

**République Algérienne Démocratique et Populaire**  
**Ministère de l'Enseignement Supérieur et de la Recherche scientifique**  
**Université Mentouri Constantine**  
**Faculté des Sciences de l'Ingénieur**  
**Département d'Informatique**

N°ordre :

N°série :

**Thèse**  
**Pour obtenir le diplôme de**  
**Doctorat en Sciences**  
**Spécialité**  
**Informatique**

---

*TITRE*

---

**Une approche basée annotation  
sémantique de documents pour la gestion  
d'une mémoire d'entreprise**

---

Présentée par :  
**SONIA GUERAICH Epouse BENMAZA**

**Soutenue le 28 / 02 / 2012**

**Devant le Jury:**

<b>M. Mahmoud Boufaïda</b>	Professeur, Université Mentouri Constantine	Président
<b>Mme Zizette Boufaïda</b>	Professeur, Université Mentouri Constantine	Rapporteur
<b>M. Med Tayeb Laskri</b>	Professeur Université Badji Mokhtar Annaba	Examineur
<b>M. Zaidi Sahnoun</b>	Professeur Université Mentouri Constantine	Examineur
<b>M. Smaine Maazouzi</b>	Maitre de Conférence, Université Skikda	Examineur

# Remerciements

*« Une thèse est une sorte d'égoïsme en soi, eh oui, la plume transformé en clavier d'ordinateur de nos jours, demande silence et solitude pour la faire danser, sur un papier blanc, qu'on ne touche jamais, mais on a l'illusion de le voir monter et descendre dans l'écran, je remercie donc les personnes qui ont compris le besoin de ma solitude, et non rien obtenu en échange. » Leoncio JIMÉNEZ CANDIA*

*La réalisation de cette thèse de doctorat demeure une expérience qui ne se répète qu'une fois dans la vie comme la naissance et la mort. Ma gratitude éternelle s'adresse à Dieu le Tout Puissant pour m'avoir éclairé le chemin et m'avoir donné le courage pour son achèvement.*

*Les personnes « tacites » et « explicites » qui ont participé de près ou de loin à sa réalisation sont toutes vivement remerciées, quel que soit, le degré de leur contribution. Néanmoins j'insisterai sur des « étoiles » sans qui je n'en serai pas à ce stade.*

*Professionnellement et personnellement, je remercie en premier lieu, le professeur Zizette Boufaïda pour ses précieux conseils, son soutien moral et ses critiques pertinentes et fructueuses surtout durant les exposés. La patience et la compréhension « féminine » éprouvée à mon égard me poussent à admirer avec le grand respect du monde ses qualités humaines et professionnelles. Tous mes respects et ma gratitude sont dressés à vous. Merci d'avoir dirigé jusqu'au bout ce travail et de m'avoir encouragé tout au long de ces années.*

*Mes remerciements sont également orientés envers le Professeur Mahmoud Boufaïda, chef de l'équipe SIBC et directeur du laboratoire LIRE pour m'avoir accueilli dans son équipe depuis les années du Magister et jusqu'à ce jour. Ses encouragements, sa franchise et ses conseils sincères ainsi que ses encouragements m'étaient un support incontestable. Je le remercie pour l'honneur qu'il me fait en présidant le jury de ma thèse. Je tiens à remercier infiniment Mr et Mme Boufaïda pour les efforts, la passion et la rigueur consentis pour procurer à notre équipe les meilleures conditions de travail et de recherche. Qu'ils trouvent ici l'expression de ma profonde reconnaissance.*

*Je tiens à exprimer mes plus vifs remerciements à mes examinateurs les Professeurs Med-Tayeb Laskri et Zaidi Sahnoun ainsi qu'au Dr Smaine Maazouzi pour avoir accepté d'examiner, de lire et d'évaluer mon travail à fin de le perfectionner et le faire évoluer.*

*Je remercie vivement le personnel de l'ENMTP spécialement le directeur Mr Berkani.K ainsi que mes collègues que ce soit du département d'informatique, du laboratoire LIRE mais surtout ceux de notre équipe SIBC permanents, doctorants et magisters. Je citerai, en particulier, Dr Salima Hacini et Dr Adla Bentellis qui m'ont beaucoup soutenu moralement. C'était une formidable expérience et des moments de plaisirs ceux vécus parmi vous que ce soit durant les discussions, les réunions ou les exposés. dans la finalisation de la rédaction. Je leur suis, à tous reconnaissante pour le soutien moral qu'ils m'ont prodigué tout le long de la réalisation de cette thèse.*

*Personnellement, je ne saurais exprimer assez de reconnaissance pour mes parents pour leur amour et soutiens qui n'ont cessé de jalonner ma vie. Merci d'avoir cru en moi, je vous dédie ma vie et humblement ce modeste effort. A chacun des membres de ma famille en particulier ma grande mère, mon mari, ma sœur et mes deux frères pour leur amour et encouragement inconditionnels.*

*Que tous ceux que mes lignes ont loupé non par ignorance mais par peur d'oubli soient aussi remerciés vivement. Que Dieu protège nos vies et notre patrie.*

**A**

***Mes origines :***

*Ma grand-mère, ma mère, mon père....*

**A**

***Mes descendants :***

*Akrame, Anis et Achraf.....*

**A**

***Mon conjoint :***

*Noureedine*

**A**

***Tous ceux à qui je pense ...***

# Résumé

Une entreprise est non seulement une unité de production de services ou de biens pour répondre aux espérances des clients, mais c'est également une unité de production de connaissances. Ces connaissances doivent répondre aux différents besoins dans les meilleures conditions de coût, de qualité et de délai. Ceci amène à réfléchir au nombre croissant de connaissances que les entreprises doivent gérer et capitaliser par leurs employés dans un contexte de concurrence exacerbée. La mémoire d'entreprise est alors, une façon de capitaliser ces connaissances et la problématique autour de cette notion se focalise essentiellement sur la question de la gestion de cette mémoire.

Afin de répondre à ce besoin de partage et de réutilisation, dans cette thèse nous proposons une approche basée annotation sémantique de documents pour la gestion d'une mémoire d'entreprise. Cette démarche supporte l'ensemble des étapes du cycle de vie de la Gestion des connaissances. Néanmoins, il n'existe pas, à notre connaissance, une démarche complète permettant de valoriser, au sens large, un capital de connaissances. Dans ce contexte, notre effort est une tentative vers cette valorisation.

Pour ce faire, nous utilisons l'annotation sémantique de documents d'entreprise. Cette approche exploite elle-même une ontologie. En premier lieu, nous nous sommes intéressés à la proposition d'un modèle pour la ME (Mémoire d'Entreprise). Ce dernier est inspiré du modèle de la « Marguerite » utilisé initialement pour la gestion d'un cycle de vie pour le KM (Knowledge Management) et non pour une ME. Dans ce modèle, nous considérons que la construction d'une mémoire d'entreprise prend en compte un ensemble de facteurs qui considèrent aussi bien les aspects humains et organisationnels que les aspects techniques et économiques. En second lieu, nous nous sommes intéressés à la conception de l'ontologie. Dans le processus d'annotation sémantique, les ontologies jouent un rôle primordial puisqu'elles modélisent les concepts, leurs attributs et les relations pour annoter le contenu des constituants documentaires de la ME. En troisième lieu, nous nous sommes focalisés sur la définition d'un processus pour l'annotation sémantique qui exploite l'ontologie

précédemment définie. Nous avons proposé CMOCA (Corporate Memory by Ontology and Collaborative semantic Annotation) architecture. Cette architecture se compose de quatre modules et exploite des bases de connaissances pour le stockage des annotations sémantiques. Le Web Sémantique d'Entreprise (WSE) est l'approche technique choisie pour matérialiser notre ME. Finalement, comme étude de cas, nous nous sommes orientées vers un type particulier de ME qu'est la Mémoire de Projet (MP). Nous définissons une MP comme une mémoire des connaissances acquise et produite au cours de la réalisation des projets. Nous nous sommes intéressés à la mémoire de projet dans le domaine de la conception. Dans ce contexte, cette MP doit permettre l'accès à des informations décrivant aussi bien les caractéristiques d'un projet que celles relatives à la résolution des problèmes rencontrés lors de la réalisation d'un projet.

Mots clés : Gestion des connaissances, mémoire d'entreprise, annotation sémantique, ontologie.

# Abstract

A company is not only a production of goods or services to meet customers' expectations, but it is also a unit of knowledge production. This knowledge must meet the different needs in the best conditions of cost, quality and timeliness. This raises to reflect on the growing knowledge that companies must manage and leverage their employees in a context of heightened competition. Corporate memory is then a way to capitalize on this knowledge and the issues around this concept focuses mainly on the issue of managing this memory.

To meet this need for sharing and reuse, this work proposes an approach based semantic annotation of documents for the management of a corporate memory. This approach supports all stages of the life cycle of Knowledge Management. However, there is not, to our knowledge, a comprehensive approach to enhancing the broad sense, a knowledge capital. In this context, our effort is an attempt towards this development.

