

Université de Montréal

**Modélisation d'une ontologie et conceptualisation d'une
application sémantique dédiée au e-recrutement dans le
domaine des technologies de l'information**

par

Michel Tétreault

Département de communication

Faculté des arts et des sciences

Mémoire présenté à la Faculté des arts et des sciences
en vue de l'obtention du grade de M.Sc.
en Sciences de la communication
option cheminement libre

Août, 2011

© Michel Tétreault, 2011

Université de Montréal
Faculté des études supérieures et postdoctorales

Ce mémoire intitulé :

Modélisation d'une ontologie et conceptualisation d'une application sémantique dédiée au
e-recrutement dans le domaine des technologies de l'information

Présenté par :
Michel Tétreault

a été évalué par un jury composé des personnes suivantes :

Milton Campos, président-rapporteur
Aude Dufresne, directeur de recherche
Michel Gagnon, co-directeur
Lorna Heaton, membre du jury
Milton Campos, membre du jury

Résumé

Ce mémoire présente les recherches et réflexions entourant la conception d'une application à base d'ontologie dédiée au e-recrutement dans le domaine des services de dotation de personnel en technologies de l'information à l'ère du Web Social. Cette application, nommée *Combine*, vise essentiellement à optimiser et enrichir la Communication Médiée par Ordinateur (CMO) des acteurs du domaine et utilise des concepts issus du paradigme technologique émergent qu'est le Web sémantique. Encore très peu discuté dans une perspective CMO, le présent mémoire se propose donc d'examiner les enjeux communicationnels relatifs à ce nouveau paradigme. Il présente ses principaux concepts, dont la notion d'ontologie qui implique la modélisation formelle de connaissances, et expose le cas de développement de *Combine*. Il décrit comment cette application fut développée, de l'analyse des besoins à l'évaluation du prototype par les utilisateurs ciblés, tout en révélant les préoccupations, les contraintes et les opportunités rencontrées en cours de route. Au terme de cet examen, le mémoire tend à évaluer de manière critique le potentiel de *Combine* à optimiser la CMO du domaine d'activité ciblé. Le mémoire dresse au final un portrait plutôt favorable quant à la perception positive des acteurs du domaine d'utiliser un tel type d'application, et aussi quant aux nets bénéfices en frais d'Interactions Humain-Ordinateur (IHO) qu'elle fait miroiter. Il avertit toutefois d'une certaine exacerbation du problème dit « d'engagement ontologique » à considérer lors de la construction d'ontologies modélisant des objets sociaux tels que ceux dont le monde du recrutement est peuplé.

Mots-clés : Web sémantique, Web social, E-recrutement, Ontologie, Communication Médiée par Ordinateur, Interactions Humain-Ordinateur, Gestion des connaissances, Interopérabilité.

Abstract

This thesis presents the research and reflections on the development of an ontology based application dedicated to e-recruitment in the field of Information Technologies' staffing services, and especially e-recruitment using Social Web platforms as sources for candidates. This application, called Combine, essentially aims to optimize and enhance the field recruiters' Computer-Mediated Communication (CMC) and uses concepts from the emerging technological paradigm that the Semantic Web represents. Rarely discussed in a CMC perspective, this submission therefore proposes to study the communications issues related to this new paradigm. It presents the main concepts, as the notion of ontology, which involves the formal modeling of knowledge, and outlines the development case of Combine. It describes how the system was designed, from the requirements analysis to the prototype evaluation, revealing the concerns, the constraints and the opportunities met along the way. After this examination, the thesis stretches out to critically assess Combine's potential to optimize the CMC in the targeted field of activity. In the end, the thesis conveys a rather favorable outcome concerning the positive perception of the field recruiters about using this type of application, and also concerning the promising benefits in Human-Computer Interactions (HCI). However, it mentions as well the exacerbation of the problem called "ontological commitment", which is to consider when building ontologies that model social objects such as those which the world of recruitment is made of.

Keywords : Semantic Web, Social Web, E-recruitment, Ontology, Computer-Mediated Communication, Human-Computer Interaction, Knowledge Management, Interoperability

Table des matières

| | |
|--|-----|
| Résumé | i |
| Abstract..... | ii |
| Table des matières | iii |
| Liste des tableaux | vii |
| Liste des figures..... | vii |
| Remerciements | xi |
| 1 Introduction | 1 |
| 2 Cadre théorique : ontologies, Web sémantique et conception centrée sur l'utilisateur.... | 6 |
| 2.1 Contexte et enjeux de l'émergence du Web sémantique..... | 6 |
| 2.2 Pertinence technique des ontologies et du WS..... | 14 |
| 2.3 Présentation de l'objet central : l'ontologie..... | 19 |
| 2.3.1 Définition informatique du concept d'ontologie..... | 19 |
| 2.3.2 Description fonctionnelle d'une ontologie..... | 21 |
| 2.3.3 Description abstraite d'une ontologie..... | 26 |
| 2.3.4 Présentation des standards relatifs aux ontologies | 29 |
| 2.3.4.1 L'URI : unité d'expression du WS..... | 31 |
| 2.3.4.2 Présentation de RDF..... | 33 |
| 2.3.4.3 Présentation de RDFS..... | 38 |
| 2.3.4.4 Présentation de OWL | 39 |
| 2.3.4.5 Présentation de SPARQL | 41 |
| 2.3.5 Techniques de modélisation d'ontologies | 42 |
| 2.4 Gestion des connaissances et interopérabilité dans les organisations : bénéfices stratégiques des ontologies | 43 |
| 2.4.1 Introduction à la gestion des connaissances | 44 |
| 2.4.2 Connaissances tacites versus connaissances explicites..... | 46 |
| 2.4.3 Révolution du numérique dans le cycle de transmission des savoirs..... | 48 |
| 2.4.4 L'interopérabilité..... | 50 |

| | | |
|---------|--|----|
| 2.5 | Perspectives CMO et IHO des ontologies et du WS | 52 |
| 2.5.1 | Introduction à la CMO..... | 52 |
| 2.5.2 | Introduction à l'IHO et à la conception de systèmes centrée sur l'utilisateur | 54 |
| 2.5.3 | Le WS et la question de l'interface utilisateur..... | 56 |
| 3 | Projet Combine : Réalisation pratique d'un prototype d'application sémantique dédiée au recrutement en TI..... | 61 |
| 3.1 | Présentation du projet Combine | 61 |
| 3.1.1 | Motivations du projet | 61 |
| 3.1.2 | Objectifs | 63 |
| 3.1.3 | Hypothèses | 64 |
| 3.1.4 | Méthodologie..... | 65 |
| 3.2 | Analyse du domaine d'application | 67 |
| 3.2.1 | Description générale du domaine d'activité à modéliser | 67 |
| 3.2.1.1 | Structure organisationnelle de l'entreprise ciblée | 68 |
| 3.2.1.2 | Défis saillants du domaine..... | 69 |
| 3.2.2 | Processus, activités et acteurs du domaine | 71 |
| 3.2.3 | Inventaire des tâches de la CdP des recruteurs..... | 73 |
| 3.2.4 | Analyse et classification des tâches critiques..... | 75 |
| 3.2.4.1 | Copie des informations..... | 75 |
| 3.2.4.2 | Documentation des communications..... | 79 |
| 3.2.4.3 | Tâches reliées à la recherche de candidats | 79 |
| 3.2.5 | Remarque générale sur l'analyse des tâches | 82 |
| 3.3 | Opportunités pour les technologies du WS d'améliorer les processus du domaine d'application..... | 83 |
| 3.3.1 | Traitement intelligent des données..... | 83 |
| 3.3.2 | Navigation et visualisation plus riches des données..... | 84 |
| 3.3.3 | Approvisionnement et intégration en temps réel des données | 85 |
| 3.3.4 | Puissance et transparence accrues de la documentation..... | 86 |
| 3.4 | Modélisation de Combine..... | 86 |

| | | |
|---------|--|-----|
| 3.4.1 | Exigences et cas d'utilisation de l'application | 87 |
| 3.4.1.1 | Exigences et cas d'utilisation de la base de connaissance..... | 87 |
| 3.4.1.2 | Exigences et cas d'utilisation de l'application sémantique | 89 |
| 3.4.2 | Modélisation ontologique du domaine d'application..... | 90 |
| 3.4.3 | Modélisation de l'application | 92 |
| 3.5 | Spécifications de Combine : caractéristiques et défis contextuels de l'application | 94 |
| 3.5.1 | L'ontologie FIC et le coupleur Requis-Acquis | 98 |
| 3.5.2 | Mécanismes d'échange de données..... | 109 |
| 3.5.2.1 | Principe du screen scraping et du furetage automatisé..... | 111 |
| 3.5.2.2 | Mise à jour des données | 117 |
| 3.5.2.3 | Intégration au Web Social | 117 |
| 3.5.2.4 | Importation et exportation des données..... | 118 |
| 3.5.3 | Dispositifs de communication | 118 |
| 3.5.4 | Aperçu du déploiement du système..... | 119 |
| 3.6 | Évaluation et appréciation du prototype de Combine | 121 |
| 3.6.1 | Évaluation de l'économie de charge de travail..... | 121 |
| 3.6.2 | Évaluation de la réceptivité de la solution Combine par les utilisateurs ciblés | 125 |
| 4 | Discussion du projet Combine..... | 128 |
| 4.1 | Développements futurs et intérêts périphériques du projet Combine | 128 |
| 4.1.1 | Développements futurs de Combine | 128 |
| 4.1.2 | Intégration aux e-Portfolios..... | 129 |
| 4.2 | Discussion critique des perspectives théoriques adoptées par le projet Combine. | 131 |
| 4.2.1 | Philosophie objectiviste et problème de l'engagement ontologique | 131 |
| 4.2.2 | Approche ascendante versus approche descendante | 133 |
| 4.2.3 | Pronostic sur l'avenir des réseaux sociaux en ligne | 135 |
| 4.2.4 | Vers un Web Pragmatique ?..... | 136 |
| 5 | Conclusion..... | 137 |
| | Bibliographie..... | 140 |

Annexe 1..... i
Annexe 2..... iii
Annexe 3..... viii
Annexe 4..... ix

Liste des tableaux

| | |
|--|----|
| Tableau 1 - Sources externes de candidatures utilisées par chaque recruteur FIC (les sources sont ordonnées par ordre de fréquentation)..... | 76 |
| Tableau 2 Tableau comparatif des techniques d'élicitation des exigences logicielles (extrait Breitman et coll., 2007)..... | iv |

Liste des figures

| | |
|---|----|
| Figure 1 - Schématisation simplifiée du World Wide Web. | 14 |
| Figure 2 - Schématisation simplifiée du manque de relief sémantique du WWW du point de vue des ordinateurs. | 15 |
| Figure 3 - Captures d'écran montrant des pages web telles que vues par des humains..... | 16 |
| Figure 4 - Captures d'écran montrant des pages web telles que vues par des ordinateurs. .. | 16 |
| Figure 5 - Schématisation simplifiée de l'interprétation d'une page par un lecteur-usager-humain. | 17 |
| Figure 6 - Schématisation simplifiée de l'interprétation d'une page par un lecteur-usager-humain qui en comprend le sens grâce à son réseau conceptuel mental. | 18 |
| Figure 7 - Schématisation simplifiée d'un web sémantique annotant des ressources documentaires web. | 19 |
| Figure 8 Schéma des rapports de synthèse entre un document, une ontologie et le monde | 22 |
| Figure 9 Exemple d'un document annoté par une ontologie sur des faits géopolitiques et corporatifs..... | 22 |
| Figure 10 Cas d'utilisation d'une ontologie pour un système de gestion de contenu | 25 |
| Figure 11 Représentation graphique des trois principales composantes d'une base de connaissances, soient : des individus, des propriétés et des classes..... | 27 |
| Figure 12 Les différents étages de la modélisation RDF | 29 |

| | |
|---|-----|
| Figure 13 - Empilage des standards relatifs au WS..... | 30 |
| Figure 14 Représentation ensembliste du caractère dual de l'URI | 32 |
| Figure 15 - Vue abstraite d'un graphe RDF | 34 |
| Figure 16 Représentation d'un triplet RDF sous forme de table..... | 35 |
| Figure 17 - Illustration de David Simonds parue dans le magazine <i>The Economist</i> et caricaturant le problème de la séquestration des données entre les sites de réseautage. | 62 |
| Figure 18 - Séquence des sous-tâches impliquées dans le transfert manuel des informations d'un système à l'autre..... | 78 |
| Figure 19 - Localisation de postes offerts et identification du type de poste par une icône codée par la couleur (interface de Combine)..... | 85 |
| Figure 20 - Analyse de l'architecture d'information d'une fiche de demande de personnel . | 91 |
| Figure 21 Diagramme UML superposant le processus de recrutement actuel aux cas d'utilisation de l'application sémantique projetée..... | 93 |
| Figure 22 - L'écran de l'interface de Combine affichant la banque de candidats..... | 96 |
| Figure 23 - L'écran de l'interface de Combine affichant la banque de postes | 97 |
| Figure 24 - Aperçu simplifié des concepts et relations définissant les concepts d'Acquis de candidat et de Requis de poste..... | 101 |
| Figure 25 - Extrait d'un graphe RDF décrivant une offre de poste | 102 |
| Figure 26 - Extrait d'un graphe RDF décrivant les acquis d'un premier candidat fictif..... | 103 |
| Figure 27 - Extrait d'un graphe RDF décrivant les acquis d'un second candidat fictif | 103 |
| Figure 28 - Portion de la taxonomie des objets de spécialisation (extrait de l'ontologie FIC) montrant la distance conceptuelle entre X10, Mercury et Groovy..... | 104 |
| Figure 29 - Boîte de configuration du coupleur requis-acquis..... | 105 |
| Figure 30 - Fiche de poste avec des candidats assignés par le coupleur | 105 |
| Figure 31 - Fiche de candidat avec un poste ouvert attribué..... | 106 |
| Figure 32 - Combine en tant que tierce application pouvant automatiser l'échange de données..... | 110 |

| | |
|--|------|
| Figure 33 - Schéma illustrant les principes du <i>screen scraping</i> et du furetage automatisé. | 111 |
| Figure 34 - Capture d'écran montrant l'intégration du bouton d'extraction sur un profil LinkedIn (l'adresse URL est magnifiée pour montrer que le bouton a été effectivement intégré sur une page en ligne)..... | 115 |
| Figure 35 - Boîte de dialogue de Combine pour configurer les destinations d'affichage de poste..... | 116 |
| Figure 36 - Capture d'écran de l'interface de l'application sociale Combine. | 118 |
| Figure 37 - Diagramme de déploiement de Combine. | 120 |
| Figure 38 - Diagramme de séquence UML du cas d'une recherche simple de trois candidats telle qu'elle est pratiquée actuellement par les recruteurs FIC. | 123 |
| Figure 39 - Diagramme de séquence UML du cas d'une recherche simple de trois candidats telle qu'elle pourrait être pratiquée en utilisant Combine en mode semi-automatique. | 124 |
| Figure 40 - Diagramme de séquence UML du cas d'une recherche simple de trois candidats telle qu'elle pourrait être pratiquée en utilisant Combine en mode automatique. | 125 |
| Figure 41 - Capture d'écran d'une maquette de l'application sociale de l'e-Portfolio. | 131 |
| Figure 42 Schéma synthétique du <i>Lexion-based Ontology Development Method</i> | vi |
| Figure 43 Ontology Development 101 Method..... | vii |
| Figure 44 - Hiérarchie des processus d'une organisation | viii |

À la reine de Tout l'Doux

Remerciements

Je tiens d'abord à exprimer ma sincère gratitude à ma directrice de recherche, Aude Dufresne, qui a su m'inspirer le sujet du présent mémoire et me transmettre sa passion pour le domaine des Interactions Humain-Ordinateur.

Je tiens ensuite à remercier Michel Gagnon, professeur à la Polytechnique de Montréal, pour sa très grande générosité de m'avoir initialement permis de participer à son cours sur les ontologies et le Web sémantique, puis de s'être gracieusement offert de co-diriger mon mémoire.

Aussi, je désire souligner tout le support moral – et disciplinaire – que m'a accordé ma copine, Catherine Forand, durant cette longue période de rédaction. Son amour et ses attentions furent le meilleur remède contre les angoisses et les doutes éprouvés en cours de route.

J'aimerais exprimer ma reconnaissance envers mes parents qui ont su m'insuffler la curiosité et la force de caractère essentielle à la réalisation de ce bel accomplissement.

J'aimerais finalement remercier tous les professeurs inspirants qu'il m'a été donné de croiser au courant de mon cheminement académique.

1 Introduction

À mesure que le Web Social, communément appelé Web 2.0, s'installe dans les pratiques quotidiennes, plusieurs sphères d'activité humaine se voient chambardées. Le monde des affaires et du travail constitue un bon exemple de domaine où les manières traditionnelles de faire se voient de plus en plus questionnées par les possibilités nouvelles du réseautage. Faute de modèles ou d'applications supportant cette nouvelle réalité, il n'est pas rare de voir les entreprises désireuses d'exploiter le potentiel – encore flou – des médias sociaux sortir de leurs processus habituels pour se mettre à expérimenter des manières de faire qui leur sont inédites. L'industrie des services de recrutement et de dotation de personnel est justement un de ces secteurs d'activité très sensibles aux nouvelles pratiques de la révolution du 2.0, et tout particulièrement le sous-secteur des services de recrutement et de dotation de personnel dans le domaine des technologies de l'information (TI). En effet, la recherche présentée ici montre que pour le groupe de sujets ciblé, des recruteurs du secteur des TI, LinkedIn, une plateforme de réseautage professionnel, est le deuxième site web le plus consulté (parmi huit sites identifiés) pour dénicher de nouveaux candidats, juste après Monster.ca, un site classique de recherche d'emplois qui vend aux compagnies des accès à son répertoire de CV. Plus généralement, l'étude tend à montrer que le réseautage est maintenant considéré comme une pratique incontournable par la culture d'entreprise de ce groupe. Bien que cette exploitation encore balbutiante de la valeur des réseaux virtuels engendre des retombées positives significatives dans l'immédiat (dans les faits, l'organisation étudiée trouve et engage bel et bien des travailleurs de qualité via les sites de réseautage), il demeure que sous des considérations plus larges ou longitudinales, elle laisse planer plusieurs incertitudes, problèmes organisationnels et débats éthiques. Dans un contexte élargi nous pouvons ainsi voir apparaître des problématiques telles que la question de la propriété des données, le manque d'interopérabilité des systèmes, les jeux politiques autour du partage de contacts, le temps consacré au maintien d'une présence sociale en ligne, etc.. Bref, cette exploitation naissante du Web social affiche une valeur potentielle encore mal définie, et donc encore largement inexploitée.

Une autre dimension de la réalité communicationnelle actuelle des recruteurs en TI, qui amplifie cette dernière problématique, est l'évolution constante, voire accélérée, des moyens de produire et de diffuser de l'information. Même les objets de contenu véhiculés, c'est-à-dire les objets concrets sur lesquels portent les informations comme les technologies, méthodologies de programmation, certifications, etc., sont assujettis à une évolution tout aussi effrénée, ce qui exacerbe le problème davantage. Dans la perspective d'une théorie sur l'*économie de l'attention* (Simon, 1996 ; Davenport & Beck, 2001), ce monde plus que riche en information dans lequel évoluent les recruteurs TI les force à toujours plus de pirouettes attentionnelles pour gérer la foison de connaissances, de systèmes d'information, de documents électroniques et de dispositifs de communication qui leur permettent de demeurer compétitifs. Mais, comme le veut la maxime populaire en ergonomie cognitive à l'effet que « trop d'information tue l'information », les recruteurs TI voient souvent la qualité de leurs interventions se dégrader à mesure qu'ils enrégimentent de plus en plus de ressources informationnelles.

Qu'est-ce qui pourrait aider ?

De donner aux gens de l'industrie une vision plus large de leur situation – technologique – actuelle pourrait être un bon début. À tous ceux qui ont encore du mal à maîtriser les concepts et la frénésie chaotique du Web 2.0 (Web social), de leur montrer l'horizon d'un Web 3.0 (Web sémantique) et ainsi leur faire voir les tangentes évolutives du web devrait les placer dans une position plus proactive. Enfin, les applications à base d'ontologies font miroiter un potentiel d'interopérabilité considérable qui pourrait effectivement s'attaquer au problème de la surcharge cognitive relatif à la multiplication et la sophistication des ressources web. En effet, les ontologies, qui constituent une des composantes techniques principales du Web sémantique (WS), sont des modèles formels de connaissances qui abstraient et standardisent les représentations d'un domaine d'activité en vue de faciliter la

gestion et l'intégration des ressources informationnelles de ce domaine, que ce soit des documents, des sites web ou des logiciels.

En ce sens, ce mémoire présente le processus de développement de **Combine**, un système basé sur des principes et technologies du WS. Il présente aussi et surtout la recherche qui a été réalisée autour du processus de son développement. Essentiellement, **Combine** intègre une ontologie de domaine, un dispositif de capture de données d'écran et un fureteur automatisé à l'intérieur d'une application d'e-recrutement dédiée à supporter les processus de recherche et de dotation de personnel en TI. Comme mandat académique, le projet **Combine** visait à mettre en évidence les avantages, pour une communauté de pratique donnée (CdP¹ ; ici les recruteurs d'un groupe de consultation et de gestion de carrières en TI), d'utiliser un système qui implémente des technologies du WS. En ce sens, le projet **Combine** était en quelque sorte une étude de marché pour sonder les besoins et la réceptivité des usagers potentiels pour une telle application. Plus précisément, **Combine** fut un cas de développement logiciel centré sur l'utilisateur, c'est-à-dire basé en premier lieu sur l'analyse des interactions humain-ordinateur du champ d'application, qui se concentrait à donner un aperçu de ce qu'une application sémantique – utilisant des techniques existantes – pourrait apporter comme solutions aux problèmes identifiés.

Comment les technologies du Web sémantique, supportées par des techniques et technologies web plus courantes, permettraient-elles d'automatiser les tâches récurrentes qu'impliquent les processus d'e-recrutement ? Telle est la question pratique de cette

¹ Une Communauté de Pratique, ou CdP, est un concept qui représente un groupe de personnes assumant des fonctions similaires autour d'intérêts communs et/ou partageant certaines ressources (Wenger, 1998). Plus largement, ce concept désigne le processus collaboratif par lequel un tel groupe d'individus élabore et partage des connaissances sur leur domaine d'activité. Dans le cadre de ce mémoire, la notion de CdP sera principalement utilisée dans son acception restreinte de *groupe de personnes* en situation de collaboration et non dans celle élargie de *processus d'apprentissage collectif* d'une communauté donnée. De plus, le choix de la notion pour référer au groupe d'utilisateurs ciblés repose davantage sur la perspective de l'application **Combine** à agir comme plateforme d'explicitation et de partage de connaissances pour ce groupe (et donc de potentiellement le révéler en tant que CdP au plein sens du terme) que sur l'observation d'une quelconque motivation explicite et actuelle des acteurs en présence à réaliser des apprentissages collectifs.

recherche qui se veut au final un pronostic sur les gains de productivité probants qu'apporterait l'usage d'un tel dispositif à composantes sémantiques d'aide au recrutement. Que se soit pour comparer et coupler de manière automatique les requis d'un poste aux acquis de candidats potentiels, de pouvoir compléter la banque de candidatures à l'aide d'inférences ou simplement de pouvoir gérer, partager et diffuser les données accumulées de manière flexible et sans heurts de compatibilité, la présente recherche tend à montrer que les ontologies et les technologies du WS peuvent faire une différence à ce propos.

Le mémoire commence par dresser son cadre théorique. Il y décrit le contexte et les enjeux de l'émergence du WS puis fait état de la pertinence technique généralement attribuée aux technologies du WS telles que les ontologies en sont. Il y décortique ensuite sous ses dimensions conceptuelle, fonctionnelle, abstraite, technique et méthodologique l'objet informatique qu'est une ontologie pour finalement problématiser la thématique et l'objet à travers les prismes disciplinaires que sont la gestion des connaissances, la communication médiée par ordinateur (CMO) et les interactions humain ordinateur (IHO).

Subséquent, le mémoire présente le projet de développement de l'application Combine. À l'intérieur de ce chapitre sont d'abord détaillés les motivations, les objectifs, les hypothèses et la méthodologie du projet de recherche entourant la conception de Combine. S'en suit un compte rendu des différentes étapes de développement de l'application, soient : l'analyse du domaine d'application, l'évaluation stratégique des opportunités pour les technologies sémantiques d'améliorer les processus du domaine analysé, la modélisation haut niveau de l'application, la définition des spécifications de Combine, et l'évaluation des bénéfices potentiels pour les utilisateurs.

Suite à cela, le mémoire discute le projet. Dans un premier temps il expose les développements futurs prévus pour Combine et les intérêts périphériques qui ont eu cours durant le projet. Puis, dans un second temps, il soulève et détaille quelques dimensions critiques liées à ce projet et aux initiatives de type WS en général.

Le mémoire se conclut finalement par un retour sur les hypothèses de recherche tout en soulignant de l'importance de développer des projets pragmatiques et localisés pour amener le Web sémantique à son plein potentiel théorique, tel que prophétisé par son père conceptuel, Tim Berners-Lee (Berners-Lee & Fischetti, 1999).

2 Cadre théorique : ontologies, Web sémantique et conception centrée sur l'utilisateur

Le cadre théorique du projet ici présenté fut essentiellement celui du Web sémantique (et de la modélisation d'ontologies), en tant que corpus conceptuel, technique et technologique visant à normaliser – pour ultimement automatiser – l'édition, la gestion et le partage des informations sur Internet. En ce sens le présent chapitre expose le contexte d'émergence de ce corpus techno-théorique, il explique les concepts et principes de fonctionnement des ontologies et autres standards technologiques fondant la vision du Web sémantique, et puis décortique les notions d'*interopérabilité* et de *gestion des connaissances* qui constituent les premières instances motivationnelles à développer des projets de type Web sémantique.

Mais encore, comme le cadre théorique adopté par le projet fut interprété, utilisé et analysé à l'aune des préoccupations disciplinaires de l'IHO (Interaction Humain-Ordinateur) et de la CMO (Communication Médinée par Ordinateur), ce chapitre présente aussi les fondamentaux paradigmatiques de ces deux domaines de connaissances, comment ils permettent de problématiser le besoin d'une « sémantisation » du web et en quoi ils peuvent contribuer à l'édification des discours et dispositifs du Web sémantique.

2.1 Contexte et enjeux de l'émergence du Web sémantique

« I have a dream for the Web [in which computers] become capable of analyzing all the data on the Web – the content, links, and transactions between people and computers. A 'Semantic Web', which should make this possible, has yet to emerge, but when it does, the day-to-day mechanisms of trade, bureaucracy and our daily lives will be handled by machines talking to machines. The 'intelligent agents' people have touted for ages will finally materialize. » (Berners-Lee & Fischetti, 1999, p.169)

Telle est la vision fondatrice du WS, formulée par Tim Berners-Lee, inventeur du World Wide Web (WWW) et directeur du World Wide Web Consortium² (W3C), dans son célèbre ouvrage *Weaving the Web: the original design and ultimate destiny of the World Wide Web by its inventor* (1999). Par ce pronostique audacieux, Berners-Lee se place vraisemblablement dans la lignée des Vannevar Bush, Joseph Carl Robnett Licklider et Douglas Engelbart, ces prophètes techniques ayant inspiré le développement de l'informatique personnelle, de l'hypertexte et d'Internet. Bien que le terme « Web sémantique » origine de cette publication de 1999, l'idée d'une socialisation humaine dans laquelle participeraient activement et intelligemment les machines, de même que la croyance en une concrétisation imminente de cet état de fait, ne sont pas neuves. Tout comme le Memex de Bush (1945) et l'OLIVER de Licklider et Taylor (1968), le Web sémantique de Berners-Lee est de prime à bord un de ces dispositifs techniques conjecturaux ébauchant une solution technique séduisante et crédible au souci millénaire – toujours réactualisé et donc jamais véritablement satisfaisable – d'une facilitation mécanique des activités humaines. Il est une de ces machines conceptuelles qui ne verra peut-être jamais le jour exactement tel qu'initialement décrite par son créateur, mais dont l'élégance des solutions projetées suffit à inspirer et influencer de manière effective les paradigmes développementaux des dispositifs techniques de son époque.

Voilà maintenant plus de 10 ans (depuis 1999) que Berners-Lee a exposé officiellement³ sa vision du WS et proposé, via le W3C, une première version du standard RDF (*Resource Description Framework*) à même de fonder cette vision (en ce qu'il est un langage qui fournit un cadre général pour décrire des ressources web et leurs contenus sous forme de métadonnées formelles utilisables par les ordinateurs). Même si depuis ce temps encore très peu d'applications utilisant intégralement les principes et technologies du WS ont su s'intégrer à l'écosystème des pratiques web et survivre à la pression sélective des impératifs

² Le W3C est un organisme de normalisation des standards, protocoles et pratiques relatifs au WWW.

³ Mais officieusement il promeut l'idée de métadonnées utilisables par les ordinateurs depuis la fondation du W3C lors du *First International World Wide Web Conference* en 1994.

toujours très économiques qui les régissent, bon nombre de chercheurs et d'organisations ont continué d'alimenter l'évolution des standards reposant sur RDF et plusieurs produits web populaires récents réutilisent ou réinterprètent ces technologies à leur profit (on a qu'à penser au protocole *open graph* de Facebook). Le WS demeure donc un concept de grand intérêt qui suscite encore beaucoup de débats dans le monde du high-tech (Anderson, 2010). Au fil des ans, c'est ainsi que RDF a atteint sa version finale en 2004, que des standards comme RDFS (*RDF Vocabulary*) et OWL (*Web Ontology Language*), finalisés eux aussi en 2004 et formalisant des modèles de données RDF, furent développés pour construire des vocabulaires et des ontologies dans le langage RDF, et que, en 2004 toujours, le standard SWRL (*Semantic Web Rule Language*) fut soumis au W3C comme langage pour construire des inférences (bien qu'il n'ait jamais été approuvé par le W3C qui lui préfère le standard RIF, moins commun). C'est ainsi aussi que GRDDL (*Gleaning Resource Descriptions from Dialects of Languages*), un mécanisme pour extraire des données RDF à partir de documents XML ou XHTML, fit l'objet en 2007 d'une recommandation du W3C, que SPARQL (*SPARQL Protocol and RDF Query Language*) devint en 2008 le langage recommandé par le W3C pour effectuer des requêtes sur les données RDF, et que RDFa (*Resource Description Framework in attributes*), aussi recommandé par le W3C en 2008, vint normaliser la manière d'intégrer des descriptions RDF à même des pages HTML. Et c'est ainsi toujours que des éditeurs d'ontologies tels que *Protégé* furent développés, et que plusieurs systèmes de gestion de base de connaissances et autres dispositifs sachant faire le pont avec des systèmes de stockage de données plus traditionnels virent le jour.

C'est par ce processus de définition, toujours en cours, des standards et outils du WS que la notion de *ressources*, qui désignait originalement dans le seul cadre de RDF « les documents en ligne à organiser » (pages web, articles de blogue, courriels, etc.), se mis à désigner aussi des objets hors ligne tels que des pizzas, des instruments médicaux ou des espèces animales. En effet, grâce aux standards RDFS et OWL, qui définissent des ressources abstraites telles que les notions de «concept», de «classe» et de «propriété», il

est maintenant possible d’asserter des connaissances en langage machine à propos de n’importe quel type d’objet, qu’il soit numérique ou physique. En ce sens, les ontologies OWL sont maintenant souvent présentées en tant qu’« outils de représentation des connaissances adaptés à l’environnement Web, permettant de transformer les données en information et les informations en savoir » (Web sémantique, 2010, sec. Historique, para. 4). Vu globalement, le WS est en ce sens une entreprise d’explicitation et de mise à profit des savoirs implicites aux données brutes en ce qu’il peut opérationnaliser des contenus électroniques sur la base des interconnexions conceptuelles existant entre les objets réels auxquels ces contenus réfèrent.

Par analogie, cette transformation des informations en connaissances formalisées par des connaissances préalablement formalisées est assez représentative de la manière dont les agents humains procèdent naturellement lorsqu’ils utilisent les informations dont ils disposent pour réaliser des opérations significatives dans un contexte donné : ils tissent les informations, selon des modèles mentaux spécifiques (connaissances préalables), en des tissus cohérents et significatifs donnant lieu à de nouvelles connaissances (Frith, 1992), à des connaissances qui peuvent servir ensuite à la performance correcte de certaines tâches et activités. Et c’est toujours à travers un tel processus d’assemblage et de référencement qu’est générée la valeur de n’importe quelle information (Dawson, 2005). Mais ce processus peut rapidement devenir fastidieux lorsque nous (les agents humains) interagissons avec des systèmes d’information complexes qui contiennent des données multivariées⁴ avec des types multiples (*datatype*)⁵. Pour permettre aux ordinateurs de nous aider à faire ce travail, nous devons leur fournir des structures de métadonnées sémantiques pour qu’ils puissent effectuer, comme nous, un traitement des informations en utilisant les domaines de connaissances auxquelles elles se rapportent.

⁴ Une donnée multivariée est une donnée possédant plusieurs dimensions ou variables dépendantes.

⁵ Le type d’une donnée réfère à la sorte de structure et/ou format de cette donnée.

Techniquement, les métadonnées sémantiques sont des données qui codent les référents (connaissances) nécessaires à l'interprétation correcte du corps d'une ressource. Le web dans sa forme actuelle, appelée *Web syntaxique* par les auteurs Karin K. Breitman et coll. (2007) à cause de son manque de « relief sémantique » vis-à-vis des ordinateurs, est comparable à un supermarché dont les denrées auraient été disposées dans des emballages sans étiquettes : pour connaître quelle denrée est dans quel emballage, il faudrait systématiquement ouvrir chacun des emballages et en goûter le contenu (Foulonneau et Riley, 2008). C'est précisément ce à quoi Google excelle. Mais à mesure que les contenus web continueront leur expansion, la recherche par mots clés à la Google perdra de son efficacité (Spivack, 2008), ou du moins demandera des dispositifs de plus en plus lourds pour la maintenir. D'autre part, la recherche par mots clés est limitée au recouvrement de documents et ne peut récupérer des informations spécifiques contenues dans les documents comme tels. L'étiquetage (*tagging*), les recherches en langage naturel, les recherches sémantiques et les inférences sont les types de recherches à venir selon Nova Spivack, visionnaire technologique et CEO chez Radar Networks⁶. Avec la venue de ces nouvelles formes de recherche, la notion de « document » va se voir lentement fragmentée en celle d'« objets d'information ».

Par rapport au parcours évolutif du caractère « intelligent » des systèmes d'information, le WS peut être vu comme marchepied du web actuel par lequel il pourra atteindre de nouvelles formes d'interactivités et se faire « Web intelligent » ou « Web de la sagesse » (Spivack, 2008 ; Zhong & al., 2003). Essentiellement, cette quête pour doter les systèmes d'intelligence et de sagesse se résume en une règle simple : interconnecter les personnes et les informations (Spivack, 2008). Pour accomplir ce travail d'interconnexion, nous devons produire et utiliser des données qui ne dépendent d'aucun système particulier pour être interprétées correctement, c'est-à-dire de données qui sont *interopérables*. En

⁶ Radar Network est une compagnie Californienne fondée en 2003 par Nova Spivack et Kristinn R. Thórisson qui développe des applications sémantiques. Twine, une plateforme de réseautage social utilisant des technologies sémantiques, est le principal produit de cette compagnie.

interconnectant les données avec des métadonnées bien formées, le WS apparaît comme une des voies les plus prometteuses pour assurer l'interopérabilité des structures d'information. L'objectif du WS est bien connu : transformer Internet en une gigantesque base de données distribuée. On peut dire du discours sur le SW qu'il s'agit là de sa vision ultime de l'interopérabilité. Toujours selon Spivack (2008), une fois concrétisée, cette vision devrait permettre la création d'une nouvelle génération de web, le « Web OS »⁷, qui deviendra la plateforme du Web intelligent, un Web par lequel les gens délègueront leurs tâches computationnelles à des agents personnels intelligents.

Le WS est un des concepts majeurs dans la construction du troisième grand palier d'interactivité d'Internet. C'est pourquoi on lui accole souvent, par métonymie, le label « Web 3.0 » de ce troisième palier. Dans cette perspective d'un versionnage du web, rétrospectivement, la nomination « Web 1.0 » tend à représenter les premières années de l'entrée en scène du World Wide Web (1990-2000) suite à l'invention d'HTTP et HTML, les standards constitutifs de ce dernier. À ce premier niveau d'interactivité, producteur et consommateur de contenu étaient deux rôles bien distincts, unis par une relation unidirectionnelle : le producteur acheminant du contenu (principalement des documents au format HTML) aux consommateurs requérants. Autour de 2000, cette relation simple devint tranquillement réversible et la distinction entre les rôles de producteur et de consommateur se brouilla. C'est le début de l'ère du contenu généré par les utilisateurs. S'étirant jusqu'à aujourd'hui, cette période, appelée « Web 2.0 », marque une multilatéralisation du web de par une amélioration de ses composantes d'avant-plan : les interfaces usager devinrent plus dynamiques et utilisables, alors que les contenus commencèrent à être dissociés de la manière dont ils sont présentés (notamment avec l'apparition du standard CSS⁸). À ce point, le WS et le Web 3.0 en général peuvent être vus

⁷ Le Web OS (*Web Operating System*) désigne une forme projetée de l'informatique où le réseau de tous les services web sémantiques composera un seul système d'exploitation global.

⁸ CSS est l'acronyme de *Cascading Style Sheets* (feuilles de style en cascade), un langage permettant de formaliser, dans un fichier indépendant, les instructions relatives à la présentation (style) d'un document balisé telle qu'une page HTML ou un fichier XML.

comme un retour à des innovations d'arrière-plan, des innovations visant à faciliter l'échange et le partage des données. En ce sens, RDF⁹ (*Resource Description Framework*), le standard central du WS, fera pour les données, et les ressources électroniques en général, ce que HTML et le Web 1.0 ont fait pour les documents (Henricks, 2007). Avec d'autres standards tels qu'OWL (*Web Ontology Language* ; langage pour construire des ontologies), SWRL (*Semantic Web Rule Language* ; langage pour manipuler des connaissances) et SPARQL (*SPARQL Protocol and RDF Query Language* ; langage pour interroger et assembler des connaissances), le WS devrait permettre, avec l'usage d'agents logiciels, une automatisation intelligente de la gestion des contenus qui donnera le signal d'envoi pour le Web Intelligent et l'horizon d'un Web 4.0 (2020-2030). Théoriquement et techniquement, le WS et ses technologies afférentes sont fonctionnellement matures, mais le concept est encore dans l'attente d'applications ultra-populaires à même d'enclencher de manière généralisée le processus de génération et de consommation des données sémantiques, et ainsi le faire émerger en tant que concept Internet dominant de la prochaine décennie.

Selon les auteurs Jorge Cardoso, Miltiadis Lytras et Martin Hepp (2005), cinq grandes caractéristiques du monde actuel constituent le « terrain de jeu » des applications à base sémantique, soit : la globalisation, le réseautage, les modèles partagés, l'intelligence collective et le paradigme de l'*open*. De ces cinq caractéristiques on peut mettre en évidence un certain nombre de défis et d'opportunités contextuels pour les applications sémantiques et, par extension, pour n'importe quel effort visant à assurer l'interopérabilité des systèmes. D'abord, contrairement aux modèles traditionnels qui ne prennent en considération qu'une perspective très étroite des objectifs du réseautage, le WS, en fournissant une infrastructure de sémantiques partagées, offre une perspective élargie dans laquelle de nouveaux besoins et opportunités d'affaires émergent. Les ontologies et les réseaux sociaux balisés par des modèles conceptuels bien définis permettraient de coupler

⁹ Tenant pour *Resource Description Framework*, RDF est un dialecte XML conçu pour exprimer des connaissances déclaratives sous forme de phrases élémentaires de type sujet-prédicat-objet que l'on appelle des triplets RDF.

les sources d'information et les services humains de sorte à créer des occasions d'affaires inédites qui appellent à la définition de nouveaux modes de réseautage des personnes, des connaissances et des entreprises. Cela en appelle aussi à une globalisation de l'information et à la définition de nouveaux contextes pour exploiter la valeur ajoutée des systèmes sémantiques. En effet, l'approvisionnement des banques d'informations locales à des dépôts de connaissances globalement partagés et la conceptualisation de nouveaux contextes d'exploitation représentent deux des valeurs phares du WS. En ce sens, le design d'ensembles d'informations et de connaissances interopérables possédant plusieurs niveaux de référencement (pour un même ensemble) permet la création de systèmes plus dynamiques et plus personnalisés. Dans la même veine, l'interopérabilité permet donc ultimement de fournir et d'ajouter de la qualité à l'information. Dans un contexte d'explosion des moyens de diffuser du contenu, le web actuel souffre de plus en plus de sa faible capacité à explorer les informations qu'il recèle de manière qualitative. Paradoxalement, le fait que les entreprises, les consommateurs, les citoyens, et tout autre type d'utilisateurs continuent de réclamer des systèmes et des infrastructures qui desservent des informations de qualité, de mieux en mieux assurée par des modèles de contrôle, crée une situation à double contrainte dans laquelle le WS devrait constituer une transformation salutaire.

En bref, le déploiement du WS dépend pour beaucoup de son adoption par l'industrie (Cardoso et coll., 2005). Pour ce faire, il est important pour les technologies du WS de s'intégrer aux modèles corporatifs actuels et d'adopter le langage du retour sur investissement en offrant des exemples d'intégration dans lesquels le WS remplit ses promesses.

