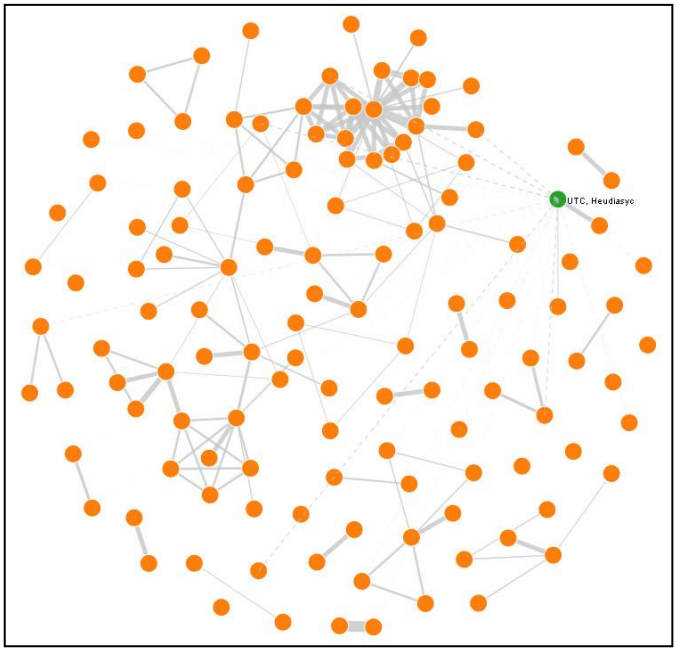
Création de nouvelles connaissances décisionnelles pour une organisation via ses ressources sociales et documentaires

Etienne Deparis

Par **Etienne DEPARIS**

Création de nouvelles connaissances décisionnelles pour une organisation via ses ressources sociales et documentaires

Thèse présentée
pour l'obtention du grade
de Docteur de l'UTC



Soutenue le 19 décembre 2013
Spécialité : Informatique

D2129

Université de Technologie de Compiègne

Heudiasyc

Thèse

En vue de l'obtention du titre de Docteur

Spécialité « Informatique »

présentée par

Étienne Deparis

Création de nouvelles connaissances décisionnelles pour une organisation via ses ressources sociales et documentaires

Thèse soutenue le 19 décembre 2013 devant le jury composé de :

M^me^	Pascale Zaraté IRIT, Université Paul Sabatier, Toulouse	Professeur, *Rapporteur*
M.	Serge Garlatti Dept. Informatique, Télécom Bretagne, Brest	Professeur, *Rapporteur*
M.	Pierre Morizet-Mahoudeaux Heudiasyc, Université de Technologie de Compiègne	Professeur, *Examinateur*
M.	Didier Bazalgette Domaine Hommes et Systèmes, Direction Générale de l'Armement, Paris	Responsable de domaine scientifique, *Examinateur*
M^me^	Cécile Bothorel Dept. Lussi, Télécom Bretagne, Brest	Enseignant chercheur, *Examinateur*

Thèse encadrée par :

M^me^	Marie-Hélène Abel Heudiasyc, Université de Technologie de Compiègne	Maître de conférence, HDR, *Directrice*
M^me^	Juliette Mattioli LRASC, Thales Research & Technology, Palaiseau	Directrice de laboratoire, *Co-encadrante entreprise*
M^me^	Gaëlle Lortal LRASC, Thales Research & Technology, Palaiseau	Chercheur, *Co-encadrante entreprise*

À Gaëlles

Remerciements

Malgré les apparences, une thèse ne s'effectue que rarement seul et c'est pourquoi j'aimerais tout d'abord remercier l'ensemble des personnes m'ayant accompagné et encouragé durant ces trois années.

Je souhaite tout d'abord remercier Pascale Zaraté et Serge Garlatti de m'avoir fait l'honneur d'accepter de rapporter cette thèse. J'aimerais remercier ensuite Pierre Morizet-Mahoudeaux et Cécile Bothorel d'avoir accepté de faire partie de mon jury de thèse.

Cette thèse existe grâce au concours de la Direction Générale de l'Armement (DGA) et en particulier de Didier Bazalgette que je veux vraiment remercier. Cette association, tant scientifique que pratique, a permis ma participation à des évènements très enrichissants comme les doctoriales ou le forum de l'innovation qui ont sans aucun doute favorisé l'enrichissement de mes recherches.

Mes remerciements vont, bien sûr et surtout, à ma directrice de thèse, Marie-Hélène Abel. Merci de m'avoir proposé ce sujet sans quoi rien de tout ceci n'aurait pu se faire, pour la confiance accordée au cours de ces trois ans et pour l'accompagnement prodigué.

Je veux également remercier mes encadrants au sein de Thales Research & Technology : Juliette Mattioli pour son soutien sans faille et Gaëlle Lortal pour son suivi, ses conseils, sa rigueur et ses encouragements.

Mes pensées se tournent ensuite vers l'ensemble des membres du laboratoire de mathématiques et techniques de la décision, devenu entre temps le laboratoire de raisonnement et d'analyse dans les systèmes complexes, d'une part, et de décision et optimisation d'autre part. Merci pour votre accueil chaleureux durant ces trois années et les échanges que nous avons pu avoir, tant intellectuels que récréatifs, nourrissant une synergie collaborative transverse *win-win* de nos activités.

J'aimerai remercier, enfin, tout ceux qui ont su me supporter et me changer les idées durant ces trois ans, aux premiers rangs desquels ma famille et les bonheurs Jocelin et Nathanaël, Maman pour la relecture globale du manuscrit un dimanche à potron-minet, Gaëlle pour les *rushs* de travail, les nuits blanches et les cafés, les élu·e·s Antoine, Camille, Louise, Thibaut, Vincent, les voisin·e·s d'Heudiasyc Kevin, Qiang, Xuan et bien sûr Adeline pour le pied à l'étrier. Sans oublier, évidemment, Berthe et Jérémy la *team* anti-blues, Pascal mon co-détenu

de cantine et de RER, les chefs Louis et Élodie et tous les autres ami·e·s que je ne pourrai malheureusement pas citer un·e à un·e ici sous peine de devoir éditer cette thèse en deux tomes. Du fond du cœur, merci.

Si finalement j'ai pu arriver au bout de cette thèse sans rien lâcher, je le dois entièrement à tes encouragements et tes attentions. Gaëlle, aucun mot ne saurait décrire l'importance que tu as à mes yeux.

Table des matières

Table des figures

Introduction

Contexte

En 2005, O'Reilly [2005] popularise le terme de Web 2.0. Ce terme a été forgé par son proche collaborateur Dale Dougherty afin d'exprimer le changement de paradigme entourant le Web. Ce dernier n'a cessé d'évoluer depuis son invention au CERN en 1993. Dougherty et O'Reilly souhaitent insister sur le changement de culture d'un Web orienté consommation — sites institutionnels, pages personnelles, etc. — à un Web orienté participation — blogs, wikis, minimessages, partage de photos, etc.

À l'aide de technologies déjà bien en place à l'aube du XXI^{e} siècle, les concepteurs de sites Web incitent progressivement leurs utilisateurs à contribuer en ajoutant leurs propres créations, voire en les concevant directement en ligne. Un nouveau moyen d'exécuter des applications sur un ordinateur apparaît au travers des applications Web qui permettent, à l'aide d'un simple navigateur Web, d'effectuer des tâches auparavant réalisées à l'aide de logiciels devant être installés sur la machine de l'utilisateur (édition de documents, gestion de boîte mail, etc.).

La facilité d'utilisation des applications Web, jointe au développement rapide des réseaux de télécommunication, démocratise l'accès au Web et rend ces applications très populaires.

L'un des facteurs de l'appropriation rapide dont a été l'objet le Web au travers de ces applications vient sans doute du fait qu'elles ont agit, parfois à leurs dépens, comme de véritables catalyseurs des pratiques de communautés humaines. Ces dernières ont trouvé dans les applications Web un moyen aisé de s'affirmer, se rassembler, communiquer.

À titre d'illustration de site ayant réussi à agréger des communautés humaines, nous pouvons citer deviantArt [1] ou MySpace [2] ou encore Facebook [3], sans doute le plus connu. Ces trois sites ont cependant connu des destinées différentes. Le site deviantArt, ouvert dès 2000, permet à ses membres de partager leurs œuvres graphiques. Il s'est affirmé rapidement comme l'une des références incontournables dans ce domaine, position qu'il possède toujours à l'heure actuelle en 2013. Fondé en 2004, MySpace tente de profiter de la popularisation du

1. http://www.deviantart.com/ (accessible le 10 novembre 2013)
2. http://www.myspace.com/ (accessible le 10 novembre 2013)
3. https://www.facebook.com/ (accessible le 10 novembre 2013)

concept de blog en offrant à ses membres un espace personnalisable et la possibilité de publier leurs créations en ligne. En particulier, la capacité de mettre en ligne des compositions musicales scelle le destin du site : par un phénomène d'entraînement, à la suite de quelques groupes de musique qui y trouvent un moyen rapide et pratique de publier leurs nouvelles compositions ou annoncer leurs dates de concerts, MySpace devient rapidement une référence contemporaine, sur le Web, de la musique.

MySpace s'impose sur le devant de la scène jusqu'en 2008, année au cours de laquelle son influence est dépassée par celle de Facebook dont la portée communautaire dépasse très largement le cadre de la musique et s'exprime au travers de ses « pages » et « groupes » aux thématiques variées. Ce changement de situation — Facebook et MySpace ont été créés la même année — peut s'expliquer de différentes manières. Retenons surtout que cela illustre bien le caractère mouvant des communautés humaines qui vont évoluer, se recomposer ou disparaître dans le temps.

Ces exemples parmi d'autres sont emblématiques de la manière dont les applications Web peuvent servir de révélateurs et d'amplificateurs à des communautés préexistantes à des échelons locaux. Tout individu participe, via les interactions qu'il entretient avec d'autres, à différentes communautés. Ces interactions sont enrichissantes pour les parties prenantes car elles permettent d'échanger des connaissances entre pairs.

Le contexte actuel de fort développement des applications Web et, en particulier, des médias sociaux suscite l'intérêt des organisations qui s'interrogent sur la capacité de ces outils à améliorer la collaboration de leurs membres via la reconnaissance des communautés numériques inhérentes à ces outils ou sur la pertinence de l'étude des données échangées sur ces outils comme sources d'informations. Cet intérêt et ce questionnement forment le point de départ de cette thèse.

Quelques définitions

Avant d'introduire plus avant nos réflexions, posons les définitions des termes courants qui seront utilisés tout au long de ce mémoire. Bien que la partie I soit l'occasion de préciser et discuter ces définitions, nous préférons en rappeler les grandes lignes dès à présent afin de faciliter la lecture de notre introduction.

Nous nous intéressons dans notre thèse à un type particulier d'applications Web ayant pris le nom de médias sociaux. Par application Web, nous désignons tous les sites internet qui, à l'instar des programmes, ou applications, pouvant être installés au sein du système d'exploitation d'un ordinateur, permettent à l'utilisateur d'effectuer des tâches à l'aide d'une interface dédiée. Les applications Web se démarquent des sites internet historiques qui n'étaient mis en

place qu'à des fins de consultation d'informations et non de productions ou d'interactions.

Concernant les médias sociaux, nous retenons la définition de Kaplan et Michael [2010] que ce terme désigne « un type d'applications Web construites à partir des idées et technologies dites Web 2.0 et permettant la publication et l'échange de contenus créés par ses utilisateurs [4]. » Cette définition limite cependant les médias sociaux au Web 2.0. Or, des applications antérieures à l'apparition du Web 2.0 peuvent être considérées comme des médias sociaux, telles Usenet ou les forums.

Le terme social dans « média social » permet d'insister sur la capacité qu'ont ces applications d'agréger des communautés humaines autour d'un outil. Ces communautés peuvent s'apparenter au réseau social des individus présents sur ces plates-formes numériques. Nous définissons le réseau social comme un ensemble d'acteurs (individus ou organisations) liés ensemble par un ensemble de relations. L'étude des interactions humaines au sein de communautés observées sous la forme de réseaux est un champ de la sociologie dont les prémices se trouvent dans les travaux d'Émile Durkheim et dans ceux de Simmel [1908]. Beaucoup plus récemment, l'étude des réseaux sociaux [Wasserman et Faust, 1994] s'est découvert un nouveau champ d'application avec l'essor du Web social [Wellman *et al.*, 1996, Erétéo *et al.*, 2008].

Nous appelons graphe social la formalisation d'un réseau social sous la forme d'un graphe composé de nœuds représentant les acteurs du réseau et de liens typés entre ces nœuds. Différents types de liens peuvent unir deux mêmes nœuds du graphe, de la même manière que différentes relations peuvent unir deux personnes au sein de leur réseau social.

Problématique sociale

Les normes sociales évoluent et il est aujourd'hui tout à fait normal de changer plusieurs fois d'organisation au cours de sa carrière. Chacune des organisations traversées représente une ou des communautés au sein desquelles la personne aura interagi. En naviguant d'une organisation à l'autre, les travailleurs se forgent un carnet d'adresses facilitant les collaborations inter-organisationnelles. La reconnaissance de cette notion de communauté et du réseau social d'une organisation apparaît comme un atout stratégique pour elle. Son réseau social et les connaissances qui y transitent sont ses meilleurs atouts dans la conception de nouveaux partenariats.

L'analyse de son réseau social permet à l'organisation de mieux identifier ses experts et ses influenceurs, les liens entre équipes et bien sûr les ponts qui la relient à ses partenaires

4. Traduction personnelle de « *Social Media is a group of Internet-based applications that build on the ideological and technological foundations of Web 2.0, and that allow the creation and exchange of User Generated Content.* » p. 61

ou concurrents. Le réseau social d'une organisation peut être considéré de deux manières distinctes :

1. Le réseau individuel composé des collaborateurs de l'organisation, des liens qu'ils entretiennent entre eux au travers de leurs travaux communs, affinités, rapports hiérarchiques, mais également avec les liens qu'ils ont pu nouer avec des individus extérieurs à l'organisation.
2. Le réseau communautaire composé des différentes communautés structurelles (services, départements, équipes de travail) ou spontanées de l'organisation et les liens qu'elles entretiennent entre elles, ou les connexions qu'elles ont pu tisser avec des communautés d'autres organisations.

L'analyse des réseaux sociaux, même en environnement professionnel, est soumise à une législation très stricte. De ce fait, nous ne nous intéresserons dans la suite de ce mémoire qu'au suivi du réseau social de l'organisation à l'échelle des communautés et de leurs connexions et non à l'échelle individuelle.

L'analyse du réseau social d'une organisation s'effectue en le matérialisant sous la forme d'un graphe. L'extraction de ce graphe peut se faire à partir des bases de connaissances de l'organisation. Ces bases contiennent habituellement de nombreux documents (propositions de réponse à un appel à projet, brevet, articles scientifiques, etc.) produits dans un contexte particulier. Les informations sur les auteurs, entre autres, font partie de ce contexte et permettent d'identifier des collaborations passées entre les auteurs, avec le document comme trace de cette collaboration. L'analyse de l'ensemble des documents de l'organisation permet de dessiner son graphe social.

Il convient alors de s'interroger sur la nature exacte de ce graphe social d'organisation, d'en identifier les composantes et de déterminer le processus d'extraction en lui-même.

Problématique décisionnelle

Nous pensons que l'analyse de médias sociaux utilisés dans l'organisation apporte plus de précision à ce graphe social. Les médias sociaux hébergent une grande activité que nous pensons plus représentative des intérêts actuels des composantes des organisations. L'analyse de l'utilisation des médias sociaux par les collaborateurs d'une organisation doit ainsi permettre d'appréhender plus finement l'activité des communautés qui composent cette organisation et donc préciser son graphe social. Il semble donc intéressant de capitaliser conjointement les informations tirées de l'activité des médias sociaux d'une organisation et les documents issus des bases documentaires de l'organisation au sein d'une seule et même base de connaissances afin d'en extraire un graphe social précis.

Cette capitalisation ne se fait pas sans but. L'intérêt premier des organisations est donc de pouvoir mobiliser les nouvelles informations issues des résultats de l'analyse du graphe social de l'organisation pour les faire intervenir dans des processus décisionnels. Ces processus décisionnels sont menés, dans les organisations, par des personnes que nous appelons décideurs. Ces derniers sont amenés à se poser fréquemment des questions comme : « avec qui puis-je nouer un partenariat ? », « quel est mon principal concurrent dans ce domaine ? », « puis-je monter un partenariat avec tel acteur ou dois-je plutôt le considérer comme un concurrent ? » Ces problématiques recouvrent ainsi classiquement les domaines tels que la détection et le suivi de tendances thématiques et l'identification des organisations travaillant sur les mêmes thématiques qui aideront le décideur à asseoir une stratégie pour son organisation.

L'ensemble des éléments présentés dans cette section nous amène aux verrous suivants :

— S'il semble intéressant de capitaliser à la fois des données provenant de médias sociaux et de bases documentaires au sein d'une même base de connaissances, il convient de s'interroger sur la manière de modéliser ces différents types de données de façon à permettre leur mobilisation homogène lors de l'extraction du graphe social d'une organisation.
— Si nous pensons que l'analyse du graphe social d'une organisation peut être source d'informations pertinentes pour les processus décisionnels d'une organisation, il convient d'identifier la nature de ces nouvelles informations et la forme qu'elles vont pouvoir prendre.

Problématique documentaire

En parallèle des vertus informationnelles sur la structure des organisations via leurs communautés internes, les médias sociaux jouent également un rôle important dans la transmission de connaissances entre utilisateurs au fil du temps en favorisant la publication et le partage d'informations entre membres. Or, les organisations cherchent justement les moyens les plus fiables de pérenniser leur capital de connaissances.

Les informations échangées sur les médias sociaux ne sont pas ou peu utilisées aujourd'hui par les organisations. L'une des raisons à cela est leur manque de prise en compte dans les processus de gestion de connaissances des organisations. Or, comme les discussions près des machines à café, ces échanges peuvent contenir des informations, être le véhicule d'une part de connaissance qui échappe à l'organisation.

La capitalisation des informations échangées sur les médias sociaux apparaît alors nécessaire pour enrichir les connaissances de l'organisation. Ces informations doivent compléter la base de connaissances de l'organisation, alimentée par ailleurs par d'autres types de don-

nées. L'alimentation d'une base de connaissances passe par un mécanisme de capitalisation des connaissances portées par les éléments qui sont effectivement enregistrés dans le système d'information d'une organisation. Cette capitalisation s'effectue en indexant les éléments en question à l'aide de concepts issus d'un référentiel élaboré et partagé au sein de l'organisation. Ce référentiel peut être plus ou moins formel et plus ou moins dynamique dans le temps.

La gestion de connaissances offre depuis de nombreuses années des solutions pour capitaliser le patrimoine documentaire des organisations. Peu de solutions, en revanche, existent à ce jour concernant les médias sociaux. Les informations qui y sont quotidiennement inscrites ou partagées restent décorrélées les unes des autres et n'entrent jamais dans les bases de connaissances des organisations.

L'ensemble des éléments présentés dans cette section nous amène à un verrou supplémentaire. Si la capitalisation des données issues des médias sociaux semble intéressante afin d'enrichir les bases de connaissances des organisations, il convient, en plus de la modélisation de la structure de ces données, de préciser le mécanisme sous-jacent d'indexation, au sein d'une seule et même base de connaissances, de différents types de données. L'utilisation et donc la définition d'une ontologie permet de répondre au besoin de modélisation des données et d'un mécanisme d'indexation.

Scénario illustratif

Afin d'illustrer notre problématique de thèse, la manière dont nous y avons répondu et les choix effectués pour y parvenir, nous suivrons le cas d'utilisation introduit dans cette section. Nous posons ici la situation initiale et présentons les différentes parties prenantes.

Nous posons comme cadre applicatif une organisation appelée « Argos » composée de différents départements dont un département de recherche appelé « Unité de Recherche et Développement. » Cette unité est elle-même organisée en différentes équipes, dont l'équipe « Web social », souvent surnommée l'équipe SoWeb.

Différentes personnes travaillent au sein de cette équipe, dont Alice, une développeuse sénior et Baptiste un ingénieur junior. L'équipe est sous la responsabilité de Denise qui a le poste de chef d'unité. Baptiste travaille actuellement en collaboration étroite avec Charles sur un livrable de projet européen. Charles a fondé sa propre petite ou moyenne entreprise (PME) qui porte son nom, Ivari corp. Il intervient en tant que consultant au sein du projet.

Argos possède un système de gestion de connaissances qui permet aux membres de ses différents départements de capitaliser les documents qu'ils traitent au quotidien. Cette capitalisation se fait au sein d'une seule et même base de connaissances afin de limiter la fragmentation du savoir de l'organisation et mieux identifier les pierres angulaires de l'organisation :

les thématiques communes à différents départements. La capitalisation prend en compte l'ensemble des documents traités par Argos, comme les notes techniques des unités de production, les propositions de brevet ou la documentation élaborée par différents services.

Afin d'améliorer la collaboration de ses membres et le partage d'informations, Argos a également mis en place des médias sociaux comme un wiki, une application de minimessages et des forums. Il s'agit de différentes applications Web avec lesquelles les membres de l'organisation peuvent créer, partager, annoter différents types de ressources.

Malheureusement, les informations produites et échangées sur ces médias sociaux ne sont pas capitalisées dans la base de connaissances d'Argos. Ces données ne sont donc pas mobilisées lors d'une requête dans la base de connaissances d'Argos. Pour pouvoir les réutiliser, il est nécessaire de passer par les outils de recherche spécifique à chaque application Web.

Au sein d'Argos, l'encadrement des équipes et la conduite de projets est déléguée à un certain nombre de responsables, comme Denise. Ces responsables sont amenés régulièrement à prendre des décisions comme le fait d'engager ou non leurs équipes dans des projets en collaboration avec d'autres équipes ou d'autres organisations et de manière plus large, de définir et faire évoluer la stratégie des équipes dont ils ont la responsabilité. Nous appelons alors ces responsables les décideurs de l'organisation.

Pour un décideur, faire le choix d'engager une ou plusieurs de ses équipes dans des projets collaboratifs dépend d'un certain nombre de paramètres comme :

— Les thématiques de la collaboration doivent être en adéquation avec les autres activités des équipes impliquées, c'est-à-dire ne pas ajouter de charge de travail supplémentaire sur les équipes en s'intégrant naturellement dans les travaux en cours ;
— Le projet doit être suffisamment porteur, c'est-à-dire ne pas disparaître des feuilles de route respectives une fois le travail engagé ;
— Le projet doit engager une réelle complémentarité des compétences : chacune des entités doit apporter quelque chose que l'autre n'a pas dans la collaboration.

Ces conditions demandent, pour être vérifiées, une bonne connaissance des positionnements théoriques des différentes entités impliquées dans le projet. Ces positionnements s'apprécient dans un environnement partenarial et concurrentiel différent, chaque entité possédant son propre réseau.

La prise de décision se fonde en partie sur l'analyse du graphe des collaborations passées des différentes entités en présence. Cette analyse permet d'identifier les tendances théoriques actuelles, les domaines porteurs etc. Elle permet également d'identifier les compétences supposées des différents partenaires sur ces domaines et *in fine* de leur accorder plus ou moins de crédit.

Argos recherche donc un moyen d'analyser le graphe des collaborations de ses différentes

entités et de leurs partenaires. L'analyse de ce graphe servira aux décideurs de l'organisation. Pour ce faire, il est nécessaire qu'Argos mette en place un système permettant d'extraire ce graphe de collaborations de sa base de connaissances. Cette extraction doit préférentiellement s'effectuer après avoir trouvé le moyen d'inclure les ressources produites et échangées sur les médias sociaux d'Argos au sein de sa base de connaissances.

Nous reviendrons régulièrement dans ce mémoire à ce scénario pour présenter plus en détail les différentes fonctionnalités du prototype développé et les choix de conception qui ont guidé sa réalisation.

Organisation du mémoire

Ce mémoire présente les principaux apports de la thèse. Ces apports se situent à différents niveaux, tant théoriques que pratiques. D'un point de vue théorique, un nouveau positionnement des notions de donnée, information, ressource, document numérique et connaissance a été défini. Ce positionnement a servi de fondation à un modèle ontologique de gestion de connaissances et d'aide à la décision. D'un point de vue pratique, un prototype de plate-forme a été réalisé, suivant les définitions apportées par le modèle théorique et ontologique. Le mémoire de thèse va permettre de revenir sur ces différents apports, leurs origines et leur maturation.

Ainsi, le mémoire balayera tout d'abord l'état de l'art des différents domaines impliqués dans nos travaux avant de présenter nos réalisations et de les discuter à la lumière des expérimentations que nous avons menées. Chacune des trois parties principales (partie I, II et III) sera introduite par un chapitre d'introduction et se clôturera par une synthèse.

La partie I, en se plaçant systématiquement dans un contexte organisationnel, explicite les différentes notions abordées dans ce mémoire. Le chapitre 2 revient tout d'abord sur la notion d'application Web et pose notre définition de média social. Cette définition s'intègre dans la notion plus générale d'écosystème numérique. La notion d'écosystème est définie dans une seconde section qui présente une définition de l'organisation, de ses composantes internes et de ses relations partenariales qui peuvent s'étudier sous la forme d'écosystèmes d'affaires. Ce chapitre se termine par la présentation de plusieurs écosystèmes numériques pour les organisations, permettant d'illustrer l'usage que ces dernières en font.

L'utilisation d'applications Web au sein des organisations est source de grandes quantités de données qui échappent aux systèmes de gestion de connaissances actuels. Le chapitre 3 présente les notions nécessaires à l'établissement de notre positionnement théorique sur la gestion de connaissances, en particulier lorsque ces connaissances proviennent de médias sociaux. Nous présenterons dans ce chapitre les différentes techniques d'étiquetage documen-

taire, dont celles issues du Web sémantique.

La capitalisation conjointe des données provenant de médias sociaux d'organisations et de bases documentaires, au sein d'une même base de connaissances, permet aux organisations d'obtenir une vision plus précise de leur réseau de collaborations. Les nouvelles connaissances induites par cette visibilité accrue sur leur réseau permet de nourir les processus de décisions internes aux organisations. Le chapitre 4 introduit les théories de la décision et des systèmes d'aide à la décision. Cette présentation se prolonge dans le cadre plus précis des systèmes interactifs d'aide à la décision (SIAD) fondés sur le principe de l'analyse des médias sociaux.

La partie II détaille notre positionnement original et les réalisations qui en ont découlé. Ainsi, le chapitre 7 présente notre positionnement théorique. Ce modèle théorique, appuyé sur les éléments présentés dans notre état de l'art, sert de fondation aux réalisations effectuées dans le cadre de cette thèse. De ce modèle théorique dérive directement le modèle ontologique présenté dans le chapitre 8. L'ontologie que nous avons conçue modélise un écosystème numérique de gestion de connaissances prenant en compte différents types de données, dont celles produites sur les médias sociaux. Cette modélisation nous a permis de développer un prototype présenté en détail dans le chapitre 9. Ce prototype s'articule autour de trois composants principaux :

1. une brique de gestion de connaissances prenant en compte différents types de données et les capitalisant au sein d'une seule base de connaissances ;
2. une brique d'analyse des données issues de la base de connaissances de la première brique ;
3. une brique de visualisation permettant d'afficher les résultats calculés par la brique d'analyse.

Afin de valider les choix effectués dans l'implémentation de notre modèle théorique sous forme ontologique et dans le développement du prototype d'écosystème numérique en découlant et donc, *in fine* afin de valider notre thèse, nous avons réalisé différentes expérimentations décrites dans la partie III. Ces expérimentations ont été menées sur deux fronts distincts.

Une première campagne au sein de l'Université de Technologie de Compiègne (UTC) a été effectuée afin de tester la brique de gestion de connaissances de notre prototype. Cette étude et ses résultats sont présentés dans le chapitre 12.

Une seconde campagne de test a pu être menée dans le cadre du laboratoire de raisonnement et d'analyse dans les systèmes complexes de Thales Research & Technology. Cette campagne avait pour mission le test de la brique inférentielle et de celle de visualisation. Le chapitre 13 présente cette étude et ses résultats. L'étude en elle-même a été découpée en deux phases : une première phase de calibrage des outils logiciels développés, présentée dans la

section 13.2 ; une seconde phase d'entretiens auprès de décideurs industriels de Thales dont les comptes-rendus sont donnés dans la section 13.3.

La partie IV nous permet de conclure ce mémoire, en revenant sur les grandes thématiques qui nous auront guidé tout au long de cette lecture. Cette partie nous donnera également l'occasion d'exprimer les perspectives de nos travaux de recherche dans le chapitre 16.

Première partie

Les connaissances échangées sur les médias sociaux des organisations au service des décideurs

1. Introduction

Cette thèse aborde des sujets et théories de différents domaines évoluant en parallèle. Nous traitons en effet des problématiques touchant les domaines des technologies de l'information et de la communication (chapitre 2), de la gestion de connaissances (chapitre 3) et de l'aide à la décision (chapitre 4).

Cette première partie est l'occasion de revenir sur ces différents domaines et positionner nos définitions des principaux termes que nous utiliserons pour décrire les travaux réalisés durant cette thèse. Cette partie nous permettra également de balayer les théories et applications pouvant nous aider à répondre aux problématiques évoquées en introduction.

Nous reviendrons ainsi sur la structuration des organisations en différentes communautés liées entre-elles par le biais de différentes collaborations. Nous verrons comment retrouver la trace de ces collaborations à l'aide des ressources sauvegardées dans les bases de connaissances des organisations. Nous verrons comment les échanges ayant lieu au sein des applications Web et des médias sociaux peuvent également être vus comme des ressources à partir desquelles il est possible de tirer des enseignements. Ces connaissances peuvent servir de sources d'information, en complément des connaissances tirées de bases documentaires des organisations, dans le cadre d'outils d'aide à la décision.

2. Organisations et médias sociaux

2.1 Les médias sociaux

Un média social se définit comme une application Web permettant à ses utilisateurs de publier et d'échanger de l'information et, ce faisant, par le biais d'intérêts partagés sur les contenus publiés et partagés, de forger une communauté. Kaplan et Michael [2010] définissent les médias sociaux comme « un type d'applications Web construites à partir des idées et des technologies dites Web 2.0 et permettant la publication et l'échange de contenus créés par ses utilisateurs[4]. » Le terme social permet également d'insister sur la capacité qu'ont ces applications d'agréger des communautés humaines autour d'un outil. L'analyse des échanges effectués au sein des médias sociaux apporte un point de vue unique sur l'activité et l'état d'une communauté humaine.

La notion d'applications Web est apparue en prolongement de la notion plus répandue de Web 2.0. Lorsqu'il crée le Web en 1989, Tim Berners-Lee propose surtout l'utilisation d'un protocole permettant d'échanger des hypertextes [Berners-Lee et Cailliau, 1992]. Ce protocole est connu sous le nom d'*HyperText Transfer Protocol* (HTTP).

Progressivement, l'avancée des technologies liées au Web a permis la conception de sites Web permettant au visiteur d'interagir directement sur le contenu qu'il consultait, voire lui permettant d'ajouter d'autres contenus. À l'instar des applications logicielles pouvant s'exécuter sur un ordinateur et permettant à leurs utilisateurs d'effectuer des tâches conduisant à la production d'objets numériques (rédaction de documents, édition de photographies, etc.) ou à la réalisation d'un service (visionnage d'un film, écoute de musique, etc.), les applications Web permettent à leurs utilisateurs d'effectuer les mêmes tâches mais au travers de leur seul « navigateur » Web.

Ainsi, les applications Web ont réduit l'écart qui pouvait exister entre le producteur de contenu — le *webmaster* — et son lecteur. Cette tendance générale du Web d'autoriser la participation des membres a pris le nom de Web 2.0 pour la différencier des anciens usages qui réduisaient l'internaute au simple rôle de lecteur passif. La rupture du Web 2.0 a donc été de permettre aux utilisateurs du Web de contribuer directement à l'enrichissement des informations qu'ils consultent.

Là où, historiquement, l'utilisateur ne pouvait que faiblement influer sur la production éditoriale d'un média au travers de commentaires, courriers, etc., les avancées technologiques du Web lui permettent désormais de s'impliquer à différentes étapes de cette production : production (rédaction de billets, soumission d'image, de vidéos, etc.), organisation (étiquetage de photographies, partage de liens, etc.), relecture (commentaires, réponses, etc.). Toutes ces tâches s'effectuant au vu et au su des autres utilisateurs, elles s'inscrivent dans une démarche sociale de participation au sein de la communauté des utilisateurs, communauté à laquelle l'utilisateur s'identifie [Fisher *et al.*, 2006].

La notion de média social apparaît donc comme une sous-partie des applications Web, nécessitant des interactions sociales. Ces interactions sociales s'apparentent majoritairement à des processus de communication ou de collaboration.

Nous définissons la communication comme le fait de transmettre de l'information d'une source bien déterminée vers une cible (ou destinataire) également déterminée. Si le destinataire répond, c'est-à-dire s'il réémet une information à destination de la source initiale, nous parlons de dialogue. Si nous considérons que nous pouvons identifier précisément une information et suivre son cheminement de sa source à son ou ses destinataire(s), nous définissons dès lors le principe de diffusion de l'information.

Pour sa part, la collaboration va désigner l'ensemble des processus nécessaires, pour différentes parties prenantes, individuelles ou organisationnelles, à la réalisation d'un objectif clairement identifié.

Les activités sociales effectuées au sein d'applications numériques soulèvent la question de leur confidentialité. Bien que le Web ait été conçu comme un vaste espace public permettant à chacun d'accéder à l'ensemble des informations y étant placées, l'utilisation des applications Web montrent les limites de cette ouverture. Les utilisateurs ajoutant de l'information sur le Web réclament la possibilité de réguler l'accès à ces informations. Cette régulation peut se décomposer en trois niveaux d'accès différents, que l'on peut également appeler audience :

La sphère privée Les informations échangées par les individus sont confidentielles et ne touchent qu'un tout petit nombre de personnes, clairement identifiées par la source de l'information.

La sphère communautaire Les informations échangées par les individus peuvent se partager plus librement, mais leur diffusion reste contrainte au sein de groupes aux frontières établies (même thématique, même service de l'organisation, etc.).

la sphère publique Les informations échangées sont librement accessibles par tous. Il s'agit d'une sorte de version standard du Web.

À chaque audience correspond également un certain degré d'attente de la source vis-à-vis de l'implication de la cible. Cette attente décroît de la sphère privée vers la sphère publique.

Ces définitions nous permettent d'envisager, selon nous, une classification des principales applications Web. Nous retenons les cinq applications Web suivantes :

Blogs Contraction de *web log*, traduit en français par cahier ou journal du Web. Un blog permet à son auteur de publier des articles, souvent présentés aux visiteurs par ordre antéchronologique, les nouveaux articles étant accessibles dès la première page. Cette application se prête bien à l'écriture d'articles d'opinion, d'articles de synthèse ou d'explication.

Forums Les forums sont des applications permettant à leurs utilisateurs d'échanger des messages organisés par fils de conversation. Cette application se prête bien à la tenue de débats ou à l'organisation d'une foire aux questions.

Wikis Les wikis sont des applications permettant la rédaction collaborative. L'une des fonctionnalités essentielles des Wikis est l'historisation des différentes versions des pages créées et modifiées par les utilisateurs. Cette application se prête bien à la création et au suivi d'une base documentaire.

Minimessages Popularisés par les applications Twitter, Facebook et assimilées, les applications de minimessages ou *microblogs* en anglais, consistent à entretenir un flux d'actualités (parfois appelé « mur » ou *wall*) courtes, voire limitées expressément.

Salon de discussions Les salons de discussions permettent à différentes personnes de discuter de manière synchrone. Lorsque le dispositif technique n'autorise qu'un maximum de deux personnes à communiquer, il est fréquent de parler de messagerie instantanée. Ce mode de communication passait auparavant par des logiciels indépendants. Néanmoins l'avancée des technologies a permis assez vite de transposer les discussions textuelles au sein d'un site Web (sous forme d'appliquette Java, puis de script AJAX), puis l'audio et la vidéo (à l'aide de technologies émergentes comme WebRTC).

Suivi des tâches, forges logicielles Dans le monde plus restreint des applications Web techniques, les forges logicielles et les applications de suivi de tâches ou tickets désignent l'ensemble des outils venant appuyer l'avancée d'un projet technique. La socialisation se retrouve, dans ce genre d'application, dans la possibilité d'interagir via des commentaires sur les tâches ou les tickets qui y sont créés et les différentes analyses pouvant être effectuées sur les soumissions de codes.

Application de réseautage Les applications de réseautage présentent un ensemble de fonctionnalités permettant d'aider les utilisateurs à gérer leur réseau de connaissances, collègues au travers d'outils de mise en relation, de fiche de profils, etc. Il s'agit fondamentalement pour les utilisateurs de s'afficher sous la forme d'une page de profil les décrivant, à partir de laquelle d'autres membres pourront réagir, se lier, etc.

Table 2.1 – Classification de quelques applications Web

Audience / Type d'interactions	**Communication**	**Collaboration**
Sphère privée	Messagerie instantanée (dialogue)	
Sphère communautaire	Blog, Salons de discussions (discours)	Wiki, Suivi de tâche (coopération)
Sphère publique	Minimessages, Réseautage (publication)	Forum (crowdsourcing)

Le tableau 2.1 classifie donc les différentes applications Web sélectionnées en les positionnant au croisement d'une audience et d'un type d'interactions particulier. Nous avons donné un nom à chacun de ces croisements, afin de mieux les repérer par la suite. Il s'agit du mot entre parenthèses dans chacune des cases du tableau.

Si la case au croisement de la sphère privée et de la collaboration reste vide dans le cadre de cette thèse, il s'agit d'un choix qui ne reflète pas la réalité. De nombreuses applications Web peuvent correspondre à ce positionnement comme les applications généalogiques ou les jeux. Nous ne les pensons cependant pas pertinentes dans le monde des organisations.

La mise en place de diverses applications Web interconnectées entre elles pour soutenir la collaboration des membres d'une organisation conduit à la formation de plates-formes logicielles complexes.

La constitution de ces environnements entraîne une meilleure cohésion de leurs composants : il est ainsi très facile de faire interagir des éléments d'une application à l'autre sur ces plates-formes. Google ou Microsoft proposent par exemple directement l'ouverture et l'édition de documents bureautiques au sein de leur suite respective dès la réception d'une telle pièce jointe au sein de leur Webmail.

L'interdépendance forte des différents composants de ces environnements fait leur force. Prises séparément, ces applications n'apporteraient pas forcément de valeur ajoutée par rapport à d'autres produits, mais insérées dans ces plates-formes complexes numériques, elles deviennent intéressantes. De telles plates-formes numériques, à l'instar des écosystèmes naturels où un réseau d'êtres vivants œuvre à la survie de l'ensemble, peuvent être appelées écosystèmes numériques.

Le terme d'écosystème [1] désigne, dans la nature, l'équilibre forgé par des êtres vivants vis-à-vis de leur milieu de manière à favoriser le maintien et le développement de la vie. Les écosystèmes sont souvent le résultat de l'association d'espèces vivantes se partageant un même environnement et réalisant, pour la survie de l'écosystème, des tâches identifiées. Sur demande de l'ONU, une étude des écosystèmes naturels a été effectuée au début des années 2000 et s'est concrétisée par le rapport Assessment [2005]. Ce rapport identifie quatre types de tâches nécessaires à la bonne marche des écosystèmes. Ces types de tâches sont appelés services :

— service de production et d'approvisionnement (ventilation, apport d'eau douce, de nutriments, etc.) ;
— service de régulation (maladie, climat, etc.) ;
— service culturel (beauté, récréation, etc.) ;
— service de support (cycles de l'eau, du carbone, photosynthèse, etc.).

Le terme d'écosystème a transpiré dans de nombreux domaines pour représenter des systèmes complexes d'acteurs dont la survie individuelle est interdépendante de celle des autres acteurs ou en tout cas d'une cohésion globale des acteurs.

Le terme d'écosystème numérique désigne, sur le Web, un ensemble d'applications Web interconnectées dont la survie individuelle dépend de la bonne santé — en termes de fréquentation, de revenus — de l'ensemble. Ces écosystèmes peuvent s'étudier à la lumière du rapport de l'ONU avec la caractérisation des quatre services utilitaires. Pour les écosystèmes numériques, les services de production sont principalement tenus par les utilisateurs eux-mêmes ; les services de régulation par les éditeurs des principales applications Web de l'écosystème ; les services culturels par la myriade d'applications Web pouvant se greffer sur une application principale ; les services support, enfin, sont tenus par les constructeurs d'appareils électroniques et les éditeurs de logiciels permettant d'accéder aux applications Web.

Google est un exemple parfait d'écosystème numérique. La messagerie Gmail peut être vue comme le service de régulation de l'écosystème : sans compte email il est en effet impossible de bénéficier de l'ensemble des services de l'écosystème, comme ajouter des articles sur son blog (Blogger), publier des photographies (Picasa), des vidéos (Youtube) ou même utiliser son téléphone portable (Android). D'un autre côté, rien ou presque ne différencie d'un point de vue fonctionnel l'application Web de publication de photographie Picasa d'une autre comme Flickr. Son « succès » et sa survie dépendent exclusivement de son appartenance à un écosystème qui lui fournit son oxygène : les utilisateurs.

Le moteur de recherche en lui-même a le rôle de service de support. La technologie d'indexation de contenu et de recherche est en effet utilisée de manière transparente au sein

1. Terme utilisé pour la première fois, d'après Wikipedia, dans Tansley, Arthur George (1935). « The Use and Abuse of Vegetational Concepts and Terms », *Ecology* Vol. 16 nº 3 : 284–307.

des différentes applications Google : pour retrouver son courrier dans Gmail, pour chercher des vidéos dans Youtube, etc. Google Doc, Youtube, Blogger ou Picasa peuvent sans conteste prendre le rôle de production en permettant à leurs utilisateurs de créer, échanger et modifier du contenu au sein de l'écosystème. Enfin, Google+ ou Youtube peuvent être vus comme assumant le rôle de service culturel, apportant une possibilité récréative aux utilisateurs de l'écosystème.

2.2 Les organisations

Dans le cadre de cette thèse nous nous intéressons à l'usage des médias sociaux au sein des organisations. Par le terme d'organisation, nous entendons principalement les entreprises, quelle que soit leur taille, les universités et les organisations étatiques ou trans-étatiques comme l'ONU, l'OTAN ou les organisations non-gouvernementales (ONG). Nous ne nous intéressons par contre pas aux associations type loi 1901, associations sportives, etc.

Avec le terme d'organisation, nous souhaitons insister sur l'intérêt que nous portons à la structure interne des entreprises. En effet, l'un des problèmes liés à la modélisation de la structure sociale des organisations semble être la difficulté de modéliser les relations entre membres de l'organisation [Bironneau et Martin, 2002] — relation de travail, hiérarchiques, jeux de pouvoir et d'influences, etc. De même, l'étude et la surveillance d'un réseau humain est soumis à une législation précise, pouvant agir comme une barrière dans certains cas. Dans cette thèse, nous laissons de côté ces considérations pour nous intéresser exclusivement à un réseau social composé d'organisations et de leurs composantes internes.

Nous considérons donc les organisations comme un réseau d'éléments emboîtés de granularité plus ou moins fine pouvant se connecter au gré des besoins stratégiques. Cette emboîtement d'« éléments » est appelé « affiliation ». L'élément le plus fin d'une affiliation est « l'équipe ». Nous appelons « entité » les objets représentant les équipes et leur chaîne d'affiliation complète.

Cependant, il ne faut pas réduire les organisations à ce seul découpage hiérarchique. Les membres de l'organisation peuvent en effet également se réunir de manière informelle, comme les groupes de collègues se retrouvant pour discuter au coin de la machine à café. Cohendet et Diani [2003] remarquent ainsi la nécessité, lorsqu'il s'agit de définir les différentes communautés évoluant au sein des organisations, d'en différencier deux types :

1. Les groupes institutionnels, créés par une entité hiérarchique, regroupent différents collaborateurs autour d'une mission déterminée. Il s'agit typiquement des équipes de travail au sein des organisations.

2. Les groupes dynamiques ou informels, créés à l'instigation des membres de l'organisation sans validation hiérarchique, regroupent différents collaborateurs autour d'un sujet précis, d'une personne ou des deux. L'ouverture d'un tel groupe aux autres membres de l'organisation peut être restreinte, dans le cadre de groupes privés. Les médias sociaux sont intéressants sur ce point par la faculté qu'ils ont de permettre l'identification de ces communautés informelles — via la liste des participants d'un forum, les participants d'un blog thématique, etc. — invisibles ou presque jusqu'alors.

Cette notion de communauté formelle ou informelle est importante pour les organisations car Gueye [2004], en citant Hutchins et Kirsh [2], souligne que le processus d'acquisition de connaissances n'est plus forcément une activité individuelle mais le fruit d'interactions au sein d'une communauté.

La mobilisation de la connaissance de l'organisation peut s'effectuer de manière volontaire au sein de communautés épistémiques [Gueye, 2004] qui ne sont autres que des groupes créatifs rassemblés de manière artificielle pour maximiser les compétences en présence. Les communautés épistémiques s'apparentent aux groupes institutionnels cités plus haut.

La connaissance peut également être mobilisée parmi celle, prégnante, qui règne au sein des communautés de pratique Wenger [1998] qui agrègent des personnes aux compétences homogènes selon leur bon vouloir dans la réalisation d'un but commun. Ces communautés s'apparentent ainsi aux groupes dynamiques décrits plus haut.

Cohendet et Diani [2003] résument ainsi (p. 705) : « Les communautés épistémiques [...] sont réellement orientées vers la création de nouvelles connaissances, et les communautés de pratique [...] sont orientées vers la réussite d'une activité, et pour lesquelles la création de connaissances est un débordement involontaire. »

La notion de groupe d'organisations ne doit pas se limiter à l'échelle individuelle. Seufert *et al.* [1999] rappellent combien les entreprises se regroupent aujourd'hui préférentiellement en partenariats industriels et ne travaillent plus seules. Nous pouvons alors ré-utiliser la notion d'écosystème définie dans la section précédente. Dans le monde des affaires, la notion d'écosystème se comprend comme un réseau d'organisations et de services mis en place pour favoriser le développement de leurs affaires et donc leur survie. Nous parlons alors d'écosystème d'affaires (*business ecosystem* [Moore, 1993]).

Nous pouvons transposer la notion de service des écosystèmes naturels aux écosystèmes d'affaires. Dans le cadre des écosystèmes d'affaires, certains partenaires vont jouer le rôle des services de production (usines, R&D, etc.), d'autres celui des services de régulation (les diffé-

2. Hutchins, Edwin (1995). « How a cockpit remembers its speeds », *Cognitive science* Vol. 19 n° 3 : 265–288 et Kirsh, David (1999). « Distributed Cognition, Coordination and Environment Design », *Proceedings of the European Cognitive Science Society.*

rentes agences étatiques, la Bourse, etc.), ou des services culturels (bureaux d'études, archives, etc.) et enfin celui des services de support (restaurants, universités, ressources humaines, etc.).

Il n'est pas rare que les organisations s'impliquent dans différents écosystèmes d'affaires, chacun pouvant avoir une portée ou un objectif différent. Comme nous l'avons vu précédemment, l'objectif intrinsèque d'un écosystème d'affaires est la survie des différents partenaires par la production de valeur. L'innovation et la création de connaissances sont perçues comme des facteurs importants de réussite par les organisations. Nous pouvons alors également parler d'écosystèmes de connaissances [Shrivastava, 1998, Bray, 2007].

À une échelle différente de la notion de communauté épistémique, les écosystèmes de connaissances s'appuient également sur l'idée que l'innovation et l'émergence de nouvelles connaissances vont être considérablement améliorées dans un environnement favorisant des interactions plus libres entre acteurs, particulièrement au sein de structures plus informelles. Les écosystèmes de connaissances peuvent être vus comme des réseaux d'organisations partageant le but commun d'améliorer ensemble leurs connaissances d'un domaine en mutualisant leurs compétences. Ils peuvent donc être vus comme un cas particulier des écosystèmes d'affaires, dont la principale valeur à créer serait de la connaissance.

Afin de permettre la bonne circulation du savoir au sein de ces écosystèmes organisationnels, la mise en place d'outils numériques est fréquente. Nous pensons que les écosystèmes numériques peuvent jouer un rôle dans l'outillage d'un écosystème de connaissances. Les écosystèmes numériques permettent en effet, au travers de leurs médias sociaux constitutifs, l'étude du réseau social organisationnel de l'écosystème d'affaires. Nous verrons dans la section suivante de quelle manière les organisations adoptent ces écosystèmes numériques et les médias sociaux.

2.3 Usage des médias sociaux en organisation

L'adoption croissante des médias sociaux rencontre encore régulièrement des freins, étant vus comme potentiellement nuisibles à la productivité d'une équipe. Il est craint en effet que l'utilisation de ces applications soit faite à des fins personnelles ou simplement pour occuper une certaine oisiveté [Ferreira et du Plessis, 2009, Shirky, 2008].

Cependant, le succès des applications Web 2.0 a mené les organisations à considérer l'usage de tels outils. Conein [2004], Pène et Dubois [2009] rappellent la possibilité d'innovation que représente les réseaux sociaux par la liberté offerte à leurs utilisateurs de s'organiser et d'échanger. Yang et Chen [2008] montrent qu'une condition à la bonne circulation de la connaissance dans l'organisation est de faciliter la mise en relation de ses membres. Cross *et al.* [2001] (p. 106) montrent d'ailleurs que la source principale d'information au sein des organi-

sations reste le contact humain et non les bases documentaires. Pène et Dubois [2009] notent toutefois le paradoxe de l'adoption des outils numériques de réseautage par les entreprises encore majoritairement construites sur des structures hiérarchiques fortes.

Nous pouvons ainsi dresser deux directions principales qui poussent les organisations à adopter de tels outils :

1. Améliorer la gestion de ses connaissances en incorporant de nouvelles sources de connaissances inexploitées.
2. Améliorer la collaboration de ses membres et de ses différentes composantes au sein de son écosystème d'affaires ou de connaissances.