The proposed approach uses semantic annotation, which itself uses an ontology. First, we looked at the proposal of a model for the ME (corporate memory). The latter is based on the model of "Marguerite" originally used to manage a life cycle for the KM (Knowledge Management) and not for ME. In this model, we believe that building a corporate memory takes into account a range of factors that consider both the human and organizational as technical and economic aspects. Second, we focused on the design of the ontology. In the process of semantic annotation, ontologies play a crucial role as they model the concepts, their attributes and relationships used to annotate the content of the constituents of the documentary ME. Third, we focused on defining a process for semantic annotation that exploits the ontology defined previously. We proposed CMOCA architecture (Corporate Memory by semantic and Collaborative Ontology Annotation). This architecture is composed from four components and uses the knowledge base for storing semantic annotations. The Semantic Web Company (WSE) is the technical approach chosen to materialize our ME. Finally, as a case study, we referred to a particular type of ME that is the Memory Project (MP). We define an MP as a memory of knowledge acquired and produced during the project implementation. We are interested in the project memory in the field of design. In this context, the MP must allow access to information describing both the characteristics of a project than on solving problems encountered during a project.

Key words: Knowledge management, corporate memory, semantic annotation, ontology.

## ملخص

الموسسة ليست فقط وحدة لانتاج الخدمات لتلبية احتيلجات زبائنها بل هي وحدة لانتاج المعارف لتلبية احتيلجات في أحسن الظروف النوعية الزمنية. هذا ما يدفع للتفكير في الطرق التي يجب انتهاجها للحصول علي أحسن المعارف الضرورية للتسيير في وسط تنافسي.

إذا فذاكرة الموسسة هي من بين الطرق للحصول علي أحسن المعارف و لكن المشكلة هي في كيفية تسيير هذه الموسسة.

الأنطولوجيا هي التمثيل المفاهيمي لميدان معين مما يسمح للجهات الفاعلة البشرية و البرامج لتقاسم المعارف. إنها توفر الوسائل اللازمة للتعبير عن المفاهيم باستخدام لغة تمثيل المفاهيم هي نتيجة لتوافق في الاراء بين مختلف الجهات المعنية و مسخذي الأنطولوجيا.

في هذه الرسالة نحن نركز علي مشكلة تسيير ذاكرة الموسسة مع الأخذ بعين الإعتبار إعطاء نموذج لهذه الذاكرة. لهذا السبب استنبطنا نموجا من "أرمين" مع إعطاء تغييرات عليه. النموذج المعطي يساهم في تسيير الذاكرة مع الأخذ بعين الإعتبار دورة حياة المعارف.

لتقديم الإجابات علي هذه المشكلة من وجهة نظر الواب المعنوي. قمنا باستعمال طريقة التعليمات المعنوية المعتمدة علي الأنطولوجيا. و كنتيجة لذلك اقترحنا الأنطولوجيا خاصة للموسسة و قاعدة هندسية

CMOCA (Corporate Memory by Ontology and Collaborative semantic Annotation)

في ظل (WSE) الواب المعنوي للموسسة .

### الكلمات المفتاحية:

تسيير المعارف, ذاكرة الموسسة, التعليمات المعنوية, الأنطولوجيا

# *Sommaire*

# Table des matières

## ***Introduction générale***

1. Contexte.....	16
2. Problématique.....	17
3. Motivation et objectifs.....	20
4. Organisation de la thèse.....	23

## ***CHAPITRE 1 : Gestion des connaissances pour l'entreprise***

1. Introduction .....	25
2. Les connaissances un enjeu pour l'entreprise .....	25
2.1 Contextes et défis .....	26
2.2 Création des connaissances dans l'entreprise.....	27
2.3 Typologies des connaissances dans l'entreprise .....	29
3. Gestion des connaissances.....	31
3.1 Définition de la gestion des connaissances .....	31
3.2 Aspects pour la gestion des connaissances .....	33
3.2.1 Aspect social.....	33
3.2.1.1 Approche organisationnelle .....	34
3.5.1.2 Approche managériale .....	34
3.2.2 Aspect technique .....	34
3.3 Quelques approches de la gestion des connaissances.....	35
3.4 Systèmes pour la gestion des connaissances.....	35
3.4.1 Les Systèmes à Base de Connaissances (SBC).....	35
3.4.2 La mémoire d'Entreprise (ME).....	36
4. Conclusion.....	36

## ***CHAPITRE 2 : Mémoire d'Entreprise et Capitalisation des Connaissances***

1. Introduction.....	38
2. Définition .....	39
3. Typologies.....	41
4. Besoins industriels en mémoire.....	42

5. Quelques méthodes de capitalisation.....	44
5.1 La méthode REX.....	44
5.2 La méthode MKSM.....	46
6. Cas spécial : Mémoire de projet .....	47
6.1 Définition de la mémoire de projet.....	47
6.2 Modèles de mémoire de projet .....	48
6.2.1 Mémoire caractéristiques de projet.....	48
6.2.2 Mémoire de logique de conception.....	48
6.3 Méthodes de capitalisation dédiées aux mémoires de projet.....	49
6.3.1 La méthode IBIS.....	49
6.3.2 L'approche QOC.....	50
6.3.3 Le système DRCS.....	51
6.3.4 Le système DRAMA.....	52
6.3.5 L'approche EMMA.....	52
6.3.6 La méthode SAGACE.....	53
6.7 Discussion.....	53
7. Conclusion.....	54

### ***CHAPITRE 3 : Du Web sémantique au Web sémantique d'entreprise***

1. Introduction.....	55
2. Principales composantes du Web sémantique .....	55
3. Ontologies .....	56
3.1 Types d'ontologies .....	58
3.2 Réutilisation des ontologies .....	58
4. Annotation classique.....	59
4.1 Définition d'une annotation .....	59
4.2 Méta-donnée ou Annotation .....	61
4.3 Quelques modèles pour l'annotation classique.....	61
4.3.1 Modèle de Denoue .....	61
4.3.2 Modèle de VERON .....	61
4.3.3 Modèle d'Annotea .....	62
4.4 Acteurs interagissant avec l'annotation classique .....	63
4.5 Cycle de vie des annotations classiques .....	63
4.6 Dimensions des annotations classiques .....	64

4.7 Quelques outils pour l'annotation classique.....	65
4.8 Synthèse .....	66
5. Annotation sémantique.....	67
5.1 Définition de l'annotation sémantique .....	67
5.2 Annotation et méta-donnée sur le web sémantique .....	68
5.3 Caractéristiques des annotations sémantiques .....	69
5.4 Processus d'annotation sémantique .....	69
5.5 Exploitation des annotations sémantiques.....	69
5.6 Méthodes et éléments de base pour l'annotation sémantique .....	70
5.7 Langages de l'annotation sémantique .....	71
6. Web sémantique d'entreprise .....	74
7. Conclusion.....	75

## ***CHAPITRE 4 : Une approche de modélisation pour la gestion d'une mémoire d'entreprise***

1. Introduction .....	77
2. Etapes de modélisation .....	78
2.1 Repérer les besoins .....	79
2.2 Construire la mémoire .....	80
2.2.1 Spécification de l'ontologie.....	81
2.2.2 Conceptualisation de l'ontologie.....	83
2.2.3 Formalisation de l'ontologie.....	90
3. Description du modèle .....	92
3.1 Pétales d'interactions .....	96
3.2 Pétales de capitalisations .....	97
3.3 Fonctionnement du modèle .....	97
4. Adaptation du modèle pour une mémoire de projet.....	98
4.1 Caractéristiques de la logique de conception .....	99
4.2 Représentation de la mémoire de projet.....	100
4.2.1 Contexte.....	101
4.2.2 Logique de conception.....	102
4.2.3 Ontologie .....	103
4.3 Modélisation de la mémoire de projet pour une entreprise.....	104
4.3.1 Modèle de cas d'utilisation .....	105

4.3.1.1	Gestion produits.....	106
4.3.1.2	Gestion DES acteurs.....	108
4.3.1.3	Gestion des documents.....	108
4.3.1.4	Gestion des processus.....	109
4.3.1.5	Gestion des ressources.....	110
4.3.1.6	Gestion de l'organisation.....	111
4.3.1.7	Gestion des versions.....	112
4.3.2	Modèle de classes .....	113
5.	Conclusion.....	114

## ***CHAPITRE 5 : Une approche pour l'annotation d'une mémoire d'entreprise***

1.	Introduction .....	115
2.	Hypothèses .....	115
2.1.	Critère 1 : Besoins en Mémoire d'Entreprise.....	116
2.2.	Critère 2 : Classification des Mémoire d'Entreprise.....	116
2.3.	Critère 3 : Modèle pour la mémoire d'entreprise.....	116
2.4.	Critère 4 : Support de documents hétérogènes.....	116
2.5.	Critère 5 : Support de documents intelligents.....	117
2.6.	Critère 6 : Support de l'évolution des documents.....	117
2.7.	Critère 7 : Support de l'évolution des documents.....	117
2.8.	Critère 8 : Centrée sur l'utilisateur.....	117
2.9.	Critère 9 : Stockage des annotations.....	117
2.10.	Critère 10 : Automatisation.....	118
3.	Modèle d'annotation.....	118
4.	Approche proposée .....	120
4.1.	Description .....	121
4.1.1.	Annotation à partir des représentations conceptuelles.....	122
4.1.2.	Annotation à partir du texte.....	125
4.1.2.1.	Etape manuelle.....	126
4.1.2.2.	Etape semi automatique.....	126
4.2.	Description de l'architecture CMOCA.....	128
4.3.	Description du fonctionnementde l'architecture CMOCA.....	131
4.3.1.	Analyses utilisées.....	131
4.3.1.1.	Analyse linguistique.....	132
4.3.1.2.	Analyse sémantique.....	132