2.2 Pertinence technique des ontologies et du WS

Tel que l'a très bien vulgarisé l'inventeur du World Wide Web (WWW) Tim Berners-Lee (1994) lors de sa plénière à la première *International World Wide Web Conference* à Genève, le web, dans sa forme hypertextuelle classique, se résume essentiellement en un ensemble de nœuds (pages HTML) interconnectés par des liens hypertextes (voir Figure 1).

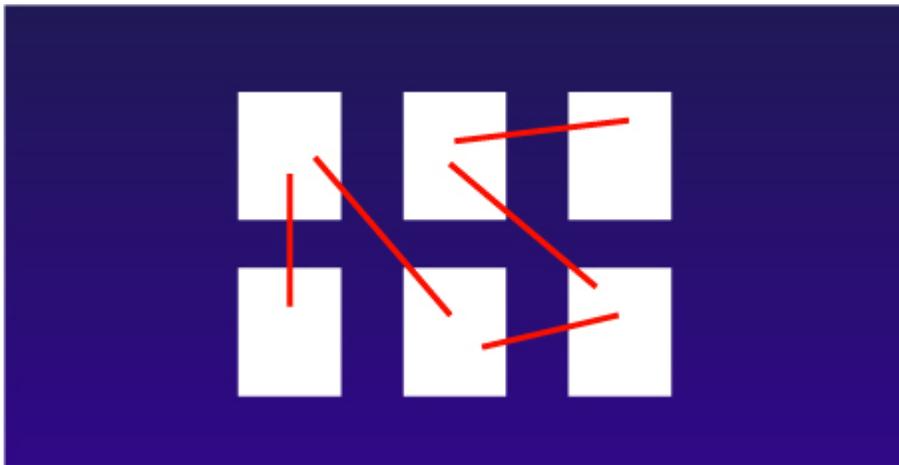


Figure 1 - Schématisation simplifiée du World Wide Web (extrait retouché de Berners-Lee, 1994).

Bien que l'arrivée de cette nouvelle forme médiatique qu'est l'hypertexte du WWW fut spectaculairement libératrice pour les usagers d'Internet, il demeure que d'un point de vue machine, ce web est totalement plat et dénué de signification (voir Figure 2)

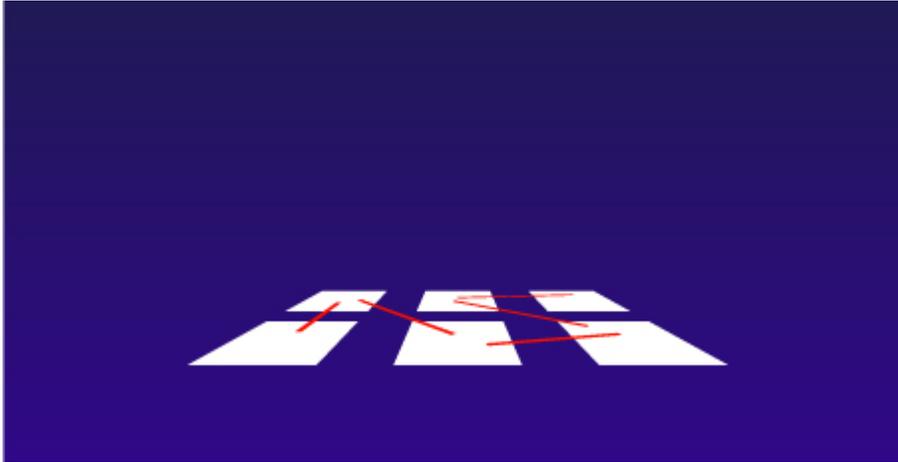


Figure 2 - Schématisation simplifiée du manque de relief sémantique du WWW du point de vue des ordinateurs (extrait retouché de Berners-Lee, 1994).

Ce qui apparaît aux humains comme étant des pages contenant des informations intelligibles (voir Figure 3), apparaît aux ordinateurs comme pur charabia de caractères et de pixels (voir Figure 4).

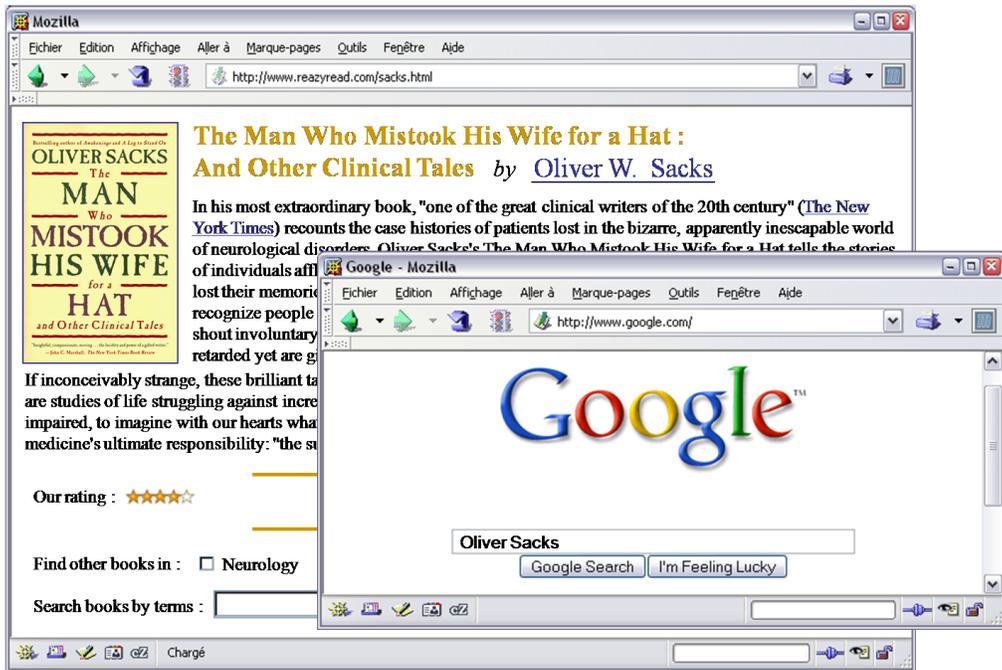


Figure 3 - Captures d'écran montrant des pages web telles que vues par des humains.

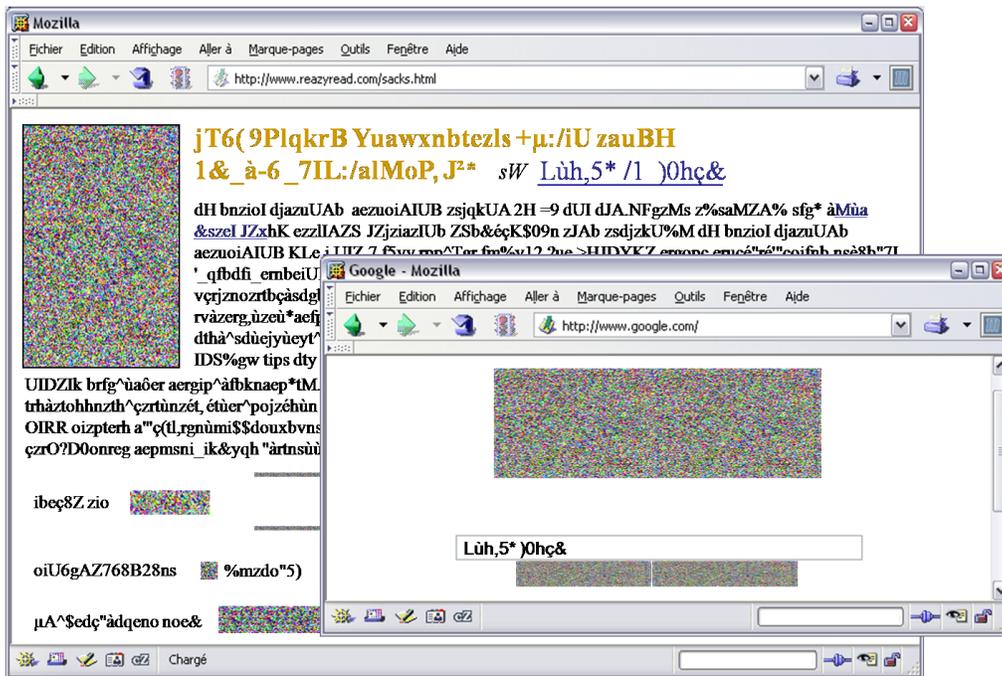


Figure 4 - Captures d'écran montrant des pages web telles que vues par des ordinateurs.

L'inconvenance d'un tel état des choses est donc que les ordinateurs ne peuvent aider les humains à générer, gérer et fureter les informations contenues sur le web que par des moyens statistiques et syntaxiques qui sont peu « intelligents » – dans le sens où ils ne peuvent rendre compte de l'intelligibilité des contenus véhiculés – et sont donc limitatifs vis-à-vis la diversification toujours croissante des formats documentaires. Pour pouvoir interpréter, et par conséquent faire sens d'un document, un lecteur doit établir des liens dits « sémantiques » entre les éléments de contenu qu'il renferme (mots, photos, vidéos, etc.) et les objets du monde réel auxquels ces éléments font référence. Par exemple (voir Figure 5), un usager qui consulte une page qui parle de maisons comprendra le sens du document que dans la mesure où il établira un isomorphisme entre le signifiant textuel qu'est le mot « maison » extrait du le document, et l'objet sensible composé de quatre murs, un toit, une porte,... qu'il aura prototypé et stéréotypé sous forme de représentation mentale (concept), ou sous forme de signifié pour parler en termes saussuriens.

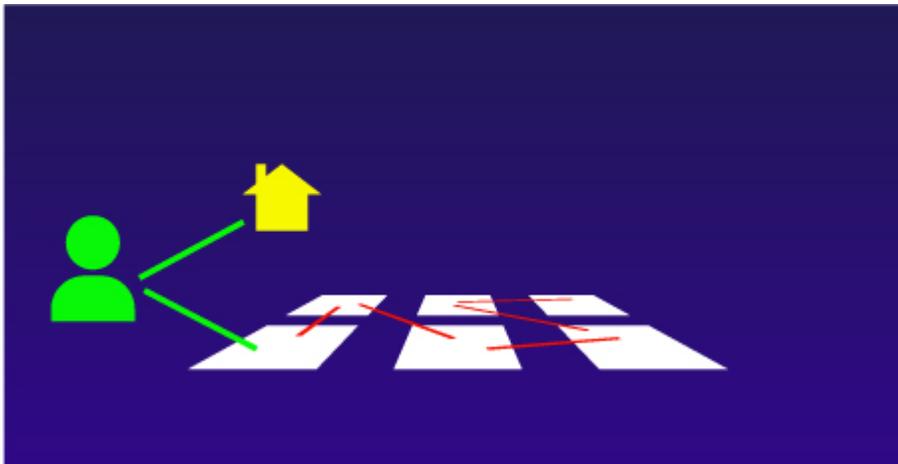


Figure 5 - Schématisation simplifiée de l'interprétation d'une page par un lecteur-usager-humain (extrait altéré de Berners-Lee, 1994).

Mais comme nous n'avons jamais un accès direct aux objets de la « réalité » empirique (Zwirn, 2000), au niveau conceptuel, les concepts (ou signifiés) se voient alors définis par les liens qu'ils entretiennent avec d'autres concepts : le sens d'un concept se résume en

conséquence à la manière dont il participe à l'écosystème conceptuel dans lequel il s'insère, à la manière dont il est interconnecté aux autres concepts d'un certain paradigme – voir même du langage dans sa totalité (voir Figure 6). Très grossièrement, on dira enfin d'un tel réseau paradigmatique qu'il forme une *ontologie*.

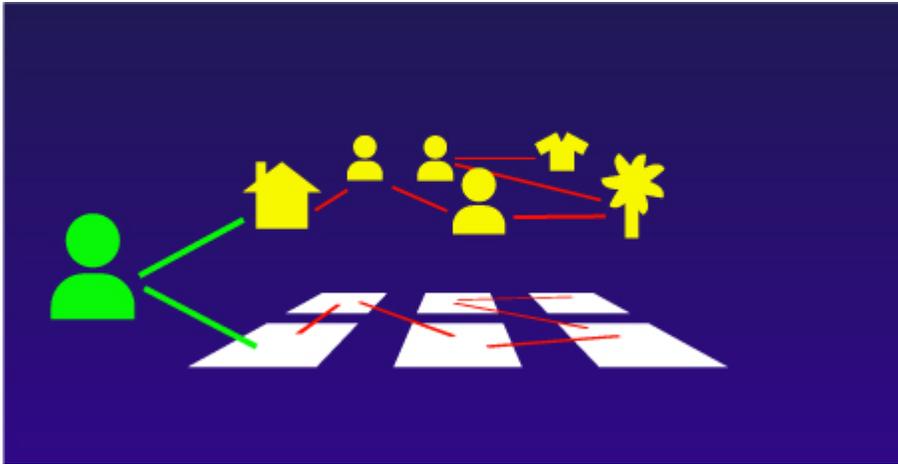


Figure 6 - Schématisation simplifiée de l'interprétation d'une page par un lecteur-usager-humain qui en comprend le sens grâce à son réseau conceptuel mental (extrait retouché de Berners-Lee, 1994).

Une fois mise en relation avec des éléments d'information contenus des pages web (ou n'importe quelle autre ressource en ligne), on pourra alors parler de l'ontologie comme composant un *web sémantique* (voir Figure 7), c'est-à-dire un réseau de « choses » permettant d'explicitier la signification des documents annotés. En fait, comme il s'agit de données (les informations traitées à un niveau conceptuel) à propos d'autres données (les informations contenues dans les documents eux-mêmes), il est de mise de dire des webs sémantiques et des ontologies qu'ils constituent des structures de *métadonnées*. À l'échelle globale, la prise en compte de l'ensemble de ces structures de métadonnées sémantiques composera la notion générale de Web sémantique (avec majuscule) en tant qu'elle dénote cette façon particulière d'organiser les informations en ligne.

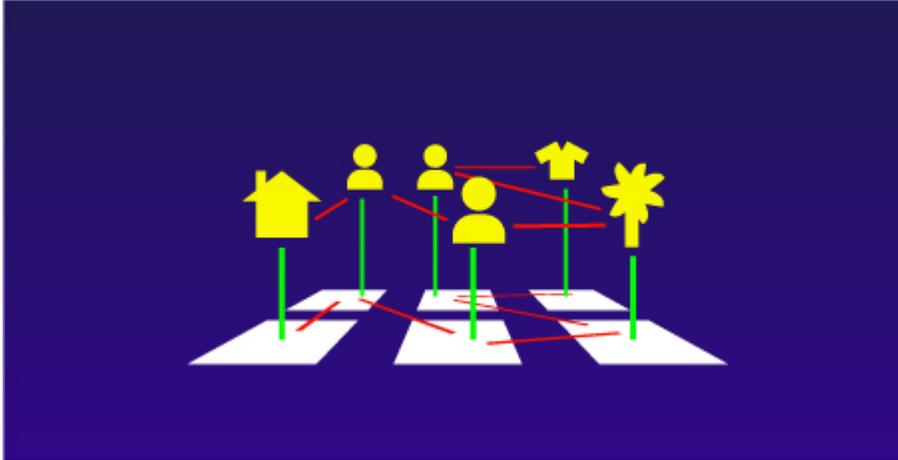


Figure 7 - Schématisation simplifiée d'un web sémantique annotant des ressources documentaires web (extrait retouché de Berners-Lee, 1994).

Dans l'optique du Web sémantique (WS), les ordinateurs (via des agents logiciels) seront dès lors habilités à assister les agents humains.

2.3 Présentation de l'objet central : l'ontologie

2.3.1 Définition informatique du concept d'ontologie

Dans l'univers informatique, le concept d'ontologie, qui est emprunté à la philosophie, a été utilisé dans plusieurs secteurs, principalement en intelligence artificielle, en représentation des connaissances, en traitement des langues naturelles, en WS et en ingénierie logicielle (Breitman et coll., 2007). Techniquement parlant, une ontologie, tel qu'elle est utilisée en informatique, peut être minimalement définie en tant qu'artéfact conceptuel qui rassemble un vocabulaire de termes, des spécifications quant à la signification de ces termes et une description de la manière dont ils sont interreliés les uns aux autres (Ushold et Jasper,

1999). Les termes en question correspondent normalement à ce que l'on appelle « classes », « concepts » ou « choses ». Les spécifications correspondent pour leur part aux « propriétés » ou « attributs » que devrait avoir un objet tombant sous un certain concept. Et les descriptions d'interrelation représentent quant à elles les « relations » que certains concepts entretiennent conceptuellement avec d'autres concepts. Les ontologies peuvent être utilisées pour décrire différents types de modèles. Elles peuvent par exemple décrire des taxonomies élémentaires comme elles peuvent décrire des modèles plus élaborés écrits dans une variante de la logique classique du premier ordre, comme la logique descriptive (Breitman et coll., 2007).

Au-delà de son simple rôle de modèle standardisant des éléments de données, les ontologies visent à expliciter la sémantique (la signification) des ressources informationnelles décrites par les données RDF. En ce sens, on peut dire des ensembles de données RDF qui réfèrent à des ontologies comme modèles qu'ils sont auto-descriptifs : la manière d'interpréter les données fait partie intégrante des données elles-mêmes et ne dépend d'aucun système particulier pour ce faire. C'est pourquoi l'usage de telles données sémantiques constitue une avancée notable en matière d'interopérabilité des systèmes. Selon la précision sémantique¹⁰ du modèle décrit par une ontologie (un glossaire, une taxonomie ou une ontologie complète), des opérations plus ou moins « intelligentes » peuvent être effectuées sur les données utilisant cette ontologie, parce que, notamment, les données sémantiques font preuve d'un potentiel d'interconnectivité accru (à cause de RDF) et que la plus ou moins grande profondeur définitionnelle de la terminologie modélisée donne aux agents logiciels le pouvoir de partager, d'échanger et/ou de transformer les données d'une manière *signifiante*. OWL (*Web Ontology Language*), qui a été utilisé pour modéliser l'ontologie de Combine, est un langage de modélisation d'ontologie qui peut utiliser RDF et RDF Vocabulary (RDFS) pour être exprimé. Ce langage fournit des opérateurs de la logique descriptive qui permettent la formalisation de connaissances déclaratives (ce que l'on

¹⁰ La précision sémantique d'une ontologie réfère au niveau de formalisme et de descriptivité du vocabulaire modélisée.

appelle des « faits »). Ces connaissances une fois formalisées de la sorte peuvent ensuite permettre aux ordinateurs d'effectuer des traitements sur les informations comme telles, et non seulement sur leur seule mise en forme comme c'est le cas avec la plupart des systèmes traditionnels. Une fois que l'ontologie est instanciée, c'est-à-dire qu'on y ajoute des entités qui sont des instances des concepts qu'elle définit, on peut alors parler de l'ensemble comme constituant une « base de connaissances ». Pour bien distinguer ces deux composantes constitutives de ce qu'est une base de connaissances, il est de mise de référer à la terminologie définie par l'ontologie (structure de classes et de propriétés) en tant qu'elle constitue une « TBox » (Boîte Terminologique) et à la collection des instances assertées de la base en tant qu'elle constitue une « ABox » (Boîte d'Assertions). Par exemple, une TBox pourrait définir le terme « chien » par la proposition « un chien est un animal » et une ABox pourrait alors utiliser ce terme pour affirmer que « Rex est une instance de chien ». Cette distinction entre TBox et ABox est importante dans la mesure où une même TBox peut servir de modèle à plusieurs ABox.

2.3.2 Description fonctionnelle d'une ontologie

Étymologiquement, la notion d'ontologie remonte à la Grèce antique et désigne une « théorie de l'existence ». Elle est constitutive de ce que l'on appelle communément une « vision du monde » : une reconstruction synthétique du monde qui sert de cadre conceptuel pour interpréter tout ce que l'on pourra en dire (du monde) subséquent. Et cette fonction de « cadre conceptuel » qu'incarne l'ontologie (dans son sens large) s'offre en tant qu'*ordre* représentatif nous permettant alors de *raisonner* sur ce qui existe. La version informatique du concept, qui fut introduite via les recherches en IA, dénote plus précisément une structure taxonomique – augmentée – qui répond au besoin d'explicitier (pour mieux contrôler) le vocabulaire d'un domaine d'application spécifique (voir Figure 8 et Figure 9). Dans son rôle informatique, une ontologie se veut donc une structure conceptuelle formelle qui possède des relations inter-conceptuelles augmentant les relations

conceptuelles d'hyperonymie/hyponymie (classe/sous-classe ; type/instance de type) qui forment son ossature taxonomique (hiérarchie de concepts structurée par la seule relation de type « est un »).

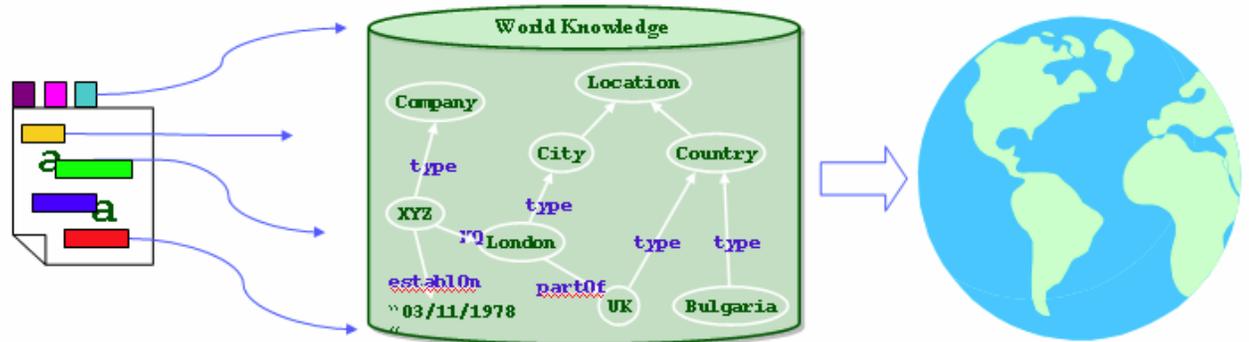


Figure 8 Schéma des rapports de synthèse entre un document, une ontologie et le monde (Extrait de Kiryakov, 2003)

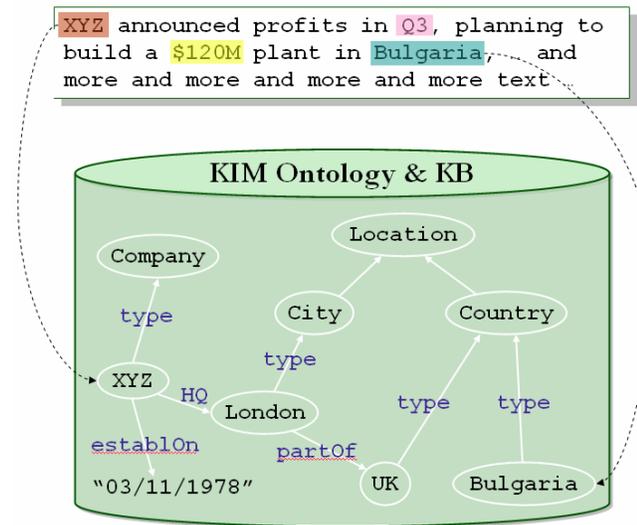


Figure 9 Exemple d'un document annoté par une ontologie sur des faits géopolitiques et corporatifs (Extrait de Kiryakov, 2003)

Dans ce contexte plus formel, comme dit plus tôt, l'ontologie est un élément constitutif de ce que l'on appelle techniquement une « base de connaissances », une sorte de base de

données où tous les champs de noms sont sémantiquement interconnectés entre eux (ce que fait l'ontologie) et sur laquelle un système de règles (de type si-alors) peut reposer pour faire des inférences logiques sur les connaissances qu'elle contient. Ceci permet donc une intercompréhension humain-machine (et machine-machine) accrue puisque l'esprit humain fonctionne sensiblement sur le même principe (interprétation par stéréotypation du monde ; Andler, 2002 ; Fortin et Rousseau, 1989 ; Pinker, 2000 ; Quine, 1977 ; Searle, 1980). En ce sens, le concept informatique d'ontologie est assurément l'une des pièces les plus prometteuses de l'artillerie actuellement mise en place pour mener à terme cette révolution qu'est le web sémantique et ainsi paver la voie pour la venue des «agents intelligents» rêvés par Tim Berners-Lee.

Un exemple concret de ce à quoi pourrait ressembler la notion d'« agent intelligent » de Berners-Lee serait de pouvoir demander à son ordinateur personnel en langage naturel « Achète un billet d'avion le moins cher possible pour un aller-retour Montréal-Lima durant mes prochaines vacances d'été », et que l'ordinateur soit capable d'aller seul sur le Net pour s'acquitter automatiquement de cette tâche. Mais pour ce faire, l'ordinateur devra avoir accès à une certaine connaissance d'arrière plan pour mener à bien l'opération. Une ontologie fait ce genre de travail. Elle pourrait en l'occurrence renseigner la machine qu'« acheter » est une relation entre une « personne » et une « marchandise », que « billet » est justement une sorte de « marchandise » (et donc achetable par une « personne »), que « cher » est une propriété relative au « prix » d'une « marchandise », qu'un « avion » offre des « vols », qu'un « vol » est une sorte de « marchandise » échangeable contre un « billet », qu'un « aller-retour » est une sorte de « vol », qu'un « vol » relie deux ou plusieurs « aéroports », que l'« aéroport » de « Montréal » se nomme « Aéroport international Pierre-Elliott Trudeau », et ainsi de suite... Sur la base de cette connaissance d'arrière-plan, les agents intelligents de l'ordinateur seraient alors habilités à faire de manière autonome toute une série de transactions telles qu'interroger l'agenda de son maître-utilisateur pour identifier la date de ses vacances, repérer et se connecter à des services de voyages en ligne, sonder et comparer les prix, procéder à l'achat de la meilleure

aubaine, consigner les informations sur le vol acheté, ajouter le voyage à l'agenda de l'utilisateur,...

Il est à savoir cependant qu'une telle connaissance d'« arrière plan » que fournit une ontologie n'agit pas dans les faits « derrière », mais en juxtaposition réflexive aux informations qui circulent. D'où la notion introduite plus tôt à l'effet que l'ontologie est une structure de *méta*-données.

Bien que la notion de métadonnées exploitables par les machines remonte jusqu'aux balbutiements d'Internet, ce n'est qu'aux alentours de 1990, avec l'entrée en scène du World Wide Web et son croisement des concepts d'adresse web (URL) et de langage par balises (HTML) que l'idée d'une « sémantisation » de l'Internet par méta-représentation devint soudainement plus concrète. Dans l'incubateur du W3C (*World Wide Web Consortium*), le web sémantique se résume en un lot de méthodes (et de standards) orientées vers la modélisation de ressources et de contenus en tout genre. Les ontologies sont une de ces formes de modélisation et s'attardent spécifiquement à la modélisation de connaissances dites *déclaratives* (ce sont des connaissances sur l'état des choses qui s'expriment avec des propositions que l'on appelle des « faits » ; elles s'opposent – de manière complémentaire – aux connaissances dites *procédures*).

Prenons l'exemple (Figure 10) d'un individu qui demanderait à un système d'information (ici un système de gestion de contenu ; CMS) d'avoir de l'information à propos de Δ . Étant équipé d'une ontologie qui possède la connaissance déclarative à l'effet que « Δ est une sorte de Δ », et qu'il se trouve que des documents sont annotés comme traitant de Δ , le système d'information peut inférer que ces documents sont en mesure d'intéresser le requérant et les lui envoyer.

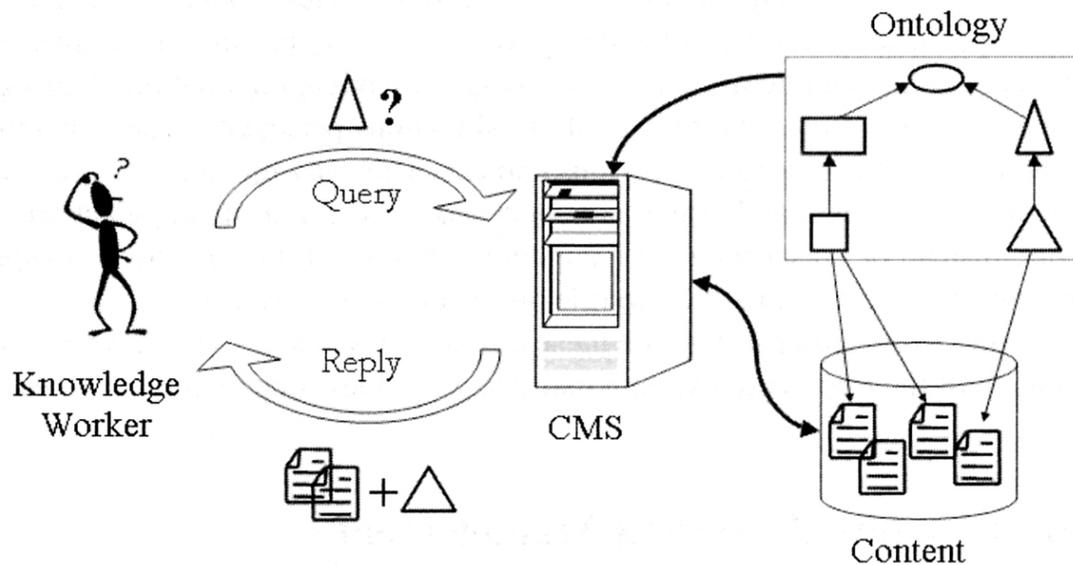


Figure 10 Cas d'utilisation d'une ontologie pour un système de gestion de contenu (tirée de Staab & Studer, 2004)

Enfin, pour donner un portrait un peu plus spectral des divers usages pertinents d'une ontologie, voici ce que Noy et McGuinness (2007), dans un cahier intitulé *Développement d'une ontologie 101*, identifient comme étant quelques bonnes raisons de créer et d'utiliser une ontologie :

- 1) *Partager la compréhension commune de la structure de l'information entre les personnes (ou les fabricants de logiciels)*
 - Améliore le partage d'information entre plusieurs usagers. Le partage d'une même ontologie par plusieurs systèmes permet à des agents informatiques d'agrèger (de déduire) des informations, non nécessairement explicites, pour fournir à un usager une réponse à une question spécifique, ou encore pour servir de données d'entrée pour programme quelconque.
- 2) *Permettre la réutilisation du savoir sur un domaine*
 - Améliore la transposition des concepts opérationnalisés d'une étude à l'autre pour un même domaine d'intérêt. Les ontologies peuvent être récupérées, modifiées et/ou fusionnées de sorte à recouvrir de proche en proche de plus vastes domaines de connaissances.

3) *Expliciter ce qui est considéré comme implicite sur un domaine*

- Améliore la discussion et l'apprentissage des postulats souvent implicites d'un domaine. Les ontologies rendent claires, et cohérentes, les connexions sémantiques de termes spécialisés liés à un domaine spécifique, souvent implicitement définis.

4) *Distinguer le savoir sur un domaine du savoir opérationnel*

- Facilite les processus de configuration des produits de type modulaire. Les ontologies permettent, par la seule explicitation fine des liens sémantiques unissant les composantes et propriétés d'un produit, de générer automatiquement, à l'aide d'un algorithme, le processus d'assemblage répondant aux choix spécifiques d'un client.

5) *Analyser le savoir sur un domaine*

- Améliore la rigueur, et l'extensibilité du vocabulaire d'un domaine. Les ontologies permettent de valider la cohérence, d'un point de vue formel, de la structure logique du réseau de connaissances qu'il constitue.(consistance définitionnelle). Elles permettent de greffer, en toute bonne logique, des parties d'autres ontologies connexes. Par auto-évaluation ou par comparaison, elles permettent la détection d'apories¹¹ et/ou de contradictions.

2.3.3 Description abstraite d'une ontologie

D'un point de vue technique, une ontologie constitue une structure conceptuelle qui, en plus de posséder des relations d'hyponymie/hyperonymie (classe/sous-classe) qui forment son ossature taxonomique, possède aussi des relations transversales définissant des rapports inter-conceptuels qui viennent se superposer à la hiérarchie de concepts structurée par la seule relation de type « est un » de sa taxonomie de base. Plus spécifiquement, il s'agit

¹¹ Une aporie est une lacune dans une construction discursive qui provoque chez le lecteur des difficultés interprétatives (embarras ou dilemme). Dans son sens fort (et plus actuel), une aporie représente une contradiction insoluble.

d'une description formelle explicite des divers concepts pertinents et relatifs à un domaine de discours spécifique.

« Une ontologie comprend des **classes** (appelées aussi **concepts**), des **attributs** applicables aux concepts pour en décrire les caractéristiques (appelés parfois **rôles** ou **propriétés**) et des **facettes** sur les attributs (appelées parfois **restrictions de rôles**). Une ontologie, complétée par l'ensemble des **instances** individuelles inscrites dans ses classes, constitue dès lors une **base de connaissances**. Il y a en réalité une frontière subtile qui marque la fin d'une ontologie et le début d'une base de connaissances. » (Noy et McGuinness, 2007, p.3).

La Figure 11 donne un aperçu graphique de ce que représentent les trois principales composantes d'une base de connaissances que sont les individus, les propriétés et les classes.

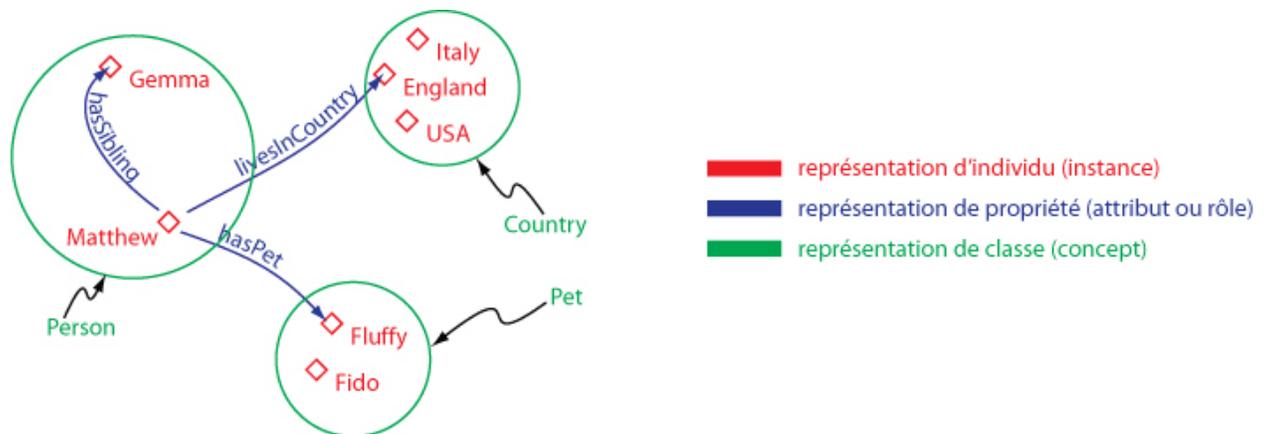


Figure 11 Représentation graphique des trois principales composantes d'une base de connaissances, soient : des individus, des propriétés et des classes (figure extraite et adaptée de « A Practical Guide To Building OWL Ontologies Using Protégé 4 and CO-ODE Tools », URL : www.co-ode.org/resources/tutorials/ProtegeOWLTutorial.pdf)

La Figure 12, tirée d'un livre sur la modélisation XML, illustre les différents « étages » de la modélisation RDF. Les ontologies étant des documents RDF, cette figure se prête donc bien pour montrer les différents niveaux d'abstraction de ce que représente une ontologie. À l'étage 0, nous avons les ressources physiques comme telles. L'étage 1 est le niveau des instances, c'est-à-dire des données représentant les ressources physiques. En ce qui a trait à l'étage 2, il s'agit de la couche où sont assertées les classes et les propriétés, de même que le domaine et le type de ces propriétés (c'est-à-dire les classes desquelles les individus peuvent être sujet de la propriété [domaine] et les classes desquelles les individus peuvent être l'objet de la propriété [type ou *range*]). Et finalement l'étage 3, qui montre les deux modes de relations caractéristiques d'une ontologie (hiérarchique et transversal), est le niveau où l'on définit les classes avec les classes et les relations précédemment assertées : on va asserter par exemple qu'une certaine classe est sous-classe d'une autre, qu'une classe se définit comme tout ce qui entretient une certaine relation spécifique avec une autre classe, et ainsi de suite (nous verrons les détails d'une telle définition des classes dans une section ultérieure).

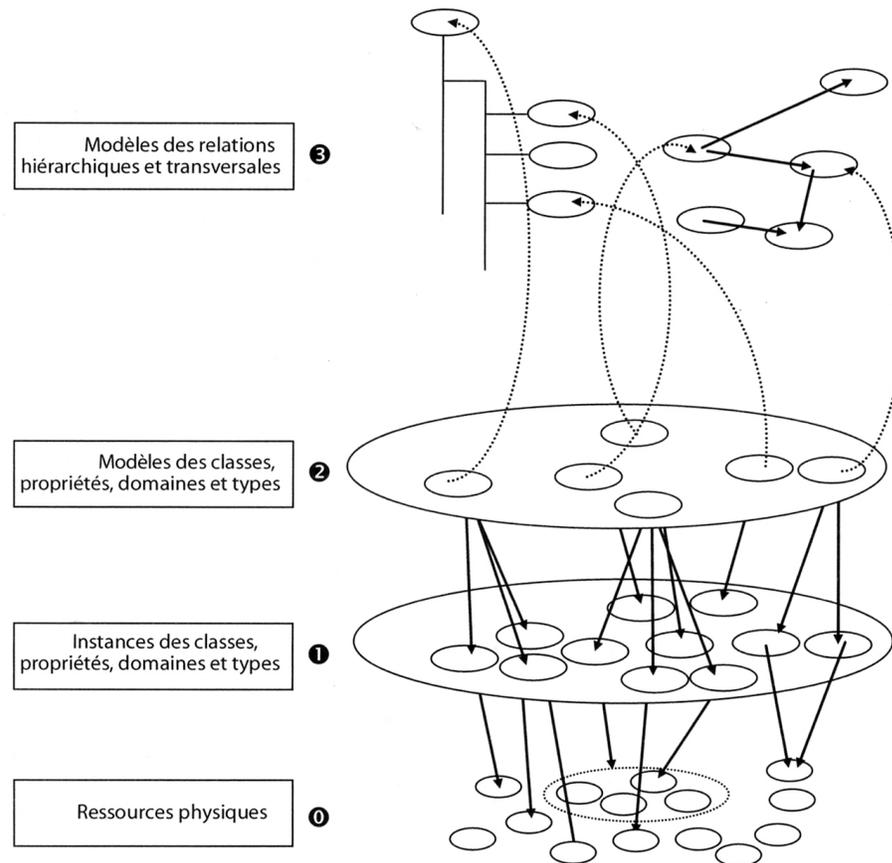


Figure 12 Les différents étages de la modélisation RDF (extrait de Lonjon et Thomasson, 2006, p.343)

2.3.4 Présentation des standards relatifs aux ontologies

Sur un plan technique, le fonctionnement et l'exploitation d'une application à base d'ontologies reposent sur la configuration de protocoles, de données et de règles observant un certain empilage de standards (voir Figure 13).

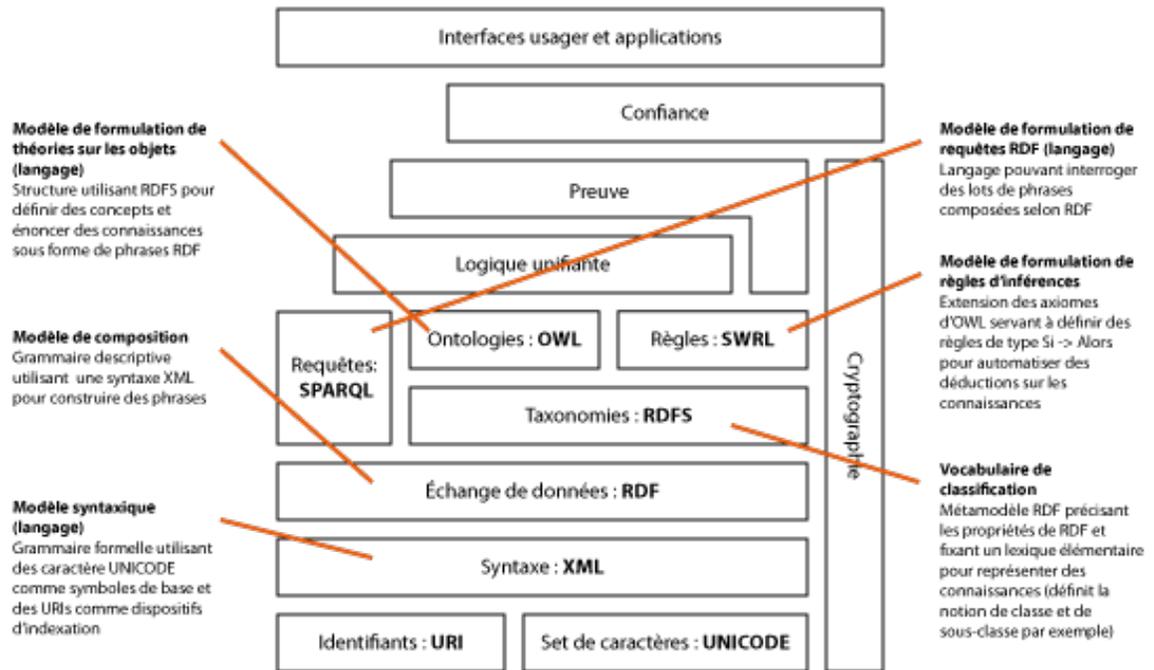


Figure 13 - Empilage des standards relatifs au WS

S'appuyant sur des caractères UNICODE comme symboles de base et sur la notion d'URI comme dispositif d'indexation, on retrouve à la base de cet empilage le langage XML¹² qui sert de modèle syntaxique (grammaire formelle) au reste de l'édifice sémantique. À ce stade, XML n'offre aucune contrainte sur la manière de l'utiliser pour coder de la «signification» : il n'impose que des contraintes syntaxiques sur ce qui peut être considéré ou non comme un document XML bien formé. Vient s'y superposer RDF, un modèle de composition qui permet de construire des phrases simples de type Sujet-Prédicat-Objet. Il est en ce sens une première couche de contraintes sémantiques. Un étage au dessus, l'on retrouve RDFS, le méta-modèle de RDF qui définit un lexique élémentaire pour représenter des connaissances. Il définit notamment les notions de «Classe» et de «Sous-classe».