La maîtrise de l'information et la gestion de connaissances sont essentielles pour les organisations qui essayent de les capturer et de les capitaliser de manière à pouvoir les réutiliser le moment venu. Il s'agit d'un enjeu stratégique dans un monde en constante évolution [Argote et Ingram, 2000, Ermine, 2000, Waltz, 2003]. En effet, il est primordial pour les organisations de pouvoir se positionner rapidement au sein de leurs secteurs d'activités.

Avec l'émergence des réseaux sociaux en ligne et la formalisation numérique des réseaux d'expertise qui auparavant restaient relativement inaccessibles au coin des machines à café, les entreprises ont pu se doter d'un formidable outil permettant d'observer et de suivre l'activité de leurs communautés constitutives. Le graphe social d'une organisation peut en effet être vu comme une carte dessinant ses forces et ses faiblesses, les experts et les influenceurs, les liens entre équipes, etc. Le réseau dépasse très souvent l'organisation elle-même par le biais de collaborations réalisées avec d'autres entreprises.

L'utilisation de médias sociaux au sein des organisations leur permet donc d'accéder à ce graphe des communautés institutionnelles ou informelles et par là même à un formidable outil de pilotage stratégique [Lazega, 1994]. Par ailleurs, l'utilisation de ces technologies leur donne accès à de nouvelles formes de connaissances, plus en prise avec la réalité quotidienne des activités de l'organisation. Cet accès va s'effectuer au travers de méthodes d'analyse des médias sociaux, sur lesquels nous revenons plus précisément dans la section 4.3.

Nous présentons dans la suite de cette section différentes plates-formes d'applications Web conçues pour les organisations. Cette liste n'est pas exhaustive mais reprend, selon nous, les environnements les plus complets et représentatifs du Web actuel. Dans un premier temps nous apportons une courte description de chacun d'entre eux, avant de croiser leurs différentes fonctionnalités dans un tableau synthétique. La figure 2.1 présente une capture d'écran par plate-forme présentée ici.

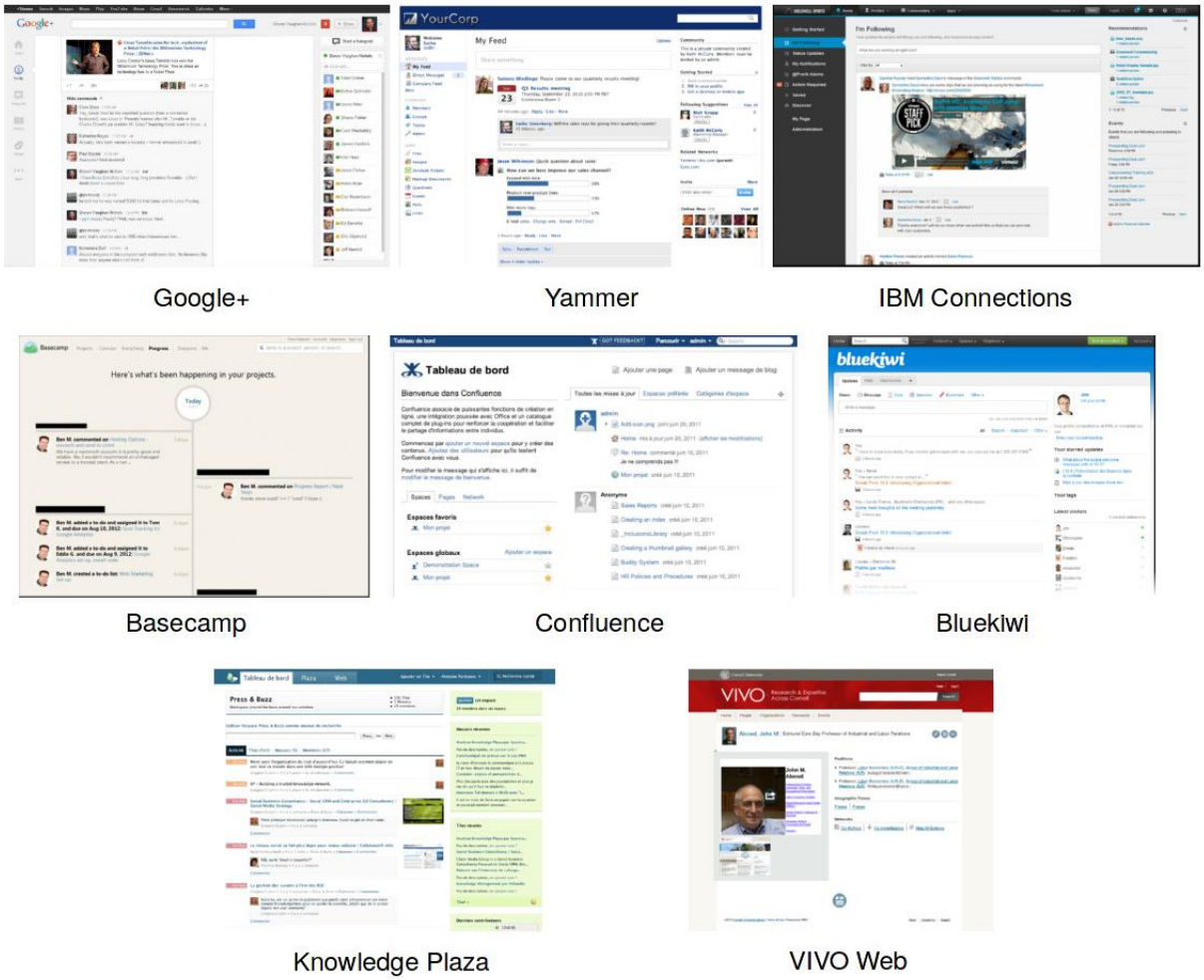


FIGURE 2.1 – Capture d'écran de quelques écosystèmes numériques pour les entreprises

2.3.1 Yammer

Yammer [3] est une *startup* proposant un ensemble d'outils permettant aux membres d'organisations de s'échanger rapidement de l'information et travailler en commun sur différents documents. Originellement conçue comme un clone de Twitter pour les entreprises, Yammer s'est depuis diversifiée en ajoutant des outils de gestion documentaire ou de *social bookmarking* [4]. La *startup* s'est fait racheter par Microsoft en juin 2012 pour l'inclure dans sa division Office et améliorer ainsi les fonctionnalités collaboratives de la suite bureautique professionnelle.

3. https://yammer.com (site accessible le 10 novembre 2013)

4. Le *social bookmarking* ou « partage de marques-pages » est une activité consistant à conserver, annoter, étiqueter des signets vers des pages Web visitées au sein d'une application Web qui permet alors de suivre les signets des autres utilisateurs et par là même découvrir de nouveaux sites. L'intérêt de ces applications, au-delà de la découverte de nouveaux contenus, est l'effort d'étiquetage et donc de classification du Web qu'elles ont apporté. Le site del.icio.us (http://delicious.com) est sans doute le site le plus connu et celui entraînant la plus grande communauté.

2.3.2 Google

Le géant du Web s'est rendu célèbre, au-delà de son moteur de recherche, par la mise à disposition gratuite d'un certain nombre d'applications pour ses utilisateurs [5]. Cette suite d'outils se compose, entre autres, de Google Docs, Gmail et Google+. Google Docs est l'équivalent d'une suite bureautique et permet l'édition à plusieurs, de manière synchrone ou asynchrone, de documents texte, tableurs ou présentations. L'utilisation d'une telle suite bureautique « dans le nuage » permet à ses utilisateurs de s'affranchir des problèmes de compatibilité de leurs habituels logiciels bureautiques respectifs et ainsi d'accéder plus facilement au contenu à partager, sur lequel collaborer. Gmail est un Webmail permettant à ses utilisateurs de consulter leurs courriers depuis n'importe quel navigateur Web. Google+, enfin, est la solution de réseautage social avec laquelle Google remplace progressivement plusieurs de ses services historiques pour offrir une expérience unifiée : Hangout remplace Talk (une application de messagerie instantanée), le mur d'activité remplace Google Buzz, Reader ou Wave (applications de partage d'informations et collaboration).

2.3.3 IBM Connections

Connections est la suite d'applications sociales pour l'entreprise proposée par le géant de l'informatique mondiale. Les applications proposées par IBM n'ont pas de nom commercial précis et désignent directement leur fonction principale. Ainsi, « Homepage » est un intranet d'entreprise permettant la publication d'articles et l'agrégation de différentes sources d'informations. « Microblogging » est un outil de minimessages pour les employés. « Profiles » est un réseau social pour les collaborateurs de l'entreprise, proposant des fiches de profil enrichies des différentes contributions que les membres peuvent effectuer avec d'autres applications de la suite d'IBM. « Communities » est une application permettant aux collaborateurs de se rassembler par centre d'intérêt. « Activities » est un outil de gestion de projet et suivi des tâches. Citons enfin l'existence de « Blogs », « Bookmarks », « Wikis », « Forum » et « Files » dont les noms sont directement évocateurs.

2.3.4 37 signals

La société 37 signals commercialise différentes applications Web liées principalement à la gestion de projets logiciels. La société s'est fait connaître avec « Basecamp », une application qui permet le suivi de projets d'une équipe à l'aide de *todo lists*, forum, partage de fichiers, gestion de versions, etc. Cette application permet à toute PME de se construire rapidement

5. L'enregistrement préalable est nécessaire pour pouvoir les utiliser.

et facilement un intranet l'aidant à suivre ses activités quotidiennes. Depuis, la société a enrichi son offre avec entre autres « Campfire », une application dédiée à la tenue de réunions virtuelles sous la forme de salons de discussions dont les retranscriptions restent accessibles après coup (ainsi que les éventuels fichiers échangés durant la conférence).

2.3.5 Atlassian

Atlassian est également une société spécialisée dans la gestion de projet informatique. Les deux applications historiques de la société sont « Jira » et « Confluence ». Jira est un outil de gestion de projet logiciel avec suivi des versions et des soumissions de codes (déposées par exemple sur la forge logicielle BitBucket acquise par Atlassian) permettant de rassembler les développeurs en équipes et suivre leurs activités. Confluence est historiquement un simple Wiki d'entreprise. Mais progressivement, Confluence est devenu un *framework* complet permettant le développement d'intranet riche d'entreprise via l'adjonction d'extensions au code de base (à la manière de Drupal par exemple). Pour concurrencer les offres similaires (Campfire de 37 signals, Chatter de Salesforces, Skype de Microsoft, etc.), Atlassian a récemment lancé « HipChat », un service de messagerie instantanée ou conférence d'entreprise.

2.3.6 Bluekiwi

Entreprise française rachetée par le groupe Atos, Bluekiwi propose une suite d'applications sociales pour les organisations autour d'un mur d'activité de leurs différents collaborateurs. Les différents outils attendus de ce type de plate-forme sont disponibles comme le partage de fichiers, la collaboration autour de documents, la création de pages Web, l'organisation de réunions, etc.

2.3.7 Knowledge Plaza

Knowledge Plaza est une entreprise basée à Louvain-la-Neuve en Belgique proposant un ensemble d'applications Web au sein d'une offre intégrée permettant d'utiliser, en tant qu'utilisateur, une même folksonomie pour identifier différentes ressources. Knowledge Plaza affirme en effet être la première organisation à avoir permis de faire du *social bookmarking* (du partage de liens, étiquetés à l'aide de mots-clés choisis par la personne initiant le partage) avec autre chose que simplement des pages Web.

2.3.8 VIVO

VIVO [Gewin, 2009] est un projet de recherche américain financé par l'université de Cornell et le *National Institutes of Health* cherchant à améliorer le partage d'informations scientifiques entre universités et en particulier la découverte de chercheurs conduisant des travaux similaires.

Le projet a développé une plate-forme Web *open source* permettant de gérer de très nombreux aspects de la vie quotidienne d'un campus universitaire et en particulier la gestion des ressources documentaires scientifiques. La pierre angulaire de la plate-forme réside dans l'ouverture des données capitalisées au travers d'une *Application Platform Interface* (API) permettant d'interconnecter différentes instances VIVO de différentes universités.

2.3.9 E-MEMORAe 2.0

E-MEMORAe 2.0 est l'environnement de gestion de connaissances pour du e-learning issues des travaux menés dans le cadre du projet MEMORAe [6]. Dans le cadre de ce projet de recherche, l'environnement E-MEMORAe 2.0 a fait l'objet d'une modélisation ontologique afin d'en décrire le fonctionnement et la manière d'y intégrer d'autres ontologies dans le cadre de cours d'université [Benayache, 2005, Leblanc, 2008]. Cette modélisation propose quelques fonctionnalités sociales comme un wiki, des forums, des salons de discussions ou encore des agendas partagés [Leblanc et Abel, 2008, 2010].

L'originalité de l'approche soutenue par ce projet tient dans la présentation constante à l'utilisateur d'une cartographie de concepts sur lesquels sont indexés les différents documents de la plate-forme.

2.3.10 Croisement des fonctionnalités

Dans le tableau 2.2 nous listons les différents écosystèmes numériques présentés précédemment et synthétisons les différentes fonctionnalités qu'ils proposent. Pour ce faire, nous utilisons la classification que nous avons définie dans la section 2.1.

Nous avons également fait le choix d'ajouter à cette synthèse les colonnes gestion documentaire et analyse des activités, qui nous semblent être des fonctionnalités importantes dans le monde des organisations.

Gestion documentaire Ensemble de fonctionnalités permettant aux utilisateurs de mettre en ligne, partager et annoter des documents produits par d'autres systèmes. La gestion

6. http://www.hds.utc.fr/memorae (accessible le 10 novembre 2013).

documentaire vient en complément des autres fonctionnalités liées à l'usage de médias sociaux.

Analyse des activités Ensemble de fonctionnalités permettant de donner à un administrateur des informations sur l'utilisation des différents outils mis en place sur la plate-forme, comme des outils d'analyse de réseaux sociaux, de synthèses d'activités, etc.

Table 2.2 – Classification de quelques environnements numériques pour les organisations

Plates-formes	**Dialogue**	**Discours**	**Publication**	**Coopération**	***Crowdsourcing***	**Gestion documentaire**	**Analyse des activités**
Yammer			X			X	X
VIVO			X			X	X
Atlassian		X		X			X
37 Signals		X		X	X		X
E-MEMORAe 2.0		X		X	X	X	
Knowledge Plaza	X	X	X	X		X	
Bluekiwi		X	X	X	X	X	X
Google	X	X	X	X	X	X	
IBM	X	X	X	X	X	X	X

Ce tableau nous permet de dresser trois catégories concernant les écosystèmes numériques s'étant développés pour les organisations :

— Les écosystèmes d'inspiration fortement documentaire : comme Yammer ou VIVO, ils ne proposent souvent que des fonctionnalités liées à la gestion électronique de documents. Les utilisateurs sont incités à partager de l'information, sans s'investir davantage. Ces écosystèmes peuvent être utilisés pour outiller la veille technologique ou commerciale d'une organisation.

— Les écosystèmes techniques : comme Atlassian ou 37 signals, ils mettent l'accent sur des outils facilitant la coopération des utilisateurs. Les outils de gestion électronique de documents sont souvent absents, au profit d'applications de gestion de projets.

— Les écosystèmes « complets » : comme Google ou IBM, ils s'attachent à offrir un large panel d'outils pour les organisations, leur permettant de répondre à leurs besoins de gestion de connaissances sur le long terme, comme à l'encadrement de projets sur le court ou moyen terme.

2.4 Synthèse

Les applications Web et les médias sociaux sont utilisés par les organisations qui voient en eux des outils numériques capables de les aider à améliorer la collaboration de leurs membres en interne, mais également avec leurs partenaires extérieurs. Il va s'agir par exemple de profiter de nouveaux moyens de communication (minimessages, blogs), d'améliorer les processus de veille (salons de discussions, forums, minimessages) et fluidifier la gestion de projets (forges logicielles, wikis, salons de discussions).

De fait, de nombreuses solutions ont été développées pour les entreprises, afin de leur permettre d'accéder rapidement et facilement à toute une gamme d'applications Web, dont des médias sociaux, organisées sous la forme d'écosystèmes numériques.

Cette thèse, du fait de son contrat d'accompagnement DGA-CIFRE, s'est déroulée à la fois dans un environnement universitaire à l'UTC et dans un environnement industriel au sein de Thales R&T. Ceci nous a permis de constater empiriquement l'usage qui était fait des applications Web et des médias sociaux au sein de ces organisations.

Dans les deux cas, des Wikis ont été mis en place pour faciliter l'accès et le partage d'informations, en particulier la documentation des différents projets entrepris et de leurs livrables. Pour Thales R&T, il s'agit également d'un moyen de capitaliser les demandes de publication et les propositions de réponse d'appels à projet. Il ne s'agit pas d'usage requérant la collaboration et la plupart des pages des deux Wikis n'a été éditée que par une seule et même personne.

Par sa vocation d'enseignement, l'UTC a été amenée à déployer une solution de plate-forme d'enseignement en ligne. Cette dernière propose, en plus des outils dédiés à l'enseignement, des outils de communications tels que des forums ou des *chats*. Ces outils sont cependant déconnectés les uns des autres et organisés en silo autour des formations proposées (d'un cours à l'autre les forums sont différents). Enfin, cet outil ne sert qu'à la partie enseignement de l'UTC et aucun lien n'existe entre les ressources hébergées par cette plate-forme et d'autres ressources hébergées ailleurs, comme les travaux de recherche, etc.

Si cette utilisation en silo paraît convenir au domaine de l'enseignement en sacralisant les ressources utilisées au sein des différents cours (UVs) et formations (ingénieur, master, formation continue, etc.), elle limite dans le même temps la possibilité d'interagir d'un cours à l'autre. Ces interactions offriraient pourtant aux utilisateurs la possibilité de découvrir et dresser des ponts thématiques là où la structuration universitaire ne le permet pas. Ces passerelles pourraient servir à réutiliser dans un autre cours les notions abordées dans un premier cours ou à mutualiser les efforts de différents étudiants d'un même cours, même s'ils ne suivent pas le même cursus. Enfin, la mixité des thématiques et des profils (chercheurs, ingénieurs, étudiants, etc.) des utilisateurs du système ouvre de nouvelles possibilités d'innovations.

Un média social a été déployé au sein de Thales R&T afin de favoriser les échanges trans-

verses au sein de l'organisation. Ce dernier prend la forme d'un mur d'activité sur lequel les utilisateurs vont pouvoir échanger des informations, des liens, des images, etc. Ces échanges s'effectuent au sein d'espaces thématiques. Un utilisateur, après s'être abonné aux espaces thématiques de son choix, pourra suivre les derniers évènements de ces espaces sous la forme d'une synthèse affichée dès la page d'accueil. Ce média social est déconnecté des autres applications de l'organisation et en particulier du Wiki. Son accès étant fortement réglementé, il n'est de plus pas possible de suivre en continue ce flux d'informations par le biais d'outils externes comme un agrégateur de flux RSS ou son téléphone portable.

L'usage des médias sociaux ajoute de nouvelles sources d'informations pour les membres des organisations. Cependant, ces sources ne sont pas encore bien intégrées dans les processus de gestion de connaissances. Les informations restent prisonnières du format utilisé par l'outil ayant servi à les créer, formant des silos hermétiques deux à deux. Les informations créées ou échangées au sein de ces applications Web ne sont donc pour l'instant que peu exploitées car difficilement accessibles, même au sein d'un même écosystème numérique.

Nous revenons dans le chapitre suivant sur les usages des organisations vis-à-vis de leur gestion de connaissances et la manière dont les informations échangées au sein des applications Web et des médias sociaux pourraient être prises en compte.

3. Organisations des connaissances

3.1 Connaissances, informations, données

Le concept de connaissance peut se définir de différentes manières selon le domaine considéré. Intéressons-nous dans un premier temps aux définitions usuelles issues de la philosophie. Les philosophes considèrent que cette notion peut se départager en trois grandes familles :

1. la connaissance propositionnelle ;
2. la connaissance objectuelle ;
3. le savoir faire.

Ces trois définitions s'accordent cependant sur l'objet étudié : il n'est pas question de définir un concept abstrait que l'on pourrait posséder comme l'on peut posséder un livre. Il s'agit plutôt de définir l'état d'un sujet éclairé par cette connaissance.

Ainsi le savoir-faire correspond à la capacité qu'a le sujet d'effectuer quelque chose [Ryle, 1978] — « je sais couper du bois », « je sais parler Allemand », etc.

La connaissance objectuelle correspond par contre au fait de connaître un objet particulier, d'en avoir entendu parler — « je connais le Louvre » (j'y suis déjà allé), « je connais H2G2 » (j'ai lu ce livre), etc. Ce type de connaissance se retrouve dans les travaux de Russell [1912] lorsqu'il disserte sur la connaissance directe (« par accointance ») que l'on acquiert par expérience personnelle — les fraises sont rouges, j'en ai déjà vu — et la connaissance rapportée (« par description ») que l'on acquiert par l'entremise de quelqu'un ou quelque chose parce qu'on ne peut pas ou plus en faire l'expérience — le chocolat provient des fèves de cacao, contenues dans une cabosse ; je n'ai jamais vu de cacaoyer.

La connaissance objectuelle se rapproche de la connaissance propositionnelle, issue de la définition de Platon qui disait que la connaissance est « une croyance vraie et justifiée. » Nous savons par exemple que fumer provoque des cancers, aux poumons principalement. Cependant il convient de noter que cette dernière définition est celle qui provoque le plus de débats autour de la notion de justification — Descartes se satisfait d'une croyance certaine, tandis que d'autres réclament une justification non défaite, ou le fruit d'une preuve formelle.

Du point de vue de la gestion de connaissances ou *Knowledge Management* (KM), il s'agit d'une interprétation et d'une appropriation par un être humain d'un ensemble d'informations. Cette définition ne considère donc plus l'état d'un sujet mais se rapproche d'une notion quantifiable ou partageable. Il ne faut néanmoins pas confondre la connaissance avec l'information ou la donnée. Ces trois notions sont clairement dissociées pour le KM [Prax, 2012] :

la donnée Est un élément brut unitaire, factuel, quantitatif, pris en dehors de tout contexte — 10 est une donnée par exemple ;

l'information En croisant différentes données ou en les plaçant dans un contexte particulier, il est possible de leur apporter un sens. On parle alors d'information. Dans le cadre de la météo, la température de 35℃ est une information.

la connaissance Est l'interprétation par l'humain d'un ensemble d'informations. Dans le cadre de la météo, à 35℃ il fait caniculaire. Médicalement, à 35℃ un patient est en hypothermie. Suivant le contexte (météo ou médecine) une personne interprétera l'information (la température de 35℃) de manière différente : c'est de la connaissance.

À ces définitions, il est courant d'ajouter celle de savoir comme étant la somme, pour un individu donné, de toutes ses connaissances et celle de compétence comme étant une connaissance évaluable par un tiers.

Au sein des organisations, la notion de connaissance est perçue comme une variable importante de leur capital. À la différence de Faust qui recherche la connaissance absolue, les organisations se concentrent sur un périmètre restreint souvent appelé cœur de métier. L'étude de l'environnement extérieur de l'organisation, dans le contexte de son cœur de métier prend le nom d'intelligence économique. Les connaissances impliquées dans ces deux notions peuvent s'exprimer de deux manières différentes [Nonaka et Takeuchi, 1995] :

1. les connaissances explicites qui ont pu faire l'objet d'une retranscription dans le monde physique sous la forme d'informations contextualisées et stockées dans une base de connaissances — cahier des charges décrivant une invention, compte-rendu d'interview d'un expert, code source d'un logiciel, etc.
2. les connaissances implicites portées par les membres de l'organisation : il s'agit du savoir progressivement acquis par les individus au gré de leurs expériences, observations personnelles. Il s'agit d'objets strictement cognitifs non encore matérialisés.

L'enjeu pour les organisations réside donc dans le passage de la connaissance tacite en connaissance explicite. Ces dernières garantissent plus sûrement une compréhension homogène de la situation, une transmission plus fiable des connaissances et, surtout, la sauvegarde de ce patrimoine.

Le passage de connaissances implicites à des connaissances explicites est matérialisé, pour le KM par un processus de documentarisation. Il s'agit de l'élaboration, par la ou les personnes possédant la connaissance, d'un document transcrivant cette connaissance sur un support donné. Nous définissons plus en détail cette notion de document dans la section suivante.

3.2 Documents et ressources numériques

Afin de maîtriser au mieux leurs connaissances, les organisations ont souvent recours à des systèmes de gestion de connaissances. Ces systèmes doivent permettre aux membres d'une organisation de capitaliser leurs documents au sein d'une base de connaissances. Une première définition de document peut se trouver dans les travaux effectués dans le cadre de la société des nations en 1937 (*in* Buckland [1998]) : « Document : Toute base de connaissances, fixée matériellement, susceptible d'être utilisée pour consultation, étude ou preuve. Exemples : manuscrits, imprimés, représentations graphiques ou figurées, objets de collection, etc. » Briet [1951] généralise un peu plus cette définition en statuant : « Un document est une preuve à l'appui d'un fait. » Buckland [1998] explicite cette définition en proposant quatre conditions qu'un objet doit remplir pour obtenir le statut de document.

1. le document doit être un objet physique (physicalité) ;
2. le document doit véhiculer une intentionnalité : il y a eu volonté de le considérer comme document (documentarisation) ;
3. Le document a été produit, construit, généré (matérialisation) ;
4. Le document doit être perçu comme tel, c'est-à-dire laisser directement transparaître sa physicalité, son intentionnalité et sa différenciation du sujet qu'il documentarise : un même objet, suivant le contexte de son utilisation ou de sa réification pourra, ou pas, être considéré comme un document (contextualisation).

La notion de contexte, vis-à-vis des documents désigne tout à la fois les conditions dans lesquelles le document a été produit — à quel besoin répond-il, s'agit-il d'une réponse ou d'une question, etc. —, sa temporalité — s'agit-il d'un document ancien ou récent, en combien de temps a-t-il été produit, etc. —, son origine — quels en sont les auteurs, les commanditaires, quelles en étaient leurs affiliations, etc. — et son environnement — dans quel type de société s'inscrit-il, etc.

Briet [1951] nous donne quelques exemples permettant de mieux appréhender les conditions encadrant la notion de document. Le plus habituel est sans doute celui de l'étoile : une étoile dans le ciel n'est évidemment pas un document, c'est une étoile. Une photographie de cette même étoile est par contre un document. Nous remarquerons qu'effectivement la pho-

tographie est un objet physique, un morceau de papier glacé que l'on peut manipuler (condition 1). Il y a bien eu volonté du photographe de fixer l'image de l'étoile sur une photographie (condition 2) et cette volonté s'est traduite par un procédé technique permettant effectivement de conserver une trace de cette étoile (condition 3). Si la photographie de l'étoile est authentifiée — oui, il s'agit bien d'une représentation de « cette » étoile — elle va pouvoir devenir un objet d'étude, en remplacement de l'étoile réelle par exemple. Cette photographie est alors perçue comme une représentation indépendante matérielle de l'objet considéré (condition 4). Nous avons bien dans ce cas une preuve à l'appui d'un fait ; si l'étoile dans le ciel est bien un « fait », la photographie apparaît comme une « preuve » de ce « fait » : cette étoile est bien dans le ciel.

Cette définition conduit à considérer le cas où une personne, étudiant cette photographie, publie un article dessus. Briet [1951] parle alors de document de second niveau : la photographie ne perd pas son statut de document, mais peut tout à fait être considérée comme l'objet d'étude d'un second document — à noter que rien n'empêche ce second document de porter tant sur le fond (étudier l'étoile au travers de la photographie) que sur la forme (le morceau de papier, son grain, l'exposition de la photographie).

Buckland [1998] s'interroge sur la notion de document numérique et remarque que la principale difficulté dans l'appréciation d'un document numérique est de pouvoir le considérer en tant que document (condition 4).

Certaines productions numériques sont triviales à documentariser. Prenons l'exemple d'un article de blog. Cet article étant sauvegardé au sein d'une mémoire informatique, il s'agit bien d'un objet physique (condition 1) : les bits concernant cet articles peuvent être identifiés et leur altération amène une altération de l'article. Les articles de blogs portent généralement sur un sujet concret ou abstrait (condition 2) de différentes manières (condition 3) : textuelle, graphique (blogs BDs, photologs), audio (podcasts) ou multimédia. Leur publication au sein d'un blog les place dans un contexte qui permet leur dissociation deux à deux et vis-à-vis du sujet original de l'article (URL différentes, mise en page). Il est intéressant de constater qu'un commentaire de blog, répondant à un article, peut alors être considéré comme un document de second niveau vis-à-vis de l'article original.

Pédauque [2003] se penche sur le cas des documents numériques et en tire une définition en trois volets : 1) le document numérique comme forme, 2) le document numérique comme signe et 3) le document numérique comme medium. Il s'agit de trois points de vue différents sur un objet documentaire, permettant d'en appréhender différents éléments.

Le document comme forme Cette définition s'établit autour d'une équation simple posant le document comme la somme d'une structuration et de données (document numérique = structure + données). Cette définition veut insister sur le caractère désormais enfoui

au sein de bases de données de la substance des documents, *a priori* inutilisable directement par les humains et l'obligation d'avoir recours à des dispositifs techniques pour pouvoir lire (au sens large) ces documents.

Le document comme signe Cette définition pose le document en réponse à la somme d'un texte informé et de connaissances. La notion de texte informé se réfère à un contenu pouvant être traité, opérationnalisé afin d'en extraire au besoin des unités d'informations. La notion de connaissance rappelle le besoin d'appropriation personnelle de l'objet considéré pour en faire un document, la connaissance étant le résultat d'une interprétation personnelle des informations issues du texte informé.

Le document comme medium Cette définition s'accroche à la notion de medium, qu'il faut comprendre au sens large, c'est-à-dire celui de phénomène social, d'élément tangible outil de communication entre personnes. C'est cette notion de medium qui permet à Pédauque [2003] de dire qu'« un document numérique est la trace de relations sociales reconstruite par les dispositifs informatiques. »

Lainé-Cruzel [2004] complète les définitions de Pédauque [2003] en apportant la notion de ressource. Selon elle, les documents numériques recouvrent deux formes d'objets numériques distincts mais compatibles : la forme « document » et la forme « ressource ». La forme document se comprend dans une logique d'archivage et va désigner des objets stables dans le temps qui obtiennent, par ce biais, une valeur de preuve. Les ressources se comprennent plutôt dans une logique d'usage et vont désigner des objets appelés à évoluer qui auront plutôt une valeur de simple renseignement.

3.3 Étiquetage documentaire

La signification d'un document numérique est portée à la fois par le fond, la forme et le contexte de son utilisation : même s'ils traitent d'un même objet, deux documents n'ont pas la même intentionnalité. Si nous prenons l'exemple de l'antilope de Briet [1951], une photographie d'une antilope dans la savane et une antilope hébergée dans un zoo sont deux documents portant sur les antilopes. Les informations portées par ces deux documents sur l'objet considéré sont cependant différentes :

— la photographie permet de définir le type d'environnement dans lequel évolue l'antilope (la savane), la couleur de sa robe, etc. ;
— l'animal dans le zoo permet de définir la taille générale des antilopes, leur régime alimentaire et autres informations vétérinaires, etc.

La gestion des documents, dans l'optique d'un partage de connaissances, doit pouvoir préserver l'ensemble des informations et leur contexte. Il s'agit de trouver le moyen de stocker

tant le sens des documents et leurs contenus — les informations qu'ils véhiculent —, que les indications sur le document lui-même. Ces indications peuvent être sauvegardées sous la forme d'étiquettes ajoutées aux documents.

Cet étiquetage permet non seulement de préciser le sens que véhicule le document, sans avoir besoin de l'interpréter ou d'en lire le contenu, mais sert également de système de classement, d'indexation des documents. Le choix d'une technique d'étiquetage des documents influe sur l'architecture d'un système de gestion de ces mêmes documents et plus particulièrement sur la manière d'accéder aux connaissances portées par ces documents.

Avec le développement des médias sociaux, beaucoup de sites communautaires, comme Flickr ou del.icio.us ont popularisé ce principe d'étiquetage, ou *tagging*. Au sein d'une interface facile d'utilisation l'utilisateur est invité à donner des mots-clés permettant de décrire le document qu'il publie. Ces *tags* peuvent être choisis aléatoirement par les auteurs ou issus d'une taxonomie préalablement définie. Au fur et à mesure de l'ajout de nouveaux documents sur le Web et de leur caractérisation à l'aide de *tags*, un utilisateur va se forger sa folksonomie. Ce terme est un mot-valise combinant l'anglais *folks* (les gens, le peuple) et le mot taxonomie. Les folksonomies sont cependant loin d'être aussi structurées que les taxonomies. Les folksonomies sont en effet comparables à des ensembles de mots non liés, tandis que les taxonomies apportent une structuration sémantique du vocabulaire employé.

Un utilisateur peut être caractérisé par sa folksonomie, qui apporte des informations sur ses goûts, ses compétences, ses habitudes. Certains médias sociaux permettent de naviguer dans les folksonomies de leurs membres, en transformant les *tags* en liens hypertextes conduisant à des listes de ressources indexées à l'aide de ces *tags*. Cette utilisation permet l'identification de communautés d'intérêt, publiant des ressources sur la même thématique.

Afin de mettre en valeur les *tags*, lorsqu'ils sont insérés dans le corps de la ressource qu'ils caractérisent, il est désormais fréquent de leur accoler un croisillon # (ou *hash* en anglais [1]). On parle alors de *hashtag*. Cette utilisation, popularisée sur Twitter, trouve ses racines sur IRC, célèbre logiciel de *chat*, qui utilisait ce moyen technique pour nommer les différents salons communautaires au sein desquels les utilisateurs pouvaient discuter. D'ailleurs, le premier *tweet* contenant un *hashtag* fait bien référence à la notion d'identification communautaire, avant celle d'étiquetage de contenu :

> « *how do you feel about using # (pound) for groups. As in #barcamp [msg] ?* » — Chris Messina [2]

Remarquons au passage que l'utilisation de mots-clés pour décrire des ressources porte une intention d'archivage, de classement et semble changer le point de vue que l'on peut

1. À ne pas confondre, en typographie, avec le symbole dièse ♯
2. « Que pensez-vous de l'utilisation de # (croisillon) pour les groupes. Comme dans #barcamp [bla] ? » (traduction personnelle). Visible le 10 novembre 2013 à l'adresse https://twitter.com/chrismessina/status/223115412

avoir de cette ressource en lui accordant alors le statut de document, selon la définition de Lainé-Cruzel [2004]. Néanmoins, disponible exclusivement au travers d'une application Web, le document numérique enrichi de *tags* reste appelé à pouvoir évoluer, être échangé, etc. Et donc à conserver de fait sa nature de ressource, toujours selon Lainé-Cruzel [2004]. Nous touchons ici la limite de cette définition de ressource et document.

3.4 Le Web sémantique et les ontologies

Avec une approche décentralisée, mais prônant et permettant l'interopérabilité, le Web sémantique offre une bonne solution pour caractériser des documents [Breslin et Decker, 2007, Halpin, 2008].

Le Web sémantique est le nom donné à la dynamique industrielle et de recherche visant à rendre le Web auto-comprenant. Tim Berners-Lee, unanimement reconnu comme l'inventeur du Web, le définit comme quelque chose qui permet de « donner aux informations une signification mieux définie et d'autoriser une meilleure coopération de l'homme et de la machine [3] » [Mikroyannidis, 2007].

Le Web sémantique propose de qualifier des données afin de les lier à différents contextes pour lesquels elles seront sources d'informations. Au sein d'une base de connaissances, cette qualification s'appuie sur un référentiel formel qui garantit l'absence d'ambiguïté dans le vocabulaire utilisé pour la qualification. Ce référentiel sera très souvent une ontologie. En informatique et plus généralement en science de l'information, les ontologies sont des ensembles structurés de concepts permettant de décrire un champ d'application. Il s'agit de « la spécification d'une vue abstraite et simplifiée du monde que l'on veut représenter dans un but déterminé [4] » [Gruber, 1993].

Le champ d'application d'une ontologie peut être extrêmement réduit ou général. Guarino [1998] propose une classification des ontologies visible à la figure 3.1 et dans les définitions suivantes.

Dans cette figure les ontologies de haut niveau (*top-level ontologies*) ou supérieures (*upper-level ontologies*) tendent à représenter des concepts très formels pouvant être partagés entre différents domaines ou applications. L'exemple le plus courant est d'y définir des notions telles que le temps ou l'espace.

Les ontologies de domaine permettent de décrire plus précisément un domaine particulier, comme la médecine, les pizzas ou les types de documents (comme BIBO [D'Arcus et Giasson,

3. Traduction personnelle de « […] *giving information a well-defined meaning, better enabling computers and people to work in cooperation* » p. 113

4. Traduction personnelle de « *A conceptualization is an abstract, simplified view of the world that we wish to represent for some purpose. […] An ontology is an explicit specification of a conceptualization.* » p. 1

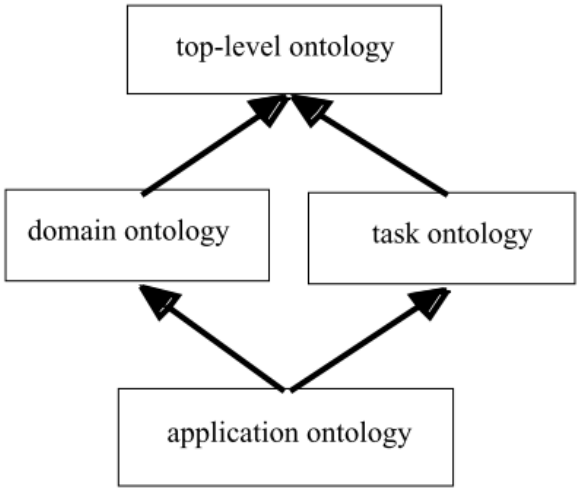


Figure 3.1 – Différents types d'ontologies (par Guarino [1998])

2009]). Cette description du domaine ne s'attache pas à décrire les activités de ce domaine, comme soigner ou manger. Ces activités sont décrites au sein d'ontologies de tâches (comme l'ontologie d'ordonnancement de Rajpathak *et al.* [2001]). Ces deux types d'ontologies se définissent indépendamment, mais souvent en spécialisant des concepts plus généraux provenant d'ontologies de haut niveau.

Les ontologies d'applications se trouvent à la réunion d'une ontologie de tâches apportant un modèle d'activités et d'une ontologie de domaine fournissant le cas d'application. La construction d'une telle ontologie s'effectue classiquement en spécialisant des concepts provenant d'une ontologie de domaine et des concepts provenant d'une ontologie de tâches.

Les ontologies d'application vont permettre par exemple de modéliser une application Web en apportant des précisions sur les rôles que vont tenir les différents utilisateurs (décrits dans une ontologie de domaine) via les différentes actions possibles sur l'application (décrites dans une ontologie de tâches). Ou encore décrire des mécanismes de classement documentaire, la notion de document étant abordée dans une ontologie de domaine et celle de classement dans une ontologie de tâche.

Les ontologies sont principalement rédigées à l'aide de standards coordonnés par le *World Wide Web Consortium* (W3C), une organisation à but non lucratif fondée par Tim Berners-Lee afin de promouvoir la compatibilité des technologies assurant le fonctionnement du Web. Le W3C fonctionne à la manière d'un consortium d'entreprises, regroupant en 2013 trois cent quatre-vingt-sept entreprises différentes, dont Microsoft, Apple, Google, IBM, mais également nombre de PME et d'universités.

Parmi les standards mis en place par le W3C permettant la rédaction d'ontologies, citons RDF (et sa variante RDF-schema) et OWL.

RDF (pour *Resource Description Framework*) est un modèle de graphe créé pour permettre la

représentation formelle des ressources disponibles sur le Web. Ceci afin de permettre leur manipulation automatique. RDF est construit sur la notion de triplet qui désigne le lien pratiqué à l'aide d'un « prédicat » entre un « sujet » et son « objet ». Le sujet représente systématiquement la ressource à qualifier. Le prédicat est une relation bien précise pouvant s'appliquer au sujet pour un type d'objet précis. L'objet, enfin, qualifie donc le sujet au travers du prédicat et peut être soit une autre ressource, soit un scalaire, en fonction de la définition du prédicat lui-même.

RDFS (pour *RDF schema*) complète la recommandation RDF en apportant les définitions de base permettant la construction d'ontologies destinées à structurer des ressources décrites en RDF. RDFS apporte ainsi par exemple la notion de `Class` permettant de décrire des objets ou concepts à instancier, ou le prédicat `subClassOf` permettant d'ordonner les différentes `Class` entre elles.

OWL (pour *Web Ontology Language*) est à voir comme une extension des deux précédentes recommandations permettant d'apporter plus d'expressivité lors de la conception d'un modèle ontologique. OWL apporte ainsi des manières d'exprimer les dépendances ou la cardinalité entre deux ressources de manière plus précise. *Web Ontology Language* (OWL) propose deux mécanismes permettant de carcatériser une ressource particulière :

Les *data properties* qui permettent de spécifier les propriétés scalaires de la ressource. Pour une ressource de type voiture, il peut s'agir par exemple d'exprimer le fait que : couleur = rouge ou nombre de sièges = 5.

Les *object properties* qui permettent de spécifier des relations entre des concepts de l'ontologie, pourvu que les types de ces concepts correspondent au domaine et au *range* de la relation. Pour une ressource A de type voiture et une ressource B de type humain, il est possible de spécifier la propriété « conduire » dont le domaine doit être un humain et le *range* une voiture. Il est alors possible d'exprimer le fait que B conduit A.

Le grand intérêt de ces standards dans la conception d'ontologie est qu'ils forcent le créateur de l'ontologie à respecter un certain niveau de formalisme. Ce niveau de formalisme est requis pour espérer pouvoir obtenir des résultats intéressants lors du traitement d'une ontologie à l'aide d'un raisonneur. L'utilité des raisonneurs est variable et va de la vérification de la conformité d'une base de connaissances avec un modèle ontologique à la détection d'incohérences dans ce même modèle ou encore au peuplement automatique d'une base de connaissances. L'utilisation d'une ontologie formelle rend ainsi beaucoup plus aisée la réutilisation des données stockées au sein d'une base de connaissances et leur traitement automatique pour en tirer de nouveaux faits qui pourront être proposés aux utilisateurs, par exemple dans le cadre d'un processus de décision.

TABLE 3.1 – Classification de quelques ontologies du Web sémantique

Ontologies	**Dialogue**	**Discours**	**Publication**	**Coopération**	***Crowdsourcing***	**Gestion documentaire**	**Gestion des utilisateurs**
sociale							
FoaF			✓			✓ (`foaf:Document`)	✓ (`foaf:Person`, `foaf:Online-Account`)
OCSO	✓ (`ocso:Conversation`, `ocso:Social-Object`)	✓ (`ocso:Conversation`, `ocso:Social-Object`)	✓ (`ocso:Social-Object`)		✓ (`ocso:Social-Object`)		✓ (`ocso:User`)
documentaire							
SIOC		✓ (`sioc:Post`)	✓ (`sioc:Item`)		✓ (`sioc:Post`, `sioc:Forum`, `sioc:Item`)		✓ (`sioc:User-Account`)
BIBO						✓ (`bibo:*`)	
VIVO		✓ (`vivo:Blog-Posting`)	✓ (`vivo:Information-Resource`)			✓ (`vivo:InformationResource`)	✓ (`vivo:Position`, `vivo:Role`, `vivo:Issued-Credential`)

Les standards du W3C forment le socle commun à de nombreuses ontologies du Web sémantique s'étant donné pour mission la modélisation des différents objets circulant sur le Web. Il s'agit là d'un intérêt supplémentaire des ontologies développées dans le cadre du Web sémantique. La recherche de solutions de modélisation standardisées permet une plus grande interopérabilité des systèmes utilisant ces standards. Dans un monde toujours plus connecté, l'utilisation de standards permet la communication de différents services informatiques, éventuellement indépendamment de leurs utilisateurs. Cette communication possible entre services est une des briques essentielles du Web de données.

Nous avons ainsi relevé quelques unes de ces ontologies que nous présentons dans le tableau 3.1 p. 40. Ce tableau reprend les différents types d'applications définis précédemment. Nous avons séparé les ontologies étudiées en deux grandes familles, selon qu'elles se présentent principalement pour modéliser un réseau social ou un système documentaire. Les sections suivantes viennent présenter plus en détails ces différentes ontologies. La modélisation d'un écosystème numérique requiert la prise en compte des utilisateurs de cet écosystème. Nous avons donc également ajouté une colonne traitant de cette problématique.

3.4.1 L'ontologie FoaF

Le projet *Friend of a Friend* (FoaF)[5] lancé en 2000, a pour but de décrire des personnes, leurs activités et les relations qui peuvent exister entre ces personnes, ou entre ces personnes et des documents. FoaF apporte ainsi des classes et relations de base pour décrire :

- des personnes : classe `foaf:Person`, relations `foaf:firstName`, `foaf:lastName`, `foaf:gender`, `foaf:phone`, etc. ;
- leurs comptes d'utilisateur : classe `foaf:OnlineAccount` et ses sous-classes (`foaf:OnlineChatAccount`, `foaf:OnlineEcommerceAccount`, etc.), relations `foaf:accountName`, etc. ;
- un réseau social : classe `foaf:Group`, relation `foaf:knows` ;
- des organisations : class `foaf:Organization`, relations `foaf:name`, `foaf:plan`, `foaf:logo`, etc. ;
- des documents : classe `foaf:Document` et `foaf:Image`, relations `foaf:title`, `foaf:topic`, `foaf:publications`.

En ce qui concerne l'ajout de métadonnées aux documents décrits à l'aide de FoaF, l'usage du standard ISO 15836 *Dublin Core* est préconisé. Ce dernier est un schéma normalisé de métadonnées, élaboré à partir de 1995 par un groupe de travail commandité par l'*Online Computer Library Center* (OCLC) et le *National Center for Supercomputing Applications* (NCSA). La

5. Dont la spécification est disponible à l'adresse http://xmlns.com/foaf/spec/

première réunion rassembla une cinquantaine de chercheurs de différents pays et eut lieu à Dublin aux USA, d'où le nom habituel de ce standard.

FoaF est l'une des premières ontologies du Web sémantique et de nombreuses autres ontologies l'utilisent en tout ou partie. De nombreuses extensions ont été écrites afin d'enrichir l'ontologie originale comme *relation FoaF*[6] qui spécialise la relation `foaf:knows` en ajoutant de nombreuses relations sociales plus précises (`foafrel:colleagueOf`, `foafrel:neighborOf`, etc.).

Dans le cadre du scénario présenté dans la section de l'introduction, nous avons considéré Alice et Baptiste comme membres de l'organisation Argos. La description d'Alice et Baptiste à l'aide de FoaF peut se faire selon le *listing* ci-après :

```
:alice a foaf:Person;
    foaf:lastName "Nonyme";
    foaf:firstName "Alice";
    foaf:gender "female";
    foaf:mbox <anonyme@orga.fr>.

:baptiste a foaf:Person;
    foaf:lastName "Atman";
    foaf:firstName "Baptiste";
    foaf:gender "male";
    foaf:mbox <batman@orga.fr>.
```

3.4.2 L'ontologie OCSO

Des travaux récents [Marie et Gandon, 2011] sur la modélisation des réseaux sociaux et plus particulièrement sur la manière qu'ils ont de rassembler des gens autour d'un objet d'intérêt ont conduit à la proposition de l'ontologie *Object Centered Sociality Ontology* (OCSO)[7]. La modélisation du réseau social s'effectue à l'aide de différentes instances de `ocso:User` représentant les utilisateurs pouvant se retrouver autour d'un `ocso:SocialObject` (qu'ils peuvent classiquement « aimer », recommander, etc.) formant ainsi un `ocso:OCSN` aux goûts partagés.

Le *listing* ci-après présente un exemple d'utilisation de cette ontologie. Cet exemple montre de quelle manière Alice et Baptiste, de notre scénario, sont parties prenantes d'une même communauté centrée sur un objet d'intérêt représenté par un article scientifique :

```
:user_alice a ocso:User.
```

6. Dont la spécification est disponible à l'adresse http://vocab.org/relationship/.html
7. Dont les spécifications sont disponibles à l'adressehttp://ns.inria.fr/ocso/

```
:user_baptiste a ocso:User.

:scientific_article a ocso:SocialObject;
    ocso:hasTitle "Médias sociaux et organisations";
    ocso:createdBy :user_baptiste.

:scientific_article_ocsn a ocso:OCSN;
    ocso:isCenteredOn :scientific_article;
    ocso:hasLikeCount 1.

:user_alice ocso:hasLiked :scientific_article;
    ocso:isMemberOf :scientific_article_ocsn.

:user_baptiste ocso:isMemberOf :scientific_article_ocsn.
```

Une approche similaire est la proposition du schéma de métadonnées sociales *Open Graph Protocol* (OGP) de Facebook [8]. Cette dernière propose de modéliser des activités sociales à l'aide de mécanismes comme RDFa pour annoter sémantiquement des documents du Web en utilisant un vocable proche de Dublin Core.

La notion d'objet d'intérêt rejoint le point de vue du document en tant que medium de Pédauque [2003] en la généralisant. La définition d'un `ocso:SocialObject` se retrouvant au centre de l'attention de différentes personne peut être comparé à un processus de documentarisation.

3.4.3 L'ontologie SIOC

L'ontologie *Semantically-Interlinked Online Communities* (SIOC) [9] et ses extensions [10] proposent une modélisation de la plupart des applications Web et la manière dont les individus, représentés par leurs comptes utilisateurs, vont pouvoir intervenir au sein de ces différentes applications. SIOC tisse des liens importants avec FoaF [Breslin *et al.*, 2009] et l'étend même lorsque nécessaire : la classe `sioc:UserAccount` est une spécialisation de `foaf:OnlineAccount`, ce qui permet de lier un compte d'utilisateur SIOC à une personne décrite à l'aide de FoaF.

