compris le projet SWOC.

Il est très important de prendre en considération cette remarque, pour se rappeler que pendant la phase de test, c'est le module, et non tout le système qui sera mis à l'essai, pour en dégager les résultats d'assignation et les comparer à ceux de l'application Web actuelle OpenConf [OpenConf].

La figure 4.18 représente un bout de code extrait du module d'assignation des articles aux évaluateurs. Ce dernier montre la récupération du thème d'un article reçu pour une conférence. On peut clairement constater que l'outil cherche sa représentation hiérarchique

dans l'ontologie afin d'interpréter son sens. Le code relatif à l'assignation de cet article à différents évaluateurs de la conférence, conformément à leurs expertises, est représenté par la figure 4.19.

```
//définir le cadre de l'article:
$subjectOfPaper = $_POST["Topic"];
$topicOfPaper="";
$mainTopicOfPaper="";

//create an InfmodelB
$infModel= ModelFactory::getInfModelB('http://InfModelF.org');

//Charger l'ontologie dans le modèle "$infModel" pour pouvoir commencer les raisonnements.
$infModel->load("C:\Program Files\EasyPHP1-8\www\Ontologie/PaperOntology.owl");

//chercher les topic (cad les classe mère de la class subject recu), le résultat est retourné sous forme de MemModel
$memModel=$infModel->find(new Resource("http://www.w3.org/2000/01/rdf-schema#.$subjectOfPaper"), $RDFS_subClassOf, null );
$it=$memModel->getStatementIterator();
//$it est de type "StatementIterator", c'est le type de retour de la fonction getStatementIterator
if ($it->hasNext())
{
    $statement=$it->next();
    $topicOfPaper= substr($statement->getLabelObject(),37);
    //chercher le mainTopic
    $memModel=$infModel->find(new Resource("http://www.w3.org/2000/01/rdf-schema#.$topicOfPaper"), $RDFS_subClassOf, null );
    $it=$memModel->getStatementIterator();
    //$it est de type "StatementIterator", c'est le type de retour de la fonction getStatementIterator
    if ($it->hasNext())
    {
        $statement=$it->next();
        $mainTopicOfPaper = substr($statement->getLabelObject(),37);
        echo "<font color = blue><B> le sujet de l'article est :\" $subjectOfPaper \",
        il s'article donc autour du thème \"$topicOfPaper \" dans le domaine du \"$mainTopicOfPaper \"<br><br></B></font>";
    }
}
}
```

Figure 4-18: interprétation du thème de l'article.


```

//définir le cadre de l'article:
$subjectOfPaper = $_POST["Topic"];
$topicOfPaper="";
$mainTopicOfPaper="";

//create an InfmodelB
$infModel= ModelFactory::getInfModelB('http://InfModelF.org');

//Charger l'ontologie dans le modèle "$infModel" pour pouvoir commencer les raisonnements.
$infModel->load("C:\Program Files\EasyPHP1-8\www\Ontologie/PaperOntology.owl");

//chercher les topic (cad les classe mère de la class subject reçu), le résultat est retourné sous form
$memModel=$infModel->find(new Resource("http://www.w3.org/2000/01/rdf-schema#.$subjectOfPaper), $RDFS_
$it=$memModel->getStatementIterator();
//$it est de type "StatementIterator", c'est le type de retour de la fonction getStatementIterator
if ($it->hasNext())
{
    $statement=$it->next();
    $topicOfPaper= substr($statement->getLabelObject(),37);
    //chercher le mainTopic
    $memModel=$infModel->find(new Resource("http://www.w3.org/2000/01/rdf-schema#.$topicOfPaper), $RDFS_
    $it=$memModel->getStatementIterator();
    //$it est de type "StatementIterator", c'est le type de retour de la fonction getStatementIterator
    if ($it->hasNext())
    {
        $statement=$it->next();
        $mainTopicOfPaper = substr($statement->getLabelObject(),37);
    }
}

```

Figure 4-19: Assignment automatique de l'outil d'un article à quelques évaluateurs.

V Conclusion

Au terme de ce chapitre, nous aurons réalisé la ré-ingénierie de l'architecture de l'application en ligne de gestion de conférence OpenConf. Nous aurons également développé l'ontologie, ainsi que les différents modules appropriés. La réalisation de ce projet, que nous venons de décrire à travers ce chapitre, a été marquée par deux points importants : Le premier est négatif et concerne les nombreuses difficultés que nous avons rencontrées pour intégrer l'outil RAP, et pour que ce dernier puisse prendre en considération l'extension OWL, afin d'effectuer ses inférences nécessaires. Cela en dit long sur les critiques qu'il y a à faire sur le Web sémantique et que nous passerons en revue dans le dernier chapitre. Le second point à noter est positif et concerne la semi-automatisation des descriptions RDF des ressources et qui évite à l'utilisateur de saisir son profil ou la description de son ouvrage si celui-ci est déjà décrit quelque part sur le Net. La nouvelle version de l'application résultante est ainsi baptisée SWOC (*Semantic Web Open Conferences*) et est basée sur l'architecture en couches du Web sémantique, proposée par Tim Berners-Lee en 2001. Au regard de cette architecture, le système SWOC répond tout à fait aux enjeux du web sémantique.

Par ailleurs, nous nous trouvons encore dans l'impossibilité de juger la pertinence de travail que nous avons réalisé et de l'apport de la nouvelle technologie du web sémantique pour les applications en ligne de gestion de conférences. Plusieurs questions se posent alors. Le chapitre suivant tentera d'y répondre à travers une auto-évaluation du projet SWOC et une analyse de la contribution de notre recherche.

Chapitre 5: Évaluation du projet

I Introduction

Après avoir mené toute cette recherche sur la nouvelle génération du web, dite « Web sémantique » et surtout après avoir effectué une ré-ingénierie d'un système existant de gestion de conférences pour intégrer cette nouvelle génération du web, nous nous devons d'évaluer le système et valider le travail réalisé, d'une part, et de discuter de l'apport du web sémantique pour les applications en ligne.

II Évaluation du système SWOC

L'évaluation d'un système informatique fait partie intégrante dans son processus de développement et intervient à différents niveaux de ce dernier. Balbo [Balbo, 1994] la définit comme un processus permettant l'appréciation complète de l'interaction personne-machine au sein d'un système visant l'utilisateur et la manière dont il agit, ainsi que le système et la manière dont il fonctionne.