¹² Tenant pour eXensible Markup Language, XML est un langage qui, tout comme HTML, est formé d'une syntaxe de type « balise » (Tag). Mais à l'instar d'HTML qui se base sur un ensemble de balises prédéfini, XML offre la possibilité de construire des structures de balise sur mesure.

Parallèlement, se fonde sur RDF le langage de formulation de requêtes SPARQL qui permet d'interroger des lots de phrases composées selon RDF. Viennent ensuite les standards OWL et SWRL. OWL est de son côté un langage offrant des contraintes sémantiques issues de la logique descriptive et sert donc en quelque sorte de modèle pour formuler des théories sur les objets, pour construire des ontologies, alors que SWRL, qui est une extension des axiomes d'OWL, constitue finalement un langage de règles permettant de définir et appliquer des inférences de type Si -> Alors sur les connaissances décrites par l'ontologie OWL. S'en suit de cet empilage les couches abstraites «Logique unifiante», «Preuve» et «Confiance» qui viennent marquer les étapes du travail applicatif qui doit être fait sur les données sémantiques pour livrer une expérience utilisateur consistante et cohérente.

2.3.4.1 L'URI : unité d'expression du WS

Dans une perspective communicationnelle, les ontologies (et par extension le WS) apparaissent comme des objets plutôt étranges et difficiles d'abord. Tantôt filtre, tantôt outil de prospection, tantôt constructeur et/ou réorganisateur de contenu, les ontologies sont – théoriquement – capables de tous les exploits en frais de gestion de contenu. Mais comment réussissent-elles à rendre compte d'autant de promesses ? Pour comprendre ce qu'est, techniquement, une ontologie, il faut commencer par saisir sa nature de « système de classification ». Décomposons l'expression. Le Larousse (2009) définit le mot « système » en tant qu'« Appareillage, dispositif formé de divers éléments et assurant une fonction déterminée » et le mot « classification » comme « Action de distribuer par classes, par catégories ». Donc, une ontologie se résume minimalement en un *dispositif formé de divers éléments qui assurent une distribution en classes et/ou catégories*.

La question qui survient d'une telle définition générale, dans la mesure où l'on veut spécifier ce qui distingue une ontologie des autres systèmes de classification, est sans conteste : Qu'est-ce qu'une ontologie distribue en classes et en catégorie? Abstraitemment,

on dira d'une ontologie qu'elle classe des ressources en général. C'est ce que l'on a vu à la section précédente. Mais concrètement, ce qu'elle classe effectivement – informatiquement – ce sont des URI (*Uniform Resource Identifier*), des identificateurs de ressource.

L'URI est une sorte d'étiquette numérique (chaines de caractères) pouvant représenter à la fois l'adresse d'une ressource sur un réseau, lorsqu'il est considéré sous sa forme d'URL (Uniform Resource Locator), et/ou le « nom propre » de cette ressource, lorsqu'il est considéré sous sa forme d'URN (Uniform Resource Name). Une URL est donc un URI qui sert d'identificateur d'emplacement, alors qu'un URN est un URI qui sert à désigner-nommer une ressource quelconque en dehors de toute référence à son emplacement.

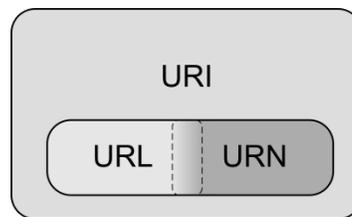


Figure 14 Représentation ensembliste du caractère dual de l'URI (extrait de <http://fr.wikipedia.org/wiki/URI>)

À titre de précision, mentionnons qu'un URI, vu sous son aspect d'URN, peut se décomposer en deux sous-parties : un URIfref (URI de référence) et un QName (nom qualifié). Un URIfref, ou identificateur de l'espace de nom, est un URI avec un fragment identificateur optionnel : le QName, qui identifie un nom particulier dans l'espace de nom identifié par l'URIfref. Il est à noter que l'URN peut s'écrire avec une syntaxe un peu différente de l'URL. Par exemple, une ressource qui utiliserait l'URI `http://www.exemple.org/item1` pour identifiant, on pourrait l'interpeler par l'URN `urn:exemple:item1`.

C'est de cette double fonction représentative (de localisation et de nomination) de l'URI que découle tout le potentiel de traiter sémantiquement les ressources web. En ce sens, elle révèle la base de connaissances à base d'ontologie en tant que dispositif de mappage d'un réseau d'URL, identifiant les ressources localisées à organiser, avec un réseau d'URN, définissant le réseau conceptuel des concepts et relations du domaine ontique auquel appartiennent les ressources à organiser. L'URI tient donc lieu dans ce portrait de pivot conceptuel permettant le mappage en question. C'est ainsi que l'annotation sémantique consiste en la mise en équivalence de deux URI, l'un utilisé pour sa capacité de pointer physiquement une ressource concrète et l'autre utilisé pour sa capacité de pointer un concept abstrait. Bref, ce jeu d'annotation que permettent les URI procède de la même manière qu'un humain apprend de nouveaux termes en langage naturel : il associe l'image visuelle du geste d'un autre humain en train de pointer une chose (une chaise par exemple) avec l'image acoustique de ce même humain en train pointer un concept par l'énoncé « Ceci est une *chaise*. ».

Chose intéressante à noter : cette constitution duale de l'URI en URL/URN n'est pas sans rappeler la structure « double face » (signifiant/signifié) de la notion de « signe » en sémiotique. En ce sens, il serait tout à fait approprié, dans le cadre d'une analyse sémiotique de la manière dont les ontologies travaillent à « faire sens » des ressources web, de prendre l'URI comme unité signifiante de base, unité ayant l'URL pour signifiant et l'URN pour signifié. D'ailleurs Tim Berners-Lee (2005) en présentant l'architecture du WS dit littéralement que les URI doivent être utilisés pour dénoter les choses.

2.3.4.2 Présentation de RDF

De cette capacité élémentaire de l'URI à constituer une sorte d'« unité signifiante » des ressources numériques sur le web on peut commencer à échafauder des structures de signifiante plus complexes qui seront en mesure d'exprimer des faits élémentaires à propos des ressources. La grammaire descriptive que constitue le standard RDF, qui définit des

règles de composition pour construire des phrases simples de type **Sujet-Prédicat-Objet**, représente un premier pas en ce sens. On dira d'une telle phrase RDF de type **S-P-O** qu'elle forme un *triplet RDF*. Le triplet RDF constitue dès lors la représentation la plus élémentaire d'une connaissance. Comme l'ont très bien posé les pères de la philosophie analytique : est une connaissance (une proposition) tout ce qui est susceptible d'être vrai ou faux. Soumettre un mot seul à une telle épreuve de vérification ne fait pas de sens. Par exemple, on ne peut dire d'un mot comme « chapeau » qu'il est vrai ou faux. Pour ce faire, il faut prédiquer le **mot** avec un **attribut** (propriété) et fixer cette dernière avec une **valeur** d'attribut. Ainsi, « le **chapeau** est de **couleur** **rouge** » est une affirmation digne de ce nom. De plus, une propriété intéressante de ce genre de structure conceptuelle triadique est la possibilité de les agréger et les représenter sous forme d'un réseau : les sujets et objets étant représentables par des nœuds que les prédicats peuvent relier en tant qu'arcs. C'est pourquoi dans la terminologie RDF, il y a souvent lieu de parler de *graphe RDF*.

Il est à remarquer, avant de poursuivre, que dans un graphe RDF il est tout à fait possible pour un sujet d'être l'objet d'un autre triplet et/ou d'un objet de devenir le sujet d'un autre triplet (voir Figure 15).

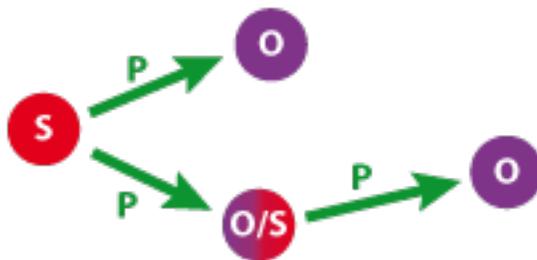


Figure 15 - Vue abstraite d'un graphe RDF

Une manière plus standard de penser le triplet RDF est sous la forme de table : la ligne représente un sujet, la colonne représente une propriété et la cellule la valeur (objet) de la propriété. (Tim Berners-Lee ; voir Figure 16). L'extraction de triplets RDF d'à partir d'une

table, ou à l'inverse la visualisation d'une banque de triplets RDF sous forme de table, fonctionne selon ce principe de conversion.

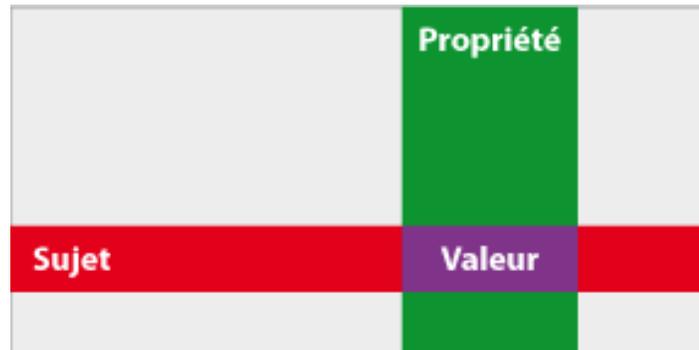


Figure 16 Représentation d'un triplet RDF sous forme de table

Une des principales forces de cette structure par triplets de RDF, par rapport à une base de données classique de type relationnelle, est que la fusion de deux ou plusieurs bases de connaissances RDF peut se faire par la simple agrégation de leurs triplets (pour autant qu'il n'y ait pas deux ou plusieurs nœuds vides ayant un même identificateur).

Pour exemplifier le tout, prenons maintenant un morceau quelconque de connaissance : « le document *page.html* a pour auteur Michel et a pour thème les ontologies ». En RDF, on pourrait exprimer cette description par les deux triplets suivants :

(*page.html*, aPourAuteur, Michel) et (*page.html*, aPourTheme, Ontologie)

Ce qui correspondrait au graphe élémentaire suivant :



Tel que susmentionné, comme RDF a été conçu originalement pour décrire des ressources web, sa syntaxe formelle utilise généralement XML (on appelle alors cette forme

RDF/XML). Mais cette dernière devient vite fastidieuse à lire pour les agents humains. Pour rendre un document RDF plus accessible, le W3C propose des notations plus abstraites que sont N3 et N-Triples (N-Triples étant un sous-ensemble de N3). Grosso modo, N-Triples permet de faire un listage simple de triplets RDF. En se rappelant que la composante signifiante élémentaire est l'URI, et en introduisant que RDF/XML accepte des littéraux¹³ pour objets. Voici ce qu'aurait l'air notre exemple en N-Triples :

```
<http://mon_domaine.ca/page.html> <http://mon_domaine.ca/Vocabulaire#aPourAuteur> "Michel".
<http://mon_domaine.ca/page.html> <http://mon_domaine.ca/Vocabulaire#aPourTheme> "Ontologie".
```

Alors qu'en notation RDF/XML¹⁴ il s'écrirait plutôt :

```
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:local="http://mon_domaine.ca/Vocabulaire#">
  <rdf:Description rdf:about="http://mon_domaine.ca/page.html">
    <local:aPourAuteur>Michel</local:aPourAuteur>
    <local:aPourTheme>Ontologie</local:aPourTheme>
  </rdf:Description>
</rdf:RDF>
```

¹³ Un littéral correspond à une entité exprimée directement par une chaîne de caractères. Au niveau du RDF, les URI peuvent être mises en relation avec des littéraux – ce qui permet d'établir des correspondances entre l'identification formelle d'entités sous forme d'URI et l'identification en langage naturel de ces mêmes entités. Dans notre exemple, les littéraux sont employés dans leurs formes de *littéral simple* (nommée aussi *forme lexicale*), mais auraient pu être typifiés avec un datatype XML (voir annexe 4). Concrètement, le typage d'un littéral consiste en l'association d'une chaîne de caractères et d'un URI représentant le type de la chaîne. Les datatypes XML sont des contraintes définissant une extension d'XML, le standard XML Schema, et permettent d'orienter le décodage des chaînes de caractères. Un exemple de littéral typé aurait été d'ajouter à notre graphe une information quant à la date de création de `page.html`. Pour ce faire on ajouterait une relation spécifiant `dateCreation` dont la valeur serait le littéral typé `"2009-02-04"^^xml:date`.

¹⁴ Des balises XML viennent encadrer la formulation des phrases RDF pour les rendre compréhensible à la machine.

Avant de poursuivre plus loin, il est important de noter que RDF, ainsi que les couches de modélisation qui lui sont supérieures, fonctionnent sur l'hypothèse d'un monde ouvert. Contrairement à une base de données de type relationnelle, l'absence d'une assertion n'est donc pas signifiante. Autrement dit, la déclaration que Michel est l'auteur de page.html n'implique pas que page.html n'a qu'un seul auteur, mais bien plutôt que page.html a au moins un auteur. À l'inverse, si on déclarait que David est aussi auteur de page.html, sans préciser que David et Michel sont deux personnes différentes, la base de connaissances consignant ces assertions ne pourrait pas déduire que page.html a deux auteurs puisque Michel et David pourraient être deux noms d'une même personne.

Dans cette forme, RDF n'impose que très peu de contraintes à la description des ressources. En ce sens, bien que très flexibles (possibilité d'agréger n'importe quoi, n'importe comment à la description d'une ressource), les descriptions dans ce format sont relativement platoniques et sans grande profondeur sémantique (faible degré de hiérarchisation des descriptions). Pour pallier cette lacune (qui en devient une que dans l'optique du besoin d'une plus grande expressivité), un vocabulaire élémentaire fut développé : RDFS (RDF Vocabulary Language). Néanmoins, le modèle RDF offre quelques ressources terminologiques élémentaires pour donner un peu de relief sémantique à la description des ressources. Une des plus importantes est la relation de typification, `rdf:type` (<http://www.w3.org/1999/02/22-rdf-syntax-ns#type>), qui permet l'identification du type d'une ressource. Par exemple on pourra exprimer que :



Aussi, RDF permet de définir des *conteneurs* et des *collections* pour lister de différentes manières des items et de réifier des triplets pour pouvoir faire des assertions à leurs propos (voir Annexe 1).

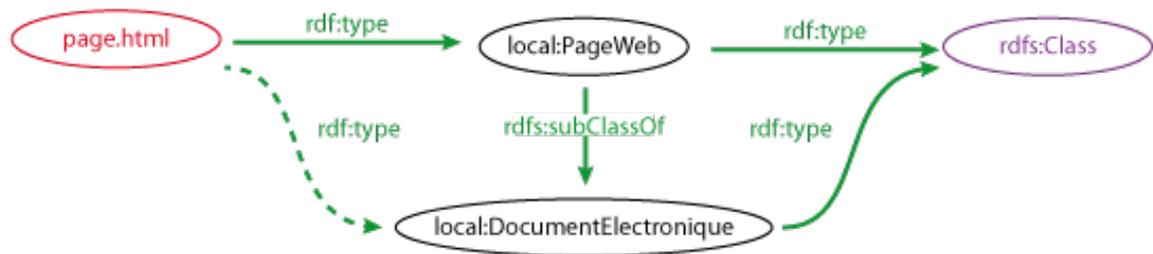
2.3.4.3 Présentation de RDFS

Tel que dit un peu plus tôt, RDFS consiste en un vocabulaire de base à partir duquel on pourra construire ultérieurement des ontologies. L'espace de nom de ce vocabulaire est `http://www.w3.org/2000/01/rdf-schema#` et son préfixe usuel est « `rdfs:` ». Les trois termes les plus saillants de ce vocabulaire sont sans conteste ceux de *classe* (`rdfs:Class`), de *sous-classe* (`rdfs:subClassOf`) et de *sous-propriété* (`rdfs:subPropertyOf`).

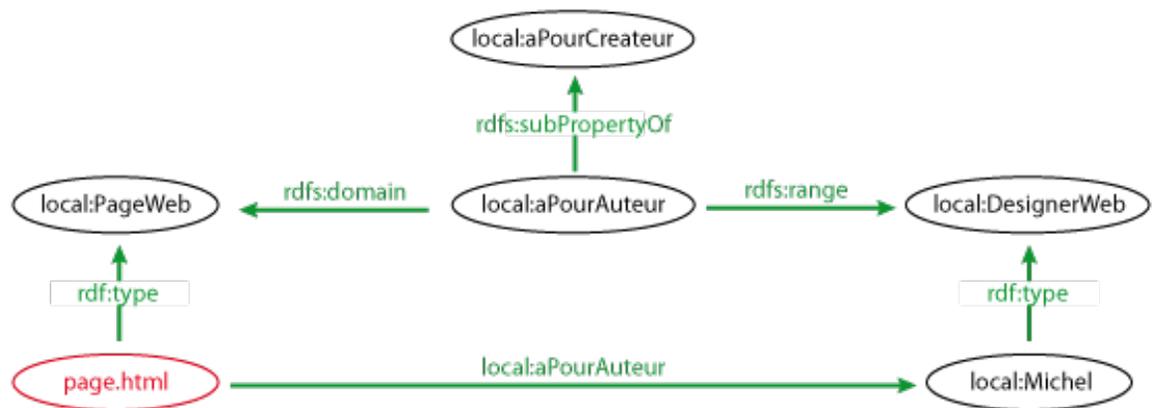
En ce qui concerne l'élément `rdfs:Class`, il s'agit d'un terme pour identifier le type des entités qui sont l'objet de la propriété `rdf:type`. Reprenons l'exemple où l'on avait déterminé que « `page.html` » était un de type « `PageWeb` ». Mais que représente exactement l'entité « `PageWeb` » ? Si on affirmait que « `page.aspx` » est aussi de type « `PageWeb` », on viendrait en quelque sorte de définir implicitement que ces deux ressources ont en commun d'être de type « `PageWeb` ». En ce sens, « `PageWeb` » n'est pas une entité simple mais une classe d'entités. On peut exprimer ce fait en déclarant que son type est `rdfs:Class`.



Subséquentement, et c'est là que cela devient intéressant, RDFS propose une propriété pour attester qu'une classe est sous-classe d'une autre : `rdfs:subClassOf`. Cette dernière propriété étant transitive, elle permet donc de faire quelques déductions : si l'on définissait que « `PageWeb` » est une sous-classe de « `DocumentÉlectronique` », on pourrait alors déduire que « `page.html` » est un document électronique.



Tout comme `rdfs:subClassOf` permet d'établir des hiérarchies de classes, la propriété `rdfs:subPropertyOf` permet d'élaborer des hiérarchies de propriétés. Mais pour ce faire il faut d'abord expliciter le caractère RDF de « ressource » de la propriété avec les propriétés RDFS `rdfs:domain` et `rdfs:range` : la première indique le type de sujet sur lequel peut porter la propriété alors que la seconde pointe vers le type d'objet auquel s'applique la propriété. Ceci permet alors de pouvoir prendre le nœud représentant la propriété réifiée pour sujet d'une propriété `rdfs:subPropertyOf` qui pointerait vers une super-propriété qui lui est plus générale. Ainsi, par rapport à notre exemple, on pourrait définir la propriété « aPourAuteur » comme sous-propriété de « aPourCreateur ».



2.3.4.4 Présentation de OWL

OWL, qui est l'acronyme de *Web Ontology Language*, désigne un vocabulaire dédié à l'assertion de connaissances sous forme d'ontologie. Pour ce faire, ce standard a été conçu

pour pouvoir être exprimable dans une syntaxe XML et reprendre des concepts et propriétés déjà définis dans RDF et RDFS. En ce sens, une base de connaissances OWL peut donc être écrite comme une collection de triplets RDF.

Un document OWL, qui décrit toujours une ontologie, se compose de *faits* et d'*axiomes*. Les faits assertent des choses sur les individus alors que les axiomes assertent des choses sur les concepts utilisés par l'ontologie.

Étant donné qu'une des motivations premières de construire une ontologie OWL est de permettre aux applications qui l'utilisent de pouvoir effectuer des inférences sur les données cumulées, OWL offre des opérateurs issus de la logique descriptive¹⁵ qui permettent de préciser et d'axiomatiser les descriptions en *classe*, *propriété* et *individu* fournis par RDF et RDFS. Il définit en l'occurrence des propriétés de propriété telles que `owl:TransitiveProperty` qui marque le caractère transitif d'une propriété, `owl:SymetricProperty` qui marque le caractère symétrique d'une propriété, `owl:FunctionalProperty` qui marque que la propriété ne peut avoir qu'une seule valeur pour un même sujet, `owl:InverseFunctionalProperty` qui marque que pour une même valeur la propriété ne peut avoir qu'un seul sujet, `owl:inverseOf` qui marque que la propriété est l'inverse d'une autre, et `owl:equivalentProperty` qui marque que la propriété est équivalente à une autre. OWL définit aussi des constructeurs de classe qui servent à décrire des classes, non seulement en fonction des relations de leurs relations taxonomiques (ce que permettait `rdfs:subClassOf`), mais aussi en fonction des relations transversales qu'ils entretiennent entre eux. En ce sens, OWL prédéfinit ainsi les classes *chose* (`owl:Thing`) et *rien* (`owl:Nothing`) qui représente respectivement l'ensemble de tous les individus et l'ensemble vide. Il définit la relation `owl:oneOf` qui permet de décrire une classe par l'énumération exhaustive des individus qu'elle contient, le

¹⁵ La logique descriptive (DL) est une famille de formalismes dédiés à représenter des connaissances déclaratives (ou connaissances terminologiques). Ces formalismes ont pour caractéristique sémantique primordiale d'être définis en logique classique du premier ordre.

terme `owl:restriction` qui permet de décrire une classe par rapport à une certaine propriété que doivent avoir les individus qu'elle contient, la relation `owl:equivalentClass` qui permet de déclarer deux classes comme étant équivalentes, la relation `owl:intersectionOf`, qui permet de définir une classe en tant qu'elle est coextensive de l'intersection de deux ou plusieurs autres classes, la relation `owl:unionOf` qui permet de définir une classe en tant qu'elle est coextensive de l'union de deux ou plusieurs autres classes, la relation `owl:complementOf` qui permet de définir une classe par sa complémentarité avec une autre, et finalement la relation `owl:disjointWith` qui permet de déclarer que l'extension d'une classe n'a aucun membre en commun avec l'extension d'une autre classe.

Mentionnons finalement que, vu le mot d'ordre du WS quant à la réutilisabilité des modèles de données, une ontologie OWL autorise l'importation d'autres ontologies OWL pour l'aider à définir ces concepts locaux via son terme `owl:imports`. Il rend ainsi possible la construction d'ontologie agrégeant des termes de vocabulaires déjà bien définis par d'autres ontologies. Il s'agit là de l'une des dix grandes exigences à l'origine de OWL et très certainement l'une de ses plus évidentes attestations de son alignement sur une philosophie de l'interopérabilité.

2.3.4.5 Présentation de SPARQL

Protocole et langage d'interrogation pour une base de données de type RDF, SPARQL est une recommandation du W3C depuis le 15 janvier 2008. Son nom est un acronyme récursif désignant *SPARQL Protocol And RDF Query Language*.

L'exigence primordiale de SPARQL était de définir un protocole basé sur HTTP et/ou SOAP pour sélectionner des instances de sous-graphe d'un graphe RDF. Ce nouveau langage, qui est inspiré et entretient des similarités techniques avec les technologies de requête XML, imite la syntaxe et les fonctionnalités du langage SQL (langage de requête

classique pour les bases de données traditionnelles de type relationnelles) pour faciliter son apprentissage. Mais à l'instar de SQL qui ne permet de faire que des requêtes sélectives (c.-à-d. de filtrer une banque de données en vue de garder seulement certaines informations spécifiques), SPARQL va plus loin et permet aussi d'effectuer des requêtes dites constructives, c'est-à-dire des requêtes visant à générer de l'information nouvelle, implicitement contenue dans la mise en relation de plusieurs banques RDF. Par exemple, quelqu'un pourrait vouloir afficher sur son site web un palmarès des bandes son les plus vendues parmi celles des 40 meilleurs films au box office. Pour ce faire, une première requête serait envoyée pour récupérer l'information à propos du top 40 des films au box office sur les serveurs d'IMDb (*The Internet Movie Database*). Les informations obtenues pourraient ensuite être réutilisées pour envoyer une seconde requête sur les serveurs d'Amazon et vérifier l'état des ventes de bandes son pour les 40 films retenus.

En somme, SPARQL est un langage de requête qui permet au WS de véritablement prendre de l'envergure puisqu'il sert à faire le pont entre les technologies du WS, dont RDF, et les plateformes web déjà existantes. Avec SPARQL le WS commence donc très tangiblement à devenir une base de données globale.

2.3.5 Techniques de modélisation d'ontologies

Karin K. Breitman, Marco Antonio Casanova et Walter Trczkowski, dans leur ouvrage *Semantic Web : Concepts, Technologies and Applications* (2007), dressent une liste des principales méthodologies de modélisation en vigueur. Suivant l'avis de John F. Sowa, les auteurs affirment que l'ingénierie des connaissances, comme toutes les autres formes d'ingénierie, doit entretenir des aspirations théoriques modestes en ce sens où l'élaboration d'une solution technique à un problème ne peut jamais relever de la pure application théorique et se voit toujours dictée par le lot de contraintes que constitue le problème lui-même. À ce propos, ils conseillent donc que, bien qu'il existe des méthodes « prêtes à

l'emploi » pour modéliser une ontologie, il est de mise de se forger – en s'inspirant de ces dernières – une méthode personnalisée en fonction des exigences du problème à résoudre.

De la liste des diverses méthodes que proposent Breitman et coll. (voir le détail de cette liste en Annexe 2) il est possible de voir émerger un pattern général en ce que chaque méthode de modélisation se découpe toujours plus ou moins en quatre grandes phases, chacune comprenant ces activités de modélisation typiques (ce découpage est inspiré de Héon et coll, 2008) :

- 1) Avant d'amorcer la modélisation comme telle il est généralement recommandé de déterminer la portée de l'ontologie en ce sens où autant l'identification des problèmes auxquels devra répondre l'ontologie seront clairement énoncés et délimités, autant il sera plus facile par la suite d'éviter l'égarement dans une modélisation trop large ou trop granulaire de l'ontologie. Une fois cette phase de définition du mandat de l'ontologie terminée, le projet de modélisation peut alors débiter.
- 2) Le projet de modélisation commence normalement par une phase d'élicitation des connaissances, c'est-à-dire par des activités de terrain cherchant à dévoiler/découvrir les concepts (et leurs ramifications) utilisés pour traiter du domaine d'activité ciblé.
- 3) S'en suit une phase de formalisation des connaissances élicitées, phase pendant laquelle l'ontologie est progressivement construite par l'application de règles de traduction sur les connaissances semi-formelles précédemment récupérées.
- 4) Et finalement, une phase de validation syntaxique et sémantique de l'ontologie vient assurer que l'ontologie ne comporte aucune inconsistance logique (validité syntaxique), d'une part, et que les définitions de concept qu'elle formalise font du sens pour les experts du domaine modélisé (validité sémantique).

2.4 Gestion des connaissances et interopérabilité dans les organisations : bénéfices stratégiques des ontologies

Jusqu'à maintenant dans ce chapitre nous avons vu ce qu'était concrètement une ontologie (dans son sens informatique), comment procéder pour en modéliser une, et quels enjeux

psychotechniques ont contextuellement guidés son évolution. Dans les deux prochaines sections nous allons présenter les attentes et perspectives que certains domaines de connaissances en communication portent sur l'ontologie comme objet sociotechnique. Cette section-ci introduira le corpus théorique qu'est la gestion des connaissances pour montrer que l'enjeu d'interopérabilité motivant l'usage d'ontologies en fait un outil stratégique de choix dans l'effort de valorisation des capitaux informationnels des organisations. La prochaine exposera les impacts et problématiques que font apparaître les technologies du WS à l'aune des champs d'études que sont la communication médiée par ordinateur (CMO) et les interactions humain-ordinateur (IHO).

2.4.1 Introduction à la gestion des connaissances

L'information est la matière brute des organisations contemporaines. Qu'elle se présente sous forme de compétences métier, de procédures, de technologies, de brevets, ou de ressources documentaires en tout genre, elle constitue, lorsque configurée et approvisionnée adéquatement, un puissant vecteur de création de richesse. Mais d'ordinaire, les informations utiles au maintien des opérations d'une organisation étant habituellement consignées de manière fragmentaire (en silo) par les différentes agences particulières composant l'organisation (employés, consultants, départements, etc.), leurs pérennités se voient constamment fragilisées, voir parfois mises en péril, par une foule d'évènements autant normaux que fortuits (démissions, mutations, maladies, ventes d'actifs, etc.) Tel est un des constats saillants du rameau théorique qu'est la *gestion des connaissances*. Ce corpus de théories transdisciplinaires, qui émergea au début des années 1990, considère par conséquent l'importance stratégique d'une bonne structuration du capital informationnel à l'intérieur des organisations.

Les pionniers de ce corpus que sont Ikujiro Nonaka (1991, 1995), Hirotaka Takeuchi (1995), Thomas H. Davenport (1992, 1997) et Baruch Lev (2001) tendent à distinguer trois niveaux ontiques des informations :

- Les données, en tant qu'elles représentent des états de fait quantitatifs et/ou catégoriques.
- Les informations comme telles, en tant qu'elles sont des agrégats de données dont les règles d'assemblage impliquent des consensus interhumains quant à leurs construction et interprétation correctes.
- Et les connaissances (savoirs), en tant qu'elles sont des informations valorisées en regard d'expertises particulières.

Dans la pratique, il est souvent difficile de dire avec exactitude à quel type appartient une information (au sens générique) – les trois types servant au final à grader le continuum d'intrication des données entre elles d'une part (données → informations) et le continuum d'intrication des informations avec les processus cognitifs les manipulant d'autre part (informations → connaissances). En ce sens, dans la perspective de la gestion des connaissances, la notion d'*information* se rapporte au travail *mécanique* d'organisation et d'interprétation des données brutes tandis que celle de *connaissance* porte plutôt sur le travail *cognitif* d'évaluation et d'appréciation des informations à l'aune d'expériences, de croyances, de théories ou de modèles particuliers (Prax, 2003). En dernière instance ce sont donc les connaissances qui supportent les prises de décision et motivent les innovations au sein d'une entreprise.

Les thèses sur la gestion des connaissances voient donc dans l'architecture spécifique des infrastructures sociotechniques générant et supportant la création et le partage de connaissance le lieu précis d'où peuvent naître les apprentissages collectifs et les intelligences collectives. La problématique de la gestion des connaissances se résume alors à la question de l'optimisation des architectures d'information dans le sens d'offrir des structures dédiées à la consignation, la création, la gestion et le partage des connaissances d'une part, et des méthodes d'accès et d'exploitation (capitalisation) de ces connaissances d'autre part. En d'autres termes, la gestion des connaissances cherche à comprendre comment les organisations assurent la pérennité et l'opérabilité des informations-connaissances qu'elles cumulent. Pour ce faire les théories de la gestion des connaissances mobilisent des concepts et principes issus de discipline telle que les sciences de l'organisation, les sciences cognitives, les sciences sociales, les sciences de l'information et

de la communication et l'informatique. Deux d'entre eux sont les concepts de *connaissance tacite* et de *connaissance explicite*.

2.4.2 Connaissances tacites versus connaissances explicites

Une des typologies classiques en gestion des connaissances est celle de la distinction entre connaissances tacites et connaissances explicites. Ces deux catégories de connaissances, généralement considérées comme les pôles complémentaires d'un même continuum des formes de connaissances par les théoriciens en gestion des connaissances (Hildreth et Kimble, 2002), servent habituellement à mettre en évidence et discuter des modes de transmission et de transformation des connaissances.

À un bout du spectre, il y a donc les connaissances tacites : il s'agit des connaissances consignées de manière informelle dans la tête des individus, sous forme de représentations mentales (Grundstein, 2004). Les connaissances tacites portent généralement sur des savoir-faire et des vécus individuels (compétences, souvenirs, trucs métier, etc.) et s'acquièrent par la pratique et l'expérience sub-symbolique des choses. Par conséquent, le stockage de telles connaissances est normalement assuré par les structures cognitives subconscientes des individus. Tel que défini originalement par Michael Polanyi (1958), le concept de connaissances tacites désigne donc concrètement tous processus cognitifs et/ou schémas de comportements non-conscients que les individus mobilisent – généralement de manière implicite – pour effectuer des actions ou des jugements dits « experts ».

À l'autre bout du spectre, il y a les connaissances explicites qui, elles, consistent plutôt en des savoirs formels objectivés (Grundstein, 2004), c'est-à-dire en des savoirs articulables textuellement sur des supports physiques extérieurs aux cerveaux des individus (manuels d'instructions, formules algébriques, rapports d'analyse, etc.). En ce sens, les connaissances explicites sont des connaissances plus facilement saisissables et manipulables par les structures cognitives conscientes des sujets connaissant que les connaissances tacites et leur

faculté à pouvoir être externalisées sous forme d'objets physiques persistants les rend beaucoup plus faciles à communiquer.

Au niveau de l'organisationnel, supposons que l'on considère le substrat des organisations dans sa dimension essentielle de *collection d'actions plus ou moins routinières opérées par des agents cognitifs*, on peut voir que la très grande part des connaissances effectives à l'intérieur des organisations (traditionnelles surtout) sont foncièrement de nature tacite. Cependant, l'aspect très subjectif de ce type de connaissances a tendance à se faire plutôt contraignant lorsque vient le temps d'en faire la transmission et de les faire passer d'une tête à une autre : leur caractère informel et non-conscient les empêche d'être externalisées avec clarté et concision. De par leur nature résolument plus tangible que les connaissances tacites, les connaissances explicites constituent en revanche la forme de connaissance la plus facile à partager collectivement. Il apparaît donc clair que l'une des stratégies privilégiées de la gestion des connaissances pour assurer la pérennité et la valorisation du capital informationnel des organisations est d'encourager la transformation des connaissances tacites en connaissances explicites de sorte à les rendre accessibles au maximum de gens pour qui ces connaissances pourraient être utiles. La problématique se résumant alors à comment arriver à codifier convenablement les connaissances tacites.

La gestion des connaissances distingue en fait quatre grands types de transformation-transmission des connaissances (Nonaka, 1991) : de connaissances tacites à connaissances tacites (transmission par socialisation ; formule maître-apprentis), de connaissances tacites à connaissances explicites (transmission par extériorisation ; diffusion en masse des savoir-faire), de connaissances explicites à connaissances explicites (transformation par combinaison, spécification, synthèse ; réutilisation et réagencement des savoirs existants), et de connaissances explicites à connaissances tacites (transformation par intériorisation ; appropriation cognitive individuelle des savoirs explicités collectivement). Selon Nonaka, ces quatre transformations s'enchaînent pour former une espèce de « spirale du savoir » où chaque cycle, en plus d'augmenter le niveau de savoir général de l'organisation, constitue des boucles de réflexivité propres à engendrer des apprentissages collectifs. Et de cycle en

cycle on pourra alors parler de ce processus en tant qu'il est substantif d'une intelligence collective.

2.4.3 Révolution du numérique dans le cycle de transmission des savoirs

Dans tout ce cycle de transmission des savoirs, c'est vraisemblablement la transformation des connaissances tacites en connaissances explicites qui est l'étape la plus critique – en regard du développement d'une intelligence collective – puisque c'est elle qui assure la *mise en commun* des savoirs. Même lorsqu'on est en présence d'un transfert de connaissances tacites à connaissances tacites il y a toujours une certaine forme d'externalisation : à savoir les gestes et paroles perceptibles du maître pour *montrer* quelque chose à son apprenti. La venue des nouvelles technologies de l'information et de la communication (NTIC) a ouvert ces dernières années tout un registre de possibilités pour supporter ce type de transformation (tacite → explicite). Mais c'est tout particulièrement l'avènement de l'informatique en réseau et du multimédia qui ont fait exploser ce registre en permettant d'explicitier et de distribuer pratiquement n'importe quel type de connaissances, à très grande échelle et à un coût plus modique que jamais. La banalisation de l'expression par des formes médiatiques « non-scripturales » comme la vidéo et la photographie, qui était jusqu'à il n'y a pas très longtemps encore l'apanage d'individus spécialement qualifiés, rend de plus en plus abordable la consignation de connaissances tacites sous forme « explicite » (c'est-à-dire sous une forme numérique et dans un format standardisé de telle sorte que les documents photo ou vidéo produits sont disséminables aussi facilement que le sont les mots depuis longtemps). La réalisation d'un tutoriel vidéo en ligne par exemple permet d'explicitier certaines connaissances tacites avec beaucoup plus d'aisance et d'économie que par la réalisation d'un document utilisant un mode strictement scriptural. Imaginez que vous deviez rédiger un article expliquant en mots seulement comment exécuter un bon swing de golf... Et l'internalisation a toutes les chances d'être pénible !

Morlen T. Hansen et coll. (2003) ont distingué, à partir de leurs études de cas, deux grands types de stratégie de gestion des connaissances : la *stratégie de codification* et la *stratégie de personnalisation*. Selon les auteurs, la stratégie de codification est celle qui insiste sur la production des connaissances explicites (tacites → explicites et explicites → explicites) par l'installation de systèmes d'information sophistiqués favorisant le stockage (codification) et la réutilisation des connaissances alors que la stratégie de personnalisation encourage plutôt l'échange et l'acquisition de connaissances tacites (explicites → tacites et tacites → tacites) par installation de moyens de communication interpersonnelle favorisant le partage d'expériences et l'interprétation des savoirs codifiés entre les différents membres de l'organisation.

Les auteurs mettent cependant en garde contre le désir d'adopter simultanément les deux stratégies puisque le succès de chacune dépend de la visée stratégique globale de l'organisation, mais aussi et surtout de la nature de ce qu'elle vend comme bien ou service. S'il s'agit d'une entreprise offrant des biens et services standardisés, que ceux-ci sont matures et que les connaissances pour résoudre les problèmes de leur production sont majoritairement explicites, la stratégie de codification est à prioriser. En revanche, s'il s'agit d'une entreprise offrant des biens et services novateurs, plutôt de type « sur mesure », et que les connaissances nécessaires à la résolution des problèmes de production sont principalement de nature tacite, la stratégie de personnalisation sera plus appropriée. Hansen et coll. déconseillent la poursuite simultanée des deux types de stratégies : cela s'avère trop souvent contre productif. Cet état des choses asserté par Hansen et coll. laisse à penser qu'une gestion rigoureuse des connaissances explicites a pour effet de faciliter la mobilisation des connaissances tacites nécessaire à la réalisation pratique des opérations courantes de l'entreprise et, qu'à l'inverse, une gestion rigoureuse des moyens de développement et de transmission des connaissances tacites a pour effet de faciliter la mobilisation des connaissances explicites nécessaires à la gouvernance et l'évolution intelligente des opérations de l'entreprise. En cherchant à contrôler rigoureusement les deux

pôles à la fois, on fait s'interférer les deux stratégies de sorte qu'elles deviennent difficilement gérables – en plus d'exiger des investissements plus importants à la base.

Néanmoins, avec l'entrée en scène des outils de type Web Social (forums de discussion, blogues, wikis, fils de nouvelles, etc.) qui permettent aux usagers de dialoguer tout en consignnant de manière plus ou moins formelle les fruits de ces dialogues, la conciliation des deux stratégies se voit de moins en moins antagonique et problématique. La notion d'« entreprise 2.0 » atteste depuis 2006¹⁶ de cette réalité émergente de l'usage du Web Social par les entreprises pour devenir plus organiques et collaboratives.

2.4.4 L'interopérabilité

L'*interopérabilité*, en ce que ce concept dénote l'étendue des compatibilités normatives entre plusieurs systèmes ou organisations qui leur permettent d'agir ensemble, d'« interopérer », constitue en gestion des connaissances un enjeu important de la structuration des dispositifs de stockage et de partage des connaissances. En effet, la normalisation des formats et standards pour documenter, modéliser et partager des informations-connaissances permet un décloisonnement des structures traditionnelles, dites « en silo », des dispositifs sociotechniques d'accumulation et d'exploitation des informations-connaissances en ce que les dispositifs localisés et distribués ne nécessitent plus de partager un même environnement – souvent limité à une agence spécifique – pour communiquer entre eux, accomplir des choses communes et partager des ressources.

Il faut distinguer ici l'interopérabilité de la compatibilité (Interopérabilité, 2011). La notion de compatibilité désigne une relation binaire entre deux ou plusieurs dispositifs spécifiques afin de marquer si oui, ou non, leur observation respective des normes en vigueur dans leur

¹⁶ Année où le terme fut proposé par Andrew McAfee du Harvard Business School dans un article publié sous l'édition printanière du MIT Sloan Management Review (2006).

environnement commun leur permet, ou non, de fonctionner ensemble et/ou d'échanger de l'information, alors que celle d'interopérabilité est plus générale et exprime plutôt le niveau global de compatibilité entre les éléments-dispositifs d'un ensemble considéré – en vertu du fait que les éléments-dispositifs en question respectent plus ou moins certaines normes externes (que les manières dont ils s'interfaçent sont plus ou moins standardisées). Lorsque la notion d'interopérabilité est appliquée à un dispositif spécifique, c'est normalement pour qualifier le potentiel d'interopérabilité qu'il projette sur le plexus systémique auquel il est abouché.