4.3.1.3. Analyse statistique.....	133
4.3.2 Détection des connaissances cruciales .....	134
4.3.2.1. Analyse morpho-syntaxique.....	134
4.3.2.2. Extraction des termes .....	134
4.3.2.3 Repérage des relations sémantiques.....	135
5. Validation.....	136
5.1. Organisme d'accueil .....	136
5.1.1. Description du problème.....	137
5.1.2. Description du scénario.....	141
5.1.3. Approche d'annotation sémantique.....	142
5.2. Environnement de développement.....	145
5.2.1. Outils du TALN utilisés.....	145
5.2.2. Outils de développement .....	152
5.2.3. Outil de création les ontologies.....	152
6. Conclusion.....	
<b><i>Conclusion générale et perspectives</i></b>	
1. Bilan .....	156
2. Limites et perspectives.....	158
<b>Bibliographie</b> .....	159
<b>Annexes</b> .....	170

## LISTE DES FIGURES

Figure 1.1.	Pôles majeurs dans l'entreprise	26
Figure 1.2.	Modèles de transformation des connaissances	28
Figure 2.1.	Principe de base de la méthode REX	44
Figure 2.2.	Triangle sémiotique de MKSM	46
Figure 2.3.	Modélisation du système de connaissances	47
Figure 2.4.	Principaux éléments de la méthode IBIS	50
Figure 2.5.	Représentation de la prise de décision selon QOC	51
Figure 3.1.	Problème de réutilisation/utilisation	59
Figure 3.2.	Relation entre l'annotation et la Méta-donnée	61
Figure 3.3.	Modèle Annotea d'une annotation	63
Figure 3.4.	Cycle de vie des annotations	64
Figure 3.5.	Schéma récapitulatif des types d'annotations	65
Figure 3.6.	Annotation sémantique	67
Figure 3.7.	Utilisation des méta-données sur le Web sémantique	68
Figure 3.8.	Exemple d'annotation sémantique en RDF	72
Figure 3.9.	Architecture du web sémantique d'entreprise	75
Figure 4.1.	Etapes de Gestion d'une mémoire	78
Figure 4.2.	Document résultat de la spécification	83
Figure 4.3.	Etapes de conceptualisation	83
Figure 4.4.	Ontologie ME	89
Figure 4.5.	Ontologie Project Mémoire	89
Figure 4.6.	Exemple de la partie terminologique	91
Figure 4.7.	Modèle de la Marguerite originale	92
Figure 4.8.	Modèle de la Marguerite étendu	94
Figure 4.9.	Modèle pour la mémoire de projet	99
Figure 4.10.	Composantes de la mémoire de projet	101
Figure 4.11.	Contexte	101
Figure 4.12.	Scénario de la logique de conception	102
Figure 4.13.	Diagramme de cas utilisation associé d'une MP	105
Figure 4.14.	Diagramme d'activités associé à la gestion de produits	106
Figure 4.15.	Diagramme d'activités associé à la gestion des acteurs	107
Figure 4.16.	Diagramme d'activités associé à la gestion de documents	108
Figure 4.17.	Diagramme d'activités associé à la gestion de processus	109
Figure 4.18.	Diagramme d'activités associé à la gestion des ressources	110
Figure 4.19.	Diagramme d'activités associé à la gestion de l'organisation	111
Figure 4.20.	Diagramme d'activités associé à la gestion de versions	112
Figure 4.21.	Diagramme de paquetages	113
Figure 4.22.	Patron décrivant une décomposition arborescente	113
Figure 4.23.	Diagramme de classes pour « Organisation »	114
Figure 5.1.	La fonction d'annotation	120
Figure 5.2.	Protocole d'appel d'offre pour annoter	121
Figure 5.3.	Définitions des différents stéréotypes utilisés pour l'annotation	123
Figure 5.4.	Exemple de stéréotypes utilisés pour l'annotation	124
Figure 5.5.	Approche d'annotation	125

Figure 5.6.	Méthode de détection	126
Figure 5.7.	L'architecture CMOCA de base	127
Figure 5.8	L'architecture CMOCA de base avec médiateur.	129
Figure 5.9	L'architecture CMOCA de base avec serveur de connaissances	130
Figure 5.10.	Etapes d'analyses	132
Figure 5.11	Ontologie conçue	139
Figure 5.12	Exemple de phrase annotée	140

## LISTE DES TABLES

Table 2.1. Table de décision DRAMA .....	52
Table 2.2. Trois visions définies dans SAGACE.....	53
Table 4.1 Liste des concepts et des relations .....	87
Table 4.2 Quelques opérateurs logique de description .....	90
Table 4.3 Composantes du modèle.....	95
Table 4.4 Typologies des problèmes .....	103
Table 5.1 Correspondances entre UML et RDF/RDFS .....	122
Table 5.2 Stéréotypes liés à la génération de l'ontologie .....	123
Table 5.3 Stéréotype lié à la génération de la base de connaissance .....	123
Table 5.4 Matrice de cadre logique .....	138
Table 5.5 Logique de zigzag .....	139
Table 5.6 matrice annotée classiquement .....	140
Table 5.7 Evaluation1 à partir des critères.....	141
Table 5.8 Evaluation2 à partir des critères.....	143
Table 5.9 Précision, rappel et mesure.....	145

# Introduction générale

*« Dès le début de la connaissance, se pose la question de la vérité. Son introduction fait du processus de la connaissance humaine un problème de savoir. ».*

*Glaserfeld*

## 1. Contexte

La Gestion des connaissances, plus connue par le Knowledge Management (KM) est aujourd'hui importante et nécessaire pour transformer et actualiser le fonctionnement du savoir et également le savoir-faire développé par le personnel. Ceci permet de créer un système interactif et compétitif de formation continue qui débouche sur une meilleure qualité des produits et services. Pour Jean-Louis Ermine [Ermine, 2000], la gestion des connaissances « est revenue d'outre-atlantique avec une vigueur et une force qui en font un des mots-clés des entreprises actuellement ».

Le KM sert à décrire des approches, des démarches, des méthodes, des outils, etc. tous dédiés à la connaissance au sein de l'entreprise. Les principales méthodologies de capitalisation des connaissances ont pour but l'analyse des Systèmes de Connaissances (SC) d'entreprise et la création de Mémoires Organisationnelles (MO). Ces deux perspectives étant indispensables à tout projet de gestion de connaissances.

La notion de la capitalisation des connaissances est utilisée pour désigner la gestion des connaissances. Ce concept a été énoncé dès 1990 chez Framatome dans le prolongement de la démarche de déploiement de l'Intelligence Artificielle et des Systèmes à Base de Connaissances [Grundstein, 1995]. Il s'agissait de pérenniser et de valoriser le savoir-faire acquis en Ingénierie des Connaissances.

La capitalisation des connaissances est donc un processus qui permet de réutiliser les connaissances d'un domaine donné, préalablement modélisées et stockées afin de mieux effectuer de nouvelles tâches [Simon, 1996]. Ces connaissances constituent un capital dont l'entreprise doit faire augmenter la valeur. De ce fait, nous allons constater le concept de capitalisation des connaissances au sens de la gestion des connaissances. Cette idée est soutenue par [Barthès, 1996] pour qui capitaliser les connaissances implique que l'on constitue un capital qui sera ensuite valorisé.

Cependant, beaucoup d'entreprises font le constat que le KM traditionnel ne suffit plus à les rendre performantes et à assurer leur pérennité dans un environnement en purement concurrentiel. En effet, les exigences des clients et du marché sont si multiples et si évolutives et le personnel de l'entreprise est si nombreux et si varié, que l'entreprise se trouve en situation de changements et remise à niveau perpétuelle. C'est la raison pour laquelle l'enjeu aujourd'hui est de se munir de nouveaux moyens de stockage et de partage pour réagir aux changements et rester à jour dans un monde en constante évolution.

Le partage de la connaissance permet à chaque acteur d'améliorer ses performances et donc celles de l'entreprise. De ce fait, pour une entreprise, la perte due à une réduction de son effectif (départ en retraite, mutation, etc.) est minorée si la gestion de son patrimoine de connaissances est jugée bonne. En archivant sa propre mémoire, l'entreprise évite aussi la répétition d'erreurs du passé et réutilise efficacement les expériences vécues.

La diffusion des connaissances via les nouvelles techniques de communication favorise la transversalité aux dépens de la hiérarchisation. La nouvelle génération du web (i.e. le web sémantique) est un moyen pour atteindre cet objectif en favorisant la sémantique et le sens de l'information. L'idée du web sémantique est de rendre le contenu des ressources diffusées sur le web interprétables par les programmes et les utilisateurs via divers moyens tels que l'explicitation, la conceptualisation et l'annotation.