Ainsi, sans directement permettre la décription un réseau social, SIOC l'autorise tout de même à l'aide des ponts ouverts vers FoaF. De même, la classe `sioc:Usergroup` (qui permet de décrire des groupes d'utilisateurs au sein d'une application Web), sans présager de la manière

8. Spécifications disponibles à l'adresse http://ogp.me/ (accessible le 10 novembre 2013)
9. Dont la spécification est disponible à l'adresse http://rdfs.org/sioc/spec/
10. Pour ne pas alourdir le cœur de l'ontologie, la plupart des spécialisations des concepts de base sont définis au sein de modules dont les spécifications sont disponible à l'adresse http://sioc-project.org/ontology#sec-modules

dont ils sont créés, laisse la liberté de l'utiliser pour modéliser des groupes sociaux spontanés. Il est d'ailleurs intéressant de noter que c'est exactement ce que propose l'ontologie OCSO qui définit sa classe `ocso:OCSN` comme étant une sous-classe de `sioc:Usergroup`.

Nous avons statué dans notre scénario qu'Argos avait déployé, pour faciliter la collaboration de ses membres, différentes applications Web. La mise en place de ce type d'applications impose la création de comptes d'utilisateur pour les différents membres de l'organisation. Ces comptes d'utilisateurs vont être répartis au sein de différents groupes d'utilisateurs, pendants numériques des différents services de l'organisation. Le *listing* ci-après exprime l'appartenance de l'utilisateur `:alice_account` (le compte d'utilisateur d'Alice) à un groupe d'utilisateurs nommé `:retdpeople` (le groupe numérique lié à l'unité de R&D d'Argos). Le lien entre le compte d'utilisateur `:alice_account` et `:alice` (définie précédemment dans la section 3.4.1) est possible parce que la classe `sioc:UserAccount` spécialise la classe `foaf:OnlineAccount`. Cette dernière peut elle-même être liée à une `foaf:Person` à l'aide de l'*object properties* `foaf:account`.

```
:alice_account a sioc:UserAccount;
    sioc:id "alice".

:alice foaf:account :alice_account.

:retdpeople a sioc:UserGroup;
    rdfs:label "Groupe des membres du département de R&D".

:retdpeople sioc:has_member :alice_account.
:alice_account sioc:member_of :retdpeople.
```

3.4.4 L'ontologie BIBO

Le projet *The Bibliographic Ontology* (BIBO) [11] a pour but de décrire des références bibliographiques pour le Web sémantique. Il permet ainsi de décrire des livres, des articles, mais aussi des conférences ou des collections de documents. Il s'appuie sur FoaF pour décrire les auteurs d'un document. La classe principale `bibo:Document` est en fait un alias de la classe `foaf:Document`.

L'ontologie BIBO tend à suivre les mêmes règles de nommage que celles employées avec Bibtex. La conversion d'une bibliographie accessible au format Bibtex selon le vocabulaire sémantique de BIBO est ainsi facilitée.

11. Dont la spécification est disponible à l'adresse http://bibliontology.com/specification

La description d'un brevet détenu par Baptiste, dans notre scénario, à l'aide de BIBO se fait, à titre d'exemple, à l'aide du *listing* ci-après.

```
:baptiste_patent a bibo:Patent;
    bibo:shortTitle "Médias sociaux et organisations";
    dc:title "Dispositif permettant la remontée automatique d'alertes à
    l'aide d'un bus de messages s'appuyant sur un média social";
    bibo:authorList ( :baptiste ).
```

3.4.5 L'ontologie VIVO

L'ontologie VIVO [Conlon *et al.*, 2003] est proposée par le projet VIVO précédemment présenté dans la section 2.3.8. Il s'agit principalement d'un *mashup* d'ontologies construit autour, pour les principales, de FoaF, BIBO, *Simple Knowledge Organization System* (SKOS), *Eagle-I ontology* (une ontologie de domaine décrivant des documents biomédicaux) et *Event ontology* (une ontologie permettant de décrire des évènements). Le projet VIVO complète néanmoins ce panorama en y ajoutant des classes et relations qui nous intéressent particulièrement :

— la modélisation d'un processus d'authentification, à l'aide de la classe `vivo:IssuedCredential`, peut être utilisée dans la gestion d'un accès commun aux différentes applications ;
— la généralisation de la classe `bibo:Document` à l'aide de la classe `vivo:InformationResource` permet de décrire de nombreuses sortes de documents différents.

À titre d'exemple, nous pouvons décrire plus précisément le brevet de Baptiste à l'aide du *listing* ci-après.

```
:soweb_project a vivo:Project;
    vivo:informationProduct :baptiste_patent;
    vivo:hasFundingVehicule :soweb_project_business_plan.

:argos a vivo:Company.
:argosrd a vivo:ResearchOrganization;
    vivo:subOrganizationWithin :argos.

:soweb_project_business_plan a vivo:Grant;
    vivo:fundingVehicleFor :soweb_project;
    vivo:administeredBy :argosrd.
```

3.4.6 Les ontologies SKOS et MOAT

L'utilisation d'une ontologie de domaine semble être un excellent moyen pour indexer des documents. Cependant, il peut arriver que la définition d'une telle ontologie de domaine ne soit pas possible et que seuls une taxonomie ou un ensemble de mots-clés soient disponibles. Dans le cadre d'une base de connaissances modélisée à l'aide d'une ontologie d'application, il est nécessaire de dresser des ponts entre des instances ontologiques et des termes issus de la taxonomie ou d'une folksonomie. À ce sujet, nous pouvons noter les travaux de Passant et Laublet [2008a] qui dressent un état de l'art des différentes ontologies existantes permettant d'encadrer et d'enrichir des folksonomies. Il en existe différentes méthodes. Notons tout d'abord la proposition de Mika [2005] qui étudie des façons d'extraire les termes les plus populaires de folksonomies d'une plate-forme Web afin de mettre en place une ontologie légère (*lightweight ontology*). Notons également la proposition de Passant et Laublet [2008b] qui réfléchissent à la mise en place de liens directs entre les *tags* d'une folksonomie et les concepts d'une ontologie formelle.

Ce dernier article présente l'ontologie *Meaning of a Tag* (MOAT), construite à partir d'une autre ontologie : SKOS. L'ontologie SKOS [12] proposée par le W3C est une ontologie de haut niveau permettant de définir des vocabulaires sémantiques. À la différence de RDF, RDFS ou OWL, SKOS ne fait pas de distinction entre classes et instances de classes. C'est-à-dire, là où des langages plus formels permettent de représenter des faits concernant différents objets et états d'un monde, SKOS ne permet que de modéliser ce monde en liant de manière générique ses différentes composantes. Ainsi, SKOS se prête davantage à l'organisation de taxonomies non formelles. À ce titre, SKOS propose un jeu de classes et propriétés génériques permettant d'enrichir les relations entre termes d'une taxonomie. Il est par exemple possible de retrouver des relations de spécialisation entre termes avec les relations `skos:broader` ou `skos:narrower`.

À titre d'exemple, nous pouvons décrire quelques termes d'une taxonomie à l'aide du *listing* ci-après.

```
1  :web a skos:Concept;
2       skos:prefLabel "Web";
3       skos:altLabel "Internet";
4       skos:broader :web2;
5       skos:broader :dataweb;
6       skos:broader :semanticweb.
7
8  :web2 a skos:Concept;
```

12. Dont la spécification est disponible à l'adresse http://www.w3.org/TR/skos-reference/

```
 9      skos:prefLabel "Web 2.0";
10      skos:narrower :web.
11
12  :dataweb a skos:Concept;
13      skos:prefLabel "Web of data"@en;
14      skos:prefLabel "Web de données"@fr;
15      skos:narrower :web;
16      skos:related :opendata;
17      skos:related :semanticweb.
18
19  :opendata a skos:Concept;
20      skos:prefLabel "Open Data"@en;
21      skos:prefLabel "Données ouvertes"@fr;
22      skos:related :dataweb.
23
24  :semanticweb a skos:Concept;
25      skos:prefLabel "Semantic web"@en;
26      skos:prefLabel "Web sémantique"@fr;
27      skos:narrower :web;
28      skos:related :dataweb.
```

3.5 Synthèse

Nous avons abordé dans le chapitre précédent l'essor des applications Web et des médias sociaux au sein des organisations. Nous avons pronostiqué que l'usage de ces applications est source de nouvelles connaissances pour les organisations. Dans ce chapitre, nous avons donc cherché à positionner, dans un premier temps, cette notion de connaissance vis-à-vis des productions numériques créées à l'aide d'applications Web ou s'échangeant sur les médias sociaux. Cette réflexion nous a mené jusqu'aux définitions de document, document numérique et ressource dont nous avons montré les limites.

Nous avons ensuite identifié dans les mécanismes d'étiquetage documentaire une solution de gestion de connaissances, lorsque cette dernière est explicitée sous la forme d'un document numérique. Nous avons alors présenté différentes techniques d'étiquetage documentaire, de préférence lié à l'environnement du Web. C'est ainsi que les travaux effectués dans le domaine du Web sémantique nous ont semblé pertinents pour servir de cadre à nos propres travaux, en vue de modéliser un environnement de gestion de connaissances prenant en compte les documents habituels et les productions des applications Web.

Dans le chapitre précédent, nous avions rappelé combien les écosystèmes numériques,

composés de différentes applications Web comme des médias sociaux, apportent une dynamique favorable au partage rapide d'informations ou à l'auto-organisation des utilisateurs. Nous avions alors postulé qu'ils semblaient à même de servir les écosystèmes de connaissances. Nous ne nous étions alors intéressé qu'à la problématique de dynamique des échanges et de libre organisation des utilisateurs. À la lumière des définitions apportées dans ce chapitre, nous allons désormais observer les propositions des écosystèmes numériques concernant la gestion de connaissances.

Le tableau 3.2 ci-après reprend le tableau 2.1 mais en s'intéressant cette fois-ci aux fonctionnalités mises en place pour permettre l'étiquetage des documents partagés au sein de ces écosystèmes.

Table 3.2 – Classification de quelques applications Web

Plates-formes	**Dialogue**	**Discours**	**Publication**	**Coopération**	***Crowdsourcing***	**Gestion documentaire**
Yammer	/	Folksonomie	Folksonomie Hashtag	/	/	Folksonomie /
Google Knowledge Plaza	/	Folksonomie	Folksonomie	Folksonomie	/	Folksonomie
Bluekiwi	/	Folksonomie	Hashtag	/	/	Folksonomie
VIVO	Ontologie	Ontologie		Ontologie	Ontologie	Ontologie
E-MEMORAe 2.0						Ontologie

Nous utilisons dans ce tableau les trois valeurs suivantes, issues des techniques d'étiquetage présentées dans ce chapitre, afin de typer les fonctionnalités d'indexation :

Ontologie les étiquettes utilisées pour décrire les documents sont des instances de classes ontologiques provenant d'une base de connaissances. L'utilisateur n'est pas à l'initiative de l'introduction d'un mot-clé dans le système. Il est simplement libre d'utiliser ceux de son choix pour décrire un document.

Folksonomie l'utilisateur est libre d'utiliser les étiquettes qu'il souhaite pour décrire un document. Ces étiquettes sont sauvegardées sous la forme d'une taxonomie lui permettant de les classer entre elles et de les réutiliser facilement sur d'autres documents.

Hashtag l'utilisateur est libre de donner les étiquettes qu'il souhaite pour décrire un document. Aucun contrôle ni consolidation n'est possible sur ces étiquettes.

Les cases vides indiquent que le type d'application Web correspondant n'est pas implémenté dans l'écosystème. Les cases comportant une barre oblique (/) indiquent que cet écosystème implémente ce type d'application Web, mais qu'aucune fonctionnalité d'étiquetage des documents n'est disponible.

Le classement des fonctionnalités d'étiquetage permet de constater un certain nombre de points problématiques vis-à-vis d'une bonne gestion des informations partagées au sein des écosystèmes.

— Mis à part au sein de Knowledge Plaza, E-MEMORAe 2.0 et VIVO, aucun référentiel unique n'est mis en place entre chaque type d'application. Les ressources produites par une application ne sont pas accessibles au sein d'une autre à l'aide des étiquettes.
— Aucun mécanisme d'indexation de contenu n'est implémenté pour certains types d'application (les cases « / »).
— L'approche de Knowledge Plaza est appréciable, bien que l'usage de folksonomies ne permette pas l'usage de raisonneurs sur la base de connaissances de l'écosystème.
— L'approche de VIVO, fondée autour d'une base de connaissances ontologique permet l'utilisation de raisonneurs. Le projet ne propose cependant pas le support de médias sociaux.
— L'environnement E-MEMORAe 2.0 semble convenir tout à fait en proposant déjà un référentiel ontologique partagé entre toutes les applications de la plate-forme.

Hormis l'environnement issu du projet de recherche E-MEMORAe 2.0, aucun des écosystèmes étudiés ne semble convenir dans le cadre d'une gestion de connaissances suffisamment formelle.

Les connaissances pouvant naître de la capitalisation des ressources intéressent les organisations. Cet intérêt est conduit en grande partie par l'idée que ces nouvelles connaissances peuvent enrichir leurs processus de décision. Nous revenons dans le chapitre suivant sur cette notion de décision et son outillage à l'aide de systèmes d'aide à la décision. Ceci afin de voir de quelle manière ces systèmes peuvent bénéficier d'un système de gestion de connaissances prenant en compte les productions des applications Web.

4. Une aide à la décision fondée sur les médias sociaux en organisations

Comme nous l'avons vu dans la section 2.3, les organisations trouvent dans les médias sociaux une source dynamique d'informations leur permettant d'améliorer la vision qu'elles possèdent d'elles-mêmes au sein de leur environnement partenarial et concurrentiel. Cette meilleure connaissance d'elle-même apporte de la précision dans le travail des décideurs de l'organisation.

Ces derniers peuvent être impliqués presque quotidiennement dans des situations pour lesquelles une décision liée à la stratégie locale ou globale de l'organisation doit être prise. Ces prises de décision sont fondées sur une bonne connaissance du domaine par le décideur, même si cette dernière n'est pas parfaite. Afin de palier les limites humaines, il est de plus en plus courant d'outiller les décideurs avec des systèmes se chargeant de compléter et étayer leur connaissances vis-à-vis de la situation pour laquelle une décision doit être prise. Nous utilisons le terme de processus de décision pour décrire le cheminement humain, éventuellement outillé, conduisant d'une situation initale à sa résolution en passant par la prise de décision en elle-même, caractérisée par le choix fait par le décideur parmi un ensemble de solutions possibles. Ces processus peuvent être outillés à l'aide de systèmes d'aide à la décision.

L'idée d'utiliser les médias sociaux comme source d'informations pour aider la prise de décision n'est pas nouvelle. Nous rappellons brièvement dans la section 4.1 le contexte de notre étude et l'apparition des médias sociaux dans le paysage de l'aide à la décision. Nous revenons ensuite sur la définition de prise de décision et des processus afférents (section 4.2), avant de voir de quelle manière l'analyse des médias sociaux peut être vue comme une source de connaissances dans un processus décisionnel (section 4.3) et de ce fait être intégrée à un système d'aide à la décision (section 4.4).

4.1 Contexte

L'explosion du nombre de *smartphones* jointe à celle du nombre de médias sociaux accessibles en mobilité sous la forme d'applications mobile ou Web a apporté de nouveaux compor-

tements sociaux, en particulier lors d'évènements impliquant un grand nombre de personnes. L'un de ces comportements est l'envoi désormais quasiment systématique de photos, avis, alertes sur les médias sociaux par les personnes présentes, avant même de veiller à leur propre sécurité, lorsqu'un imprévu potentiellement désagréable se présente [Starbird et Palen, 2010]. La figure 4.1 est une comparaison de deux photographies prises lors du même évènement mais à huit ans d'écart : l'élection d'un nouveau pape. La seconde photographie, prise lors de l'élection du nouveau pape François exprime parfaitement ces nouvelles habitudes mobiles sociales.

FIGURE 4.1 – Élection d'un pape à huit ans d'intervalle

Des études [Sakaki, 2010, Vieweg *et al.*, 2010, Crooks *et al.*, 2012] montrent ainsi que les médias sociaux (et en particulier Twitter) peuvent tout à fait être utilisés pour observer les comportements sociaux en réponse à une crise. Différentes méthodes [Singh *et al.*, 2010, Starbird et Palen, 2010] ont été proposées pour permettre la détection d'événements à partir de l'activité sociale d'une population.

Suite à l'élection présidentielle de 2007 au Kenya, de nombreuses violences ont éclaté dans tout le pays. Des journalistes locaux ont alors mis en place une plate-forme permettant de recueillir par SMS ou mail des témoignages anonymes des abus subis par les populations. Ces témoignages ont été synthétisés au sein d'une application cartographique affichant au fur et à mesure l'état des exactions commises à travers le pays. Cette application a grandement

facilité la tâche de coordination des ONG sur place. Suite à ces évènements et devant le succès de la chose, la plate-forme s'est généralisée sous le nom « d'Ushahidi » (qui veut dire témoin en swahili)[1]. Sous le patronage de la fondation à but non-lucratif du même nom, d'autres solutions ont émergé, proposant un ensemble d'applications permettant la coordination en temps de crise :

Ushahidi composant le plus ancien de la fondation — qui lui a d'ailleurs donné son nom. Il s'agit d'une plate-forme permettant d'afficher sur une carte interactive des témoignages géo-localisés et autres informations issues de divers canaux d'informations (SMS, email, RSS, Twitter). Cette plate-forme a été plébiscitée dans les pays du Sud car elle permet le suivi d'informations dans des conditions extrêmes (conçue pour se connecter à l'aide de simple puces GSM comme des *feature-phones*).

SwiftRiver composant permettant la surveillance continue de différents canaux d'informations (SMS, email, RSS, Twitter), l'indexation (dans une base de données relationnelles) et l'analyse de leur contenu permettant la mise en relation d'informations déconnectées (SMS justifiant des propos tenus sur Twitter, etc.).

CrowdMap il s'agit simplement d'une version *Software as a Service* (SaaS) de la plate-forme Ushahidi.

La plate-forme Ushahidi a depuis été utilisée avec succès lors des tremblements de terre d'Haïti et du Chili (2010) ou d'un incendie de forêt géant en Russie la même année. Spécialisée dans l'acheminement et le classement de messages, cette plate-forme apporte une bonne vision des évènements en cours et permet la compréhension globale de la crise observée. Néanmoins, elle manque d'atouts concernant la coordination des réponses éventuelles.

À ce sujet notons l'apparition, suite au Tsunami de 2004 au Sri Lanka, de la plate-forme logicielle **Sahana**[2]. Cette suite logicielle *open source* utilisable d'un simple navigateur s'est depuis illustrée lors des évènements dramatiques tels que le tremblement de terre au Pakistan en 2005, l'ouragan Sandy aux États-Unis d'Amérique ou les feux de forêt au Chili en 2012. La plate-forme propose les deux composants suivants :

Eden outil historique de la plate-forme, il est conçu pour permettre la coordination de différentes organisations sur place suite à une crise et assurer le suivi de leurs missions. En plus d'outils classiques de gestion de projets — enregistrement des organisations ou personnes volontaires, gestion du matériel, des prêts, inventaire des objets trouvés, tâches à effectuer, etc. —, ce logiciel possède sa propre solution de suivi de situation en permettant la réception de SMS, tout comme SwiftRiver.

1. Site officiel de la fondation http://ushahidi.com/ (accessible le 10 novembre 2013)
2. Site officiel de la fondation qui s'occupe du développement de la plate-forme : http://sahanafoundation.org/ (accessible le 10 novembre 2013)

Vesuvius composant permettant le suivi des personnes disparues — de la même manière que Google *person finder*[3] — ainsi que la bonne répartition des blessés selon les places disponibles dans les hôpitaux des alentours.

D'autres initiatives ont pu profiter du Web pour agréger une forte communauté de volontaires et proposer des solutions au suivi de crise. Ainsi, la communauté CrisisCommons[4] propose aux internautes de contribuer à leur Wiki en ajoutant des ressources — logiciels, documentations, conseils, etc. — à leur base de connaissances.

Enfin, il ne faut pas oublier l'initiative globale d'ouverture des données publiques pour les rendre accessibles sur le Web. Stigmatisé par l'arrivée d'acteurs tels que WikiLeaks[5] ou les actions d'Aaron Swartz[6], l'*open data* est cependant un mouvement ayant progressivement gagné ses lettres de noblesse via la publication d'un certain nombre de référentiels étatiques[7, 8]. En permettant le libre accès à nombre de bases de données publiques, le croisement d'informations devient possible et permet la découverte de nouvelles informations.

4.2 La prise de décision

La définition la plus courante de la décision est le processus par lequel une ou plusieurs personnes — appelées décideurs — vont être amenées à effectuer un choix entre plusieurs solutions possibles pour résoudre le problème auquel elles sont confrontées [Parthenay, 2008]. La prise de décision intervient dans de nombreux domaines et en particulier, pour ce qui nous intéresse, dans les domaines de l'intelligence économique [Link-Pezet *et al.*, 1999] ou la stratégie militaire [Fewell et Hazen, 2005]. Auger *et al.* [2006] rappellent alors une définition de l'aide à la décision : apporter une aide au décideur au cours des différentes étapes de la prise de décision. Il peut s'agir par exemple de l'outiller pour faciliter la mobilisation des bonnes informations nécessaires à la compréhension du domaine étudié ou encore de lui présenter ces mêmes informations sous des formes synthétiques permettant la mise en valeur des informations clés.

Le processus de décision a été maintes fois décrit dans la littérature. Nous suivons la définition donnée par Courtney [2001] représentée dans la figure 4.2.

Cette modélisation du processus de décision se résume en quatre grandes étapes discutées par Fewell et Hazen [2005] et résumées dans l'acronyme Observer, Orienter, Décider,

3. http://google.org/personfinder/global/home.html (accessible le 10 novembre 2013)

4. Site officiel : http://crisiscommons.org/ (accessible le 10 novembre 2013)

5. Association à but non lucratif s'étant donné pour mission de publier les fuites d'informations normalement protégées — typiquement par le secret d'État — afin d'alerter les citoyens sur les agissements possiblement non éthiques de leurs gouvernements. Site officiel : http://wikileaks.org/ (accessible le 10 novembre 2013)

6. Activiste américain s'étant suicidé le 11 janvier 2013. Il est connu (entre autre, ses réalisations étant trop nombreuses pour être toutes évoquées ici) pour avoir tenté de mettre à disposition des internautes l'ensemble

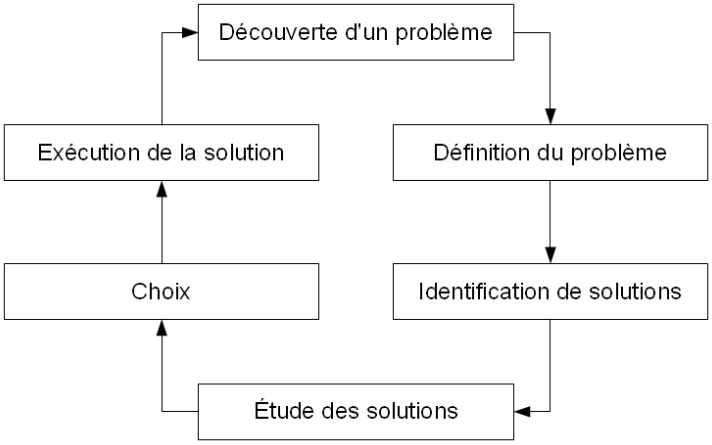


FIGURE 4.2 – Schéma d'un processus de décision d'après Courtney [2001]

Agir (OODA). Pour éviter l'emboîtement — parler d'étape de décision dans un processus de décision —, il conviendrait mieux d'utiliser le terme de « choisir » plutôt que « décider » pour la troisième étape. C'est ce terme que nous choisissons d'adopter pour la suite, tel qu'affiché dans la figure 4.3.

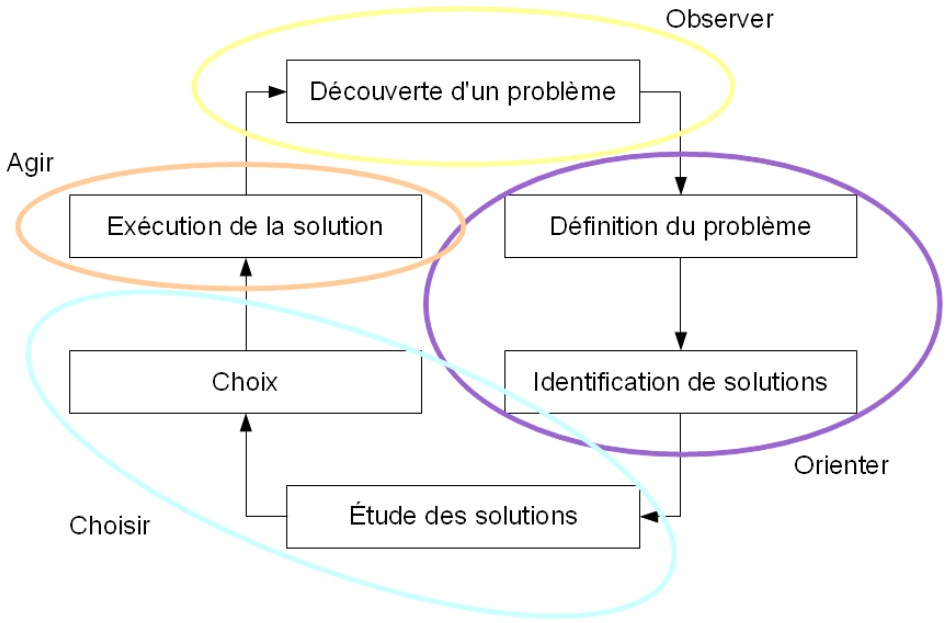


FIGURE 4.3 – Schéma du processus de décision de la figure 4.2 intégré à l'approche OODA de Fewell et Hazen [2005]

Observer Il s'agit pour le décideur d'obtenir une compréhension globale de la situation actuelle tout en ayant à disposition les outils nécessaires pour permettre la fouille de situations similaires du passé.

des archives de JSTOR pour permettre le libre accès au savoir

7. http://www.data.gov/ (accessible le 10 novembre 2013)
8. http://www.etalab.gouv.fr/ (accessible le 10 novembre 2013)

Orienter L'analyse de la situation actuelle et des précédentes permet la définition des solutions alternatives pouvant conduire à la résolution du problème étudié.

Choisir Ayant toutes les cartes en main (*situation awareness*), le décideur peut désormais choisir la meilleure option. Ce choix s'effectue à l'aide de comparaisons sur des métriques dépendantes du problème étudié.

Agir Il s'agit simplement d'exécuter le plan choisi à l'étape précédente, tout en prenant garde de bien conserver toutes les issues de cette exécution, afin de les capitaliser et pouvoir les utiliser à nouveau dans un autre processus de décision.

Dans le cadre de cette thèse, nous nous intéressons à l'étape d'orientation. Il s'agit de permettre la mobilisation des bonnes informations depuis diverses sources de données pour donner toutes les cartes en main au décideur. Ces sources de données peuvent être de natures variées, selon qu'il s'agit de bases de données relationnelles, non-relationnelles, de bases de connaissances comme les *triplestores* ou même ce que l'on appelle des sources ouvertes.

Le terme de source ouverte est ambigu. Nous ne voulons pas parler ici du libre accès au code source d'un logiciel, tel que prôné par le mouvement du logiciel libre, mais de l'étude d'informations diffusées sur des supports publiquement accessibles. Il va s'agir par exemple des médias (journaux, radio, Web, etc.), des registres officiels (cadastres, états-civils, etc.), ou plus récemment également des informations disponibles sous l'égide du mouvement d'*open data*. Les médias sociaux, tels que nous les avons définis dans la section 2.1, sont considérés comme des sources ouvertes.

La principale problématique des sources ouvertes reste bien entendu leur agrégation et leur exploitation, du fait qu'elles sont par définition de nature hétérogène. Leur prise en compte dans des processus décisionnels demande un traitement relativement lourd. La consultation de sources ouvertes peut cependant être vue comme une réponse raisonnable à la constitution, tout aussi coûteuse, d'un corpus d'informations homogènes. De plus, les sources ouvertes sont réputées plus à même de refléter la dynamique de l'actualité et, ce faisant, plus intéressantes dans le cadre de décisions stratégiques, commerciales ou militaires (voir à ce propos [Steele, 1997]).

Dans la section suivante nous nous penchons plus précisément sur l'exploitation d'une source ouverte particulière : les médias sociaux. Avec l'avènement du tout numérique, la plupart des sources ouvertes possèdent désormais un équivalent en ligne, simplifiant en partie leur exploitation. Comme nous l'avons vu, les médias sociaux peuvent être vus comme une manifestation numérique d'activités sociales réelles. Leur analyse doit permettre d'obtenir des informations assez fidèles sur les groupes sociaux impliqués.

4.3 L'analyse des médias sociaux

L'analyse des médias sociaux bénéficie des travaux effectués depuis de nombreuses années dans le domaine de l'analyse des réseaux sociaux ou *Social Network Analysis* (SNA) en anglais. Dans leur livre, Degenne et Michel [2004] rappellent cet historique, tout comme Beaudoin [2009] qui offre un panorama bibliographique synthétique, tout en se questionnant sur les nouveautés éventuelles introduites par le Web dans ce domaine.

L'analyse des réseaux sociaux, en sociologie, consiste en l'étude d'une communauté, représentée sous la forme d'un réseau dont les nœuds sont les différents acteurs (personnes mais aussi institutions) et les liens des relations qu'ils peuvent entretenir deux à deux (influence, collaboration, parenté, hiérarchie, localité, etc.). Il s'agit ainsi de formaliser l'état social d'un groupe d'acteurs pour en faire ressortir des caractéristiques à analyser.

La SNA introduit différents concepts que nous retrouverons dans la suite de ce mémoire :

Graphe la notion de graphe se définit comme la matérialisation d'un réseau social sous la forme d'un ensemble de nœuds reliés deux à deux par un ensemble de liens.

Graphe dirigé / non dirigé la notion de direction d'un graphe définit si les liens du graphe ont un sens, c'est-à-dire s'ils vont d'un nœud source vers un nœud cible (graphes dirigés) ou au contraire si les liens entre les nœuds représentent des relations mutuellement supportées (graphes non dirigés).

Densité du graphe la notion de densité donne la probabilité, pour tout nœud, d'être connecté à l'ensemble des autres nœuds du réseau. Une densité de 1 désigne ainsi un graphe au sein duquel tous les nœuds sont connectés à tous les autres. Une densité de 0,5 désigne un graphe au sein duquel chaque nœud sera, en moyenne, connecté à la moitié des nœuds.

Degré d'un nœud nombre de liens d'un nœud. Dans le cadre d'un graphe dirigé, on pourra parler du degré entrant et du degré sortant d'un nœud.

Chemin dans le graphe un chemin dans un graphe entre deux nœuds est une suite de liens à parcourir pour se rendre de l'un à l'autre des nœuds en question, en visitant pour cela d'autres nœuds.

Bond ou étape lorsqu'on décrit un chemin entre deux nœuds d'un graphe, on utilisera fréquemment le terme de bond ou étape pour désigner le passage d'un nœud à un autre lors du parcours dudit chemin.

Distance d'un chemin la distance d'un chemin dans un graphe représente le nombre de bonds nécessaires pour le parcourir.

Plus court chemin au sein d'un graphe, le plus court chemin entre deux nœuds est le chemin qui requiert le plus faible nombre de bonds successifs pour passer de l'un à l'autre des

nœuds en question, autrement dit le chemin ayant la plus faible distance entre ces deux nœuds.

Diamètre le diamètre d'un graphe représente, parmi tous les plus courts chemins possibles du graphe, la distance la plus longue.

Pour les organisations, l'analyse des réseaux sociaux revêt un caractère stratégique [Lazega, 1994]. Différentes mesures de l'importance d'un nœud au sein d'un graphe existent. Cette mesure de l'importance d'un nœud dans un graphe est le reflet de l'importance d'une organisation dans son environnement. La plus simple est la notion de centralité. Un nœud central dans un graphe non dirigé est un nœud accaparant une majorité de liens. De fait, ce nœud devient un point de passage obligé lors des parcours dans le graphe.

Les médias sociaux facilitent en partie l'étude de certains réseaux sociaux. En tant qu'outils de collaboration, c'est-à-dire d'outils permettant à différentes personnes de réaliser un but commun que ce soit par l'échange ou la création de ressources en ligne, les médias sociaux servent de révélateur au réseau social constitué des différents collaborateurs. De nombreux états de l'art ont été effectués dans ce domaine, en particulier ceux de van Duijn et Vermunt [2006], Erétéo *et al.* [2009b], Combe *et al.* [2010].

Erétéo *et al.* [2009a] énumèrent les nombreux algorithmes du domaine, les mesures communément attendues d'un outil de SNA et comparent ces outils entre eux. Ces travaux ont été menés autour du projet SemSNA [Erétéo *et al.*, 2011] pour représenter de manière sémantique le processus d'analyse d'un réseau social. Ce *framework* apporte une modélisation sémantique d'un réseau social permettant d'étiqueter les relations du réseau, l'importance (centralité, degré, etc.) des nœuds considérés etc.

Parmi les métriques de SNA ayant acquis dernièrement une certaine notoriété, nous nous sommes penchés sur le *Edgerank*. Le *Edgerank* est une valeur définie par un algorithme mis au point par Facebook, visible dans la formule (4.1) qui permet de ranger les différents objets devant apparaître sur le *wall* d'un utilisateur A, en fonction du nombre de *edges* que ces objets auront provoqués. Le terme d'Objet, dans le vocabulaire de Facebook, représente tous les types de publications pouvant être ajoutées sur le *wall* d'un utilisateur. Il peut s'agir d'un message de statut, du partage d'une photographie, d'un lien vers un site Web, d'un lien vers une chanson, etc. Le terme de *edge*, de son côté, désigne les interactions possibles, au sein de Facebook ou au travers des APIs mises en places, sur les objets. Il peut s'agir d'un clic sur le bouton « j'aime », de l'ajout d'un commentaire, de l'ajout d'une étiquette sur une photographie, etc. Il est intéressant de noter que la plupart des *edges* génèrent en retour de nouveaux objets avec lesquels il est possible d'interagir : l'ajout d'un commentaire crée un objet commentaire, le fait de cliquer sur un bouton « j'aime » crée un objet « untel aime telle chose », etc.

La formule (4.1) dépend des trois paramètres suivants :

$$\sum_{edges\ e} u_e w_e d_e \tag{4.1}$$

- u_e exprime le degré d'affinité existant entre A et B. Cette valeur, calculée par ailleurs, n'est pas expliquée par le réseau social.
- w_e est un facteur de pondération de l'*edge* considéré en fonction de son type. Le partage d'un article a ainsi plus de poids qu'un clic sur un lien « j'aime » par exemple.
- d_e est un facteur de pondération du résultat final qui décroît en fonction de la date de l'*edge*. Plus ce dernier est ancien, moins son intérêt est grand.

À la différence de la plupart des mesures de SNA, le *Edgerank* s'attache ainsi à accorder un score à un lien entre deux nœuds et non à un nœud. Le graphe social de Facebook mêle en effet deux types de nœuds : les utilisateurs et les objets. En classant les différents objets à apparaître sur le *wall* d'un utilisateur, le *Edgerank* permet d'exprimer la force du lien entre un nœud utilisateur et un nœud objet. En complément d'un score de centralité, qui permet de classer les acteurs du réseau par importance, le *Edgerank* permet ainsi de déterminer les relations les plus solides pouvant exister entre les acteurs du graphe.

La SNA se heurte néanmoins à un facteur de taille dans les raisonnements : la dimension temporelle des réseaux. La plupart des travaux du domaine portent sur des structures de graphes statiques, sans prendre en compte le côté dynamique dans le temps de ces structures. Shekhar et Oliver [2010] proposent une méthode pour représenter l'évolution temporelle de ces réseaux.

Tantipathananandh *et al.* [2007] proposent un *framework* permettant d'identifier des communautés au sein de réseaux sociaux dynamiques. Lahiri et Berger-Wolf [2008] proposent une méthode pour identifier des patrons d'interactions périodiques et, par extension, proposent de prédire l'évolution des réseaux dynamique [Lahiri et Berger-Wolf, 2007]. Dans un monde en constante évolution, où la mobilité des collaborateurs va jouer un rôle de plus en plus grand, il va devenir stratégique de s'intéresser à l'évolution des réseaux.

De nombreux travaux ont également cherché à étudier le phénomène de diffusion d'informations au sein d'un réseau. Aspnes *et al.* [2006] ont montré via l'étude de la transmission d'un virus informatique au sein d'un réseau d'ordinateurs de quelle manière les principes issus de la théorie des jeux pouvaient être utilisés pour prédire sa propagation au sein d'une communauté. [Lahiri et Cebrian, 2010] proposent des méthodes à base d'algorithmes génétiques. [Ugander *et al.*, 2012] tentent de prouver que la diffusion d'informations suit les mêmes règles que les modèles établis de contagion virale en utilisant le développement de Facebook en cas d'usage.

La représentation mathématique des réseaux sociaux permet d'utiliser les mêmes mé-

thodes que celles utilisées pour les ressources Web pour parcourir un graphe de données sociales et l'indexer. Cette indexation demande cependant un export préalable du graphe à étudier hors de son contexte d'application et n'apparaît donc pas pertinent dans le cadre d'une indexation continue du Web. En revanche cette technique se révèle intéressante dans le cadre de l'étude *a posteriori* d'un évènement particulier.

Divers logiciels permettent de réutiliser ces données collectées au sein d'écosystèmes numériques pour les visualiser. Citons Pajek [de Nooy *et al.*, 2005], Gephi [Bastian *et al.*, 2009], ou encore Tulip [Auber *et al.*, 2012]. Ces logiciels permettent classiquement de visualiser les données récupérées et d'exécuter des algorithmes sur ces données. Combe *et al.* [2010] abordent ces outils dans leur état de l'art et couvrent à l'aide de tableaux l'ensemble des algorithmes et possibilités offertes par chacun d'entre eux.

Quelques applications permettent de mesurer l'influence d'un utilisateur au sein de son réseau social. Cette influence est alors mesurée au travers de ses activités sur les différents médias sociaux et le nombre de contacts qu'il entretient. Citons à ce titre les deux applications Web Klout et PeerIndex [9]. Ces applications requièrent d'un utilisateur qu'il les lie à ses comptes ouverts sur divers médias sociaux.

Cette liaison entre différentes applications Web est rendue possible par l'existence d'APIs proposées par ces applications. Néanmoins, chaque application proposant sa propre API incompatible avec les autres rend l'interconnexion des applications Web fastidieuse et dans la pratique, plutôt que d'offrir un écosystème ouvert d'applications, provoque l'existence parallèle de *walled gardens* [Halpin, 2008].

L'une des principales critiques envers ces applications est qu'elles ne tiennent pas compte de l'influence réelle que peut avoir une personne. Ainsi, plus un utilisateur possède de comptes différents et produit du contenu, plus il a de chances d'avoir un score élevé. À l'inverse, même si une personne est reconnue internationalement, tant qu'elle ne publie pas, son score n'évoluera pas. L'exemple couramment donné est celui de Barack Obama : comme il ne publie que rarement et ne possède pas un profil par média social existant, son score est inférieur à nombre de blogueurs dont l'influence réelle dépasse rarement le cadre d'Internet, voire de leurs centres d'intérêts.

Si l'analyse des réseaux sociaux apparaît comme une source d'informations intéressante pour les organisations dans le cadre de processus de décisions commerciales ou stratégiques, la SNA ne suffit pas au processus. Il faut en effet croiser ces informations avec d'autres sources pour en faire ressortir différents choix qui pourront être proposés au décideur. Dans des environnements complexes, ces derniers recourent de plus en plus à des systèmes d'aide à la décision.

9. Accessibles aux adresses http://klout.com/home et http://www.peerindex.com/

4.4 Les systèmes d'aide à la décision

Dans des environnements de plus en plus complexes, les processus de décision se sont outillés afin de venir épauler les décideurs sur chacune des étapes du cycle, voire en automatiser certaines. Le fait d'outiller le processus de décision, que ce soit à l'aide de systèmes évolués ou plus simplement de méthodologies venant appuyer le processus cognitif du décideur, entre dans ce qui est appelé couramment l'aide à la décision. Les spécificités liées à l'évolution des théories de la décision vers une méthodologie de l'aide à la décision sont traitées par Tsoukiàs [2008].

La quantité d'informations à traiter étant constamment en augmentation, les décideurs des organisations se sont tournés depuis longtemps vers des systèmes informatiques dédiés pour outiller leurs prises de décision. Bien que l'outil informatique prenne une place toujours plus importante, le choix d'une solution finale reste une étape humaine au sein du processus de décision. L'interactivité entre l'humain et la machine est donc un élément indispensable des systèmes d'aide à la décision. On parle alors de système interactif d'aide à la décision (SIAD) ou *Decision Support System* (DSS). Arnott et Pervan [2005], Zaraté [2005] proposent un état de l'art des systèmes d'aide à la décision.

Une première approche des DSS permet de décrire leur fonctionnement de bout en bout au travers de trois blocs fonctionnels [Shim *et al.*, 2002, Boreisha et Myronovych, 2008] :

1. une base de données servant de source de connaissances sur les situations observées au cours du processus de décision. Cette base de données doit recouvrir différentes sources distinctes permettant d'enrichir la connaissance générale du système. Cela passe par l'enregistrement de données provenant de capteurs, de systèmes de gestion de connaissances, etc.
2. une interface permettant à l'utilisateur — le décideur — d'interagir avec le système (à l'aide de cadrans ou formulaires intégrés à un tableau de bord).
3. une partie logique servant à effectuer des traitements ordonnés par le décideur à l'aide de l'interface sur la base de données. Ces traitements permettent d'extraire, en fonction de la situation ou de la requête, de proposer au décideur des solutions à partir desquelles il fondera sa décision.

Nous préférons à cette description celle de Darren Knight (2002) citée par Auger *et al.* [2006]. Cette présentation d'un SIAD suit en effet les quatre étapes de la boucle OODA présentée avant. La figure 4.4 représente cette seconde approche. Celle-ci sépare de plus le processus de décision en trois domaines ou niveaux d'abstraction :

Le monde physique au sein duquel se déroule des évènements que l'on souhaite suivre, étudier,

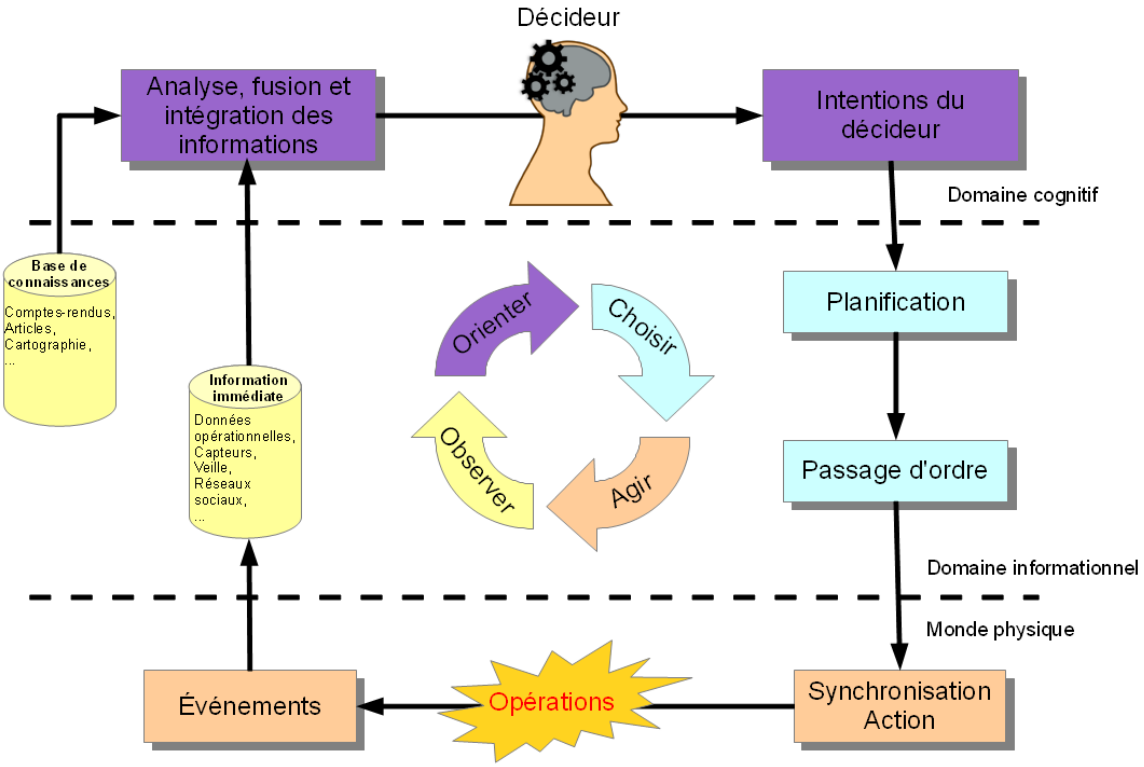


FIGURE 4.4 – Schéma des processus interne à un SIAD par Knight (2002)

mémoriser. C'est aussi dans ce domaine que s'exprimera, en action, le fruit du processus de décision.

Le domaine informationnel au sein duquel l'information est théoriquement — si l'on s'abstrait bien sûr des problématiques techniques / institutionnelles d'autorisation d'accès aux données — partagée et accessible par toutes les parties prenantes. Cet accès est le même pour tous et les informations traitées à ce niveau ne dépendent d'aucun facteur extérieur.

Le domaine cognitif Le choix final, qui va différencier une prise de décision d'une autre, en influant principalement sur les actions à entreprendre, appartient au domaine cognitif. Il s'agit en effet pour le décideur de pouvoir interpréter les informations communément disponibles à l'échelle informationnelle pour obtenir la connaissance nécessaire à l'évaluation des solutions envisageables. Cette interprétation des informations étant nécessairement humaine, le choix de la solution sera également influencé par l'expérience du décideur. Le choix effectué étant ainsi intimement lié au décideur, on parle d'intention du décideur.

Il apparaît de plus en plus important, face au flot croissant d'informations auquel le décideur doit faire face, d'automatiser, du moins en partie, les processus d'analyse, de fusion et

d'intégration de ces mêmes informations, afin de lui en présenter une version synthétique qui lui fera gagner temps et précision.

Les SIAD profitent ainsi des dernières avancées en termes de fouille de données (*datamining*) et d'analyse multidimensionnelle (« cube OLAP »). La fouille de données a pour but de trouver des corrélations entre différentes variables au sein de grandes masses de données et donc d'apporter de nouvelles connaissances à l'utilisateur. De son côté, l'*On-Line Analytical Processing* (OLAP) est un ensemble de techniques ne cherchant qu'à faciliter la présentation d'une multitude d'informations selon un nombre *a priori* non restreint de critères à l'utilisateur. Ces différentes technologies ont bénéficié d'avancées majeures liées aux progrès effectués dans le domaine de l'informatique répartie — plate-forme de calcul distribué (Hadoop [10], MapReduce[32], etc.), bases de données sous forme de graphe (Neo4j[32], DEX[32], etc.), bases de données non relationnelles (Kyoto Cabinet[32], Voldemort[32], Cassandra[32], Hbase[32], etc.). Ces progrès permettent d'intégrer de nombreuses sources de données dans les SIAD, dont les sources ouvertes de l'*open data* ou des APIs sociales (Twitter, Facebook, etc.).

L'utilisation intensive des applications Web et plus particulièrement des médias sociaux, même au sein des organisations, nous incite à imaginer une nouvelle famille d'outils d'aide à la décision : les *Social Decision Support System* (SDSS). Nous ne voyons pas les SDSS comme des systèmes de vote évolués [Turoff *et al.*, 2002], mais plutôt comme des SIAD fondés sur l'étude des interactions sociales d'utilisateurs au sein d'un écosystème numérique d'applications Web. Ceci requiert un croisement de théories que nous illustrons dans la figure 4.5 :

— Utilisation d'une base de connaissances pour modéliser les différents types d'interactions sociales pouvant être rencontrées sur les médias sociaux ;
— Utilisation des méthodes issues de l'analyse des gros volumes de données, les médias sociaux générant des quantités astronomiques d'informations ;
— Utilisation des théories sociales des systèmes d'aide à la décision de groupes.

Les écosystèmes numériques d'applications Web que nous avons décrits dans la section 4.1 peuvent, par exemple, être considérés comme des SDSS. Ils répondent en effet dans leurs implémentations aux six conditions de Gachet (2002) rappelées dans Zaraté [2005] et en particulier celles traitant de l'indépendance des données, de l'utilisation intensive du Web et du support de la mobilité, même dans les zones à faible couverture (quitte à passer par des moyens détournés comme l'usage de téléphones cellulaires).

10. Toutes ces technologies possèdent une fiche Wikipedia donnant accès à leur site officiel. Toutces projets sont actifs au 8 août 2013.