Il est à distinguer aussi l'interopérabilité syntaxique de l'interopérabilité sémantique (Interopérabilité, 2011). À la base, l'interopérabilité syntaxique adresse le problème de la communicabilité des données et des systèmes sous l'angle de la compatibilité des formes d'encodage. ASCII (*American Standard Code for Information Interchange*), qui est la norme la plus répandue pour coder en binaire les caractères alphanumériques, exemplifie parfaitement la notion d'interopérabilité syntaxique est ce qu'elle rend possible la copie de texte d'un système à un autre. À un niveau syntaxique plus élevé, XML, qui est un langage de balisage générique, constitue aussi un standard syntaxique favorisant une interopérabilité « mécanique » des systèmes électroniques de communication. Mais au-delà de la simple capacité technique de communiquer (échange de signaux + compréhension réciproque des normes d'encodage et de décodage), les systèmes ont besoin de référents communs – définis par des modèles sans ambiguïtés – pour pouvoir effectuer automatiquement des traitements significatifs avec les informations communiquées. Lorsqu'un tel besoin est satisfait, on dira alors des systèmes partageant de tels référentiels qu'ils font preuve d'interopérabilité sémantique. Une ontologie constitue justement un type de référentiel pouvant mener à de l'interopérabilité sémantique, non seulement entre les systèmes s'y référant, mais aussi entre les agents humains qui l'utilisent (puisque la terminologie modélisée par l'ontologie formalise généralement celles utilisées par les acteurs du domaine d'application).

En somme, cette distinction entre interopérabilité syntaxique et interopérabilité sémantique nous est ici importante dans la mesure où elle nous permet de comprendre les visées particulières, en matière d'interopérabilité, fondant les projets de développement de type Web sémantique, de même que la manière dont ces visées s'échafaudent sur celles, plus syntaxique du WWW. En effet, l'invention des standards http et HTML à la base du WWW fut une avancée formidable en frais d'interopérabilité syntaxique pour ce qui est de l'échange et du partage de documents sur Internet. Les ontologies et le Web sémantique voulant doubler cette facilité d'échange d'une facilité d'intercompréhension des informations contenues par ces documents, ils apparaissent donc aux yeux de la gestion des connaissances en tant que nouveaux vecteurs de force notables dans le développement et l'entretien des cycles de transmission et de transformation des savoirs.

2.5 Perspectives CMO et IHO des ontologies et du WS

Si d'un point de vue organisationnel les ontologies et le WS sont perçus comme des outils pleins de promesses en matière de gestion des connaissances, d'un point de vue interactionnel ils ne sont pas sans soulever une foule d'enjeux et problématiques quant à la manière dont ils doivent s'intégrer dans les pratiques communicationnelles et donc ultimement à la manière dont ils devront s'exprimer dans les interfaces. Cette section introduira alors, pour faire la lumière sur ces enjeux, les champs d'études que sont la CMO et les IHO.

2.5.1 Introduction à la CMO

Partant du constat que les technologies de l'information, et les dispositifs électroniques en général, ont pour but global de supporter la *co-ordination* des activités humaines par le traitement automatique des informations (c'est-à-dire la structuration ordonnée d'activités distinctes en vue de les harmoniser), et que toute coordination implique toujours

préalablement, ou simultanément, une forme ou une autre de communication (Kleinbaum, Stuart & Tushman, 2008), on peut donc voir l'ensemble des technologies de l'information en tant qu'il constitue un médium en propre. L'étude des communications utilisant un tel « support » dit électronique constitue dès lors la préoccupation de base de la *Communication Médinée par Ordinateur* (CMO). Bien que ce champ d'études ne considérait initialement que les transactions communicationnelles réalisées à partir d'ordinateurs – d'où le nom –, la diversification contemporaine des dispositifs électroniques de communication a tôt fait d'élargir cette considération.

Pour sujet de focalisation principal, la CMO s'intéresse avant tout aux conditions et aux effets psychosociologiques inhérents à l'usage des dispositifs électroniques de communication, afin, notamment, de mettre en lumière en quoi cette forme de communication possède des caractéristiques communes et/ou inédites par rapport aux formes plus traditionnelles de communication. L'engagement des interlocuteurs, la dynamique des communautés (virtuelles), la désinhibition des participants et la formation des relations sociales sont parmi les dimensions psychosociales où les études en CMO remarquent généralement les transformations les plus significatives.

À la base, la CMO distingue généralement deux modes fondamentaux de communication médiée par ordinateur : le mode synchrone, lorsque tous les interlocuteurs investissent en même temps un même espace virtuel pour communiquer ; et le mode asynchrone, lorsque la communication se déroule en dehors de telles contraintes d'espace – commun –, mais surtout de temps.

Qu'importe le mode, une communication médiée par ordinateur implique toujours une forme ou une autre d'interaction des interlocuteurs avec le médium pour communiquer via celui-ci. Et lorsque le focus de recherche s'attarde spécifiquement à observer ces formes d'interaction, on dira alors de cette observation qu'elle relève du domaine des *Interactions Humain-Ordinateur*.

2.5.2 Introduction à l'IHO et à la conception de systèmes centrée sur l'utilisateur

En ce qui concerne l'*Interaction Humain-Ordinateur* (IHO), appelé encore souvent Interaction Humain-Machine (IHM) ou Interface Homme-Machine (IHM), il s'agit d'un champ d'études qui s'attarde, comme on vient de dire, à comprendre les phénomènes dits interactifs qui prennent place à la frontière entre les dispositifs électroniques et leurs utilisateurs humains. L'objet central de ce domaine s'exprime alors à travers le concept d'*interface usager*. La notion d'interface dénotant « une limite commune entre deux systèmes, permettant des échanges entre ceux-ci » (Larousse), c'est ainsi que l'IHO enrégimente des théories et perspectives issues de disciplines telles que les sciences informatiques, les sciences cognitives, la sociologie, l'anthropologie, la bibliothéconomie et le design, pour expliquer l'interaction humain-ordinateur en tant qu'elle résulte de la confrontation des caractéristiques psychosociologiques des utilisateurs d'une part aux caractéristiques computationnelles et représentationnelles des dispositifs électroniques de l'autre. D'une certaine manière, il faut voir que l'IHO, bien qu'elle soit une composante essentielle de toute CMO, n'implique pas nécessairement de fins communicationnelles comme telles : elle porte avant tout sur le régime de contrôle d'un usager opérant une machine que sur le résultat, communicationnel ou autre, de l'opération (Jensen, 1999).

De plus, Douglas C. Engelbart, un des pionniers du domaine (et de l'informatique en général), nota que la notion d'interface humain-machine, bien que relativement récente, exprime une réalité qui existe depuis que l'humain utilise des outils (Engelbart, 1962). Pour lui, l'interface humain-ordinateur constitue une forme particulière des interfaces humain-artéfact en général qui, elles, adviennent lorsque des processus humains explicites sont couplés à des processus artéfactuels explicites en vue d'exécuter des processus composites, c'est-à-dire des cascades d'actions issues de la participation conjointe de plusieurs agents. L'interface humain-artéfact constitue dès lors un lieu d'échange d'énergie (principalement

sous forme d'informations dans le cas de machines complexes) qui permet d'instaurer une interaction coopérative entre l'humain et son outil.

De manière globale, l'IHO se préoccupe donc de minimiser les disparités et conflits entre les modèles cognitifs des usagers humains relativement à l'exécution de certaines tâches et les structures d'information des systèmes guidant la compréhension que les ordinateurs ont de l'intentionnalité des usagers vis-à-vis ces mêmes tâches. Pour ce faire l'IHO déploie des objectifs spécifiques tels que (« Human-computer interaction », 2010) :

- proposer et affiner des méthodologies et processus pour designer des interfaces ;
- rester à l'affût des méthodes d'implémentation (toolkits logiciels, bibliothèques, algorithmes, etc.) ;
- proposer et affiner des techniques pour évaluer et comparer les interfaces ;
- développer de nouvelles formes d'interfaces et/ou d'interactions techniques ;
- développer des théories et modèles, descriptifs et prédictifs, de l'interaction ;

Au-delà de sa visée heuristique élémentaire qui cherche à rendre compte et à comprendre les phénomènes interactifs entre humains et ordinateurs (volet recherche du domaine), l'IHO se voit aussi souvent imputable de l'évaluation et/ou de l'amélioration du caractère « utilisable » des dispositifs électroniques (volet pratique du domaine). Sont comprises dans cette notion d'*utilisabilité* celles d'ergonomie, d'accessibilité, d'efficacité, d'efficience et de satisfaction. Il s'agit en quelque sorte du pendant professionnel du domaine qui devient alors *design d'expérience usager*, *design d'interaction*, ou plus traditionnellement *architecture d'information*. Plus l'analyse IHO se fait tôt dans le processus de conception des systèmes, plus on dira de cette conception qu'elle est *centrée sur l'utilisateur* : le design du système étant alors guidé, et non seulement ajusté, par la prise en compte des besoins et comportements des usagers, de leur contexte organisationnel particulier et des contenus qu'ils utilisent (Morville & Rosenfeld, 2006).

2.5.3 Le WS et la question de l'interface usager

En regard aux techniques et technologies particulières du WS, l'IHO et la CMO demeurent encore largement muets. Très peu de recherches IHO ou CMO ont été menées pour comprendre ou créer des interfaces et des systèmes de communication virtuelle basés sur ce type de techniques et technologies. Seules quelques problématiques très spécifiques, ou très générales, ont été abordées. Côté IHO, les recherches en WS traitent quasi exclusivement du problème de la visualisation des données sémantiques – qui sont généralement peu abordable dans leur forme fondamentale de « graphe ». Côté CMO, les recherches en WS tablent souvent sur des cas d'utilisation de systèmes sémantiques à des fins de communication, mais ces derniers sont habituellement plus illustratifs des principes généraux du WS que sujets précis à des analyses psychosociologiques rigoureuses.

Considérant, avec Kiryakov, Popov, Terziev, Manov et Ognyanoff (2003), que le corpus des technologies du WS est parvenu à une certaine maturité fonctionnelle (dans sa philosophie globale et ses éléments techniques), mais que les applications sachant en intégrer les potentialités à un niveau intermédiaire font encore grandement défaut, ou du moins qu'aucune n'a réussi à acquérir une popularité digne d'enclencher une dissémination généralisée du concept de WS (Henricks, 2007), il est à se demander si le développement d'applications à base d'ontologies ne manque pas de vision en matière de design d'interaction et d'expérience usager. Soit, les techniques et technologies du WS présentent un potentiel puissant et novateur en matière de gestion des informations, mais tant qu'elles n'auront pas clarifié et normalisé la manière dont les interfaces-usager exposent leurs fonctionnalités et les rendent accessibles aux utilisateurs, elles demeureront sous le joug paralysant d'une perception négative qui les range dans la classe des technologies de type « vaporware ». Janna Anderson, professeur de communication à l'Université Elon, affirme en ce sens qu'elles sont difficiles à appréhender par les usagers ordinaires d'Internet : « The concept of the semantic web has been fluid and evolving and never quite found a concrete expression and easily-understood application that could be grasped readily by ordinary

Internet users. » (Anderson, 2010, para. 1). En effet, la plupart des applications sémantiques actuellement disponibles requièrent bien souvent de l'utilisateur qu'il comprenne un minimum de notions relatives au WS pour qu'il en apprécie les fonctionnalités. Mais comme ces notions revêtent un caractère plutôt abstrait, la facilité d'apprentissage des interfaces sémantiques est donc rarement au rendez-vous.

Au niveau de l'interface graphique (*graphical user interface* ou GUI en anglais) des systèmes électroniques, il faut remarquer que la disposition des éléments graphiques sur le médium à deux dimensions qu'est l'écran d'ordinateur procède déjà d'un régime sémantique qui lui est propre, en ce que les proportions, les alignements, les regroupements, les isomorphismes, les encadrements, etc., suggèrent d'entrée de jeu certains rapports sémantiques spécifiques comme des relations d'association ou de composition entre les éléments représentés (Tidwell, 2006). Considérant ce fait, nous pouvons alors déduire que ce « langage » de la graphique a toutes les chances d'entrer en dissonance avec celui des données sémantiques lorsque la présentation de ces dernières ne prend pas en considération cette tendance naturelle des cerveaux humains à interpréter les compositions graphiques (ce que les théories de la Gestalt ont bien démontré). En effet, la structure des données sémantiques, qui tirent leur force de leur capacité à représenter des objets sur un maximum de dimensions grâce à RDF, suit une logique plutôt réticulaire alors que l'écran, avec ses deux dimensions (plus une troisième qu'il ne peut que simuler au mieux), suit plutôt une logique de grille et se voit donc d'une certaine manière limitative pour les données sémantiques qui doivent alors user de l'aspect dynamique de l'interface pour s'exprimer correctement. C'est d'ailleurs cette compréhension intuitive qu'a l'esprit humain de la logique de la grille qui aurait été un facteur déterminant dans l'adoption massive des bases de données de type relationnelles, qui fonctionnent selon un modèle tabulaire et non hiérarchique ou réticulaire (Graziosi, 2010).

Mais, comme le note Graziosi (2010) : « l'information, qui est par nature interreliée et non « contenue dans des tables », est par conséquent mieux représentée sous forme de graphe » (ma traduction ; Graziosi, 2010, section *Relational won the battle*, para. 3). Cette remarque

de Graziosi, à propos des systèmes d'organisation des données, semble être un écho contemporain de celle que fit Vannevar Bush à propos des systèmes d'organisation des informations-documents de son époque dans son article prophétique de 1945, *As We May Think*, dans lequel il jeta les bases conceptuelles de l'ordinateur personnel, de l'hypertexte et d'Internet en imaginant le memex, un dispositif de traitement automatisé des informations (microfilms) dont le fonctionnement des structures de stockage et d'accès reflèterait mieux le mode de fonctionnement – foncièrement associatif – de l'esprit de ses usagers humains :

« Our ineptitude in getting at the record is largely caused by the artificiality of systems of indexing. When data of any sort are placed in storage, they are filed alphabetically or numerically, and information is found (when it is) by tracing it down from subclass to subclass. It can be in only one place, unless duplicates are used; one has to have rules as to which path will locate it, and the rules are cumbersome. Having found one item, moreover, one has to emerge from the system and re-enter on a new path.

The human mind does not work that way. It operates by association. With one item in its grasp, it snaps instantly to the next that is suggested by the association of thoughts, in accordance with some intricate web of trails carried by the cells of the brain. » (Bush, 1945, section 6)

Bush table ici sur la nécessité pour les dispositifs techniques de classement documentaire de se doter d'interface offrant des structures de navigation plus ergonomiques vis-à-vis le penchant naturel des cerveaux humains à traiter les informations en tant qu'elles sont interreliables, associables. Puis, lorsqu'il poursuit en présentant son fameux memex, sorte de dispositif hypertextuel qui permettrait de créer des accès rapides entre plusieurs documents distincts et donc de les naviguer de manière associative, Bush admet néanmoins le rôle toujours nécessaire et fondamental du principe de l'index pour atteindre rapidement un enregistrement précis :

« There is, of course, provision for consultation of the record by the usual scheme of indexing. If the user wishes to consult a certain book, he taps its code on the keyboard, and the title page of the book promptly appears before him, projected onto one of his viewing positions. » (Bush, 1945, section 6)

La position dominante du moteur de recherche Google, qui constitue d'une certaine façon le grand « code book » du WWW, témoigne à cet égard de la persistance et du rôle encore bien central des structures indexicales dans nos manières de localiser et de fureter des informations (Google étant un gigantesque index jumelant l'URL des ressources web à des mots clés représentatifs de leur contenu).

L'index, en tant que liste ordonnée, est une très vieille technologie qui est sans doute plus profondément ancrée dans nos modes cognitifs d'élicitation et d'organisation des informations que des technologies plus sophistiquées, et plus récentes, qui offrent des modes de structuration arborescente ou rhizomiques (par opposition au mode linéaire de la liste). Au-delà de la simple limitation dimensionnelle de l'écran, il faut donc aussi voir que les structures simples comme l'index, résolument linéaires, sont cognitivement plus ergonomiques que les formes plus sophistiquées qui rompent avec l'éloquence signalétique de la linéarité. En réalité, on peut préciser, avec Christeaens, Wets et Vanthienen (1998), en soulignant que ce caractère ergonomique des informations présentées sous forme listée, et par extension sous forme tabulaire, est relatif à la nature plus ou moins interprétative et séquentielle de la tâche de l'utilisateur : lecture des valeurs discrètes de chaque donnée, comparaison de la valeur des données, etc..

Puisqu'elles consistent en des listes d'enregistrements (eux-mêmes listant des valeurs de champs), on peut dire des tables d'une base de données relationnelle qu'elles appartiennent à la même famille technologique que l'index. Un graphe RDF tombe en revanche sous la classe des dispositifs rhizomiques. En ce sens, les efforts de développement du WS, en promouvant l'adoption de RDF comme format de donnée, constituent en quelque sorte la poursuite de la vision de Bush, mais au niveau des modes d'exploration et d'association des données. *Linked Data*, un sous-concept méthodologique du WS et projet cher de Berners Lee, propose justement une interconnexion globale les différentes sources ouvertes de données sur le web.

En somme, bien que la structure des données sémantiques soit plus représentative de la façon dont le cerveau humain se représente le monde, l'expression visuelle la plus respectueuse et concise de ces données qu'est le graphe n'est pas un type de pattern graphique que les agents humains sont traditionnellement habiles à décoder. Le défi des technologies du WS vis-à-vis les interfaces usager est donc qu'ils ne doivent pas tenter de leur imposer leur régime graphique propre, mais plutôt assumer leur rôle de technologies d'arrière-plan (*back-end*) et de travailler à les optimiser, c'est-à-dire de travailler à les rendre mécaniquement plus simples et ergonomiques tout en soutenant des expériences utilisateurs aussi efficaces sinon plus. Dans l'économie de l'attention à l'intérieur de laquelle nous sommes de plus en plus immergés, le ratio entre l'effort de consultation d'un média et la puissance de son effet devient un enjeu critique pour les solutions numériques qui tentent de survivre à l'écosystème des pratiques modernes. Une stratégie qui apparaît en ce sens gagnante pour les projets de développement de systèmes sémantiques est de créer des systèmes qui permettent de déléster l'utilisateur de certains traitements cognitifs qu'il doit ordinairement opérer de son côté de l'interface pour les faire exécuter par la machine, derrière l'interface. D'un point de vue IHO, les technologies sémantiques prouvent donc leur valeur au mieux lorsqu'elles sont utilisées pour créer des interfaces avenantes à même d'alléger les dialogues entre la machine et l'utilisateur. Et c'est précisément cette perspective des choses qui a guidé le développement de Combine, notre application de recrutement sémantique.

3 Projet Combine : Réalisation pratique d'un prototype d'application sémantique dédiée au recrutement en TI

3.1 Présentation du projet Combine

Pour comprendre plus concrètement les enjeux technologiques, économiques, CMO et IHO du concept de WS évoqué jusqu'ici, nous avons fait l'exercice de développer une application à base d'ontologie. Nous avons développé en l'occurrence Combine, une application sémantique qui prend pour requis les besoins particuliers d'une communauté de pratique (CdP) spécifique, soient ceux d'un groupe de recruteurs oeuvrant dans le domaine des TI.

3.1.1 Motivations du projet

Les contenus générés par les usagers et le réseautage social sont maintenant de plus en plus perçus par les communautés de recruteurs comme des sources d'information crédibles, utiles et efficaces, surtout par les recruteurs évoluant dans un champ professionnel où il y a rareté de main-d'œuvre. Tel est la réalité des recruteurs en TI qui doivent fréquemment recourir à des méthodes de recrutement de type « chasseur de têtes » (c'est-à-dire qu'ils doivent recruter des travailleurs qui ne sont pas dans le circuit de la recherche d'emploi) pour atteindre leurs objectifs. Mais, tel qu'observé, développer et gérer un réseau social étendu et efficace demande un investissement considérable de temps et d'attention. Les politiques de séquestration des données (*data lock-in*) de la plupart des principaux sites de réseautage rendent difficile pour les recruteurs d'exploiter plusieurs de ces sites en parallèle ou encore de les faire échanger des informations entre eux et avec leur système

d'information maison (voir Figure 17 qui caricature le problème). Et ce n'est que la pointe de l'iceberg en termes de problèmes d'interopérabilité rencontrés par les recruteurs dans leurs interactions quotidiennes avec leurs ressources électroniques. En fait, dans une perspective IHO, une quantité fulgurante de temps et d'attention est perdue par les recruteurs dans la manutention manuelle des données ; des ressources précieuses qui pourraient être résolument mieux investies dans l'entretien et l'accroissement la présence sociale en ligne, dans la documentation des candidats, et/ou dans toute autre activité qui augmente directement la valeur ajoutée du service de recrutement. Nous pensons donc que les technologies du WS pourraient contribuer à éliminer – ou du moins amoindrir significativement – les problèmes d'interopérabilité spécifiques aux processus de ce champ d'activité.

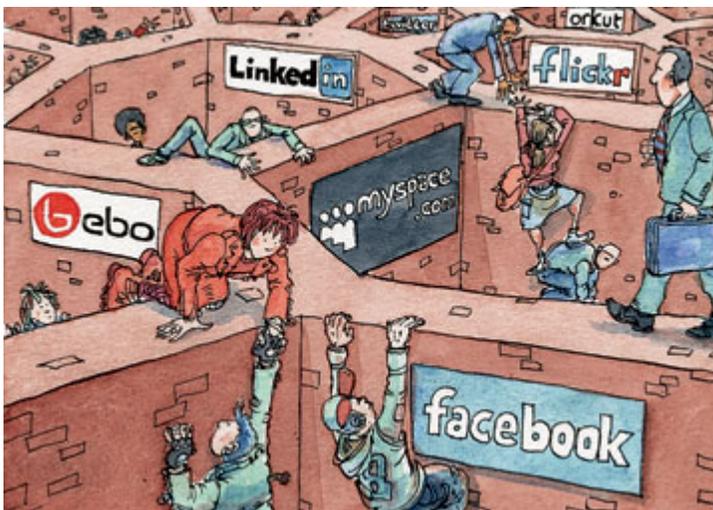


Figure 17 - Illustration de David Simonds parue dans le magazine *The Economist* et caricaturant le problème de la séquestration des données entre les sites de réseautage (*The Economist*, 2008).

Un autre aspect du travail des recruteurs qui pourrait aussi être facilité par l'utilisation d'une application sémantique est l'évaluation et la sélection de candidatures. En effet, une bonne part des ressources attentionnelles des agents recruteurs est généralement mobilisée

pour l'exécution de cette activité. Même si normalement les tâches impliquées par cette activité sont supportées par les dispositifs électroniques à la disposition des recruteurs, nous avons pu remarquer qu'il n'en demeure pas moins que les recruteurs doivent réaliser beaucoup d'interactions avec leurs ressources numériques pour mener à bien ce type de tâche. Nous croyions donc qu'une application sémantique pourrait aussi aider à cet égard.

3.1.2 Objectifs

En tant que *démonstration critique* des bénéfices potentiels pour une CdP donnée d'adopter les technologies du WS, le projet Combine devait rencontrer des objectifs pratiques relatifs à la conceptualisation de l'application comme telle, et des objectifs théoriques portant sur la discussion et la problématisation du cas de développement.

En ce qui a trait aux objectifs pratiques, deux groupements principaux sont à distinguer : la modélisation de l'application d'e-recrutement et l'évaluation du retour sur investissement potentiel de l'application modélisée. Dans le groupe des objectifs se rapportant à la modélisation du système on retrouvait :

- L'analyse du contexte d'utilisation (soit le domaine du recrutement en TI) ;
- L'élaboration des spécifications de l'application (inventorisation, comparaison et configuration des solutions techniques existantes + modélisation UML¹⁷ de l'architecture du système) ;
- La construction du prototype de l'application (modélisation de l'ontologie + prototypage de l'interface graphique)

Dans le groupe des objectifs liés à l'évaluation du système on retrouvait :

- L'évaluation du potentiel d'optimisation des processus de travail par l'application ;

¹⁷ UML est l'acronyme pour *Unified Modeling Language*. UML constitue un langage graphique développé initialement pour assister la conception de logiciels orientés objet, mais qui avec le temps est devenu extensivement applicable à la représentation de données et de processus en tout genre.

- L'évaluation de l'intérêt des usagers potentiels (les recruteurs) à utiliser une telle application

En ce qui concerne les objectifs théoriques, le projet avait pour mandat général d'exposer, de définir et de discuter les concepts, les problèmes et les opportunités rencontrés tout au long du développement de l'application. À la base, la recherche s'est donc attardée à définir et à analyser la notion de « sémantique » tel qu'elle est *mécanisée* par les standards du WS (ce qui a été exposé précédemment dans la partie du cadre théorique). D'un point de vue méthodologique, la recherche s'est aussi attardée à comprendre l'art d'éliciter et de modéliser des connaissances, compréhension qui aura ensuite guidé l'élaboration de la stratégie de modélisation de la base de connaissances de Combine, en fonction des fonctionnalités espérées du système. Mais considérant le mandat du projet Combine d'intégrer des technologies et techniques du Web 2.0 et du Web 3.0, cet objectif de discuter le développement de l'application à la lumière de l'IHO et de la CMO a donné lieu à des réflexions sur des thématiques telles que :

- Les dimensions éthiques et légales de l'extraction ou du partage ouvert d'informations personnelles (profilage) ;
- Le problème de l'engagement ontologique dans la définition et l'opérationnalisation de concepts sociaux ;
- L'avenir des réseaux sociaux à l'ère du WS

3.1.3 Hypothèses

Dans la perspective de communication humaine adoptée par ce projet, nous avons mis de l'avant deux propositions qui devaient être validées, ou invalidées, par l'évaluation de l'utilité – projetée par les concepteurs et perçue par les usagers potentiels – de l'application modélisée :

1. L'usage de l'application prototypée devrait améliorer les interactions humain-ordinateur des acteurs du domaine par une gestion plus efficiente des ressources informationnelles et documentaires du domaine.
2. L'usage de l'application prototypée devrait améliorer la communication médiée par ordinateur de la CdP ciblée par la formalisation et le partage d'un savoir d'avant et d'arrière-plan¹⁸.

De plus, en tant que ce projet embrasse aussi – et évidemment – les perspectives du WS, fut posée comme hypothèse corolaire que : « il est possible d'utiliser une ontologie pour gérer les échanges d'information entre des SIS non enrichis sémantiquement, telles les applications sociales ». L'étude de faisabilité technique de Combine se devait donc de supporter cette affirmation. En d'autres termes, le projet devait démontrer qu'il est possible d'extraire des informations en provenance d'applications de type Web 2.0 pour les intégrer, en utilisant une ontologie, dans la base de connaissance (base de données sémantiques) d'une application d'e-recrutement.

3.1.4 Méthodologie

Pour en répondre des objectifs pratiques fixés, dans ses grandes lignes, la conceptualisation de l'application Combine a procédé selon cinq temps, soit :

1. L'analyse du domaine d'application
2. L'étude stratégique des opportunités du domaine
3. La modélisation de l'ontologie et de l'application
4. La définition des spécifications de l'application et des enjeux contextuels de développement
5. L'évaluation de l'application vis-à-vis les utilisateurs ciblés

¹⁸ La distinction d'un savoir d'avant-plan d'un savoir d'arrière-plan fait ici référence à la distinction entre la ABOX et la TBOX de la base de connaissances en ce que la ABOX représente les données sémantiques consolidant les informations usuelles relatives aux offres de poste et aux candidats qui seront manipulées par l'utilisateur directement via l'interface de l'application et que la TBOX représente le modèle ontologique définissant les concepts utilisés par la ABOX et qui sera utilisé par la machine pour traiter de manière automatique les données sémantiques, derrière l'interface.

Certaines de ces phases de développement ont nécessité la réalisation d'activités d'enquête sur le terrain pour être menées à bien. En l'occurrence, il s'agit des activités d'analyse du domaine d'application, de la modélisation de l'ontologie et de l'évaluation de l'application.

En ce qui concerne la stratégie d'enquête sur le terrain pour réaliser l'analyse du domaine d'application et de modélisation de l'ontologie, une multitude d'activités a été déployée afin d'en éliciter les ressources, les processus, les connaissances et les besoins nécessaires à la performance des activités quotidiennes de l'entreprise. Ces activités d'enquête comprenaient :

- L'analyse de contenu (demandes de personnel, courriels, CVs, etc.)
 - o Identification et catégorisation des objets métiers (profils de candidat, entrevues, notes d'entrevue, requis de poste, etc.)
 - o Identification et catégorisation des acteurs du domaine (candidats, consultants, clients, recruteurs, etc.)
- L'analyse des processus de travail standardisés
- L'observation en situation réelle de travail
- L'analyse des systèmes d'information et de communication utilisés

Par rapport aux objectifs d'évaluation de l'application, deux autres activités d'enquête ont été menées pour mesurer l'éventuelle réceptivité du système par les usagers et son efficacité potentielle :

- Administration d'un questionnaire pour obtenir du feedback des usagers ciblés sur l'application après une visite guidée du prototype.
- Développement et application d'une métrique utilisant des diagrammes de séquence ULM qui modélisent et comparent des scénarios d'utilisation types pour donner un portrait avant-après des processus réels versus ceux virtuellement « optimisés » par l'application et ainsi quantifier les économies (en temps et étapes) attendues du nouveau système.

3.2 Analyse du domaine d'application

3.2.1 Description générale du domaine d'activité à modéliser

Le domaine d'activité auquel tâchera de s'adapter l'application Combine est celui des organisations vendant des services de consultation et de dotation de personnel en TI. Pour des fins heuristiques, cette recherche a donc pris pour modèle une PME montréalaise d'environ 300 employés qui vend justement de tels services. Nous l'appellerons ici « FIC », acronyme de son plein nom. Plus précisément encore, comme l'application Combine cherche à répondre aux problématiques issues des nouvelles pratiques de recrutement, la recherche s'est focalisée sur le sous-domaine d'activité propre à la CdP des recruteurs qui s'adonnent à du e-recrutement social, c'est-à-dire la CdP des recruteurs qui utilisent des ressources électroniques de type Web 2.0 dans l'exécution de leurs processus de travail.

FIC, qui possède un pied-à-terre à Paris, Québec et Boston, fut fondé en 1992 et possède une certification ISO 9001¹⁹. Elle figure parmi les 10 plus notables des quelque 250 entreprises du genre à Montréal. CGI, IBM (LGS), DMR, Code 6, Cofomo et Systématix sont en ce sens les concurrents les plus immédiats de FIC. Se décrivant non pas comme une firme ou une agence, mais plutôt comme un *groupe*, les consultants-employés FIC possèdent un large spectre d'expertise :

- Gestion de processus
- Gestion de Projets ; PMP (Project Management Professional)
- Intelligence d'affaires (BI ; Business Intelligence)
- Entrepôt et bases de données
- Infrastructure, télécom et réseautique
- Sécurité informatique
- Développement d'applications et de systèmes de type ERP (Enterprise Resource Planning)
- Technologies Web et Internet

¹⁹ Les processus de recrutement de la compagnie ont été forgés selon ce standard ISO pour pouvoir essentiellement devenir fournisseur de clients qui sont titulaire de cette certification ISO 9001.

En plus de fournir des services de consultation, FIC offre aussi à ses clients des services d'embauche et de pré-embauche²⁰.

3.2.1.1 Structure organisationnelle de l'entreprise ciblée

FIC est une entreprise familiale assez jeune et, par rapport à la typologie des organisations d'Henry Mintzberg (2004), il s'agit à première vue d'une organisation de type *entrepreneuriale* qui présente une structure assez simple, plutôt informelle (ligne hiérarchique peu développée), flexible, et où le chef d'entreprise possède un leadership fort qui repose sur son talent à promouvoir avec charisme sa vision personnelle des choses. En même temps, certains aspects de l'entreprise – l'autonomie des employés-consultants²¹ notamment – pourraient inciter à ranger FIC dans la case des organisations dites *professionnelles* : la structure centrale de l'entreprise (service de recrutement + direction de comptes + gestion des opérations), qui regroupe environ une trentaine de personnes²², peut être vue en tant que support logistique²³ de l'ensemble des consultants éparpillés un peu partout dans les organisations clientes. En ce sens, les employés non consultants constituent une sorte d'organisation dans l'organisation, et c'est cette sous-organisation (ou plutôt *sur-organisation*) qui peut à juste titre être taxée d'entrepreneuriale. Par ailleurs, on peut dire de FIC que, globalement, elle tient le rôle de support logistique vis-à-vis de ses organisations clientes. En leur fournissant des ressources humaines qualifiées qui viendront supporter les opérations de ces dernières, ce que vend concrètement FIC est un service de recherche et de dotation d'expertises en technologies de l'information.

²⁰ Le service de pré-embauche offre la possibilité à une entreprise cliente d'embaucher un candidat à titre de consultant d'abord, puis, si elle en est satisfaite, d'employer ce dernier en tant que permanent.

²¹ Par rapport à la structure organisationnelle de FIC et non nécessairement par rapport à celles des entreprises clientes dans lesquelles ils œuvrent à titre de consultant.

²² Voir annexe 3 pour visualiser l'organigramme, ainsi que les détails des deux grands processus qu'ils ont pour mission de soutenir (vente + recrutement).

²³ Issue de la terminologie de Mintzberg, l'expression « support logistique » désigne une unité de personnel dont le mandat est de fournir certains services spécifiques au reste de l'entreprise, à l'interne.

3.2.1.2 Défis saillants du domaine

Croissance perpétuelle des connaissances tacites

Suivre l'évolution effarante des divers objets de spécialisation (technologies et institutions) constitue certainement un des défis les plus importants du domaine. La plupart des gens qui œuvrent dans le monde du recrutement/dotation en TI possèdent des formations en gestion et/ou en relations industrielles, ce qui fait que la connaissance des nombreuses techniques, technologies, compagnies et/ou certifications peuplant l'univers des TI, bien que tout à fait pertinente à la bonne réalisation des objectifs processuels en ce domaine, constitue pour les agents-recruteurs un savoir toujours tacite et parcellaire. Très fréquemment, les recruteurs et les gestionnaires de comptes se voient obligés d'assister à des séances de présentation de métiers TI, de chercher sur le web des informations par rapport aux à une technologie ou d'une expertise particulière, ou encore de s'informer à propos des caractéristiques et interconnexions des diverses organisations peuplant le champ d'activité (clients, fournisseurs, institutions académiques, etc.). En ce sens, la valeur d'un travailleur en gestion des ressources humaines dans le domaine des TI est assez directement liée à la maîtrise plus ou moins profonde, de même qu'à l'étendue qu'il a, de tous ces éléments de connaissance du domaine. Il est à préciser toutefois que ce qui est valorisé est l'aspect transversal (l'étendue) de cette maîtrise : plus important que l'affinement de chaque élément de connaissance TI, il s'agit d'être en mesure d'avoir une vue d'ensemble du domaine. Cette primauté de la vue d'ensemble sur la compréhension détaillée du corpus de connaissances en TI découle du fait qu'une des responsabilités d'un gestionnaire de talents TI est d'aider les clients à formuler leurs besoins de ressources en TI, et donc la capacité de cerner les modes d'intégration transdisciplinaires des différents sous-domaines des TI se voit être un atout crucial (bien que, paradoxalement, une connaissance plus fine d'un domaine particulier soit nécessaire au bon établissement de ses relations interdisciplinaires). La clé du succès par rapport à ce défi particulier réside donc dans l'équilibre entre une capacité d'exploration analytique des domaines de connaissances en TI et une aptitude à synthétiser et systématiser les éléments d'analyse recueillis.

Fragmentation des sources d'information

L'éclatement actuel des formes de communication bouleverse sensiblement (pour ne pas dire radicalement) les manières habituelles de prospection, de rétention et de diffusion des informations du domaine. Classiquement, le téléphone est le moyen de communication privilégié. Sous quelques aspects, il demeure, et demeurera probablement encore longtemps, la manière la plus directe, rapide et la plus universelle de joindre quelqu'un. Néanmoins, le courriel, qui depuis plusieurs années déjà complémente quasi systématiquement l'appel téléphonique, s'est vu augmenté d'une multitude de nouvelles ressources web. On parle d'ailleurs maintenant de « e-recrutement » pour marquer cet état des choses. Et, avec la montée en puissance de la culture du 2.0, les ressources web en question se rapportent maintenant de plus en plus à celles qui forment le corpus constitutif d'un genre web particulier : le web social.

Agglomérant à peu près toutes les formes médiatiques préexistantes, les sites de réseautage social professionnel comme LinkedIn, Viadeo et Plaxo (de même que les wikis et les blogues) donnent lieu aujourd'hui à une source d'informations autant inédite que porteuse d'opportunités. Par exemple, un usage déjà en vigueur, par rapport au recrutement de candidats pour un poste où il y a une rareté de main-d'œuvre, est celui d'avoir accès à des candidats potentiels qui ne sont pas nécessairement en recherche d'emploi, c'est-à-dire qui n'affichent aucune présence active dans les circuits médiatiques traditionnels de la recherche d'emploi (banques de cv, profil sur sites de recherche d'emploi, liste de regroupements professionnels, etc.). Il demeure cependant que les informations dénichées via ces nouveaux médias, lorsque jugées pertinentes, sont habituellement copiées, reformatées, puis stockées dans le système d'information particulier de la firme. Le hic, c'est qu'en faisant cela, non seulement est impliqué un effort de transcription certain, mais on fige en quelque sorte les informations recueillies et leur mise à jour doit être exécutée de façon tout aussi manuelle que pour leur saisie. Et cela, c'est sans compter l'effort de vigilance pour rester simplement à l'affut des changements de profils des candidats suivis.

En ce sens, on retranche un aspect très attrayant des informations en provenance du web social : leur actualisation dynamique et distribuée par le fait que les candidats peuvent tenir à jour eux-mêmes leurs renseignements personnels et professionnels.

Concurrence et raccourcissement des délais

De plus, il est à considérer que la prolifération des moyens de communication susmentionnée accroît la pression concurrentielle déjà féroce du milieu et a pour effet systémique de raccourcir radicalement les temps de réaction pour répondre à la réception du besoin d'un client (poste à combler). Dans les hautes sphères de ce créneau d'activité, une structure dite d'« accréditation » existe dans laquelle de gros clients acheminent toujours simultanément et exclusivement leurs demandes en personnel aux mêmes cinq à sept firmes « accréditées » chez eux, ce qui vient considérablement amplifier le phénomène en ne laissant en moyenne que dix à vingt minutes aux firmes contactées pour combler le besoin. Cela fait en sorte que le recrutement de candidats se voit forcé de devenir de plus en plus proactif et prospectif. La rareté de certaines candidatures, par rapport à des postes de haut niveau, ou encore exigeant une spécialisation en forte demande, exotique ou vétuste, est une autre réalité du domaine qui convie aussi à une capacité de prospection accrue.

3.2.2 Processus, activités et acteurs du domaine

D'une manière très générale, on peut dire presque trivialement que les recruteurs FIC – les membres de la CdP ciblée – sont responsables du macroprocessus²⁴ qu'est le recrutement de candidats. Tel que l'ont clairement démontré nos observations sur le terrain, l'activité centrale de ce macroprocessus, celle qui monopolise le plus l'horaire et l'attention des

²⁴ L'analyse des processus ici menée découpe conceptuellement la notion en regard à la hiérarchie des processus d'une organisation de Suzanne Rivard et Jean Talbot (2008) qui considère l'aspect processuel d'une organisation comme un emboîtement hiérarchique de macroprocessus, de sous-processus, d'activités et de tâches (voir Annexe 3 pour un exemple graphique)

recruteurs, et qui est perçue comme étant la plus critique par ces derniers, est sans conteste la recherche de candidats. Mais tel que mentionné plus tôt, l'arrivée des sites de réseautage social, en tant qu'ils constituent de nouvelles sources de prospection, a beau aider les recruteurs à trouver des candidatures intéressantes, il n'en reste pas moins qu'elle provoque une surcharge cognitive, en ce que le développement et la maintenance d'un réseau social en ligne requièrent beaucoup d'attention et de temps. En ce sens, la considération et l'exploitation de ces nouvelles ressources (les sites de réseautage) telles qu'elles sont effectivement pratiquées sur le terrain, font apparaître l'activité de recherche de candidats comme étant sous-optimale – par rapport principalement à l'aspect encore très manuel des tâches que cette activité implique. L'un des besoins les plus saillants de la CdP étudiée, d'un point de vue IHO, était donc de contrer la surcharge informationnelle issue de l'activité de prospection des sites de réseautage. En regard des technologies du WS, on peut voir plusieurs optimisations possibles des processus d'e-recrutement social à même d'amoindrir cette problématique. En d'autres termes, il y a un besoin réel pour un système supportant les recruteurs dans leurs tâches de gestion de leurs réseaux de contacts en ligne (par rapport à leurs objectifs de recrutement).