Étant donnée la taille de l'application et s'agissant surtout de la ré-ingénierie d'un système existant et non pas le développement d'une application sur mesure, seuls les aspects en relation avec nos objectifs seront soumis à cette évaluation. Les critères qui seront alors retenus en priorité sont: La conformité de l'assignation des articles aux évaluateurs, la pertinence de la recherche de ressources, et le temps de réponse aux requêtes. En génie logiciel, cette phase est appelée contrôle de qualité. Elle suit immédiatement le cycle de développement d'un système informatique et est d'une importance capitale, dans la mesure où elle permet de rendre compte des éventuelles anomalies (bugs, non-conformité par rapport au cahier de charge, etc.) et de valider le projet. Il existe différentes techniques d'évaluation d'un système informatique. Celles-ci sont classées en deux approches : empirique et analytique [Senach, 1990].

II.1 Approche empirique :

Dans cette section, il s'agit d'évaluer le système informatique après la phase d'implémentation, et ce, de façon partielle, à travers un prototypage, ou de façon totale via le diagnostic d'usage (Analyse statique et métrique du logiciel, évaluation itérative, etc.) ou les tests de conception (Observation, Questionnaire, Interviews, etc.) [Fadhel, 2005].

II.2 Approche analytique :

Contrairement à la première approche, celle-ci est définie par [Fadhel, 2005] comme une évaluation à priori de l'implémentation du système informatique en question. Les méthodes qui en émanent peuvent alors être classées en deux catégories selon qu'elles soient formelles ou informelles. La première est elle-même divisée en deux modèles, à savoir :

- ***Le modèle de qualité*** : Il est défini par [Paterno, 2000] comme un modèle qui permet d'identifier les propriétés formelles d'une application ayant un effet sur les performances des utilisateurs, et vise à établir une correspondance entre ces propriétés et les difficultés d'utilisation d'interfaces.
- ***Le modèle prédictif*** : Ce modèle permet de prédire les performances d'utilisation à partir de la conception réalisée. Il s'agit de prévoir les performances sans passer de par une phase de test.

Quant aux types de méthodes informelles les plus répondues, elles sont au nombre de deux, selon [Senach, 1990]. Il s'agit de l'expertise où nous sollicitons les connaissances d'un expert, et les grilles d'évaluation.

III Apport de la technologie

Il s'agit dans cette section d'évaluer la contribution technique de la technologie du web sémantique aux applications de gestion de conférences. Cela se déroulera de deux manières possibles selon les cas d'utilisations. En effet, dans le cas d'un apport d'une nouvelle fonctionnalité inexistante auparavant, nous n'aurons aucun support de comparaison. La tâche consistera alors à parler de sa plus-value, et surtout à tester la cohérence après son intégration. Nous pourrions également mesurer le gain en productivité ou la convivialité de l'outil, mais cela ne constitue pas le principal objectif de notre étude. Le second cas de figure, consiste, quant à lui, à reprendre un cas d'utilisation, tel que l'assignation des articles aux évaluateurs et essayer de l'optimiser. La tâche consistera dans ce cas à évaluer la nouvelle version du module repris, et ce, selon une des approches préalablement citées.

III.1 Automatisation de la saisie d'information

Il s'agit véritablement d'une nouvelle fonction, ajoutée grâce à la technologie du web sémantique, qui n'existait dans aucune des applications en ligne, et qui permet d'éviter de re-saisir les données pour une personne ayant déjà un profil quelque part sur le web.

En effet, nous étions guidés principalement par les nouvelles possibilités d'automatisation sur le web et une meilleure collaboration entre les hommes et les machines, qui deviennent alors aptes à comprendre la sémantique de l'information et à interpréter son sens. Aujourd'hui, l'application SWOC évite aux utilisateurs potentiels de saisir à chaque fois des informations qui les concernent et qui les décrivent. Il suffit juste que la personne soit définie par un profil FOAF sur le réseau web et qu'elle définisse l'URL du fichier en question, pour que toutes les informations nécessaires soient récupérées. Il en est de même pour les ouvrages soumis par les auteurs. L'éditeur peut désormais stocker toutes les informations nécessaires (ISBN, Titre, auteur, Etc.) sur un fichier RDF au lieu de les conserver dans une base de données. L'auteur n'aura plus alors qu'à spécifier son chemin URL pour récupérer toutes les informations lors de la soumission pour la conférence

Les figures 5.1 et 5.2 représentent des captures d'écran du formulaire de saisie d'un nouveau membre du CP.


Canadian Semantic Web Working Symposium 2007	
OpenConf Conference Management System	
OpenConf Home	Email Chair
<h3>Reviewer Committee Signup</h3> <p>Thank you for agreeing to be a reviewer. The Review Committee is a key part of the conference organization. Its role is to review and comment on paper, thus providing the input to the the Program Committee which makes the final decision on which papers are accepted and rejected.</p> <p>Note: Members of the Review Committee see unpublished work of other authors. Your professional ethics preclude disclosure to any other party the contents of the papers you read.</p> <hr/> <p>If this person already has a FOAF profile, please enter it in this area: <input type="text"/> </p> <p><i>First/Given Name</i> <i>Last/Family Name</i></p> <p>Name: <input type="text"/> <input type="text"/></p> <p>Email: <input type="text"/></p> <p>Organization: <input type="text"/></p> <p>Country: <input type="text"/></p> <p>Telephone: <input type="text"/></p> <p>Web Address: <input type="text"/></p> <hr/> <p>Username: <input type="text"/></p> <p>Password: <input type="text"/></p> <p>Retype Pwd: <input type="text"/></p>	

Figure 5-1: capture d'écran d'un formulaire de saisie d'un nouveau membre du CP (extrait du système SWOC).

<p>Areas of Interest:</p> <ul style="list-style-type: none"><input type="checkbox"/> Ontology Management<input type="checkbox"/> Ontology Alignment<input type="checkbox"/> Ontology Learning and Metadata Generation<input type="checkbox"/> Multimedia and Semantic Web<input type="checkbox"/> Semantic Annotation of Data<input type="checkbox"/> Semantic Web Trust, Privacy, Security and Intellectual Property Rights <p>Other Areas: <input type="text"/></p> <hr/> <p>Comments to Chair: <input type="text"/></p> <hr/> <p><input type="button" value="Submit Form"/></p> <hr/> <p style="text-align: center;">Powered by OpenConf Copyright ©2002-2006 Zakon Group LLC</p>
--


Figure 5-2: Suite du formulaire de saisie d'un nouveau membre du CP (extrait du système SWOC) .