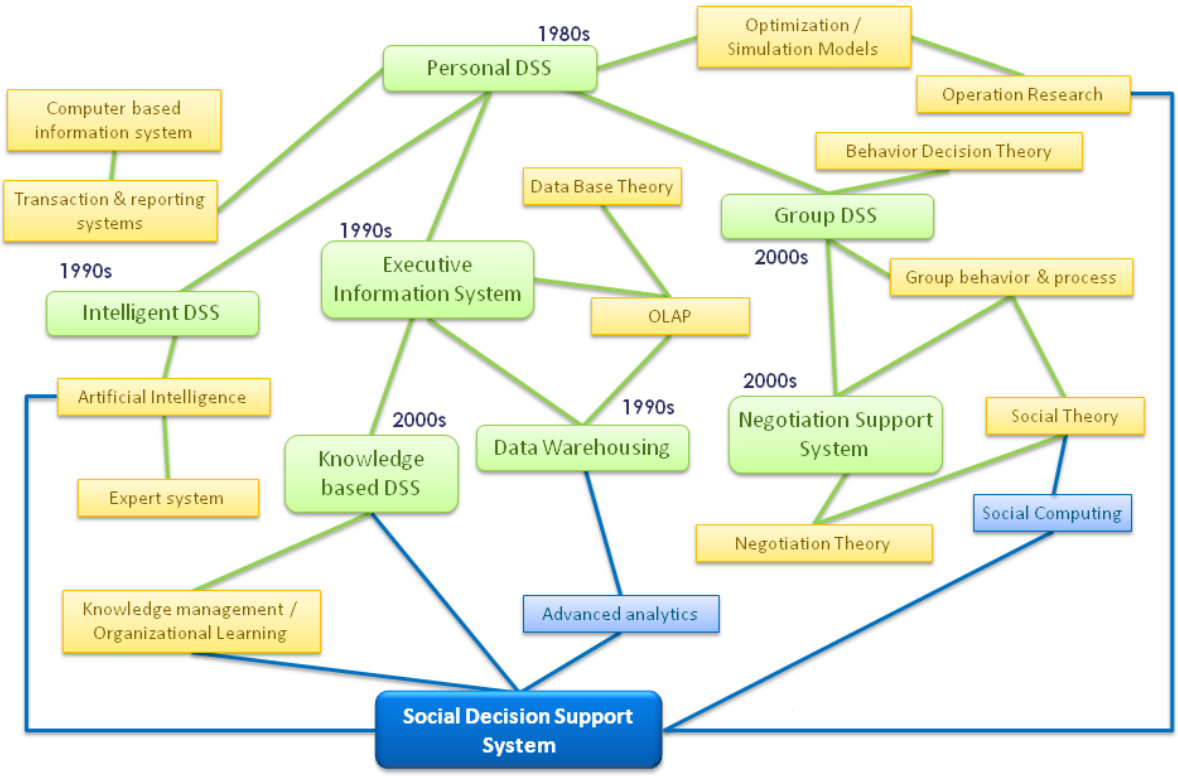


FIGURE 4.5 – Évolution possible des courants de l'aide à la décision d'après Arnott et Pervan [2005], p. 6

4.5 Synthèse

Comme nous l'avons vu dans ce chapitre, quelques systèmes peuvent d'ores et déjà être considérés comme des SDSS. Cependant, ces systèmes sont avant tout conçus à des fins de gestion de crises. Nous constatons alors une limite importante dans l'implémentation actuelle des SDSS. Les systèmes développés actuellement ne permettent pas, ou le permettent, mais difficilement, de capitaliser les informations extraites des données sociales étudiées au cours d'un processus de décision en vue d'une réutilisation postérieure. Ce type d'outil n'est pour l'instant bien souvent installé que pour une crise et désinstallé lorsque le danger immédiat est passé. Il ne reste après la crise que les rapports synthétiques sur les évènements, sans possibilité de rejeu sur les données sociales utilisées au cours des processus décisionnels.

Or, l'étude sur le long terme des données sociales peut apporter de nouvelles informations qui ne sont pour l'instant pas exploitées. Si cette exploitation peut être discutable dans le cadre de la gestion de crises, elle nous semble primordiale dans le cadre d'une veille technologique, stratégique ou concurrentielle en vue de processus de décision organisationnels. La capitalisation sur le long terme des données issues des médias sociaux doit donc être mise en place.

5. Synthèse

Les organisations se tournent davantage vers les applications Web et les médias sociaux. Ces derniers sont appréciés pour la facilité et la vitesse accrue qu'ils permettent dans les tâches collaboratives, en facilitant la mise en relation des membres de l'organisation. Ils permettent aux membres de s'échanger des informations qui s'inscrivent dans des processus de veille collaborative, de curation ou de simple communication en complément du téléphone ou des mails. La facilité de partage des informations permet, même dans des cas d'utilisation avant tout personnels (veille, gestion de documents, etc.) de mettre à disposition des autres utilisateurs les informations utilisées.

De la même manière que les organisations s'entrecroisent et travaillent ensemble au sein d'écosystèmes d'affaires, les applications Web utilisées s'insèrent bien souvent dans des écosystèmes numériques qui tendent à offrir une expérience complète aux utilisateurs. Avec la convergence des outils sur l'internet, c'est tout le système d'information (SI) des organisations, avec sa gestion de connaissances et ses nouveaux moyens de communication et de collaboration, qui se trouve impliqué au sein d'un même écosystème numérique.

Les organisations cherchent à conserver et valoriser leur capital de connaissances. Pour ce faire, elles cherchent à expliciter ces connaissances sous forme de documents afin de pouvoir les capitaliser au sein d'une base de connaissances et permettre leurs réutilisations. Les connaissances sont personnelles et proviennent de l'interprétation par une personne d'informations lui étant communiquées. Le transport de ces informations implique la notion de document et en informatique celle de document numérique. Ces définitions sont fondées sur une approche formelle du document, à la forme et aux intentions clairement identifiées. Cette approche se fonde sur des modèles qui ne sont plus cohérents avec la structure et les besoins apportés par l'utilisation des médias sociaux.

Toutes les activités que vont réaliser les membres de l'organisation au sein des applications Web représentent une manne d'informations pour l'organisation. Ces informations, si elles sont correctement capitalisées de manière à favoriser leur réinterprétation, peuvent apporter de nouvelles connaissances à l'organisation :

— connaissances plus précises de ses propres activités, au travers des discussions en cours, thématiques abordées dans la veille de ses membres, etc. Ces connaissances nécessitent

de pouvoir capitaliser les activités des membres et leurs contenus au sein d'une base de connaissances.

— connaissances plus précises sur son environnement partenarial et concurrentiel au travers des liens que vont entretenir ses membres avec des membres d'autres organisations. Ces connaissances nécessitent de pouvoir capitaliser au sein d'une base de connaissances les informations sur le réseau social de l'organisation.

La plupart des écosystèmes numériques actuels ne permettent pas facilement la capitalisation des informations échangées et partagées au sein d'une base de connaissances, comme les informations sur le réseau social d'une organisation. Et quand bien même ces écosystèmes proposent une méthode de capitalisation, cette dernière ne correspond pas au besoin de structures formelles fortes des organisations. Nous avons considéré dans notre thèse que la capitalisation de connaissances pouvait s'effectuer à l'aide de méthodes d'étiquetage ou d'indexation des documents. Les principaux écosystèmes utilisés actuellement par les organisations ne proposent, au mieux, que la gestion d'ensembles d'étiquettes, ou *tags*, comme les folksonomies. Ces dernières sont rarement partagées d'une application à l'autre au sein d'un même écosystème numérique. Or, les organisations ayant à gérer de grosses quantités de données pouvant provenir de différents services, le référentiel utilisé pour mener à bien cette tâche d'étiquetage des documents nécessite un niveau de formalisme que nous trouvons dans les ontologies.

La maîtrise des connaissances d'une organisation aide, comme nous l'avons vu, ses décideurs à effectuer leurs choix. Or, le manque de capitalisation des données issues des médias sociaux remarqué dans le cadre général de la gestion de connaissances se remarque également dans le cas plus particulier des systèmes d'aide à la décision. Ces derniers, bien que proposant des outils permettant l'analyse des médias sociaux, ne permettent pas la capitalisation sur le long terme des informations qu'ils identifient au sein de leurs bases de connaissances.

Nous postulons que la réunion de différents éléments nous permet d'entrevoir les grandes lignes d'un écosystème numérique venant en appui d'une organisation souhaitant piloter son écosystème de connaissances. Intégrée avec d'autres organisations dans son tissu partenarial, l'organisation souhaite bénéficier d'une plate-forme lui permettant d'améliorer son capital de connaissances. Pour ce faire, l'écosystème numérique utilisé doit prendre à la fois en compte les informations issues des bases documentaires et de médias sociaux utilisés dans l'organisation pour stimuler la collaboration de ses communautés constitutives. L'écosystème numérique utilisé pour soutenir cette gestion de connaissances doit également autoriser la réutilisation des informations capitalisées dans la base de connaisances dans les processus décisionnels liés au pilotage de l'écosystème de connaissances.

Si les SIAD doivent tirer partie de connaissances issues à la fois de bases documentaires et d'échanges ayant lieu au sein de médias sociaux, il est nécessaire de proposer un modèle

théorique décrivant le cycle de gestion de connaissances adapté à l'utilisation conjointe de bases documentaires et de médias sociaux. Ce modèle doit identifier les étapes nécessaires à la création de connaissances depuis les données enregistrées par divers systèmes informatiques et à la sauvegarde de ces mêmes connaissances pour une réutilisation ultérieure.

Deuxième partie

Modélisation et développement d'un écosystème applicatif source de connaissances pour l'aide à la décision dans les organisations

« *Taxonomy is always a contentious issue because the world does not come to us in neat little packages* »
— Stephen Jay Gould, « The mismeasure of man », 1981 p.158

6. Introduction

La partie I a permis de dessiner le cadre théorique dans lequel vient se placer notre thèse. Il s'agit de pouvoir utiliser les informations circulant sur les médias sociaux d'une organisation de manière à en tirer des connaissances qui pourront être utilisées dans des processus de décision.

Cette notion de connaissance porte à la fois sur les informations échangées au sein des médias sociaux d'une organisation et également sur les informations du réseau de collaboration de l'organisation qui peuvent être tirées de l'analyse de ces médias sociaux d'un point de vue structurel.

Nos recherches ont suivi un certain cheminement qui nous a mené à une proposition originale que nous présentons dans le chapitre 7. Ce chapitre détaille notre modèle du cycle de gestion de connaissances prenant en compte à la fois les informations provenant des bases documentaires et des médias sociaux mis en place dans l'organisation. À chaque étape de ce cycle de vie des données, informations, ressources et documents correspond un ensemble de fonctionnalités à mettre en œuvre au sein d'un écosystème numérique. Cet écosystème numérique doit permettre la gestion des connaissances de l'organisation et non simplement la gestion de ses documents, comme le font déjà les systèmes présentés dans notre état de l'art.

La mise en œuvre de notre modèle théorique présenté dans le chapitre 7 passe par une modélisation ontologique que nous détaillons dans le chapitre 8. Notre modélisation s'est attachée à représenter un écosystème numérique d'applications Web au service d'une organisation. Cet écosystème lui permet de capitaliser les connaissances de ses membres à l'aide d'un référentiel unique partagé.

Les informations capitalisées au sein de cet environnement sont considérées comme autant de traces des relations sociales des membres de l'organisation et par extension de l'organisation elle-même. La base de connaissances constituée par cet environnement décrit ainsi la place que tient l'organisation vis-à-vis de l'écosystème au sein duquel elle intervient. Cette place au sein de son réseau social de collaborations peut-être calculée, en fonction des thématiques qu'elle aborde, via l'analyse des données enregistrées dans la base de connaissances. Ces nouvelles informations stratégiques sont transmises aux décideurs de l'organisation afin

d'être utilisées dans un processus de décision.

Tout ceci nous a guidé dans la réalisation du prototype d'un écosystème numérique, comportant à la fois une brique de gestion de connaissances, une brique d'analyse des données indexées dans la base de connaissances (appelée également brique inférentielle) et une brique permettant l'affichage des résultats des analyses (appelée également brique de visualisation). Nous présentons ce prototype dans le chapitre 9.

7. Modèle théorique

Les informations issues des médias sociaux prennent de plus en plus de place dans les organisations, en plus des articles scientifiques, journalistiques ou brevets, et poussent les organisations à les considérer. Nous avons cherché à montrer dans cette thèse que ces informations pouvaient servir de sources de connaissances, en vue de nourrir un processus de décision d'une organisation.

Nous avons vu dans notre état de l'art que les organisations s'outillent à l'aide de systèmes interactifs d'aide à la décision (SIAD) pour encadrer leurs processus de décision. Ces SIAD servent les décideurs de l'organisation en leur permettant d'identifier les différents choix de réponses possibles à un problème en puisant dans les bases de connaissances de l'organisation. Il ne s'agit pourtant matériellement que de bases de données stockant des éléments qui, pris séparément, sont tout à fait abscons. L'intérêt des SIAD et de leurs bases de connaissances est double.

Il s'agit d'une part de conserver l'information permettant de présenter ces données à un décideur de l'organisation de telle manière qu'il en tire la connaissance voulue par la personne ayant organisé l'enregistrement des données en question.

Il s'agit d'autre part de profiter de la manne que constitue l'ensemble des données enregistrées pour présenter des données liées ensemble de manière originale à un décideur. Ces liens entre données sont inférés à l'aide de règles renseignées par un décideur ou un expert de l'organisation et lui permettent ainsi de tirer de nouvelles connaissances des données enregistrées dans la base de connaissances.

Nous avons vu par ailleurs dans notre état de l'art que les principaux éléments constitutifs de ces bases de connaissances étaient avant tout des données documentaires ajoutées par les membres de l'organisation. Nous avons montré que la définition de document dans le monde numérique pouvait souffrir d'un positionnement relativement flou qu'il convient de fixer, du moins pour nos besoins propres. Cette définition doit, de plus, permettre de faire le lien entre la notion de document numérique et celle de connaissance.

S'il ne semble s'agir ici que d'un problème autour du sens du mot document, nous souhaitons ouvrir ce positionnement pour considérer également les productions effectuées au sein

des médias sociaux. Nous avons montré dans notre état de l'art que ces productions pouvaient être vues comme les traces d'une activité sociale. Elles semblent alors, à ce titre, entrer dans la définition de document numérique de Pédauque [2003], lorsqu'il parle du document vu comme medium.

Cependant, les médias sociaux dans leurs formes et leurs usages exposent les objets numériques considérés à une grande dynamique. Les productions des médias sociaux sont-elles alors des ressources, au sens de Lainé-Cruzel [2004] ? Et dans ce cas sont-elles incompatibles avec leur stockage au sein d'une base de connaissances, qui demanderait une stabilité propre à l'archivage ?

Nous faisons évoluer ce positionnement et en particulier nous revisitons la notion de ressource dans le monde numérique. La figure 7.1 page suivante illustre notre modèle théorique original. Pour nous, dans le monde numérique, une ressource est la matérialisation numérique d'un ensemble d'informations, à comprendre selon la définition donnée par Prax [2012]. En effet, l'utilisation d'un environnement numérique peut cacher aux utilisateurs les niveaux d'accès aux données et aux informations. À la recherche d'une information dans le système, l'utilisateur se verra proposer en réponse un ensemble de données re-contextualisées (grâce à la structure de la base de données et les liens placés entre les données y étant sauvegardées) et mises en forme (selon les spécifications implémentées dans le système utilisé). Nous appelons ces résultats des ressources. Notre définition de ressource s'éloigne de celle de Lainé-Cruzel [2004] pour revenir vers le point de vue du « document comme forme » de Pédauque [2003].

Dans l'environnement numérique, les ressources s'affichent à l'écran de l'utilisateur et n'ont aucune autre mission que de lui présenter de l'information. L'utilisateur peut alors, dans un acte volontaire d'appropriation de la ressource, la percevoir en tant que document. Ce phénomène d'appropriation se rapproche du point de vue du document comme signe de Pédauque [2003]. Notre définition procède donc à l'identification de trois étapes intermédiaires entre la notion de donnée et celle de connaissance. L'appropriation d'une connaissance par un individu se décline selon qu'il la conserve simplement en mémoire ou qu'il souhaite la repartager, enregistrant dans le SI de son organisation la ressource support de cette connaissance. Nous considérons que l'interprétation des informations mises en forme dans une ressource est nécessaire à la découverte de nouvelles connaissances et fait partie intégrante du phénomène d'appropriation de la ressource en document. Nous parlons dans ce dernier cas de documentarisation, c'est à dire de l'action selon laquelle un utilisateur ayant interprété une ressource et souhaitant la conserver sous cette forme, l'enregistre intentionnellement dans la base de connaissances de l'organisation. Ce document peut désormais être utilisé dans le SI comme n'importe quelle autre donnée.

Notre définition implique deux choses :

— La notion de document est désormais personnelle et liée à celle de connaissance. La do-

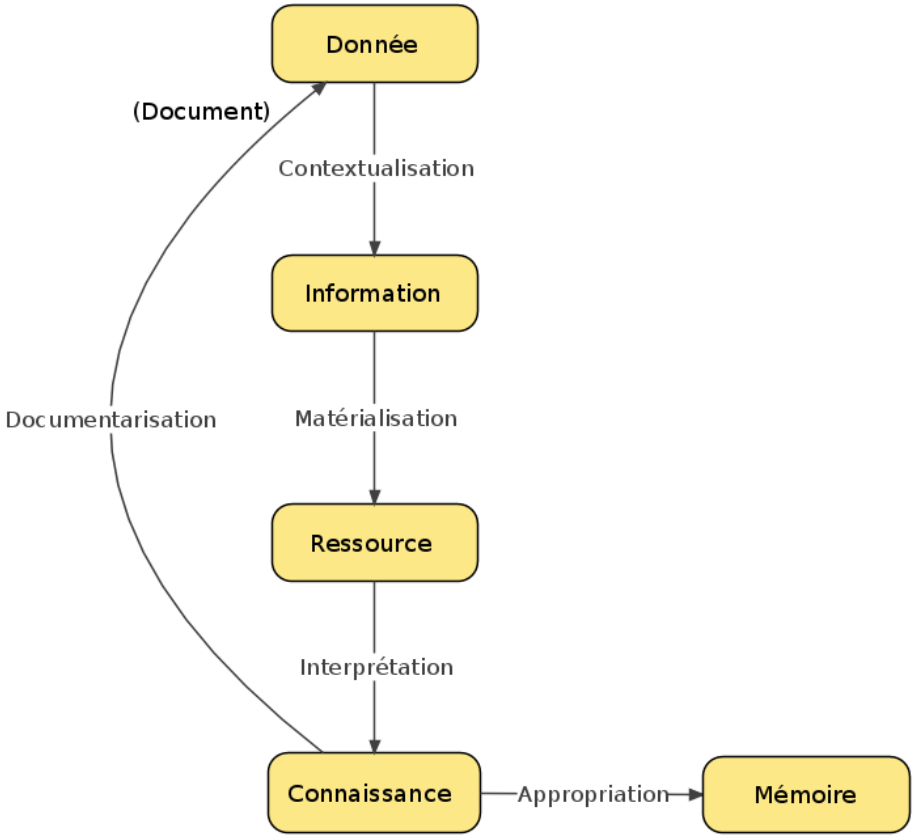


FIGURE 7.1 – Schéma de situation de la notion de ressource

cumentarisation d'une ressource est le fait d'une personne bien particulière qui en tirera de la connaissance. Deux personnes différentes n'interpréteront pas une ressource de la même façon et n'en tireront donc pas forcément la même connaissance.

— le document devient une sorte de cliché figé d'un ensemble d'informations mises en forme et qu'une personne s'est appropriées. En tant qu'objet perçu comme un document, ce dernier peut être considéré comme un élément autosuffisant et non réductible : en enlever une partie casse la perception que pouvait en avoir l'utilisateur l'ayant documentarisé. À ce titre les documents apparaissent comme un type de données possible à prendre en compte.

Le fait de considérer un document comme une simple donnée, permet de faciliter sa réutilisation ultérieure. Un utilisateur va sauvegarder un document dans la base de données de l'organisation, document qui pourra, en tant que donnée, être re-contextualisé et apparaître en tant qu'information à matérialiser au sein d'une autre ressource.

En reprenant notre scénario, nous allons étudier deux cas distincts permettant d'illustrer notre positionnement sur les ressources. Ces deux cas décrivent la manière dont ce positionnement nous autorise à considérer la manipulation de connaissances issues de données provenant d'applications sociales ou documentaires.

Prenons tout d'abord le cas de Baptiste, l'ingénieur, qui se heurte à un problème d'implémentation technique de ses travaux. Il hésite en effet, pour un nouveau projet lui ayant été affecté et touchant au domaine du Web mobile, entre le développement d'applications natives et dédiées aux différentes plates-formes mobiles ou le développement d'une application Web utilisant les dernières fonctionnalités mobiles de la spécification HTML5. Afin de confronter son point de vue à celui des autres membres de son laboratoire, il rédige un message sur le forum de son équipe et l'associe manuellement aux différents concepts « Web App », « Mobilité » et « HTML5 » pour faciliter sa découverte. La figure 7.2 illustre la forme que la création d'une telle question pourrait adopter sur un média social de type forum. Ce message est désormais une donnée parmi d'autres pour le système. Alice, après avoir consulté ce message, y répond. Cette réponse est également considérée comme une donnée par le système, indépendante du message original de Baptiste.

Figure 7.2 – Maquette graphique de la création d'une question par Baptiste

Lorsque Charles — son contrat de prestataire avec Argos l'autorisant également à bénéficier des outils du SI de l'organisation — consulte le fil de discussion associé au message de Baptiste, le système puise dans la base de données les différents messages en question, les recontextualise (les considère comme liés ensemble, associe chaque message à son auteur, à sa date de production, aux thématiques liées, etc.) et les affiche à l'écran en suivant les règles de mise en page du forum. Du point de vue de Charles, il lui est tout à fait possible de ne considérer qu'une réponse en particulier ou le message original ou l'ensemble du fil de discussion. En fonction de quoi, il va par exemple pouvoir ré-indexer un message particulier ou l'ensemble du fil de discussion sur une autre thématique que celles choisies par Baptiste. Ou bien il va pouvoir sauvegarder la discussion sous la forme d'un fichier PDF. En l'occurence, il remarque que la question de Baptiste est un altermoiement qui revient fréquement ces derniers temps et auquel, en sa qualité de conseil, il doit apporter régulièrement le même type de réponse.

Il indexe donc l'ensemble du fil de discussion avec le concept « Smartphone » pour diffuser plus largement encore le débat. Ce faisant, Charles s'approprie la ressource matérialisant la discussion, qui prend à ses yeux le statut de document. La figure 7.3 montre l'apparence que pourrait prendre le fil de discussion engendré par Baptiste au sein d'un média social de type forum. La figure 7.3 affiche à l'aide de la petite bulle marquée « 1 » l'emplacement visuel du nouveau concept choisi par Charles pour indexer le fil (et non les messages individuels) de discussion.

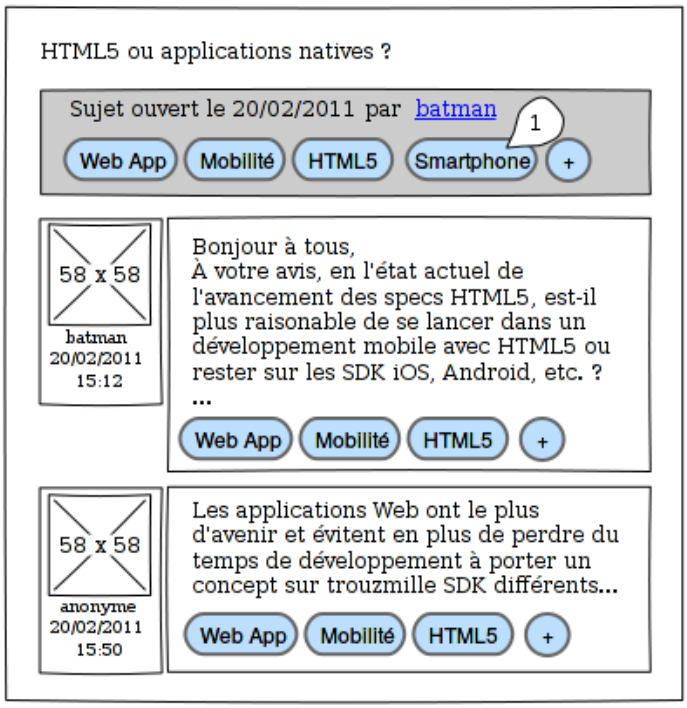


FIGURE 7.3 – Maquette graphique d'un fil de message sur un forum

Prenons ensuite le cas d'Alice et Baptiste s'échangeant différentes ressources au sein des différentes applications Web déployées par l'organisation Argos. Ces activités génèrent de nombreuses données qui sont enregistrées dans les bases de données d'Argos — dates de connexion au système, textes tapés, images téléversées, liens visités, etc. L'ensemble de ces données est disponible pour les décideurs d'Argos, comme Denise, via une application de supervision. Cette application va recueillir des informations dans la base de données, qu'elle présente ensuite au décideur au sein de cadrans sous la forme de tableaux ou graphes. Le contenu de ces cadrans dépend tout à fait des requêtes effectuées par le décideur. Les ressources générées (les graphes ou les tableaux, qui ne sont que des matérialisations des informations issues de la base de données) peuvent alors être considérées comme un document à part entière, ou assemblées au sein d'un document plus important en tant que compte-rendu d'activité. Ce dernier pourra alors être sauvegardé dans la base de données en vue d'une réutilisation ultérieure.

Ainsi, l'émergence de connaissances est due à un processus de documentarisation de ressources générées par le système d'information d'une organisation. Au cours de ce processus, les membres de l'organisation vont s'approprier les ressources que le système leur affiche à l'écran. Cette appropriation des ressources les conduit à interpréter les informations constitutives de ces mêmes ressources, aidant les membres de l'organisation à se créer de nouvelles connaissances.

Comme nous l'avons vu, le système lui-même re-contextualise les données qu'il présentera à l'utilisateur de deux manières différentes :

- À partir des informations fournies par les utilisateurs du système. Ces informations sont alors enregistrées sous la forme de données et de liens entre ces données qui permettront plus tard au système de reconstituer l'information en question. Le fait qu'un utilisateur indexe une ressource avec un concept de la base de connaissances du système permet :
  - d'aider cet utilisateur à s'approprier cette ressource. Il en tire du même coup une certaine connaissance qui lui est propre.
  - d'ajouter une nouvelle information au système sous la forme du lien unissant désormais les données impliquées dans la ressource en question et le concept choisi par l'utilisateur.
- À partir des informations inférées par le système à partir des données emmagazinées dans sa base de connaissances. Tout l'intérêt des organisations réside dans ces inférences qui vont leur apporter de nouvelles informations et donc, après appropriation par les utilisateurs, de nouvelles connaissances. Cet intérêt est particulièrement vrai en ce qui concerne les SIAD intégrés aux SI des organisations afin de soutenir les décideurs de l'organisation.

L'analyse des données, leur nettoyage, leur croisement est un processus qui peut-être outillé et partagé entre différents décideurs, d'autant que l'analyse des médias sociaux est appelée à devoir manipuler de grands jeux de données. Ce travail étant effectué directement sur la base de connaissance du système, il semble ainsi plus logique de positionner cette étape dans le domaine informationnel des SIAD, tels que nous les avons définis dans notre état de l'art. Le système va alors générer un certain nombre de ressources constituées à partir des informations issues de la base de données et va les présenter à un décideur. La tâche principale du décideur serait donc l'interprétation des différentes solutions proposées sous la forme de ressources par le SIAD, ce qui lui apporterait les connaissances nécessaires qui, une fois jumelées à son expérience personnelle, appuieront son choix au travers de son intention.

Nous amendons ainsi selon le schéma de la figure 7.4 le schéma des processus internes à un SIAD de Knight (2002), visible p. 62 dans la figure 4.4. Nos modifications ont consisté à descendre la boîte notée (1) du domaine cognitif, personnel au décideur, au domaine informa-

tiel du fait que cette tâche peut être confiée directement à un système informatique. Si cette tâche n'est plus directement confiée au décideur, une étape supplémentaire, notée (2) dans la figure 7.4 est nécessaire pour permettre au décideur d'interpréter les ressources présentées par le système. L'interprétation des ressources, également nourrie des intentions du décideur, l'aide à se créer ou mobiliser des connaissances sur lesquelles il appuiera sa résolution.

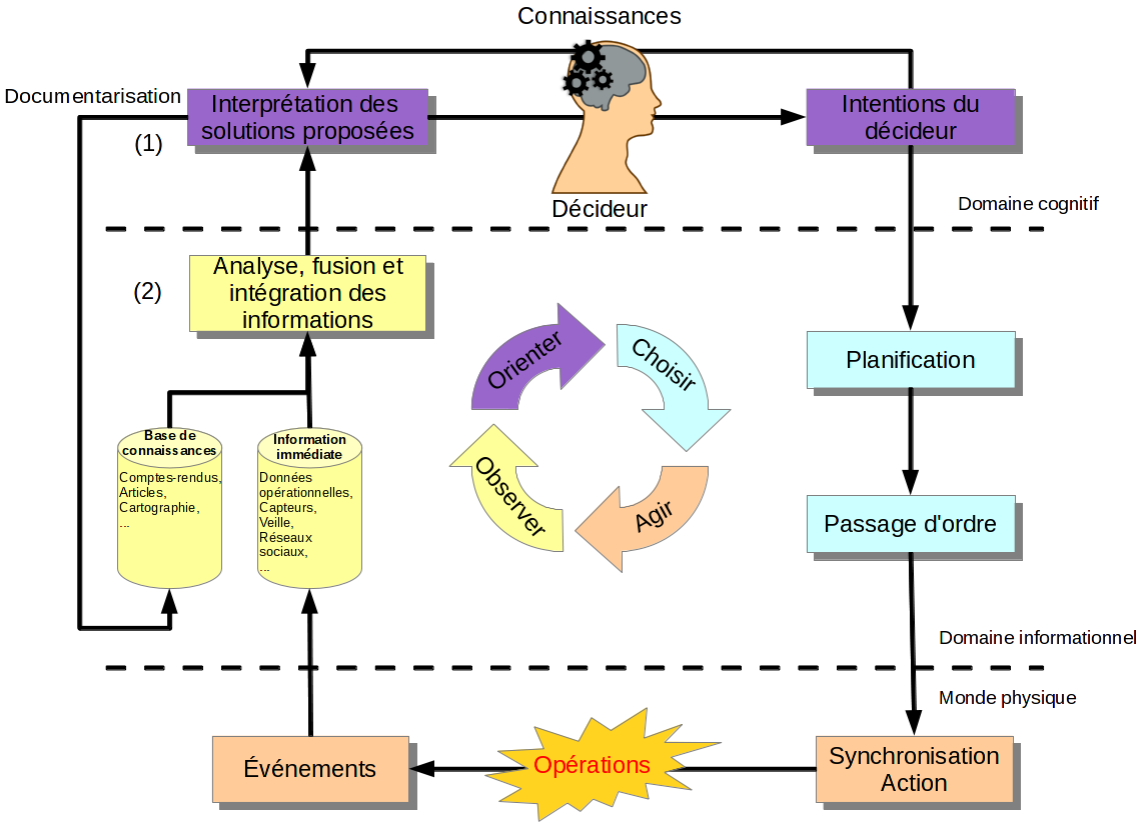


FIGURE 7.4 – Réinterprétation du fonctionnement d'un SIAD

Afin d'avoir des résultats d'analyse de données pertinents, il est nécessaire d'organiser la capitalisation des connaissances à l'échelle de l'organisation complète. Cette capitalisation requiert l'utilisation d'un référentiel unique lors de la phase d'indexation des ressources produites ou échangées par les membres de l'organisation. Ainsi, quels que soient l'application d'origine ou le service d'origine de l'utilisateur du système, la ressource documentarisée sera plus facilement accessible, partageable et le système pourra utiliser des données provenant de deux services différents dans ses analyses. Il devient alors plus facile d'obtenir une vision générale des communautés œuvrant au sein de l'organisation au travers des informations échangées et capitalisées dans la base de connaissances de l'organisation.

8. Modélisation ontologique d'un écosystème numérique

Dans ce chapitre, nous développons notre modélisation d'un écosystème numérique dont le but est de venir soutenir les processus de décision d'une organisation via une bonne gestion de ses connaissances. Comme nous l'avons statué, cet écosystème doit à la fois :

- Permettre la production ou l'ajout de ressources par ses utilisateurs ;
- Permettre la modification ou la lecture de ces ressources ;
- Permettre l'indexation de ces ressources à l'aide d'un référentiel unique et partagé afin d'autoriser leur réutilisation.

Ces deux derniers points sont importants puisqu'ils permettent l'appropriation des ressources par les utilisateurs de l'écosystème et ainsi la création de connaissances.

Les trois nécessités précédentes nous conduisent à devoir modéliser :

- Différentes applications Web et les ressources pouvant y être produites, ajoutées, échangées, qu'elles soient issues de l'application Web elle-même ou qu'il s'agisse de documents bureautiques importés au sein de l'environnement (section 8.1) ;
- Les utilisateurs de ces applications et la manières dont ils vont pouvoir s'organiser au sein de l'environnement et ce en lien, ou non, avec leur structuration réelle au sein de l'organisation (section 8.2) ;
- Le mécanisme d'indexation des ressources produites, ajoutées et échangées au sein de l'écosystème à l'aide d'un référentiel unique et partagé (section 8.3).

Dans le cadre de nos travaux, nous nous intéressons à une forme particulière de réutilisation des ressources documentarisées et ajoutées à la base de connaissances. Comme nous l'avons vu, les documents numériques peuvent être vus comme des traces sociales. Ces traces permettent de dessiner le réseau de collaborations de l'organisation les ayant collectées. Ainsi, nous nous sommes penché sur la modélisation d'un tel réseau de collaborations (section 8.4).

Nous avons cherché, pour la réalisation de notre modèle, à réutiliser au mieux les travaux existants du domaine et l'avons donc conçu comme un alignement d'ontologies issues des travaux du Web sémantique. Nos travaux se sont effectués dans la continuité du projet

MEMORAe dont l'approche a été présentée dans la section 2.3.9. Celle-ci propose un modèle théorique et une plate-forme de collaboration et de gestion de connaissances pouvant être utilisés par des organisations.

Nos travaux étendent la précédente ontologie développée dans le cadre du projet MEMORAe. Notre évolution du modèle a principalement consisté à faire correspondre les concepts en place aux propositions du Web sémantique et à l'étendre afin de prendre en compte d'autres catégories de médias sociaux. Ces standards du Web sémantique étudiés sont des ontologies construites à l'aide du langage OWL. Notre propre modélisation a donc également pris la forme d'une ontologie écrite en OWL.

La précédente ontologie de l'approche MEMORAe [Ghebghoub *et al.*, 2010] ayant pour nom memorae-core, c'est naturellement que nous avons décidé de nommer cette évolution memorae-core 2 [1]. Dans la suite du mémoire et à des fins de facilité de lecture, nous utiliserons le *namespace* `mc2` pour identifier les classes et relations de ce nouveau modèle ontologique. La figure 8.1 donne une vision graphique globale de l'ontologie ayant été réalisée dans le cadre de cette thèse et dont nous allons présenter les différentes composantes dans les sections suivantes.

8.1 Modélisation des documents numériques

Comme nous l'avons vu dans l'état de l'art, il existe déjà un certain nombre d'écosystèmes numériques regroupant diverses applications Web. Ceux-ci nous permettent d'identifier les applications Web les plus souvent proposées aux utilisateurs. Nous avons souhaité pouvoir traiter un maximum de ces fonctionnalités. Ainsi, en plus des fonctionnalités d'import de documents bureautiques numériques, nous avons modélisé les médias sociaux suivants : blogs, minimessages, wikis et forums.

L'étude des différentes ontologies présentées dans la section 3.4 nous a permis de constater qu'il n'existait pas à l'heure actuelle de modèle permettant de décrire un écosystème numérique dans sa globalité en apportant un degré de précision satisfaisant. Cependant, chacune des ontologies présentées nous apporte un fragment de solution.

L'ontologie la plus complète que nous avons étudié est l'ontologie issue du projet VIVO (section 3.4.5). Cette dernière est elle-même déjà un alignement manuel d'autres ontologies comme FoaF ou BIBO. Cet alignement, complété de ses propres classes et relations, permet de représenter de nombreuses sortes de ressources documentaires (articles scientifiques, rapports, brevets, etc.). L'ontologie VIVO présente cependant un manque important à nos yeux.

1. L'ontologie complète a été publiée sous la forme d'un fichier turtle à l'adresse suivante : http://www.hds.utc.fr/memorae/core2.ttl (accessible le 10 novembre 2013)

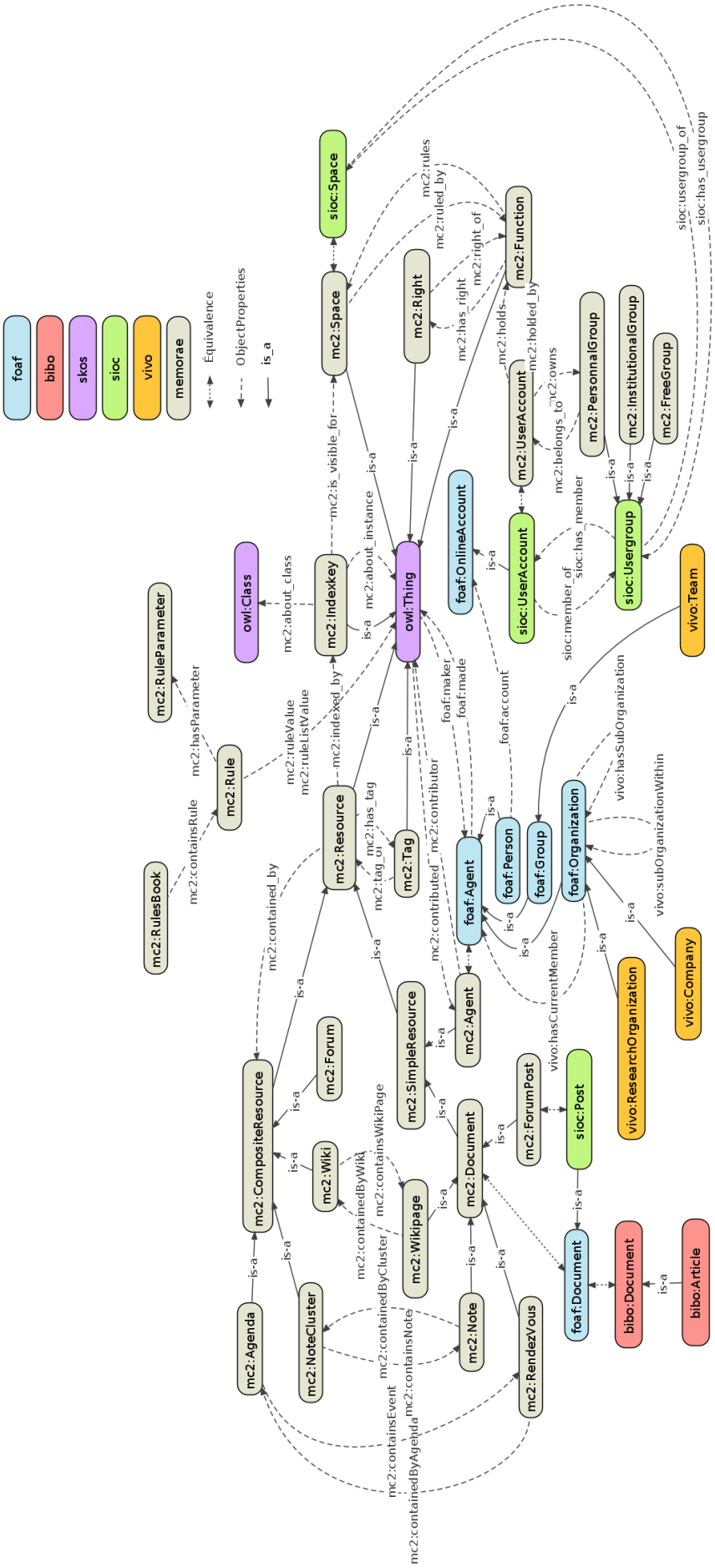


FIGURE 8.1 – Représentation graphique de l'ontologie memorae-core 2 réalisée

Elle ne propose pas de solution de représentation des ressources pouvant être produites au sein des médias sociaux. Nous reprenons leur modélisation pour l'intégrer à notre propre vision en la complétant.

En positionnant les ressources comme la matérialisation numérique de données recontextualisées, nous posons la question de la forme de cette matérialisation. En fonction du type initial de données (scalaire, chaîne de caractère, image, etc. ou même document) et des conditions de leur contextualisation — rappel de l'horodatage d'une donnée, reconstruction de liens sémantiques entre données (auteur, géolocalisation, etc.) ou de liens structurels (ordre d'apparition de données dans une séquence) —, les ressources peuvent prendre différentes formes. Ces formes sont variées et vont de la retranscription brute des informations (une image suivie du nom de son auteur, un texte affiché à l'écran, etc.) à l'élaboration d'éléments graphiques complexes (diagrammes agrégeant des données, *mashups*, etc.). Nous pouvons ainsi classer les différentes formes de ressources selon deux familles distinctes :

Les ressources simples qui permettent de matérialiser une information atomique. Cette dernière est le résultat de connexions entre différentes données.

Les ressources composées qui permettent de matérialiser en un seul résultat un ensemble d'informations différentes.

Pour illustrer ces définitions, nous pouvons prendre l'exemple d'un forum. Les messages s'échangeant au sein de ce type d'application Web sont des ressources simples. Chaque message est composé de différents éléments comme du texte, le nom de l'auteur, la date de publication et éventuellement des images. Chacun de ces éléments est une donnée qui, prise séparément, n'a pas énormément de sens. Leur réunion dans le contexte d'un message du forum apporte une information. À l'inverse, les fils de messages sont des ressources composées. En plus des données comme le titre général du fil de message, les thématiques globales associées à ce fil, le statut (ouvert ou fermé) du fil, ce dernier agrège différents messages qui peuvent être considérés indépendamment comme des ressources.

Nous avons donc décidé de différencier dans notre modélisation les ressources dites simples des ressources composées. Nous avons créé une classe `mc2:Resource` que nous avons spécialisée en `mc2:SimpleResource` et `mc2:CompositeResource`.

La notion de ressource composite est absente de VIVO. Le point d'ancrage de cette ontologie va donc se situer avec la classe `mc2:SimpleResource`. VIVO utilise directement la hiérarchie de classe de l'ontologie BIBO pour décrire les ressources de leur plate-forme. Notre classe `mc2:Document`, spécialisation de `mc2:SimpleResource` est ainsi déclarée équivalente à `bibo:Document` (classe principale de l'ontologie BIBO) et par jeu d'équivalence à `foaf:Document`.

Cette équivalence nous permet d'accéder à l'ensemble des classes de l'ontologie BIBO, comme des spécialisations de notre classe `mc2:SimpleResource`. Citons à titre

d'exemple les classe `bibo:Book`, `bibo:AcademicArticle`, `bibo:Patent`, `bibo:Thesis` ou encore `bibo:Slideshow`.

Cette liaison ne nous permet cependant pas, comme nous l'avons dit, de prendre en compte les ressources produites au sein de medias sociaux. Pour ce faire, nous avons créé de nouvelles classes en spécialisant `mc2:SimpleResource`. Pour ce faire nous nous sommes appuyé cette fois-ci sur l'ontologie SIOC et plus particulièrement le module type [2]. Nous avons ainsi mis en place les équivalences suivantes :

- `mc2:BlogPost` pour représenter les articles publiés sur des blogs, équivalente à la classe `sioc:BlogPost` ;
- `mc2:WikiPage` pour représenter les pages de Wiki, équivalente à la classe `sioc:WikiArticle` ;
- `mc2:ForumPost` pour représenter les messages sur les forums, équivalente à la classe `sioc:BoardPost`.

Enfin, nous avons créé de toute pièce la classe `mc2:ShortMessage` pour représenter les minimessages issus de plates-formes comme Twitter. Cette dernière n'est en effet pas présente dans la spécification du module type de SIOC.

La plupart des ressources des médias sociaux s'intègrent au sein d'applications plus larges, que nous avons définies comme des spécialisations de `mc2:CompositeResource` et là aussi notées équivalentes à leur homologues SIOC :

- `mc2:Blog` équivalente à `sioc:Weblog` ;
- `mc2:Forum` équivalente à `sioc:MessageBoard` ;
- `mc2:Wiki` équivalente à `sioc:Wiki`.

Le fonctionnement des applications Web que sont les médias sociaux repose sur la production de contenus par leurs utilisateurs. Notre modélisation ne pourrait être complète sans avoir correctement décrit ces utilisateurs. Nous présentons dans la section suivante notre modélisation des utilisateurs et leur organisation au sein d'un environnement numérique.

8.2 Modélisation des utilisateurs de l'environnement

Le statut d'écosystème provient de l'existence d'interdépendances entre ses différentes applications. La dépendance la plus évidente correspond à l'utilisation d'une méthode unique d'authentification sur les différentes applications. L'authentification unique, que l'on retrouve d'ailleurs au sein de la plupart des écosystèmes numériques existants, permet entre autres l'existence de fonctionnalités transverses comme la possibilité d'ajouter des commentaires de manière uniforme sur les différentes applications de l'écosystème.

2. http://sioc-project.org/ontology#sec-modules-types (accessible le 10 novembre 2013)

Le principe d'authentification repose sur la notion d'identifiant permettant de repérer de façon certaine les actions d'une personne au sein d'un environnement numérique. Ce qui permet entre autres de pouvoir lui attribuer ses productions ou encore personnaliser son expérience utilisateur. Si le système mémorise les actions d'un utilisateur numérique, c'est bien à une personne réelle que nous souhaitons les attribuer, de manière à pouvoir associer les comportements sociaux numériques à des communautés bien réelles au sein de l'organisation. Notre modélisation doit donc prendre en compte ces deux nuances à savoir :

— décrire l'identité numérique d'une personne, appelée également compte d'utilisateur ;
— décrire l'identité physique d'une personne, associée à un compte d'utilisateur ou non (dans le cas d'auteurs de documents ajoutés au système par exemple).

Comme nous l'avons vu, la principale ontologie permettant de décrire des personnes sur le Web sémantique est l'ontologie FoaF (section 3.4.1). Cette dernière fournit à la fois la classe `foaf:Person` qui décrit des êtres humains physiques, mais également la classe `foaf:OnlineAccount` qui décrit le compte d'utilisateur d'une personne. Des instances de ces deux classes peuvent être liées à l'aide de l'*object property* `foaf:account` qui associe une personne à son ou ses comptes utilisateur personnels.

Cette fois-ci, ce sont les lacunes dans les propriétés permettant de décrire précisément un compte d'utilisateur qui nous ont poussé à créer notre propre classe `mc2:UserAccount`. Cette classe est équivalente à la classe `foaf:OnlineAccount`. Nous lui avons défini en particulier la *data property* `mc2:password_hash` permettant d'affecter un mot de passe à un utilisateur.

Au sein des médias sociaux, il est fréquent que les utilisateurs puissent se regrouper en communautés sur des sujets variés. Nous n'avons pas réutilisé OCSO pour décrire ces groupes d'intérêt car nous souhaitions apporter quelques précisions à cette notion de groupe. En particulier, notre écosystème numérique étant conçu pour les organisations, nous voulions pouvoir modéliser à la fois des communautés de pratiques informelles, mais également des regroupements d'utilisateurs imposés par une hiérarchie, tels que définis dans la section 2.2. Les ontologies FoaF, SIOC ou VIVO ne permettent pas de représenter les groupes formels et informels d'une organisation au sein d'un environnement numérique.

Pour cela nous spécialisons la classe `sioc:Usergroup` à l'aide des deux sous-classes suivantes :

— `mc2:Institutionalgroup` permet de décrire les groupes institutionnels. Un groupe institutionnel est créé par une entité hiérarchique et regroupe différents collaborateurs autour d'une mission déterminée. Il s'agit typiquement des équipes de travail au sein des organisations.
— `mc2:Freegroup` permet de décrire les groupes spontanés. Un groupe dynamique ou spontané est créé à l'instigation d'un ou plusieurs membres de l'organisation sans validation hiérarchique. Il peut s'agir de communautés de pratique Wenger [1998] ou épis-

témiques [Gueye, 2004] autour d'un sujet précis, d'une personne ou des deux. L'ouverture d'un tel groupe aux autres membres de l'organisation peut être restreinte, dans le cadre de groupes privés.

La participation d'un utilisateur au sein d'une communauté s'effectue à l'aide des *objects properties* `sioc:member_of` et `sioc:has_member` qui permettent de lier des instances de `sioc:Usergroup` et `sioc:UserAccount`. SIOC ayant défini `sioc:UserAccount` comme une spécialisation de `foaf:OnlineAccount`, afin que notre modèle reste valide nous avons décidé de rendre notre classe `mc2:UserAccount` équivalente à la classe `sioc:UserAccount`.

Le listing en turtle ci-après est un exemple d'instanciation de notre modèle. Puisque notre modèle est fondé sur quelques ontologies présentées dans la section 3.4, cet exemple comporte de nombreuses similarités avec les exemples déjà montrés dans cette section. Ce listing et les suivants utilisent des espaces de noms déclarés selon le listing disponible dans la section B.1 en annexe.

Nous détaillons dans cet exemple de quelle manière Baptiste, l'ingénieur de notre scénario, est décrit, ainsi que son compte d'utilisateur. Cet exemple montre également la modélisation du groupe numérique d'utilisateurs correspondant à l'unité de R&D au sein de laquelle Baptiste travaille.

```
1  :baptiste a foaf:Person;
2      foaf:lastName "Atman";
3      foaf:firstName "Baptiste";
4      foaf:gender "male";
5      foaf:mbox <batman@orga.fr>.
6
7  :baptiste_account a mc2:UserAccount;
8      mc2:password_hash "R2HDq2xsZSBCLiBqZSB0J2FpbWUgOik=";
9      sioc:id "baptiste".
```

Nous avons déclaré `mc2:UserAccount` équivalent à `sioc:UserAccount` ce qui nous permet d'utiliser les propriétés `sioc:id` et `sioc:member_of`.

`mc2:UserAccount` est une spécialisation de `foaf:OnlineAccount` ce qui permet d'utiliser la propriété `foaf:account` entre une `foaf:Person` et un `foaf:UserAccount`.

```
1  :baptiste foaf:account :baptiste_account.
```

`mc2:InstitutionalGroup` est une sous-classe de `sioc:UserGroup` ce qui permet l'utilisation des propriétés `sioc:has_member`, `sioc:usergroup_of` et `sioc:has_usergroup`.

```
:retdpeople a mc2:InstitutionalGroup;
    rdfs:label "Groupe des membres de l'unité de R&D";
    sioc:has_member :baptiste_account.

:baptiste_account sioc:member_of :retdpeople.
```

Les actions effectuées par les utilisateurs permettent de révéler les communautés de l'organisation au travers de leur contexte thématique. Plus exactement, une communauté de pratique sur un thème donné pourra être identifiée par le fait que ses membres partagent et produisent des ressources sur ce thème. Au sein d'une organisation composée de nombreux services plus ou moins connectés entre eux, il est nécessaire, pour rendre lisible la cartographie de ses communautés constitutives, d'utiliser un référentiel unique entre toute ces composantes. Ce référentiel sera utilisé pour indexer les ressources échangées ou produites par les membres de l'organisation. Nous décrivons dans la section suivante la modélisation du mécanisme d'indexation défini dans notre ontologie.