Bien qu'il s'agisse du problème primaire auquel s'adresse l'application Combine, il y a dans les faits plusieurs autres activités du processus de recrutement qui pourraient aussi être facilitées par l'installation d'un système sémantique. En amont, il y a la transmission des demandes de personnel du client au directeur de compte, et puis du directeur de compte au recruteur. Au niveau du processus de recrutement comme tel, il y a des activités comme : l'entrée des demandes de personnel dans le SI maison (un service web nommé Taleo), l'acquisition de connaissances sur les expertises et technologies relatives aux demandes de personnel à traiter, la publication des offres de postes sur des SI externes, le développement et l'entretien d'une présence sociale en ligne (sur les sites de réseautage), la recherche de candidats dans le SI maison, la recherche de candidats sur les SI externes (sites d'emplois, banques de CVs, sites de réseautage, etc.), l'entrée des profils des nouveaux candidats dans le SI maison, la mise à jour des profils, la documentation des candidats (références +

entrevues), et la communication avec les candidats. En aval, le recrutement implique : l'obtention de l'approbation des candidats sélectionnés par le directeur de compte, le formatage des CVs des candidats sélectionnés pour fins d'envoi aux clients (cette tâche est souvent déléguée à un adjoint administratif), l'envoi des dossiers de candidatures aux clients, l'organisation des entrevues-client (le client rencontre le candidat), et l'obtention du feedback à propos de l'entrevue-client, celui des clients comme celui des candidats. Comme on peut le constater, le processus de recrutement implique cinq types d'acteurs : les recruteurs, les directeurs de compte, les clients, les candidats et les adjoints administratifs.

3.2.3 Inventaire des tâches de la CdP des recruteurs

Étant donnée la perspective IHO de cette recherche, les activités générales précédemment identifiées se devaient d'être détaillées en tâches spécifiques sur lesquelles notre solution sémantique pourra effectivement jouer. De l'enquête menée sur le terrain, il a été possible d'observer (pour les activités directement impliquées dans le processus central de recrutement) qu'il y avait des tâches récurrentes telles que :

- Entrer les demandes de personnel dans le SI maison
 - o Copier les descriptions de postes des courriels des directeurs de compte (contenant les descriptions de postes transmis par le client) dans le formulaire « nouvelle demande de personnel » du SI maison
 - o Paramétrer l'état des demandes de personnel
- Acquérir des connaissances sur les expertises et technologies relatives aux demandes de personnel
 - o Lire de la documentation en ligne à propos des technologies ou expertises impliquées par le requis du poste
 - o Ajouter les nouvelles technologies ou expertises en tant que tags dans le SI maison
- Publier les offres de postes sur des SI externes
 - o Copier les descriptions de poste du SI maison dans les formulaires d'affichage d'emplois des sites de recherche d'emplois et/ou de réseautage
- Développer et entretenir une présence sociale en ligne
 - o Envoyer et solliciter des invitations à joindre de nouveaux contacts
 - o Joindre des groupes professionnels

- Chercher des candidats sur le SI maison
 - o Convertir les requis de poste en recherche par mots clés
 - o Raffiner la recherche
 - o Consulter les CVs des candidats
 - o Sélectionner les candidats avec le meilleur *fit*²⁵ par rapport à la demande de personnel
- Chercher des candidats sur les sites de recherche d'emplois traditionnels (SI externes)
 - o Convertir les requis de poste en recherche par mots clés
 - o Raffiner la recherche
 - o Contacter les candidats trouvés
 - o Créer de nouveaux profils de candidats dans le SI maison avec les informations sur les candidats trouvés
- Chercher des candidats sur les sites de réseautage (SI externes)
 - o Convertir les requis de poste en recherche par mots clés
 - o Demander à ses contacts s'ils connaissent des candidats potentiels pour un poste donné
 - o Raffiner la recherche
 - o Contacter les candidats trouvés
 - o Créer de nouveaux profils de candidats dans le SI maison avec les informations sur les candidats trouvés
- Entrer le profil des nouveaux candidats dans le SI maison
 - o Copier les informations des candidats issues de sources externes dans le SI maison
 - o Tagger les profils de candidat (à propos de leurs expertises ou technologies maîtrisées ; méthode de catégorisation des candidatures propre à la CdP et orientée par les possibilités techniques de leur SI maison)
- Tenir à jour les profils
 - o Demander aux candidats leur CV à jour
 - o Joindre les CVs aux profils des candidats
 - o Modifier le statut des candidats à mesure qu'ils progressent dans le processus d'embauche
- Documenter les candidats
 - o Acquérir et vérifier les références des candidats
 - o Fixer et noter les dates et lieu des entrevues avec les candidats
 - o Prendre et consigner des notes à propos des entrevues avec les candidats
- Communiquer avec les candidats
 - o Garder des traces des communications avec les candidats

²⁵ Le *fit* est le terme usuel qu'utilise la CdP pour parler de l'adéquation entre une candidature et le requis d'un poste.

3.2.4 Analyse et classification des tâches critiques

Pour des fins de concision l'analyse des tâches ici présentée ne traite que des tâches récurrentes jugées les plus problématiques et les regroupe en fonction du type de problème qu'elles occasionnent.

3.2.4.1 Copie des informations

De toutes les tâches récurrentes, copier des informations d'un système à un autre est certainement la plus triviale et la plus redondante d'entre toutes. Le fait est que cette tâche n'est pas perçue comme apportant une valeur ajoutée directe au processus global de recrutement, elle est souvent négligée. Pour être plus précis il faut dire que cette négligence ne se pose pas tellement en termes de l'exécution même de la tâche qu'en termes d'un différentiel marqué entre ce qui pourrait être avantageusement copié et ce qui est effectivement copié. Par exemple, même s'ils sont tenus par les processus officiels de l'entreprise de diffuser les offres de poste sur tous les sites externes identifiés comme pertinents, les recruteurs ne les affichent jamais systématiquement sur tous ces sites à la fois, faute de temps pour remplir tous les formulaires web spécifiques à chacun de ces sites. Très souvent même, ils n'afficheront pas du tout l'offre de poste et préféreront se mettre à la recherche de candidats dans les banques de CV et les réseaux sociaux dès la réception du besoin en personnel. Copier de l'information en vue de tenir à jour des dossiers et/ou des profils de candidat est aussi un autre cas d'espèce de cette tâche que les recruteurs exécutent avec minimalisme : ils se contentent généralement de faire les mises à jour essentielles des seuls dossiers en cours d'étude, relativement aux seules demandes actives de personnel.

De la manière spécifique dont le recrutement est pratiqué chez FIC, les recruteurs interagissent avec plusieurs SI pour accomplir les tâches qui leur incombent. L'enquête sur le terrain montre qu'ils utilisent en moyenne six différents systèmes (SI maison + sites web externes) sur une base régulière pour trouver des candidats. Le Tableau 1 montre en ce sens

la liste ordonnée des sources web de candidatures les plus fréquentées quotidiennement de chaque recruteur FIC. La moyenne est de 4,6 sites externes utilisés par recruteur. Il est à remarquer aussi que la majorité des recruteurs n'investissent qu'un seul site de réseautage et qu'aucun n'en investit plus de deux. En comptant tous les sites sur lesquels ils affichent des offres de poste ou sur lesquels ils vont chercher des informations sur des spécialités spécifiques (logiciel interne de gestion de la relation client, Wikipedia, sites niches d'affichage de postes, etc.) deux ou trois SI pourraient gonfler ce nombre et porter à huit ou neuf le nombre total de SI utilisés.

Tableau 1 - Sources externes de candidatures utilisées par chaque recruteur FIC (les sources sont ordonnées par ordre de fréquentation)

| REPONDANT | SOURCE1 | SOURCE2 | SOURCE3 | SOURCE4 | SOURCE5 | SOURCE6 |
|-----------|----------|----------|------------|------------|------------|---------|
| 1 | Monster | LinkedIn | Itjob | Agent Solo | Jobboom | 0 |
| 2 | Monster | LinkedIn | Agent Solo | Viadeo | 0 | 0 |
| 4 | LinkedIn | Monster | Aqiii | Jobboom | Agent Solo | Itjob |
| 5 | Monster | LinkedIn | Jobboom | Beljob | Aqiii | 0 |
| 6 | Monster | LinkedIn | Jobboom | 0 | 0 | 0 |

Étant donné que les interactions des recruteurs avec tous ces systèmes sont principalement motivées par un besoin de faire circuler de l'information d'une plateforme à une autre (les informations sur les postes du SI maison vers les sites externes de recherche d'emplois et les informations sur les candidats des sources web externes vers le SI maison), et comme les recruteurs ne disposent pas d'autres moyens techniques que le presse-papier de leurs ordinateurs personnels pour ce faire, nous avons observé que les échanges d'information entre les systèmes utilisés nécessitent toujours l'intervention manuelle d'un recruteur.

Le problème avec cette réalité est qu'il y a une quantité considérable de temps dépensé à la manipulation de telles informations de bas niveau²⁶. Plus encore, étant exécutée

²⁶ La notion d'information de bas niveau, qui s'oppose à celle d'information de haut niveau, dénote une information possédant un faible niveau d'abstractivité et de complexité. En bref, elle se rapporte à tout ce qui est de l'ordre des données élémentaires.

manuellement, cette tâche ne peut être accomplie que de manière sérielle. Cela veut dire que les morceaux d'information ne peuvent être transférés qu'un à un et vers un seul système à la fois. La Figure 18 illustre en ce sens la séquence des sous-tâches impliquées. Pour commencer, le recruteur doit visualiser et copier un premier morceau d'information du système source. Dans un deuxième temps, il doit localiser et atteindre l'interface du système cible. Puis, il doit mentalement convertir le morceau d'information en fonction des champs de formulaire spécifiques du système cible, qui diffèrent souvent légèrement de ceux du système source. Et finalement, il doit entrer ou copier l'information dans la boîte de texte du champ de formulaire approprié et réitérer la séquence jusqu'à ce que tous les morceaux d'information aient été transférés. (Sur la figure, chaque morceau d'information est représenté par un carré gris et la tonalité de gris code le format de données spécifiques des différents SI en présence.)

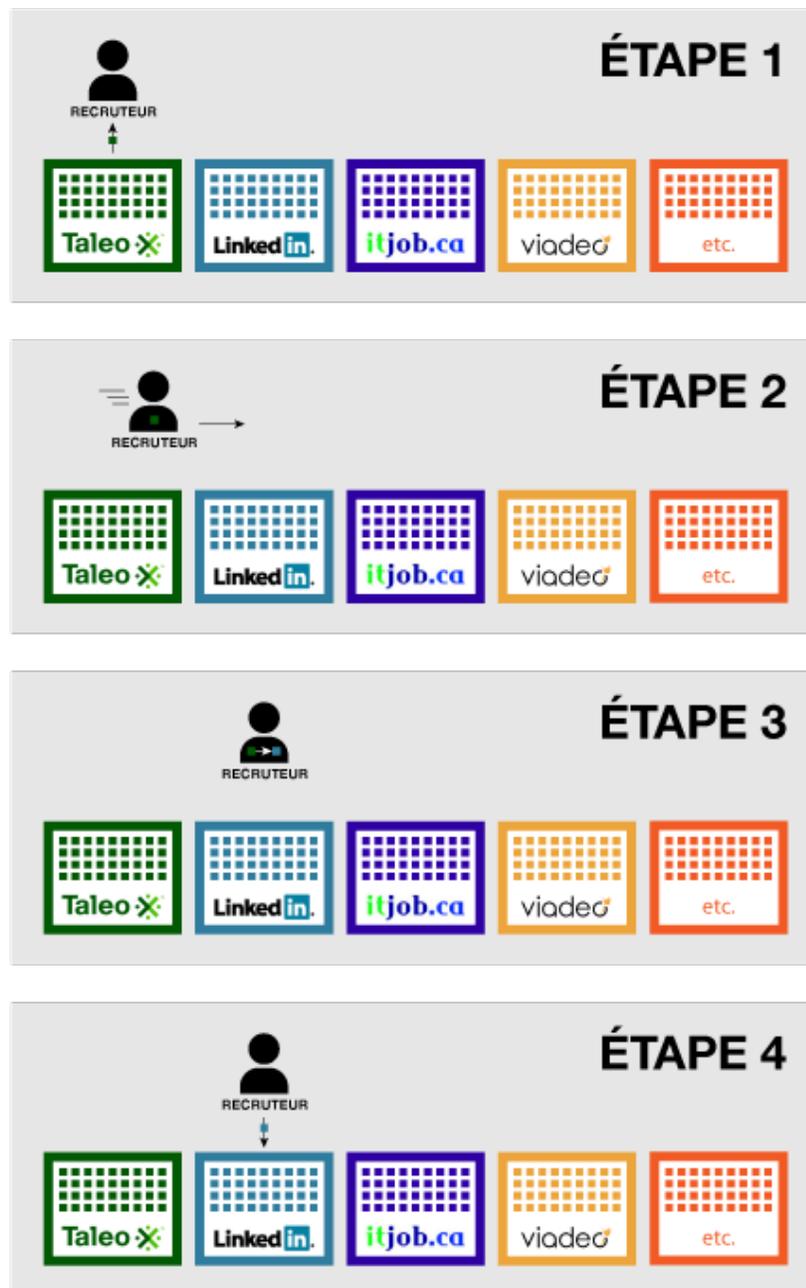


Figure 18 - Séquence des sous-tâches impliquées dans le transfert manuel des informations d'un système à l'autre.

3.2.4.2 Documentation des communications

Une autre tâche négligée qui touche au problème de la copie manuelle de l'information précédemment soulevé est celle de garder des traces des communications avec les candidats (ou autres acteurs). Tout comme la copie d'information, elle n'est pas considérée comme une tâche apportant de la valeur ajoutée concrète et immédiate. Elle est donc souvent omise ou minimalement réalisée. Bien que le SI maison des recruteurs met à leur disposition des fonctionnalités permettant de consigner de telles informations, le fait est que les recruteurs FIC emploient des canaux de communication qui sont extérieurs au SI maison (téléphone cellulaire, téléphone fixe, boîte de courriels locale, boîte de courriels en ligne, etc.), et qui eux-mêmes conservent automatiquement certaines traces des communications passées, diminue considérablement le sentiment d'obligation des recruteurs de bien documenter chacune de leurs communications dans le SI maison tel qu'indiqué par la consigne de l'entreprise, d'autant plus que cette documentation doit être effectuée encore là de manière manuelle.

3.2.4.3 Tâches reliées à la recherche de candidats

À l'opposé, les recruteurs tendent à considérer comme plus importantes les tâches relatives à la recherche de candidats : la valeur ajoutée est évidente. Elles sont sans conteste les plus complexes des tâches, et du même coup les plus consommatrices de temps. Le fait que les recruteurs sont plus que conscients de la centralité et de l'importance cruciale de ce lot de tâches peut être pointé comme étant l'incitation première des recruteurs à négliger d'autres tâches jugées dès lors plus périphériques. Tel est le cas des deux tâches susmentionnées.

Pour vraiment comprendre cet ensemble de tâches critiques du processus de recrutement, un peu plus de déconstruction est nécessaire.

Primo, il s'agit des tâches les plus exigeantes cognitivement parce qu'elles impliquent beaucoup de connaissances d'arrière-plan et d'aptitude à raisonner (à faire des inférences logiques), et surtout les tâches de recherche sur les SIs à proprement parler (recherche par mots-clés + raffinement de recherches + évaluation et sélection des candidats trouvés). Les connaissances d'arrière-plan en question sont les connaissances dont le recruteur a besoin pour comprendre les différents termes composant les requis de poste, et tout particulièrement ceux spécifiant le type d'expertise et/ou de compétences techniques demandées. Autrement dit, les recruteurs doivent connaître les définitions générales des expertises et technologies, de même que les relations sémantiques entre les différentes instances de celles-ci, pour pouvoir juger de la pertinence d'une candidature. Mais les recruteurs, qui ne sont pas des spécialistes TI de formation, ont besoin d'un temps considérable pour acquérir une familiarité minimale de toutes les spécialités TI. Même les recruteurs d'expériences se voient régulièrement confrontés à des spécialités qui leurs sont inédites, parce qu'elles sont nouvelles, très anciennes ou exotiques. Bien que les connaissances d'arrière-plan à propos des spécialités techniques soient les plus critiques par rapport au type de recrutement effectué chez FIC (recrutement en TI), toute une panoplie d'autres connaissances sont à l'œuvre dans la réalisation de cette tâche de recherche de candidats (connaissances des types d'aptitudes organisationnelles, des secteurs d'activité industrielle, des entités géopolitiques, des entités horaires, etc.). C'est donc sur la base de leurs connaissances d'arrière-plan que les recruteurs sont plus ou moins habilités à faire des déductions et des comparaisons à propos des objets d'information dont ils disposent. Par exemple, les recruteurs doivent déduire à partir des requis de poste quels sont les meilleurs critères de pertinence avant d'entreprendre la recherche comme telle. Ils ont à déduire subséquemment quels sont les meilleurs mots clés par rapport aux critères choisis. Après, et seulement après, la recherche peut techniquement commencer.

Dans un deuxième temps, selon la quantité et la qualité des résultats de recherche retournés par le système, le recruteur choisi généralement parmi trois options : a) raffiner la recherche en ajoutant des mots clés supplémentaires s'il y a trop de résultats ; b) essayer d'autres

critères, et par le fait même d'autres mots clés, si les résultats sont insuffisants ; ou c) quitter les SI maison et commencer de nouvelles recherches sur les SI externes (sites de recherche d'emplois ou de réseautage) si toutes les options de recherche sur le SI maison ont été épuisées, ou encore si le recruteur juge qu'il trouvera de meilleures candidatures ailleurs.

Il est à noter que pour évaluer l'aspect qualitatif des résultats de recherche, les connaissances d'arrière-plan et l'aptitude au raisonnement du recruteur demeurent décisives. Dans les faits, cette évaluation de l'adéquation des candidats vis-à-vis les postes à combler implique du recruteur qu'il ouvre, lise et compare, à l'aune d'un requis du poste spécifique, chaque profil de candidat trouvé. Et cette sous-tâche n'est pas aussi simple qu'elle en a l'air : on ne peut se contenter de faire une plate comparaison mot à mot. Une bonne comparaison est une comparaison faite à un certain niveau d'abstraction, supérieur au niveau strictement littéral. À titre illustratif, prenons le cas d'un recruteur qui, en cherchant à combler un poste requérant des compétences de programmation en HTML et CSS, tombe sur le profil d'un candidat stipulant uniquement que celui-ci est programmeur Ajax. Le recruteur considèrera ce candidat seulement s'il sait que, conceptuellement, le terme « Ajax » tient pour *Asynchronous JavaScript and XML*, un corpus de techniques en programmation web qui utilise des standards tels que HTML, CSS, JavaScript, DOM, XML et XMLHttpRequest, et donc qu'un candidat qui programme en Ajax a forcément des compétences en HTML et CSS. C'est en l'occurrence le genre de déductions qui pourraient être réalisées par une ontologie couplée à un moteur d'inférence. L'ontologie FIC que nous avons développée pour Combine a été conçue en ce sens.

De plus, il est aussi à noter que, quand le recruteur ne trouve pas de candidatures de qualité directement sur le SI maison, il doit atteindre une autre interface de recherche (normalement celle d'un des SI externes) et y répéter le processus de recherche. En ce sens, d'avoir un système qui pourrait lancer des requêtes sur tous les SI à partir d'une interface de recherche unique serait d'une grande valeur, pas seulement en frais d'économie de temps, mais aussi pour l'aspect pratique de pouvoir comparer les résultats de recherche de

toutes les sources sur un même écran. Si toutes les sources d'information stockaient leurs données dans des dépôts RDF ouverts, cette interface de recherche centralisée serait très facile à construire et à opérer en ce qu'elle utiliserait directement des requêtes SPARQL. Mais de tels dépôts étant encore pratiquement inexistant, des techniques plus indirectes doivent être utilisées pour obtenir un tel système centralisé. Il sera expliqué plus loin en quoi consistent ces techniques alternatives.

3.2.5 Remarque générale sur l'analyse des tâches

Dans l'ensemble, il est loisible de remarquer que les deux principales problématiques soulevées lors de cette analyse sont les revers complémentaires, qui s'alimentent mutuellement, d'une problématique plus générale qu'est le manque d'interopérabilité entre les différentes ressources logistiques, communicationnelles et documentaires utilisées. D'un côté il y a les tâches reliées à la manutention des informations de bas niveau qui sont négligées au profit de tâches plus cognitives se rapportant au traitement d'informations plus abstraites qui elles, d'un autre côté, se voient compliquées et alourdies par le manque de consolidation, d'actualisation et d'uniformité des informations de bas niveau sur lesquelles elles s'appuient.

3.3 Opportunités pour les technologies du WS d'améliorer les processus du domaine d'application

Dans cette section, nous confronterons les problématiques qui se sont dégagées de la précédente analyse du domaine d'application aux diverses potentialités des technologies sémantiques exposées à l'intérieur du cadre théorique. Se faisant, nous ferons apparaître des fenêtres d'opportunités spécifiques à l'intérieure desquelles il sera possible pour notre application sémantique de faire une différence positive.

Tels que révélés par l'analyse des tâches, les principaux besoins de la CdP des recruteurs FIC sont de :

- Faciliter les échanges d'informations (réduction de la manutention manuelle des données)
- Améliorer la mise à jour des données
- Améliorer la documentation des communications
- Offrir un support cognitif pour la recherche de candidats
- Réduire le nombre d'interfaces de recherche
- Accélérer les recherches
- Améliorer la qualité et la profondeur des recherches

Tous ces besoins constituent dès lors les exigences auxquelles l'application à base d'ontologies devait répondre, et donc sur la base desquelles les spécifications de Combine furent pensées.

3.3.1 Traitement intelligent des données

En formalisant les connaissances d'arrière plan du domaine d'activité, nous avons vu qu'une ontologie (la TBox) permet la constitution, ainsi que l'exploitation intelligente de connaissances d'avant-plan (la ABox : les données sémantiques comme telles) et qu'elles peuvent même créer de nouvelles informations à partir de l'intégration raisonnée d'ensembles d'informations fragmentaires. Ensemble, elles rendraient donc possible le support de tâches cognitivement accaparantes comme acquérir, suivre et interrelier des informations en provenance de sources externes distribuées.

Par exemple, une ontologie du recrutement, qui comprendrait des connaissances déclaratives telles que la liste de tous les objets de spécialisation TI et leurs relations (langages de programmation, logiciels, composants matériels, types de service, etc.) et/ou une liste de tous les objets d'attestation de compétence et leurs relations (formations, certifications, diplômes, expériences, etc.) permettrait à une application d'e-recrutement implémentant ces représentations formalisées de coupler automatiquement des offres de poste à des profils de candidat sur la base d'une évaluation de la proximité sémantique de ces objets.

3.3.2 Navigation et visualisation plus riches des données

Étant donné que la structure réticulaire (graphe) sous-jacente au format RDF utilisé par les systèmes sémantiques jouit d'un potentiel d'interconnectivité qui va au-delà de la structure tabulaire des bases de données traditionnelles, les applications basées sur RDF possèdent une capacité accrue de lier et intégrer les données qui permet au final à l'utilisateur de les explorer et les visualiser de manières bien plus riches qu'avec une base de données traditionnelle. Contrairement aux applications standards fonctionnant sur des bases de données relationnelles qui ne peuvent rendre compte de l'aspect multidimensionnel d'une donnée, les applications à base de RDF peuvent facilement expliciter et gérer toutes les variables relatives à une donnée : l'objet ou le prédicat d'un triplet²⁷ RDF pouvant en tout temps devenir le sujet d'un autre triplet. Ce type de structure de données facilite donc par conséquent la programmation d'interfaces qui permettent des interactions intelligentes avec les données présentées à l'écran comme la navigation par facettes ou l'exploration de graphe par exemple. En ce sens, une application sémantique d'e-recrutement pourrait présenter ses informations dans une variété de contextes visuels : la visualisation des CVs sur une ligne du temps, le positionnement de la localisation des postes offerts sur une carte géographique avec des icônes dont la couleur code le type de poste (voir Figure 19 qui

²⁷ Un triplet RDF est un morceau de connaissance représenté par une phrase élémentaire de forme sujet-predicat-objet.

montre un écran de Combine qui réalise une telle exploration contextuelle des données), etc.

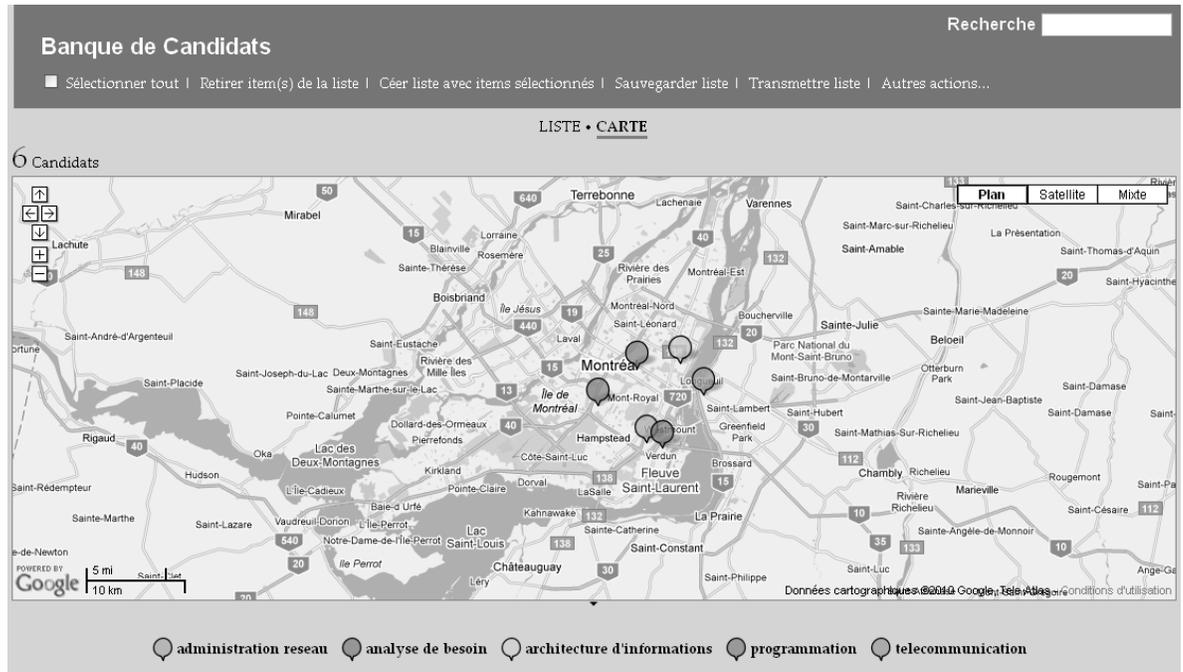


Figure 19 - Localisation de postes offerts et identification du type de poste par une icône codée par la couleur (interface de Combine)

3.3.3 Approvisionnement et intégration en temps réel des données

Un autre avantage des systèmes basés sur RDF est qu'ils permettent une exploitation « en direct » des sources d'information distribuées. Dans un Web sémantique pleinement développé, la copie et la recopie d'informations d'un système à un autre seront inutiles puisque les sources d'information donneront des accès directs aux requêtes en provenance de systèmes externes, donnant par conséquent à ces derniers l'opportunité de pouvoir toujours utiliser les données les plus à jour, directement de la source. Ce faisant, cela éliminera le besoin de copier et synchroniser les données entre les systèmes : les sources de

données sémantiques fourniront les informations requises sur demande, automatiquement. De plus, il sera possible d'agréger des informations en provenance de sources multiples et localement distribuées avec une requête unique. Cela fait miroiter au domaine du recrutement en TI un horizon corporatif dans lequel les dépôts sémantiques de CVs, les réseaux sémantiques socioprofessionnels et autres ressources web utiles possédant des structures sémantiques appropriées seront interrogeables à partir d'une seule interface de requêtes. La beauté d'une telle vision tient au fait que les données employées par un tel système seraient toujours aussi à jour que celles des systèmes sources.

3.3.4 Puissance et transparence accrues de la documentation

La possibilité pour les recruteurs de laisser des traces transparentes sur les objets d'information qu'ils manipulent par des annotations sémantiques constitue un autre point positif en faveur d'un système sémantique de recrutement. D'avoir un dispositif d'annotation sémantique permettrait aux recruteurs, ainsi qu'aux autres acteurs du domaine, de greffer des commentaires (ou des notes) sur n'importe quel objet métier du système et même sur les commentaires eux-mêmes (un peu à la manière dont on peut commenter presque tout sur Facebook). Plus encore, en permettant aux recruteurs d'avoir un accès en temps réel à ces annotations, le système deviendrait une plateforme de communication à part entière (les annotations pourraient être suivies et répondues sur des fils de nouvelles, ou directement sur les objets d'annotation).

3.4 Modélisation de Combine

La présente section traduit les opportunités précédemment énoncées en des exigences plus spécifiques pour l'application sémantique conceptualisée.

3.4.1 Exigences et cas d'utilisation de l'application

Dans un premier temps il est à distinguer deux catégories d'exigences et de cas d'utilisation : ceux se rapportant à la modélisation de la base de connaissances comme telle, et ceux se rapportant à la conceptualisation d'une application particulière exploitant cette dernière. La première définit les potentialités générales de la constitution d'une ontologie, la seconde définit le potentiel d'intégration ergonomique d'une application sémantique dans la pratique spécifique qu'est le recrutement tel qu'il est pratiqué chez FIC.

3.4.1.1 Exigences et cas d'utilisation de la base de connaissance

La question à se poser ici est : Pourquoi une base de connaissances ? À quels besoins²⁸ généraux, et ultimement spécifiques, la réalisation puis l'application d'une ontologie répondront-elles ? Il en va du conseil méthodologique de Noy et McGuinness (2007) qui stipule que de dresser une liste des questions de compétence auxquelles la base de connaissances devra être en mesure de répondre devrait aider à déterminer le focus et la granularité²⁹ de l'ontologie à construire. Dans notre cas, la question éminente qui a guidé nos choix de modélisation fut la suivante : « Quel(s) profil(s) de candidat, parmi tous ceux disponibles sur les sources identifiées, correspond(ent) le mieux à telle ou telle demande de personnel ? ».

De cette question centrale, à savoir quel est le meilleur candidat pour une demande de personnel donné, on a pu définir le cas d'utilisation principal de la base de connaissances, et par extension de l'application sémantique en ce qu'elle devait pouvoir coupler automatiquement chaque demande de personnel X à une liste de candidats potentiels ordonnée en fonction de critères présélectionnés (adéquation de l'expertise, proximité

²⁸ À l'intérieur du mémoire, la notion de « besoin » est comprise/problématisée en terme de processus « optimisable », par opposition à son acception usuelle qui dénote un « quelque chose qui fait défaut », un « quelque chose qui manque ».

²⁹ La granularité d'un modèle dénote le niveau de raffinement conceptuel de ce modèle.

géographique, années d'expérience, emplois précédents, disponibilités, etc.). Ce cas d'utilisation principal de l'ontologie prévu pour Combine exploite donc résolument l'opportunité susmentionnée d'un traitement plus intelligent des données.

De cette exigence pour l'ontologie de devoir opérer un tel couplage ont découlé deux implications méthodologiques. D'abord, l'élicitation terminologique du domaine et la modélisation de l'ontologie se devaient d'être focalisées sur les termes que sont le « Requis de poste » et l'« Acquis du candidat », en ce qu'ils constituent les deux concepts centraux dans l'acte d'évaluer l'adéquation d'un candidat par rapport à un poste donné. Et puis, la construction des règles d'inférence – pour opérer cette évaluation – exigea de l'enquête qu'elle débroussaille les raisonnements que mettent en œuvre les acteurs du domaine dans la sélection et la gestion des candidats potentiels, pour déterminer ce qu'ils entendent d'un point de vue logico-sémantique lorsqu'ils font des déclarations décisives du genre « Le candidat X est le *meilleur* pour la job Y ».

En vue d'exploiter l'opportunité susmentionnée à l'égard d'un approvisionnement et d'une intégration en temps réel des données, la question de compétence suivante fut aussi énoncée : « À quoi correspond l'information X d'un certain SI, dans les autres SI utilisés ? ». À cette compétence de conversion des terminologies propres à chacun des SI qui seront fédérés par l'application sémantique, on a pu définir quelques cas d'utilisation intéressants tels que :

- L'uniformisation des informations dans tous les SI
- La consolidation et complétion automatique des informations fragmentaires
- La mise à jour automatique des informations
- La présentation et la recherche d'information multisource sur une même interface

En ce qui concerne l'opportunité d'une documentation plus puissante et plus transparente, de même que celle d'une navigation et d'une visualisation plus riches des données, il s'agit là d'opportunités inhérentes à la nature même de l'ontologie et des données sémantiques, et

n'ont donc pas eu d'implications particulièrement importantes au niveau de la modélisation de l'ontologie FIC. Ces opportunités sont en ce sens le reflet du caractère extensible et flexible des données sémantiques, caractère que devra exploiter l'application sémantique comme telle.

3.4.1.2 Exigences et cas d'utilisation de l'application sémantique

Dans la perspective d'un développement logiciel, la question à se poser est : Comment se servir de la base de connaissances pour créer des dispositifs de gestion de contenu efficaces ? En l'occurrence, l'application particulière qui était à développer (assister la tâche de prospection des recruteurs) permet de préciser cette notion de « dispositif » contenue dans la précédente question de sorte à transformer cette dernière en : Comment, via l'exploitation d'une base de connaissances, créer des « postes de pilotage » capables de gérer ergonomiquement de vastes réseaux de contacts issus du web social ? Bref, Combine se devait de permettre aux recruteurs d'être en mesure de piloter leurs réseaux socioprofessionnels respectifs ou communs, malgré le fait que ceux-ci se composent de gens dont les avatars numériques proviennent de sites web différents. Encore là, on peut décortiquer cette notion de « pilotage » en ce que l'application développée devait offrir un tableau de bord qui permette à l'utilisateur de visualiser globalement l'information contenue par son réseau en fonction de différentes métriques et paramètres d'une part, et devait offrir des commandes permettant la transformation, le partage et la communication de ces informations d'autre part. En frais de communication des informations, on peut considérer le système dans son ensemble pour voir que l'application a pour mandat général d'interfacer le SI maison de l'entreprise avec les réseaux socioprofessionnels égo-centrés³⁰ de chacun des recruteurs. Il agira donc un peu comme une espèce de mémoire tampon entre les sites de réseautage, considérés comme banques – disparates et officieuses – de candidats

³⁰ En théorie des réseaux sociaux de communication on dit qu'un réseau social est de type égo-centré lorsque tous les nœuds (personnes) du graphe-réseau ont un lien direct (1^{er} degré de contact) avec une même et unique tierce personne. Personne de laquelle on pourra dire qu'il s'agit de SON réseau social personnel.

potentiels sur le net, et la banque de candidats-employés officiels de l'entreprise. Comme nous l'avons vu, les conseillers en recrutement doivent actuellement réaliser cette opération de manière « artisanale », c'est-à-dire qu'ils doivent récolter manuellement les informations pour les entreposer tout aussi manuellement.

3.4.2 Modélisation ontologique du domaine d'application

Si l'on se rapporte au développement de Combine à titre d'exemple, l'identification et la catégorisation des divers acteurs et objets métier du domaine d'application correspondent dans l'ordre à l'inventoriage des éléments de la ABox (identification des instances), suivi de la modélisation de la TBox (formalisation de la structure de classes subsumant les instances identifiées). En ce qui concerne l'inventoriage des éléments de la ABox de la base de connaissances de Combine, nous avons déconstruit les principales ressources documentaires utilisées par les recruteurs (analyse de l'architecture d'information de leurs contenus). Figure 20 montre un exemple de déconstruction, en l'occurrence la déconstruction (partielle) d'une fiche de demande de personnel FIC. Comme chaque champ et section de la fiche était déjà bien étiqueté, il était relativement facile dans ce cas d'identifier les instances qui la composent, mais ce ne fut pas le cas de tous les documents. Dans un sens, l'identification des instances de la ABox – l'étiquetage des éléments de contenu – pouvait déjà la voie pour la modélisation de la TBox en ce que les libellés utilisés pour identifier les morceaux de contenu ont pu servir en tant que concepts spécifiques de l'ontologie.

Demande de personnel : PCO Senior « [Retour à la liste](#) » [Envoyer la demande vers Combine](#)

[Modifier](#) [Supprimer](#) [Dupliquer](#) [Redémarrer le processus d'approbation](#) [Aperçu imprimable](#) [Envoyer](#) [Dupliquer en tant que modèle](#) [Formes](#)

Prochaines étapes : [Candidats \(8\)](#) / [Chercher](#) / [Rouvrir](#) / [Fermer - Comblé par Facilité](#) / [Fermer - Besoin annulé](#) / [Fermer - Candidat non-disponible](#) / [Fermer - Candidats non-retenus](#) / [Fermé - Offre non-conclue](#) / [Fermer](#) / [Refusé](#) / [En attente d'approbation](#)

Information sur la demande de personnel :

| | |
|--|---|
| Client : Jeux Quebec Nom du client | Date de dépôt : 09-11-03 |
| Contact client : Mario Griroux | Date de début : 09-11-16 Date de début du poste |
| Titre : PCO Senior Titre du poste | Durée : 6 mois renouvelable |
| État : En attente - Retour client | LegacyReqID : |
| Détenteurs : Forge, Catherine | LegacyReqCreateDate : |
| Lapierre, Martin | |
| Date d'ouverture : 09-10-27 | Prime : 100\$ |

Détails

| | |
|--|---|
| Profils : | Taux client : |
| Technos : | Horaire : |
| Niveau du poste : Senior Niveau du poste | Localisation : Région de Montréal Localisation du poste |
| Nombre de candidats : 1 | Lieu mandat : |
| Type de contrat : Consultation | Anglais requis : |
| Direct, S.T. : Direct | Candidats réservés : Imed Chouchani |
| Taux candidat : | |

Description

Description du poste

Description : Le responsable de la planification et du suivi se rapporte au directeur du projet et son rôle est de supporter celui-ci dans le suivi et le contrôle du projet. Il interviendra dans l'application et l'évolution des processus de gestion de projet. Il assistera le directeur de projet dans la planification et l'exécution de son projet afin de livrer dans les délais et à l'intérieur des budgets définis, et ce, selon les normes d'assurance qualité en vigueur. Son rôle sera de gérer les volets suivants :

- Suivre et maintenir le plan global du projet pour l'ensemble des étapes (démarrage, définition, réalisation, mise en place et formation);
- Analyser les écarts et participer à la mise en place de solutions;

Figure 20 - Analyse de l'architecture d'information d'une fiche de demande de personnel

Si l'identification des instances de la ABox fut une activité presque triviale, il s'en est allé autrement de la modélisation de la TBox en ce qu'elle impliquait de prendre des décisions comportant une part d'arbitraire importante (problème de l'engagement ontologique ; nous en traiterons plus loin). Comme pour n'importe quel modèle de données, l'efficacité de la TBox d'une ontologie est une question d'équilibre entre sa simplicité et sa précision (Lonjon et Coll., 2006).

3.4.3 Modélisation de l'application

La Figure 21 présente schématiquement comment devait s'intégrer les cas d'utilisation de Combine à l'intérieur du circuit fonctionnel qu'est le processus de recrutement et modélise à haut niveau la manière dont l'application devra soutenir ces cas d'utilisation. Sur ce graphique, les bonshommes allumettes représentent les principaux acteurs du domaine et la chaîne des pastilles en vert pale est représentative du circuit élémentaire des activités constitutives d'un processus de recrutement. À l'intérieur de ce circuit, on peut voir que l'accent a été mis sur les activités propres au conseiller en recrutement, soit la recherche de candidats, l'étude et le suivi de candidatures, la sélection de candidats et la proposition de candidats au client. Les pastilles en vert foncé quant à elles consignent des tâches/activités plus concrètes incluses dans les grandes activités du conseiller en recrutement. Ce sont principalement ces tâches que l'application combine se devra de supporter. On parle donc ici du fait de rechercher de nouvelles candidatures, de consigner des candidatures, de documenter les candidatures et finalement d'apparier les besoins-client (postes à combler) aux candidatures disponibles. Au centre, il y a tout ce qui est de l'ordre du « dispositif » : c'est-à-dire, tout ce qui a trait aux systèmes d'information et aux fonctionnalités réalisées par ceux-ci. Plus spécifiquement, les boites représentent les dispositifs/systèmes comme tels, la case avec une page comme icône le document qu'est l'ontologie, et les pastilles ovales orangées les tâches/fonctionnalités nouvelles que devrait permettre de réaliser l'application sémantique, soit la présentation des informations du système sous forme de tableau de bord et l'agrégation (consolidation) des données dans la base de connaissances. Les éléments en jaune se rapportent à des ressources internes de la firme de recrutement, ceux en bleu aux ressources externes et ceux en orange aux ressources et fonctionnalités relatives à l'application sémantique. En ce qui concerne les flèches pointillées avec une tête blanche pleine et qui n'ont pas de libellé, elles dénotent une relation de « réalisation », c'est-à-dire que l'élément source réalise, fait ou accomplit l'élément de destination.

3.5 Spécifications de Combine : caractéristiques et défis contextuels de l'application

Le nom de l'application ici modélisée, Combine, a été choisi du fait de ses allusions aux principales fonctionnalités de l'application. Primo, il y a *combine*, l'impératif présent du verbe « combiner », qui dénote l'action d'associer deux ou plusieurs choses ensemble et qui représente la fonctionnalité centrale de Combine qui consiste à associer automatiquement des offres de poste à des candidatures en fonction de la proximité sémantique des requis de poste et des acquis de candidat. Secundo, il y a le nom anglais *combine* (moissonneuse-batteuse en français) qui représente une machine servant à faire la récolte. Ici la métaphore de la moissonneuse-batteuse illustre le concept général du système en ce qu'il récolte automatiquement des informations à propos de candidats potentiels sur le Web Social et des informations à propos des offres de poste sur le SI maison en utilisant un mécanisme de cueillette (requêtes SPARQL + *screen scraper*), qu'il déconstruit et organise intelligemment les informations récoltées en utilisant des mécanismes de broyage et de filtrage (ontologie OWL + moteur d'inférence), et finalement qu'il exporte l'information traitée à des tiers systèmes en utilisant un mécanisme de transbordement (scripts PHP + API RAP).