Il en est de même pour les articles soumis via le site de la conférence. L'auteur n'a qu'à saisir normalement toutes les informations relatives à son ouvrage, et le système génère automatiquement une description DC, qu'il enrichit à travers ses connaissances. Le DC, au même titre que le FOAF, est un vocabulaire d'annotation des ressources web. Il repose sur le langage RDF. Ici encore, le système évite à l'auteur de remplir tout le formulaire, si son article est déjà décrit quelque part sur le web, sous le vocabulaire DC. Dans le cas échéant, Il n'aura qu'à spécifier le chemin URL de son ouvrage.

Les figures 5.3 et 5.4 illustrent le formulaire de saisie des informations qui accompagnent la soumission d'un ouvrage (article, poster ou autre).

Paper Submission

Please review the entire form before starting to fill it out to ensure you have all the required information. Blinded papers must be submitted at the same time as you fill out this form.

If this document already has a DC Description, please enter it in this area: 

General Information

Paper Title

Student Paper? Yes No

Author(s) Information

Author #1	Name:	<input type="text"/>	<input type="text"/>
	Organization:	<input type="text"/>	
	Country:	<input type="text"/>	
	Email:	<input type="text"/>	
Author #2	Name:	<input type="text"/>	<input type="text"/>
	Organization:	<input type="text"/>	
	Country:	<input type="text"/>	
	Email:	<input type="text"/>	

Figure 5-3: capture d'écran d'un formulaire de soumission d'un document à la conférence (extrait du système SWOC).

Contact
Author:

Alternate
Email: *Either an alternate email address,*

Telephone: *or a telephone number is required*

Technical Areas

Ontology Management
 Ontology Alignment
 Ontology Learning and Metadata Generation
 Multimedia and Semantic Web
 Semantic Annotation of Data
 Semantic Web Trust, Privacy, Security and Intellectual Property Rights

Content

Keywords:

Abstract:

Paper: **Format:**

Make sure paper is completely blinded (i.e., no authors info).

Figure 5-4: Suite du formulaire de soumission d'un document à la conférence (extrait du système SWOC).

Toutes ces métadonnées permettent au système d'interpréter la sémantique des ressources et d'optimiser certaines tâches, qui font défaut dans les systèmes actuels, notamment l'affectation des articles aux différents évaluateurs et la recherche d'un document ou d'une personne.

III.2 Assignment des articles aux évaluateurs

Il est très important de rappeler que l'assignation des articles a été développée comme un module indépendant. La figure 5.5 récapitule l'interaction entre les différentes composantes concernées, à savoir, les articles à assigner, les évaluateurs, et l'ontologie qui représente une modélisation des connaissances.

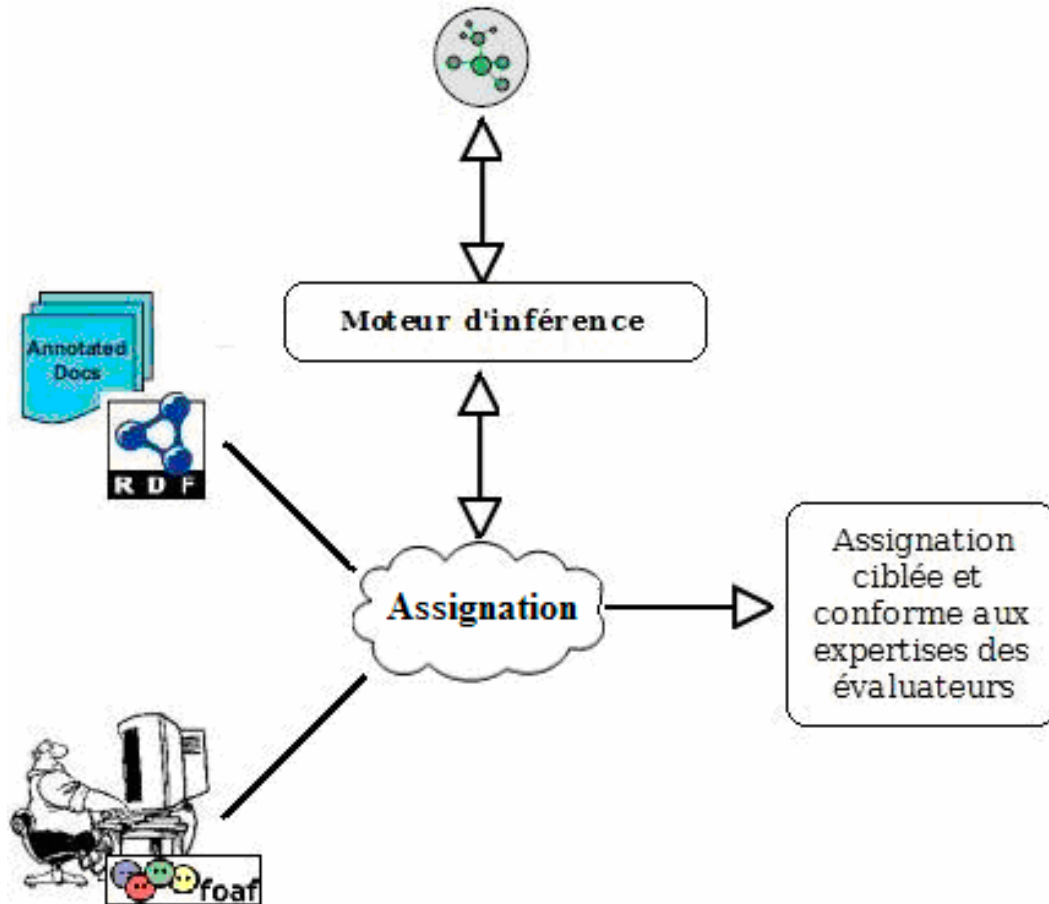


Figure 5-5: Processus d'assignation des articles aux évaluateurs.