8.3 Modélisation du mécanisme d'indexation

La fonctionnalité centrale de notre écosystème est la mise en place d'un référentiel unique partagé par les différentes applications de l'écosystème. Ce référentiel prend la forme d'une ontologie de domaine dont les instances permettent l'indexation des différentes ressources produites ou partagées au sein de l'écosystème.

8.3.1 Définition du référentiel partagé

Comme nous l'avons vu dans état de l'art, la plupart des écosystèmes numériques actuels ne proposent, au mieux, qu'un système à base de *tags* personnel à chaque utilisateur pour indexer les contenus produits ou ajoutés dans l'écosystème. Ces *folksonomies* sont souvent limitées à une application particulière de l'écosystème. Les différentes ressources des différentes applications ne sont ainsi pas facilement accessibles de manière globale par un système d'indexation partagée.

Par ailleurs, les folksonomies simples ne permettent pas de mettre en relation les différents mots-clés utilisés pour décrire les documents de l'écosystème. C'est pourquoi l'approche MEMORAe propose à l'organisation souhaitant déployer un écosystème numérique de connaissances de préalablement définir une ontologie de son domaine d'activité. C'est cette ontologie qui servira de référentiel partagé avec lequel les différents documents de l'écosystème seront indexés. L'utilisation d'une ontologie partagée permet en effet de répondre aux deux problèmes évoqués précédemment :

— Les ressources de l'écosystème sont accessibles de manière globale et décentralisée via leur indexation avec l'ontologie, quelle que soit l'application de départ ayant servi à les générer.
— Les mots-clés utilisés pour indexer les documents dans l'écosystème peuvent être mis en relation (que ce soit en les spécialisant ou en les associant sémantiquement) et définis de manière précise. Nous parlons alors de concepts.

L'intégration de l'ontologie de domaine d'une organisation au sein du modèle memorae-core 2 s'effectue en transposant la racine de cette ontologie de domaine (la classe `owl:Thing`) sur la classe `skos:Concept` de manière à ce que toutes les classes de l'ontologie deviennent *de facto* des spécialisations de la classe `skos:Concept`. Cette spécialisation autorise une approche simple et large de notre principe de référentiel partagé :

— Lorsque le référentiel partagé est une ontologie exprimée en OWL, sa transposition sous la classe `skos:Concept` nous permet de faire hériter les instances de cette ontologie de propriétés intéressantes, comme les propriétés `skos:prefLabel` et `skos:altLabel`, qui permettent d'exprimer les différents termes nommant un concept, éventuellement en différentes langues. Pour exploiter ce référentiel, en plus des instances classiques de l'ontologie nous instancions chacune des classes en un concept générique dont l'*Uniform Resource Identifier* (URI) est forgée à partir de l'URI de la classe mère à laquelle nous adjoignons le suffixe `_general`. Cette instance particulière permet aux utilisateurs de l'écosystème d'indexer un document, non pas sur une instance précise de l'ontologie, mais directement par la notion portée par la classe elle-même. Par exemple, l'ontologie de domaine de l'organisation Argos définit la classe `MessageBoard` qui décrit tous les systèmes de publication de messages sur le Web. Elle se spécialise en `Forum` ou `Wall`. La classe `MessageBoard` est instanciée en `PHPBB`, `SimpleMachineForum` ou encore `VanillaBoard` qui sont des exemples de systèmes de publication de messages. Les instances `MB_RetD` ou `MB_Compta` existent également pour décrire les systèmes de messages déployés respectivement pour le département de R&D et la comptabilité. Enfin, l'instance `MessageBoard_general` permet de décrire tout système de publication de messages.
— Il se peut que la définition d'une ontologie formelle ne soit pas possible ou coûte trop cher pour une organisation. Dans ce cas, l'utilisation de SKOS est également justifiée. SKOS est construite à l'aide du modèle de données *Resource Description Framework* (RDF), tout comme OWL, ce qui permet à des taxonomies SKOS de s'interfacer à des ontologies OWL, comme notre modèle memorae-core 2. Ainsi, la personne en charge de la définition du référentiel partagé peut se contenter d'établir une taxonomie de termes, SKOS lui permettant tout de même d'exprimer des relations de spécialisation (`skos:broader` et `skos:narrower`) ou d'association (`skos:related`) entre les concepts du référentiel. SKOS nous permet ici d'apporter une profondeur sémantique suffisante

au référentiel partagé, en comparaison de l'usage de simples mots-clés ou *tags*.

Dans la pratique, la mise en œuvre du référentiel partagé au sein de notre modèle passe donc par la manipulation d'instances SKOS, quel que soit le degré de formalisation choisi par l'organisation. Nous présentons dans la section suivante les classes que nous avons définies pour permettre l'utilisation d'un référentiel partagé tel que nous l'avons défini ici.

8.3.2 Indexation des documents

Dans notre approche, le référentiel partagé permettant d'indexer les documents est fondé sur la spécialisation ou l'instanciation de la classe `skos:Concept`. Il nous a donc semblé pertinent de concevoir une nouvelle classe `mc2:IndexKey`. En effet, les propriétés héritées des autres modèles que nous utilisons qui auraient pu nous servir ne possédaient pas le bon domaine ou la bonne cible d'application. La relation `foaf:topic` par exemple ne s'applique qu'à des `foaf:Document` alors que nous cherchons à indexer tout type de `mc2:Resource`. Ainsi, notre classe `mc2:IndexKey` associe une `mc2:Resource` à une instance de type `skos:Concept`.

L'utilisation d'une classe particulière pour mettre en place une relation sémantique entre un objet à indexer et un concept est rendue nécessaire, dans notre cadre organisationnel, par le besoin de n'effectuer une indexation que dans un cadre spécifique. Ainsi, notre modèle d'indexation ne définit pas une simple relation binaire, mais une relation ternaire. Ce cadre spécifique d'indexation exprime le fait que l'indexation d'un document s'effectue exclusivement pour un espace partagé appartenant à un groupe d'utilisateur particulier du système.

Cette notion d'espace partagé est importante dans le domaine des organisations où le problème de la sensibilité des documents est prégnante. Les espaces de partages nous permettent de cloisonner les productions entre communautés au sein d'une organisation, tout en les laissant bénéficier du même référentiel partagé pour indexer ces mêmes productions. Dans notre modélisation, chaque `sioc:Usergroup` se voit ainsi affecter un unique espace de partage décrit à l'aide de la classe `mc2:SharingSpace`. Cet espace unique permet aux membres du groupe de s'échanger des ressources entre eux, sans que ces dernières soient visibles des autres utilisateurs de la plate-forme. La classe `mc2:SharingSpace` est définie comme étant une spécialisation de la classe `sioc:Space`. Ceci nous permet d'associer chaque espace de partage à un groupe d'utilisateur à l'aide des propriétés inverses `sioc:usergroup_of` et `sioc:has_usergroup`.

Nous présentons ci-après un exemple de ce mécanisme d'indexation, suite de l'exemple de la section précédente. Baptiste souhaite désormais capitaliser un article scientifique obtenu à la suite d'une recherche sur internet. L'article a pour titre « Web Squared : Web 2.0 Five Years On » et traite du Web 2.0, de l'internet des objets ou encore des données ouvertes. Il se trouve que ces trois sujets sont déjà définis dans le référentiel partagé de l'organisation de Baptiste. Ce référentiel partagé a été élaboré sous la forme d'une ontologie de domaine dont l'espace

de nom est noté `argos:`. Le listing de la section B.2 dans les annexes montre un extrait de ce référentiel et présente l'organisation de ces trois concepts.

Le listing ci-après illustre pour sa part l'indexation de l'article selon les trois concepts choisis par Baptiste. Comme nous l'avons vu, Baptiste fait partie de l'unité de R&D de son organisation, représentée par le groupe numérique d'utilisateurs `:retdpeople`. L'exemple suivant illustre donc la manière dont l'indexation va s'effectuer, pour les trois concepts choisis, au sein de l'espace de partage correspondant à ce groupe d'utilisateurs.

```
:oreilly a foaf:Person;
    foaf:firstName "Tim";
    foaf:lastName "O'Reilly".

:battelle a foaf:Person;
    foaf:firstName "John";
    foaf:lastName "Battelle".

:oreilly2009 a bibo:AcademicArticle;
    dc:title "Web Squared: Web 2.0 Five Years On"@en;
    foaf:maker :oreilly;
    foaf:maker :battelle.

:retd_space a mc2:Space;
    sioc:has_usergroup :retdpeople.

:retdpeople sioc:usergroup_of :retd_space.

:idk42 a mc2:IndexKey;
    mc2:about_instance :web2_general;
    mc2:about_instance :wot_general;
    mc2:about_instance :opendata_general;
    mc2:is_visible_for :retd_space;
    mc2:index :oreilly2009.
```

Cet exemple permet également de bien noter la nuance importante entre les utilisateurs de l'écosystème (`mc2:UserAccount`), qui ne sont en fait que des identifiants numériques et les personnes réelles (`foaf:Person`) qui peuvent, au besoin, posséder différents comptes d'utilisateur ou pas du tout. Ce sont bien les `foaf:Person` qui sont notées comme auteurs/créateurs des ressources circulant au sein de l'écosystème numérique. Plus exactement, la relation `foaf:maker` a pour *range* la classe `foaf:Agent`, une personne morale (`foaf:Organization` ou `foaf:Group`) pouvant tout à fait être considérée comme l'auteur d'une ressource.

Si l'indexation des ressources partagées ou produites par les utilisateurs d'un écosystème numérique permet d'identifier des communautés au sein d'une organisation, il convient de proposer également une définition sémantique de ces communautés. Nous apportons notre définition dans la section suivante.

8.4 Modélisation d'un réseau d'organisations

Comme nous l'avons vu dans la section 2.2, les écosystèmes de connaissances peuvent être vus comme des réseaux d'organisations partageant le but commun d'améliorer ensemble leurs connaissances d'un domaine en mutualisant leurs compétences.

Si un écosystème numérique, tel que décrit précédemment, peut servir à gérer les connaissances et plus exactement les ressources échangées dans le cadre d'un écosystème de connaissances, il convient d'apporter une attention particulière à la modélisation du réseau d'organisations.

Ce dernier est non seulement constitué de liens de collaboration entre différentes organisations, mais également, au sein de ces organisations, de liens entre leurs différentes composantes, telles que nous les avons définies dans la section 2.2.

Toute la difficulté de la démarche provient de la maintenance de la cohérence de l'ensemble des affiliations capitalisées dans la base de connaissances. Il s'agit en effet de pouvoir détecter deux affiliations composées d'éléments communs (deux laboratoires au sein d'un même service), sans pour autant mélanger des entités comportant des éléments aux noms similaires, bien que différents dans la pratique (les « services informatiques » de différentes entreprises ne sont bien sûr pas liés).

De plus, il n'existe pas de solution unique et efficace permettant de modéliser une chaîne d'affiliation. Ces dernières peuvent contenir un nombre variable d'éléments (un seul dans le cadre des PME à une dizaine dans le cadre de grosses multinationales) et il n'existe pas de règle sémantique fixe sur la manière de nommer ces différents éléments. Certaines organisations sont structurées en services, découpés en départements contenant diverses équipes, tandis que d'autres, au contraire, sont structurées en départements regroupant des services.

Ces règles de nommage font partie intégrante de l'identité de l'organisation. Quand bien même, dans le cadre de notre base de connaissances, nous utiliserions une ontologie formelle pour désigner les différents éléments des affiliations enregistrées en base, il nous faudrait conserver l'information supplémentaire liée aux habitudes locales de nommage des éléments. De cette seule manière les utilisateurs retrouveraient bien les qualificatifs utilisés lors de l'enregistrement de l'affiliation.

Pour toutes ces raisons, nous considérons comme nécessaire, dans la modélisation du réseau de collaborations d'un écosystème de connaissances, d'enregistrer l'affiliation complète des entités de deux manière distinctes :

1. d'une part la vision que peut en avoir un utilisateur de la base de connaissances, c'est-à-dire les mots qu'il utilisera préférentiellement pour désigner tel ou tel élément d'une affiliation ;
2. d'autre part à l'aide d'un typage défini strictement au préalable, ce qui nous permettra d'intégrer les entités au sein de la base de connaissances, comme n'importe quelle ressource.

Le deuxième point revient à utiliser une modélisation ontologique des organisations. De nombreux projets d'ontologies [Fox, 1992, Uschold *et al.*, 1996, Reynolds, 2010a] se sont intéressés à la modélisation des organisations. Reynolds [2010b] propose un panorama relativement complet, agrémenté d'exemples d'implémentations. Ces ontologies s'intéressent à la nature des organisations et à leur fonctionnement en définissant par exemple des modèles d'activité. Cependant leur structure interne n'est pas aussi explicitée que notre besoin le nécessite.

Le projet VIVO, déjà présenté dans la section 2.3.8, propose une autre modélisation offrant une plus grande liberté de nommage des composantes internes des organisations. Incluse directement dans l'ontologie VIVO présentée dans la section 3.4.5, cette modélisation spécialise l'ontologie FoaF et la classe `foaf:Organization` en proposant par exemple les classes `vivo:University`, `vivo:AcademicDepartment`, `vivo:PrivateCompany` ou encore la classe `vivo:Team` comme spécialisation de la classe `foaf:Group`.

Pour représenter les liens d'affiliation d'une entité, l'ontologie VIVO propose les *object properties* `vivo:hasSubOrganization` et `vivo:subOrganizationWithin`. Ces deux propriétés ne peuvent cependant s'utiliser que sur des `foaf:Organization`. L'affiliation d'une `vivo:Team` à un `vivo:Department`, par exemple, s'effectuera à l'aide de la propriété `vivo:hasCurrentMember` qui s'applique à tout type de `foaf:Agent`.

L'utilisation de l'ontologie VIVO pour modéliser les affiliations des entités ne nous permet pas de répondre directement à notre problématique de sauvegarde du vocable utilisé par le membre ayant ajouté l'organisation à la base de connaissances. Pour en tenir compte, nous utilisons la *data properties* `mc2:affiliationElementName` qui s'applique aux `foaf:Organization` et aux `vivo:Team` et qui contient le nom utilisé par l'utilisateur pour désigner l'élément considéré sous la forme d'un `xsd:string`.

À titre d'exemple, la figure 8.2 illustre de manière imagée l'une des deux affiliations de Charles, le consultant de notre scénario. Dans le cadre d'un projet conduit par Argos, il a pu signer un contrat de prestation entre sa propre société et l'équipe SoWeb d'Argos le faisant

évoluer pour un temps sous une double affiliation. Le *listing* de la section B.3 en annexe illustre de manière plus complète ses deux affiliations.

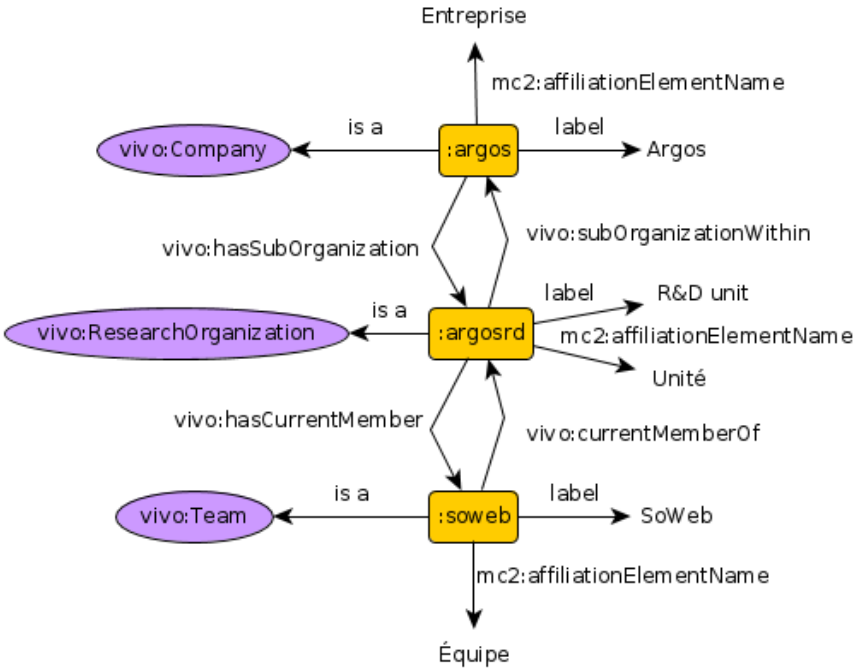


FIGURE 8.2 – Illustration de la chaîne d'affiliation de l'équipe Web Social au sein de laquelle Charles intervient

9. Développement du prototype à partir du modèle

9.1 Survol de l'architecture du prototype

Au cours de cette thèse et afin de tester les différentes théories étudiées, un prototype a été mis en place. Ce prototype a pris la forme d'une plate-forme, que nous avons nommée SIDEKICK [1], qui se compose de différentes briques qui seront décrites dans cette partie.

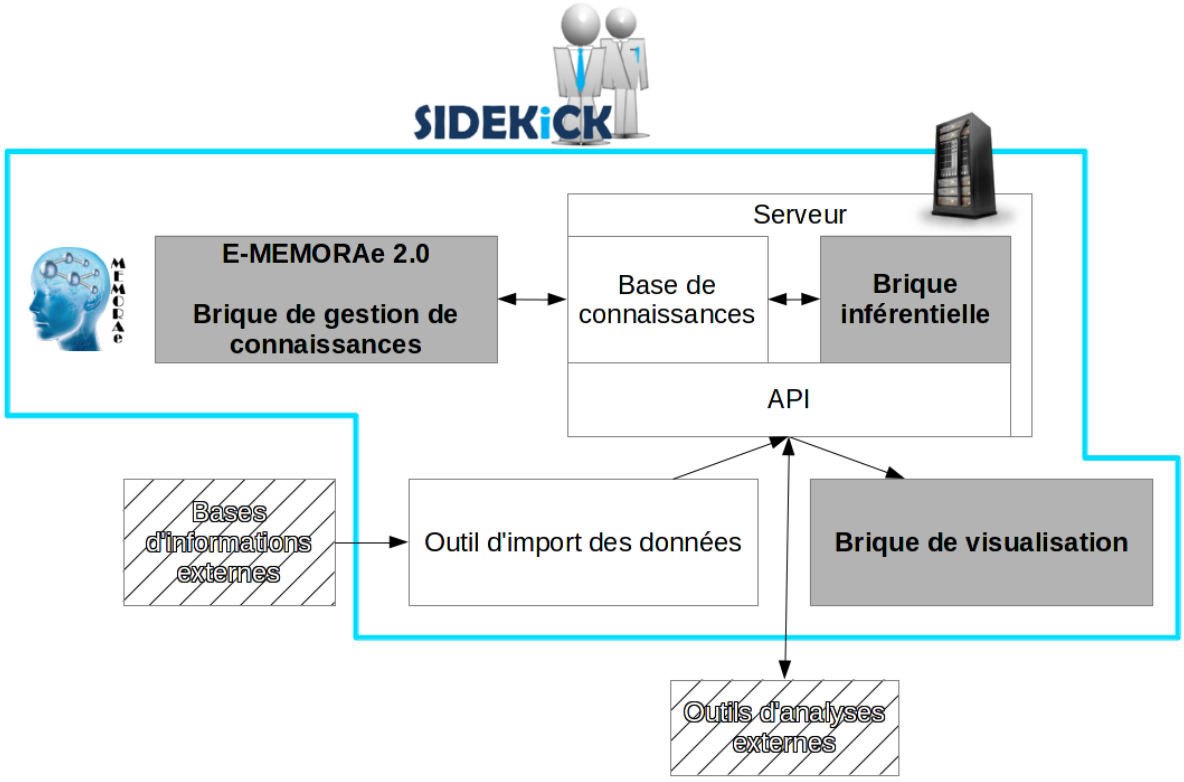


FIGURE 9.1 – Architecture globale de notre prototype

1. Le nom de cette plate-forme est un acronyme pour *Software for socIal Decision support Enabled by Knowledge Interpreted from Collected Key information*, ce qui signifie logiciel d'aide à la décision fondée sur des connaissances interprétées à partir d'informations clés.

Nous présentons ici l'architecture globale de cette plate-forme. Cette dernière est représentée dans la figure 9.1. L'objectif de cette plate-forme d'expérimentation est d'offrir une solution complète au service des organisations, allant d'un environnement de gestion de connaissances pour les membres de l'organisation au système d'analyse des données échangées au sein de l'environnement de gestion de connaissances pour les décideurs.

La plate-forme peut être décomposée en un certain nombre de modules fonctionnels dont les trois briques essentielles qui feront l'objet, pour chacune d'entre elles, d'une section particulière :

— La brique de gestion de connaissance (section 9.2) ;
— La brique inférentielle (section 9.3) ;
— La brique de visualisation (section 9.4).

La brique de gestion de connaissances (section 9.2) permet à n'importe quel membre de l'organisation de bénéficier d'un écosystème d'applications Web dont les ressources produites ou échangées sont capitalisées selon un référentiel partagé unique à toutes les applications. Parmi les applications Web installées, l'accent a été mis sur la mise à disposition de médias sociaux (application de minimessages, forum, wiki) de manière à favoriser la collaboration et recentraliser des services habituellement externalisés (hébergés hors de l'organisation). De cette manière, la base de connaissances peut s'enrichir des informations circulant dans ces applications. Cette brique a la particularité d'être la seule possédant un accès direct à la base de connaissances de la plate-forme. L'environnement de gestion de connaissances et la base de connaissances ont été conçues à partir de la modélisation développée dans le chapitre 8.

Au sein de la brique de gestion de connaissances, les membres de l'organisation, à l'aide de leur compte utilisateur personnel, vont pouvoir réaliser différentes tâches :

— tenir un flux de minimessages, dont chaque élément est considéré comme une ressource par l'environnement ;
— éditer à plusieurs des ressources textuelles (Wikis) ;
— intervenir sur des forums ;
— planifier des évènements ;
— partager des ressources avec d'autres membres, en les plaçant au sein de mémoires partagées. L'accès à ces ressources est contrôlé par le biais de vérifications sur l'appartenance des utilisateurs aux groupes associés à ces mémoires partagées.

Les données ajoutées ou échangées au sein de la base de connaissances sont observées par la brique inférentielle de la plate-forme (section 9.3). La brique inférentielle fonctionne selon deux modes différents :

— un mode automatique, à l'aide de tâches planifiées sur le serveur hébergeant la plate-forme, qui scrute en permanence les informations ajoutées ou retirées de la base de

connaissances et réagit en fonction de règles décrites à l'avance par un décideur.
— un mode manuel qui permet de générer, au besoin, des jeux de données particuliers (par exemple, ceux utilisés par la brique de visualisation ou pour l'extraction de données au format Gephi). Ces jeux de données sont construits à l'aide des mêmes heuristiques que celles utilisées pour déclencher les règles du mode automatique.

Les principales heuristiques ayant été implantées au sein de cette brique inférentielle se fondent sur la notion de concepts. Ces concepts sont issus du référentiel partagé de la base de connaissances avec lesquels les différentes ressources sont indexées. Il peut s'agir de ressources ajoutées ou créées via la brique de gestion de connaissances ou importées via l'outil d'import de données. Les heuristiques, qui ont été conçues pour servir les décideurs de l'organisation pour laquelle la plate-forme a été déployée, sont les suivantes :

— détection de tendances émergentes parmi les concepts du référentiel partagé de la plate-forme ;
— détection de partenaires potentiels autour d'un concept de la plate-forme ;
— génération d'un rapport listant les opportunités de montage de projets de recherche, en fonctions des règles ajoutées à la plate-forme.

Les résultats de ces heuristiques sont accessibles pour les décideurs à l'aide de notre application d'analyse des médias sociaux : la brique de visualisation (section 9.4). La brique de visualisation affiche les résultats de la brique inférentielle au sein de cadrans permettant :

1. D'afficher le graphe des collaborations des différentes organisations indexées dans la base de connaissances et son évolution, éventuellement en fonction d'un ou plusieurs mots-clés.
2. D'afficher diverses statistiques d'utilisation de mots-clés par les différentes organisations indexées dans la base de connaissances.
3. D'afficher l'évolution des tendances d'utilisation des mots clés. La possibilité de prédire l'émergence de mots-clés a également été mise en œuvre.

Les trois briques principales sont visibles dans la figure 9.1 sous la forme de boîtes colorées en gris. Afin d'assurer le fonctionnement de la plate-forme, des modules utilitaires ont été conçus. C'est le cas par exemple de l'API qui permet la communication entre les différentes briques du système. Cette API permet également à d'autres applications de venir interagir avec le système, que ce soit pour obtenir des informations issues de la base de connaissances, avant ou après traitement par la brique inférentielle, ou pour venir peupler la base de connaissances.

C'est le cas par exemple de l'outil d'import de données. Cet outil permet d'ajouter automatiquement et massivement de nouvelles informations à la base de connaissances depuis des sources externes. Il s'agit principalement de pouvoir importer des informations bibliographiques provenant de sources externes (HAL, Web of Knowledge) ou d'outils personnalisés.

Les ressources importées par ce biais sont indexées selon le même référentiel que les ressources produites par la brique de gestion de connaissances. Cet outil d'import des données est appelé à n'être utilisé que par des personnes clairement identifiées parmi les membres de l'organisation. Son usage principal est le peuplement initial de la base de connaissances, afin que les utilisateurs de la brique de gestion de connaissances ou de la brique de visualisation puissent dès le début bénéficier d'une expérience optimale de la plate-forme.

Dans les sections suivantes nous revenons plus en détail sur les briques principales de la plate-forme, qui n'ont été qu'évoquées dans la présente section.

9.2 Brique de gestion de connaissances

Les travaux effectués dans le cadre du projet MEMORAe, décrit dans le chapitre 2.3.9, ont conduit non seulement à la modélisation d'un environnement de gestion de connaissances, mais également à la réalisation du prototype d'un tel environnement qui a pris le nom de « plate-forme E-MEMORAe 2.0 ».

L'approche de la gestion de connaissances suivie par le projet utilise un référentiel partagé pour capitaliser différents types de ressources. L'environnement développé dans le cadre du projet MEMORAe permettait ainsi déjà l'utilisation d'un référentiel partagé pour capitaliser des ressources. Son utilisation nous a offert une base de départ pour nos propres travaux.

Cependant, la modélisation du projet MEMORAe ne prenait pas en compte toutes les ressources produites au sein des médias sociaux. Du même coup, l'environnement E-MEMORAe 2.0 à partir duquel nous sommes reparti ne proposait pas toutes les fonctionnalités sociales que nous souhaitions proposer.

Deux premières versions successives de la plate-forme avaient été précédemment réalisées en PHP/HTML et s'appuyaient sur une base de données MySQL. Nous avons fait évoluer cette plate-forme en lui ajoutant les fonctionnalités manquantes. Malgré les changements induits par la conception d'une troisième version de la plate-forme, nous avons conservé le nom E-MEMORAe 2.0. Le chiffre est en fait utilisé comme un clin d'œil vis-à-vis du Web 2.0 et non plus comme un numéro de version.

Le développement de cette troisième version de la plate-forme a été rendu nécessaire par les changements que nous avons fait au modèle sous jacent, décrit dans le chapitre 8. Le prototype que nous avons développé sert en effet directement de démonstrateur à notre modèle et a donc été réalisé entièrement à partir de ce modèle.

Les versions précédentes de la plate-forme E-MEMORAe incluaient déjà des fonctionnalités que nous utilisons comme les forums, la cartographie sémantique ou les espaces de partages Leblanc et Abel [2010]. Des fonctionnalités existantes comme le module de chat ou

l'agenda partagé n'ont pas encore été portées sur la nouvelle version de la plate-forme, bien qu'elles aient été adaptées et incluses dans notre nouveau modèle. D'autres fonctionnalités qui n'existaient pas sur les précédentes versions ont été ajoutées dans le cadre de cette thèse (wiki ou application de minimessages) ou dans le cadre de travaux attenants, mais en prenant appui sur notre nouveau modèle (annotations [Penciuc, 2012], module d'analyse des traces d'interactions [Li, 2013]).

Le développement de l'interface du nouveau prototype a été entrepris juste avant le début de cette thèse afin de doter la plate-forme de la puissance des interfaces Web dites riches. Cette nouvelle interface a été développée avec le *toolkit* Adobe Flex. Cette interface s'appuie sur un *backend* en PHP dont le rôle est d'assurer la complète compatibilité avec le modèle de données des versions précédentes de la plate-forme et se connecte donc à la même base de données MySQL.

L'une des tâches de cette thèse a donc consisté à enrichir cette nouvelle interface et le *backend* mis en place de manière à prendre en compte les changements apportés au modèle ontologique de la plate-forme. La section 9.2.1 apporte une présentation générale de la plate-forme réalisée. La section 9.2.2 décrit de quelle manière le référentiel partagé est présenté aux utilisateurs de la plate-forme. La section 9.2.3 aborde l'organisation des utilisateurs en groupe et l'accès de ces groupes à leurs espaces de partage respectifs, au sein desquels les utilisateurs vont pouvoir indexer des documents. La mise en pratique de l'indexation est décrite dans la section 9.2.4. Enfin, les sections 9.2.5, 9.2.6 et 9.2.7 présenteront successivement les fonctionnalités de forum, wiki et minimessage de la plate-forme réalisée dans le cadre de cette thèse.

9.2.1 Interface d'accueil

L'intérêt de la plate-forme réside dans son intégration de différentes applications Web telles qu'un forum ou un wiki, permettant aux membres de l'organisation de capitaliser leurs échanges de la même manière que les autres ressources documentaires, selon un même référentiel.

La figure 9.2 présente une capture d'écran de la plate-forme réalisée. Le référentiel commun utilisé pour indexer les ressources y prend la forme d'une cartographie de concepts affichée en permanence de manière centrale pour l'utilisateur. Autour de cette cartographie s'articulent deux boîtes destinées à présenter les informations concernant le focus de la carte ou celles de l'utilisateur courant.

La boîte utilisateur — à droite sur la figure 9.2 et détaillée sur la figure 9.3 — contient principalement :

— le bloc de commandes permettant à l'utilisateur de modifier son profil ou se déconnecter ;

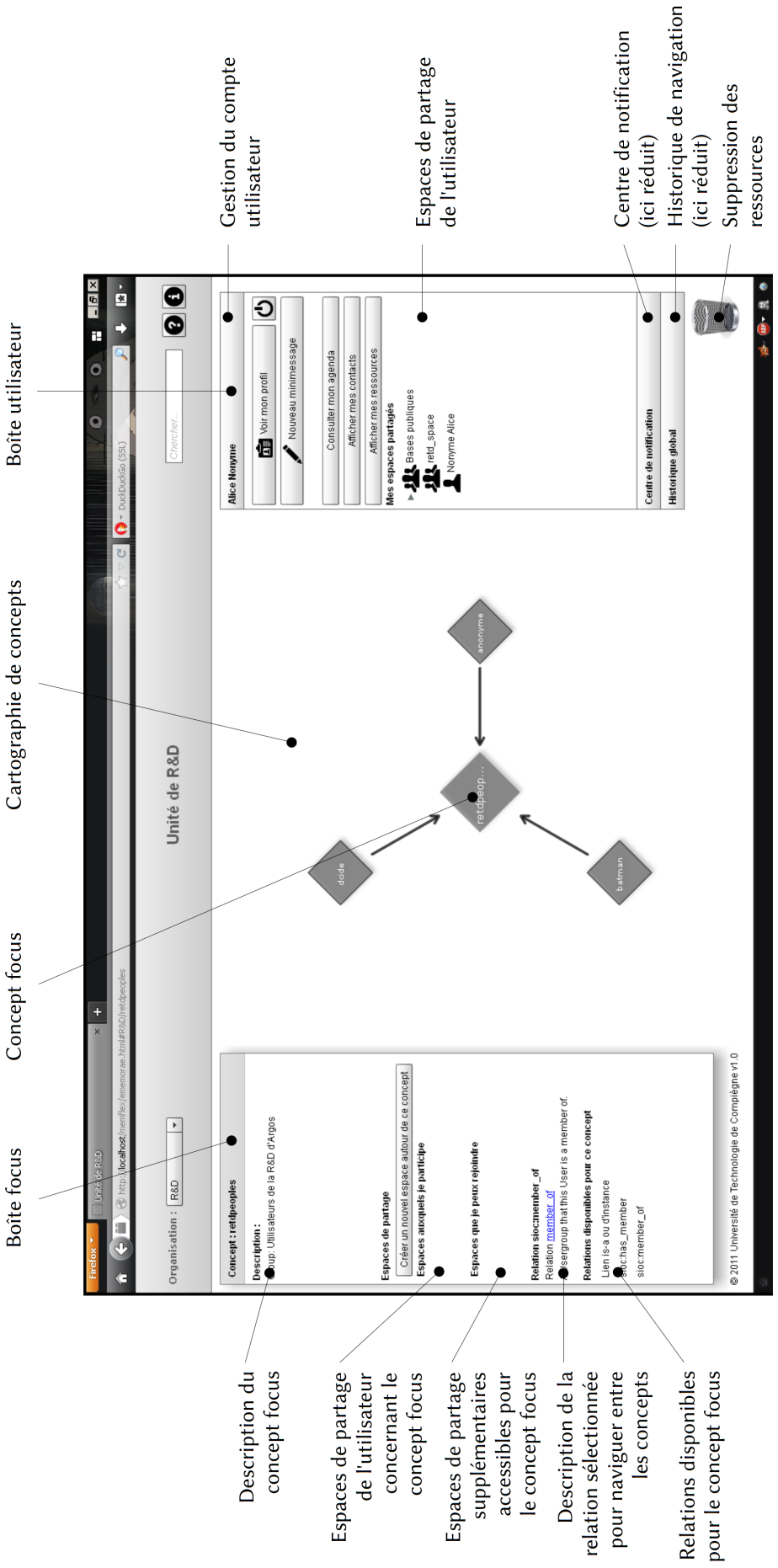


FIGURE 9.2 – Capture d'écran de l'environnement E-MEMORAe 2.0

— le bloc d'accès aux différents espaces de partage auxquels il participe.

FIGURE 9.3 – Détail de la boîte utilisateur

La boîte focus — à gauche sur la figure 9.2 — contient les informations concernant le concept focus actuellement affiché au centre de la cartographie. La boîte focus est donc contextuelle.

Outre l'affichage du label et de la description du concept focus, la boîte focus donne accès aux différentes relations disponibles pour ce concept. La cartographie de concepts étant en fait une représentation d'une ontologie écrite en OWL, ces différentes relations sont, entre autres, les *Object Properties* associées à ce type de concept. Le fait de cliquer sur le nom d'une de ces relations recharge la cartographie centrale pour ne plus montrer que les relations de ce type.

La partie centrale de la boîte focus permet d'afficher les espaces de partage concernant le concept focus, c'est-à-dire les espaces de partage liés aux groupes créés autour de ce concept. Elle permet également de créer un groupe libre souhaitant échanger autour du concept focus. À sa création, ce groupe est constitué d'un seul membre : l'utilisateur créateur. Les autres membres pourront le rejoindre au fil de l'eau.

9.2.2 Cartographie sémantique

L'apport principal de l'approche MEMORAe et de la plate-forme E-MEMORAe 2.0 réside dans la cartographie de concept présentée en permanence à l'utilisateur au centre de l'écran (visible au centre de la figure 9.2). La cartographie est réalisée à l'aide du composant Flex

Relation Browser, développé par M. Stefaner [2], qui permet la navigation au travers d'un réseau sémantique.

La cartographie sémantique représente les différentes classes de l'ontologie memorae-core 2 et leurs instances, ainsi que les classes et instances du référentiel partagé et enfin les relations existantes entre les différentes instances. Les classes ontologiques sont représentées sous la forme de disques de couleur grise portant le label de la classe en leur centre. Les instances sont représentées sous la forme de losanges de couleur grise également, avec le label correspondant au centre. La figure 9.4 présente ces différents types de nœuds en situation.

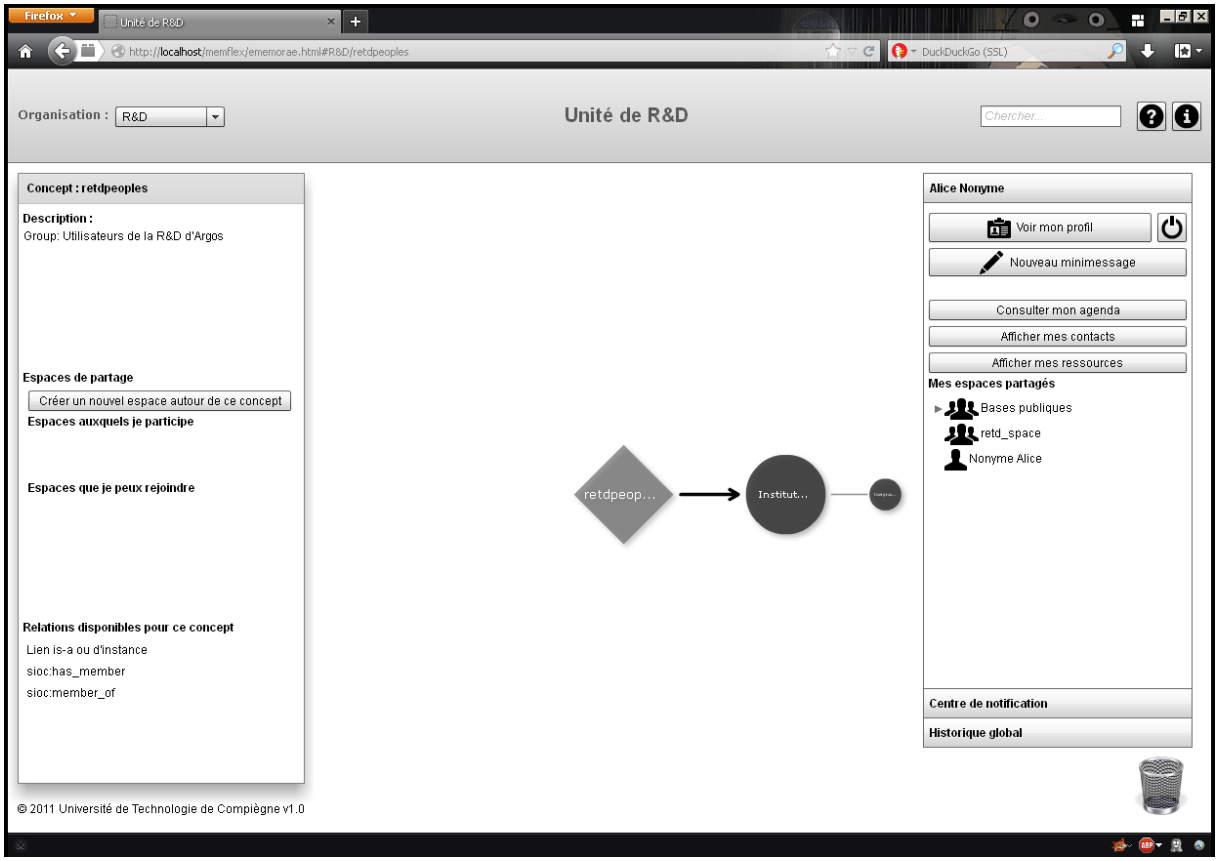


FIGURE 9.4 – Différents types de nœuds

La cartographie positionne toujours un nœud (disque ou losange) au centre de l'écran. C'est ce que nous appelons le nœud ou concept focus. Le fait de cliquer sur n'importe lequel des autres nœuds visibles à l'écran vient le placer au centre de la carte : il devient le nouveau concept focus. Le contenu de la boîte focus et des éventuels espaces de partage ouverts est alors rafraîchi.

Cette action permet de se déplacer au sein de la cartographie. L'environnement E-MEMORAe 2.0 permet deux modes de déplacement distincts :

La navigation verticale Ce mode de navigation correspond au parcours la cartographie séman-

2. http://moritz.stefaner.eu/projects/relation-browser/ (accessible le 10 novembre 2013)

tique en passant de nœud en nœud exclusivement à l'aide des relations dites verticales. Il s'agit des relations de spécialisation `rdfs:subClassOf` et des relations de typage (dans le cas d'une instanciation de classe OWL) `rdf:type`.

La navigation horizontale Ce mode de navigation permet, après avoir choisi une relation d'association particulière, de parcourir la cartographie sémantique selon ce point de vue. Les relations d'association comprennent toutes les *objects properties* instanciées dans la base de connaissances dont les propriétés `skos:related` ou encore, au niveau du modèle, les relations `owl:equivalentClass`. Le choix du mode de navigation ou de la relation à parcourir en mode navigation horizontale s'effectue dans le bas de la boîte focus (détail à la figure 9.5) et dépend du concept focus : toutes les relations ne sont pas disponibles pour tous les concepts.

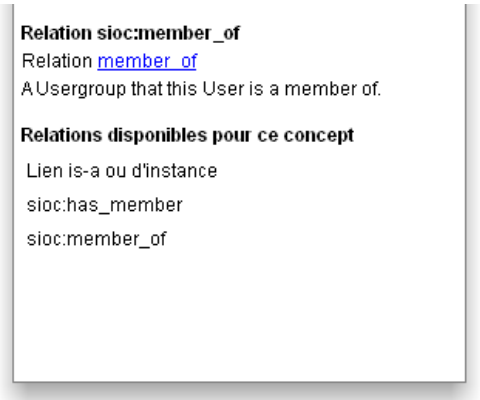


FIGURE 9.5 – Détail de la boîte focus montrant le choix de relations horizontales

9.2.3 Groupes d'utilisateurs et espaces de partage

Nous avons vu dans les sections 8.2 et 8.3 que notre modèle memorae-core 2 permet l'expression de groupes d'utilisateurs. À chacun de ces groupes est associé un espace de partage unique. Ces espaces de partages sont hermétiques et inaccessibles aux utilisateurs qui ne font pas partie du groupe associé. Les membres du groupe d'utilisateur A ne peuvent pas accéder aux documents enregistrés dans l'espace de partage du groupe B. À moins, bien sûr, de faire également partie de ce groupe B.

Chaque utilisateur peut voir l'ensemble des espaces de partage auxquels il a accès sous la forme d'une liste dans la boîte utilisateur à gauche de l'écran. Un clic sur un élément de cette liste permet d'ouvrir une fenêtre liée à l'espace de partage cliqué. Les fenêtres des espaces de partage permettent à l'utilisateur d'accéder aux données qui y sont placées. La capitalisation d'un document s'effectue en effet exclusivement pour un espace de partage donné. La figure 9.6 montre un espace de partage et les différentes données s'y trouvant.

Il existe deux espaces de partage particuliers :

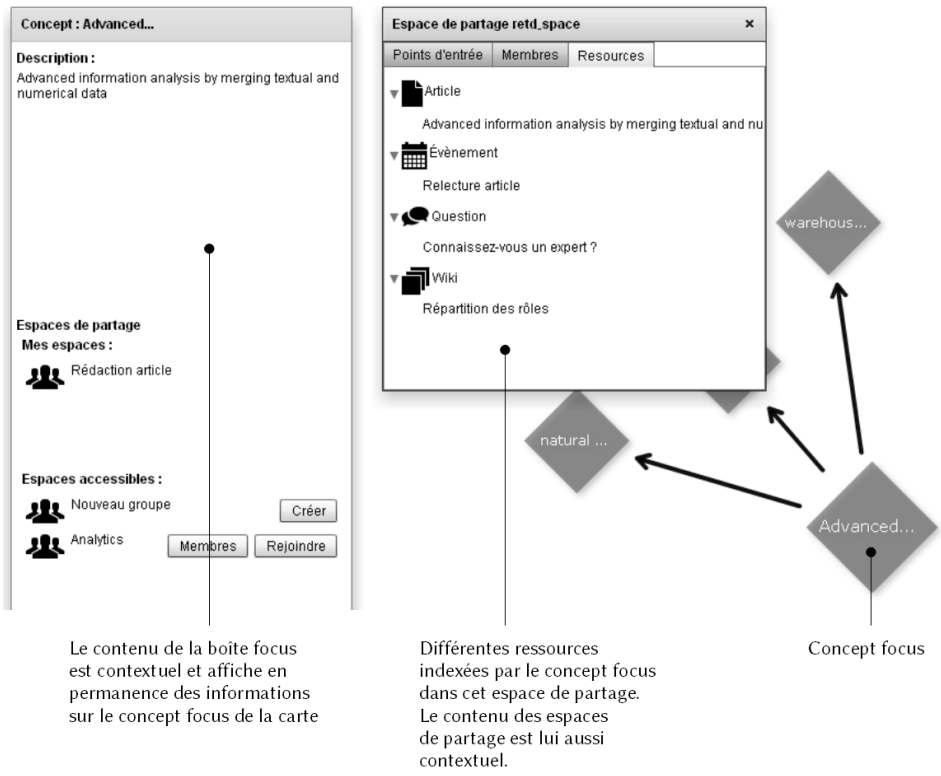


FIGURE 9.6 – Un espace de partage et différents types de ressources

— L'espace de partage commun à toute l'organisation. Cet espace est associé au groupe d'utilisateur comprenant l'ensemble des utilisateurs de l'organisation. Il permet de capitaliser des documents pouvant être consultés par n'importe qui.
— L'espace de partage privé. Chaque utilisateur se voit affecté un espace de partage qui lui est propre. Cet espace est donc associé à un groupe particulier ne contenant qu'un seul membre. Cet espace permet à chaque utilisateur de capitaliser pour lui-même des documents sans qu'aucun autre utilisateur ne soit au courant.

Le contenu des espaces de partage est contextuel et n'affiche que les ressources indexées pour le concept focus. Ainsi, l'accès aux ressources ne se fait pas à l'aide d'un moteur de recherches, mais suite à la navigation volontaire d'un utilisateur au sein de la cartographie sémantique, dans le référentiel commun de son organisation. Cette vue décentralisée de l'utilisateur sur les ressources, en fonction du concept focus, permet de n'afficher à l'utilisateur que les informations qui l'intéressent vraiment.

9.2.4 Capitalisation de documents

La capitalisation de documents consiste, au sein de l'environnement E-MEMORAe 2.0, à associer un document à un concept du référentiel partagé et rendre cette association visible pour un groupe d'utilisateurs particulier. Ce faisant, une nouvelle ressource apparaît alors

dans l'espace de partage du groupe en question lorsque le concept focus est l'un des concepts pour lequel la ressource est indexée (comme visible sur la figure 9.6). Cette association peut se faire de deux manières différentes, selon le type de ressource considéré.

S'il s'agit d'un document provenant d'un outil externe à la plate-forme (traitement de texte, création graphique, capture vidéo, etc.) il est nécessaire de l'ajouter au préalable à l'environnement en le téléversant dans une zone spéciale. Cette zone spéciale est personnelle et contient l'ensemble des ressources ajoutées ou créées par un utilisateur. Ce dernier peut alors les indexer avec un ou plusieurs concepts du référentiel partagé par une action de glissé déposé de cette zone spéciale vers le concept de leur choix de la cartographie sémantique. Si le concept souhaité n'est pas visible à ce moment là sur la carte, il est nécessaire de s'y rendre au préalable.

Les ressources d'un utilisateur sont indéfiniment listées au sein de la zone spéciale. De fait, une ressource n'est jamais réellement supprimée de l'environnement. L'action de suppression (en déplaçant une ressource d'un espace de partage vers la corbeille située en bas à droite de l'écran) supprime simplement l'association existant entre la ressource, l'espace de partage en question et le concept focus. Il s'agit de supprimer un contexte particulier de la ressource.

Il ne faut pas confondre cette zone spéciale et l'espace de partage privé des utilisateurs. En effet cette zone spéciale contient l'ensemble des ressources ajoutées ou créées par un utilisateur (dont des ressources non encore indexées) indépendamment du concept focus, tandis que l'espace de partage privé contient exclusivement des ressources capitalisées et ne les présente qu'en fonction du concept focus. En particulier, l'espace privé d'un utilisateur peut contenir des ressources qu'il n'a pas créées mais qu'il a tirées d'un espace de partage public. L'indépendance par rapport au concept focus de cette zone spéciale permet de capitaliser un document sur différents concepts, même si ces derniers ne sont pas visibles en même temps à l'écran.

L'environnement permet également de produire directement des ressources à l'aide d'applications Web. L'accès à ces applications s'effectue depuis la fenêtre d'un espace de partage particulier. Ceci permet une première indexation automatique de la nouvelle ressource créée suite à l'utilisation d'une application. La ressource est indexée pour le concept focus de la cartographie sémantique au sein de l'espace de partage via lequel l'application avait été ouverte.

Assez classiquement un utilisateur aura accès à différents espaces de partage (au minimum son espace privé et l'espace public de l'organisation). En fonction du concept focus, il lui est tout à fait possible de changer le niveau d'accès d'une donnée en la copiant d'un espace de partage restreint vers un espace plus ouvert ou inversement. Cette action se réalise en effectuant un glissé-déposé de l'espace de partage d'origine vers celui de destination.

9.2.5 Les forums

L'environnement E-MEMORAe 2.0 offre une fonctionnalité de type forum. Pour cela, un utilisateur va tout d'abord poser une question à laquelle d'autres utilisateurs ou lui-même pourront répondre. La question et chacune de ses réponses sont considérées au sein de l'environnement comme autant de données différentes qui peuvent être documentarisées en les capitalisant par différents concepts et rendues visibles au sein de différents espaces de partage. Les réponses restent néanmoins toujours liées à leur question d'origine.

L'ouverture d'une ressource de type question du forum s'effectue sous la forme d'une fenêtre qui présente à l'utilisateur à la fois la question d'origine et l'ensemble des réponses déjà ajoutées par les membres de l'organisation. L'ouverture d'une ressource de type réponse ouvre le même genre de fenêtre, mais met en exergue la réponse demandée. Celle-ci se trouve donc *de facto* remise dans le contexte de la question originale.

9.2.6 Les wikis

L'environnement E-MEMORAe 2.0 permet de créer des pages de wiki. Il s'agit de ressources textuelles créées directement au sein de l'environnement et auxquelles n'importe quel utilisateur peut contribuer, pourvu qu'il soit membre de l'espace de partage dans lequel la page résultante est capitalisée. Chaque modification est enregistrée comme une révision supplémentaire de la ressource originale.