L'interface centrale de Combine est constituée de deux écrans principaux : la banque de candidats (voir Figure 22) et la banque des postes (voir Figure 23). Ces deux écrans partagent une même structure générale mais diffèrent dans les options d'action qu'ils offrent relativement aux items spécifiques qu'ils contiennent. La sélection de la banque, qui fait basculer l'interface d'un écran à l'autre, s'effectue dans le cadran supérieur gauche du corps principal de l'interface (le tableau sous l'espace d'entête comprenant le logo de l'application et le menu global). Dans le cadran juste en dessous se trouvent les filtres relatifs à la banque affichée qui permettent à l'utilisateur de fureter les items de la banque (qui se trouvent dans le cadran inférieur gauche) via une classification par facettes personnalisables. Le filtrage par facettes des items est un système de navigation et de

recherche qui trie sélectivement les items en fonction de leurs caractéristiques. Par exemple, à partir de la banque des postes, on pourrait demander à voir seulement les postes de *programmeur* (facette 1), qui ont pour objet de spécialisation le *langage Java* (facette 2), qui sont relatifs au secteur *bancaire* (facette 3) et qui exigent un *minimum de 10 ans d'expérience* (facette 4). Dans l'entête du panneau principal (cadran supérieur gauche) se trouvent toutes les actions qui peuvent être exécutées sur les items telles que les regrouper en une liste, retirer des items de la liste, sauvegarder une liste, envoyer une liste, et autres actions plus spécifiques telles que changer le statut des candidats ou des postes, importer et exporter des données, publier des offres de poste et, bien sûr, lancer le coupleur requis-acquis pour matcher des candidats à des postes sélectionnés ou vice versa.



Banques
Candidats | Postes

Banque de Candidats

Selectionner tout | Retirer item(s) de la liste | Créer liste avec items sélectionnés | Sauvegarder liste | Transmettre liste | Autres actions...

Recherche

LISTE • CARTE

Filtres

Listes Sauvegardées

- 4 listeCandidats 1
- 3 listeCandidats 2
- 4 listeCandidats 3

Organisation

- 3 Bell
- 1 Hydro-Quebec
- 2 Videotron

Spécialisation

- 1 administration reseau
- 1 analyse de besoin
- 1 architecture d'informations
- 2 programmation
- 1 telecommunication

6 Candidat

Trier par : libellés et posteOccupé, puis par... • Grouper selon le tri

Actions sur la fiche

| Candidat | Coordonnées | Acquis du candidat | Annotations et documents | Poste(s) ouvert(s) attribué(s) |
|--|--|---|---|---|
| <p>1. Bob Taylor</p> <p>Technos: Axure et Enterprise Architect</p> <p>CV: cv.doc</p> <p>bob.taylor@videotron.ca</p> <p>Profils: web: LinkedIn</p> | <p>Architecte d'Informations chez Bell // Consultant FIC</p> | <p>Technos: Axure et Enterprise Architect</p> <p>CV: cv.doc</p> <p>Référence(s): Référence 01</p> <p>Notes d'entrevue: NotesEntrevue 01</p> | <p>Annotations et documents</p> <p>CV: cv.doc</p> <p>Référence(s): Référence 01</p> <p>Notes d'entrevue: NotesEntrevue 01</p> | <p>CIO - Videotron</p> <p>(afficher la liste, transmettre la liste)</p> |
| <p>2. Guy Dubois</p> <p>Technos: Windows server</p> <p>guy_dubois@hotmail.com</p> <p>Profils: web: LinkedIn</p> | <p>Administrateur reseau senior chez Bell // Consultant FIC</p> | <p>Technos: Windows server</p> <p>CV: cv.doc</p> <p>Référence(s): Référence 01</p> <p>Notes d'entrevue: NotesEntrevue 01</p> | <p>Annotations et documents</p> <p>CV: cv.doc</p> <p>Référence(s): Référence 01</p> <p>Notes d'entrevue: NotesEntrevue 01</p> | <p>CIO - Videotron</p> <p>(afficher la liste, transmettre la liste)</p> |
| <p>3. Julie Saugé</p> <p>Technos: Eclipse</p> <p>jsauges@gmail.com</p> <p>Profils: web: LinkedIn</p> | <p>Programmeur Java junior chez Hydro-Quebec // Consultant FIC</p> | <p>Technos: Eclipse</p> <p>CV: cv.doc</p> <p>Référence(s): Référence 01</p> <p>Notes d'entrevue: NotesEntrevue 01</p> | <p>Annotations et documents</p> <p>CV: cv.doc</p> <p>Référence(s): Référence 01</p> <p>Notes d'entrevue: NotesEntrevue 01</p> | <p>CIO et Analyse de besoins - Videotron et Molson</p> <p>(afficher la liste, transmettre la liste)</p> |

Figure 22 - L'écran de l'interface de Combine affichant la banque de candidats



Recherche

Banques
Candidats | Postes

Listes Sauvegardées

- 4 ListePostes 1
- 7 ListePostes 2
- 5 ListePostes 3

Recruteur attribué

- 2 Catherine Forge
- 10 Martin Lapierre

État de la demande liée au poste

- 3 Actif
- 1 Actif - Prospection
- 6 Fermée/écoulé -
- Complète/écoulé par Facilité/écoulé

Poste assigné à

- 1 Bob Taylor
- 2 Julie Saugé
- 1 Nadia Cadman

Banque de Postes

■ Sélectionner tout | Retirer item(s) de la liste | Créer liste avec items sélectionnés | Sauvegarder liste | Transmettre liste | Autres actions...

LISTE • CARTE • HISTORIQUE DES POSTES

10 Poste

Trier par : dateDépot, organisation, domaineSpécialisation, et libellés, puis par... • Grouper selon le tri

2009-11-16 (1)

■ **PCO Senior chez Hydro-Québec** (source:)

| Information sur le poste | Information sur la demande | Requis du poste | Candidat(s) assigné(s) |
|--|--|---|------------------------|
| Début du mandat: 2009-12-01 Fin du mandat: 2010-05-01 | État: Actif + Date d'ouverture: 2009-11-10 Date de dépôt: 2009-11-16 | Technos: NS Project + Secteur d'activité: énergie Niveau du poste: senior | |

1.

2009-10-20 (3)

■ **Analyste de besoins chez Molson** (source:)

| Information sur le poste | Information sur la demande | Requis du poste | Candidat(s) assigné(s) |
|--|--|--|--|
| Début du mandat: 2009-11-11 Fin du mandat: 2012-12-21 | État: Actif + Date d'ouverture: 2009-10-01 Date de dépôt: 2009-10-20 | Technos: IBM Rational + Secteur d'activité: alimentaire Niveau du poste: intermédiaire | Julie Saugé (afficher la liste, transmettre la liste) |

1.

■ **Programmeur dotNET senior chez Via Rail** (source:)

| Information sur le poste | Information sur la demande | Requis du poste | Candidat(s) assigné(s) |
|--|--|---|--|
| Début du mandat: 2010-01-01 Fin du mandat: 2012-09-26 | État: Actif + Date d'ouverture: 2009-10-01 Date de dépôt: 2009-10-20 | Technos: Visual Studio et dotNET + Secteur d'activité: | Valérie Godbout et Nadia Cadman (afficher la liste, transmettre la liste) |

2.

Figure 23 - L'écran de l'interface de Combine affichant la banque de postes

Nous allons maintenant présenter les dispositifs techniques supportant cette interface.

3.5.1 L'ontologie FIC et le coupleur Requis-Acquis

Première des choses, regardons le modèle de donnée de Combine : l'ontologie FIC (ontologie éponyme de l'entreprise ciblée). L'ontologie FIC, qui fut modélisée sous forme d'un fichier OWL, vise à formaliser les connaissances d'arrière-plan requises pour la constitution et la maintenance d'un savoir d'avant plan : le dépôt RDF. Par rapport à l'analogie de la moissonneuse-batteuse, on peut dire de l'ontologie qu'elle servira de schème organisationnel unificateur pour les composantes de cueillette, de broyage-filtrage et de transbordement de l'application. Pour ce faire, les connaissances d'arrière-plan formalisées devaient comprendre tous les objets métier importants du domaine d'application qui ont été identifiés en tant qu'éléments de la ABox. Cette identification a mené à l'assertion de classes spécifiques telles que :

Acteurs :

- Recruteur
- Directeur de compte
- Client
- Contact-client
- Candidat
- Consultant
- Etc.

Concepts abstraits :

- Demande de personnel
- Poste
- Requis (de poste)
- Acquis (de candidat)
- Projet
- Domaine de spécialisation
- Objet de spécialisation
 - o Logiciel

- Matériel
- Standard
- Secteur d'activité industrielle
- Élément temporel
- Etc.

Élément acquis/requis :

- Compétence
 - Compétence technique
 - Compétence linguistique
 - Compétence organisationnel
 - Etc.
- Diplôme/certification
- Expérience
- Formation
- Position corporative
- Salaire
- Position de poste
- Etc.

Communications, documents et médias :

- Chaine de courriels
- Chaine d'appels téléphoniques
- Conversation SMS
- Entrevue
- Contrat
- Description
 - Description de candidat
 - CV
 - Lettre de présentation
 - Note d'entrevue
 - Description de poste
 - Description de projet
 - Etc.
- VCard
- Site web
- Etc.

Plusieurs autres objets périphériques utiles ont aussi été identifiés.

La modélisation de la TBox de l'ontologie FIC fut en grande partie conduite par la question de compétence évoquée plus tôt : « Quel(s) profil(s) de candidat, parmi tous ceux disponibles sur les sources identifiées, correspond(ent) le mieux à telle ou telle demande de

personnel ? » De cette question, nous avons établi que la partie névralgique de l'ontologie sur laquelle nous devons mettre l'accent était l'ensemble des éléments qui définissent les acquis de candidat et les requis de poste, ces derniers étant après tout des acquis de candidat *exigés* par tel ou tel poste (telle une espèce d'acquis virtuel). Comme la fonctionnalité principale de l'application est le couplage des profils candidat aux demandes de personnel (offres de poste), les concepts et relations les plus importants et critiques à définir pour ce faire sont donc sans conteste ceux se rapportant aux objets pivots que sont le *Requis de poste* et l'*Acquis de candidat*.

La Figure 24 donne un aperçu simplifié de cette portion critique de l'ontologie. En général, on peut observer que les concepts *AcquisCandidat* et *RequisPoste* sont des enfants-frères d'un concept général d'*Acquis* qui est défini en tant qu'il se compose d'éléments d'acquis (*AcquisElement*). Ces éléments d'acquis sont ensuite définis par des propriétés d'objet qui les lient à des concepts catégoriels et attributifs (objet de spécialisation, type de compétence, secteur d'activité industriel, etc.), ainsi que par des propriétés de donnée qui permettent de les noter et/ou de les annoter (par exemple l'ontologie FIC définit une propriété de donnée nommée *noteSur10* qui permet au recruteur d'attribuer un score évaluatif pour chaque élément d'acquis d'un candidat ou d'une offre de poste).

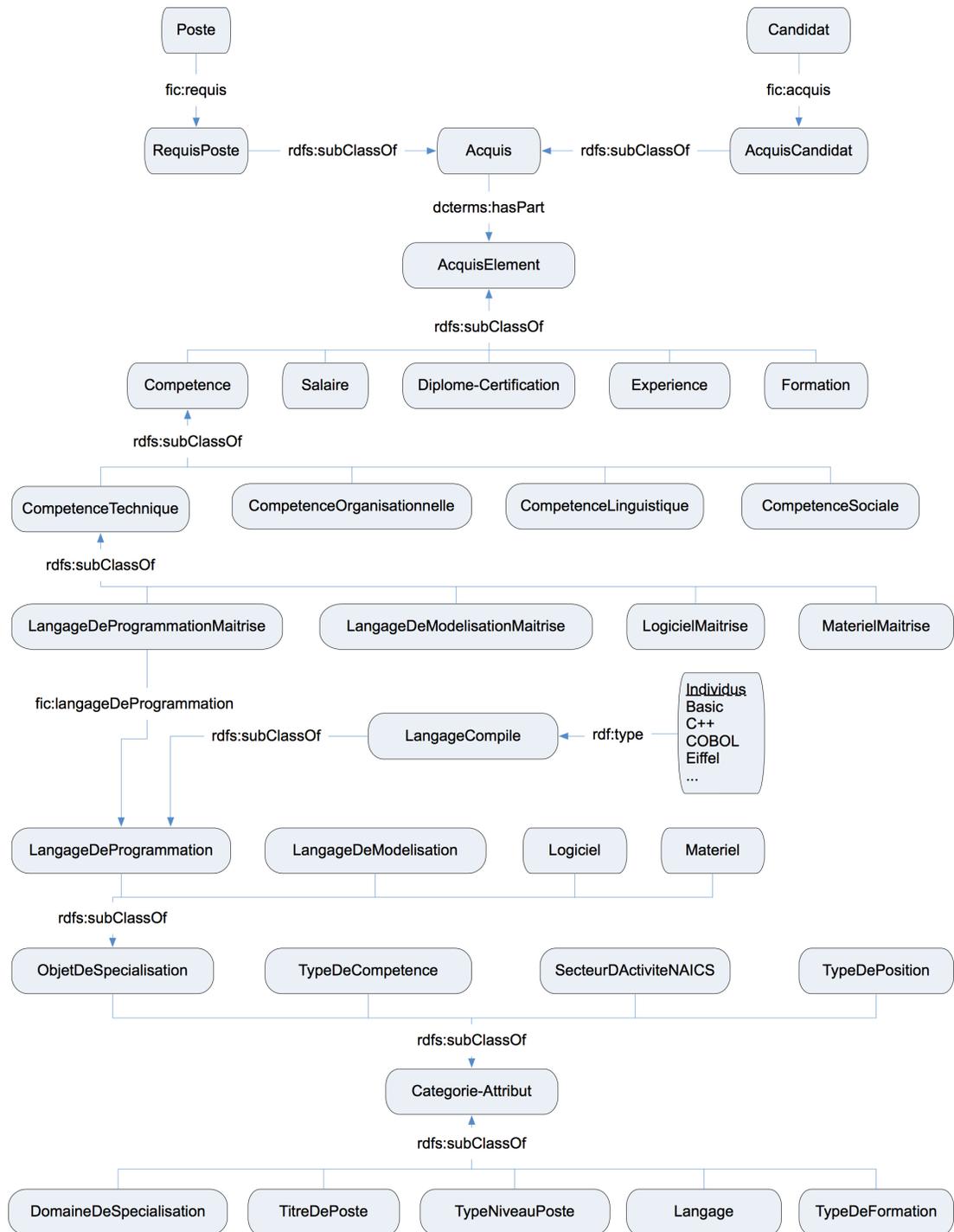


Figure 24 - Aperçu simplifié des concepts et relations définissant les concepts d'Acquis de candidat et de Requis de poste

Pour évaluer la similarité sémantique entre des requis de poste et des acquis de candidat en vue d'associer automatiquement les candidats et les postes selon une adéquation optimale, Combine utilise des algorithmes inspirés des travaux de Leila Yahiaoui (Yahiaoui et coll., 2006) et Christian Bizer (Bizer et coll., 2005). Ces algorithmes calculent des coefficients de similarité entre les concepts qui partagent une même sémantique (ontologie) en regardant leur distance hiérarchique.

Pour illustrer ce procédé, considérons la description RDF d'une offre de poste (Figure 25), ainsi que les descriptions RDF de deux candidats (Figure 26 et Figure 27) :

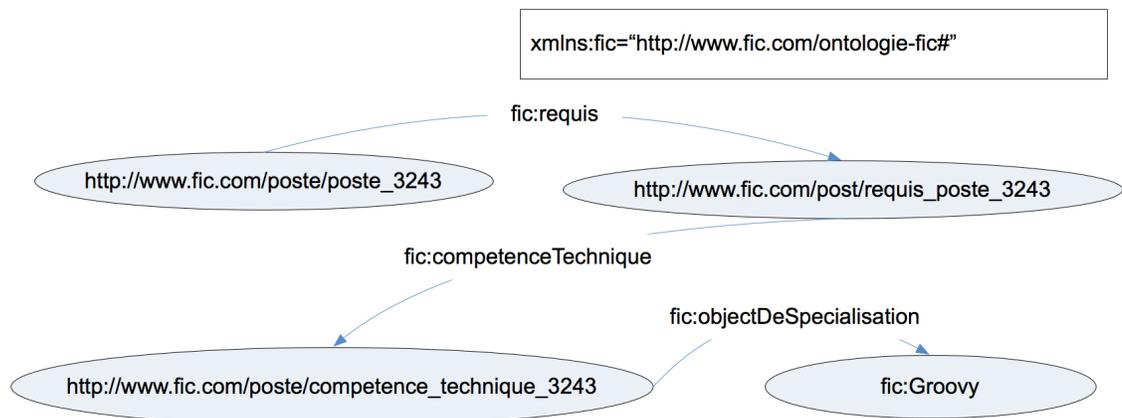


Figure 25 - Extrait d'un graphe RDF décrivant une offre de poste

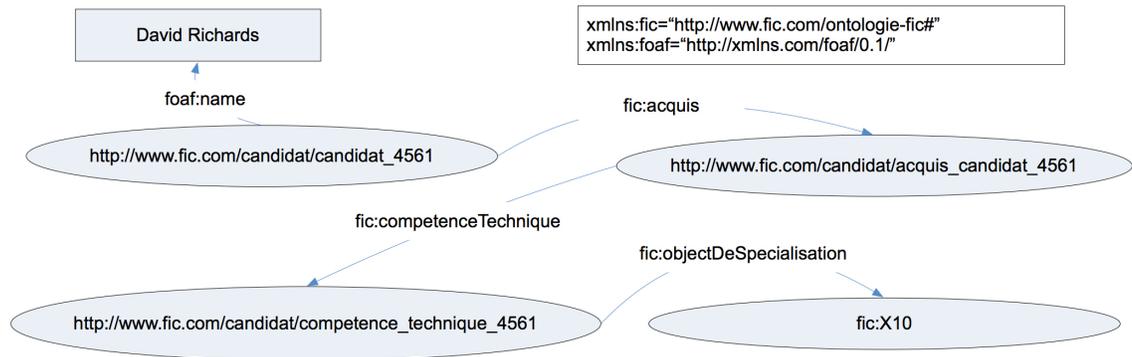


Figure 26 - Extrait d'un graphe RDF décrivant les acquis d'un premier candidat fictif

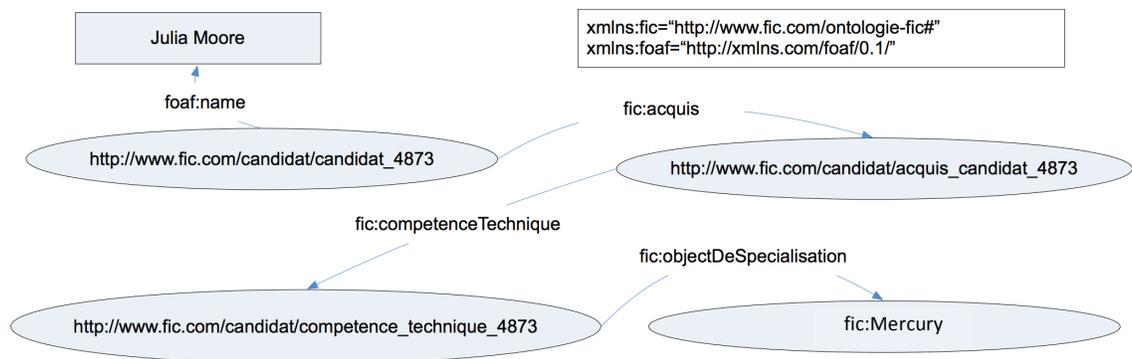


Figure 27 - Extrait d'un graphe RDF décrivant les acquis d'un second

Sur la base d'une requête de la part d'un recruteur, le coupleur sémantique évaluera la proximité sémantique entre chacun des éléments d'acquis de l'offre de poste et des éléments d'acquis correspondants des deux candidats. Cette évaluation de la correspondance des acquis pourrait inclure les compétences techniques requises versus les compétences techniques acquises, les formations requises versus les formations suivies, les expériences requises versus les expériences acquises, etc. Si l'on regarde maintenant l'objet de spécialisation du requis de poste et ceux des acquis des deux candidats, on peut voir, en considérant la portion de l'ontologie FIC dressant la hiérarchie des objets de spécialisation (Figure 28), que l'objet de spécialisation du premier candidat (X10) est sémantiquement

plus proche de l'objet de spécialisation du requis du poste (Groovy) que l'objet de spécialisation du second candidat (Mercury) puisque X10 et Groovy partagent tous deux un même concept parent, soit la classe des langages de programmation basés sur Java.

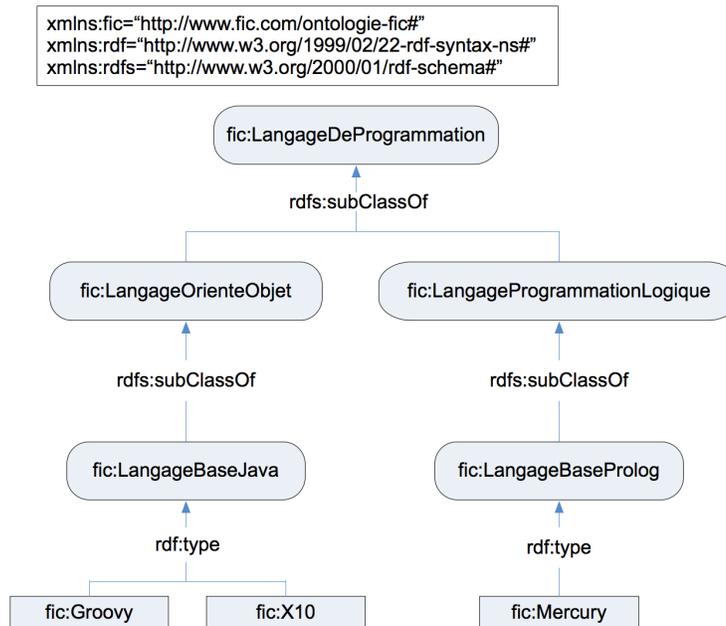


Figure 28 - Portion de la taxonomie des objets de spécialisation (extrait de l'ontologie FIC) montrant la distance conceptuelle entre X10, Mercury et Groovy.

La somme des scores attribués à chaque élément d'acquis évalué donnera finalement le score total des acquis d'un candidat en regard au requis du poste pour lequel la requête de couplage fut lancée. Cependant, il sera toujours possible pour le recruteur de moduler l'importance de chacun des éléments dans le calcul du score total : avant d'exécuter le couplage, Combine demande à l'utilisateur (via une boîte de dialogue ; voir Figure 29) de configurer le poids relatif qu'il désire que chaque élément d'acquis ait dans la comparaison conceptuelle en vue du couplage – le poids étant un facteur par lequel le coupleur multiplier le score de chaque type d'élément comparé, leur donnant ainsi différentes importances dans la sommation du score total.

Poids* accordé à chaque critère :

| | |
|---------------------------|---|
| Hot List | 1 |
| Domaine de spécialisation | 1 |
| Secteur d'activité | 1 |
| Technologies maîtrisées | 1 |
| Proximité géographique | 1 |

* Le poids est un nombre indiquant l'importance d'un critère dans le couplage automatique du requis des postes aux acquis des candidats.

Nombre de postes par candidat : 3

mémoriser les configurations

Figure 29 - Boîte de configuration du coupleur requis-acquis.

Puis, en dernière instance, le coupleur ordonnera les candidats en fonction du score total de chacun, sélectionnera les meilleurs (en raison du nombre demandé par le recruteur), et annotera ces candidats à la fiche du poste pour lequel la requête de couplage fut lancée (voir Figure 30). Aussi, il annotera réciproquement ce poste à la fiche de chacun des candidats sélectionnés (voir Figure 31).

| | | |
|--|--|---|
| <input type="checkbox"/> CIO chez Videotron (source : ) <input type="button" value="Proposer Candidats"/> | | |
| Information sur le poste Début du mandat : 2009-11-10 Fin du mandat : 2015-09-23 | Information sur la demande État : Actif - Prospection + Date d'ouverture : 2009-10-01 Date de dépôt : 2009-10-20 | Requis du poste Technos : Axure et MS Project + Secteur d'activité : télécommunication Niveau du poste : senior |
| Candidat(s) assigné(s) Serge Laroche, Julie Sauge, et Bob Taylor (afficher la liste, transmettre la liste) | | |

Figure 30 - Fiche de poste avec des candidats assignés par le coupleur

Bob Taylor / Architecte d'informations chez Bell / Consultant FIC + Actions sur la fiche

| Coordonnées | Acquis du candidat | Annotations et documents | Poste(s) ouvert(s) attribué(s) |
|---|---|---|--|
| 1-514-987-6325 SMS bob.taylor@videotron.ca Profil(s) web : LinkedIn | Technos : Azure et Enterprise Architect + Secteur d'activité : télécommunication | CV : cv.doc + Référence(s) : Référence 01 + Notes d'entrevue : NotesEntrevue 01 + | CIO - Videotron (afficher la liste, transmettre la liste) |

Figure 31 - Fiche de candidat avec un poste ouvert attribué

Notons pour clore cette démonstration que le couplage fonctionne aussi bien dans le sens inverse : le recruteur peut tout aussi bien lancer la requête de couplage à partir de la fiche d'un candidat pour trouver, et annoter automatiquement à sa fiche, les postes ouverts qui correspondent le mieux à son profil.

Par ailleurs, pour décrire les éléments d'acquis utilisés par le coupleur l'utilisation de relations lexicales de type méronymique³¹ et holonymique³² fut une autre stratégie déployée pour faciliter l'évaluation de l'adéquation entre candidatures et demandes de personnel (les relations lexicales consistent en des termes dénotant des rapports sémantiques entre des concepts tels que les expressions « hasPart » ou « related » par exemple). C'est pourquoi associer l'ontologie FIC à un vocabulaire explicitant de telles relations lexicales, comme SKOS notamment, devrait faciliter l'évaluation de la similarité sémantique entre les requis de poste et les acquis de candidat puisqu'elle s'effectuera non seulement sur la base des rapports taxonomiques hiérarchiques (de type hyponymique/hyperonymique ou classe/sous-classe) entre les concepts décrivant les postes et les candidats, mais aussi sur la base de leurs rapports transversaux et compositionnels. L'exemple du Ajax évoqué plus haut serait un cas de figure où une telle évaluation méronymique serait de mise : HTML, CSS, XML, DOM, JavaScript et XMLHttpRequest ne sont pas des sous-classes du concept

³¹ Une relation lexicale de type méronymique dénote un rapport partitif dans le sens de la partie au tout entre un terme A et un terme B. Ainsi, on dira que le terme « roue » est un méronyme du terme « automobile ».

³² À l'inverse de la relation lexicale de type méronymique, la relation d'holonymie dénote un rapport partitif dans le sens du tout à la partie entre un terme A et un terme B. Ainsi, alors du terme « automobile » qu'il est un holonyme du terme « roue ».

Ajax (HTML n'est pas *une sorte* d'Ajax), ils en sont des composantes (HTML *fait partie* d'Ajax).

D'une perspective de la conventionalité lexicale, et telle que le suggère la vaste majorité des méthodologies de modélisation de vocabulaire contrôlé, l'élaboration de l'ontologie FIC a pris avantage de la réutilisation de vocabulaires communs déjà existants. Ceci devrait avoir pour résultat d'augmenter la communicabilité de l'application avec les dépôts RDF externes qui emploient en tout ou en partie ces vocabulaires. En d'autres termes cette récupération conceptuelle de l'ontologie FIC permettra à Combine de pouvoir intégrer tel quel des données RDF en provenance de systèmes externes utilisant les terminologies réutilisées, puisque les données extraites seront codées selon exactement les mêmes concepts. Concrètement, le développement de l'ontologie FIC a donc identifié et implémenté quelques uns de ces vocabulaires répandus et la sélection de ceux-ci a été faite en regard de la manière dont chacun capturerait des pans conceptuels de l'ontologie en leur fournissant des concepts primitifs³³ ou synonymiques.

Il en résulte que l'ontologie FIC réutilise sept vocabulaires génériques et bien standardisés.

Le premier vocabulaire utilisé est la spécification *Friend of a Friend* (FOAF). Il est vraisemblablement le plus populaire des vocabulaires RDF pour décrire de manière générale des personnes, des groupes, des organisations, les relations entre ceux-ci, de même que des ressources générales telles que des comptes, des documents, des informations de contact, etc. Ce standard RDFS, basé à <http://xmlns.com/foaf/0.1/>, fut utile pour définir les acteurs du domaine.

Une autre terminologie majeure employée est le *Dublin Core Metadata Initiative Metadata Terms* (DCMI Metadata Terms; ou simplement dcterms) qui a pour URI

³³ Un concept primitif est un concept qui n'a d'autre concept parent dans une ontologie que celui le plus générique de « chose » subsumant tous les autres concepts.

<http://purl.org/dc/terms/>. Ce vocabulaire est une version plus développée du très populaire DCMI Element Set (<http://purl.org/dc/elements/1.1/>) qui rassemble une liste de 15 relations dédiées à la description de ressources documentaires et informationnelles. Le schéma de métadonnées Dublin Core fait l'objet d'une norme internationale (ISO 15836) depuis 2002 et aida le projet avec le balisage de plusieurs types d'items périphériques aux objets d'information que sont les descriptions, les notes, les commentaires, les besoins, etc., bref, toutes les ressources documentaires relatives aux acteurs et descriptions de poste.

Tenant pour *Simple Knowledge Organization System*, SKOS fut un autre vocabulaire pratique, surtout en ce qui a trait au support du couplage sémantique des requis de poste aux acquis de candidat tel évoqué plus haut. Comme SKOS est un vocabulaire qui propose des concepts servant à construire des taxonomies et des thésaurus, il fournit des relations conceptuelles explicites utiles pour définir les différences et similarités sémantiques entre les éléments constitutifs des requis de poste et des acquis de candidat. L'URI de SKOS est le <http://www.w3.org/2004/02/skos/core>.

À la croisée des trois premières ressources ontologiques, SIOC (*Semantically-Interlinked Online Communities*) est construit à partir de FOAF, Dublin Core et SKOS et apporte des concepts qui supportent la définition de ressources génériques relatives aux communautés en ligne telles que les wikis, les blogs, les articles de blog (posts), etc. Par rapport à l'ontologie FIC, SIOC offre des concepts primitifs servant à définir les objets communicationnels et médiatiques. SIOC se trouve à l'URI <http://rdfs.org/sioc/ns>.

iCalendar est un format de données calendrier utilisé pour échanger des informations à propos de demandes de réunion, de tâches et d'évènements. Étant donné que les emplois sont en soi des évènements, que les demandes de personnel ont des dates d'ouverture et de complétude, et que les entrevues se tiennent sur une certaine plage horaire, le vocabulaire iCalendar (ou iCal) est tout désigné pour décrire n'importe quel objet qui possède une dimension temporelle. De plus, le standard iCalendar, qui a

<http://www.w3.org/2002/12/cal/ical> pour URI, est implémenté dans une variété étendue de systèmes comme Apple iCal, Lotus Notes, Mozilla Calendar, Windows Calendar, Microsoft Outlook et Google Agenda, pour ne nommer qu'eux. Ce standard est aussi compatible avec la norme hCalendar de Microformat.

Dans le but de faciliter la communicabilité des informations sur les contacts, le vocabulaire du VCard (<http://www.w3.org/2006/vcard/ns>) fut aussi employé. Bien que FOAF couvre déjà la quasi-totalité des informations de contact du standard VCard, l'alignement des concepts redondants entre les deux vocabulaires devrait permettre à Combine de gagner en interopérabilité.

Puis finalement, l'ontologie FIC enrôle une taxonomie des secteurs d'activité industrielle appelée *North American Industry Classification System* (NAICS), qui est disponible en format OWL à l'adresse <http://www.srdc.metu.edu.tr/ubl/contextOntology/naics.owl>. Comme la taxonomie NAICS propose des catégories industrielles qui regroupent les secteurs d'activité en fonction de leurs proximités sémantiques, elle sera particulièrement valorisée par le coupleur requis/acquis.

3.5.2 Mécanismes d'échange de données

Tel que vu plus tôt, l'échange de données entre les systèmes utilisés par les recruteurs FIC est exécuté manuellement, par les recruteurs eux-mêmes. C'est pourquoi Combine fut développée dans l'optique de devenir une sorte de tierce application universelle (méta-application) qui automatiserait au maximum l'échange et la gestion des informations basses (voir Figure 32). Cette exigence de l'application interpelle l'ontologie FIC et ces triplets RDF correspondant en ce qu'ils sont conçus pour faire de mieux : servir de référent abstrait commun. Voyons maintenant plus en détails le problème de l'échange de données exposées au chapitre précédent en vue d'expliquer la solution proposée par Combine en cette matière.

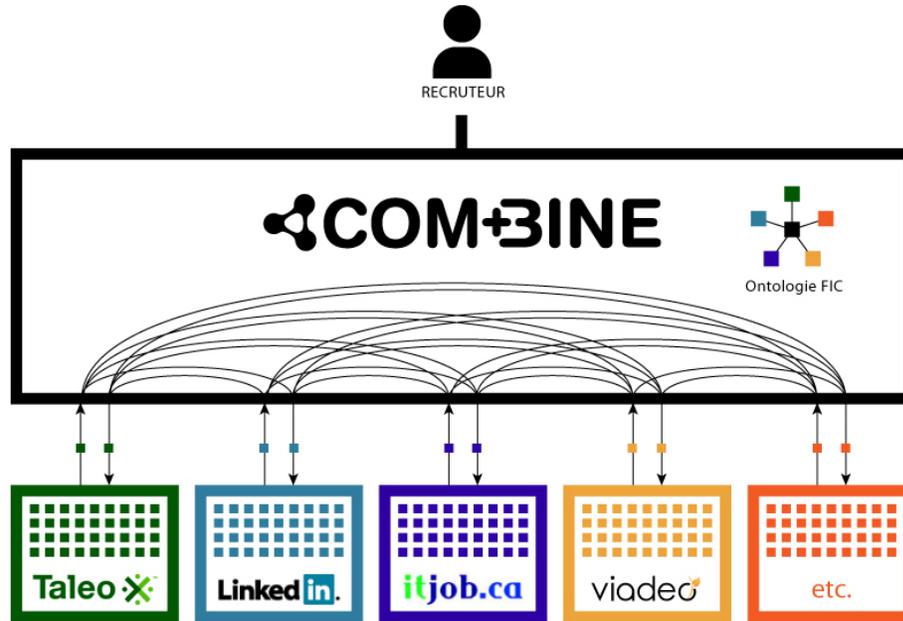


Figure 32 - Combine en tant que tierce application pouvant automatiser l'échange de données.

Dans le scénario idéal d'un monde électronique pleinement sémantisé, toutes les informations relatives aux candidats (coordonnées + informations professionnelles) seraient accessibles via des requêtes SPARQL, directement sur les dispositifs personnels en ligne des candidats ou sur des dépôts ouverts de données. Mais la réalité est que ces ensembles de données RDF ouverts et accessibles sont pratiquement inexistants à l'heure actuelle. Pire, les principaux acteurs qui font des affaires sur la base d'une médiation de telles informations sont par principe contre un tel écoulement des données à l'extérieur de leurs bases de données. Ce fait, que l'on appelle le syndrome de la séquestration des données, constitue *de facto* l'un des principaux obstacles au développement du SW et de la notion d'interopérabilité en générale.

3.5.2.1 Principe du screen scraping et du furetage automatisé

Tant que les modèles d'affaires traditionnels des entreprises électroniques n'évoluent pas significativement donc – de sorte à embrasser pleinement le paradigme de l'*open* –, on ne pourra jamais s'attendre de ceux qui tirent des recettes de la consignation et de la mise à disposition de contenus qu'ils donnent libre accès à leurs données brutes. Les usagers, tels les recruteurs FIC, ne sont normalement autorisés qu'à manipuler les pages HTML construites à partir de ces données brutes. Voilà pourquoi Combine a été dotée de fonctionnalités de *screen scraping* pour extraire automatiquement des données encadrées dans des pages web et de *furetage automatisé* pour emplir des formulaires web ou pour exécuter les *screen scrapers*, automatiquement. Figure 33 schématise en ce sens les principes généraux de ces deux techniques telles qu'elles sont implémentées dans Combine.

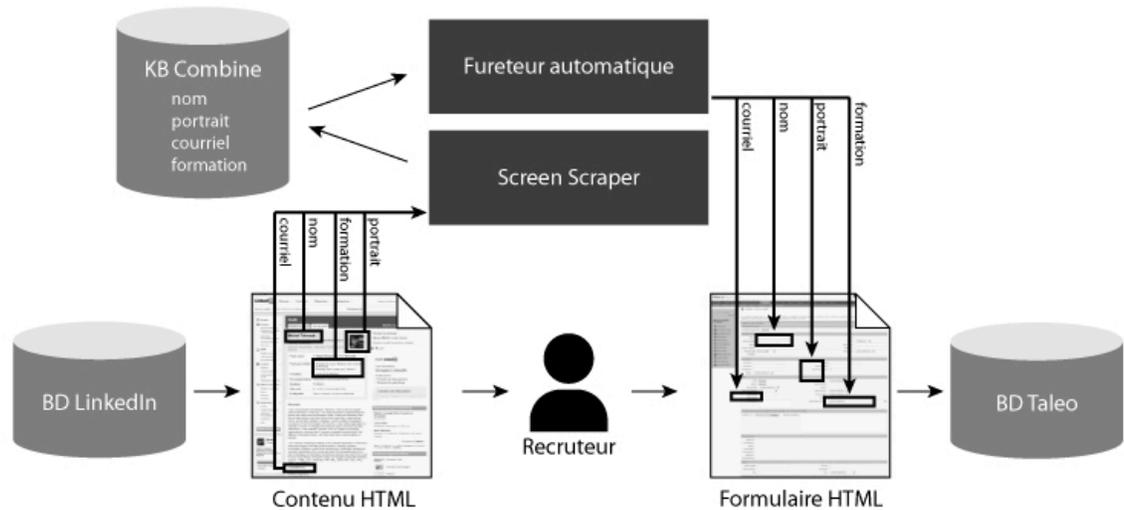


Figure 33 - Schéma illustrant les principes du *screen scraping* et du furetage automatisé.

Sur le code HTML d'une page web, un fureteur-client augmenté (Firefox + Greasemonkey) applique du code JavaScript qui extraira des éléments d'information spécifiques de la page pour les envoyer ensuite dans la base de connaissances de Combine. De là, un fureteur automatisé (fureteur web commandé par des macro-commandes) pourra se saisir des

informations extraites, atteindre – par lui-même – une page spécifique contenant un formulaire web, remplir les champs de ce formulaire avec les informations appropriés, puis soumettre le formulaire. Ce n'est clairement pas une solution aussi élégante et robuste que d'utiliser uniquement des requêtes SPARQL, mais cette façon de faire a au moins l'avantage d'être une manière autonome d'acquérir et d'exporter de l'information. Toutefois, le *screen scraping* est une pratique polémique qui requière quelques réflexions préliminaires avant de s'y adonner.

De la perspective business des concessionnaires de contenus, le *screen scraping* apparaît comme une forme d'ingénierie inversée, et par extension un acte de piratage. La plupart des contrats d'utilisation stipulent précisément que l'information présentée à l'utilisateur ne peut être copiée ou utilisée commercialement.

Mais encore, les contrats d'utilisation sont souvent très agressifs et alambiqués en regard de ce que l'utilisateur a le droit de faire, mais surtout en regard de ce que l'utilisateur n'a pas le droit de faire avec les informations contenues dans les pages générées et dispensées par les serveurs de ces organisations à but très lucratif. Les contrats des sites de réseautage comme LinkedIn ou Facebook laissent entendre que tous les utilisateurs qui utilisent leurs sites cèdent *de facto* aux compagnies propriétaires de ces sites les droits des contenus qu'ils y consignent. Mais, tel que le posent clairement l'Office de la propriété intellectuelle du Canada et le U.S. Copyright Office, les *informations factuelles* ne peuvent être en aucun cas sujets de droit d'auteur (seulement la manière dont elles sont présentées le peuvent). Considérant que les coordonnées et les informations professionnelles d'une personne sont justement des informations factuelles, les contrats d'utilisation des sites de réseautages apparaissent pour le moins abusifs à ce propos, pour ne pas dire légalement invalide.

La seule véritable préoccupation légale se situe plutôt au niveau de la collecte et de l'usage d'informations personnelles. Au Canada la *Loi sur la protection des renseignements personnels et les documents électroniques* stipule que : « Toute personne doit être informée

de toute collecte, utilisation ou communication de renseignements personnels qui la concernent et y consentir» (article 4.3 ; Troisième principe). C'est donc dire qu'en principe, pour pouvoir légalement collecter et utiliser à des fins commerciales des informations personnelles à propos de candidats potentiels, le recruteur-collecteur est tenu d'informer correctement les personnes ciblées de ce qu'il a l'intention de faire comme usage de ces informations et d'obtenir leurs accords. En fait la loi précise qu'un tel accord ne devient vraiment impératif que lorsque les informations collectées se voient effectivement utilisées. En d'autres mots, une collecte d'informations personnelles sans accord des personnes visées ne devient techniquement illégale qu'à partir de l'instant où ces informations sont utilisées pour certaines fins. Quoi qu'il en soit, il est déjà dans les pratiques des recruteurs FIC de demander de tels accords; sur les réseaux sociaux, ils font généralement cette demande par le biais du message généralement inclus dans les demandes de contact.