Afin d'évaluer la performance du module d'assignation des articles aux évaluateurs de l'outil SWOC, nous avons adopté l'approche empirique, et plus particulièrement la méthode de l'interview pour la collecte de données qui nous serviront d'appui pour en tirer des conclusions. Il s'agit donc de calculer la moyenne de satisfaction pour l'assignation des articles avec l'outil OpenConf et de la comparer à celle du module que nous venons de développer afin

de tirer des conclusions quant à l'approche Web sémantique dans les applications de gestion de conférences.

Cette phase se déroule sur deux étapes. La première consiste à interroger des experts et des conférenciers ayant travaillé sur un outil Open source de base pour nous faire part de leurs appréciations. La seconde étape, quant à elle, consiste à effectuer une simulation du module que nous avons développé. Nous avons demandé à certaines personnes de tester l'application pour nous, en jouant le rôle d'évaluateurs, et ce, à défaut de pouvoir faire participer les mêmes experts qui ont utilisé OpenConf.

Une fois que les résultats de satisfaction sont calculés pour les deux outils, il ne restera plus qu'à les interpréter en tirant des conclusions.

III.2.1 Degré de satisfaction pour le module d'assignation de l'outil OpenConf.

L'approche retenue pour cette étape consiste à soumettre un questionnaire aux différents évaluateurs ayant travaillé avec le module d'assignation automatique des articles de l'outil OpenConf lors d'une conférence de renommé intitulé « European Workshop on Multi-Agent Systems (EUMAS) 2006 », organisé au Portugal. Les données brutes collectées expriment ainsi le niveau de satisfaction de chacun de ces experts pour chaque article qui lui a été affecté, et ce, à travers une note sur une échelle à trois valeurs :

- (1) : L'article est peu conforme à l'expertise de l'évaluateur
- (2) : L'article est plus au moins conforme à l'expertise de l'évaluateur
- (3) : L'article est très conforme à l'expertise de l'évaluateur

Parmi les 70 articles soumis pour cet évènement, nous en avons retenu un échantillon de 38. Nous avons ainsi pu calculer, pour chacun de ces ouvrages, le degré de satisfaction de leurs assignations, et ce, du point de vue conformité du thème par rapport aux expertises des évaluateurs qui les leurs ont été affectés. Il est à noter que chacun des articles de l'échantillon a été assigné à 3 ou 4 évaluateurs selon la disponibilité des évaluateurs.

A partir des données recueillies, nous avons déduit la moyenne de satisfaction pour l'assignation de l'ensemble des articles, en se basant ainsi sur la méthode de l'échantillonnage.

Le tableau représenté par La figure 5.6 exprime les résultats ainsi calculés pour l'ensemble des articles de l'échantillon prélevé.

ID Article	NB Evaluation	Eval. 1	Eval. 2	Eval. 3	Eval. 4	Niveau de satisfaction	pourcentage de satisfaction
A1	3	2	3	2	0	2,33	77,78
A2	3	3	1	1	0	1,67	55,56
A3	4	2	2	3	2	2,25	75,00
A4	4	2	2	2	3	2,25	75,00
A5	4	2	2	2	1	1,75	58,33
A6	3	2	2	1	0	1,67	55,56
A7	3	3	2	3	0	2,67	88,89
A8	3	2	2	2	0	2,00	66,67
A9	4	2	2	2	2	2,00	66,67
A10	3	3	2	2	0	2,33	77,78
A11	3	1	1	3	0	1,67	55,56
A12	4	1	1	2	3	1,75	58,33
A13	4	1	2	2	3	2,00	66,67
A14	4	3	3	2	2	2,50	83,33
A15	4	2	3	1	2	2,00	66,67
A16	4	1	1	2	2	1,50	50,00
A17	4	1	2	1	2	1,50	50,00
A18	3	1	2	3	0	2,00	66,67
A19	4	2	2	2	2	2,00	66,67
A20	3	2	2	1	0	1,67	55,56
A21	3	2	2	1	0	1,67	55,56
A22	4	1	2	1	2	1,50	50,00
A23	4	1	1	3	3	2,00	66,67
A24	3	1	1	2	0	1,33	44,44
A25	3	3	3	1	0	2,33	77,78
A26	3	2	2	1	0	1,67	55,56
A27	4	2	2	2	3	2,25	75,00
A28	4	3	3	2	2	2,50	83,33
A29	4	2	2	2	3	2,25	75,00
A30	3	3	2	1	0	2,00	66,67
A31	4	2	2	1	1	1,50	50,00
A32	3	3	2	2	0	2,33	77,78
A33	4	1	2	1	2	1,50	50,00
A34	4	3	2	2	3	2,50	83,33
A35	4	2	2	2	3	2,25	75,00
A36	4	2	2	2	2	2,00	66,67
A37	3	2	2	2	0	2,00	66,67
A38	3	2	2	1	0	1,67	55,56
Niveau moyen de satisfaction pour l'assignation de l'ensemble des articles :						1,97	65,57

Figure 5-6: Calcul de la moyenne de satisfaction de l'assignation des articles par l'outil OpenConf.

L'avant dernière colonne de ce tableau exprime le niveau de satisfaction de chaque article, selon l'échelle préalablement définie au début de la section. Le résultat est calculé sur la base de la formule suivante : $(\sum (\text{Notes des évaluateurs}))/\text{Nombre d'évaluateur}$. La dernière colonne, quant à elle exprime, ce chiffre en terme de pourcentage, selon la formule : $(\text{Niveau de satisfaction}/3)*100$

III.2.2 Degré de satisfaction pour le module d'assignation de l'outil SWOC

Dans cette deuxième étape de la phase d'évaluation du module d'assignation, nous nous sommes proposés d'effectuer une série de tests de l'outil que nous venons de mettre en place. L'approche consiste à utiliser le même échantillon d'articles et de l'assigner automatiquement à quelques évaluateurs. Il aurait été plus logique et pertinent de faire appel aux services des mêmes experts qui nous ont donné leurs appréciations pour l'outil OpenConf, mais faute de disponibilité, nous avons demandé à quelques personnes de jouer le rôle d'évaluateurs et de nous communiquer leur degré de satisfaction pour chaque article reçu.

Il est très important de noter que nous avons suivi la même logique de raisonnement que celle employée pour l'outil OpenConf.

Le tableau représenté par La figure 5.7 exprime les appréciations des évaluateurs par rapports aux assignations faites par l'outil SWOC, et ce du point de vue conformité du thème par rapport à leur expertise.