L'ouverture d'une ressource de type wiki n'affiche que la dernière révision en date de la page dans une fenêtre dédiée. Les anciennes révisions sont accessibles depuis un bouton dans cette fenêtre. Chaque révision, indépendamment, peut être indexée pour un concept particulier du référentiel partagé et ainsi suivre son propre cycle de documentarisation (qui peut impliquer de nouvelles modifications, une nouvelle indexation etc.). Chaque révision reste néanmoins liée à sa révision précédente, ce qui permet de remonter son historique. Une même révision peut être à l'origine de deux (ou plus) ressources différentes dont les futures révisions seront divergentes.

9.2.7 Les minimessages

Nous avons développé une application de minimessage afin d'améliorer le partage rapide d'informations entre utilisateurs de la plate-forme. Pour mémoire, ce type de message peut être composé de plusieurs éléments remarquables que sont les mentions et les *hashtags*. Nous les illustrons à l'aide du minimessage suivant :

> @alice #yammer vient de se faire racheter par #microsoft http://bit.ly/KAlhGq #km #social_network

Ce message, envoyé par Baptiste au cours de sa veille quotidienne à Alice, présente ainsi les éléments suivants :

— une « mention » `@alice` qui permet d'interpeller l'utilisateur répondant au login d'Alice sur l'application de minimessage ;
— différents *hashtags* permettant d'étiqueter le message ;
— une *Uniform Resource Locator* (URL) pointant vers une ressource externe.

L'environnement E-MEMORAe 2.0 reconnaît automatiquement les *hashtags* et les mentions des messages. Lors de l'enregistrement d'un tel message, des liens sont automatiquement créés vers les instances de la classe `skos:Concept` dont l'un des `skos:prefLabel` correspond à un *hashtag* détecté ou vers les instances de la classe `mc2:UserAccount` dont le `sioc:id` correspond à une mention détectée. Les minimessages sont ainsi automatiquement indexés sur les bons concepts du référentiel partagé. Ainsi, dans notre exemple, le minimessage sera directement indexé par les concepts Yammer, Microsoft, KM et Social Network du référentiel et le concept représentant Alice au sein de la base de connaissances.

Les minimessages étant considérés comme des ressources publiques, ils sont systématiquement capitalisés de manière à être accessible par tous, c'est-à-dire pour le groupe d'utilisateur institutionnel comportant tous les utilisateurs de l'écosystème numérique. Notre message d'exemple sera donc visible dans l'espace de partage `:retd_space` lié au groupe `:retdpeople`, tel que définis dans la section 8.3.

L'utilisation des mentions dans les minimessages permet à leurs expéditeurs d'informer rapidement un autre utilisateur. Dans notre exemple de minimessage précédent, l'utilisatrice « alice » se verra informée par une notification de l'ajout de ce minimessage. Cette notification apparaîtra dans le centre de notification dont l'accès est visible sur la figure 9.2.

9.3 Brique inférentielle

Nous inscrivons la recherche de nouvelles connaissances et plus particulièrement l'étude de l'émergence de nouvelles connaissances dans le domaine de l'aide à la décision. En effet, nous cherchons à valider le fait qu'au-delà d'augmenter le capital de connaissances de l'organisation, les nouvelles connaissances pouvant émerger de l'utilisation des médias sociaux de l'organisation peuvent entrer dans un processus de décision.

Nous avons décidé de concevoir notre outil d'aide à la décision sous la forme d'une brique inférentielle appuyée sur une base de connaissances alimentée par les productions documentaires d'une organisation mais aussi par des informations issues de médias sociaux.

Cette brique réalise trois fonctions qui seront successivement présentées :

— extraction d'un graphe de collaboration de la base de connaissances ;

— suivi et prédiction de tendances thématiques ;
— proposition de consortiums pouvant servir dans l'élaboration d'une réponse à un appel à projet européen.

9.3.1 Cartographie des collaborations

Ce module doit permettre de représenter sous la forme d'un graphe, fait de nœuds et de liens, le réseau social de collaboration des différentes organisations dont les informations ont été enregistrées dans la base de connaissances. Le graphe du réseau présenté à l'utilisateur est bien une vue particulière sur les données enregistrées dans la base de connaissances. Chaque nœud de ce graphe représente un partenaire différent, tandis qu'un lien entre deux nœuds doit représenter la force de la relation unissant les deux partenaires.

Au sein de notre écosystème, nous considérons trois sources d'informations pour extraire le graphe social des entités d'un écosystème de connaissances :

— Les informations directement données par les utilisateurs du système. Une application de l'écosystème permet d'ajouter des entités à la base de connaissances et de préciser les collaborations qu'elle a pu avoir, que ce soit en déclarant des relations de co-auteurs ou des collaborations contractuelles, qui peuvent être représentées à l'aide de la classe `vivo:Agreement` et ses sous classes.
— Les informations extraites des affiliations des auteurs des documents indexés au sein de l'écosystème. VIVO permet en effet de décrire l'appartenance de personnes au sein d'équipes ou d'organisations. Nous suivons ainsi la définition de document numérique de Pédauque [2003] qui, dans sa forme « en tant que medium », amène l'idée qu'un document peut être vu comme la trace d'une collaboration. Nous déclarons alors qu'il y a eu collaboration entre les entités représentées par les affiliations des auteurs du document considéré. Nous rappellons qu'avec notre positionnement théorique les ressources créées au sein des médias sociaux peuvent être documentarisées et sont donc également considérées par notre méthode d'extraction d'informations.
— Les informations extraites des affiliations des personnes interpellées à l'aide de mentions au sein des documents [3].

Le calcul de la force des relations est effectué à l'aide d'une version modifiée de l'algorithme d'*edgerank* utilisé sur Facebook, dont nous rappelons l'expression originale dans la formule (9.1). Cette formule permet de ranger les différents objets devant apparaître sur le *wall* d'un utilisateur A en fonction de paramètres que nous avons déjà définis dans la section 4.3 :

u_e degré d'affinité entre l'utilisateur A et l'auteur de l'objet considéré ;

3. Voir à ce sujet l'explication de ce que sont exactement les mentions dans la section 9.2.7.

w_e facteur de pondération du score final en fonction du type d'objet considéré ;

d_e facteur de pondération du score final en fonction de l'ancienneté de l'objet considéré.

$$\sum_{edges\ e} u_e w_e d_e \tag{9.1}$$

À la différence du graphe social de Facebook, composé d'utilisateurs s'organisant autour d'objets de différents types (page Web, messages personnels, liens, photographies, commentaires, etc.), notre graphe est lui composé d'organisations et des liens qu'elles entretiennent au travers de publications communes, de comptes-rendus, de brevets, etc. Du fait de nos données, les paramètres de la formule vont avoir, dans notre implémentation, les significations suivantes — où A et B désignent deux entités ayant collaboré au travers de différents *edges* e représentant donc les différents liens ou actes de collaboration passés :

u_e est le coefficient de corrélation d'un acte de collaboration à un certain concept k du référentiel commun, donné par le décideur. Dans le cadre d'une analyse générique, c'est-à-dire lorsqu'aucun concept n'est donné par le décideur, ce coefficient prend simplement la valeur 1. Dans le cadre de l'analyse des collaborations portant sur un concept k, ce coefficient permet de mettre en valeur les liens impliquant effectivement ce concept et de mettre en retrait les autres liens. Le coefficient est alors calculé à l'aide de la formule (9.2). Dans cette formule, nous posons K comme étant l'ensemble des concepts k_i utilisés pour étiqueter la collaboration l étudiée. Nous posons ki_{match} comme ayant une valeur de 1 si $k_i = k$ et 0 sinon. Nous permettons au décideur de régler finement l'importance relative qu'il veut donner à la correspondance thématique à l'aide de la constante d'appréciation a_k qui ne sera prise en compte que dans le cas des collaborations portant sur le concept k.

$$u_e = \sum_{k_i \in K} \frac{ki_{match} a_k}{Card(K)} \tag{9.2}$$

w_e donne l'importance relative du type de collaboration (publication, brevet, etc.). Nous avons fixé ces valeurs selon le type de publication dans le code source de notre prototype. Ces dernières sont bien sûr modifiables en fonction de l'importance relative que l'on souhaite donner à un type de collaboration. Nous utilisons pour l'instant les valeurs suivantes :

- pour le développement conjoint ou la réalisation d'un projet : 4 ;
- pour l'écriture d'un brevet ou le suivi d'une thèse : 3 ;
- pour la rédaction d'une publication : 2 ;

— pour tout autre fait de collaboration, à commencer par les réunions de travail informelles : 1.

d_e est le coefficient de dépréciation temporel d'un acte de collaboration définit par la formule (9.3). Pour ce calcul, nous posons $t1$ comme étant la date de fin de cette collaboration et $t2$ la date courante au moment du calcul — date pouvant varier pour mesurer l'évolution de la force d'un lien de collaboration à travers le temps. Nous autorisons le décideur à influer sur le coefficient de dépréciation temporel en lui permettant de fixer la constante c_t. Cela permet de contrôler l'influence du temps vis-à-vis des autres coefficients de la formule (9.1).

$$d_e = \frac{c_t}{(t2 - t1) + 1} \tag{9.3}$$

Ainsi, notre formule développée finale, notée S_{AB} (pour Score de la collaboration entre A et B), peut prendre deux formes selon que nous souhaitons procéder à un classement des relations entre acteurs sur un concept en particulier (formule (9.4)) ou de manière générique (formule (9.5)).

$$S_{AB} = \sum_{edges\ e} \overbrace{\left(\sum_{k_i \in K} \frac{ki_{match} a_k}{Card(K)} \right)}^{u_e} \times w_e \times \overbrace{\frac{c_t}{(t2 - t1) + 1}}^{d_e} \tag{9.4}$$

$$S_{AB} = \sum_{edges\ e} w_e \times \overbrace{\frac{c_t}{(t2 - t1) + 1}}^{d_e} \tag{9.5}$$

Le calcul d'*Edgerank* se prête bien pour le calcul de l'importance d'un lien pouvant effectivement exister entre deux entités. Il requiert néanmoins l'existence d'une trace de collaboration entre les deux entités considérées, ce qui n'est bien sûr plus possible lorsque l'on cherche à évaluer la proximité thématique de deux entités n'ayant encore jamais collaboré.

Nous exprimons cette proximité thématique sous la forme de liens potentiels unissant deux entités. Ces liens vont avoir une importance plus ou moins grande en fonction du nombre de concepts partagés par les deux entités et la proximité temporelle de leurs travaux sur ces concepts.

Pour un concept k et deux entités A et B donnés, nous définissons ces liens potentiels à l'aide de leur score noté SP_{AB} (pour Score de la collaboration Potentielle entre A et B) selon la formule (9.6), qui utilise les paramètres suivant :

— la durée des dernières périodes de collaboration de A et B sur k ;
— le type des dernières périodes de collaboration de A et B sur k ;

— la durée de travail parallèle, ou recouvrement, sur k de A et B, au cours de leurs derniers travaux respectifs sur k, si elle existe ;
— l'ancienneté de cette période de recouvrement, si elle existe.

$$SP_{AB} = \frac{gp}{pp} \times \frac{recouvrement}{ancienneté} \tag{9.6}$$

La formule (9.6) finale requiert de comparer les deux durées de travaux de A et B sur k, afin d'identifier la durée de travail la plus longue. Nous utilisons alors les termes de gp (grande période) et pp (petite période). Ces dernières peuvent concerner indifféremment A ou B.

9.3.2 Suivi et prédiction de tendances

L'utilisation d'un même référentiel pour capitaliser différents types de ressources et en particulier celles circulant sur les médias sociaux permet d'observer les tendances des concepts du référentiel en fonction de leurs utilisations au sein de la base de connaissances.

La notion d'émergence d'une tendance est définie comme étant le fait que la fréquence d'utilisation d'un concept ne dépasse pas une courbe de seuillage sur une certaine période avant de croiser cette courbe à une date, appelée date d'émergence.

La difficulté consiste à choisir la courbe de seuillage. Dans le cadre de nos travaux, nous considérons deux types de seuillage : statique et dynamique.

Le seuillage statique ou constant consiste à définir une fois pour toute une valeur de référence pour toute la période de temps observée. Le seuillage dynamique consiste à faire varier, pour chaque pas de l'échantillonnage sur la période observée, la valeur du seuil par une formule quelconque.

Nous avons implémenté trois méthodes de seuillage dynamique au sein de notre brique inférentielle :

Courbe définie par le décideur La courbe de seuillage suit un profil défini à la main par le décideur qui va affecter une valeur pour chaque pas de l'échantillonnage. Dans notre algorithme, si le décideur ne renseigne pas de valeur pour un pas quelconque, la valeur du pas précédent sera utilisée.

La valeur maximale à la date Pour chaque pas de l'échantillonnage, la courbe de seuillage prend pour valeur le score le plus élevé des concepts pour ce pas. Cette formule permet de détecter les concepts sortant réellement du lot à leur date d'émergence, sans s'occuper réellement de leur évolution précédente.

La moyenne à la date Pour chaque pas de l'échantillonnage, la courbe de seuillage prend pour valeur la moyenne des scores des différents concepts pour ce pas. Cette formule permet

de ne considérer comme candidat à l'émergence que des concepts dont l'évolution serait dans un premier temps « noyée dans la masse. »

La détection d'émergence de concepts s'effectue en fixant la date d'émergence voulue. Ceci permet d'identifier les mots clés émergeant effectivement à cette date, c'est-à-dire dépassant la courbe de seuillage. Sont alors exclus les concepts dont l'évolution avait déjà fait croiser la courbe de seuillage avant la date d'émergence étudiée, ou n'émergeant que plus tard.

Notre heuristique de suivi de tendances et de détection d'émergence de concepts autorise la fixation d'une date d'émergence dans le futur. Nous cherchons à proposer des prédictions d'émergence future de concepts. Cette prédiction pose néanmoins deux problèmes :

— en l'absence de données, de quelle manière peut-on définir une courbe de seuillage dynamique dans le futur ?
— de quelle manière peut-on prédire l'évolution de l'utilisation d'un concept ?

Nous répondons dans notre heuristique à la première question en muant la courbe de seuillage dynamique en courbe statique lors de l'étude des tendances futures. La valeur retenue est alors la dernière valeur calculée de la courbe de seuillage, en fonction des données dont nous disposons réellement : la courbe de seuillage est stagnante dans le futur.

Pour répondre à la seconde question, nous utilisons trois profils d'évolutions différents.

Évolution stagnante L'évolution du score d'utilisation du concept reste fixée à la dernière valeur disponible dans les données de la base de connaissances.

Évolution conservative L'évolution du score d'utilisation du concept va se maintenir selon le même coefficient directeur qu'entre les deux dernières valeurs calculées.

Évolution moyenne L'évolution du score d'utilisation du concept va suivre le coefficient directeur moyen de son évolution sur l'ensemble de la période dont les données sont disponibles dans la base de connaissances.

Dans le cadre d'une analyse manuelle de l'émergence des concepts, c'est au décideur de faire ses choix de courbe de seuillage et de profil de prédiction.

Dans le cadre d'une analyse automatique, la courbe de seuillage retenue est la courbe suivant la valeur moyenne des scores des différents concepts à chaque pas de l'échantillonnage. Pour une date d'observation préalablement définie dans le futur par un décideur, notre heuristique va tenter, en testant les différents profils d'évolution l'un après l'autre et en les combinant, de vérifier pour chaque concept du référentiel commun de la base de connaissances s'il peut émerger ou pas à la date donnée. Si tel est le cas, un rapport est alors généré et décrit l'évolution que devrait suivre le concept pour émerger. Cette courbe pourra alors servir de référence dans les mois suivants pour vérifier si l'émergence se vérifie et prendre les mesures appropriées.

9.3.3 Proposition de consortiums pour montage de projets

Notre heuristique permettant la génération de rapports sur l'identification de partenaires potentiels dans le cadre du montage d'un projet repose en grande partie sur les heuristiques présentées dans la section 9.3.1. En effet ce sont ces scores de proximité autour de concepts déterminés qui nous aident à définir la pertinence d'une alliance.

Le montage d'un projet de recherche ou industriel suit par ailleurs un certain nombre de règles définies par les organismes financeurs. Ces règles peuvent varier d'une année à l'autre et posent donc la question de leur implémentation au sein d'un système informatique.

Comme notre plate-forme prototype est modélisée à partir d'ontologies, il nous a semblé pertinent de modéliser les différentes règles impliquées dans le montage d'un projet sous la forme d'une ontologie. De cette manière, les différentes règles pourront être sauvegardées dans la même base de connaissances que les autres ressources. Cette ontologie complète notre modèle memorae-core 2 et définit les classes suivantes, qui spécialisent la classe `owl:Thing`.

`mc2:Rule` Il s'agit d'une règle atomique. Les règles sont composées de `mc2:RuleParameter`.

`mc2:RuleParameter` Il s'agit du paramètre d'une règle. Chaque paramètre peut avoir une ou plusieurs valeurs. Ces valeurs peuvent être d'autres objets de la base de connaissances (liés à l'aide le l'*object property* `mc2:ruleValue`) ou des scalaires (liés à l'aide de la *data property* `mc2:ruleScalarValue`).

`mc2:RulesBook` Cette classe, pouvant être traduite par « cahier de règles » s'instancie pour décrire un objet agrégeant un ensemble de `mc2:Rule` afin de décrire les prérequis d'un appel à projet.

Autour de ces classes, nous avons défini les *object properties* suivantes :

`mc2:containsRule` qui permet d'associer les différentes règles à un cahier de règles.

`mc2:ruleListValue` qui permet d'associer une liste de valeur (des objets de la base de connaissances) à une règle.

`mc2:ruleValue` qui permet d'associer un autre objet de la base de connaissances à une règle.

`mc2:hasParameter` qui permet de définir les paramètres d'une règle.

Nous avons également défini les *data properties* suivantes :

`mc2:ruleScalarValue` qui s'applique à un objet de type `mc2:RuleParameter` et permet de définir sa valeur scalaire.

`mc2:ruleTarget` qui s'applique aux `mc2:Rule` et permet de définir si la règle s'applique à un partenaire particulier ou au consortium complet. En effet, notre travail porte sur la

vérification qu'un ensemble d'organisation forme un consortium viable pour un projet. Les conditions à vérifier s'appliquent ainsi autant aux organisations elles-mêmes (pays d'origine, type d'organisation, etc.) qu'au potentiel consortium (nombre de participants, proportion d'académiques, etc.). Les valeurs possibles pour cette propriété sont « partner » si la cible est un partenaire ou « consortium » si la cible est une instance de la classe `vivo:Consortium`.

`mc2:ruleAction` qui s'applique également aux règles et permet de définir l'état de sortie de la règle. Dans le cas d'une règle définissant un état acceptable pour le consortium, la valeur sera « ok ». Dans le cas d'une règle définissant un état non-acceptable, la valeur sera « ko ».

Dans le cadre de nos travaux nous avons principalement travaillé sur la modélisation des règles encadrant le montage de projets FP7 [4] et DGA ASTRID [5]. Ainsi, les règles encadrant le montage d'une réponse à un appel à projet européen type FP7 pourront s'exprimer comme dans le listing suivant.

```
:fp7ictrules a mc2:RulesBook.

:iceland a vivo:Country.
:switzerland a vivo:Country.

:associated_countries a mc2:RuleParameter;
    mc2:ruleListValue (:iceland, :switzerland).

:member_countries a mc2:RuleParameter;
    mc2:ruleListValue (:deutschland, :france).

:rule1 a mc2:Rule;
    mc2:ruleTarget ”consortium”;
    mc2:hasParameter [
        a mc2:RuleParameter;
        mc2:ruleScalarValue 3.
    ];
    mc2:hasParameter :associated_countries.
```

4. Le *seventh Framework Programme* ou septième programme-cadre européen est un programme de financement à l'initiative de l'Union Européenne pour soutenir et encourager la recherche au sein de l'Union. Site officiel de CORDIS, l'agence coordonnant les programmes-cadres européens et accessible le 10 novembre 2013 : http://cordis.europa.eu/fp7/home_fr.html

5. Le programme de financement ASTRID vise à soutenir des projets de recherche et d'innovation sur des thématiques d'intérêt pour la Défense. Ces projets sont financés par la DGA.

Les notions d'appel à projet et de projet existent déjà dans l'ontologie VIVO et nous permettent de tisser directement des ponts entre ces règles et le reste de la base de connaissances :

```
:cordis a vivo:FundingOrganization.

:fp7ictpropal a vivo:ResearchProposal.

:fp7ictrules a mc2:RulesBook.

:fp7ictproject a vivo:Grant;
    vivo:grantAwardedBy :cordis;
    vivo:supportedInformationResource :fp7ictpropal;
    mc2:requireRules :fp7ictrules.
```

Après avoir modélisé à l'aide de cette ontologie quelques cadres contractuels au montage de projets industriels ou de recherche, la génération d'un rapport consiste à trouver le ou les consortiums les plus à même de remporter l'appel à projet considéré en fonction du cahier de règles lié à cet appel. Pour ce faire, nous ordonnons les différentes relations effectives ou potentielles unissant deux à deux les organisations enregistrées dans la base de connaissances. Cet ordonnancement s'effectue à l'aide du score de *Edgerank*, tel que défini dans la section 9.3.1, associé à chacune des relations. Ce score est bien sûr calculé et peut donc varier pour chacune des thématiques impliquées dans une relation donnée. Une heuristique s'exécute alors pour identifier, pour chacune des thématiques, les regroupements d'organisations qui maximisent les scores de *Edgerank* tout en respectant les règles de constitution des consortiums. Les solutions possibles, si elles existent, sont alors transmises via la brique de visualisation aux décideurs de l'organisation.

9.4 Brique de visualisation

Jusqu'à présent nous avons pu voir de quelle manière les utilisateurs de notre prototype pouvaient peupler une base de connaissances à l'aide de la brique de gestion de connaissances (section 9.2). Notre positionnement théorique stipule que nous pouvons considérer cette base de connaissances comme une base de données stockant à la fois des données et leurs contextes, permettant de les re-matérialiser sous la forme de ressources.

La brique inférentielle présentée précédemment (section 9.3) se positionne avant la phase de matérialisation des données. Elle ne traite en effet que des informations, c'est-à-dire des données re-contextualisées issues de la base de connaissances. Le but de ce traitement, comme

nous l'avons vu, est de trouver de nouvelles relations entre les informations déjà enregistrées dans la base.

La brique de visualisation que nous avons ajoutée à notre prototype permet la matérialisation de ces nouvelles informations et leur consultation par les décideurs d'une organisation. Cette brique de visualisation, conçue comme une application Web intégrée à l'écosystème applicatif que forme notre prototype, a été réalisée à l'aide de la bibliothèque javascript d3js [6]. Elle présente au sein d'un tableau de bord unifié différents cadrans :

— le graphe des collaborations passées ou possibles des différentes entités enregistrées dans la base de connaissances de l'organisation ;
— des diagrammes séries affichant les tendances mesurées ou prédites de l'utilisation des concepts du référentiel partagé de l'organisation dans l'indexation de documents (ce qui permet, par extension, de mesurer l'activité d'une thématique) ;
— des diagrammes circulaires présentant les statistiques d'utilisation des concepts du référentiel partagé pour indexer les productions des différentes entités du graphe de collaboration ;
— un lien vers l'outil de génération de rapport de proposition de consortiums.

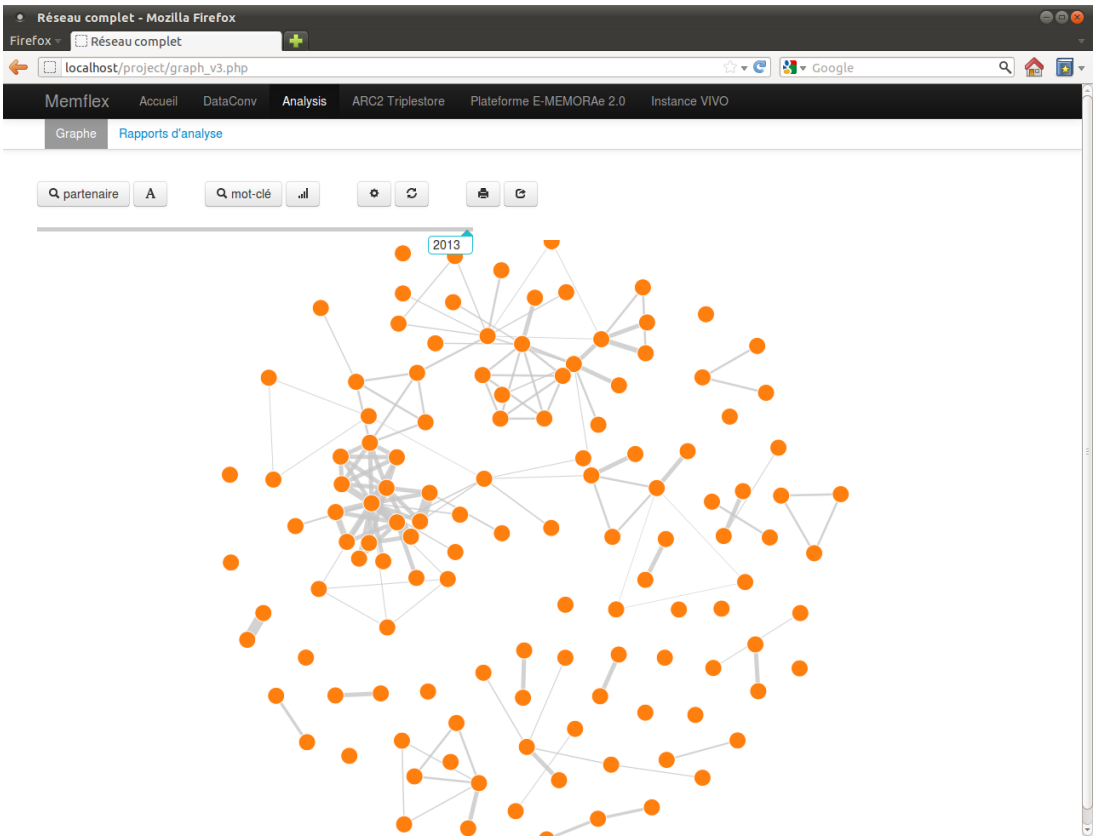


FIGURE 9.7 – Vue générale de la brique de visualisation

La figure 9.7 présente la vue générale de la brique de visualisation au démarrage. À ce

6. Site officiel de la librairie d3js : http://d3js.org/ (accessible le 10 novembre 2013)

moment là, seul le graphe de collaboration est visible, les autres cadrans n'étant ouverts qu'au besoin. Ils viennent alors se positionner au dessus du graphe de collaboration.

Chacun de ces cadrans possède des outils permettant aux décideurs d'interagir avec les informations matérialisées sous leurs yeux, de les mettre en forme selon leurs préférences et éventuellement de les exporter. Ce faisant, les décideurs s'approprient les ressources affichées à l'écran, processus via lequel ils vont pouvoir tirer de nouvelles connaissances qu'ils pourront capitaliser au sein de la base de connaissances de l'organisation en y sauvegardant le document, fruit de leur appropriation des ressources présentées.

Nous illustrons plus en détail la présentation de chacun de ces cadrans avec notre scénario. Dans ce dernier, nous posons que l'équipe de recherche SoWeb de l'unité de R&D de l'organisation Argos est régulièrement représentée au sein de la communauté scientifique d'Ingénierie des Connaissances (IC). Les membres de l'équipe SoWeb ont progressivement peuplé la base de connaissances de la plate-forme prototype avec les actes de la conférence IC disponibles sur HAL [7], soit 206 articles allant de 1998 à 2012. Denise, en sa qualité de décideur (comme elle est, rappelons-le, responsable de l'équipe SoWeb), a accès à la brique de visualisation.

Le cadran affichant le graphe des collaborations peut lui permettre d'étudier, par exemple,

7. http://hal.archives-ouvertes.fr/ (accessible le 10 novembre 2013)

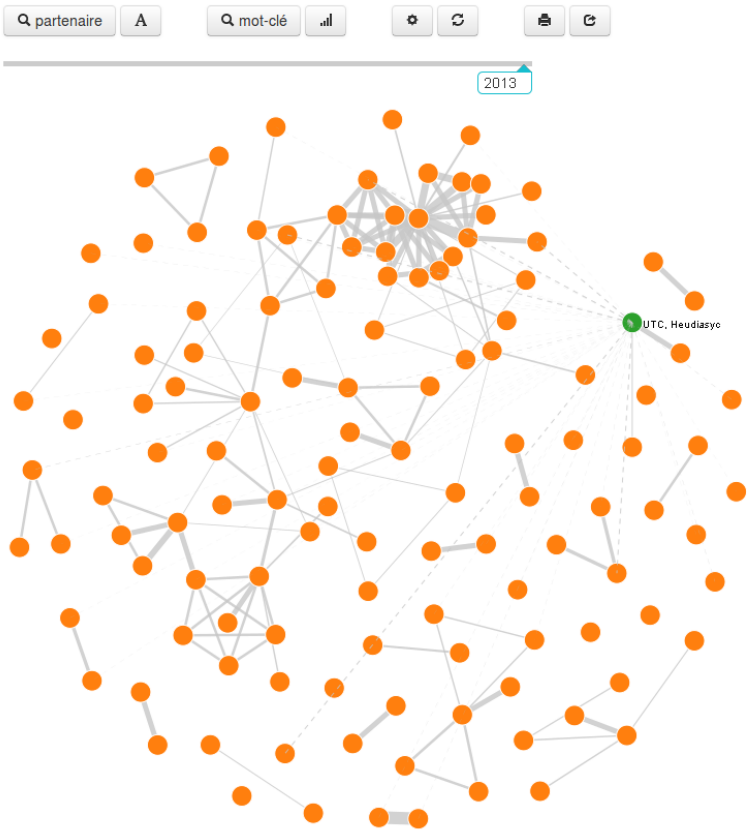


Figure 9.8 – Graphe des collaborations dans la communauté IC

les relations du laboratoire Heudiasyc de l'UTC. Après avoir sélectionné le nœud correspondant à ce laboratoire avec l'outil de recherche dédié, le cadran affichera un graphe similaire à celui de la figure 9.8. La vue actuelle du graphe de collaboration peut être exportée à tout moment vers le logiciel Gephi.

Les informations utilisées pour générer le graphe de la figure 9.8 ont été calculées par la brique inférentielle au fur et à mesure de l'ajout de nouvelles informations dans la base de connaissances. Chaque nœud du graphe représente une entité (telle que définie dans la section 8.4) et la taille des liens pleins reliant les nœuds deux à deux est définie par notre formule du *Edgerank* (section 9.3.1). L'algorithme de *Edgerank* a permis de calculer l'importance des liens factuels ou potentiels entre chaque organisation du réseau. Les liens potentiels apparaissent en pointillés et leur taille est calculée à l'aide de la formule (9.6) section 9.3.1.

Le graphe des collaborations lui permet également de visualiser les thématiques sur lesquelles ont porté ces collaborations. Si Denise souhaite observer l'évolution de ces différentes thématiques sur la période pour laquelle des données existent, elle utilise le cadran de suivi des tendances dont la figure 9.9 présente l'interface.

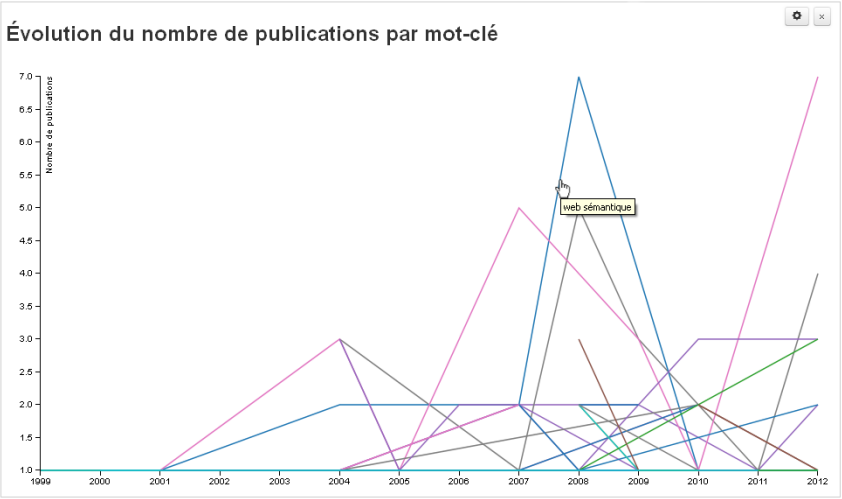


FIGURE 9.9 – Courbes d'évolution des concepts utilisés dans la base de connaissances

L'évolution des thématiques est fondée sur le suivi du nombre de ressources ayant été capitalisées par année, pour chaque concept du référentiel partagé. Chaque courbe représente donc le nombre de ressources en ordonnées, par année en abscisse. Il est possible de réduire le nombre de courbes affichées en même temps sur le diagramme. Le label du concept du référentiel partagé lié à chaque courbe s'affiche au survol de la souris sur sa courbe. Un clic sur l'une de ces courbes permet de mettre en valeur les collaborations ayant porté sur cette thématique dans le graphe des collaborations.

Une commande permet d'affiner l'étude du comportement des courbes du diagramme. Il s'agit d'une boîte de dialogue permettant de fixer la valeur des différents paramètres nécessaires à la détection de l'émergence d'une thématique par rapport à l'ensemble des concepts

observés. Cette boîte de dialogue est visible dans la figure 9.10. Cette détection de l'émergence suit l'heuristique décrite dans la section 9.3.2. Après avoir entré les paramètres voulus dans la boîte de dialogue, le cadran de suivi de tendances est rechargé et ne présente plus que les courbes d'évolution des concepts considérés comme émergents d'après notre définition. Le graphique présenté dans la figure 9.11 affiche les mots-clefs émergents de la communauté IC.

FIGURE 9.10 – Boîte à outils de paramêtrage de la détection d'émergence de tendances

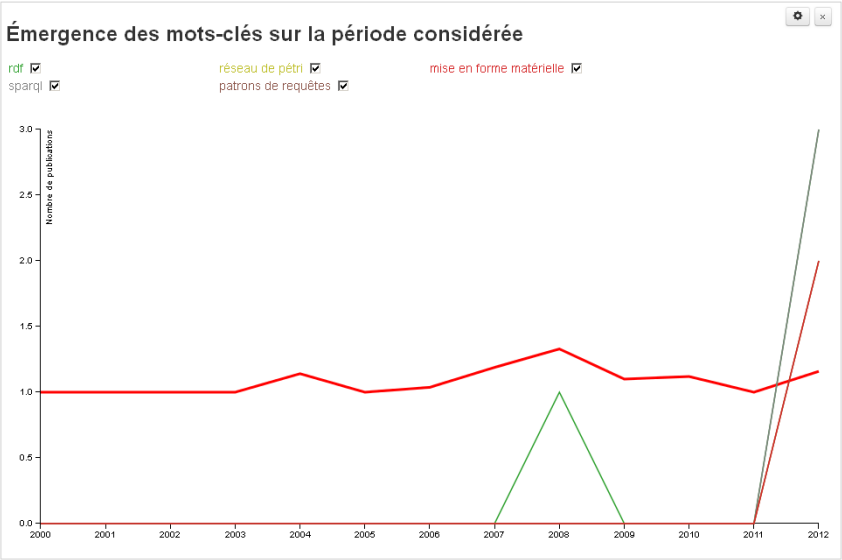


FIGURE 9.11 – Concepts émergents au sein de la communauté IC

Si maintenant Denise souhaite se concentrer, non plus sur l'utilisation générale des concepts du référentiel partagé, mais plutôt sur leur utilisation par un partenaire donné, il lui faut tout d'abord sélectionner ce partenaire au sein du graphe de collaboration. Ceci fait, le cadran des statistiques d'utilisation des concepts par une entité du graphe va s'ouvrir, présentant un diagramme circulaire permettant, année par année, de constater la part d'utilisation relative de chacun des concepts utilisés cette année là par cette entité (figure 9.12). La sélection de l'année s'effectue à l'aide de la tirette de sélection d'année du graphe de collaboration (visible sur la figure 9.8 au dessus du réseau de collaborations). L'utilisation de cette tirette va permettre à Denise d'observer à la fois l'évolution thématique d'une entité, mais également l'évolution de

ses relations.

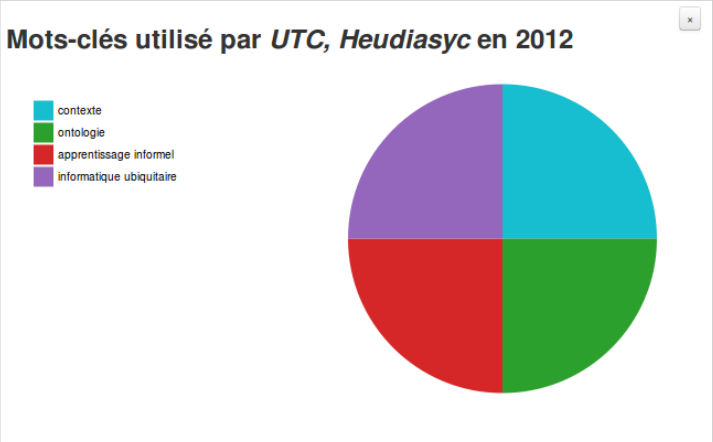


FIGURE 9.12 – Vue du cadran présentant les statistiques d'utilisation des concepts pour une organisation et une année données

Enfin, la sélection d'une entité au sein du graphe de collaboration permet également à Denise d'accéder à l'outil de génération de rapports de proposition de consortiums. Cet outil se sert à la fois des règles de modélisation d'appel à projet décrites dans la section 9.3.3 et du même calcul de la force des liens avérés ou potentiels entre deux entités du graphe de collaboration. Il s'agit de pouvoir générer un nouveau type de ressources à partir d'informations extraites de la base de connaissances.

Après avoir sélectionné le type d'appel à projet à surveiller et éventuellement réduit l'observation à un ou plusieurs concepts précis du référentiel partagé, Denise lance la génération du rapport. Ce dernier prend la forme d'une suite de paragraphes. La figure 9.13 présente l'interface de l'outil de génération de rapports, avec à gauche le formulaire permettant d'entrer les paramètres souhaités et à droite un exemple de rapport.

Chaque paragraphe d'un rapport est dédié à un type d'appel à projet et un concept particulier, parmi ceux sélectionnés. Le paragraphe rappelle tout d'abord les règles à respecter pour valider le montage du consortium et liste ensuite les entités issues du graphe de collaboration qui semblent pertinentes pour cet appel et ce concept. La notion de pertinence correspond à notre heuristique décrite dans la section 9.3.3. Le rapport peut ensuite être imprimé en partie ou en totalité. De même, chaque paragraphe (correspondant donc à une proposition particulière) peut également être sauvegardé en texte brut dans la base de connaissances.

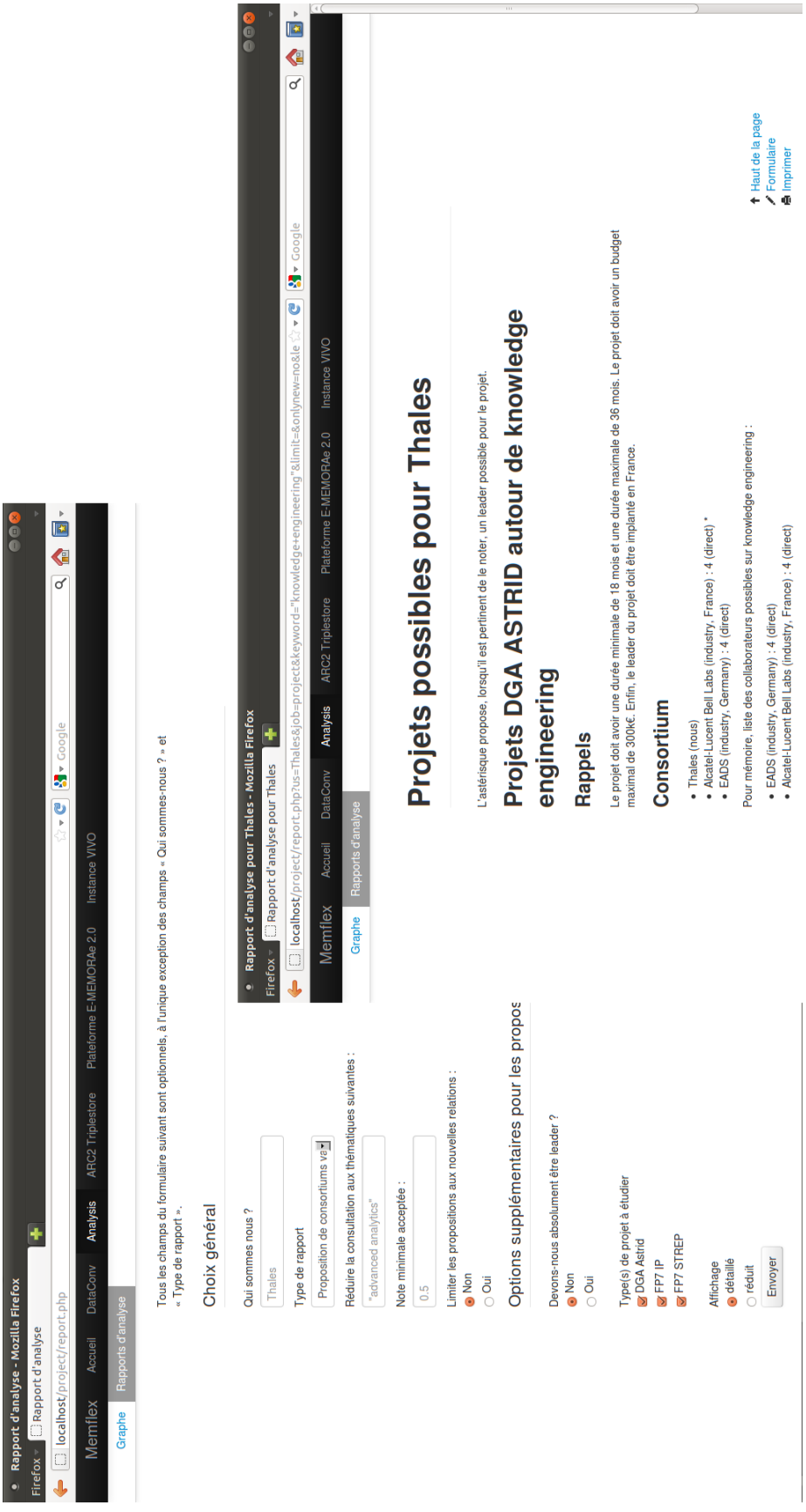


FIGURE 9.13 – Présentation de l'interface de génération de rapport de montage de consortiums

10. Synthèse

L'écosystème numérique réalisé dans le cadre de notre thèse permet aux décideurs d'une organisation d'accéder à différentes informations en adéquation avec leurs besoins. Cet accès s'effectue soit à l'aide d'une brique de gestion de connaissances permettant aux décideurs de naviguer parmi les données partagées par les membres de leur organisation, soit à l'aide d'une brique de visualisation leur permettant d'interagir de différentes manières sur ces données via les cadrans mis à leur disposition.

Ces deux briques permettent aux décideurs de prendre en considération à la fois des données documentaires et des données issues des médias sociaux déployés au sein de leur organisation. Ces données sont en effet enregistrées dans une même base de connaissances. Cette dernière peut être initialisée à l'aide d'un outil d'import de données, permettant de faire démarrer très tôt les heuristiques de la brique inférentielle. Ces différentes heuristiques permettent d'apporter un nouveau regard sur les données enregistrées dans la base de connaissances de l'organisation, que ce soit au travers de l'extraction du réseau des collaborations de l'organisation ou la génération de rapports pour aider au montage de consortiums.

Le fait d'avoir différent moyens pour accéder aux informations enregistrées au sein d'une seule et même base de connaissances permet de faire varier la matérialisation de ces données re-contextualisées selon les besoins en visualisation des décideurs. Ceci permet d'outiller leur appropriation des ressources ainsi générées par le système et la documentarisation éventuelle des connaissances acquises.

La partie III présente les résultats des expérimentations que nous avons mises en place pour valider notre modèle théorique et son implémentation au travers de l'écosystème numérique que nous venons de présenter.

Troisième partie

Tests et utilisations de notre modélisation et de nos réalisations

« *Prediction is very difficult, especially about the future.* »

— Attribuée à Niels Bohr

11. Introduction

Différents tests ont été menés pour évaluer le fonctionnement et la pertinence de notre plate-forme vis-à-vis de notre modèle théorique, ainsi que la qualité du support apporté aux décideurs d'une organisation.

La brique de gestion de connaissances a été évaluée en milieu universitaire afin d'étudier sa réception auprès du public dans le cadre d'un travail collaboratif. Les résultats de cette étude sont présentés dans le chapitre 12 et valident les hypothèses quant à l'intérêt et l'implémentation de l'approche suivie pour cette brique.

Une seconde campagne d'expérimentations a été menée pour tester le comportement de la brique inférentielle et de la brique de visualisation. Ces résultats sont présentés dans le chapitre 13.

Afin de mener a bien cette seconde campagne, nous avons tout d'abord cherché à constituer un jeu de données de test pour peupler la base de connaissances de la plate-forme. L'origine de ces données et leur contenu sont décrits dans la section 13.1. Les données ainsi collectées nous ont permis d'étalonner la brique inférentielle. Ce banc d'essais est décrit dans la section 13.2.

Cette seconde campagne s'est déroulée dans le cadre plus global de travaux de veille partenariale et concurrentielle :

— Une première étude s'est concentrée sur l'analyse temporelle des collaborations d'un département de recherche. Ces connaissances sont importantes pour la définition d'une stratégie de partenariat dans le montage de projets collaboratifs.
— Une seconde étude s'est focalisée sur un domaine de recherche particulier « la fusion d'information » et a cherché à identifier les thématiques émergentes de ce domaine et les acteurs les portant, afin de mieux comprendre l'écosystème de cette discipline.

Enfin, une dernière série d'évaluations, présentée dans la section 13.3 et prenant la forme d'entretiens auprès d'un expert et d'un chef de projets, a permis de tester la brique inférentielle au travers de l'expérimentation plus large de la brique de visualisation et de ses différents cadrans.

12. Utilisation de la plate-forme de gestion de connaissances

La plate-forme E-MEMORAe 2.0 que nous avons développée, support à un environnement de gestion de connaissances, a été utilisée par des étudiants dans le cadre d'un cours enseigné à l'UTC et portant sur les méthodes et outils de capitalisation de connaissances. Durant quatre mois au semestre d'automne 2012, une quarantaine d'étudiants a réalisé une veille technologique sur les thématiques des environnements de gestion documentaire d'entreprises et celle plus large du Web de données. Il leur a été imposé dans un premier temps de définir le périmètre de leurs recherches au moyen d'une ontologie. Celle-ci permet de définir les notions clés identifiées pour effectuer leur veille. Dans un second temps il leur a été demandé d'utiliser la plate-forme E-MEMORAe 2.0 pour collaborer, organiser et capitaliser au fil de l'eau le fruit de leurs recherches. Celles-ci pouvaient être effectuées grâce aux outils de leur choix.

L'ensemble des étudiants a été divisé en quatre groupes distincts, chaque groupe travaillant sur un sujet donné. Au sein d'un groupe, les étudiants devaient s'organiser pour réaliser leur veille et produire le rapport associé.

À cette fin, ils pouvaient se servir des fonctionnalités Web 2.0 pour échanger autour des ressources capitalisées et/ou des concepts de leur ontologie. Si l'organisation en groupes formels ne leur suffisait pas, ils avaient la possibilité de créer des groupes supplémentaires à leur convenance (membres et sujet libres).

À la fin du semestre un questionnaire leur a été donné, nous permettant d'obtenir des retours sur la manière dont ils avaient utilisé la plate-forme. Le questionnaire a la forme d'un formulaire en ligne. Les questions posées sont disponibles pour information dans l'annexe C.1. La participation à l'étude était libre et les réponses anonymes, aussi aucun entretien postérieur n'a été réalisé. Nous avons pu recueillir 21 témoignages.

Les questions étaient de trois formes : échelle de Likert — échelle d'évaluation sur 5 valeurs de 1 - pas du tout ou nul à 5 - beaucoup ou très intéressant —, question à choix multiple où une seule réponse était autorisée et question ouverte. Dans le cadre des questions graduées, les réponses présentées ci-dessous sont exprimées en pourcentage.

Les questions étaient réparties en cinq catégories : l'utilisation des outils de la plate-forme (12.1), le type des ressources partagées (12.2), le partage et l'indexation de ressources (12.3), l'ergonomie générale de la plate-forme (12.4) et l'intérêt global pour l'approche MEMORAe (12.5).

12.1 Utilisation des outils

Concernant les outils de la plate-forme E-MEMORAe 2.0, 48% des répondants ont utilisé le forum (réponses allant de 2 à 4), 52% le Wiki (réponses allant de 2 à 5) et 62% l'outil d'annotation (réponse de 2 à 5). Ils sont par contre 71% à avoir utilisé beaucoup (5) un traitement de texte externe à la plate-forme et 62% à avoir utilisé un outil de cartographie mentale (Freemind, etc.) pour construire leur ontologie (réponses de 3 à 5).

Un dernier champ libre leur permettait de nous faire connaître d'autres logiciels utilisés dans leur veille. La suite Google Docs en ressort très majoritaire.

12.2 Types de ressources partagées

Les réponses à cette section sont assez uniformes. Chacun des types proposés (articles académiques, articles encyclopédiques, articles d'opinions, supports de présentations) a été capitalisé par au moins 40% des répondants avec une valeur de jugement de 4 ou 5 (nombreuses fois à beaucoup). Seuls les livres scientifiques (ressources peu pratiques à lire en quatre mois et à partager numériquement) et les comptes-rendus de réunion ont été très peu, voire pas du tout (score de 1 ou 2) capitalisés par plus de 40% des répondants. Ce dernier point peut s'expliquer par une absence fréquente de rédaction de compte-rendu.

12.3 Partage des ressources

Les résultats montrent que les répondants ont très peu partagé de ressources entre différents groupes d'utilisateurs : 86% des répondants n'ont partagé leurs ressources que dans un seul groupe. L'indexation s'est révélée relativement pauvre : 67% des répondants déclarent n'avoir indexé des ressources que par un concept. Seuls 33% ont effectué une multi-indexation (au maximum par 3 concepts).