Pour être justes, nous devons noter que certaines clauses des contrats d'utilisation des sites de réseautage reconnaissent ces dispositions légales. Par exemple le contrat de LinkedIn stipule que l'utilisateur ne peut « recueillir, utiliser ou transférer toute information, y compris, mais sans limitation, des informations personnelles identifiables obtenues de LinkedIn *sauf comme expressément permis* dans le Contrat ou *par le propriétaire de telles informations* » (mes italiques). Ceci étant, il n'en reste pas moins que ces mêmes contrats semblent interdire du même souffle l'usage de tout moyen visant à copier cette information. Toujours selon le contrat de LinkedIn, il est défendu d'« utiliser des logiciels ou appareils manuels ou automates, robots de codage ou autres moyens pour accéder, explorer, extraire ou indexer toute page Internet sur le site ».

D'un strict point de vue communicationnel, un site de réseautage social est définitivement plus un dispositif de communication interpersonnelle qu'un dispositif de communication de masse où le propriétaire du médium diffuse de manière unidirectionnelle des contenus originaux. Sous cet angle, le contrôle que tente d'imposer les propriétaires de sites de

réseautage sur ce que leurs usagers ont le droit ou non de faire avec les informations échangées grâce à leur médiation technique est comparable à une compagnie de téléphone qui voudrait interdire l'enregistrement des conversations téléphoniques qu'elle supporte ou à un service postal qui essaierait de proscrire la photocopie des lettres qu'il achemine.

En mettant de côté ce débat éthique, il n'en demeure pas moins le fait que les sites comme LinkedIn sont très sensibles aux comportements de navigation non-humains et tendent à bannir tout utilisateur dont le compte a été utilisé par un agent logiciel. La manière dont Combine « racle » les pages web avec ses *screen scrapers* prend au sérieux un tel état des choses et a été pensé se sorte à fonctionner de façon très discrète et parcimonieuse. Concrètement, Combine utilise Greasemonkey, une extension Firefox qui permet l'application de code JavaScript sur des pages web en ligne. L'application de code sur une page permet alors d'en altérer la présentation et/ou d'en manipuler le contenu. Comme cette opération se déroule du côté client (sur l'ordinateur de l'utilisateur), il en résulte qu'elle est totalement intraçable et non-intrusive par rapport aux serveurs d'où originent les pages traitées. En d'autres mots, le *screen scraping* de Combine n'utilise pas de robot (*web crawler*) qui explorerait automatiquement et aléatoirement le web social en vue de collecter un maximum d'informations, il se concentre seulement de supporter le recruteur dans sa gestion des informations en provenance des sites et contacts qu'il connaît déjà ; un peu à l'image d'un assistant personnel qui, une fois son maître-recruteur entré en contact avec un candidat potentiel, se charge des détails techniques relatifs à l'acquisition et au suivi des informations sur ce candidat, en utilisant un fureteur web, comme n'importe quel usager humain.

Étant donné que nous avons choisi d'utiliser Greasemonkey pour appliquer le codage d'extraction, les *screen scrapers* de Combine sont écrits en JavaScript. Ils utilisent HTML

DOM³⁴ pour récupérer l'information et XMLHttpRequest³⁵ pour envoyer cette information au serveur de Combine. À l'avant-plan de l'application (*front end*), cette solution est implémentée en tant que simple bouton que Greasemonkey ajoute à toutes les pages web qu'il reconnaît – par son URL – comme étant une page « raclable », bouton duquel le recruteur pourra déclencher le codage d'extraction comme tel (voir Figure 34)

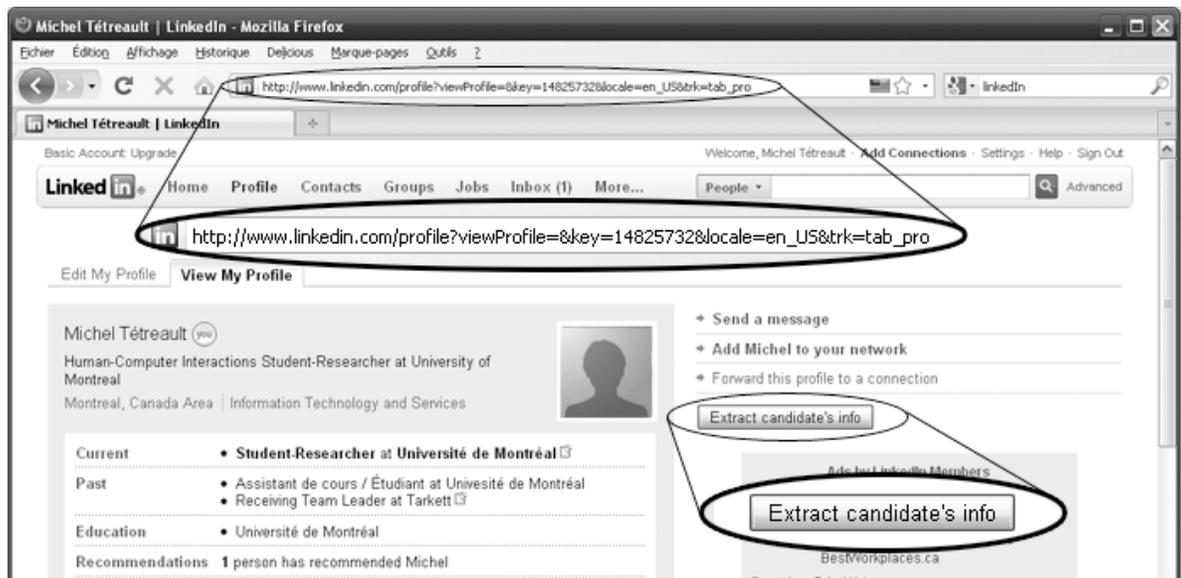


Figure 34 - Capture d'écran montrant l'intégration du bouton d'extraction sur un profil LinkedIn (l'adresse URL est magnifiée pour montrer que le bouton a été effectivement intégré sur une page en ligne).

En ce qui concerne le furetage automatisé, qui sera utilisé principalement pour faire de la saisie de données automatique (emplir des formulaires web par exemple), il est compris dans Combine sous forme d'un fureteur web qui peut être manipulé par des macro-fonctions programmables (iMacros d'iOpus). Le principe derrière le furetage automatisé de Combine est relativement simple. À partir des connaissances de la base (triplets RDF) une

³⁴ DOM est l'acronyme de *Document Object Model*, un modèle qui sert à décrire et manipuler des langages de balisage tels que HTML et XML.

³⁵ XMLHttpRequest est un protocole dédié à l'échange de données entre une page web et un serveur.

requête SPARQL construira un fichier XML dont les balises (structure sémantique du document) correspondent aux champs d'entrée spécifiques d'un formulaire web ciblé. Combine active ensuite la macro appropriée du fureteur automatique qui se saisira alors du fichier XML, utilisera les valeurs des champs du document XML pour compléter les champs du formulaire web, puis soumettra le formulaire. Pour permettre aux recruteurs d'utiliser cette fonctionnalité, l'interface graphique de Combine offre différentes options d'exportation sur chaque objet ou groupe d'objets contenu dans sa base (les plus développés sont principalement celles liées aux offres de poste et aux profils de candidat). Par exemple, un recruteur pourrait choisir de publier une certaine offre de poste sur plusieurs sites d'affichage d'emplois à la fois. Pour ce faire, il n'aurait qu'à cliquer sur un bouton à cet effet directement sur la fiche du poste (ou encore dans le menu des actions sur les items). Sur un clic de ce bouton, Combine ouvre une boîte de dialogue (voir Figure 35) qui permet de sélectionner les sites sur lesquels on désire afficher le poste. Sur le clic du bouton « afficher », le fureteur automatique se chargera d'atteindre et d'emplir les formulaires appropriés. Avec cette composante de Combine, toute la dissémination des offres de poste, ou encore l'enregistrement des fiches candidat dans le SI maison, pourraient même être faites de manière complètement automatique, en arrière-plan – dans la mesure où Combine est configurée comme tel. Elle pourrait aussi être utilisée pour publier des offres de poste sur les statuts des pages profil des leurs recruteurs assignés. Cette fonctionnalité restaurerait donc puissamment le respect des principales tâches négligées des processus standardisés FIC – analysées précédemment.

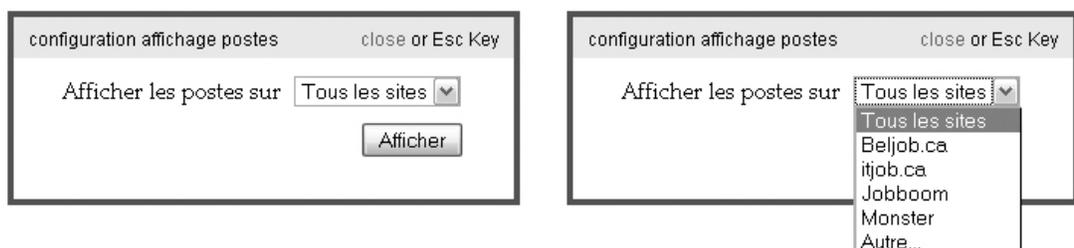


Figure 35 - Boîte de dialogue de Combine pour configurer les destinations d'affichage de poste.

3.5.2.2 Mise à jour des données

Les mises à jour des profils de candidat sont aussi exécutables par le fureteur automatique. Pour ce faire Combine tire profit des fils RSS dispensés par les sites de réseautage comme déclencheur pour lancer les mises à jour. LinkedIn, par exemple, envoie un item RSS à chaque fois qu'un contact modifie son profil. Naturellement, à cause des politiques de séquestration des données, ces types d'envoi RSS ne sont pas très verbeux, mais comme ils contiennent normalement les URLs des profils modifiés, Combine peut signaler au fureteur automatisé d'atteindre les adresses en question pour que ce dernier racle les pages correspondantes et rafraichisse la base de connaissances.

3.5.2.3 Intégration au Web Social

Dans l'esprit de faciliter l'échange de données entre Combine et les plateformes de type Web Social, nous avons pensé développer une application sociale (avec OpenSocial) qui permettrait aux recruteurs de publier en détails, et de manière interactive, toutes les offres de poste dont ils sont tributaires, directement sur leur profil 2.0 (voir Figure 36 pour un aperçu de l'interface). Des candidats potentiels pourraient alors appliquer aux postes affichés à même l'interface de cette application. Ils auraient aussi l'option de partager une offre de poste spécifique via le service de leur choix (Courriel, Twitter, Facebook, Google, etc.) ou encore de suivre les offres de poste de ce recruteur en s'abonnant à un fil RSS.

Jim Sample
Recruiter at FIC
Montreal, Canada Area

→ Contact Jim Sample
→ Add Jim Sample to your network

Current • Recruiter at FIC

Past • Assistant de cours / Étudiant at Université de Montréal
• Receiving Team Leader at Tarkett

Education • Université de Montréal

Recommended 👍 1 person has recommended Jim

Connections 🌐 9 connections

Industry Information Technology and Services

Websites • 11 of M I Communication Dep

Public profile powered by: **LinkedIn**
Create a public profile: [Sign In](#) or [Join Now](#)

Job Offers **FIC**

- Project manager
- Business analyst
- PHP Programmer

Follow offers: [RSS](#) [RDF](#)

[View Jim Sample's full profile:](#)

Figure 36 - Capture d'écran de l'interface de l'application sociale Combine.

3.5.2.4 Importation et exportation des données

Finalement, Combine est équipée de dispositifs généraux d'importation et d'exportation. Il peut donc prendre et convertir (ou annoter) en RDF n'importe quelle sorte de documents structurés et non-structurés (XML, JSON, HTML, DOC, XLS, PDF, etc.). Symétriquement, il peut convertir ses données sémantiques dans tous ces formats et les transférer vers d'autres acteurs ou systèmes du domaine (collègues, directeurs de compte, clients, CRM, etc.).

3.5.3 Dispositifs de communication

Une autre des caractéristiques pratiques de Combine est l'usage qu'il fait des schèmes URI tels que les protocoles *mailto:* et *callto:* pour assister et documenter les communications des recruteurs. Avec le protocole *mailto:*, Combine peut fabriquer de courriels

personnalisés et envoyer automatiquement des messages spécifiques. Il peut en l'occurrence construire des courriels pré-formatés en fonction de thématiques récurrentes : offrir un poste, demander un CV à jour, requêter des références, etc. Un seul clic sur l'adresse courriel du candidat à contacter (ou sur une des options de courriels pré-formatés) et hop, le message est prêt à être envoyé (ou peut être envoyé automatiquement si Combine est configurée comme tel). Couplé à l'usage d'un système de téléphonie IP, le protocole *callto*: permettrait de lancer des appels téléphoniques simplement en cliquant sur le numéro de téléphone apparaissant à l'écran. En plus de sauver le temps normalement nécessaire à la composition du numéro, cette caractéristique permet de garder des traces de tous les appels placés, et même de les enregistrer. Cela devrait régler le problème du manque de temps des recruteurs pour documenter les communications avec les candidats.

3.5.4 Aperçu du déploiement du système

Figure 37 montre un aperçu graphique de toutes les caractéristiques de Combine que nous venons de présenter. On peut y voir les composantes principales du système et leurs interactions.

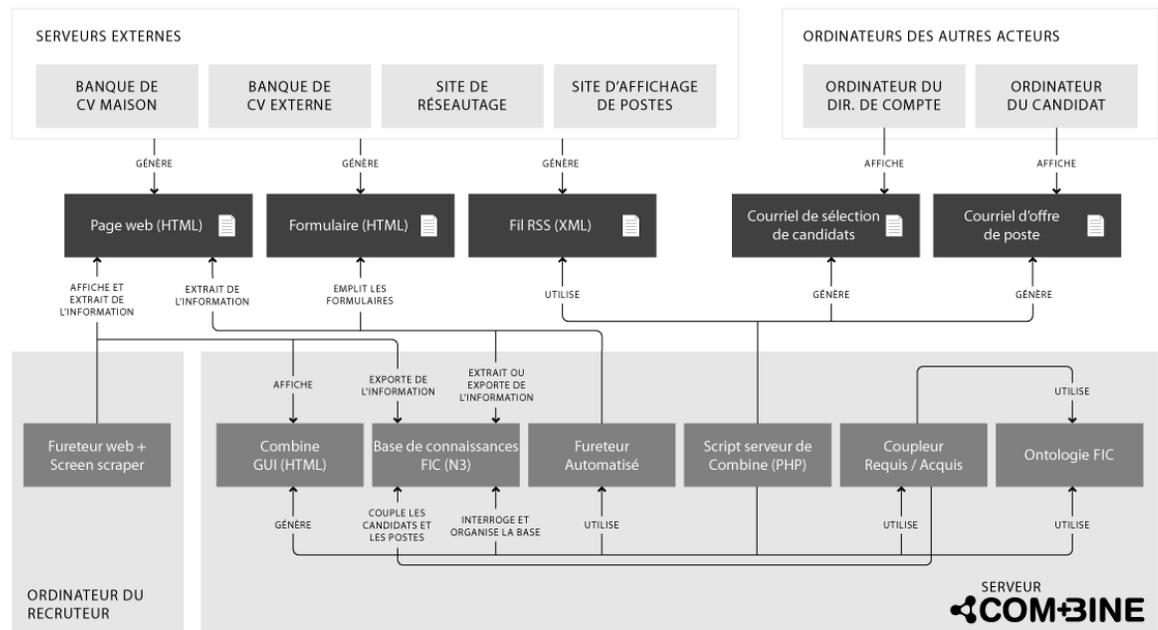


Figure 37 - Diagramme de déploiement de Combine.

3.6 Évaluation et appréciation du prototype de Combine

Tel que mentionné précédemment, un des objectifs principaux de cette recherche était d'évaluer l'utilité potentielle de l'application Combine et ses chances d'adoption en termes d'acceptation des usagers ciblés. Deux activités ont été déployées pour ce faire : la comparaison de diagrammes de séquence UML par rapport à des scénarios d'usage types pour évaluer quantitativement l'économie de charge de travail qu'apporterait Combine (donnant au final les portraits *avec* ou *sans Combine* de la manière dont les recruteurs interagissent avec leurs ressources électroniques), et puis l'administration d'un questionnaire aux membres de la CdP ciblée pour évaluer qualitativement leur réceptivité à la solution numérique que constitue notre application.

3.6.1 Évaluation de l'économie de charge de travail

Étant donné que l'activité d'analyse par comparaison des portraits d'interactions avec ou sans Combine fut plutôt complexe à réaliser, nous allons détailler ici seulement le scénario d'utilisation le plus critique en tant qu'il s'agit de l'activité centrale du processus de recrutement : la recherche de candidat. Le scénario en question est celui d'une recherche simple de trois candidats, de l'entrée de la demande de personnel dans le SI maison jusqu'à la remise d'une sélection de trois candidats au directeur de compte. La comparaison des diagrammes UML représentant ce scénario a montré qu'en bout de ligne Combine permettrait de réduire le nombre d'opérations directes avec les divers SIs utilisés d'un minimum de 15 ou 16 actions pour le scénario actuel d'une recherche « manuelle » sans Combine (voir Figure 38) à un maximum de 5 actions dans le cas projeté d'une même recherche qui emploierait Combine en mode semi-automatique, c.-à-d. dans un mode où combine demande au recruteur de valider chaque étape importante du processus de recherche (voir Figure 39). Ce nombre pourrait même tomber à 0 dans le cas d'une recherche avec Combine en mode pleinement automatisé (voir Figure 40) : la recherche

étant déclenchée automatiquement dès qu'une nouvelle demande de personnel est enregistrée dans le SI maison. Le diagramme de la Figure 38 présente ici un cas tel qu'il est effectivement pratiqué chez FIC, mais « idéal », en ce que les résultats de la première tentative de recherche sont présumés à la fois pertinents et suffisants et où le recruteur examine et sélectionne uniquement les trois premiers profils de candidat retournés par le SI maison. Les flèches en gras des diagrammes, qui ont été prises en compte dans la comparaison, représentent les actions directes du recruteur avec le système. Puisque le recruteur constitue lui-même une ressource comme une autre dans le processus de recrutement, les actions réflexives de ce dernier (choix, évaluation, ordonnancement, etc.) se devaient d'être aussi prises en compte. De fait, c'est surtout ces actions réflexives que l'application Combine permettrait au recruteur de déléster.

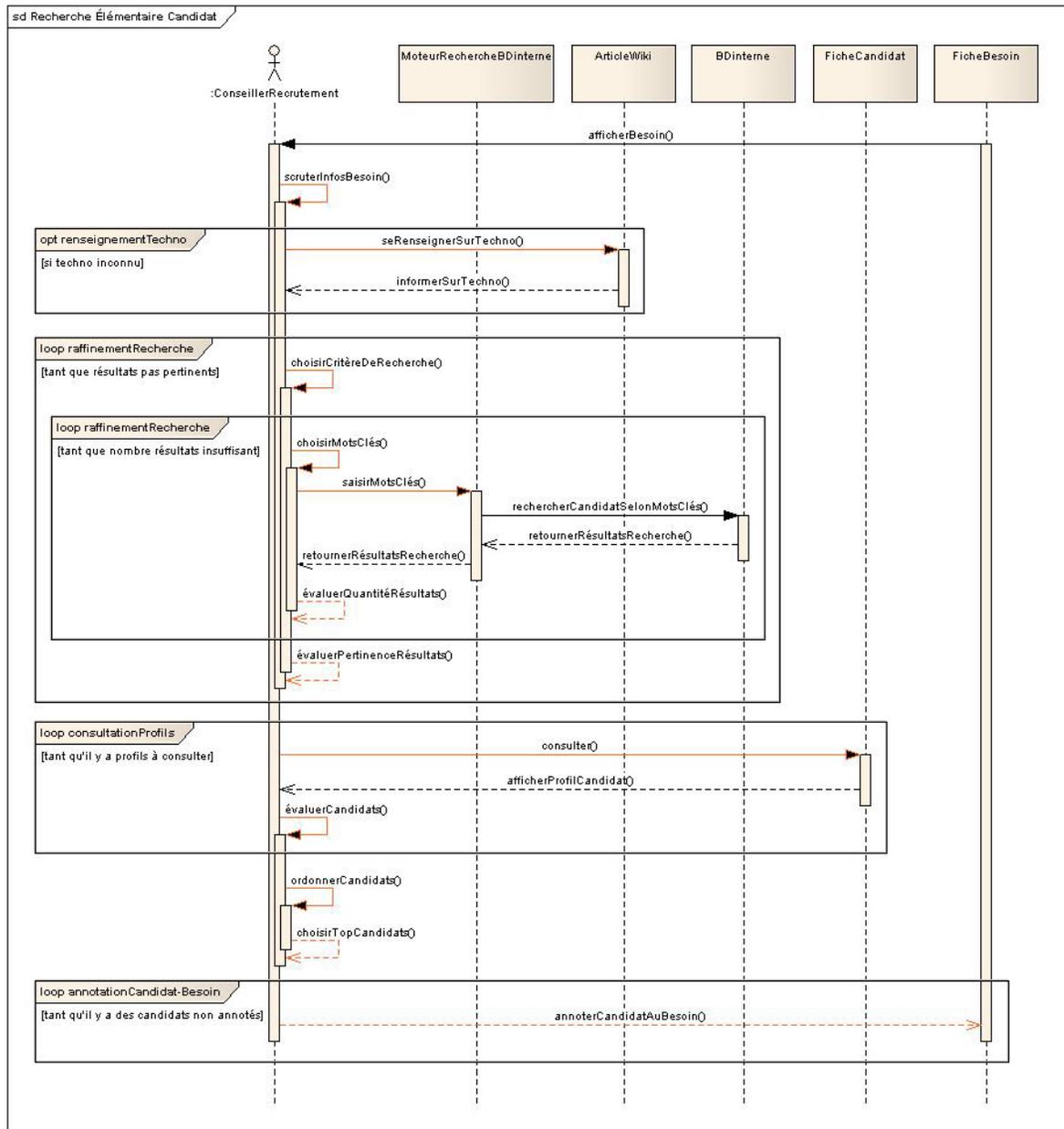


Figure 38 - Diagramme de séquence UML du cas d'une recherche simple de trois candidats telle qu'elle est pratiquée actuellement par les recruteurs FIC.

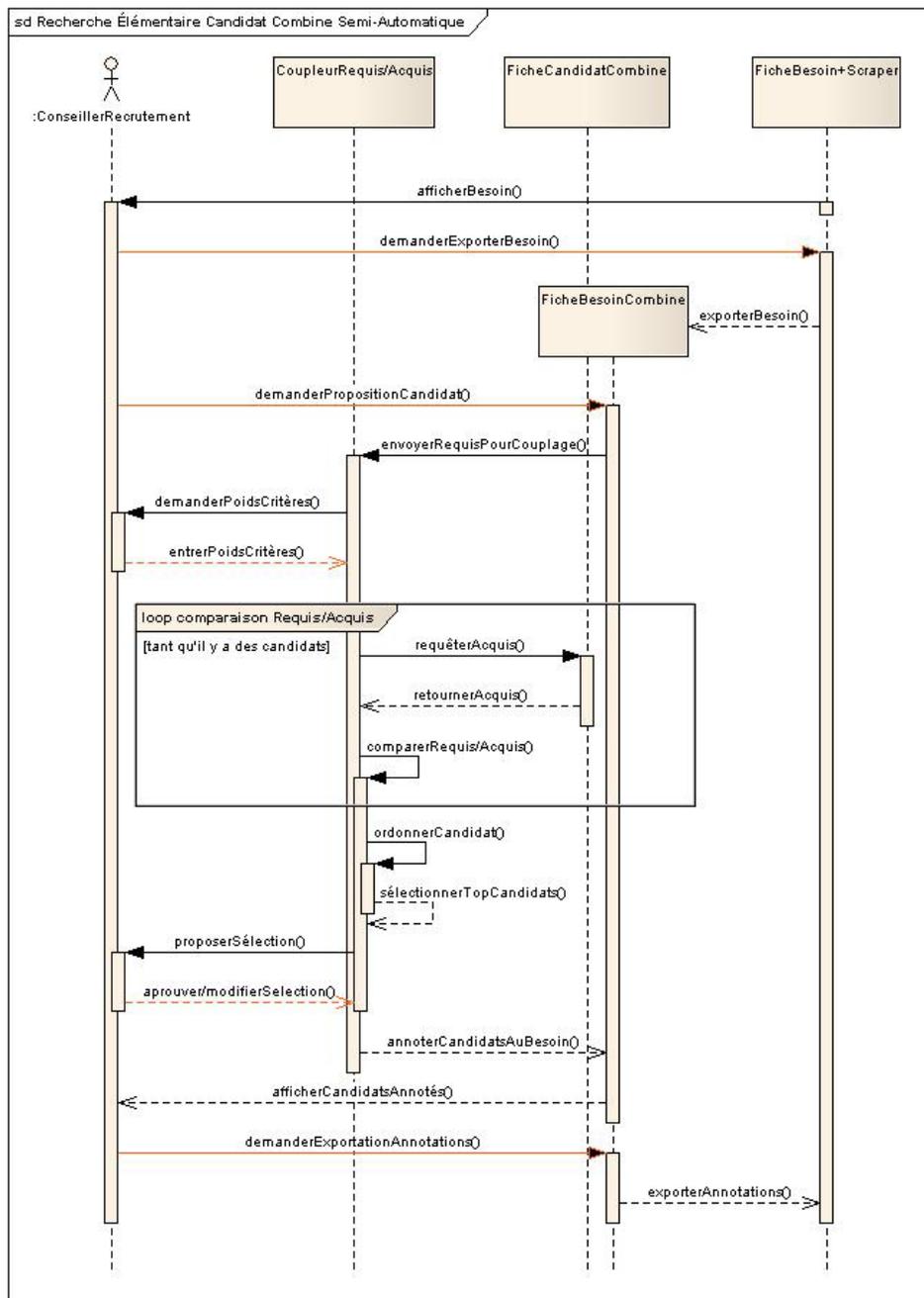


Figure 39 - Diagramme de séquence UML du cas d'une recherche simple de trois candidats telle qu'elle pourrait être pratiquée en utilisant Combine en mode semi-automatique.

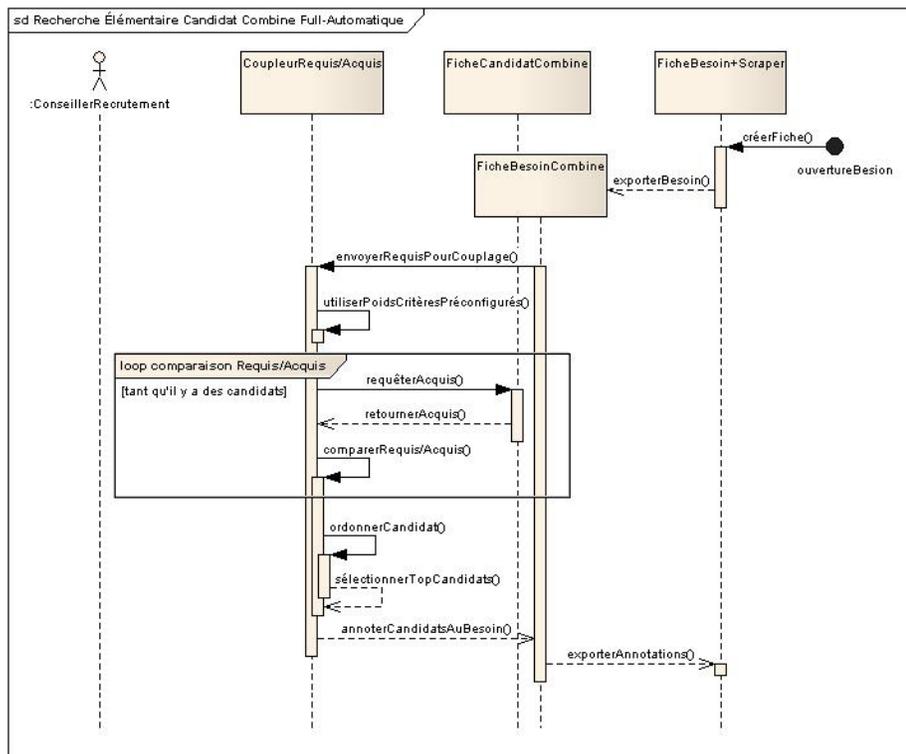


Figure 40 - Diagramme de séquence UML du cas d'une recherche simple de trois candidats telle qu'elle pourrait être pratiquée en utilisant Combine en mode automatique.

3.6.2 Évaluation de la réceptivité de la solution Combine par les utilisateurs ciblés

En ce qui concerne l'administration du questionnaire pour sonder l'appréciation et les perceptions des utilisateurs ciblés par rapport à Combine et ses fonctionnalités, les résultats obtenus furent, là aussi, plutôt encourageants.

Cette activité s'est déroulée comme suit : après avoir assisté à une présentation du prototype de Combine (présentation des caractéristiques + démonstration de quelques cas d'usage), les recruteurs FIC devaient compléter un court questionnaire (voir Annexe) pour recueillir leurs appréciations.

En réponse aux questions portant sur leur appréciation générale de la solution présentée, les recruteurs ont manifesté unanimement leur désir d'utiliser un tel système : ils ont tous répondu par l'affirmative à la question « Utiliseriez-vous une telle application ? ». Ce désir exprimé était principalement basé sur leur perception de l'utilité potentielle du système et de l'adéquation de ce dernier à leur réalité de travail.

Aux questions plus spécifiques, ils ont pu détailler ces perceptions. En ce qui a trait aux avantages perçus, nous avons pu remarquer que les plus récurrents étaient :

- Amélioration générale de la vitesse de travail
- Interconnexion des outils existants
- Amélioration générale de l'aisance de travail
- Accélération des recherches
- Accélération des tâches redondantes
- Fusion de l'information
- Amélioration de la mise à jour des profils de candidat
- Facilité de personnalisation
- Accélération de l'affichage de poste
- Accélération de l'entrée de données
- Adaptation de l'application au contexte de la consultation TI
- Centralisation des opérations

En réponse aux questions cherchant à recueillir les perceptions négatives de Combine, les recruteurs ont exprimé des appréhensions telles que :

- Crainte d'être remplacé par le système
- Crainte quant à la fiabilité du coupleur requis-acquis
- Doutes sur le fonctionnement de la base de connaissances
- Dédoublage de systèmes (avec le SI maison)

Sans vouloir minimiser ces dernières perceptions négatives, nous pouvons arguer que Combine a été conçue dans l'optique d'augmenter les compétences du recruteur, pas de les remplacer. En considérant que Combine fut conceptualisée plutôt comme un poste de pilotage qu'un simple dispositif automatisé, le jugement du recruteur demeurera toujours au cœur du système. En ce qui concerne les craintes de fiabilité du coupleur, tout ce que nous pouvons dire à ce point est que nous partageons ce souci et que nous testerons rigoureusement cette fonctionnalité lors des étapes subséquentes du développement du système (si jamais le projet se poursuivait).

Somme toute, les activités de validation menées pour évaluer la valeur ergonomique de Combine et son appréciation auprès de la clientèle cible indiquent de manière relativement éloquente qu'il s'agit bien là d'une application qui a des chances de succès en matière d'adoption et qu'elle permettrait effectivement d'optimiser les processus de travail du domaine d'activité ciblé.

4 Discussion du projet Combine

4.1 Développements futurs et intérêts périphériques du projet Combine

Les appréhensions des recruteurs recueillies lors de l'évaluation du prototype indiquent déjà des pistes de réflexion et des travaux futurs à faire sur l'application. Dans la mesure où le projet Combine se poursuivrait, nous voyons aussi plusieurs options qui seraient pertinentes de développer.

4.1.1 Développements futurs de Combine

Nous avons l'intention de produire des règles de transformation des connaissances qui pourraient considérablement enrichir la manière dont l'application acquiert, complète et assemble ses informations. Un langage de règles tel que SWRL (*Semantic Web Rule Language*) pourrait être utilisé pour ce faire. Ces règles permettraient aux logiciels d'inférences d'effectuer des déductions sur la base de connaissances dans le but de synthétiser ou d'analyser les objets d'information qui s'y trouvent.

Aussi, nous allons certainement chercher des sources de connaissances générales externes et fiables sur lesquels pourraient être lancées des requêtes SPARQL pour nourrir l'ontologie. Une de ces sources sera probablement DBpedia, un dépôt de données sémantiques extraites de la populaire encyclopédie en ligne Wikipedia. Les concepts les plus changeants et expansifs de l'ontologie FIC tels que les « ObjetsDeSpécialisationTI » ou « Organisation » bénéficieront beaucoup de cette externalisation (au niveau de la maintenance principalement).

L'étalement de l'application à des CdPs adjacentes (directeurs de compte, clients, candidats), ou encore la conceptualisation de manières de rendre la maintenance de l'application accessible à ses propres usagers (c.-à-d. de rendre possible la construction et la modification des ontologies, des règles, des *screen scrapers*, des macro-commandes, ou n'importe quel paramètre sémantique, directement sur l'interface usager) sont deux autres développements potentiels qui pourraient aboutir sur des améliorations significatives du système, en apportant à la fois de l'uniformité et de la flexibilité.

Éventuellement, il serait également bien pour Combine d'inclure des dispositifs de traitement du langage naturel ou de reconnaissance des visages en ce qu'ils permettraient d'extraire des données sémantiques d'artéfacts non-structurés tels que les CVs, les descriptions de poste (courriels) ou les portraits de candidat.

4.1.2 Intégration aux e-Portfolios

Durant le déroulement de cette recherche, l'Université de Montréal lança un projet pan-universitaire visant à développer un dit *e-Portfolio* qui pourrait supporter l'acquisition et la promotion des compétences des étudiants. Le e-Portfolio se veut donc à la base une plateforme électronique de suivi et d'attestation du développement des compétences des élèves. Côté suivi de l'acquisition des compétences, il prévoit permettre aux étudiants de pouvoir réfléchir à leurs apprentissages et réalisations (fonction formative et sommative du portfolio). Côté promotion et attestation des compétences, il prévoit permettre aux étudiants de pouvoir présenter de manière structurée les aptitudes acquises lors de leur formation (fonction « vitrine » du portfolio). Ces deux pans constituent les exigences globales du e-Portfolio. Malgré le fait que le projet Combine et celui du e-Portfolio n'avait aucune affiliation officielle, les chercheurs des deux côtés ont démontré des intérêts réciproques pour leurs travaux respectifs : les promoteurs du e-Portfolio étaient curieux d'en apprendre sur la vision d'employabilité sous-tendue par le projet Combine, tandis que ceux de

Combine fouinaient et questionnaient la manière dont les compétences seraient formalisées dans l'e-Portfolio, en espérant et militant pour que ce dernier utilise un format sémantique.

De la perspective heuristique de Combine, il serait d'une grande valeur de voir l'e-Portfolio de l'Université de Montréal accessible via des requêtes SPARQL. Les deux systèmes bénéficieraient grandement d'une telle interopérabilité. Un portfolio sémantique pourrait, par exemple, permettre :

- Une scrutabilité du portfolio par des agents logiciels externes (logiciels de recrutement)
- Une diffusion automatique du profil académique des étudiants via des applications sociales (voir Figure 41)
- Une certification par l'institution académique des diplômes ou des crédits cumulés par les étudiants
- La réception, sur le portfolio des étudiants, d'offres d'emploi personnalisées en fonction des compétences acquises
- Des suggestions automatiques de cours à suivre, ou de compétences à développer en fonction des objectifs de carrière des étudiants
 - o affichage du niveau des acquis par rapport aux types de postes désirés
 - o explication (du point de vue de l'employabilité) de l'importance d'acquérir telle ou telle compétence

Michel Tétreault
Human-Computer Interactions Student-Researcher at
University of Montreal
Montreal, Canada Area

→ **Contact Michel Tétreault**
→ **Add Michel Tétreault to your network**

Current • **Student-Researcher at Université de Montréal**

Past • Assistant de cours / Étudiant at Université de Montréal
• Receiving Team Leader at Tarkett

Education • Université de Montréal

Recommended 👤 1 person has recommended Michel

Connections 🌐 9 connections

Industry Information Technology and Services

Websites • U of M | Communication Dep.

Public profile powered by: **LinkedIn**
Create a public profile: **Sign In** or **Join Now**

e-Portfolio Université **um**
de Montréal

Portfolio des acquis de **Michel Tétreault**

- Baccalauréat en Communication
- Maîtrise en Communication

Suivez mon développement: **RSS** **RDF**
...ou parlez s'en à quelqu'un! **SHARE**

View Michel Tétreault's full profile:

Figure 41 - Capture d'écran d'une maquette de l'application sociale de l'e-Portfolio.

4.2 Discussion critique des perspectives théoriques adoptées par le projet Combine

4.2.1 Philosophie objectiviste et problème de l'engagement ontologique

Le WS, ainsi que la modélisation des connaissances en général, sont basés sur la supposition que la réalité existe indépendamment de la conscience humaine, et que nous pouvons par conséquent extraire des connaissances *objectives* de nos perceptions de celle-ci (Kiryakov et coll., 2003). Ce type de présupposé anhypothétique fait partie d'une philosophie dite *objectiviste*. Comme le pose Kiryakov :

« Modelling knowledge about the world requires some assumptions about its nature, as well as about the nature of the “observer” who is expected to use, understand, and rely on the models. [...] Ontologies aid the sharing of knowledge on the basis of the assumption that there is a

single reality and the sharing is a matter of aligning the way different people or systems “think” about it. » (Kiryakov and coll., 2003, p.489)

Mais cet alignement est dans les faits assez difficile à réaliser, et c’est là tout le véritable défi d’interopérabilité du WS. Ce n’est pas juste une question de conflits de politiques et de modèles d’affaires ; c’est toute une problématique de définir des référents objectifs qui servent des buts et intérêts toujours subjectifs. Voilà pourquoi le WS a de la difficulté et tarde à s’imposer à une échelle globale. Le philosophe américain Willard Van Orman Quine a appelé ce problème le problème de l’engagement ontologique à cause de l’aspect arbitraire qu’il y a à définir des propriétés d’objet qui rencontrent les perceptions de tout un chacun. Et ce problème gagne considérablement en amplitude lorsqu’il s’agit de formaliser des objets sociaux comme nous l’avons fait avec l’ontologie FIC. En fait, les consensus concernant la définition de concepts sociaux sont difficiles à obtenir parce que les concepts sociaux ont des référents opaques – en ce que nous ne pouvons pas « voir » les objets qu’ils pointent puisque ceux-ci sont de l’ordre de l’intangible –, contrairement aux concepts naturels qui, eux, possèdent des référents transparents – pointant des matérialités physiques « visibles », c'est-à-dire des choses localisées dans l’espace-temps et donc appréhendables par les sens (Dennett, 1997). Par exemple, nous ne pouvons porter à la vue des objets sociaux comme un *poste* ou un *projet* pour les appréhender empiriquement puisqu’ils ne possèdent aucune présence physique. Ils sont des objets virtuels, des abstractions toutes *méta*-physiques. Ils utilisent des choses physiques pour supports symboliques – arbitraires – soit, mais n’existent que par la consensualité des croyances et attitudes des gens à l’égard de leurs définitions conceptuelles, ou par la reconnaissance des gens que leurs usages et représentations physiques satisfont cette définition (Searle, 1997).

En tant qu’ils demeurent des représentations humaines, mêmes les concepts naturels, qui possèdent des référents plus transparents (garant d’une certaine indépendance des choses pointées par rapport à leur nomination), conservera toujours néanmoins une certaine opacité référentielle résiduelle, principalement dans la manière dont nous définissons et

catégorisons leurs propriétés et/ou dans la manière dont nous leur attribuons certaines fonctions.

Une façon de comprendre les concepts sociaux à l'aune du problème de l'engagement ontologique est de constater que leurs définitions sont constitutives – et non représentative – des objets qu'ils « représentent » (Searle, 1997). En ce sens, les objets sociaux sont plus des projections conceptuelles sur la « réalité » que des objets réels sujets à des représentations conceptuelles. En d'autres mots, l'existence des objets sociaux ne tient qu'à la manière dont ils sont définis et utilisés. Qu'est-ce qu'un poste sinon qu'une définition en termes de tâches et responsabilités ainsi que la manière dont une personne interprètera cette définition et agira en conséquence ? Pour cette raison, de la manière dont nous définissons un concept social découle toujours d'importantes conséquences sur l'objet social résultant (Bowker et Star, 2000). Sur le front du WS, ce problème se révèle à travers l'existence de plusieurs ontologies qui décrivent de manières différentes les mêmes domaines de connaissances. Les recherches par rapport aux techniques dites d'« alignement d'ontologies » se doivent d'être intensifiées pour contrer cette inconvenance (l'alignement d'ontologie dénote le processus de comparer et mettre en équivalence les concepts similaires d'ontologies distinctes). En même temps, les projets de développement du WS doivent commencer à répondre à des préoccupations d'ordre plus local pour mettre en marche le processus de génération et utilisation des données sémantiques. Les applications de type SW actuelles ne proposent que des fonctionnalités et modèles très généraux qui rendent difficilement compte des besoins spécifiques des acteurs localement situés et intéressés.

4.2.2 Approche ascendante versus approche descendante

Comme l'affirme Spivack : « The Semantic Web is a web of semantic webs » (Spivack, 2008, p.28). À cause du problème de l'engagement ontologique, nous devons comprendre et assumer que les ontologies les plus significantes pour une CdP donnée demeureront toujours

les ontologies modélisées localement selon les objets, contextes et besoins particuliers de cette CdP. Pour activer Le WS, nous devrions donc nous concentrer plus sur la constitution de webs sémantiques pragmatiques et localisés que sur le simple développement et la promotion de modèles génériques qui ne servent que des fins générales (Alani, 2008). Cette position mène à la vision d'un WS pleinement développé dans lequel chaque individu possèdera ultimement son propre web sémantique personnel qui se maillera, par alignement d'ontologies, aux autres webs sémantiques ambiants. Dans ce portrait, les ontologies de haut niveau (plus abstraites et générales) pousseront naturellement de l'alignement de plusieurs webs sémantiques particuliers, du maillage graduel de multiples webs sémantiques de bas niveau. Ceci constitue dès lors une perspective très émergentiste (ascendante) de modéliser et promouvoir les systèmes sémantiques. En ce sens, le projet Combine a fait le pari, en choisissant de développer un web sémantique bien ciblé, que les gains en frais d'interopérabilité seront supérieurs à long terme que ceux qu'aurait générés le développement d'un système sémantique fédéré à grande échelle. Considérant à la notion d'interopérabilité, il faut donc voir que les modèles sémantiques génériques conçus dans l'optique d'organiser de très larges éventails de ressources numériques (ce qui ne peut se faire qu'en faisant abstraction des contextes particuliers d'utilisation) insistent plus sur la dimension « inter » du concept que sur sa dimension d'« opérabilité » qui lui est, dans un sens, plus fondamentale. En effet, c'est cette dernière dimension qui est garante de la valeur utile des systèmes. Valeur sans laquelle on ne peut espérer séduire les usagers potentiels et sur laquelle il faut donc capitaliser en premier lieu. Seulement ensuite qu'il peut devenir intéressant de développer l'aspect « inter » de du potentiel d'interopérabilité des systèmes.