ID Article	NB Evaluation	Eval. 1	Eval. 2	Eval. 3	Eval. 4	Niveau de satisfaction	pourcentage de satisfaction
A1	4	3	3	2	2	2,50	83,33
A2	4	3	2	2	1	2,00	66,67
A3	4	2	2	3	3	2,50	83,33
A4	4	2	3	2	3	2,50	83,33
A5	4	3	3	2	3	2,75	91,67
A6	3	1	1	3	0	1,67	55,56
A7	4	3	2	3	2	2,50	83,33
A8	3	2	2	3	0	2,33	77,78
A9	4	3	3	2	2	2,50	83,33
A10	4	3	3	2	3	2,75	91,67
A11	4	2	2	3	3	2,50	83,33
A12	3	2	2	3	0	2,33	77,78
A13	4	3	3	2	3	2,75	91,67
A14	4	2	3	1	2	2,00	66,67
A15	4	1	3	3	2	2,25	75,00
A16	4	2	1	3	2	2,00	66,67
A17	3	1	2	3	0	2,00	66,67
A18	4	3	1	3	3	2,50	83,33
A19	3	2	2	3	0	2,33	77,78
A20	3	3	2	2	0	2,33	77,78
A21	4	2	3	3	2	2,50	83,33
A22	3	1	2	2	0	1,67	55,56
A23	4	2	2	3	3	2,50	83,33
A24	4	3	2	2	3	2,50	83,33
A25	3	3	3	2	0	2,67	88,89
A26	3	2	2	3	0	2,33	77,78
A27	4	2	2	2	3	2,25	75,00
A28	4	2	3	3	2	2,50	83,33
A29	4	2	2	2	3	2,25	75,00
A30	3	3	3	2	0	2,67	88,89
A31	4	2	1	3	3	2,25	75,00
A32	4	3	3	3	2	2,75	91,67
A33	4	3	2	3	2	2,50	83,33
A34	4	3	2	2	3	2,50	83,33
A35	3	2	3	3	0	2,67	88,89
A36	4	3	2	3	2	2,50	83,33
A37	4	3	3	2	3	2,75	91,67
A38	4	2	3	3	3	2,75	91,67

Niveau moyen de satisfaction pour l'assignation de l'ensemble des articles :

2,41

80,26

Figure 5-7: de la moyenne de satisfaction de l'assignation des articles par l'outil SWOC.

III.2.3 Interprétation des résultats

À la lumière des résultats obtenus dans le tableau de la figure 5.6, la moyenne de satisfaction de l'assignation de l'ensemble des articles, avec l'outil OpenConf, est de 1,97 sur un maximum de 3, soit 65,57 %. Cette moyenne est faible pour un outil qui se veut très performant pour l'organisation des conférences.

Par ailleurs, le second tableau, qui représente le niveau de satisfaction des assignations avec l'outil que nous venons de développer, indique une augmentation considérable de cette moyenne générale, qui passe alors de 65,57 % à 80,26 %. La figure 5.8 retrace graphiquement les données calculées dans les tableaux précédents, pour mettre davantage l'accent sur la variation de la moyenne de satisfaction générale pour les deux outils que nous comparons.

Cette satisfaction n'est pas encore à son maximum, certes, mais l'approche Web sémantique a largement contribué dans l'amélioration des résultats d'assignation. Le raisonnement basé sur la sémantique des données et la connaissance du domaine modélisé en ontologie y sont pour beaucoup.