La durée finalement assez courte du semestre explique peut-être que les étudiants n'aient pas eu le temps ou l'opportunité de plus partager leurs ressources ou les indexer de manière plus large. Une étude supplémentaire sur une période plus longue permettrait de confirmer ou amender ce résultat.

12.4 Ergonomie générale

Cette catégorie présentait des questions ouvertes aux utilisateurs, dont voici la synthèse.

L'apparence actuelle de la plate-forme a semblé trop éloignée de ce que les étudiants ont l'habitude d'utiliser. L'absence d'une vue globale sur un flux d'activité — comparable au *Wall* de Facebook ou Yammer, à la *Timeline* de Twitter — et l'obligation de passer exclusivement par la cartographie pour accéder aux ressources ont souvent déstabilisé les étudiants. Au moment de la mise à disposition de la plate-forme, cette dernière ne permettait pas par exemple d'avertir les utilisateurs que de nouvelles activités avaient eu lieu autour des thématiques qu'ils suivaient.

Cette fonctionnalité de flux d'activité a depuis été ajoutée en permettant la mise en place d'alertes suite à l'indexation de ressources au sein de la plate-forme, ainsi que la mise en place du service de minimessage. Ce dernier permet de retrouver un environnement de type *Wall* sur lequel les utilisateurs échangent des informations, sans pour autant négliger la capitalisation de ces mêmes informations.

Les étudiants ont regretté par ailleurs la difficulté de maintenir convenablement la structure de la cartographie. Cette difficulté viennent principalement du fait que la plate-forme a été conçue originellement pour afficher une ontologie et non la concevoir.

Enfin, ils sont nombreux à avoir également regretté l'absence de possibilité d'avoir une vue globale sur leur ontologie. Cette limitation est liée à l'implémentation actuelle du prototype qui n'affiche la cartographie des concepts que de proche en proche et pour un type de relation donné. Elle devrait cependant être rapidement levée par la mise en place d'un second prototype utilisant de nouvelles technologies.

12.5 Intérêt pour l'approche

La plupart des avis que nous avons pu recueillir étaient très favorables à l'utilisation d'un outil sémantique pour organiser un travail de réflexion et d'analyse tel que les étudiants avaient eu à réaliser. Les répondants sont ainsi 57% à trouver la navigation sur la cartographie de concept intéressante, 67% à apprécier le fait de pouvoir se constituer en groupe libre et 81% à apprécier pouvoir capitaliser différents types de ressources selon le même référentiel (scores de 4 ou 5).

Enfin, bien que l'ergonomie générale de la plateforme ait pu les perturber (ils sont 71% à trouver que l'environnement n'est pas un bon outil de collaboration en l'état), ils sont 90% à avoir trouvé l'approche et son implémentation innovante.

13. Utilisation de la brique inférentielle et de la brique de visualisation

Afin d'initialiser notre outil d'analyse, nous avons constitué plusieurs jeux de données artificiels censés représenter des données amassées au fil du temps au sein d'un écosystème numérique utilisé dans le cadre d'un écosystème de connaissances. Ces données et leurs provenances sont décrites dans la section 13.1. Elles ont pu être importées au sein de la plate-forme prototype à l'aide de la brique d'import de données évoquées dans la section 9.1.

Suite à cet import, nous avons pu étudier le comportement de nos heuristiques sur ces données, comme relaté dans la section 13.2. Les résultats apportés par ces tests nous ont permis d'identifier les limites de l'approche suivie et d'envisager des améliorations.

Deux entretiens ont été menés au sein du laboratoire de raisonnement et d'analyse dans les systèmes complexes de Thales R&T à la suite d'une expérimentation de la brique d'analyse et de la brique de visualisation. Nous présentons dans la section 13.3 le compte-rendu de ces expérimentations et des *interviews* liées.

13.1 Sources de données utilisées pour initier la base de connaissances

Nous présentons dans cette section les différentes méthodologies utilisées pour récupérer ces données et leur format source à traiter. Ce traitement conduit à l'insertion des données après transformation dans une base de connaissances modélisée à l'aide de notre ontologie memorae-core 2.

Nous avons récoltés les trois jeux de données suivants :

Données de l'enquête STI : issues d'une enquête portant sur les collaborations de différentes équipes de travail, ces données mettent l'accent sur les liens sociaux pouvant unir ces mêmes équipes.

Données « fusion » : issues d'un travail de veille documentaire au sein de Thales, ces données sont plus fiables du point de vue de la thématique étudiée et plus « propres » dans leur forme.

Données Web of Knowledge : issues d'une grosse base documentaire, ces données permettent d'ouvrir le champ d'observation sur des collaborations externes à Thales. Elles permettent également d'appréhender des problématiques liées à la taille des données à traiter.

13.1.1 Données de l'enquête STI

Contexte

Un premier recueil des données a été effectué au sein du groupe de recherche en Sciences et Techniques de l'Information (STI) de Thales R&T. Ce groupe de recherche abrite quatre laboratoires de recherche différents. Le groupe de recherche a également à sa disposition deux experts assumant un rôle transverse.

Chaque laboratoire aborde différents thèmes fixés par la hiérarchie. Dans STI, ces thèmes peuvent être vus comme des équipes de travail sur une thématique au sein des laboratoires, avec toutefois pour particularité qu'un membre d'un laboratoire donné peut participer à différent thèmes tant que ceux-ci restent au sein de son laboratoire.

Le but de ce recueil de données est d'obtenir un graphe des collaborations entre les différentes équipes de STI mais également avec d'éventuels partenaires extérieurs. Il a ainsi été décidé d'interroger les responsables de chaque équipe afin qu'ils remplissent un questionnaire permettant de lister les dernières collaborations ayant impliqué leur équipe.

Ce questionnaire a la forme d'un fichier Excel dans lequel chaque ligne correspond à une collaboration différente et dont les colonnes permettent de décrire cette collaboration : définition des différents acteurs, forme et date de la collaboration, etc. L'explication détaillée du format est donnée dans la section 13.1.1 suivante.

Après soumission du formulaire aux quinze responsables d'équipe ou experts du groupe de recherche, huit formulaires ont été retournés, exprimant un ensemble de cent quarante collaborations impliquant cent seize partenaires (dont les équipes de STI).

Format du fichier de réponse

Nous considérons que les thèmes de recherche officiels du groupe de recherche STI peuvent être assimilés à des équipes de travail. Ces équipes se répartissent au sein de laboratoires et constituent notre élément de granularité le plus fin.

En effet, les données entrées ne sont pas nominatives. Les partenaires ne sont pas des individus mais des équipes, des laboratoires, des entités, des institutions appartenant potentiellement à des structures de plus haut niveau. Les champs de description des partenaires

sont inspirés de la structure hiérarchique interne de TRT. Cette structuration ne pouvant pas s'appliquer à tous les partenaires, les sondés étaient libres de laisser les champs inconnus ou ne s'appliquant pas vides.

Les « topics » sont les sujets officiels sur lesquels les effectifs d'un thème travaillent. Ils sont définis dans des notes d'organisation par la hiérarchie du groupe Thales.

Les « keywords » sont des concepts libres pouvant être scientifiques, techniques — algorithme, matériel, outil, etc. Le principe du « keyword » est de permettre de spécifier un sujet précis ayant lié deux partenaires ensemble lorsque l'intitulé des thèmes ou topics n'est pas représentatif. L'intitulé du keyword peut rester vide.

Un partenaire peut collaborer avec un thème sur un « topic » ou « keyword » selon différents vecteurs : *via* des échanges scientifiques informels, des publications, des tâches dans des projets, des thèses, des stages, des brevets, du développement logiciel, etc. Pour certains vecteurs — publication, juridique, études et échanges — un champ « type » permet de les préciser. Enfin il était possible, si aucune description de vecteur n'était appropriée, d'utiliser sa propre dénomination.

Une collaboration possède une temporalité précisée dans les dernières cases de chaque ligne. Une date de début et de fin doit être entrée. Si la collaboration est toujours d'actualité, la date de fin peut être omise. S'il s'agit d'une collaboration ponctuelle — typiquement une rencontre informelle en marge d'un colloque — la même date devait apparaître dans les deux cases. Le format de date n'était pas contraint.

Chaque ligne du tableau représente un partenariat précis entre un thème STI et une autre équipe sur un certain « keyword », au cours d'une période donnée et selon un certain vecteur. Si le thème collabore autour du même « keyword » mais avec différents partenaires, le fichier contiendra autant de lignes que le thème a de partenaires sur ce « keyword ». Idem, si la collaboration avec un partenaire prend différentes formes (vecteurs), plusieurs lignes lui seront affectées. Les tableaux ci-après représentent un exemple de ligne pouvant apparaître dans le tableau.

Groupe	Laboratoire	Thème	Topic	Keyword	…
STI	LRASC	xx	xx	social network	…

…	Topic	Thème	Laboratoire	Service	Département	Entité	Institution	…
…	KM	ICI	HEUDIASYC	/	Informatique	/	UTC	…

…	Vecteur	Type	Date début	Date fin
…	Études	Thèse	10/2010	10/2013

13.1.2 Données « fusion »

Les données fusion se présentent sous la forme d'un fichier Excel servant à maintenir à jour une bibliographie centrée sur la thématique de la fusion d'informations. Le format utilisé dans ce fichier n'exprime pas directement des liens de collaboration. En effet, les lignes définissent seulement que telle entité a publié telle ressource sur tels sujets à telle date. Cependant, comme le fichier est exhaustif sur les différents auteurs d'une ressource, pour une même ressource plusieurs lignes peuvent exister : une ligne par auteur ou créateur de la ressource. Ce faisant, en ré agrégeant les auteurs autour des ressources, il est possible d'identifier des actes de collaboration.

Chaque ligne du tableau représente donc un acte de publication de la part d'un acteur identifié à l'aide des colonnes « Legal Entity » et « Country ». Ce format ne permet pas d'obtenir la même granularité que les données provenant de l'enquête STI décrites dans la section 13.1.1.

Les publications sont décrites à l'aide des colonnes :

IP Type Type général de la publication, comme par exemple « conférence », « journal », « produit », etc.

Publication media Cadre général au sein duquel la publication est parue, comme le nom du journal, de la conférence, etc.

Title Titre de la publication considérée.

Publication year Année de la publication.

URL Lien vers un fichier informatique représentant la publication.

Enfin, chaque publication est étiquetée à l'aide de concepts au sein de deux colonnes différentes :

Keywords mots-clés libres ;

Topic classification officielle au sein de Thales.

Les tableaux ci-après représentent un exemple de ligne pouvant apparaître dans le tableau Excel.

Keyword(s)	Topic	Fusion Level	Legal Entity	Country	…

…	IP Type	Publication year	Publication media	Title

Après un premier traitement du fichier permettant de reconstituer les collaborations exprimées via la participation à une même ressource, nous avons pu extraire 1014 collaborations impliquant 513 partenaires.

13.1.3 Données Web of Knowledge

Une dernière campagne de recueil de données a été effectuée à l'aide du moteur de recherche *Web of Knowledge*[1]. Comme pour les données fusion décrites dans la section 13.1.2, il s'agissait de réunir une bibliographie sur la thématique fusion. La requête passée sur le moteur de recherche interrogeait spécifiquement les champs « Topic » et « Title » des ressources indexées sur les mots clés suivants : *information fusion*, *semantic fusion*, *decision fusion*, *soft fusion*, *information correlation* et *information retrieval*.

Parmi les différents fonds documentaires existants, nous avons choisi *Web of Knowledge* car c'est un des rares fonds qui présente à la fois une base documentaire suffisamment garnie et la possibilité d'exporter, entre autres informations sur les ressources, l'affiliation des auteurs, nécessaire du fait que notre chaîne de traitement ne travaille que sur les organisations et non sur les humains.

Le format d'export de *Web of Knowledge* est basé sur le format RIS[2] qui se présente sous la forme d'un fichier texte déroulant un couple clé-valeur à chaque ligne.

Après un premier traitement du fichier permettant de reconstituer les collaborations exprimées via la participation à une même ressource, nous avons pu extraire 12341 collaborations impliquant 11144 partenaires.

13.2 Tests de calibrage de la brique inférentielle

Afin de vérifier le bon fonctionnement de notre brique inférentielle, une série de tests a été réalisée. Ces tests avaient pour but de mesurer la robustesse de la solution technique utilisée. Nous avons organisé nos données de test en quatre groupes distincts :

— les données issues de l'enquête STI seules (section 13.1.1) ;
— les données issues de la bibliographie « fusion » seules (section 13.1.2) ;
— les données issues de l'export *Web of Knowledge* seules (section 13.1.3) ;
— l'ensemble des trois sources de données ci-dessus réunies.

13.2.1 Test de robustesse

Pour chacun des quatre groupes précités, le test de robustesse a consisté à recueillir différentes valeurs liées à l'exécution des heuristiques sur notre serveur. L'analyse des médias

1. Grand fond documentaire édité par Thomson Reuters et accessible à l'adresse http://isiknowledge.com le 10 novembre 2013.

2. format développé par la société Reference Manager (http://www.refman.com/support/risformat_intro.asp, site accessible le 10 novembre 2013) et personnalisé par Thomson Reuters sous le nom d'*ISI Export Format* ou CIW.

sociaux implique la manipulation d'énormes quantités de données (les fameuses *big data*). Il convenait donc de s'assurer si notre solution, en l'état, pouvait être utilisée en conditions réelles ou pas.

La base de connaissances de notre prototype repose sur l'utilisation d'un triple-store, afin de bénéficier des avantages de notre modélisation ontologique. S'agissant d'un prototype, notre choix s'est rapidement porté sur le triple-store ARC2 [3]. Ce dernier, loin d'être une solution intégrée et optimisée pour de grands jeux de données, a l'avantage de reposer sur les mêmes briques technologiques que le reste de notre prototype (à savoir PHP et MySQL), ce qui a grandement facilité son déploiement.

Nous avons alors mesuré trois grandeurs représentatives de l'efficacité de nos heuristiques au quotidien :

— Temps de traitement des fichiers de données originaux pour le convertir en OWL afin de les importer dans la base de connaissances ;
— Temps de traitement des données de la base de connaissances pour calculer le *Edgerank* des relations entre les différentes organisations de la base de connaissances ;
— Et en ce qui concerne la brique de visualisation, le temps d'affichage du graphe calculé.

L'ensemble des mesures a été réalisé sur une machine fonctionnant avec la distribution GNU/Linux Ubuntu 10.10. Il s'agit d'une version 32 bits du système qui exploite un processeur Intel Core 2 Duo E7500 et 1,9Gio de mémoire vive. Les logiciels impliqués dans le fonctionnement du prototype sont Apache en version 2.2.16, PHP en version 5.3.3, MySQL en version 5.1.61 et ARC2 à la version du commit d08f8fb8ed. Tout ces logiciels sont utilisés dans leur configuration standard, sauf PHP dont nous avons modifié les variables suivantes afin d'empêcher l'arrêt prématuré de nos heuristiques pour cause de dépassement mémoire ou temporel :

— `max_execution_time = 0`
— `memory_limit = 1024M`

Nous obtenons alors les temps reportés dans le tableau 13.1. Nous avons replacé dans ce tableau la taille des différents jeux de données pour comparaison. Les temps sont donnés en secondes et en minutes'secondes. La case « nombre de liens considérés » est à comprendre au sens des *edges* développés dans la section 9.3.1, c'est-à-dire qu'une relation entre deux partenaires (ou deux nœuds dans le graphe) est la résultante d'un ensemble de liens. Chaque *edge* ou lien représente une trace de collaboration (ici il s'agit quasiment à chaque fois d'un papier coécrit). La ligne « temps d'extraction du graphe » exprime le temps nécessaire à la brique inférentielle pour calculer un score de *Edgerank* pour l'ensemble des relations (pouvant donc agréger plusieurs liens) unissant deux à deux des partenaires de la base de connaissances.

3. Site officiel et de suivi de développement du projet, accessible le 10 novembre 2013 : https://github.com/

TABLE 13.1 – Temps d'exécution de nos heuristiques sur les différents jeux de données

Valeurs	**Données enquête STI**	**Données étude fusion**	**Données *Web of Knowledge***	**Données totales**
Nombre de partenaires	116	513	11144	/
Nombre de liens considérés	453	1014	12341	/
Temps de conversion source vers OWL	3s	14s	2530s (42'10)	/
Temps d'extraction du graphe	19s	152s (2'32)	/	/
Temps de rendu du graphe (brique de visu)	11s	33s	/	/

Après un certain nombre d'essais successifs, il s'est avéré qu'il était impossible d'obtenir une réponse dans un temps acceptable (inférieur à 3 heures) en ce qui concerne l'extraction du graphe lors de l'étude du jeu de données issues du portail *Web of Knowledge*.

De même, les tentatives de concaténation de l'ensemble des données en notre possession n'ont pas abouti, ce qui explique la dernière colonne vide du tableau.

13.2.2 Test de pertinence des résultats

Afin de tester la qualité de notre heuristique de prédiction d'émergence des concepts, nous avons suivi la procédure de test suivante. Pour les deux premiers jeux de données précédemment décrits, nous avons tout d'abord identifié les concepts considérés comme émergent la dernière année pour laquelle nous avions des données. Notre précision temporelle sur tous

semsol/arc2/wiki

nos jeux de données est en effet de l'ordre de l'année. Pour le jeu de données provenant de l'enquête STI, cette détection a donc consisté à identifier les concepts émergeant en 2012, du fait que nous possédions des données allant de 2004 à 2012.

Dans un second temps, nous avons sciemment réduit les données en notre possession d'un an. Ainsi, le calcul ne pouvait plus s'effectuer, pour le jeu de donnée STI, que sur la période 2004–2011. Nous avons alors demandé à notre heuristique de prédire les concepts pouvant émerger en 2012, en ne nous servant que de ce nouvel échantillon de données amputé d'un an.

Table 13.2 – Résultats de l'heuristique de prédiction de l'émergence de concepts

Valeurs	**Données enquête STI**	**Données étude fusion**
Période complète T d'observation	2004–2012	2007–2012
Nombre de concepts détectés émergents à T_{max}	2	25
Nombre de concepts prédits émergents à partir de $T_{max} - 1$	13	tous
… dont nombre de faux positifs	12	/
… dont nombre de faux négatifs (concepts non prédits)	1	0
Nombre de concepts prédits émergents à partir de $T_{max} - 2$	17	tous
… dont nombre de faux positifs	16	/
… dont nombre de faux négatifs (concepts non prédits)	1	0

Enfin, dans un troisième temps, nous avons encore réduit d'un an nos jeux de données pour effectuer une prédiction sur deux ans. L'ensemble des résultats obtenus ont été reportés dans le tableau 13.2.

Dans le cadre de l'enquête STI, un seul des deux concepts ayant effectivement été mesurés comme émergents lors de l'analyse sur les données complètes a pu être prédit. Et ce, même en reculant à $T_{max} - 2$ la date d'arrêt de prise en compte des données.

Dans le cadre des données « fusion », la mauvaise qualité des résultats, à commencer par le grand nombre d'émergences détectées la dernière année, provient sans doute de la nature des données. Le nombre de ressources exploitables (et donc naturellement le nombre de concepts) augmente en effet avec les années.

13.3 Expérimentation de la brique de visualisation

Suite aux vérifications de bon fonctionnement de nos heuristiques, nous les avons testées auprès de deux collaborateurs de Thales R&T. Plus exactement, nous avons présenté et testé la brique de visualisation, cette dernière utilisant directement les heuristiques de la brique inférentielle et leurs résultats. Cette expérimentation a pris la forme d'un entretien oral avec chacune des deux personnes interrogées. Cet entretien a systématiquement eu lieu après une démonstration des différentes fonctionnalités de la plate-forme. Bien que la discussion au cours de l'expérimentation ait été très libre, toute l'expérience suivait un protocole établi à l'avance (disponible dans l'annexe C.2). Les personnes interrogées n'ont cependant jamais eu accès à ce protocole qui restait dans la main de l'interrogateur.

Nous souhaitions, à l'aide de ces entretiens, valider la pertinence de l'outil et des heuristiques sous-jacentes dans :

— un travail de veille technologique et concurrentielle ;
— un processus décisionnel lié à l'établissement d'une stratégie concurrentielle ;
— un processus décisionnel lié au montage de projet.

Nous avons alors identifié deux types de profils au sein de l'établissement où nous sommes intervenus dans le cadre de cette thèse :

— le profil chef de projet, qui intervient lors de la phase de montage d'un projet cadre comme les projets européens, ANR, ou DGA et en particulier lors de l'établissement de la proposition de consortium ;
— le profil expert technologique, qui intervient à différents niveaux dans la stratégie globale de l'établissement et dont les activités de veille technologique et concurrentielle sont primordiales pour l'organisation.

Nous présentons dans les deux sous-sections suivantes le compte-rendu de chacune des *interviews* réalisées.

13.3.1 Entretien avec un chef de projet

Le premier utilisateur interrogé est chef de projet senior au sein du laboratoire de raisonnement et analyses dans les systèmes complexes et, à ce titre, monte et gère plusieurs projets en parallèle. Il est également souvent mobilisé lors de réunions de lancement d'appel à projets pour représenter les intérêts du laboratoire et identifier la pertinence de l'appel quant à ces mêmes intérêts.

L'expérimentation s'est déroulée en deux parties. Il s'agissait dans un premier temps de faire une démonstration de l'outil, avant que l'entretien en lui-même ne débute. L'ensemble

de l'expérience s'est passé dans le bureau de l'utilisateur, sur son poste de travail, la brique de visualisation étant accessible à l'aide d'un simple navigateur Web.

Présentation de l'outil

La présentation de la brique de visualisation et de ses fonctionnalités a duré quarante minutes. Au cours de cette présentation, quelques bugs logiciels ont été remarqués (incompatibilités liées au navigateur utilisé, problèmes de rechargement du graphe) ainsi que des problèmes ergonomiques (labels des nœuds ou courbes mal positionnées, incohérence du mécanisme de recherche d'un cadran à l'autre, choix des couleurs pour mettre en valeur les concepts dans le graphe de collaboration, manque de documentation intégrée au logiciel).

Au-delà des problèmes logiciels simples, trois demandes ont été formulées pour améliorer la brique de visualisation et par extension les heuristiques mises en place.

— Dans l'outil de génération de rapport de montage de consortium, présenter les organisations candidates sous la forme d'un graphe de collaboration et non sous la forme d'une liste. Cette demande touche exclusivement la brique de visualisation, sans modifier les heuristiques de calcul des liens de collaboration.
— Modifier le comportement de la recherche et de la mise en valeur des concepts au sein du graphe de collaboration ou du cadran de suivi de tendance. L'heuristique actuelle cherche en effet à synthétiser au mieux la recherche formulée. Si l'utilisateur entre « fusion » par exemple, l'heuristique va agréger sous une seule thématique « fusion » toutes les informations indexées par « information fusion », « soft fusion » ou encore « radar fusion ». Il est alors impossible de connaître en un coup d'œil les composants de cet agrégat, autrement qu'en procédant à un inventaire manuel en fonction des informations récupérées. L'idée est alors de présenter à l'utilisateur l'ensemble des concepts identifiés et les informations liées, sans les agréger. Cette demande nécessite des modifications aux heuristiques utilisées dans la brique inférentielle.
— Revoir la présentation du cadran de détection d'émergence de thématique. Le mode de prévision automatique, tel qu'il est implémenté est inutile. Par contre, le chef de projet aurait souhaité pouvoir visualiser facilement les réponses à des questions formulée comme « Sur la période de temps comprise entre telle et telle année, est-ce que la thématique X peut-être considérée comme émergente ? ». Cette demande nécessite, en ce qui concerne le mode de prévision automatique de l'émergence, des modifications aux heuristiques utilisées dans la brique inférentielle. En ce qui concerne les questions sur le suivi d'une thématique donnée, tant que la période demandée se situe à l'intérieur des bornes fixées par les données disponibles, alors la demande ne nécessite pas de modification particulière. Néanmoins, si la période dépasse ces mêmes bornes, la remarque précédente sur le mode de prévision automatique s'applique.

Par ailleurs, en rebondissant sur sa demande concernant l'amélioration du fonctionnement de la recherche et la mise en valeur de thématiques au sein de la brique de visualisation, le chef de projet a insisté sur la nécessité d'utiliser une ontologie, ou au moins une taxonomie, plutôt que des mots-clés libres pour définir l'ensemble des thématiques abordées par les différents partenaires du réseau. Il a souligné le fait que cela permettrait d'effectuer des recherches à des niveaux de précisions divers, en se servant des relations de spécialisations entre les termes du référentiel utilisé. Par exemple, commencer par étudier le réseau de collaboration portant sur la thématique de fusion, avant de passer à l'étude du réseau portant sur la thématique de soft fusion, cette dernière thématique étant un raffinement de la thématique générale de fusion.

Prise en main de l'outil

La seconde partie de l'entretien a consisté par une prise en main de la brique de visualisation par le chef de projet interrogé. Cette prise en main était dirigée par l'intermédiaire de questions que l'interrogateur lui posait.

La première série de questions servait à mesurer la compréhension des différentes fonctionnalités de navigation dans le temps et dans l'espace au sein du cadran affichant le graphe des collaborations. Il s'agissait d'identifier d'une part des relations anciennes entre partenaires, d'autre part des relations récentes mais fortes et enfin des relations régulières. La subtilité provenait du fait que ces trois types de relations pouvaient s'exprimer de prime abord sous la forme d'un lien plus fort entre deux partenaires et que seule l'utilisation du curseur de navigation dans le temps permettrait de différencier ces différents types de relations. Bien qu'ayant formellement identifié ces trois types de relation comme étant les liens les plus forts entre partenaires, le chef de projet n'a pas pu expliquer une méthode permettant d'identifier plus formellement chaque type de relation demandé. Il n'a pas utilisé le curseur de navigation temporel.

Une seconde série de questions portait sur l'utilisation de la mise en évidence de thématiques et de partenaires dans le graphe. Cette série de questions n'a pas posé de problème particulier, le chef de projet ayant réussi du premier coup à mettre en valeur la thématique demandée et à nommer des partenaires ayant travaillé sur cette thématique, mais pas ensemble.

La dernière question liée à l'utilisation du cadran de visualisation des thématiques émergentes a posé problème. Il s'agissait de constater si une thématique pouvait être considérée comme émergente et, si oui, à quelle date. La personne interrogée, bien qu'ayant compris qu'il fallait se servir de ce cadran, n'a pas réussi de prime abord à le paramétrer correctement, pensant qu'il fallait d'abord isoler la thématique en question sur le graphe de collaboration. Il a cependant réussi à répondre à la question au bout de quelques tâtonnements. Il s'agit ici clairement d'un problème ergonomique qui n'avait pas été envisagé au préalable.

La série de question concernant l'outil de génération de rapport de montage de consortium n'a pas posé de problème particulier.

Habitudes et autres outils utilisés

La dernière série de questions devait permettre d'en connaître plus sur les habitudes de la personne interrogée. À la question de donner les sources habituelles permettant d'identifier de potentiels partenaires pour des projets technologiques ou commerciaux, le chef de projet nous a cité, par ordre d'importance :

1. Les experts de son organisation ;
2. Les salons professionnels ;
3. Les précédentes propositions de son organisation en réponse à des appels à projets ;
4. Les avis de différents partenaires connus ;
5. Les chambres de commerces et d'industries.

L'entretien s'est terminé par une demande d'impression générale sur la plate-forme, telle qu'elle avait été présentée, manipulée, ressentie. La principale impression du chef de projet interrogé est que cet outil répond bien à un souci qu'il rencontre régulièrement, à savoir pouvoir identifier un partenaire potentiel pour compléter un possible consortium dans le cadre d'une réponse à un appel à projet. En ce sens l'expérience a été très positive. Il regrette du coup de ne pouvoir plus interagir avec le système pour entrer des paramètres supplémentaires afin d'affiner les entités présentées, que ce soit au sein du graphe des collaborations comme de l'outil de génération de rapport, par exemple pour n'observer que ses concurrents, ou qu'un type particulier de partenaire (PME, académique, etc.). Il a enfin pointé du doigt les faiblesses de l'outil et en particulier le manque d'informations concernant les calculs effectués ou les fonctionnalités disponibles qui ne favorisent pas la confiance dans les résultats affichés, du fait qu'il était difficile d'en comprendre la provenance.

13.3.2 Entretien avec un expert

Le second entretien s'est déroulé avec un expert technologique du laboratoire de raisonnement et analyses dans les systèmes complexes. Cet entretien devait se dérouler, comme le précédent, dans le bureau de la personne interrogée, directement sur son ordinateur, mais suite à une incompatibilité de version du navigateur internet, l'entretien s'est finalement effectué à l'aide de la machine utilisée pour le développement de la plate-forme.

Présentation de l'outil

La présentation des différentes fonctionnalités s'est déroulée de la même manière que le premier entretien. Les remarques formulées par l'expert, identiques au premier entretien ne sont pas rediscutées ici. L'expert a porté une attention particulière à l'outil de visualisation d'émergence de thématiques. Cette personne a exprimé le souhait de pouvoir obtenir différentes formes de réponse pour cet outil. En plus du graphe qui permet de répondre à la question « quelles sont les thématiques ayant émergé en telle année », elle aurait souhaité obtenir un rapport listant, en une fois et par année des thématiques ayant émergé sur la période considérée. Enfin, l'expert aurait souhaité pouvoir poser directement au système une question du type « En quelle année telle thématique a-t-elle émergée » et obtenir en réponse « jamais » ou une date particulière.

Prise en main de l'outil

La première série de questions de type exercice permettant la prise en main par la personne interrogé de l'outil ne lui a pas posé de problème particulier. À la différence du chef de projet interrogé précédemment, l'expert s'est souvenu de la tirette de navigation dans le temps et s'en est servi correctement pour identifier des relations anciennes ou régulières entre entités sur le graphe des collaborations.

Habitudes et autres outils utilisés

La série de questions ouvertes clôturant l'entretien a permis de relever, comme pour le premier entretien, les différentes sources actuellement consultées pour chercher des partenaires potentiels sur des projets de recherche ou industriels. Les sources principalement consultées par l'expert sont ainsi :

1. les fiches de renseignements mises à disposition par l'Union Européenne,
2. les publications scientifiques et les brevets déposés par les partenaires ou concurrents,
3. les activités de réseautage de l'expert, entre autre grâce aux réunions de lancement des appels à projets.

Par ailleurs, le manque le plus criant relevé par l'expert sur les outils actuellement disponibles pouvant l'épauler dans sa veille sur les activités de ses domaines d'expertise est le manque d'automatisme dans la collecte des données. L'expert regrette que les outils ne soient pas capables de s'alimenter seuls auprès, par exemple, des différents organismes ouvrant des appels à projets, afin de lui apporter les informations les plus fraîches. L'expert a en ce sens apprécié l'idée de la base de connaissances partagée par tous les membres de l'organisation permettant une meilleure couverture des actualités.

14. Synthèse

Nos expérimentations concernant les trois briques de notre plate-forme ont amené des résultats complémentaires pour son évaluation tant quantitative que qualitative. Différentes dimensions de la plate-forme ont pu être testées : son utilisabilité, son fonctionnement et son usage.

L'expérimentation effectuée dans le cadre universitaire nous a permis de mesurer l'acceptabilité de l'approche proposée pour la gestion de connaissances. Si l'approche a intéressé les utilisateurs ou leur a plu, ils ont regretté que celle-ci se démarque trop fortement des autres plates-formes utilisées habituellement.

Le calibrage des briques inférentielle et de visualisation a nécessité l'élaboration d'un jeu de données d'expérimentation pour pallier le manque d'utilisation de la brique de gestion de connaissances et approvisionner la base de connaissances. Les données utilisées pour cette études s'intéressaient aux collaborations pouvant exister au sein d'un écosystème industriel. Ces collaborations ont été relevées à l'aide d'une enquête auprès de décideurs de Thales R&T et d'une extraction de données bibliographiques.

L'utilisation de grands jeux de données nous a conduit à considérer le problème de qualité des données — problème qui se serait également manifesté, sinon plus, en utilisant des données issues automatiquement de médias sociaux. Par ailleurs, nous nous sommes aperçu au cours de l'élaboration de nos jeux de données que la valeur temporelle des documents numériques est discutable. Il est courant qu'une publication scientifique relate des faits s'étant produit six mois ou plus auparavant. La datation effective des travaux devient alors ambiguë, de même que la détection de travaux parallèles de deux organisations sur les mêmes thématiques. C'est pour encadrer cette datation que l'appui sur des données issues de médias sociaux, plus précisément datées, aurait été appréciable.

Enfin, nos choix techniques de prototypage rapide à l'aide du langage PHP ont effectivement permis la mise en place rapide d'applications Web pour tester nos propositions, mais se sont également révélés contraignant en termes de capacité de calcul. Les résultats de la brique inférentielle et de la brique de visualisation mettent en évidence des goulots d'étranglement à dimensionner et prendre en compte pour faire passer le prototype à l'échelle :

— puissance du processeur et mémoire vive allouée à PHP sur le serveur pour le calcul par la brique inférentielle du *Edgerank* sur de grands graphes ;
— utilisation d'un triple-store plus robuste qu'ARC2 pour les grands jeux de données pour s'abstraire, à ce niveau là déjà, des limitations liées à l'utilisation de PHP et MySQL ;
— puissance du processeur et de la mémoire vive disponible au navigateur internet sur le poste client pour améliorer le rendu des grands graphes par la bibliothèque javascript d3js.

Néanmoins, les entretiens réalisés pour évaluer la brique de visualisation et à travers elle la brique inférentielle ont été très positifs. Les décideurs interrogés ont souligné l'intérêt de l'outil proposé pour permettre la mise en évidence de partenaires potentiels pour des projets collaboratifs. Cette mise en évidence est calculée à partir des données de tous ordres ajoutées à une base de connaissances.

Notre positionnement autour de la notion de document numérique et de ressource et la manière dont nous avons modélisé les différents types de ressources nous permettent d'utiliser toutes sortes de données. L'utilisation d'heuristiques conçues pour étudier des phénomènes sociaux sur différents types de données, dont celles issues de bases documentaires et bien sûr celles issues de médias sociaux, apporte des résultats intéressants quant à l'extraction d'une vue globale du réseau de collaboration d'une organisation et de son évolution dans le temps. Cette extraction, et sa mise en forme de manière à pouvoir être présentée à des décideurs, a fortement intéressé les décideurs interrogés qui y voient un moyen d'accéder facilement à de nouvelles connaissances qu'ils n'acquéraient auparavant qu'au prix d'un long travail manuel.

Les retours obtenus lors de nos entretiens auprès des décideurs de Thales nous permettent de valider l'idée que les ressources issues des médias sociaux peuvent être sources de connaissances, en l'occurrence en apportant une vision claire sur un réseau d'organisations.

Quatrième partie

Conclusion et perspectives

« I see little commercial potential for the Internet for the next 10 years. »

— Attribué à Bill Gates lors du COMDEX de 1994

15. Conclusions

Nous avons montré dans cette thèse que les échanges effectués au sein des médias sociaux d'organisations peuvent être source de connaissances pour les organisations. Nous avons inscrit cette création de nouvelles connaissances dans le cadre de la prise de décisions en montrant de quelle manière ces connaissances peuvent nourrir un processus de prise de décisions.

Suite à la grande démocratisation des technologies numériques et en particulier des nouveaux moyens de communication et collaboration, le Web a évolué en un environnement de production et de consommation de contenus en masse. Cette production et cette consommation de contenus s'effectuent désormais principalement au travers de sites Web, comparables à des applications, permettant le renforcement d'une certaine forme de lien social en facilitant la rencontre numérique de communautés de pratique ou de communautés épistémiques. Il s'agit des médias sociaux.

Les organisations s'intéressent grandement à cette évolution du Web et l'apparition des médias sociaux qui leur permettent d'améliorer la collaboration de leurs membres en facilitant les échanges d'informations. Les organisations espèrent pouvoir accéder aux informations circulant sur les médias sociaux de manière à capitaliser ce savoir qui échappe encore à leurs bases de connaissances. La capitalisation de ces connaissances permettrait de nourrir leurs SIAD d'une nouvelle forme de connaissances. Cependant le domaine de la gestion des connaissances opère sur des modèles adaptés à la gestion documentaire qui ne sont plus adaptés à la gestion des connaissances pouvant être tirées des médias sociaux. Il nous a semblé nécessaire d'étudier un moyen de rendre accessible la connaissance résidant au sein des médias sociaux d'une organisation. Pour ce faire, nous avons travaillé sur les différentes notions de donnée, information, ressource, document et connaissance et la création de manière homogène de connaissances à partir des données disponibles sur les médias sociaux et les bases documentaires des organisations.

Nous avons postulé que cette création de connaissance est issue d'un processus personnel à chaque décideur des organisations. En s'appropriant les différentes ressources à sa disposition, un décideur va les interpréter et en tirer des connaissances. Pour nous, la documentarisation est ce processus d'appropriation/interprétation. Nous considérons que l'appropriation des ressources provenant de fonds documentaires ou de médias sociaux peut s'effectuer, au

sein d'un environnement numérique, de différentes façon via l'annotation ou le commentaire de la ressource ou encore son indexation à l'aide de concepts choisis. Ce faisant, le décideur affecte aux ressources considérées le statut de document les différenciant, à ses yeux, des autres ressources du système.

Nous avons également vu, lors de notre exploration de la définition des documents numériques, que ces derniers pouvaient être considérés comme la trace d'une collaboration. Nous nous servons de cette approche pour déterminer le graphe social d'une organisation. Cette tentative semble d'autant plus pertinente que les nouveaux outils numériques sociaux que sont les médias sociaux investissent progressivement les organisations. Ces outils sont la source d'énormes quantités de données qui ne sont pas, ou mal, réutilisées par les organisations. Nous avons donc proposé un modèle théorique permettant de répondre aux besoins de représentation et de gestion des connaissances sociales et documentaires des organisations.

Ce modèle passe par la définition d'un cycle de vie des données proposant différentes étapes les faisant évoluer jusqu'à la création de nouvelles connaissances. Nous nous sommes attaché, avec ce cycle de documentarisation, à définir la contextualisation des données en informations qui, une fois mises en forme, matérialisées aux yeux des utilisateurs sous la forme de différentes ressources numériques, pourront être appropriées par ces mêmes utilisateurs. Une fois documentarisée, la ressource devenue document est appelée à être sauvegardée dans la base de connaissances de l'organisation. Le document peut alors être vu comme une donnée qui, contextualisée, apporte un sens particulier (la connaissance) à l'utilisateur qui est libre de lui-même la percevoir en tant que document.

Lorsqu'elles sont issues de processus sociaux tels que la communication ou la collaboration, les ressources numériques documentarisées sont riches d'enseignements concernant le graphe social des parties prenantes. Le graphe social d'une organisation peut s'apprécier à différents niveaux, de l'individuel au structurel, sous la forme de groupes ou communautés de tailles diverses. Nous nous sommes intéressé dans cette thèse à l'analyse du graphe social de l'organisation à l'échelle de ses groupes et communautés en mettant de côté l'échelle individuelle.

Nous avons étudié la manière dont le graphe social d'une organisation, reconstitué à partir des informations issues d'une part de l'analyse des médias sociaux et d'autre part des bases documentaires de l'organisation, pouvait entrer dans un processus décisionnel.

Il est fréquent, dans les organisations, que les décideurs s'outillent de système d'aide à la décision. Le but de ces systèmes est de les aider à identifier différentes solutions possibles à un problème et procéder à un choix parmi ces solutions. L'identification de différentes solutions s'effectue en puisant des informations au sein de bases de connaissances. Les processus de décision nécessitent donc l'élaboration de telles bases de connaissances.

L'édification de ces bases de connaissances passe habituellement par la capitalisation des

documents de l'organisation comme les notes de synthèse de réunion, les brevets, les articles scientifiques, les réponses à appel d'offre, etc. Au cours de cette thèse nous avons cherché à intégrer des éléments issus des médias sociaux aux bases de connaissances utilisées pour la décision.

Cette intégration conjointe de différents types de données nous a conduit à penser et réaliser une modélisation sémantique de ce que nous avons appelé des ressources et d'une plate-forme permettant la capitalisation de ces ressources. Cette modélisation a pris la forme d'une ontologie appelée memorae-core 2.

Un prototype a été développé à destination des membres d'une organisation, à partir de cette modélisation. Il s'agissait de réaliser un environnement unifié au sein duquel n'importe quel membre d'une organisation peut trouver différentes applications Web lui permettant de créer, échanger et capitaliser différents types de ressources. Certaines de ces applications sont des médias sociaux : le forum, le wiki ou l'application de minimessage. Toutes les ressources créées ou échangées sur cette plate-forme prototype sont capitalisées au sein d'une seule et même base de connaissances à l'aide d'un même référentiel.

Cette indexation conjointe de différents types de ressources nous a permis de mettre en place une brique inférentielle sur la plate-forme afin de pouvoir tirer de nouvelles informations de celles déjà à disposition. Le but de cette brique inférentielle est d'alimenter un processus de décision en apportant de nouvelles informations, obtenues à l'aide d'heuristiques sur la base de connaissances.

L'heuristique principale mise en œuvre au cours de cette thèse est l'adaptation de l'algorithme de *Edgerank* de Facebook. Ce dernier nous permet d'affecter un score à chacune des traces de collaboration inférées de la base de connaissances par le biais des ressources qui y sont capitalisées. La somme des scores de chacune des traces entre deux partenaires détermine une note qualifiant le degré de proximité de ces deux partenaires.

Cette notation nous permet, par le biais d'une brique de visualisation à destination des décideurs d'une organisation, d'extraire le graphe des collaborations de l'organisation utilisant la plate-forme prototype. Il s'agit d'un graphe dynamique, dont l'évolution peut être présentée en recalculant la note de chacune des relations pour une date donnée. Ces évolutions sont mesurées en fonction des différentes thématiques abordées par les ressources de la base de connaissances et *in fine* par l'organisation.

L'observation de ces thématiques fait partie des activités de veille stratégique des décideurs de l'organisation. Ces derniers ont besoin de connaître le positionnement de leur organisation sur les marchés identifiés par ces thématiques. C'est dans cette optique que la brique inférentielle que nous avons réalisée effectue le suivi de l'évolution de ces thématiques (autrement appelé suivi de tendances).

Les apports de notre thèse peuvent donc s'appréhender selon deux angles complémentaires. Il s'agissait d'abord de répondre à une problématique de gestion de connaissances au sein des organisations et en particulier la prise en compte des informations échangées au sein des médias sociaux. L'apport se situe ici dans la proposition d'un modèle théorique des notions de donnée, information, ressource, document et connaissance, défendue dans le chapitre 7 et son implémentation sous la forme d'une ontologie présentée dans le chapitre 8. Il s'agissait ensuite d'étudier la problématique de réutilisation des connaissances capitalisées selon l'ontologie apportée par notre thèse. Nous nous sommes placé dans le cadre des systèmes d'aide à la décision dans les organisations. L'apport se situe ici dans l'élaboration d'un système d'aide à la décision pour les organisations, puisant dans ses bases de connaissances afin d'aider les décideurs à se créer de nouvelles connaissances. Ce système est conçu pour outiller un processus de veille stratégique et concurrentielle ainsi que les phases d'identification de consortiums lors du montage de projets de recherche ou industriels.

Notre étude nous permet finalement d'affirmer que les médias sociaux, et par extension une approche guidée par l'utilisation de données issues des médias sociaux de la gestion de connaissances dans les organisations, permettent de soutenir les processus de prise de décision dans les organisation via la création, par les décideurs, de nouvelles connaissances.

16. Perspectives

Nos travaux nous ont permis de dresser des passerelles entre différents domaines scientifiques afin d'en étudier les interactions possibles. Au travers de l'étude de l'impact des médias sociaux sur les systèmes d'aide à la décision, nous avons pu aborder les domaines de la gestion de connaissances, du Web sémantique, de la théorie des graphes et de l'aide à la décision.

Les expérimentations que nous avons pu réaliser dans le cadre de cette thèse sont encourageantes et permettent de dégager des voies d'améliorations, tant d'un point de vue théorique que pratique. En reprenant les différents apports explicités dans le chapitre 15, nous pouvons identifier deux directions principales pour nos perspectives :

— Évolution du modèle théorique supportant la création de connaissances à partir des données échangées au sein des médias sociaux et de son implémentation sous la forme de l'ontologie memorae-core 2 ;
— Amélioration des heuristiques fondées sur l'utilisations des connaissances ayant pu être capitalisées en suivant notre approche théorique et ouverture du système à d'autres utilisations.

16.1 Évolution de la théorie

Notre travail de positionnement sur la notion de document numérique a porté principalement sur la possibilité de réutilisation de données en provenance de médias sociaux à des fins d'analyses dans le cadre de l'aide à la décision et en particulier dans les processus de veille technologique et concurrentielle. Ce faisant, notre cycle de vie de la donnée vers la connaissance est fortement dépendant d'une implémentation sous la forme d'un système numérique. En particulier, la phase d'appropriation des ressources permettant la création de connaissances n'a été définie et évaluée que dans un cadre numérique. Dans ce cadre, il est aisé de proposer différentes actions aux utilisateurs pour outiller leur interprétation et leur appropriation des ressources. Nous avons présenté dans ce sens les processus d'indexation documentaire ou de génération de rapports personnalisés.

Il semble donc intéressant de réfléchir à la généralisation de notre positionnement théo-

rique. La notion de ressource, par exemple, a-t-elle un sens en dehors du monde numérique ? Peut-on identifier cette étape intermédiaire entre l'information et la connaissance dans des objets courants ? Nous pensons qu'une telle transposition, qui reste cependant à valider, est possible. Il y a par exemple le livre, objet physique pouvant prendre différentes formes (ressource) pour présenter le même contenu (information) au lecteur, contenu duquel il pourra tirer de nouvelles connaissances.

Une autre voie de prospection concerne l'évolutivité de notre approche. Celle-ci favorise l'utilisation d'ontologies formelles pour capitaliser les connaissances de l'organisation. Or, les organisations sont des environnements dynamiques qui, bien que réclamant des systèmes fiables et pérennes de gestion de connaissances, ne peuvent se satisfaire d'une plate-forme trop rigide. Il convient donc d'étudier les différentes manière d'autoriser la révision de l'ontologie de domaine supportant la capitalisation des connaissances de l'organisation. Nous pensons que cette évolution peut s'effectuer en mettant en œuvre deux niveaux distincts d'indexation :

— Un premier niveau utilisant l'ontologie de domaine élaborée par les experts de l'organisation et structurant les concepts actuels et bien documentés de son domaine d'affaire ;
— Un second niveau plus permissif composé de la réunion des différentes folksonomies élaborées par les membres de l'organisation au cours de leurs travaux quotidiens. L'observation et l'analyse des tendances thématiques au sein de ces folksonomies (par exemple à l'aide de l'outil présenté dans cette thèse) permet d'identifier les futurs concepts clés du domaine d'affaire de l'organisation à intégrer à l'ontologie formelle.

16.2 Devenir du système développé

L'analyse des médias sociaux oblige à se confronter très tôt à la problématique de la taille des corpus à manipuler. Notre système étant fondé sur l'exécution régulière d'heuristiques sur une base de connaissances appelée à grandir, il semble intéressant de se pencher sur l'évolution de nos algorithmes pour les rendre opérationnels sur des infrastructures de calculs répartis.

Le peuplement lui-même de la base de connaissances soulève des interrogations. Il est idéalement le fruit du travail collaboratif de nombreuses personnes différentes. L'activité parallèle de ces personnes génère de nombreuses données qui peuvent s'avérer redondantes, incomplètes voire incohérentes. Il semble alors pertinent d'introduire des mécanismes de nettoyage de la base de connaissances, de manière à améliorer les résultats de nos propres inférences. Nous avons mis en place une API permettant de greffer différents modules à la base de connaissances et pouvant travailler en avance de phase par rapport à nos propres heuristiques.

Nous avons pu expérimenter la jonction de notre plate-forme avec des travaux mettant en œuvre des algorithmes de fusion d'informations [Laudy *et al.*, 2013, Lortal *et al.*, 2013b]. Cette

fusion d'informations permet de nettoyer ou compléter les affiliations des entités détectées ou ajoutées à la base de connaissances, ce qui améliore du même coup les résultats de nos propres heuristiques (moins de doublons ou de nœuds isolés).

Une autre collaboration a été entreprise dans le cadre du stage de master de Jean-Philippe Attal cherchant à étudier l'apport potentiel de la théorie des jeux à l'analyse des réseaux sociaux. Il était principalement question de mettre en œuvre des heuristiques permettant de mesurer l'intérêt et le coût potentiel du montage de différents consortiums d'organisations. Les premiers éléments de ces travaux ont pu être présentés dans un poster [Lortal *et al.*, 2013a].

Par ailleurs, nous avons étudié la possibilité d'améliorer la précision de notre heuristique établissant des prévisions d'évolution de l'utilisation de concepts de la base de connaissances. Notre étude a porté sur les mécanismes liés à l'apprentissage automatique, afin de mieux prendre en compte les subtilités dans l'évolution de l'utilisation de chacun des concepts. Des expérimentations ont été menées dans ce sens au cours du stage de master de John Charbit que nous avons pu suivre. Cette approche a apporté des résultats intéressants.

Elle a cependant permis de constater une limite théorique importante dans son utilisation. L'apprentissage ne peut en effet se baser que sur des données déjà en notre possession. Si l'étude de l'évolution de l'utilisation d'un concept au sein d'une communauté d'intérêt à la composition stable est correcte, dès lors que cette communauté devient dynamique, en intégrant ou perdant de manière aléatoire des membres, les heuristiques fondées exclusivement sur des méthodes d'apprentissage se révèlent extrêmement médiocres. Il convient alors de considérer des méthodes permettant de suivre l'évolution des communautés [Tantipathananandh *et al.*, 2007, Crossman *et al.*, 2009, Reda *et al.*, 2011, Stattner, 2012] et ainsi intégrer cette dynamique dans les heuristiques d'évolution d'utilisation des concepts.

Bibliographie

Linda Argote et Paul Ingram. Knowledge transfer : A basis for competitive advantage in firms. *Organizational Behavior and Human Decision Processes*, 82(1) :150 – 169, May 2000. doi : 10.1006/obhd.2000.2893.