Les ontologies de domaine sont devenues très populaires, en prouvant leur efficacité pour la représentation des connaissances dans le domaine de l'ingénierie de connaissances. L'annotation sémantique est une technique qui se base sur les ontologies et qui permet à l'annotateur d'incorporer l'information sémantique dans les documents. Tous deux, l'ontologie et l'annotation sémantique sont considérés comme des composants du web sémantique.

## **2. Problématique**

Depuis quelques années, les entreprises ont observé une véritable effervescence dans l'industrie, dans la recherche et dans les médias autour de la mémoire d'entreprise, de la capitalisation et de la gestion des connaissances. Mais les réflexions stratégiques dans les grands groupes industriels pour une meilleure exploitation de leur patrimoine de connaissances dans le nouvel ordre économique mondial témoignent d'un réel besoin industriel. Le partage du travail nécessite un partage de connaissances au sein des entreprises et ce besoin peut devenir encore plus crucial avec la réduction du temps de travail. Pour survivre et réussir les entreprises doivent non seulement gérer aux mieux leur capital de connaissances mais aussi innover et créer de nouvelles connaissances. En outre, un véritable partage de connaissances doit améliorer les conditions de travail des membres de l'entreprise [Dieng et al., 2006].

La capitalisation des connaissances dans une entreprise nécessite la gestion des ressources afin de faciliter leur accès et leur réutilisation [O'Leary, 1998]. Elle permet de favoriser la croissance, la transmission et la conservation des connaissances dans une organisation [Steels, 1993]. Ce problème très complexe peut être abordé de plusieurs points de vues : socio-organisationnels, économiques, techniques, humains et légaux [Barthès, 1996].

La mémoire d'entreprise est une nouvelle facette du KM qui tente d'apporter une solution aux problèmes qui perturbent le processus de capitalisation. C'est une sorte de diagnostic établi sur l'entreprise à partir de besoins réels. Notons la diversité des sources envisageables : spécialistes humains, documents papiers ou électroniques existants (notes, rapports, documents contractuels, documentation technique, comptes rendus de réunions...), messages échangés par courrier électronique, dictionnaires, nomenclatures, bases de données, thésaurus. Dans [Van Heist et al., 1996], la mémoire d'entreprise est définie comme la « représentation explicite, persistante et désincarnée des connaissances et des informations dans une organisation ». Elle peut inclure les connaissances sur les clients, les stratégies de vente, les produits, les résultats financiers etc., ces connaissances peuvent provenir de bases de données, de documents électroniques, de rapports, de spécification de produits etc.,

De nombreux travaux de recherche se sont penchés sur la gestion de la mémoire d'entreprise et l'ont proposé à différents niveaux et sous différentes formes. Dans la littérature, nous trouvons à titre d'exemple : la mémoire de groupe/d'équipe qui assiste une petite affaire d'équipe [Nunamaker et al., 1991], [Morrison, 1993], la mémoire de discussions qui préserve l'évolution de la conception d'un produit ou la discussion de politique dans la fabrication [Konda et al., 1992], la mémoire environnementale qui permet d'assister le processus de fabrication en interaction avec l'environnement de l'organisation [Elofson et Konsynski, 1991], la mémoire documentaire qui fournit un accès à un ensemble ciblé de documents décrits [Huhns et al., 1989] et la mémoire de projet utilisée pour l'assistance d'un projet habituellement de longue durée et avec des participants répartis [Lynch et Chen, 1992]. Pour ce dernier type de mémoire, peu sont les travaux qui l'ont abordé [Bekhti, 2003]. Mais en pratique, cette mémoire est plus répandue que la mémoire d'entreprise. Elle existe depuis plus longtemps et est notamment largement développée en conduite de projets informatiques. « Bénéficiant d'une tradition plus longue, les organisations mettraient plus d'avantages en pratique la mémoire de projet que la mémoire d'entreprise » [Pomian, 1996].

Dans la littérature sur la mémoire d'entreprise, la mémoire de projet est un volet restreint d'un exercice beaucoup plus large de capitalisation de tout un ensemble d'expériences diversifiées réalisées au sein de l'entreprise par l'ensemble des gestionnaires, voire l'ensemble du personnel [Girod, 1995]. En d'autres termes, les projets ne sont qu'un domaine et qu'une occasion parmi d'autres de documenter et de préserver les savoirs et savoir-faire des acteurs de l'entreprise les savoirs et savoir-faire

des acteurs de l'entreprise. Selon les auteurs, la mémoire de projet revêt toutefois une importance particulière. D'une part, contrairement aux autres activités de l'entreprise, les projets étant déterminés et ciblés dans le temps, lorsqu'ils sont terminés, les équipes sont dissoutes et les pratiques ne sont plus accessibles.

Le principal intérêt est de pouvoir, en théorie du moins, orchestrer les différents types de mémoire d'entreprise selon le contexte d'utilisation et les évolutions de l'environnement. Ils doivent analyser des problèmes et des solutions relatifs à chacune des étapes du cycle de vie d'une mémoire d'entreprise à savoir, la détection des besoins, la construction, la diffusion, l'utilisation, l'évaluation et l'évolution [Dieng et al., 2005]. La majorité des travaux ont l'avantage d'apporter un degré de modélisation pour chaque type de mémoire d'entreprise. Cependant, ils ont l'inconvénient de compliquer les modèles et de recourir à des formalismes qu'un expert ne peut ni suivre ni maîtriser.

De nombreuses études s'accordent sur l'intérêt de l'usage des ontologies dans le domaine de l'ingénierie des connaissances. Les ontologies sont devenues nécessaires pour pallier à cette problématique [Gandon, 2002]. Ces études montrent que les ontologies permettent des représentations formelles des connaissances car elles s'appuient sur le raisonnement jugé important dans le contexte de la gestion des connaissances.

Mais, comme nous nous intéressons à la concrétisation de la mise en œuvre de la mémoire d'entreprise, nous avons constaté que l'annotation sémantique issue initialement d'une gestion documentaire présentée par une ontologie peut se révéler particulièrement intéressante dans le contexte d'une mémoire d'entreprise. Cette technique consiste à associer à une ressource des métadonnées et des concepts de l'ontologie choisis par l'annotateur en adéquation avec le contenu de la ressource à décrire. L'annotateur devrait être un expert ou un spécialiste du domaine, afin de garantir un certain niveau de pertinence quant à la description des ressources du domaine.

La problématique majeure que cette thèse tente de traiter est de participer au mouvement de recherche pour proposer un modèle de mémoire d'entreprise qui sera projetée sur une mémoire de projet d'une façon plus adaptative. L'objectif est que la modélisation soit la plus proche possible des experts et des acteurs de l'entreprise. Ces derniers ont pour principales fonctions la construction et la définition d'une mémoire

d'entreprise qui aura une ampleur vitale sur le déroulement de l'entreprise. L'exploitation de cette mémoire via les techniques nouvelles du web sera possible par une approche basée sur l'annotation sémantique des documents pertinents de l'entreprise. De ce fait, les ressources de cette mémoire sont similaires à celles du Web, elles ont besoin d'être annotées sémantiquement pour mieux découvrir et utiliser les connaissances de ces ressources. Nous partageons la démarche des travaux qui proposent de matérialiser une mémoire sous forme d'un Web Sémantique d'entreprise [Dieng et al., 2005]. Les principales composantes de ce web sémantique sont les *ressources*, les *ontologies* qui décrivent le vocabulaire partagé par les différentes communautés de l'entreprise et les *annotations sémantiques* qui décrivent des méta-données sur les ressources en se basant sur les concepts et les relations de l'ontologie.

### **3. Motivation et objectifs**

La gestion des connaissances de l'entreprise est une problématique complexe et pluridisciplinaire nécessitant la prise en compte des aspects organisationnels, méthodologiques et techniques. Dans notre contexte de recherche, nous proposons de toucher ces trois aspects organisationnels, méthodologiques et techniques. L'objectif de ce travail est de contribuer à offrir à l'entreprise un outil de capitalisation pour lui permettre d'être plus réactive aux changements rapides et incessants du marché actuel et ce en lui permettant de définir plus aisément des mémoires d'entreprises facilement adaptables à l'évolution socio-économique.

De nouvelles fonctions ont vu le jour avec l'avènement du KM ce qui a révolutionné les modes d'organisation des entreprises. Dans l'entreprise, il y a rarement un spécialiste de la connaissance. Bien souvent, de nouveaux postes sont donc à créer, tels que: le Knowledge Manager (KM) et le Chief Knowledge Officer (CKO).

Cette thèse présente un modèle pour la mémoire d'entreprise inspiré du modèle de la « Marguerite ». Ce modèle prend en compte un ensemble de facteurs qui considèrent aussi bien les aspects humains et organisationnel que les aspects techniques et économiques.