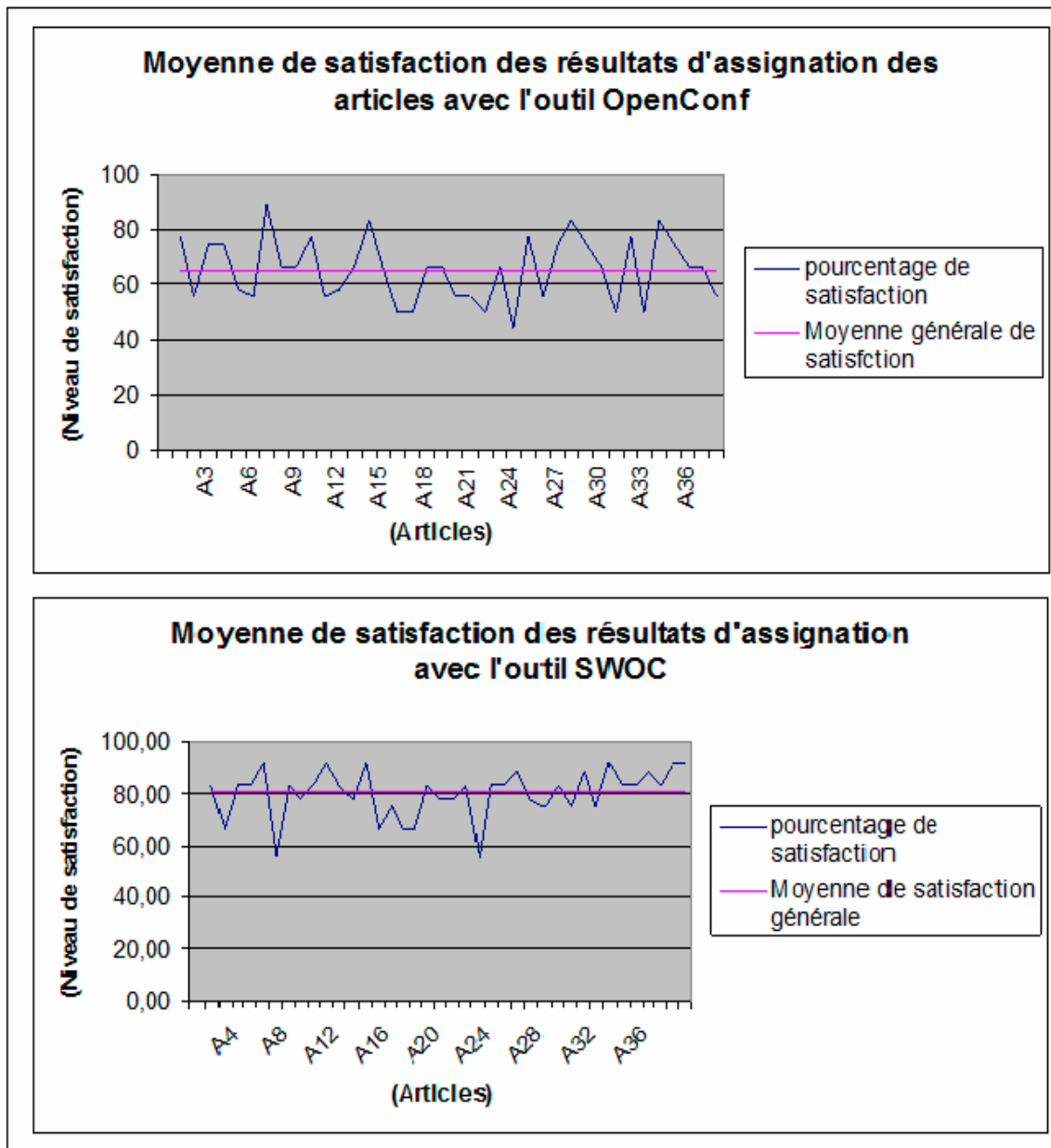


Figure 5-8: Courbes des moyennes de satisfaction de l'ensemble des articles pour les Outils OpenConf et SWOC.

Outres les résultats d'assignation plus performant de l'application, nous notons la baisse de la note « 1 » pour les assignations, souvent synonyme de non-conformité presque totale du thème de l'article avec l'expertise de l'évaluateur.

Au terme de ce qui a été réalisé jusqu'à maintenant, nous sommes en mesure d'affirmer l'apport considérable de la technologie du Web sémantique dans les applications en ligne de gestion de conférences. Cette contribution se traduit notamment à travers les nouvelles fonctionnalités qu'elle a apportées et la gestion plus intelligente dont elle a fait preuve au niveau des résultats d'assignation. Tous les objectifs fixés au début du présent document ont été ainsi atteints, et la problématique a été résolue.

Cependant, il y a encore bon nombre de critiques autour de cette technologie émergente, et qui seront abordées dans la section suivante.

IV Critiques sur le Web sémantique

Malgré la contribution considérable du Web sémantique pour une gestion plus intelligente du contenu, ainsi qu'une meilleure collaboration entre les hommes et les machines à des fins de nouvelles possibilités d'automatisation, cette technologie fait l'objet de nombreuses critiques de la part de certains chercheurs qui lui reprochent principalement son manque de pragmatisme. En effet, Clay Shirky [Shirky, 2005], par exemple soutient que la gestion d'ontologies n'est pas une sinécure et propose d'autres alternatives plus adaptées à l'Internet, tel que le « folksonomies ».

D'autres analystes pensent que les principes d'organisation sur lesquels ils reposent ne sont applicables qu'à l'échelle d'un système d'information fermé, tel qu'un réseau Intranet d'une entreprise, mais il ne fonctionnera jamais sur un réseau aussi étendu que la gigantesque toile du Web pour des raisons techniques et sociales [Shirky, 2005].

Par ailleurs, bon nombre de critiques ont été faites à l'égard du W3C, lui reprochant de consacrer trop de moyens au développement du Web sémantique, et ce, au détriment d'autres technologies comme les services Web [ZdNet, 2002]. Tim Berners-Lee, le fondateur et premier responsable du W3C, a expliqué pourtant lors d'une conférence Gartner, que ces critiques étaient infondées puisque les réflexions du W3C sur le Web sémantique, qualifiées par certains de "théoriques", pouvaient parfaitement être aussi très utiles pour les performances des Services Web. Il s'agit selon lui de deux technologies complémentaires. D'ailleurs,

l'argument qu'il présente à cet effet est que le Web sémantique est basé sur les assertions RDF, qui permettent de modéliser et décrire le monde réel, et pas seulement des documents, et ainsi faciliter les liens qui enrichissent un Service Web [Chiaramonti, 2003].

Il est très important de noter également, que malgré les efforts considérables menés autour du Web sémantique par les milieux universitaires et les centres de recherche, la sphère logicielle professionnelle reste encore assez restreinte¹. Les industriels restent en effet très sceptiques à l'idée de déployer leurs applications sous une nouvelle technologie Web qui n'a pas encore fait ses preuves. D'ailleurs, Claver Age [ClaverAge], cabinet de conseil en architecture technique, reproche aux éditeurs de bases de connaissances lancés dans le créneau d'être très marginaux dans leurs solutions et paradoxalement peu orientés Web. Ils justifient cela par la jeunesse des recommandations.

V Conclusion

A la lumière de tout le travail réalisé et en se basant sur l'évaluation de notre projet, en particulier, ainsi que sur ce qui a été lu comme critiques dans les différents ouvrages, nous pouvons dire que la technologie de Web sémantique, de façon générale, est en pleine croissance et a largement contribué à une gestion plus efficace du contenu. Malgré la taille très volumineuse de cette gigantesque toile du web, la technologie du web a permis de s'ouvrir sur de nouvelles possibilités d'automatisation, une meilleure réutilisation des données et un plus large partage des connaissances. Néanmoins, il s'agit d'une technologie qui se cherche encore et qui n'a toujours pas fait ses preuves de manière suffisante pour convaincre les industriels de son apport. D'autres travaux de recherches devraient également être menés pour explorer de plus près les couches supérieures de l'architecture présentée par Tim Berners-Lee en 2001, et y trouver des solutions.

¹<http://www.clever-age.com/veille/clever-link/le-web-semantique-en-entreprise-comment-et-a-quels-niveaux.html>

Chapitre 6: Conclusion générale

Proclamé technologie du futur, en 2001, par son créateur Tim Berners-Lee (Berners-lee, 2001), le web sémantique, propose une nouvelle plateforme permettant une gestion plus intelligente du contenu, à travers sa capacité de manipuler les ressources sur la base de leurs sémantiques. Le web sémantique favorise ainsi les coopérations Homme/Machine et permet de s'ouvrir à de nouvelles possibilités d'automatisation sur le Web.

C'est dans ce contexte que vient s'inscrire le projet SWOC. Ce dernier consiste à effectuer la ré-ingénierie d'une application en ligne existante de gestion des conférences, en intégrant la technologie du Web sémantique. Ce projet est une réponse à des besoins réels liés à l'amélioration de la gestion et l'identification des ressources, de la recherche d'informations et de l'assignation automatique des articles aux évaluateurs.