David R. Arnott et Graham Pervan. A critical analysis of decision support systems research. *Journal of Information Technology*, 20(2) :67 – 87, 2005. doi : 10.1057/palgrave.jit.2000035.

James Aspnes, Kevin Chang et Aleksandr Yampolskiy. Inoculation strategies for victims of viruses and the sum-of-squares partition problem. *Journal of Computer and System Sciences*, 72(6) :1077 – 1093, 2006. doi : 10.1016/j.jcss.2006.02.003.

Millennium Ecosystem Assessment, editor. *Ecosystems and Human Well-being : Synthesis*. Island Press, Washington, DC, USA, 2005.

David Auber, Daniel Archambault, Romain Bourqui, Antoine Lambert, Morgan Mathiaut, Patrick Mary, Maylis Delest, Jonathan Dubois et Guy Mélançon. The tulip 3 framework : A scalable software library for information visualization applications based on relational data. Technical report, Inria, Research Centre Bordeaux – Sud-Ouest, January 2012.

Alain Auger, Denis Gouin et Jean Roy. Decision support and knowledge exploitation technologies for c4isr. Technical Report September, Recherche et développement pour la défense Canada, Valcartier, Canada, 2006.

Mathieu Bastian, Sebastien Heymann et Mathieu Jacomy. Gephi : An open source software for exploring and manipulating networks. In *Proceedings of the 3rd International AAAI Conference on Weblogs and Social Media*, May 2009.

Valérie Beaudoin. Panorama sur les réseaux sociaux en sciences humaines. In *Séminaire Business Intelligence du BILab*, March 2009.

Ahcene Benayache. *Construction d'une mémoire organisationnelle de formation et évaluation dans un contexte elearning : le projet MEMORAe*. PhD thesis, Université de Technologie de Compiègne, Compiègne, France, 2005.

Tim Berners-Lee et Robert Cailliau. World wide web. In Catharinus Verkerk et Wojciech Wójcik, editors, *Proceedings of the Conference on Computing in High-Energy Physics*, pages 69 – 74, Annecy, France, 1992. CERN. ISBN 9290830492.

Laurent Bironneau et Dominique Philippe Martin. Modélisation d'entreprise et pratiques de management implicitement liées aux erp : enjeux conceptuels et études de cas. *Finance Contrôle Stratégie*, 5 (4) :29 – 50, December 2002.

Yuri Boreisha et Oksana Myronovych. Web-based decision support systems as knowledge repositories for knowledge management systems. *Ubiquitous Computing and Communication Journal*, 3(Special Issue of IKE'07 Conference) :22 – 29, 2008.

David Bray. Knowledge ecosystems : A theoretical lens for organizations confronting hyperturbulent environments. In Tom McMaster, David Wastell, Elaine Ferneley et Janice I. DeGross, editors, *Organizational Dynamics of Technology-Based Innovation : Diversifying the Research Agenda*, IFIP Advances in Information and Communication Technology, chapter 31, pages 457 – 462. Springer, 2007.

John G. Breslin et Stefan Decker. The future of social networks on the internet : The need for semantics. *Internet Computing*, 11(6) :86 – 90, November 2007. ISSN 1089-7801. doi : 10.1109/MIC.2007.138. URL http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=4376234.

John G. Breslin, Uldis Bojārs, Alexandre Passant, Sergio Fernández et Stefan Decker. Sioc : Content exchange and semantic interoperability between social networks. In *W3C Workshop on the Future of Social Networking*, Barcelona, España, 2009. URL http://www.w3.org/2008/09/msnws/papers/sioc.html.

Suzanne Briet. *Qu'est-ce que la documentation.* Editions documentaires et industrielles, Paris, France, 1951.

Michael Buckland. What is a « digital document » ? *Document numérique*, 2(2) :221 – 230, 1998.

Patrick Cohendet et Morad Diani. L'organisation comme une communauté de communautés croyances collectives et culture d'entreprise. *Revue d'économie politique*, 113(5) :697 – 720, 2003.

David Combe, Christine Largeron, Előd Egyed-Zsigmond et Mathias Géry. A comparative study of social network analysis tools. In *Proceedings of the 2nd International Workshop on Web Intelligence and Virtual Enterprises*, page 12p., Saint-Étienne, France, 2010.

Bernard Conein. Communautés épistémiques et réseaux cognitifs : coopération et cognition distribuée. *Revue d'économie politique*, 113 :141 – 159, 2004.

Michael Conlon, Katy Borner, Curtis Cole, Jon Corson-Rikert, Ellen J. Cramer, Valrie I. Davis, Kristi Holmes, Gerald Joyce, Dean B. Krafft, Leslie McIntosh et Richard Noel. Semantic modeling of scientist : The vivo ontology, 2003. URL https://wiki.duraspace.org/display/VIVO/VIVO+Ontology.

James F. Courtney. Decision making and knowledge management in inquiring organizations : toward a new decision-making paradigm for dss. *Decision Support Systems*, 31 :17 – 38, 2001.

Andrew Crooks, Arie Croitoru, Anthony Stefanidis et Jacek Radzikowski. #earthquake : Twitter as a distributed sensor system. *Transactions in GIS*, In Press, Accepted Manuscript :Available online, November 2012. ISSN 13611682. doi : 10.1111/j.1467-9671.2012.01359.x. URL http://gisagents.blogspot.fr/2012/04/earthquake-twitter-as-distributed.html.

Rob Cross, Andrew Parker, Lawrence Prusak et Stephen P. Borgatti. Knowing what we know : Supporting knowledge creation and sharing in social networks. *Organizational Dynamics*, 30(2) :100 – 120, November 2001. ISSN 0090-2616/01.

Jacob Crossman, Robert Bechtel, Van Parunak et Sven Brueckner. Integrating dynamic social networks and spatio-temporal models for risk assessment, wargaming and planning. In *Proceedings of the Network Science Workshop*, West Point, NY, USA, 2009.

Bruce D'Arcus et Frédérick Giasson. Bibo specifications, 2009. URL http://bibliontology.com/specification.

Wouter de Nooy, Andrej Mrvar et Vladimir Batagelj. *Exploratory Social Network Analysis with Pajek*. Number 27 in Structural Analysis in the Social Sciences. Cambridge University Press, March 2005.

Alain Degenne et Forsé Michel. *Les réseaux sociaux. une approche structurale en sociologie*. Armand Colin, Paris, 2nde edition, 2004.

Étienne Deparis, Marie-Hélène Abel, Gaëlle Lortal et Juliette Mattioli. Knowledge capitalization in an organization social network. In Joaquim Filipe et Kecheng Liu, editors, *Proceedings of the International Conference on Knowledge Management and Information Sharing*, pages 217 – 222, Paris, France, October 2011a. SciTePress. ISBN 978-989-8425-81-2.

Étienne Deparis, Marie-Hélène Abel et Juliette Mattioli. Capitalisation des échanges informels en entreprise via les réseaux sociaux. In Hakim Hacid, Cécile Favre et Ludovic Denoyer, editors, *Actes de l'atelier Le Web Social*, pages 35 – 42, Brest, France, January 2011b.

Étienne Deparis, Marie-Hélène Abel et Juliette Mattioli. Modeling a social collaborative platform with standard ontologies. In Kokou Yetongnon, Richard Chbeir et Albert Dipanda, editors, *Proceedings of the Knowledge Acquisition, Reuse and Evaluation Workshop of the 7th SITIS Conference*, pages 167 – 173, Dijon, France, November 2011c. IEEE Computer Society. ISBN 978-0-7695-4635-3. doi : 10.1109/SITIS.2011.58.

Étienne Deparis, Marie-Hélène Abel et Juliette Mattioli. Modélisation d'une plate-forme sociale de collaboration à l'aide de standards du web sémantique. In *Actes de la 12^{e} Conférence Internationale Francophone sur l'Extraction et la Gestion de Connaissances*, pages 565 – 566, Bordeaux, France, 2012. Poster.

Étienne Deparis, Marie-Hélène Abel, Gaëlle Lortal et Juliette Mattioli. Gestion de connaissance pour l'entreprise prenant en compte les interactions sociales. In *Actes de la 31^{e} conférence d'informatique des organisations et systèmes d'information et de décision*, pages 335–350, Paris, France, May 2013a.

Étienne Deparis, Marie-Hélène Abel, Gaëlle Lortal et Juliette Mattioli. Modélisation sociale d'une organisation et usage de son réseau. In *Actes de la conférence d'ingénierie des connaissances*, Lille, France, July 2013b. Poster.

Étienne Deparis, Marie-Hélène Abel, Gaëlle Lortal et Juliette Mattioli. Designing a system to capitalize both social and documentary resources. In *Proceedings of the 17th IEEE International Conference on Computer Supported Cooperative Work in Design*, Whistler, BC, Canada, June 2013c. IEEE SMC Society.

Étienne Deparis, Marie-Hélène Abel, Gaëlle Lortal et Juliette Mattioli. Information management from social and documentary sources in organizations. *Computers in Human Behavior*, 30 :753–759, January 2014.

Guillaume Erétéo, Michel Buffa, Fabien Gandon, Patrick Grohan, Mylène Leitzelman et Peter Sander. A state of the art on social network analysis and its applications on a semantic web. In John G. Breslin, Uldis Bojārs, Alexandre Passant et Sergio Fernández, editors, *Proceedings of the Social Data on the Web Workshop of the 7th International Semantic Web Conference*, Karlsruhe, Deutschland, 2008. CEUR Workshop Proceedings.

Guillaume Erétéo, Fabien Gandon, Michel Buffa et Patrick Grohan. Analyse des réseaux sociaux et web sémantique : un état de l'art. Technical report, Projet ISICIL, Appel ANR CONTINT 2008, 2009a.

Guillaume Erétéo, Fabien Gandon, Olivier Corby et Michel Buffa. Semantic social network analysis. *The Computing Research Repository*, abs/0904.3 :5p., 2009b. URL http://arxiv.org/abs/0904.3701.

Guillaume Erétéo, Freddy Limpens, Fabien Gandon, Olivier Corby, Michel Buffa, Mylène Leitzelman et Peter Sander. Semantic social network analysis, a concrete case. In Ben Kei Daniel, editor, *Handbook of Research on Methods and Techniquesfor Studying Virtual Communities : Paradigms and Phenomena*, chapter 7, pages 122 – 156. IGI Global, 2011. ISBN 9781609600402. doi : 10.4018/978-1-60960-040-2.cho07.

Jean-Claude Ermine. La gestion des connaissances, un levier stratégique pour les entreprise. In Pierre Tchounikine, editor, *Actes des 11es Journées Francophones d'Ingénierie des Connaissances*, Toulouse, France, 2000. URL http://www.irit.fr/IC2000/ACTES/texte-ermine.html.

Amanda Ferreira et Tanya du Plessis. Effect of online social networking on employee productivity. *South African Journal of Information Management*, 11(1) :1 – 16, 2009. doi : 10.4102/sajim.v11i1.397.

M. P. Fewell et Mark G. Hazen. Cognitive issues in modelling network-centric command and control. Technical report, Defence Science and Technology Organisation - DoD Au. Gov., Edinburgh, Australia, 2005.

Danyel Fisher, Marc Smith et Howard T. Welser. You are who you talk to : Detecting roles in usenet newsgroups. In *Proceedings of the 39th Hawaii International Conference on System Sciences*, volume 3, pages 1 – 10, Kauai, HI, USA, 2006. IEEE Computer Society. ISBN 0-7695-2507-5. doi : 10.1109/HICSS.2006.536.

Mark S. Fox. The tove project : A common-sense model of the enterprise. In F. Belli et F.J. Radermacher, editors, *Industrial and Engineering Applications of Artificial Intelligence and Expert Systems*, volume 604 of *Lecture Notes in Artificial Intelligence*, pages 25 – 34. Springer Verlag, 1992. URL http://www.eil.utoronto.ca/enterprise-modelling/tove/index.html.

Virginia Gewin. Networking in vivo. *Nature*, 462(7269) :123, November 2009. doi : 10.1038/nj7269-123a. URL http://www.nature.com/naturejobs/science/articles/10.1038/nj7269-123a.

Ouafia Ghebghoub, Marie-Hélène Abel, Claude Moulin et Adeline Leblanc. Building and use of a lom ontology. In Faiez Gargouri et Wassim Jaziri, editors, *Ontology Theory, Management and Design : Advanced Tools and Models*, chapter 7, pages 162 – 177. IGI Global, 2010. ISBN 9781615208593. doi : 10.4018/978-1-61520-859-3.ch007.

Thomas R. Gruber. Toward principles for the design of ontologies used for knowledge sharing. Technical report, Stanford University, August 1993. Available as Technical Report KSL 93-04, Knowledge Systems Laboratory.

Nicola Guarino. Formal ontology and information systems. In *Proceedings of the 1st International Conference on Formal Ontologies in Information Systems*, pages 3 – 15, Trento, Italia, June 1998. IOS Press.

Charles Gueye. Nouvelle forme organisationnelle et approche conventionnaliste du partage des connaissances : le recours a la notion de communauté épistémique. In *Actes du 7e Congrès International Francophone en Entrepreneuriat et PME*, Montpellier, France, October 2004.

Harry Halpin. Beyond walled gardens : Open standards for the social web. In John G. Breslin, Uldis Bojārs, Alexandre Passant et Sergio Fernández, editors, *Proceedings of the Social Data on the Web Workshop of the 7th International Semantic Web Conference*, Karlsruhe, Deutschland, October 2008.

Andreas M. Kaplan et Haenlein Michael. Users of the world, unite ! the challenges and opportunities of social media. *Business horizons*, 53(1) :59 – 68, 2010. doi : 10.1016/j.bushor.2009.09.003.

Mayank Lahiri et Tanya Y. Berger-Wolf. Structure prediction in temporal networks using frequent subgraphs. In *Proceedings of the IEEE Symposium on Computational Intelligence and Data Mining*, pages 35 – 47, Honolulu, HI, USA, 2007. IEEE Computer Society.

Mayank Lahiri et Tanya Y. Berger-Wolf. Mining periodic behavior in dynamic social networks. In Fosca Giannotti, Dimitrios Gunopulos, Franco Turini, Carlo Zaniolo, Naren Ramakrishnan et Xindong Wu, editors, *Proceedings of the 8th International Conference on Data Mining*, pages 373 – 382, Pisa, Italia, 2008. IEEE Computer Society. ISBN 978-0-7695-3502-9.

Mayank Lahiri et Manuel Cebrian. The genetic algorithm as a general diffusion model for social networks. In *Proceedings of the 28th International AAAI Conference on Artificial Intelligence*, pages 494 – 499, Atlanta, GA, USA, 2010. AAAI.

Sylvie Lainé-Cruzel. Documents, ressources, données : les avatars de l'information numérique. *Revue Information-Interaction-Intelligence*, 4(1) :105–119, 2004.

Claire Laudy, Étienne Deparis, Gaëlle Lortal et Juliette Mattioli. Multi-granular fusion for social data analysis for a decision and intelligence application. In *Proceedings of the 16th International Conference on Information Fusion*, Istanbul, Turkey, July 2013.

Emmanuel Lazega. Analyse de réseaux et sociologie des organisations. *Revue française de sociologie*, 35(2) :293 – 320, 1994.

Adeline Leblanc. *Environnement de collaboration et mémoire organisationnelle de formation dans un contexte d'apprentissage*. PhD thesis, Université de Technologie de Compiègne, December 2008.

Adeline Leblanc et Marie-Hélène Abel. A forum-based organizational memory as organizational learning support. *International Journal of Digital Information Management*, 6(4) :303–312, 2008.

Adeline Leblanc et Marie-Hélène Abel. E-memorae2.0 : A web platform dedicated to organizational learning needs. In *Proceedings of the 3rd World Summit on the Knowledge Society, WSKS 2010*, pages 306 – 315, Corfou, Grèce, 2010.

Qiang Li. *Modélisation et exploitation des traces d'interactions dans l'environnement de travail collaboratif*. PhD thesis, Université de Technologie de Compiègne, June 2013.

Jo Link-Pezet, Redouane Berkane et Didier Mottay. Intelligence économique et décision. In Société Française de Bibliomètrie Appliquée, editor, *Actes du colloque sur les systèmes d'information élaborée*, L'Île-Rousse, France, 1999.

Gaëlle Lortal, Étienne Deparis, Claire Laudy, Stefano Moretti et Juliette Mattioli. Un bouquet de théories pour une plateforme de décision stratégique à base de réseaux sociaux. In *Actes de la conférence d'ingénierie des connaissances*, Lille, France, July 2013a. Poster.

Gaëlle Lortal, Claire Laudy, Étienne Deparis et Juliette Mattioli. Approaches for multi-granular fusion for social data analysis for a decision and intelligence application. In *Proceedings of the 10th International Conference on Modeling Decisions for Artificial Intelligence*, Barcelona, España, November 2013b.

Nicolas Marie et Fabien Gandon. Social objects description and recommendation in multidimensional social networks : Ocso ontology and semantic spreading activation. In *Proceedings of the 3rd IEEE International Conference on Privacy, Security, Risk and Trust of the 3rd IEEE International Conference on Social Computing*, pages 1415 – 1420, Boston, MA, USA, October 2011. IEEE Computer Society.

Peter Mika. Ontologies are us : A unified model of social networks and semantics. In Yolanda Gil, editor, *Proceedings of the 4th International Semantic Web Conference*, volume 3729 of *Lecture notes in computer science*, pages 522 – 536, Galway, Ireland, 2005. Springer Science & Business. ISBN 9783540297543. doi : 10.1.1.142.7750.

Alexander Mikroyannidis. Toward a social semantic web. *Computer*, 40(11) :113 – 115, November 2007. ISSN 0018-9162. doi : 10.1109/MC.2007.405.

James F. Moore. Predators and prey : a new ecology of competition. *Harvard Business Review*, 71(3) :75 – 86, 1993.

Ikujirō Nonaka et Hirotaka Takeuchi. *The knowledge-creating company : how Japanese companies create the dynamics of innovation.* Oxford University Press, New York, NY, USA, 1995. ISBN 9780195092691.

Tim O'Reilly. What is web 2.0, design patterns and business models for the next generation of software. Technical report, O'Reilly, September 2005. URL http://oreilly.com/web2/archive/what-is-web-20.html.

Claude Parthenay. Herbert simon : rationalité limitée, théorie des organisations et sciences de l'artificiel. In Didier Chabaud, Jean-Michel Glachant, Claude Parthenay et Yannick Perez, editors, *Les grands auteurs en théorie des organisations*, pages 64 – 93. Éditions Management et Société, management edition, 2008.

Alexandre Passant et Philippe Laublet. Ontologies pour le web 2.0. In *Actes des 19es Journées Francophones d'Ingénierie des Connaissances*, pages 99 – 110, 2008a.

Alexandre Passant et Philippe Laublet. Meaning of a tag : A collaborative approach to bridge the gap between tagging and linked data. In *Proceedings of the WWW2008 Workshop on Linked Data on the Web*, volume 369, Beijing, China, 2008b. CEUR Workshop Proceedings.

Diana Penciuc. *Identification et intégration des éléments de connaissance tacite et explicite dans un processus de développement par solution de référence.* PhD thesis, Université de Technologie de Compiègne, 2012.

Jean-Yves Prax. *Le manuel du Knowledge Management — Mettre en réseau les hommes et les savoirs pour créer de la valeur.* Dunod, 3e édition edition, 2012. ISBN 978-2100575589.

Sophie Pène et Cathy Dubois. Sémiologie de l'hypertravail, communication d'exploration et réseaux sociaux. In *Actes du colloque H2PTM*, 2009.

Roger T. Pédauque. Document : forme, signe et medium, les re-formulations du numérique. RTP-DOC (RTP 33 « Documents et contenu : création, indexation, navigation » – STIC – CNRS), July 2003. URL http://archivesic.ccsd.cnrs.fr/sic_00000511.

Dnyanesh Rajpathak, Enrico Motta et Rajkumar Roy. A generic task ontology for scheduling applications. In *Proceedings of the 3rd International Conference on Artificial Intelligence*, Las Vegas, NE, USA, 2001.

Khairi Reda, Chayant Tantipathananandh, Andrew Johnson, Jason Leigh et Tanya Y. Berger-Wolf. Visualizing the evolution of community structures in dynamic social networks. In *Proceedings of the 13th IEEE Symposium on Visualization*, pages 1061 – 1070, Bergen, Norge, 2011. IEEE Computer Society.

Dave Reynolds. An organization ontology, 2010a. URL http://www.w3.org/TR/vocab-org/.

Dave Reynolds. An organization ontology : Survey. http ://www.epimorphics.com/web/wiki/organization-ontology-survey, May 2010b.

Bertrand Russell. *The problems of philosophy*. Williams and Norgate, London, 1912. URL http://www.ditext.com/russell/russell.html.

Gilbert Ryle. *La notion d'esprit*. Payot, 1978.

Takeshi Sakaki. Earthquake shakes twitter users : Real-time event detection by social sensors. In *Proceedings of the 19th International Conference on World Wide Web*, pages 851 – 860, Raleigh, NC, USA, April 2010. ACM Press. ISBN 9781605587998.

Andreas Seufert, Georg von Krogh et Andrea Back. Towards knowledge networking. *Journal of Knowledge Management*, 3(3) :180 – 190, 1999. doi : 10.1108/13673279910288608.

Shashi Shekhar et Dev Oliver. Computational modeling of spatio-temporal social networks : A time-aggregated graph approach. In Michael Goodchild et Kathlyn Carly, editors, *Proceedings of the Spatio-Temporal Constraints on Social Networks Workshop*, pages 6 – 10, Santa Barbara, CA, USA, 2010.

J. P. Shim, Merrill Warkentin, James F. Courtney, Daniel J. Power, Ramesh Sharda et Christer Carlsson. Past, present, and future of decision support technology. *Decision Support Systems*, 33 :111 – 126, 2002.

Clay Shirky. *Here comes everybody – the Power of Organizing Without Organizations*. Penguin Group, 2008. ISBN 978-0-713-99989-1.

Paul Shrivastava. Knowledge ecology : Knowledge ecosystems for business education and training. http ://www.facstaff.bucknell.edu/shrivast/KnowledgeEcology.html, 1998. URL http://www.facstaff.bucknell.edu/shrivast/KnowledgeEcology.html.

Georg Simmel. *Soziologie*. Duncker & Humblot, Leipzig, Deutschland, 1908.

Vivek K. Singh, Mingyan Gao et Ramesh Jain. Situation detection and control using spatio-temporal analysis of microblogs (poster). In *Proceedings of the 19th International Conference on World Wide Web*, pages 1181 – 1182, Raleigh, NC, USA, April 2010. ACM Press. ISBN 9781605587998. doi : 10.1145/1772690.1772864. URL http://portal.acm.org/citation.cfm?doid=1772690.1772864.

Kate Starbird et Leysia Palen. Pass it on ?: Retweeting in mass emergency. In *Proceedings of the 7th International Conference on Information Systems for Crisis Response and Management*, Seattle, WA, USA, May 2010.

Erick Stattner. *Contributions à l'étude des réseaux sociaux : propagation, fouille, collecte de données*. PhD thesis, Sciences et Technologies de l'Information – Université des Antilles et de la Guyane, December 2012.

Robert D. Steele. Open source intelligence : What is it ? why is it important to the military ? In *Proceedings of the 6th International Conference and Exhibit on Global Security and Global Competitiveness : Open Source Solutions*, volume 2, Washington, DC, USA, September 1997. Open Source Solutions, Inc.

Chayant Tantipathananandh, Tanya Y. Berger-Wolf et David Kempe. A framework for community identification in dynamic social networks. In *Proceedings of the 13th International Conference on Knowledge Discovery and Data Mining*, pages 717 – 726, San Jose, CA, USA, 2007. ISBN 9781595936097.

Alexis Tsoukiàs. From decision theory to decision aiding methodology. *European Journal of Operational Research*, 187 :138–161, 2008.

Murray Turoff, Starr Roxanne Hiltz, Hee-kyung Cho, Zheng Li et Yuanqiong Wang. Social decision support systems (sdss). In *Proceedings of the 35th Hawaii International Conference on System Sciences*, volume 1, pages 81 – 90, Washington D.C., USA, 2002. IEEE Computer Society. ISBN 0-7695-1435-9. doi : 10.1109/HICSS.2002.993863.

Johan Ugander, Lars Backstrom, Cameron Marlow et Jon Kleinberg. Structural diversity in social contagion. *Proceedings of the National Academy of Science of the United States of America*, 109(16) :5962 – 5966, April 2012. doi : 10.1073/pnas.1116502109.

Mike Uschold, Martin King, Stuart Moralee et Yannis Zorgios. The enterprise ontology, August 1996. URL http://www.aiai.ed.ac.uk/project/enterprise/enterprise/ontology.html.

Marijtje A. J. van Duijn et Jeroen K. Vermunt. What is special about social network analysis ? *Methodology*, 2(1) :2 – 6, 2006. doi : 10.1027/1614-1881.2.1.2.

Sarah Vieweg, Amanda L. Hughes, Kate Starbird et Leysia Palen. Microblogging during two natural hazards events : What twitter may contribute to situational awareness. In *Proceedings of the ACM Conference on Human Factors in Computing Systems*, pages 1079 – 1088, Atlanta, GA, USA, April 2010.

Edward Waltz. *Knowledge Management in the intelligence enterprise*. Artech House, 2003. ISBN 1580534945.

Stanley Wasserman et Katherine Faust. *Social Network Analysis : Methods and Applications*. Cambridge University Press, 1994.

Barry Wellman, Janet Salaff, Dimitrina Dimitrova, Laura Garton, Milena Gulia et Caroline Haythornthwaite. Computer networks as social networks : Collaborative work, telework, and virtual community. *Annual Review of Sociology*, 22 :213 – 238, 1996.

Étienne Wenger. *Communitites of Practice : Learning, Meaning and Identity*. Cambridge University Press, Boston, MA, USA, 1998.

Stephen J. H. Yang et Irene Y. L. Chen. A social network-based system for supporting interactive collaboration in knowledge sharing over peer-to-peer network. *International Journal of Human-Computer Studies*, 66(1) :36 – 50, January 2008. ISSN 10715819. doi : 10.1016/j.ijhcs.2007.08.005. URL http://linkinghub.elsevier.com/retrieve/pii/S1071581907001139.

Pascale Zaraté. *Des Systèmes Interactifs d'Aide à la Décision aux Systèmes Coopératifs d'Aide à la Décision : Contributions conceptuelles et fonctionnelles*. Habilitation à diriger des recherches, Institut National Polytechnique de Toulouse, Toulouse, France, December 2005.

Cinquième partie

Annexes

A. Publications

Nous avons régulièrement communiqué sur l'avancée de nos travaux au sein de différents congrès. Nous avons pu intervenir lors des ateliers « le Web social » [Deparis *et al.*, 2011b] et « *Knowledge Acquisition, Reuse and Evaluation* » [Deparis *et al.*, 2011c]. Nous avons également participé aux conférences « *Knowledge Management and Information Sharing* » [Deparis *et al.*, 2011a], « Extraction et la Gestion de Connaissances » [Deparis *et al.*, 2012], « informatique des organisations et systèmes d'information et de décision » [Deparis *et al.*, 2013a], « ingénierie des connaissances » [Deparis *et al.*, 2013b] et « *Computer Supported Cooperative Work in Design* » [Deparis *et al.*, 2013c]. Enfin, un article a été soumis et accepté dans la revue *Computers in Human Behavior* [Deparis *et al.*, 2014].

B. Fragments de code turtle

B.1 Déclaration des espaces de nom

Tous les extraits de code turtle des sections suivantes et du manuscrit sont a considérer comme étant situés au sein d'un fichier global ayant déclaré un certain nombre d'espaces de noms de la manière suivante.

```
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#>.
@prefix foaf: <http://xmlns.com/foaf/0.1/>.
@prefix sioc: <http://rdfs.org/sioc/ns#>.
@prefix bibo: <http://purl.org/ontology/bibo/>.
@prefix mc2: <http://www.hds.utc.fr/memorae/memorae-core2.owl#>.
@prefix argos: <http://argos-incorporated.demo/domain-ontology#>.
# baseURI: http://argos-incorporated.demo/knowledge-base#
```

B.2 Représentation d'un extrait du référentiel partagé

```
argos:Web a owl:Class;
    rdfs:subClassOf skos:Concept.

:web_general a argos:Web;
    skos:prefLabel "Web";
    skos:altLabel "Internet";
    skos:broader :web2_general;
    skos:broader :wot_general;
    skos:broader :dataweb_general;
    skos:broader :semanticweb_general.

argos:Web2 a owl:Class;
    rdfs:subClassOf argos:Web.

```

```
15  :web2_general a argos:Web2;
16      skos:prefLabel ”Web 2.0”;
17      skos:narrower :web_general.
18
19  argos:WoT a owl:Class;
20      rdfs:subClassOf argos:Web.
21
22  :wot_general a argos:WoT;
23      skos:prefLabel ”Web of things”@en;
24      skos:prefLabel ”Internet des objets”@fr;
25      skos:narrower :web_general.
26
27  argos:DataWeb a owl:Class;
28      rdfs:subClassOf argos:Web.
29
30  :dataweb_general a argos:DataWeb;
31      skos:prefLabel ”Web of data”@en;
32      skos:prefLabel ”Web de données”@fr;
33      skos:narrower :web_general;
34      skos:related :opendata_general;
35      skos:related :semanticweb_general.
36
37  argos:OpenData a owl:Class;
38      rdfs:subClassOf skos:Concept.
39
40  :opendata_general a argos:OpenData;
41      skos:prefLabel ”Open Data”@en;
42      skos:prefLabel ”Données ouvertes”@fr;
43      skos:related :dataweb_general.
44
45  argos:SemanticWeb a owl:Class;
46      rdfs:subClassOf :web.
47
48  :semanticweb_general a argos:SemanticWeb;
49      skos:prefLabel ”Semantic web”@en;
50      skos:prefLabel ”Web sémantique”@fr;
51      skos:narrower :web_general;
52      skos:related :dataweb_general.
```

B.3 Représentation d'une chaîne d'affiliation

```
1  @prefix vivo: <http://vivoweb.org/ontology/core>.
2  @prefix foaf: <http://xmlns.com/foaf/0.1/>.
3  @prefix mc2: <http://www.hds.utc.fr/memorae/memorae-core2.owl#>.
4
5  :civari a foaf:Person;
6      foaf:lastName ”Ivari”;
7      foaf:firstName ”Charles”.
8
9  :ivaricorp a vivo:Company;
10     foaf:name ”Ivari Corporation”;
11     rdfs:label ”Ivari Corp.”;
12     vivo:hasCurrentMember :civari;
13     vivo:currentlyHeadedBy :civari.
14
15 :soweb a vivo:Team;
16     foaf:name ”Équipe Web social”;
17     rdfs:label ”SoWeb”;
18     mc2:affiliationElementName ”Équipe”;
19     vivo:hasCurrentMember :civari;
20     vivo:currentlyHeadedBy :dcide;
21     vivo:currentMemberOf :argosrd.
22
23 :argosrd a vivo:ResearchOrganization;
24     foaf:name ”Unité de R&D”;
25     rdfs:label ”R&D unit”;
26     mc2:affiliationElementName ”Unité”;
27     vivo:hasCurrentMember :soweb;
28     vivo:subOrganizationWithin :argos.
29
30 :argos a vivo:PrivateCompany;
31     foaf:name ”Argos Inc.”;
32     mc2:affiliationElementName ”Entreprise”;
33     vivo:hasSubOrganization :argosrd.
```

C. Questionnaires utilisés pour les expérimentations

C.1 Questionnaire utilisé avec les étudiants

Bonjour,

Au cours du semestre écoulé vous avez pu utiliser la plateforme E-MEMORAe 2.0 dans le cadre de votre travail de veille technologique. Nous vous remercions de prendre quelques minutes pour répondre au questionnaire suivant, qui servira à améliorer la plate-forme et l'approche utilisée pour la gestion de connaissances.

Lors de votre veille, quel(s) outil(s) avez-vous utilisé ?

- L'outil forum de la plate-forme (échelle de Likert)
- L'outil wiki de la plate-forme (échelle de Likert)
- L'outil d'annotation de la plate-forme (échelle de Likert)
- Un logiciel de traitement de textes externe à la plate-forme (échelle de Likert)
- Un logiciel de présentations externe à la plate-forme (échelle de Likert)
- Un logiciel de tableur externe à la plate-forme (échelle de Likert)
- Un logiciel de création d'images mentales externe à la plate-forme (échelle de Likert)
- Un autre outil (champ de texte libre)
- Si vous avez utilisé des logiciels externes, transformiez-vous vos documents en PDF avant l'envoi vers la plate-forme ? (oui/non)

Lors de votre veille, quel(s) type(s) de ressource avez-vous consulté ET partagé sur la plate-forme ?

- Des articles académiques (échelle de Likert)
- Des livres scientifiques (échelle de Likert)

— Des articles encyclopédiques (échelle de Likert)
— Des articles d'opinions (échelle de Likert)
— Des supports de présentation (échelle de Likert)
— Des comptes-rendus de réunions (échelle de Likert)
— Un autre type de ressource (champ de texte libre)
— D'après vous, quelle est la fonctionnalité qui manque le plus à la plateforme ? (choix multiple parmi « Édition de documents bureautiques », « Gestion d'un agenda », « Messagerie privée », « chat », « Import d'informations de mes réseaux sociaux », « Import d'informations de solutions cloud (telles que Dropbox, SkyDrive, Google Drive, etc) », « Autre » (champ de texte libre dans ce cas))

De quelle manière accédiez-vous à la plate-forme ?

— Depuis chez-vous (échelle de Likert)
— Depuis l'UTC, mais d'une machine personnelle (échelle de Likert)
— Depuis l'UTC, mais d'une machine partagée en salle informatique (échelle de Likert)

Le partage d'informations

— Lors de l'ajout d'une ressource à la plate-forme, vous l'indexiez en moyenne (choix unique parmi « À 1 seul concept », « À 2 ou 3 concepts », « À 4 ou 5 concepts », « Entre 5 ou 10 concepts », « Plus de 10 concepts », « Je n'ai pas indexé de ressource », « Autre » (champ de texte libre dans ce cas))
— Lors de l'ajout d'une ressource à la plate-forme, vous la partagiez en moyenne (choix unique parmi « À 1 seul groupe », « À 2 groupes », « À 3 groupes ou plus », « Je n'ai pas partagé de ressource », « Other » (champ de texte libre dans ce cas))
— Lors de l'ajout d'une ressource à la plate-forme (choix unique parmi « Vous l'indexiez d'abord dans votre espace privé avant de la partager », « Vous l'indexiez directement dans un ou plusieurs espaces partagés », « Les deux réponses ci-dessus s'appliquent à peu près également à votre cas », « Vous ne pouvez pas dire », « Je n'ai pas partagé de ressource »)

Intérêt de l'approche Memorae

— Quel est pour vous l'intérêt de pouvoir naviguer au sein d'une cartographie de concept (échelle de Likert)
— Quel est pour vous l'intérêt de pouvoir créer des groupes libres autour d'une thématique donnée (échelle de Likert)

— Quel est pour vous l'intérêt de pouvoir accéder à tous les types de ressources possible pour une thématique donnée ? (échelle de Likert)
— Quel est pour vous l'intérêt de pouvoir accéder à tous moments à toutes vos ressources accessibles via vos différents groupes pour une thématique donnée ? (échelle de Likert)
— Avez-vous apprécié le fait de pouvoir visualiser en parallèle l'ensemble des ressources accessibles indexées sur un sujet (focus de la carte) et distribuées dans vos différents espaces ? (échelle de Likert)

Ergonomie de la plate-forme

— Que pensez-vous de la navigation au sein de la cartographie ? (échelle de Likert)
— Que pensez-vous de la création / l'organisation en groupe libre autour d'une thématique ? (échelle de Likert)
— Que pensez-vous de l'ajout de ressource ? (échelle de Likert)
— Que pensez-vous de l'ajout d'une question sur le forum ? (échelle de Likert)
— Que pensez-vous de l'ajout d'une réponse à une question sur le forum ? (échelle de Likert)
— Que pensez-vous de l'ajout d'une page de wiki ? (échelle de Likert)
— Que pensez-vous de la recherche d'un concept ? (échelle de Likert)
— Que pensez-vous de la recherche d'une ressource donnée ? (échelle de Likert)

Pour conclure

— Pensez-vous d'une manière général que la plateforme E-MEMORAe 2.0 est un bon outil de collaboration ? (oui/non)
— Pensez-vous d'une manière général que la plateforme E-MEMORAe 2.0 est innovante ? (oui/non)
— Pouvez-vous justifier votre choix ? (question ouverte)
— Merci de donner dans l'espace ci-dessous toutes les remarques complémentaires que vous aimeriez nous faire remonter. (question ouverte)

C.2 Protocole utilisé au sein de Thales R&T

Présentation de l'outil de visualisation

— Expliquer que les nœuds sont des partenaires et les liens des agrégations de différentes collaborations.

— Montrer l'affichage du label des partenaires de manière globale et au survol de la souris
— Rechercher un partenaire avec l'outil dédié (*Univ. Paris 6*)
— Afficher son réseau réduit
— Montrer l'évolution du graphe dans le temps
— Revenir au graphe global et montrer le détail d'un lien
— Montrer comment afficher les liens potentiels
— Faire une recherche sur un mot-clé
— Effacer le mot-clé
— Afficher les stats d'utilisation de mot-clé pour un partenaire
— Cliquer sur un de ces mots-clés et montrer la colorisation du graphe.
— Revenir sur le graphe général et lancer les fenêtre d'évolution des mots-clés
— Lancer une recherche d'émergeance.

Présentation de l'outil de génération de rapport

— Expliquer les différents éléments du formulaire
— Générer un rapport simple
— Générer un rapport de consortium pour un mot-clé

Exercices

Sur le graphe

— Identifier des relations anciennes entre partenaires
— Identifier des relations récentes mais fortes entre partenaires
— Identifier des relations régulières
— Mettre en évidence le mot-clé *particle filter*
— Lister les différents partenaires ayant déjà travaillés sur ce mot-clé
— Identifier deux partenaires ayant tous deux travaillé sur *particle filter* mais n'ont pas de collaborations commune sur le sujet.
— Identifier les partenaires les plus « proches » sur la thématique *particle filter*
— Vérifier si le mot-clé est considéré comme émergeant ou pas, et éventuellement à quelle date.

Avec l'outil de rapport

— Donner une liste de consortium répondant à un mot-clé (*particle filter*) en vue réduite.
— Donner un avis sur la pertinence des consortiums proposés

— Donner la même liste, mais cette fois-ci en vue complète.
— Discuter le choix des partenaires effectués en fonction des disponibilités
— Détecter les incongruités
— Proposer de nouvelles règles pour les types de projets implémentés (type de partenaires, nombre réel de partenaires à prendre en compte pour débuter un consortium ?)

Questions subsidiaires

— Quelles sont les sources habituellement consultées pour trouver de l'information à propos de partenaires éventuels ?
— Outils utilisés pour compiler les informations consultées (Excel, Tableau software, etc.) ?
— Quel serait l'outil ultime pour aider à définir et monter un consortium ?

D. Formats intermédiaires de notre outil d'analyse de réseaux de collaborations

Le convertisseur permet de générer différents formats de sortie :

OWL Ce générateur permet de transformer les données soumises en entrée selon le formalisme imposé par notre ontologie. Les classes de notre ontologie sont instanciées et les individus ainsi décrits sont exportés sous la forme de triplet RDF au sein d'un fichier OWL.

JSON Ce générateur permet de transformer les données soumises en entrée selon un formalisme permettant de s'interfacer facilement avec la bibliothèque de visualisation d3.js. Le détail de ce format est décrit à la section D.1.

Synthesizer Ce générateur est une modification légère du précédent générateur JSON. Il s'agit ici de sortir un fichier permettant d'échanger avec d'autres briques logicielles externes comme le module de Claire Laudy. Le détail de ce format est décrit à la section D.2.

Gephi Ce générateur permet d'exporter les données soumises en entrée dans un format compréhensible pour le logiciel d'analyse Gephi.

D.1 Format de sortie JSON

La sortie JSON suit le format suivant.

links Tableau contenant l'ensemble des relations du réseau étudié. Bien que les relations ne soient pas dirigées, afin de se conformer au modèle de données attendu par la bibliothèque de visualisation d3.js que nous utilisons, les notions de « source » et « target » sont utilisées, même si nous n'utilisons pas leur propriété de direction.

source index dans le tableau *nodes* (voir ci-après) de l'une des deux extrémités de la relation considérée.

target index dans le tableau *nodes* (voir ci-après) de l'autre extrémité de la relation considérée.

id identifiant unique de la relation (peut-être utilisé pour manipuler cette relation de façon certaine dans l'application). Il s'agit d'une somme md5 (donc un nombre hexadécimal).

linktitle label de la relation.

size taille de la relation calculée d'après les paramètres entrés dans l'interface de configuration du réseau.

infos justification du calcul de « size », sous la forme d'un fragment de page html directement exploitable en sortie d'un appel Ajax.

keywords tableau rappelant l'ensemble des mots clés caractérisant la relation considérée.

nodes Tableau contenant l'ensemble des nœuds du réseau étudié.

entity label du nœud considéré.

country_origin et country_operating respectivement le pays d'origine du nœud considéré et le pays de son activité effective. Par exemple, Amazon France est considéré comme une organisation dont le « country_origin » sont les USA et le « country_operating » est la France.

type type de l'organisation considérée, parmi les valeurs suivantes : industriel, académique, pme.

date_incomming année de première apparition dans le réseau.

visible visibilité du nœud dans le réseau. Par défaut, seuls les nœuds possédant au moins une relation avec un autre nœud sont affichés. Les nœuds solitaires, bien que transmis dans les données, ne sont pas affichés.

affiliation tableau représentant la hiérarchie d'affiliation de l'organisation considérée. L'ordre des éléments dans ce tableau est important : ils sont indexés de l'élément le plus général au plus particulier. Chaque élément comporte les propriétés suivantes :

name Nom de l'entité à ce niveau de l'affiliation.

user-type Typage de cette entité à ce niveau de l'affiliation, tel qu'entré par l'utilisateur lors de l'enregistrement de l'organisation.

keywords tableau de mots-clés caractérisant ce nœud.

id rappel de l'index de ce nœud au sein du tableau *nodes*, tel qu'utilisé par exemple dans le tableau *links*.

mergein champ spécial utilisé dans le cadre d'une fusion d'informations, pour signifier le nouveau nœud (ajouté en fin de tableau *nodes*) résultant d'une fusion entre le nœud considéré et un autre nœud.

phpids Rappel des identifiants originaux des entités utilisées pour concevoir le tableau d'affiliation du nœud considéré sous la forme d'un simple tableau.

Voici un extrait de fichier JSON formaté par le convertisseur :

```
{"links":[
  {
    "source": 1, // index de l'une des extrémité de la relation
    "target": 0,
    "id": "6b3c",
    "size": 3.667,
    "linktitle": "Lien entre Delft Univ. Technology et TNNL",
    "infos": "",
    "keywords":[
      "multitarget particle filter",
      "track extraction",
      "Markov chain",
    ]
  },
  {...},
  ...
],
"nodes":[
  {
    "entity":"TNNL",
    "country_origin":"Netherlands",
    "country_operating":"Netherlands",
    "type":"",
    "date_incoming":"2008",
    "visible": true,
    "affiliation":[
      {"name":"TNNL", "user-type":"Legal Entity"},
      ...
    ],
    "keywords":["multitarget particle filter", "track extraction"],
    "mergein":0,
    "id":"0",
    "phpids":[""]
  },
```

```
35    {...},
36    ...
37  ]}
```

D.2 Format de sortie pour le Synthesizer

Autant le format JSON précédent permet de représenter un graphe dont les liens sont des agrégations de collaborations plus ou moins affectées d'un coefficient, autant ce format-ci ne sert qu'à transmettre les liens bruts unissant des entités deux à deux.

De ce fait, la spécification des nœuds dans le tableau *nodes* est la même que précédemment. Par contre, la spécification des éléments du tableau *links* est modifiée, du fait qu'il ne s'agit plus d'agrégats mais d'actes de collaboration.

Chaque élément possède ainsi les propriétés suivantes :

source, target, keywords et id Même signification que précédemment.

start_date, end_date Date de début et date de fin de la collaboration considérée.

format_date Format utilisé pour décrire les dates précédemment citées. La valeur de ce champ est soit « year » si la seule information disponible est une année, « timestamp » si l'information de date est plus complète. Il s'agit alors d'un timestamp unix (nombre de millisecondes depuis le 1er janvier 1970).

achievement État final de la collaboration si disponible. Cette propriété sert principalement pour l'instant dans le cadre de l'analyse des réponses d'appel à projet, à savoir si la proposition en question était passée ou pas.

vecteur Grand type de collaboration. Il peut s'agir de publication, produit, etc. Il s'agit en fait de l'expression des propriétés « IP type » ou « Vecteur » des données en entrée, quand disponibles.

type Un raffinement de la notion précédente de vecteur. Si le Vecteur étudié est « publication », son type pourra être « journal » ou « conférence ».

title Label de la collaboration. Il s'agira fréquemment du titre d'une publication sous-jacente. Cette propriété correspond à la propriété « title » des formats d'entrée, quand disponible.

cause et causename Ces propriétés permettent de caractériser le cadre plus grand dans lequel a eu lieu la collaboration. Si l'acte considéré est un article de conférence, « cause » donnera classiquement le nom de la conférence, tandis que « causename » donnera le titre des *proceedings* dans lequel l'article a été publié.

Voici un extrait de fichier JSON formaté par le convertisseur :

```
{”nodes”:[
  {
    ”id”:0,
    ”entity”:”TCS”,
    ”country_origin”:”France”,
    ”country_operating”:”France”,
    ”type”:””,
    ”affiliation”:[
      {”name”:”TCS”, ”user-type”:”Legal Entity”},
      ...
    ],
    ”keywords”:[”data processing”, ”geographic”, ...],
    ”id”:”0”
  },
  ...
],
”links”:[
  {
    ”source”:3,
    ”target”:2,
    ”start_date”:”2008”,
    ”end_date”:”2008”,
    ”format_date”:”year”,
    ”achievement”:”accepted”,
    ”keywords”:[”particle filter”, ”track extraction”, ...],
    ”vecteur”: ”publication”,
    ”type”:”conference”,
    ”title”:”2D spatial model matching using HRR multi-radar data”,
    ”cause”:”International Conference on Information Fusion”,
    ”causename”:”International Conference on Information Fusion”,
    ”id”:”6b3c8d2cbd0194f476a4c07edff3314f”
  },
  ...
]}
```

E. Acronymes

AJAX *Asynchronous Javascript And XML*

API *Application Platform Interface*

BIBO *The Bibliographic Ontology*

C2 *Command and Control*

CRM *Customer Relationship Management*

DGA Direction Générale de l'Armement

DSS *Decision Support System*

ERP *Enterprise Resource Planning*

FoaF *Friend of a Friend*

HTTP *HyperText Transfer Protocol*

IA intelligence artificielle

IC Ingénierie des Connaissances

ICI Information, Connaissance, Interaction

IANA *Internet Assigned Numbers Authority*

ISO *International Organization for Standardization*

KM *Knowledge Management*

LRASC Laboratoire de Raisonnement et d'Analyse dans les Systèmes Complexes

MOAT *Meaning of a Tag*

OCSO *Object Centered Sociality Ontology*

OGP *Open Graph Protocol*

OLAP *On-Line Analytical Processing*

ONG organisation non-gouvernementale

OODA Observer, Orienter, Décider, Agir

OWL *Web Ontology Language*

PME petite ou moyenne entreprise

RDF *Resource Description Framework*

RDFS *RDF schema*

RSS *Really Simple Syndication*

SaaS *Software as a Service*

SDSS *Social Decision Support System*

SI système d'information

SIAD système interactif d'aide à la décision

SIG système d'informations géographiques

SIOC *Semantically-Interlinked Online Communities*

SNA *Social Network Analysis*

SKOS *Simple Knowledge Organization System*

STI Sciences et Techniques de l'Information

URI *Uniform Resource Identifier*

URL *Uniform Resource Locator*

UTC Université de Technologie de Compiègne

W3C *World Wide Web Consortium*

Ce document a été rédigé à l'aide du mode Org de l'éditeur de texte GNU Emacs et mis en page avec le logiciel de composition typographique LaTeX.
version du 17 février 2014 — 02 :09

Titre Création de nouvelles connaissances décisionnelles pour une organisation via ses ressources sociales et documentaires

Résumé L'aide à la décision se fonde sur l'observation d'un environnement évolutif dont on scrute les évènements. Ces évènements peuvent être de différentes natures, dont les connexions qui peuvent se créer au sein d'un réseau d'acteurs. L'observation des bases documentaires ne semble plus suffisante pour nourrir l'aide à la décision. En effet, les nouveaux outils de communication et de collaboration, dont l'usage se répand rapidement au sein des organisations, sont sources de nouvelles formes d'informations peu ou mal utilisées par les systèmes actuels d'aide à la décision des organisations.
L'objectif de la thèse est de concevoir une plate-forme (modélisation et développement) pour les organisations permettant à leurs membres de bénéficier de médias sociaux et à leurs décideurs de bénéficier d'outils d'aide à la décision prenant en compte tous les types de ressources circulant sur cette plate-forme.

Mots-clés Gestion de connaissances, Médias sociaux, Réseaux sociaux, Web sémantique, Applications Web, Systèmes d'aide à la décision, Analyse des réseaux sociaux

Title Creation of New Decisional Knowledge for an Organization by Analysing its Social and Documentary Resources

Abstract Decision support is partly based on the observation of a dynamic and mutating environment (Situation Awareness). The events of such environments can be of different types, including new relations created within a network of actors. We think classical documentary databases are no longer sufficient to serve situation awarness. The quick spread and adoption of new communication and collaboration tools in organizations, bring new kind of information, like the social network of the organization, which are currently not or badly taken in account by organizational decision support systems.
The aim of this thesis is to design a platform, which provides to organizations both the social media to help their members to collaborate and the decision tools, which take into account all types of information exchanged in the platform.

Keywords Knowledge Management, Social Media, Social Networks, Semantic Web, Web Applications, Decision Support System, Social Networks Analysis