Afin de matérialiser cette mémoire, nous avons choisi d'utiliser une ontologie de domaine pour l'entreprise. La conception d'une ontologie d'entreprise s'est avérée une tâche importante. Et dans la but, d'exploiter cette ontologie, nous avons opté pour un processus d'annotation sémantique des documents pertinents. De ce fait, l'ontologie

modélise les concepts, leurs attributs et les relations utilisés pour annoter le contenu documentaires de la ME.

Nous avons exploité l'ingénierie documentaire pour faciliter la formalisation des savoir-faire d'experts. Ceci consiste à présenter les connaissances sous une forme échangeable entre les individus qui peuvent manipuler des applications hétérogènes. L'utilisation d'XML comme support de formalisation et d'explicitation des savoir-faire offre l'avantage de structurer les connaissances selon des balises. Ces dernières sont fournies par les métadonnées et les concepts préalablement créés à partir de l'ontologie du domaine.

Dans cette acceptation, l'ontologie est utilisée comme un modèle de représentation conceptuelle des savoir-faire. Ces derniers peuvent également être présentés par des représentations conceptuelles telles que des diagrammes UML. C'est ainsi que la formalisation des savoir-faire peut reposer à la fois sur l'XML, les représentations conceptuelles et les ontologies.

Nous avons proposé CMOCA (Corporate Memory by Ontology and Collaborative semantic Annotation) architecture. Cette architecture se compose de quatre modules et exploite des bases de connaissances pour le stockage des annotations sémantiques. Le Web Sémantique d'Entreprise est le cadre technique général choisi pour matérialiser la mémoire.

Comparée aux travaux existants, notre proposition se positionne au carrefour de différents domaines tels que la gestion des connaissances, les représentations conceptuelles et le web sémantique. L'objectif de notre travail est donc, de construire une mémoire d'entreprise basée sur une ontologie du domaine exploitée par l'annotation sémantique et dont, un utilisateur de l'entreprise peut en tirer profit. Cet utilisateur exprime son besoin en connaissances structurées. Il sera à la fois un « client » et un « serveur » en matière de capitalisation. S'il aura besoin de connaissances, la ME pourra lui fournir ce qu'il veut. Et s'il se dotera de nouveaux savoirs faire qu'il voudra faire bénéficier, la ME lui permettra de les faire diffuser. Dans les deux cas, la réutilisation et le partage des connaissances sont des facteurs inhérents. Nous avons pensé que, pour faire responsabiliser les acteurs de l'entreprise dans le processus de la gestion des connaissances « moderne », il vaudrait mieux exploiter les techniques du web sémantique. Ceci permettra de vulgariser ces techniques dans le monde de nos entreprises.

Notre contribution intervient donc, sur trois parties principales :

- Une première partie propose une approche de modélisation pour la gestion d'une mémoire d'entreprise. Le fondement théorique de notre modèle repose sur le modèle de la Marguerite de Jean-Louis Ermine [Ermine, 1996]. Nous motivons

notre choix par le fait qu'il permet de guider, l'analyse de la complexité. Cette complexité est due à son couplage entre le système de connaissances qui gère le patrimoine de connaissances et l'environnement interne et externe de l'organisation. Mais, cette gestion de la complexité ne comprend pas le cycle de vie de la gestion des connaissances. Notre idée est d'incorporer via des pétales du cycle de vie, le cycle de développement d'une mémoire d'entreprise. Ensuite, au fur et à mesure du développement, nous considérons la capitalisation via des pétales dédiées. Nous avons dissocié le partage de la capitalisation. Nous avons également intégré la sélection dans le noyau. Ce qui permet d'approcher dans une vision globale, les mécanismes d'adaptation et d'évolution dans un contexte économique. De notre point de vue ceci permettra de garantir le maintien intégral de l'organisation.

- Une deuxième partie propose une approche d'annotation sémantique qui utilise une ontologie de domaine. Ceci permettra de mettre l'accent sur l'intérêt de l'usage des ontologies dans la gestion des connaissances et de montrer leur apport dans les processus de modélisation et de restitution des connaissances. Nous exploitons les ontologies pour la représentation des connaissances et l'utilisation de celles-ci pour l'annotation et la manipulation des ressources de connaissances formalisées. L'expérience montre qu'il est recommandé pour la formalisation de l'ontologie dans le cadre d'une application industrielle, d'utiliser un formalisme suffisamment expressif basé sur une syntaxe accessible. Ainsi, notre utilisation du langage de définition d'ontologies pour le Web, OWL, répond à ces critères. D'une part, OWL est doté d'une sémantique issue des logiques de description qui offrent une expressivité suffisamment précise, et d'une autre part, la syntaxe d'OWL est basée sur l'XML
- Une troisième partie présente l'architecture du système supportant le modèle ainsi que l'approche, afin de les intégrer dans une démarche globale pragmatique et opérationnelle.
- Une étude de cas a été réalisée pour expérimenter les étapes de la modélisation dans une entreprise nationale ENMTP de Ain Smara, filiale de Constantine. Cela permet d'exposer notre contribution à la résolution du problème posé concernant la gestion des connaissances de l'ENMTP. Cette contribution est concrétisée à travers une mémoire d'entreprise (spécialement une mémoire de projet) bâtie autour d'une ontologie du domaine. Pour la gestion des connaissances au sein de l'ENMTP, nous avons observé qu'il existe des problèmes de perte, d'accès et de

gestion des connaissances, ce qui a amené l'entreprise à perdre un capital important de connaissances ces dernières années. Chacun de ces facteurs est décrit et illustré par des exemples extraits de l'étude de l'existant.

#### **4. Organisation de la thèse**

Ainsi notre thèse vise à proposer une démarche pour la capitalisation des connaissances via la proposition d'une démarche d'annotation sémantique qui gère la composante mémoire d'entreprise.

La thèse est structurée en six (6) chapitres principaux. Elle débute par une introduction générale dans laquelle, nous avons fourni une description du contexte, de la problématique et des objectifs et elle se termine par une conclusion générale dans laquelle, nous avons présenté le bilan atteint et les objectifs à atteindre.

L'état de l'art inventorie les concepts de base manipulés et présente un panorama sur les méthodes de la gestion des connaissances et du web sémantique dédiées aux entreprises dans le but d'introduire la notion de la mémoire d'entreprise dans la définition des systèmes de capitalisation de connaissances d'entreprises. Il comprend quatre chapitres.

Le premier chapitre, intitulé *La gestion des connaissances pour l'entreprise*, permet de décrire l'état de l'art de la gestion des connaissances et d'introduire les différentes propositions de modélisation sur la base de modèles sociotechnique selon l'approche organisationnelle fondée sur le concept *knowledge creating-company*, l'approche biologique fondée sur le concept de l'arbre de connaissance, et l'approche managériale. Il introduit aussi la notion des connaissances industrielles induisant des savoirs et des savoirs-faire dans les entreprises industrielles.

Le deuxième chapitre, intitulé *Mémoire d'Entreprise et Capitalisation des connaissances*, présente des typologies relatives à la mémoire d'entreprise, ainsi que quelques méthodes de capitalisation. La mémoire de projet qu'est un type particulier est présentée avec quelques caractéristiques.

Le troisième chapitre, intitulé *Web sémantique pour l'entreprise et annotation* est divisé en deux parties. La première est consacrée à présenter le web sémantique d'entreprise en mettant l'accent sur l'intérêt de l'usage des ontologies dans la gestion des connaissances. Son objectif est de montrer l'apport des ontologies dans les processus de modélisation et de restitution des connaissances. Cet apport, est ensuite mis en évidence dans la deuxième partie consacrée à l'annotation classique puis

sémantique. Ce qui permet l'utilisation des ontologies pour l'annotation, l'indexation et la manipulation des ressources de connaissances formalisées.

Deux chapitres viennent, ensuite, dévoiler la contribution à une gestion d'une mémoire d'entreprise via l'annotation basée sur l'exploitation d'une ontologie de domaine. Ils sont présentés en détail comme suit :

Le quatrième chapitre, explique l'approche de modélisation basée sur le modèle de la Marguerite en insistant sur ses concepts de base, à savoir les pétales de connaissances, d'interaction et d'annotation.

Le cinquième chapitre, présente en détail l'Approche d'annotation sémantique suivi ainsi que l'architecture et le fonctionnement du système CMOCA proposé et sa validité via une étude de cas consacrée à l'expérimentation et à la réalisation d'un outil prototype qui implémente l'ontologie du domaine et l'outil de consultation de la mémoire d'entreprise.

La conclusion générale rappelle le bilan de nos travaux de recherche et des résultats obtenus. Le bilan est axé sur les propositions de la thèse à partir de la problématique vécue par l'entreprise. Enfin, les perspectives et les travaux futurs sont également énoncés.

# ***CHAPITRE 1***

## ***Gestion des connaissances pour l'entreprise***

# Gestion des connaissances pour l'entreprise

*« Définir le Knowledge Management ou sa traduction littérale, la "gestion des connaissances", est un art difficile car il fait appel à deux notions abstraites : la "gestion" et les "connaissances" qui ne peuvent donner naissance qu'à un concept lui-même abstrait dont l'existence même pourrait paraître étonnante. Gérer n'est pas produire et correspond à de tâches d'organisation et la "connaissance" n'est pas l'information mais se rapproche plus du savoir-faire au sens classique »  
[Tisseyre, 99].*

## 1. Introduction

Dans ce chapitre, nous allons présenter la problématique se focalisant sur la notion du "knowledge management". Cela pousse à réfléchir sur le nombre croissant des connaissances que les entreprises doivent capitaliser par leurs employés dans un contexte de concurrence serré et exacerbée.