La mise en place de ce système a suivi une méthodologie précise et laborieuse, qui nous a permis de conduire ce projet de recherche dans les règles de l'art. En effet, nous avons entamé le travail par une phase de documentation sur la technologie du web sémantique, autour duquel s'articule tout le projet, qu'on appelle état de l'art. L'idée de cette phase est d'effectuer une revue de littérature complète et à savoir plus sur ce domaine, notamment en esquissant les modèles de représentation des connaissances et les langages de description tels que le RDF, le OWL et autre [Fadhel, 2001].

La seconde phase de la méthodologie adoptée consiste à couvrir le domaine d'application, qu'est la gestion des conférences, et d'exposer clairement la problématique à résoudre. Cette étape vise particulièrement à introduire tous les besoins fonctionnels d'une application de gestion de conférences et à présenter en détail tout le processus de gestion des conférences. Nous avons également effectué une analyse complète des acteurs (utilisateurs potentiels), ainsi que des différents cas d'utilisations du système

La troisième phase est d'une importance capitale et consiste à réaliser la migration du système de gestion de conférence existant, vers une plateforme Web sémantique, et ce, en

effectuant une ré-ingénierie de son architecture. Le système final répond tout à fait aux enjeux du Web sémantiques et aux standards fixés par le W3C.

L'intégration du web sémantique dans le contexte des systèmes de gestion de conférences et incontestablement bénéfique à plus d'un titre. Cette nouvelle technologie a répondu à tous les besoins enregistrés et qui s'articulent autour des faiblesses retenues par notre étude sur les applications actuelles. L'apport de cette nouvelle technologie se concrétise par une recherche plus précise de l'information, une assignation plus conforme des articles aux différents évaluateurs et l'automatisation de certaines tâches, faites jusqu'à maintenant à la main, tel que la description du profil d'une personne. Ce projet a permis, en soi, de s'ouvrir sur une gestion plus intelligente du contenu, une meilleure réutilisation et un plus grand partage des connaissances sur le Web.

Cependant, le système SWOC reste ouvert à de nouvelles perspectives, dont la priorité doit être accordée à une représentation plus fidèle du contexte. L'aptitude du système à répondre conformément et efficacement aux requêtes de l'utilisateur est intimement liée au degré de précision de l'ontologie du domaine. Plus le domaine est couvert, plus le système est riche en connaissances. Par ailleurs, il est fortement recommandé de revoir de plus près les différentes couches supérieures de l'architecture en « gâteau » du web sémantique proposée par Tim Berners-Lee, à savoir, la couche de confiance, la couche de Preuve et la couche de la logique. Du fait que celles-ci n'ont été que rarement abordées dans les différents ouvrages qui nous ont servi d'appui, il est vivement recommandé de reprendre la revue de littérature et essayer de l'approfondir davantage. Enfin, nous pourrions reprendre ultérieurement l'ergonomie de l'application en apportant plus d'efficacité et de convivialité aux interfaces graphiques.

Bibliographie

- [Balbo, 1994] : S. Balbo, Evaluation ergonomique des interfaces utilisateur : un pas vers l'automatisation, thèse de doctorat, Université Joseph Fournier, Grenoble 1, 1994.
- [Benayache, 2005]: Ahcène Benayache, Construction d'une mémoire organisationnelle de formation et evaluation dans un contexte e-learning: Le projet MEMORAE
- [Berners-Lee & al, 2001]: T.Berners-Lee, J. Hendler, O. Lassila, the semantic web, Scientific American, 284(5), mai 2001, pp.34-43
- [Berners-Lee, 2002]: T.Berners-Lee, Prepare for Next-Gen Web Now: http://boston.internet.com/news/article.php/2001_1013111, avril 2002.
- [Berners-Lee, 2004]: T.Berners-Lee, Semantic Web emerges as commercial-grade infrastructure for sharing data on the web, W3C Issues RDF and OWL recommendation
- [Borst, 1997]: Borst, W.N. (1997, September 05). Construction of Engineering Ontologies for Knowledge Sharing and Reuse, 231 pp., ISBN 90-365-0988-2.
- [Bourigault, 2001]: Didier Bourigault, Guiraude Lame, « Une ontology du droit Français pour la reformulation de requêtes sur le Web », Tal. VolumeX- n° X/2001.
- [Chiamonti, 2003] : « Semantic Web ou Services Web ? », Claude Chiamonti, VendrEDI. 16 Mai 2003, <http://xmlfr.org/actualites/decid/030516-0001>
- [Chimaera]: <http://www-ksl.stanford.edu/software/chimaera/>
- [ClaverAge] : <http://www.clever-age.com/veille/clever-link/le-web-semantique-en-entreprise-comment-et-a-quels-niveaux.html>
- [ClearLead, 2003]: ClearLead, Free Internet Access - Internet User Statistics, ClearLead Inc., 2003, http://www.clearleadinc.com/site/internet_free_access.htm
- [ClearLead, 2003]: ClearLead, Free Internet Access - Internet User Statistics, ClearLead Inc., 2003, http://www.clearleadinc.com/site/internet_free_access.htm
- [ConfTool]: www.conftool.net
- [CyberChair]: www.cyberchair.org
- [DAG-Edit]: <http://amigo.geneontology.org/dev/java/dagedit/docs/index.html>
- [Dardailler, 2002]: Daniel Dardailler, « Du Web sémantique aux Web Services, nous automatisons le travail des applications» article paru au « journal du net » le 08 février 2002, propos recueillis par Antoine Crochet Damais. http://solutions.journaldunet.com/itws/020208_it_w3c.shtml
- [Decker, 1999]: Decker, M. Erdmann, D. Fensel, and R. Studer. 1999. Ontobroker: Ontology Based Access to Distributed and Semi-Structured Information.