Bien que la modélisation de type ascendante se doit d'être considérée plus sérieusement, l'utilisation de techniques de type descendante, comme d'adopter, de créer et de réutiliser des vocabulaires comprenant des concepts généraux, demeure néanmoins d'une importance significative pour faire marcher le WS sur une échelle étendue (dans la mesure où les techniques d'alignement d'ontologies ne sont pas encore véritablement au point). Du point de vue de notre étude, les meilleures pratiques de modélisation sont celles qui commencent

par des activités de modélisation de type ascendante, pour bien ancrer le système sémantique développé dans le contexte particulier du champ d'application, et qui incorpore par la suite des activités de modélisation descendantes, pour maximiser l'interopérabilité du système. En d'autres mots, même si le concept du WS porte de belles promesses en matière d'interopérabilité, il est de mise de garder à l'esprit que ces promesses ne sont que des buts généraux en direction desquels les applications sémantiques actuelles doivent tendre graduellement – à travers leur offre de solutions bien pragmatiques et localisées. Peut-être les techniques descendantes ont pour vocation appropriée de fournir aux gens des moyens communs et utilisables de modéliser, gérer, et entretenir leurs propres webs sémantiques.

4.2.3 Pronostic sur l'avenir des réseaux sociaux en ligne

Par rapport à l'évolution du réseautage social, on peut imaginer un futur où il ne sera plus tenu ou performé par des plateformes centralisées comme c'est le cas actuellement, mais en tant que résultant de l'interconnexion d'objets partagés tels que des conversations, des images, des événements, des travaux, etc. Les relations sociales électroniques résulteront simplement du chevauchement des webs sémantiques personnels et les réseaux sociaux seront déduits de la prise en compte des objets communément partagés donnant lieu à ces chevauchements. Les objets partagés deviendront en ce sens le médium fondamental des relations sociales en ligne, et reflèteront plus fidèlement du même coup la manière dont fonctionnent les relations sociales hors-ligne. Ce qui est vrai des relations sociales est aussi vrai pour n'importe quelles institutions : elles sont la cristallisation, entre plusieurs individus, de connaissances, de modèles et d'objets communs, de même que leurs règles d'usage. Par conséquent, le réseautage social sémantique devrait se montrer beaucoup plus flexible et subtil que l'actuel Web 2.0 qui connecte les personnes en les forçant à faire des déclarations catégoriques, et *a priori*, sur la nature de leurs relations : « Êtes-vous des amis ou pas ? ».

4.2.4 Vers un Web Pragmatique ?

Tout comme pour la linguistique qui distingue trois pans de son champ d'études, soit la syntaxe, la sémantique et la pragmatique, l'évolution du web laisse présager un horizon où l'aspect sémantique des transactions digitales sera généré et interprété par des structures capables de rendre compte des contextes d'utilisation situés et localisés. « La signification est l'usage » disait le philosophe viennois Ludwig Wittgenstein. Dans cette optique, on peut voir que la dimension sémantique que tentent d'insuffler les technologies du WS au web syntaxique pour le rendre plus intelligent demeure relativement fragile vis-à-vis la très grande disparité des cas d'utilisation possibles. Comme nous venons de le voir avec le problème de l'engagement ontologique, il n'est pas aisé de créer des référents universellement acceptés, surtout lorsque l'on a affaire à des objets sociaux.

C'est donc toute une mécanique d'indexicalisation qui resterait à mettre en place, pour que les ontologies se génèrent d'elle-même sur la base de l'observation de la manière dont les informations sont utilisées. Dans le cadre du projet Combine, un premier pas dans cette direction serait que les agents recruteurs puissent indiquer au système que telle ou telle information sur l'interface correspond à tels ou tels concepts de l'ontologie et même de créer de nouveaux concepts au besoin, un peu à l'image du système de signets sémantiques Semantic Turkey proposé par Donato Griesi et coll. (2007). Pour l'instant, Combine est conçue pour utiliser des *screen scrapers* et macro-commandes de saisie construits à l'avance, ce qui implique de prévoir du temps de maintenance à chaque fois que le code HTML des pages utilisées change. Bref, d'orienter les systèmes dans le sens qu'ils soient capables « d'apprendre » directement de l'utilisateur et non du programmeur est, à notre avis, une première étape logique vers l'auto-génération des ontologies et des données sémantiques.

5 Conclusion

Harith Alani et coll. (2008), dans un article intitulé *Building a Pragmatic Semantic Web*, ont énoncé quatre principes pour mettre en confiance les entreprises à prendre le virage sémantique : minimiser les perturbations avec les infrastructures existantes ; utiliser de petites ontologies bien ciblées ; montrer la valeur ajoutée ; et préserver la provenance et la confidentialité. Voilà tout ce sur quoi le projet Combine était fondé. Premièrement, nous avons observé et analysé les interactions humain-ordinateur de la CdP ciblée de manière à faire de Combine une application qui maximise son adéquation avec l'écosystème des pratiques et ressources existantes de cette CdP. Deuxièmement, nous avons choisi de restreindre notre modélisation ontologique sur un processus bien spécifique de l'organisation étudiée : le recrutement. Nous aurions pu choisir de modéliser une ontologie et une application qui aurait aidé différentes CdP à communiquer plus efficacement, mais comme le web sémantique qu'aurait généré cette ontologie n'aurait pas été assis sur des sous-systèmes sémantiques bien ciblés et orientés sur des tâches précises, il aurait probablement été perçu simplement comme un système additionnel qui n'apporte aucun gain concret en termes de productivité. Troisièmement, nous avons effectivement convaincu les recruteurs de la valeur ajoutée d'utiliser une application sémantique telle que Combine. Finalement, la base de connaissances de Combine a été pensée de sorte à suivre rigoureusement les sources d'information et de donner le plein contrôle aux recruteurs quant au niveau de confidentialité de chaque objet d'information (qu'est-ce qui vient d'où et qui peut voir quoi).

À cause de sa perspective IHO, ce projet était plus à propos du design de l'avant-plan de l'application que du codage et de la mise à l'épreuve de ses technologies d'arrière-plan. En fait, il s'agissait de la première phase de ce projet de développement et, considérant les rétroactions positives de la part de notre CdP ciblée, nous considérons sérieusement d'aller de l'avant avec la seconde phase du projet et de produire l'application conceptualisée.

Si l'on se rapporte aux hypothèses de recherche émise au début du projet, nous pouvons affirmer que l'utilisation de notre application devrait en effet améliorer les interactions humain-ordinateur des acteurs du domaine en :

- Automatisant les opérations redondantes
- Supportant les processus de recherche et décisions
- Réduisant le nombre des étapes de la plupart des tâches

Nous pouvons aussi affirmer que l'utilisation de notre application devrait effectivement améliorer la CMO de la CdP en :

- Facilitant le partage et l'annotation des objets d'information
- Automatisant et maximisant la diffusion des messages (courriels, affichage de poste, etc.)
- Intégrant toutes les technologies de communication utilisées par la CdP

En somme, le projet de développement qui fut le sujet de ce mémoire, propose un processus d'intégration des sources d'information basé sur des mécanismes d'extraction et de déduction pour récupérer et gérer les structures d'information via des ontologies. Ce développement centré sur l'utilisateur visait en premier lieu à démontrer les bénéfices d'utiliser des ontologies en dépit des contraintes résultant du manque d'une masse critique d'utilisateurs et d'entreprises qui génèrent et diffusent des données sémantiques. Aussi, en dépit du fait que ce développement centré sur l'utilisateur a limité d'une certaine manière le plein potentiel d'interopérabilité pour lequel les systèmes sémantiques sont habituellement reconnus (en tant qu'ils sont typiquement utilisés pour normaliser l'échange de données entre de multiples systèmes et organisations), il a démontré avec succès des retours sur investissement potentiels et intéressants pour sa CdP ciblée. Au-delà des contraintes des entreprises par rapport aux questions sensibles de la propriété des données et des politiques de séquestration des données, les efforts en vue d'installer une interopérabilité sémantique à grande échelle se butent à l'incapacité des acteurs particuliers à utiliser et tirer avantage des données sémantiques. Même si Combine ne ferait pas réaliser de gains d'interopérabilité significatifs et immédiats – dans l'esprit exact d'une interopérabilité qui ferait travailler les systèmes et/ou les organisations de concert – elle habiliterait l'entreprise FIC à exploiter la

valeur ajoutée des données sémantiques et, par conséquent, placerait FIC dans une position avantageuse pour réaliser des gains d'interopérabilité majeurs lorsque le Web 3.0 prendra son véritable envol. Ce qu'il faut bien comprendre ici est qu'avant même de penser à partager des données sémantiques, les acteurs particuliers doivent d'abord apprendre à les générer et à les utiliser, avantageusement.

Bibliographie

- Alani, H., Hall, W., O'Hara, K., Shadbolt, N., & Szomszor, M. (2008). Building a Pragmatic Semantic Web. *IEEE Intelligent Systems*. 23(3), 61-68.
- Anderson, J. (2010). *Those Who Understand the Semantic Web Are Split On its Future*. Repéré le 14 juillet 2010, à <http://www.pewinternet.org/Press-Releases/2010/Semantic-Web.aspx>
- Andler, D. (2002) « Processus Cognitifs » -extrait-, dans Andler, D., A. Fagot-Largeault & B. Saint-Sernin *dirs.* (2002) *Philosophie des sciences I*, Paris : Gallimard, pp. 226-296.
- Berners-Lee, T. (1994, mai). *W3 Future Direction* [notes de plénière]. Communication présentée au First International World Wide Web Conference, Genève, Suisse. Repéré le 17 août 2008, à <http://www.w3.org/Talks/WWW94Tim/>
- Berners-Lee, T., & Fischetti, M. (1999). *Weaving the Web: the original design and ultimate destiny of the World Wide Web by its inventor*. San Francisco: HarperBusiness.
- Berners-Lee, T. (2005). Putting the Web back in the Semantic Web, *ISWC – RuleML conference Galway 2005*, Repéré le 20 décembre 2009, à [http://www.w3.org/2005/Talks/1110-iswc-tbl/#\(1\)](http://www.w3.org/2005/Talks/1110-iswc-tbl/#(1))
- Berners-Lee, T., Hollenbach, J., Lu, K., Presbrey, J., Pru d'ommeaux, E. & schraefel, m. c. (2007) *Tabulator Redux: Writing Into the Semantic Web. Rapport technique ECSIAMeprint14773*, Electronics and Computer Science, University of Southampton. Repéré le 28 septembre 2009, à <http://eprints.ecs.soton.ac.uk/14773/1/tabulatorWritingTechRep.pdf>
- Bizer, C., Heese, R., Mochol, M., Oldakowski, R., Tolksdorf, R., & Eckstein, R. (2005). The Impact of Semantic Web Technologies on Job Recruitment Processes. *Wirtschaftsinformatik*. 1367-1382.
- Bowker, G. C., Leigh Star, S. (2000) *Sorting Things Out: Classification and Its Consequences*. Cambridge : The MIT Press.

- Breitman, K. K., Casanova, M.A., & Truszkowski, W. (2007). *Semantic Web: Concepts, Technologies and Application*. London: Springer-Verlag.
- Bush, V. (1945, juillet). As We May Think. *The Atlantic*. Repéré le 29 juin 2010, à <http://www.theatlantic.com/magazine/archive/1969/12/as-we-may-think/3881/>
- Cardoso, J., Hepp, M., & Lytras M. (2008). The Future of The Semantic Web for Enterprises. in J. Cardoso and al. (Ed.), *The Semantic Web: Real World Applications from Industry* (pp. 2-37). New York: Springer Science + Business Media.
- Christiaens, H., Wets, G., & Vanthienen, G. (1998). *Developing a Decision Support System Using the Cognitive Fit Approach: An Evaluation* (Rapport no. NR 9819). Leuven : Université Catholique de Leuven.
- Computer-mediated communication. (2010, 15 juin). Dans *Wikipedia, The Free Encyclopedia*. Repéré le 25 juillet 2010, à http://en.wikipedia.org/w/index.php?title=Computer-mediated_communication&oldid=368155150
- Davenport, T. H., & Beck, J. C. (2001). *The Attention Economy: Understanding the New Currency of Business*. Boston: Harvard Business School Press.
- Davenport, T. H. (1992). *Process Innovation: Reengineering Work Through Information Technology*. Boston: Harvard Business School Press.
- Davenport, T. H. (1997). *Information Ecology: Mastering the Information and Knowledge Environment*. New York: Oxford University Press.
- Dawson, R. (2005). *Developing Knowledge-Based Client Relationships*. Oxford: Elsevier Butterworth Heinemann.
- Dennett, D. C. (1997). *Kinds Of Minds: Toward An Understanding Of Consciousness*. New York: Basic Books.
- Economist [The] (2008). Everywhere and nowhere. May 19, 2008. http://www.economist.com/business/displaystory.cfm?story_id=10880936. 2008.

- Engelbart, D.C. (1962). *Augmenting human intellect: a conceptual framework*. Menlo Park, CA: Stanford Research Institute. (Summary report AFOSR-3233) Repéré le 28 juin 2010, à <http://www.dougenelbart.org/pubs/augment-3906.html>
- Foulonneau, M., Riley, J. (2008). *Metadata for Digital Resources: Implementation, Systems Design and Interoperability*. Oxford: Chandos Publishing Limited.
- Fortin, C. et Rousseau, R. (1989). *Psychologie cognitive : Une approche de traitement de l'information*. Québec : Télé-Université du Québec.
- Frith, U. (1992). *L'énigme de l'autisme*. Paris : Éditions Odile Jacob.
- Graziosi, F. (2010) *Databases: relational vs object vs graph vs document*. Repéré le 29 Juin 2010, à http://www.cbsolution.net/roller/ontarget/entry/databases_relational_vs_object_vs
- Griesi, D., Paziienza, M. T., & Stellato, A. (2007) Semantic Turkey: a Semantic Bookmarking Tool (System Description). Repéré le 28 Juin 2010, à http://art.uniroma2.it/publications/docs/2007_ESWC07_Semantic%20Turkey%20-%20a%20Semantic%20Bookmarking%20tool.pdf
- Grundstein, M. (2004). « De la capitalisation des connaissances au management des connaissances dans l'entreprise ». Paris : Open Access publications from Université Paris-Dauphine, urn:hdl:123456789/2620, 18 p.
- Hansen, M. T., Nohria, N., & Tierney, T. (2003). « Quelle est votre stratégie de gestion du savoir ? ». dans L. P. Cohen (Ed.). *Le management du savoir en pratique*. Paris : Édition d'Organisation. pp. 117-149.
- Héon, M., Paquette, T., & J. Basque. (2008). Transformation of Semi-formal Models into Ontologies According to a Model Driven Architecture, *Centre de recherche LICEF*. Repéré le 12 Decembre 2009, à www.licef.telug.uqam.ca/Portals/29/docs/pub/onto/JFO2008-Heon.pdf
- Henricks, M. (2007). The Third Wave. *Entrepreneur Magazine*. Juin 2007.

- Hildreth, P.J. & Kimble, C. (2002). The duality of knowledge, *Information Research*, Vol. 8, No. 1, Octobre 2002.
- Human-computer interaction. (2010, Juin 22). Dans *Wikipedia, The Free Encyclopedia*. Repéré le 25 Juillet 2010, à http://en.wikipedia.org/w/index.php?title=Human%E2%80%93computer_interaction&oldid=369563954
- Interopérabilité. (2011, mai 5). *Wikipédia, l'encyclopédie libre*. Repéré le 25 Juin 2011, à <http://fr.wikipedia.org/wiki/Interopérabilité>
- Jensen, J. F. (1999). 'Interactivity' – Tracking a New Concept in Media and Communication Studies. Dans P. Mayer (Dir.), *Computer Media and Communication: A Reader* (pp. 160-187). Oxford: Oxford University Press.
- Klanten, R., Bourquin, N., Tissot, T., Ehmann, S., & Van heerden, F. (2008). *Data Flow: Visualising Information in Graphic Desing*. Berlin: Gestalten.
- Kleek, Max., Shrobe, H. E. (2007) A Practical Activity Capture Framework for Personal, Lifetime User Modeling. Dans C. Conati, K. Mccoy, et G. Paliouras (Dir.), *Proceedings of the 11th international conference on User Modeling (UM '07)*. Berlin, Heidelberg : Springer-Verlag, (pp. 298-302). DOI=10.1007/978-3-540-73078-1_33 http://dx.doi.org/10.1007/978-3-540-73078-1_33
- Kiryakov, A., Popov, B., Terziev, I., Manov, D., & Ognyanoff, D. (2003). Semantic Annotation, Indexing, and Retrieval. Dans D. Fensel, K. Sycara-Cyranski, K. Sycara, & J. Mylopoulos (Éd.), *The Semantic Web, ISWC 2003: Second International Semantic Web Conference, Vol. 2870, 2003* (pp. 484-499). Berlin: Springer-Verlag.
- Kleinbaum, A. M., Stuart, T. E., & Tushman, M. L. (2008). *Communication (and Coordination?) in Modern, Complex Organization*. Repéré à <http://www.hbs.edu/research/pdf/09-004.pdf>. 69 p.
- Laughery, R. K. J., Lebiere, C., & Archer, S. (2006). « Modeling Human Performance in Complex Systems ». dans S. Gavriel (Ed.), *Handbook of Human Factors and Ergonomics* (Third Edition).(pp. 965-996). Hoboken: Wiley

- Lev, B. (2001). *Intangibles: Management, Measurement, and Reporting*. Harrisonburg: R. R. Donnelley and Sons.
- Lévy, P. (2006) Pour un modèle scientifique des communautés virtuelles. Dans S. Proulx & coll. (Dir.) *Communautés virtuelles : Penser et agir en réseau* (pp. 91-106). Québec : Les Presses de l'Université Laval.
- Licklider, J.C.R., & Taylor, R. (1968, avril). The Computer as Communication Device. *Science & Technology*, avril 1968, 21-31.
- London, A., Thomasson, J.-J., Maesano, L. (2006). *Modélisation XML*. Paris : Eyrolles.
- Maes, P. (1994). Agents that Reduce Work and Information Overload. *Communications of the ACM*, 37(7), 31-40.
- McAfee, Andrew (2006). « Enterprise 2.0: The Dawn of Emergent Collaboration », dans *MIT Sloan Management Review* Vol. 47, No. 3, pp. 21-28
- Mintzberg, H. (1989). *Mintzberg on Management: Inside Our Strange World of Organization*. New York: Hungry Minds Inc.
- Morville, P., Rosenfeld, L. (2006). *Mintzberg on Management: Inside Our Strange World of Organization*. Sebastopol, CA : O'Reilly Media.
- Nonaka, I., Takeuchi, H. (1995). *Information Architecture for the World Wide Web*. New York : Oxford University Press, 304 p.
- Nonaka, I. (1991). "The knowledge creating company". *Harvard Business Review* 69 (6 Nov-Dec): 96–104. <http://hbr.harvardbusiness.org/2007/07/the-knowledge-creating-company/es>.
- Noy, N. F., McGuinness, D. L. (2007) *Développement d'une ontologie 101 : Guide pour la création de votre première ontologie* (Anila Angjeli trad.), Stanford : <http://protege.stanford.edu/>.
- Pinker, S. (2000). *Comment fonctionne l'esprit*, Paris : Éditions Odile Jacob.

- Polanyi, M. (1958). *Personal Knowledge: Towards a Post- Critical Philosophy*. Chicago : University of Chicago Press.
- Prax, J.-Y. (2003). *Le Manuel Du Knowledge Management: Une Approche De 2e Generation*. Paris : Dunod.
- Quine, W.V.O. (1977). *Le mot et la chose*. Paris : Champs Flammarion, 399 p.
- Rivard, S., Talbot, J. (2008). *Le développement de systèmes d'information*. Québec : Presses de l'Université du Québec.
- Searle, J. R. (1980). "Minds, Brains, and Programs", extrait de *The Behavioral and Brain Sciences*, vol.3 Cambridge : Cambridge University Press, pp. 417-457
- Searle, J. R. (1997). *The Construction of Social Reality*. New York: Free Press.
- Simon, H. A. (1996). *The Sciences of the Artificial*. Cambridge: The MIT Press.
- Sloterdijk, P. (2006). *Le Palais de cristal*. Paris: Hachette Littératures.
- Spivack, N. (2008). Semantic Web Talk, *Nova Spivack: Minding the Planet*, Retrieved December 20, 2009, from http://novaspivack.typepad.com/nova_spivacks_weblog/2008/02/video-of-my-sem.html
- Staab, S., Studer, R. (ed.) (2004). *Handbook on Ontologies*. Berlin : Springer-Verlag.
- Takeuchi, H. (1995), *The knowledge creating company: how Japanese companies create the dynamics of innovation*, New York: Oxford University Press, pp. 284
- Tidwell, J. (2006). *Designing Interfaces: Patterns for Effective Interaction Design* (2e éd.). États-Unis : O'Reilly Media.
- Uniform Resource Identifier. (2010, juillet). *Wikipédia, l'encyclopédie libre*. Repéré le 15 juillet 2010, à

http://fr.wikipedia.org/w/index.php?title=Uniform_Resource_Identifier&oldid=54825342.

Uschold, M. Jasper, R. (1999). A Framework for Understanding and Classifying Ontology Applications. Compte rendu de l'atelier du IJCAI-99 *Ontologies ans Problem-Solving Methods*, Stockholm.

Web sémantique. (2010, 15 juin). Dans *Wikipedia, L'encyclopédie libre*. Repéré le 25 juillet 2010, à http://fr.wikipedia.org/wiki/Web_sémantique

Wenger, E. (1998) *Communities of Practice: Learning, Meaning and Identity*. Cambridge: Cambridge University Press.

Wilks, Y. (2008). The Semantic Web: Apotheosis of Annotation, but What Are Its Semantics?, *IEEE Intelligent Systems*, 23(3), 41-49.

Yahiaoui, L., Boufaïda, Z., & Prié, Y. (2006, Decembre). *Semantic Annotation of Documents Applied to E-Recruitment*. Paper presented at *SWAP 2006*, Pisa, Italia.

Zhong, N., Liu, J., & Yao, Y. (Ed.). (2003). *Web Intelligence*. Berlin: Springer-Verlag.

Zwirn, H. (2000). *Les limites de la connaissance*. Paris : Éditions Odile Jacob.

Annexe 1

Définition des notions RDF de conteneur et de collection

La catégorie des conteneurs présente trois sous-classes que sont `rdf:Bag`³⁶, qui permet de déclarer qu'une ressource est un groupe d'items sans ordre, `rdf:Seq`, qui permet de déclarer qu'une ressource est un groupe d'items ordonné, et `rdf:Alt`, qui permet de déclarer qu'une ressource est un groupe d'alternative dont une seule peut-être choisie à la fois. Pour énumérer les membres d'un conteneur, il suffit de les relier à la ressource-conteneur par des relations de type `rdf:_n` (`rdf:_1`, `rdf:_2`, `rdf:_3`, ...) dans le cas où l'on désire contrôler manuellement la numération des items, sinon, la relation `rdf:li` permet de générer une numération automatique.

En ce qui concerne les collections, à l'instar des conteneurs qui ne sont pas considérés comme des ensembles fermés, les collections constituent des déclarations de listes en tant qu'ensembles d'items formant des tous complets. Pour asserter une collection on commence par asserter `rdf:List` comme portant sur une certaine ressource (la collection en question). Puis on prédique la première ressource de la liste avec l'URI `rdf:first`. Après, on asserte immédiatement que la collection contient d'autres items en asserant `rdf:rest`. On ajoute l'item suivant en réutilisant `rdf:first...` et ainsi de suite jusqu'à récursivement tous les items aient été inscrits. Une collection vide sera représentée par l'URI `rdf:nil`.

Finalement un dernier type d'assertions que permet de construire RDF sont des assertions à propos de triplets RDF. Pour ce faire il faut déclarer un triplet comme étant une ressource de type (`rdf:type`) `rdf:Statement`. Ce procédé se nomme la réification. Puis on précise le contenu de ce triplet avec les propriétés `rdf:subject`, `rdf:predicate` et

³⁶ Lorsqu'on utilise seulement le préfixe « `rdf:` » c'est pour représenter de manière abrégée l'URIref « `http://www.w3.org/1999/02/22-rdf-syntax-ns#` »

`rdf:object`. Mais pour bien identifier le triplet on peut ajouter l'attribut `rdf:ID` à la propriété et donner un nom particulier à la proposition que constitue le triplet réifié.

Annexe 2

Revue descriptive des méthodologies de modélisation d'ontologie listées par Breitman et coll. (2007)

Méthode Uschold et King

Il s'agit d'une méthode somme toute assez simple où l'identification de scénarios motivés (motivated scenarios) guide essentiellement le développement de l'ontologie. Quatre sous-étapes composent cette méthode, soit : 1- l'identification des buts et de la portée de l'ontologie ; 2- la construction de l'ontologie ; 3- l'évaluation de l'ontologie ; 4- la documentation du processus de construction. Cette méthode est toutefois critiquée pour son manque d'étapes intermédiaires dans l'explicitation des classes et relation de l'ontologie.

Méthode Toronto Virtual Enterprise (TOVE)

En contraste avec le manque de formalisme de la méthode précédente, la méthode TOVE met de l'avant une étape subséquente à l'élaboration de scénarios qui préconise la formulation de questions spécifiques auxquelles l'ontologie devra être en mesure de répondre. Subséquemment, ces questions doivent être ensuite traduites dans un langage de représentation des connaissances en passant la formulation d'une terminologie formelle. En somme, les étapes de la méthode TOVE sont : 1- description de scénarios ; 2- formulation informelle de questions de compétence ; 3- spécification des termes de l'ontologie dans un langage de représentation des connaissances ; 4- traduction formelle des questions de compétence ; 5- spécification des axiomes (définition terminologique de l'ontologie) ; 6- vérification de la complétude de l'ontologie.

Méthode Methontology

Très descriptive, cette méthode suggère, de manière comparative et non directive, un corpus d'activités propices au développement d'une ontologie (qui sont pour la plupart issues des méthodes d'élaboration d'exigences en génie logiciel). On peut y distinguer (dans ce corpus) trois grandes catégories d'activités : les activités de management

(identification, planification, contrôle de la qualité,...), les activités « orientées développement » (étude de faisabilité, spécification, formalisation, implémentation,...), et les activités de support (acquisition de connaissances, alignement, documentation, évaluation, intégration, agrégation, configuration,...). Bien que la méthode propose un certain ordre d'enchaînement procédural de ces activités, l'intérêt réside dans le fait qu'elle table sur les avantages et désavantages des diverses techniques d'élicitation conceptuelle :

Tableau 2 Tableau comparatif des techniques d'élicitation des exigences logicielles (extrait Breitman et coll., 2007)

| Technique | Avantages | Désavantage |
|--|---|--|
| Entrevue | <ul style="list-style-type: none"> - Contact direct avec les acteurs - Validation immédiate | <ul style="list-style-type: none"> - Connaissances tacites - Différences culturelles |
| Lecture dynamique de documents | <ul style="list-style-type: none"> - Sources d'information accessibles - Volume d'information important | <ul style="list-style-type: none"> - Information dispersée - Effort considérable |
| Questionnaires | <ul style="list-style-type: none"> - Traitement statistique standardisé | <ul style="list-style-type: none"> - Réponses limitées - Peu d'interaction |
| Atelier d'exigences (groupe de réflexion) | <ul style="list-style-type: none"> - Opinions multiples - Créativité collective | <ul style="list-style-type: none"> - Dispersion - Coût |
| Observation | <ul style="list-style-type: none"> - Faible coût - Faible complexité | <ul style="list-style-type: none"> - Relatif à la personne qui fait l'observation - Exposition superficielle à l'univers d'information |
| Analyse de protocole | <ul style="list-style-type: none"> - Identification de faits non observables - Compréhension accrue des faits et événements | <ul style="list-style-type: none"> - Accent sur la performance (ce qu'on dit n'est pas ce qu'on fait) |
| Anthropologie | <ul style="list-style-type: none"> - Vision interne/externe - Contextualisation | <ul style="list-style-type: none"> - Demande beaucoup de temps - Faible systématisation |
| Dépôts d'exigences non-fonctionnelles | <ul style="list-style-type: none"> - Réutilisation des connaissances - Anticipation des enjeux d'implémentation - Défie l'identification | <ul style="list-style-type: none"> - Coût de mise sur pied élevé - Fausse impression de complétude |

Chose intéressante à noter de cette approche, c'est qu'elle suggère d'employer un langage de modélisation intermédiaire, UML, pour mener à bien la phase de conceptualisation de l'ontologie. Bref, on peut résumer cette méthodologie en disant qu'elle invite à la réutilisation des méthodes de conceptualisation déjà bien éprouvées en génie logiciel.

Méthode KACTUS

Fortement axée sur l'idée de « réutilisation » des connaissances, KACTUS milite en faveur de l'intégration dans le processus de développement d'une ontologie d'étapes d'analyse de ce que les autres ontologies déjà existantes ont à offrir comme concepts pertinents. Plus spécifiquement, cette méthode propose de sélectionner des catégories appartenant à des ontologies de haut niveau et de les raffiner jusqu'à ce qu'elles atteignent une représentativité acceptable du domaine d'application particulier.

Méthode de développement d'ontologie basée sur le lexique

Résolument la plus étoffée des procédures présentées, la méthode de développement d'ontologie basée sur le lexique (*Lexicon-Based Ontology Development Method* ; Breitman et coll., 2007) propose essentiellement de prendre pour point de départ la constitution d'un lexique étendu du langage du domaine (*Language Extended Lexicon* ; LEL) à partir duquel il faut appliquer un processus – semi-formel – de construction d'ontologie. En gros, LEL est un lexique qui liste en précatégories (verbe, sujet, objet et situation) les termes récurrents de l'univers de discours du domaine à modéliser pour les préparer à leur modélisation formelle via le processus de construction (Po) que propose la méthodologie (ce processus est schématisé en Figure 42). La structure ontologique (O) résultante de ce processus correspond au quintuple (C, R, H^c, rel, A^o) ou :

- C et R sont des ensembles disjoints appelés ensemble des *concepts* et ensemble des *relations*
- $H^c \subseteq C \times C$ est une *hiérarchie de concepts* ou *taxonomie*
- $rel : R \rightarrow C \times C$ est une fonction qui relie des concepts de manière non taxonomique
- A^o est un ensemble d'axiomes ontologiques exprimés dans un langage logique approprié

Pour obtenir ce lexique étendu, l'auteur suggère de pratiquer une ou plusieurs des activités d'élicitation telles que celles exposées dans le Tableau 2.

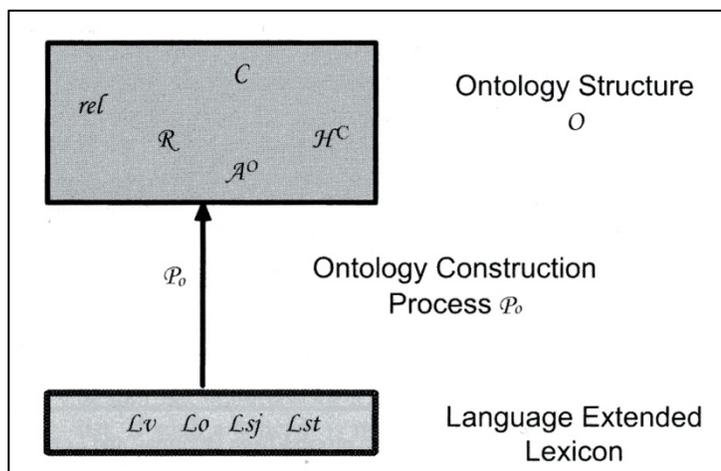


Figure 42 Schéma synthétique du *Lexion-based Ontology Development Method* (tiré de Breitman et coll. 2007)

Étant donné qu'il s'agit d'une méthode qui s'illustre pour sa clarté, sa systématique et sa formule très inductive, la modélisation de l'ontologie à l'intérieur de ce projet utilisera pour principal protocole de formalisation le processus de construction d'ontologie que cette méthode propose. Ce processus, comprend pour principales étapes :

1. Lister les termes alphabétiquement et en les séparant selon leur type : verbe (Lv), objet (Lo), sujet (Lsj) et état (Lst).
2. Préparer trois listes : les concepts (C), les relations (R) et les axiomes (A^O).
3. Transformer les verbes en relation : créer une nouvelle relation pour chaque verbe dont il n'y a pas déjà une relation qui exprime la même chose.
4. Transformer les objets et les sujets en concept (C) et définir ceux-ci avec les relations.
5. Évaluer pour chaque terme de type « état », pour savoir s'il doit donner lieu à un axiome, un concept ou une relation.
6. Construire la taxonomie de l'ontologie (l'arborescence classe/sous-classe) en fonction de la centralité relationnelle de chaque concept (nombre de relations pour un même concept) et en fonction des relations communes.

Développement d'ontologie 101

Proposée par Noy et McGuinness (2007) comme « guide » pour construire sa première ontologie, cette méthode peut être considérée comme une espèce de résumé synthétique de toutes les autres précédemment exposées. Entre autres choses elle insiste sur le fait que la modélisation d'une ontologie est un processus essentiellement itératif dans lequel les procédés de type top-down peuvent alterner avec des procédés de type bottom-up sans problème, voir préférablement. Voici en l'occurrence un arbre que Breitman, Casanova et Truszkowski ont confectionné pour donner un aperçu général des différentes étapes de la méthode :

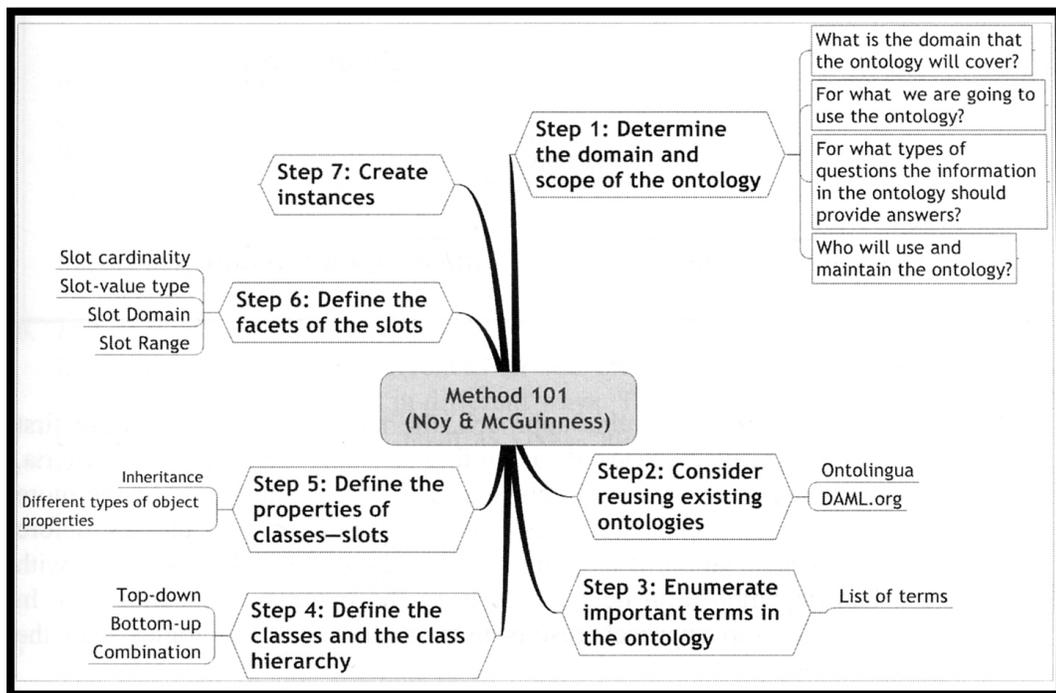


Figure 43 Ontology Development 101 Method (tiré de Breitman et coll. 2007)

Annexe 3

Hiérarchie des processus d'une organisation (exemple ; Rivard et Talbot, 2008)

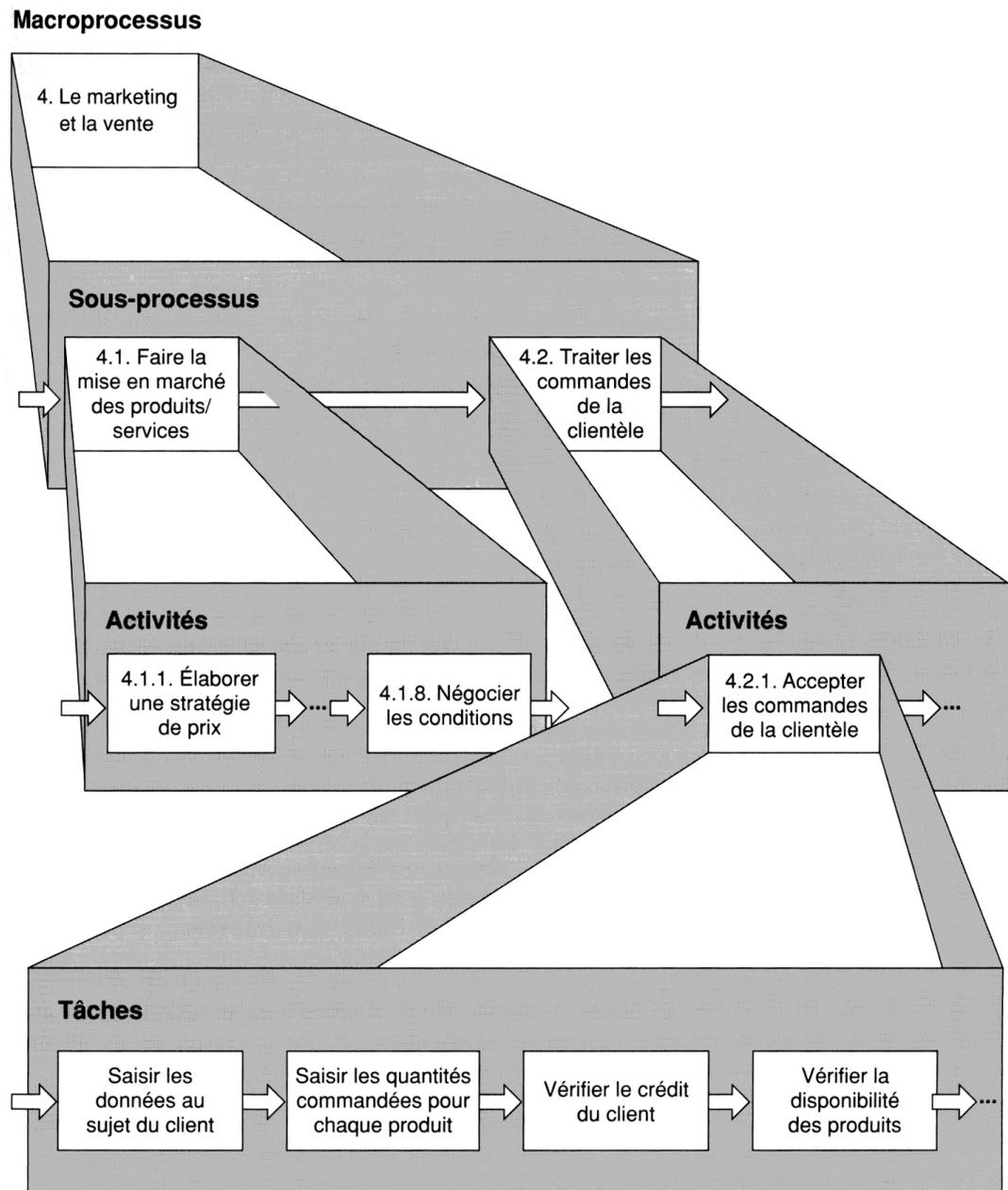


Figure 44 - Hiérarchie des processus d'une organisation (Rivard et Talbot, 2008)

Annexe 4

Questionnaire d'évaluation du prototype Combine.

- 1- Y a-t-il une ou plusieurs tâches dans vos processus de travail que vous trouvez plus difficiles ou inutilement compliquées ? Si oui, lesquelles ?
- 2- Quels sont les principaux sites web que vous utilisez pour trouver ou prospecter de nouvelles candidatures ? Classez-les par ordre de fréquence d'utilisation : du plus utilisé au moins utilisé.
- 3- Qu'est-ce qui vous a plu ou impressionné de l'application *Combine* ?
- 4- Qu'est-ce qui vous a déçu, déplu ou inquiété de cette application ?
- 5- Quelle fonctionnalité manquante trouvez-vous que l'application devrait avoir ?
- 6- Avez-vous trouvé des fonctions superflues ? Si oui, lesquelles ?
- 7- Utiliseriez-vous une telle application ? Pourquoi ?
- 8- Croyez-vous que l'utilisation de cette application pourrait améliorer la vitesse, l'aisance ou encore la qualité de votre travail ? Si oui, expliquez brièvement comment ?