En effet, les connaissances dans l'entreprise constituent un capital "immatériel" qu'il convient de gérer au même titre que le capital financier. Cette capitalisation est sensible aux problèmes liés notamment aux employés, tels que les départs massifs en retraite, les licenciements, les mutations etc., de même qu'au contexte général d'évolution. Une grande partie de l'information « capital de l'entreprise », n'est pas intégrée dans la partie informatisée de son système d'information. Elle se trouve sur des documents papiers ou dans les cerveaux des experts; dans ce dernier cas, elle n'est donc pas exploitable par les autres.

La mise en place d'un programme de gestion des connaissances demande du temps. Il faut recueillir, et de façon permanente, la matière première. Les connaissances tacites doivent être formalisées pour pouvoir jouer leur rôle de données pour les processus (informatisés ou à informatiser) de l'entreprise.

## 2. La connaissance, un enjeu pour l'entreprise

L'ensemble des connaissances d'une entreprise forme un capital intellectuel très volatile. Pour Jean-Louis Ermine [Ermine, 200], la gestion des connaissances « est revenue d'outre-atlantique, sous le vocable de Knowledge Management (KM), avec une vigueur et une force qui en font un des mots-clés des entreprises actuellement ». Il s'agit d'un mot pour décrire des approches, démarches, méthodes, outils, etc.

Le KM englobe la gestion du savoir et savoir-faire également le savoir-faire développé par le personnel, de manière à créer un système interactif de formation continue qui débouche sur une meilleure qualité des produits et services, et de ce fait il accroît la compétitivité de l'entreprise. Le KM permet de sauvegarder ce savoir et/ou savoir-faire. En favorisant la diffusion des informations, il favorise la transversalité aux dépens de la hiérarchisation. Si la gestion est jugée bonne, tout salarié « quittant un jour son entreprise » suite à une réduction des effectifs, départ en retraite, etc. Pour l'entreprise, la perte est minorée. En archivant sa propre mémoire, l'entreprise évite aussi la répétition d'erreurs du passé; le partage de la connaissance permet à chaque acteur d'améliorer ses performances et donc celles de l'entreprise. Ce concept du KM n'est pas limité à un secteur de l'entreprise ni même à un type d'entreprise.

## 2.1 Contexte et défis

Sous la pression d'une concurrence internationale de plus en plus vive, l'entreprise cherche à améliorer sans relâche le triptyque Qualité, Coûts Délais, ce qui l'amène à optimiser son organisation et à modifier en profondeur la culture de ses différents acteurs. Si elle veut se développer, l'entreprise doit tenir compte d'au moins trois pôles majeurs dont les intérêts divergent (Figure 1.1).

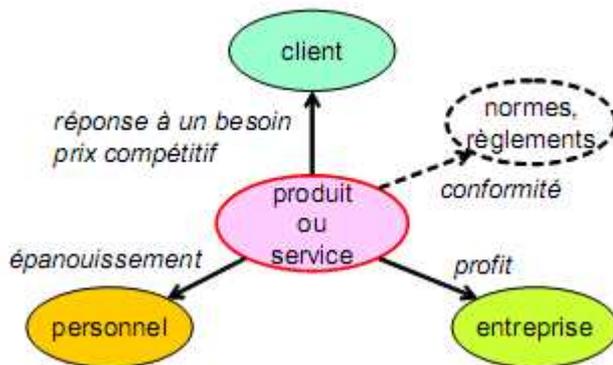


Figure 1.1 : Pôles majeurs dans l'entreprise

En premier lieu, l'entreprise propose un produit et/ou un service à des clients réels et potentiels; elle se doit de leur faire une offre correspondant à un besoin, ceci à un prix compétitif. Dans ce contexte, on voit par exemple les constructeurs automobiles proposer à leurs clients des gammes de véhicules renouvelées de plus en plus rapidement, possédant des fonctions nouvelles, associées à des offres de service très riches (financement, entretien, assurance...).

En second lieu, pour prospérer l'entreprise doit dégager un profit, faute de quoi elle disparaît. Bien qu'il s'agisse d'une évidence, il ne faut pas oublier la véritable mutation culturelle (passage à une logique concurrentielle) que doivent opérer certaines entreprises. Le troisième pôle enfin est celui du personnel, les collaboratrices et collaborateurs de l'entreprise, qui doivent s'épanouir dans leur travail, progresser, se sentir véritablement acteurs des progrès de l'entreprise ; la mise en pratique de tels principes se heurte encore aujourd'hui à de nombreux obstacles.

Ces exigences, auxquelles il convient d'ajouter le respect de normes précises en termes de qualité, sécurité et de protection de l'environnement, obligent les entreprises à prendre en considération de nombreux paramètres internes et externes. La maîtrise de ces paramètres demande l'intervention de disciplines nombreuses et variées issues des sciences économiques, des sciences humaines et sociales, des sciences de l'ingénieur. Pour améliorer son efficacité, l'entreprise est souvent amenée à remplacer une organisation vue principalement sous l'angle de structures « métiers » verticales au profit de processus transversaux orientés vers le client [Midler, 2002]. Il devient donc nécessaire pour l'entreprise d'assurer la coordination de tous ses moyens (méthodes, outils, modèles) et activités hétéroclites pour les mettre en synergie, ce qui constitue un véritable défi compte tenu de la complexité croissante des systèmes industriels.

## **2.2 Création des connaissances dans l'entreprise**

Nonaka et Takeuchi [Nonaka et Takeuchi, 1995] ont formalisé la création des connaissances dans l'entreprise en identifiant quatre modes de création et de transfert de connaissances. Ils partent de l'hypothèse que la connaissance est créée à partir des différentes interactions possibles entre connaissances tacites et connaissances explicites (cf. figure 1.2). Pour ces auteurs, la création des connaissances dans une entreprise intervient à trois niveaux qui sont : le niveau individuel, le niveau du groupe et le niveau de l'entreprise.

Il s'agit au niveau individuel de l'autonomie des individus et au niveau du groupe, il s'agit de l'interaction et du dialogue. Le niveau de l'entreprise, concerne la compétition pour l'accès aux ressources. De fait, les connaissances organisationnelles se développent particulièrement dans le temps au cours de deux dimensions : la dimension épistémologique (tacite / explicite) et la dimension ontologique (individus, groupes, organisations). Le processus de création de la connaissance se joue alors dans les différentes circulations entre l'individu et le collectif, les connaissances tacites et les connaissances explicites.

- Socialisation (Tacite vers Tacite) : la conversion de connaissances tacites en connaissances tacites se fait par interaction entre individus, ou socialisation. Cette dernière représente le processus de transmission de connaissances tacites. Il s'agit donc de transmettre des modèles mentaux ou des compétences techniques. Cette transmission peut très bien se faire sans échanges verbaux. En effet, la transmission d'un tour de main s'effectue généralement par l'observation, l'imitation et surtout la pratique. Comme le soulignent les auteurs, l'expérience est la clef pour acquérir des connaissances tacites.

- Externalisation ou articulation (Tacite vers Explicite) : l'externalisation est le processus de conversion du tacite à l'explicite. Elle consiste à rendre transmissibles et exploitables les savoirs tacites créés dans l'organisation. L'externalisation permet de valoriser les compétences acquises en multipliant leurs usages. Mais comme elle repose sur une codification des savoirs, elle favorise leur diffusion et imitation et réduit leur valeur. C'est le paradoxe de la valeur : en exploitant le savoir, ce qui permet le passage de connaissances tacites en connaissances explicites, sous la forme de concepts, modèles ou hypothèses. La modélisation d'un concept est très souvent déclenchée par le dialogue et l'échange avec d'autres individus.

- Internalisation (Explicite vers Tacite) : l'internalisation est le processus de conversion de connaissances explicites en connaissances tacites. Typiquement, cette conversion est un processus d'apprentissage avec des supports, documents, manuels. L'internalisation des connaissances explicites est un processus d'appropriation par expérimentation. Les compétences explicites sont progressivement traduites, par essai/erreur et interaction, en compétences tacites qui permettent l'application des connaissances explicites.