- [Desmontils, 2002]: Desmontils E. & Jacquin C. Annotation et indexation de documents électroniques : application au web. Cours CNAM, Ingénierie des systèmes d'information, <http://www.sciences.univnantes.fr/info/perso/permanents/desmontils/CNAM/>.
- [Dubin Core]: <http://dublincore.org/>
- [Fadhel, 2005]: Fadhel Ben Jaâfar, Une plate-forme de services web gouvernementaux en ligne dans un environnement web sémantique, Université Laval, Québec, 2005.
- [Fensel, 2002]: Fensel D. Ontology Language, v.2 (Welcome to OIL). Deliverable 2, On-To-Knowledgeproject, 2002. <http://www.ontoknowledge.org/download/del2.pdf>
- [Fernandez, 1997] FERNANDEZ, M., GOMEZ-PEREZ, A. and JURISTO, N. METHONTOLOGY: From Ontological Art Towards Ontological Engineering. AAAI-97 Spring Symposium on OntologicalEngineering, Stanford University, March 24-26th, 1997.
- [FOAF]: <http://www.foaf-project.org/>
- [Gomez-Pérez, 1999]: Gomez-Pérez, A., Benjamins, V. R., Overview of Knowledge Sharing and Reuse Components: Ontologies and Problem-Solving Methods, Proc. CAI-99 Workshop on Ontologies and Problem-Solving Methods (KRRS), 1999.
- [Gruber, 1993] Gruber T R. A Translation Approach to Portable Ontology Specifications. Knowledge Acquisition, 1993, 5 :199-220
- [Gruber, 1995] Gruber T R. Towards Principles for the Design of Ontologies Used for Knowledge Sharing. International Journal of Human Computer Studies. 1995, 43: 907 ~ 928.
- [Guenther, 2001] Rebecca Guenther and Jacqueline Radebaugh. Understanding Metadata. NISO Press, 2001. ISBN: 1-880124-62-9
- [Kietz, 2000] Jorg-Uwe Kietz, Raphael Volz, Alexander Maedche , Extracting a Domain-Specific Ontology from a Corporate Intranet, Proceedings of the Fourth Conference on Computational Natural Language Learning and of the Second Learning Language in Logic Workshop, Lisbonne, 2000
- [Lambrix, 2003]: Evaluation of ontology development tools for bioinformatics Patrick Lambrix, Manal Habbouche and Marta Perez Department of Computer and Information Science, Linköpings universitet, 581 83 Linköping, Sweden Received on October 7, 2002; revised on November 22, 2002; accepté on February 26, 2003.
- [Lévy, 2003]: Pierre Lévy, « Le futur Web exprimera l'intelligence collective de l'humanité » paru dans « le journal du net », le 25 août 2003. Propos recueillis par Pierre Lombard. http://www.journaldunet.com/itws/it_plevy.shtml

- [Macdche, 2001] Alexander Macdche and Steffen Staab. *Ontology Learning for the Semantic Web*. IEEE Intelligent Systems, March/April, 2001.
- [Marshall, Decker, M. Erdmann, D. Fensel, and R. Studer. 1999. *Ontobroker: Ontology Based Access to Distributed and Semi-Structured Information*. In R.
- [Meersman et al., 1998]: editors, *Database Semantics: Semantic Issues in Multimedia Systems*, pages 351–369. Kluwer Academic Publisher. MARSHALL C. C., *Toward an ecology of hypertext annotation*. In *ACM Hypertext*, 1998
- [Natalya, 2002]: Natalya F. Noy and Deborah L. McGuinness. “*Ontology Development 101: A Guide to Creating Your First Ontology*”
- [Neeches et al., 1991] voir mémoire de benayache
- [NetCraft, 2005]: Netcraft is an Internet services company, http://news.netcraft.com/archives/2005/10/04/october_2005_web_server_survey.html
- [OilEd]: <http://oiled.man.ac.uk/>
- [OntoText] : <http://www.ontotext.com/>
- [OpenConf]: www.openconf.org
- [Paterno, 2000]: F. Paternò, *Model Based Design and evaluation of Interactive applications*, Springer-Verlag, London, 2000.
- [PROTEGE] : <http://protege.stanford.edu/>
- [Psyché et al, 2003]: PSYCHE V., MENDES O., BOURDEAU J. (2003). *Apport de l'ingénierie ontologique aux environnements de formation à distance*, revue STICEF, Numéro spécial : *Technologies et Formation à distance*, Vol 10.
- [Sattler, 2003]: Ulrike Sattler. *Description Logics for Ontologies*, Avril 2003.
- [Senach, 1990]: B. Senach, *Evaluation ergonomique des interfaces Homme-Machine: une revue de la littérature*, Rapport n°1180, INRIA, 1990.
- [Shirky, 2005]: « *L'ontologie est surfaite* », Clay Shirky, Elanceur, 13 Mars 2006, <http://www.elanceur.org/Articles/OntologieSurfaite.html>
- [Uschold et al., 1996]: Uschold, M. et Gruninger, M. (1996) "Ontologies : Principales Methodes and Applications Knowledge", *Engineering Review*, vol. 11, n°2, juin 1996.
- [W3C Recommendation, 2004]: *OWL Ontology Web Language Overview*, W3C Recommendation, 10 Feb, 2004, <http://www.w3.org/TR/owl-features/>
- [W3C Technology, 2001]: *Metadata and Resource Description*. W3C Technology and Society domain. <http://www.w3.org/Metadata/>
- [W3C, 2004]: <http://www.w3.org/2004/01/sws-pressrelease.html.en> 2004

[W3C]: World Wide Web Consortium. [Http://www.w3.org](http://www.w3.org)

[Xavier Borderie, 2002]: Article Le web sémantique extrait du journal du net
http://developpeur.journaldunet.com/tutoriel/xml/021115xml_websemantique1a.shtml

[Yang, 2006]: Kun YANG, A conceptual Framework for semantic Web-based ecommerce, mémoire de maîtrise, Université Laval, Québec, 2006.

[ZdNet, 2002] : Article « L'éclosion des services web coince sur les standards », Paul Festa, ZDNet US, 12 Mars 2002.

Dictionnaire des acronymes

OWL: Ontology Web Language

OWL-S: Ontology Web Language Schema

RDF: Resource Description framework

RDF-S: Resource Description framework Schema

XML: eXtensible Meta Language

XML-S: eXtensible Meta Langage Schema

URI: Uniform Resources Identifier

DAML: Darpa Agent Markup Language

OIL: Ontology Inference Layer

FOAF: Friend Of A Friend

DC: Dublin core

SW: Semantic Web

XSL: Extended Stylesheet Language

XSLT: Extended Stylesheet Language Transformations

HTML: Hyper Text Meta

PHP: Hypertext PreProcessor

DL: Descriptive language

W3C: World Wide Web Consortium

SWOC: Semantic Web Open Conferences