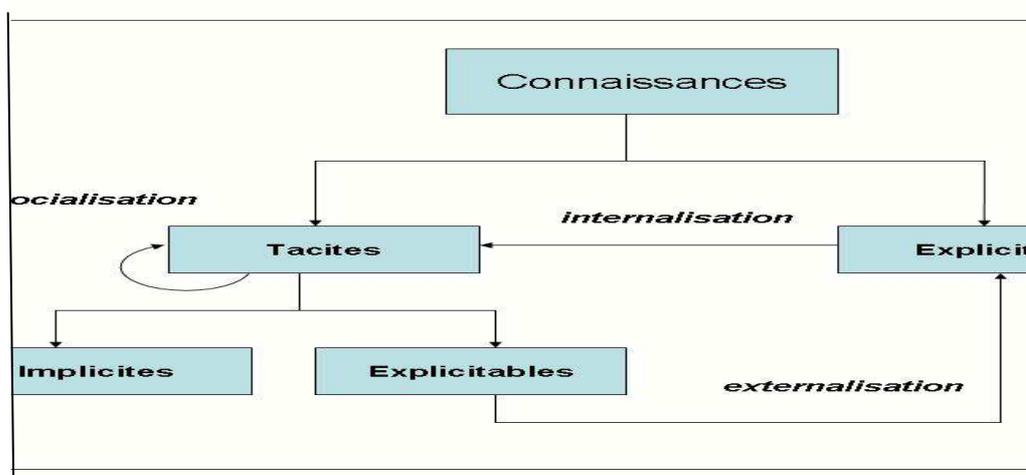


Figure 1.2. Modèles de transformation des connaissances

- Combinaison (Explicite vers Explicite) : la combinaison est un processus de création de connaissances explicites à partir de la restructuration d'un ensemble de connaissances explicites acquises par différents canaux de communication. La théorie de la création de la connaissance développée par Nonaka et Takeuchi considère que la première fonction de l'entreprise est de créer un avantage concurrentiel basé sur le savoir collectif et que le rôle des managers est d'orienter les activités de création de la connaissance. Le modèle de création des connaissances repose donc sur la distinction entre savoir tacite et savoir explicite. Le savoir tacite est enraciné dans l'action, dans les habitudes, dans un contexte spécifique. Cette transmission peut très bien se faire sans échanges verbaux. En effet, la transmission d'un tour de main s'effectue généralement par l'observation, l'imitation et surtout la pratique. Comme le soulignent les auteurs, l'expérience est la clef pour acquérir des connaissances tacites.

## **2.3 Typologie des Connaissances**

### **- Connaissances explicites – Connaissances implicites**

La plupart des auteurs présentent les connaissances selon une approche dichotomique : les connaissances tacites et les connaissances explicites. Cette classification a été popularisée par Nonaka et Takeuchi dans leurs théories de création de connaissances: la spirale de connaissances et par [Polanyi, 1996]. Ils distinguent deux types de connaissances :

- Les connaissances explicites : « les connaissances explicites se réfèrent à la connaissance qui peut être exprimée sous forme de mots, de dessins, d'autres moyens « articulés » notamment les métaphores » [Polanyi, 1996].
- Les connaissances tacites : « les connaissances tacites sont les connaissances qui sont difficilement exprimables quelle que soit la forme du langage » [Polanyi, 1996].

### **Autres typologies**

Il existe bien d'autres classifications des connaissances. Nous en énumérerons brièvement certaines, en remarquant que la distinction tacite/explicite demeure la plus citée. Karl Popper distingue trois mondes de connaissances [Karl, 1972] :

- Premier Monde de la connaissance : il s'agit de l'ensemble des structures encodées dans les systèmes physiques cela permet à ces objets de s'adapter à leur environnement.

- Second monde de la connaissance : croyances validées dans les esprits humains, concernant le monde, le beau, le vrai.
- Troisième monde de la connaissance : formulations linguistiques validées concernant le monde, le beau, le vrai.

Michel Saloff-Coste distingue trois niveaux dans l'intégration d'une connaissance : savoir, savoir-faire, savoir-être [Saloff, 2000].

- Le savoir est une conceptualisation fondée sur l'abstraction et qui prétend à des énoncés universels. On la rencontre en philosophie et en science par exemple. Le savoir est clairement formulé et contextualisé dans l'ensemble des connaissances implicites nécessaires à sa bonne compréhension.
- Le savoir-faire se développe parallèlement à l'expérience d'un collaborateur. Il résulte de la mise des connaissances acquises par ce dernier dans le cadre d'une activité particulière.
- Le savoir-être : au-delà du savoir-faire, quand on analyse finement les raisons de l'excellence dans une matière on découvre un véritable « savoir-être ». Ce savoir-être est souvent complètement implicite et donc non-dit. Il fait référence à un certain nombre de valeurs morales, éthiques ou esthétiques que la personne véhicule sans les avoir analysées. Ce sont des représentations du monde, des « paradigmes » qui sont transmis de manière informelle, le plus souvent à travers les relations sociales et culturelles, ou qui sont élaborées plus ou moins inconsciemment par l'individu ou l'entreprise dans un désir de se dépasser.

Dans le contexte des études organisationnelles contemporaines, Blackler [Blackler, 1995], fournit cinq types de connaissances:

- « Embrained Knowledge » : Ce sont des connaissances abstraites fortement dépendantes des compétences conceptuelles et des habilités cognitives d'un individu.
- « Embodied Knowledge » : Il s'agit des connaissances personnifiées. Cette catégorie de connaissances est orientée vers l'action. Elle dépend de la présence physique d'un individu, de sa sensibilité et de sa personnalité. Ce type de connaissances est acquis par l'apprentissage dans des contextes spécifiques.
- « Encultured Knowledge » : Elles se réfèrent au processus par lequel les membres d'un groupe atteignent le niveau d'une compréhension commune et partagée. C'est un processus fortement dépendant des valeurs culturelles et sociales des acteurs qui y participent. Il peut être supporté par des formulations explicites d'idées, par des métaphores, par des discussions de groupes etc.

- « Embedded Knowledge » : Il s'agit des connaissances imbriquées. Elles résident dans des routines organisationnelles systématiques. Elles sont analysées par l'exploration des relations existantes, par exemple, entre les technologies, les rôles des acteurs, les processus, les procédures formelles et les routines organisationnelles émergentes. Au final, ce sont des connaissances émergeant des interactions de l'intégralité des éléments organisationnels.
- « Encoded Knowledge » : Il s'agit des connaissances codées. Ce sont des informations acheminées sous forme de symboles, elles prennent forme dans des documents, des manuels, des livres, des rapports etc. et désormais sous forme électronique.

Quant aux auteurs Nahapiet et Ghoshal [Nahapiet et al., 1998], ils distinguent deux catégories de connaissances : les connaissances pratiques, basées sur l'expérience et les connaissances théoriques dérivées à partir de réflexions et d'abstractions faites sur ces expériences.

Ryle [Ryle, 1949] différencie les connaissances procédurales des connaissances déclaratives. Les connaissances procédurales concernent les compétences, les routines organisationnelles et le savoir-faire. Les connaissances déclaratives concernent le développement de faits et de propositions.

Ces différentes manières de classer et typer les connaissances, ne sont évidemment pas les seules, mais la distinction explicite/tacite demeure toujours perceptible.

### **3. Gestion des connaissances**

#### **3.1 Définitions de la gestion des connaissances**

Proposer une terminologie commune au KM est une tâche vraiment ambitieuse. En effet, ce concept émergent n'a pas encore connu de définition stable. Le KM répond à plusieurs définitions plus au moins étendues selon le domaine des intervenants qui l'utilisent. Certaines définitions peuvent parfois être réductrices et refléter les lacunes et erreurs du passé et du présent. Nous avons sélectionné, dans cette section des définitions de KM proposées dans la littérature du KM. Notre objectif est de relever les principales questions qui résultent de tentatives typiques de définitions de KM et qui conduisent vers la définition d'une mémoire d'entreprise/.

« La gestion des connaissances est l'objectif de formaliser les connaissances tacites afin de les rendre mobilisables et opérationnelles au niveau de l'organisation entière » [Ermine, 2003].

Dans ce qui suit , nous allons présenter un ensemble de définitions du KM issues de la littérature.

- Swan [Swan et al., 1999] définit le KM comme étant un « ensemble de démarches, méthodes et outils de création, de collecte, de formalisation, de capitalisation, de partage, et d'utilisation de connaissances. Sa finalité est de favoriser les processus collectifs d'apprentissage et d'innovation ».
- Wiig [Wiig, 1999] définit le KM comme suit : « le Knowledge Management concerne la compréhension, le management systématique, la construction explicite et délibérée de connaissances et son application. Il s'agit d'un management efficace des processus centrés connaissances ». Il souligne que le KM doit être perçu selon quatre perspectives:
  - la perspective technologique qui concerne la capture, la manipulation et la localisation des connaissances;
  - la perspective organisationnelle (Enterprise Effectiveness Focus) qui concerne l'utilisation optimale des connaissances pour améliorer l'efficacité des processus opérationnels de l'organisation;
  - la perspective capitale intellectuelle (Intellectual Asset Focus) qui met en œuvre les méthodologies d'exploitation du capital intellectuel pour optimiser la valeur économique d'une organisation ;
  - la perspective Humaine (People Focus) s'attache à créer un environnement de partage de connaissances pour les employés de l'organisation.
- Dave Snowden [Snowden, 2000] définit le Knowledge Management comme « l'identification, l'optimisation et le management actif du capital intellectuel, qu'il soit sous forme de connaissances explicites contenues dans des artefacts, ou sous forme de connaissances tacites possédées par des individus ou des communautés. L'optimisation du management des connaissances explicites est réalisée à travers la consolidation et l'amélioration de l'accès aux artefacts. L'optimisation du management des connaissances tacites est réalisée à travers la création de communautés partageant et cultivant ces connaissances. Le Knowledge Management, concerne la création de processus de management et d'infrastructures pour réunir les artefacts et les communautés dans une écologie commune qui va maintenir la création, l'utilisation et la rétention des connaissances ».
- D'après [Ramon, 2001] « Le terme Knowledge Management est le processus à travers lequel une entreprise utilise son intelligence collective pour accomplir ses objectifs stratégiques ».

Cette définition conduit à souligner quatre dimensions du KM :

- Le KM est un processus qu'il comprend des phases et des composants,
- Le KM implique que la connaissance collective doit amener à l'action,
- Le KM implique l'intelligence collective qui implique l'importance du rôle du partage de connaissances,
- Le KM n'est pas une fin en soi, il est fortement lié aux objectifs stratégiques de l'organisation qui le pratique.

L'ensemble des définitions citées ci-dessus se focalisent principalement sur les concepts suivants d'une manière explicite ou implicite : processus, management, méthodes etc., connaissance.

Un processus de gestion des connaissances se définit tout d'abord en fonction de l'approche de gestion des connaissances adoptée. Il peut également se définir à partir des fonctionnalités attendues du système de gestion des connaissances à mettre œuvre. La gestion des connaissances s'effectue alors grâce aux interactions entre les différentes facettes décrites ci-dessous [Grundstein, 2002] :

- repérage des connaissances cruciales,
- préservation des connaissances,
- valorisation des connaissances,
- actualisation des connaissances.

***Notre contribution : le KM est un moteur des changements contextuels, organisationnels et technologiques qui mettent en évidence tout le processus du cycle de vie de la connaissance au sein de l'entreprise, incluant la création, la collecte, l'organisation, la diffusion, l'utilisation et l'exploitation des connaissances. Le KM nécessite le passage des connaissances personnelles vers des connaissances collectives partagées largement dans une organisation.***

### **3.2. Aspects pour la gestion des connaissances**

Il existe plusieurs aspects pour la gestion des connaissances, nous pouvons citer les suivants :

#### **3.2.1. Aspect social**

Dans l'aspect social de la gestion des connaissances, la connaissance est perçue comme un système, d'une part organisé à travers un système de connaissances, et d'autre part structuré à travers la gestion des connaissances. Pour explorer cet aspect nous nous attacherons à retrouver la signification de la connaissance dans le travail et à justifier l'importance de cette connaissance comme un facteur clé dans l'organisation du travail coopératif. Cet aspect comprend :

#### **3.2.1.1. Approche organisationnelle de la gestion des connaissances**

D'un point de vue organisationnel l'aspect humain : l'individu, le groupe, l'entreprise est à la base de la création des connaissances nouvelles et d'apprentissage organisationnel pour l'entreprise. Autrement dit, la connaissance est à rapprocher d'un système social qui existe dans un domaine humain au travers du dialogue et du langage des acteurs (individu, groupe, entreprise).

#### **3.2.1.2. Approche biologique de la gestion des connaissances**

Dans l'approche organisationnelle, les mécanismes de création de connaissances nouvelles et d'apprentissage organisationnel sont caractérisés. L'objectif de l'approche biologique sera de retrouver ces mécanismes, mais cette fois dans le cadre d'un système naturel qui existe dans un domaine biologique (les relations sont au niveau cellulaire) au travers du "dialogue" et du "langage mais non pas au niveau humain mais au niveau cellulaire, organisé dans le système nerveux. Puis de faire l'hypothèse que ces mécanismes au niveau cellulaire peuvent être généralisés au niveau humain à travers un système social. La validation de cette hypothèse reste restreinte à ce jour au niveau des analogies.

#### **3.2.1.3. Approche managériale de la gestion des connaissances**

Dans cette perspective, la gestion des connaissances est perçue comme un problème de gestion, que pour comprendre et résoudre, il faut (1) diagnostiquer ; (2) évaluer ; et (3) recommander des actions à entreprendre.

#### **3.2.2. Aspect technique de la gestion des connaissances**

Pour l'approche technique du KM, les mécanismes de création de connaissances nouvelles et d'apprentissage organisationnel sont matérialisés à travers des méthodes et des outils pour la collecte ou l'acquisition, l'extraction et la diffusion ou le partage des

connaissances. Cet aspect permet également, l'utilisation des nouvelles technologies de l'information et de la communication du Knowledge Management (les NTIC du KM).

Nous pensons que l'approche organisationnelle et managériale ainsi que l'aspect technique de la gestion des connaissances nous permettent de cerner le contexte environnemental de la connaissance utilisée pour la gestion d'une mémoire d'entreprise.

### **3.3. Quelques approches de la gestion des connaissances**

Nous avons trouvé en littérature deux grands types d'approches :

1. l'approche orientée information,
  2. l'approche orientée connaissances.
- L'approche orientée information pour la gestion des connaissances utilise les outils informatiques facilitant le travail coopératif et la communication entre les différents collaborateurs de l'entreprise (outils de groupware). Elle se concentre sur l'amélioration de la gestion et de l'échange d'information sans se soucier des frontières organisationnelles ou professionnelles.
  - L'approche orientée connaissances se base sur la gestion d'une mémoire d'entreprise. L'élaboration d'une telle mémoire correspond à une phase de capitalisation de l'ensemble (ou d'un sous-ensemble) des connaissances de l'organisation. Cette approche, très liée aux recherches effectuées en Ingénierie des Connaissances, se base sur une étape de capitalisation consistant à recenser puis à modéliser des connaissances. Cette approche permet de traiter des problèmes liés à l'optimisation du fonctionnement de l'entreprise. Cela peut concerner la gestion des compétences, des ressources (matérielles et humaines), des processus, des relations avec les clients ou prestataires, etc.

Nous pouvons bénéficier des deux approches, du moment où elles ne sont pas antagonistes et peuvent être au contraire assez complémentaires. Pour la gestion d'une mémoire d'entreprise, le travail coopératif est utile notamment s'il inclut une gestion des retours d'expériences.

### **3.4. Systèmes pour la gestion des connaissances**

Nous pouvons distinguer plusieurs systèmes de gestion des connaissances, les plus importants sont les suivants :

#### **3.4.1. Les Systèmes à Base de Connaissances (SBC)**

Les Systèmes à Base de Connaissances permettent le stockage et la consultation de

connaissances, le raisonnement automatique sur les connaissances stockées (sans préjugé sur le type de raisonnement à mener), la modification des connaissances stockées (ajout ou suppression de connaissances), ainsi que le partage de connaissances entre systèmes. Dans cette approche qui s'insère dans le courant de la pensée systémique, la connaissance se perçoit comme un élément qui contient de l'information avec trois points de vue, la structure, la fonction et l'évolution.

### **3.4.2. La Mémoire d'Entreprise (ME)**

L'approche mémoire d'entreprise vise à localiser et à rendre visible les savoirs et savoir-faire de l'entreprise en vue de leur exploitation par les employés. Ce terme récent résulte « d'une prise de conscience par un nombre croissant d'entreprises que les connaissances détenues par les employés constituent un capital immatériel qu'il convient de gérer au même titre que le capital financier » [Dieng et al., 2005].

Nous proposons d'utiliser le deuxième type tout en considérant les trois points de vue de la connaissance et qui sont la structure, la fonction et l'évolution.

## **4. Conclusion**

La connaissance constitue un levier stratégique pour l'entreprise, car c'est à la fois un capital économique, une ressource stratégique, un facteur de stabilité et un avantage concurrentiel décisif. Un certain nombre de fonctionnalités du KM commencent à être intégrées dans des environnements tels que les ERP (Entreprise Resource Planning), les EPM (Entreprise Performance Management) ou les outils de SCM (Supply Chain Management).

Si l'on applique l'outil QQQQCP au management des connaissances dans cette dernière optique, on peut répondre aux questions comme suit :

- Qui ? Le KM concerne potentiellement tous les acteurs de l'entreprise;
- Quoi ? On s'intéresse exclusivement à la connaissance utile au développement de l'entreprise, en lien avec ses objectifs stratégiques;
- Où ? Là où la connaissance est requise, ce qui nécessite d'établir une cartographie des compétences de l'entreprise et des connaissances sur lesquelles elles s'appuient;
- Quand ? Dans la continuité, ce type de démarche s'apparentant davantage aux approches qualité (donc dans la durée) qu'au re-engineering (modification brutale et profonde) ;

- Comment ? Il s'agit de mettre en œuvre des démarches adaptées au contexte;
- Pourquoi ? Parce qu'il faut préserver le capital intellectuel de l'entreprise.

Nous avons privilégié, dans le cadre de nos travaux, l'orientation basée sur la gestion des connaissances pour résoudre les problèmes liées aux besoins d'avoir une mémoire d'entreprise.

Dans le prochain chapitre, nous allons aborder la définition de la notion de mémoire d'entreprise, quelques typologies qui nous semblent pertinentes et les besoins en d'en bâtir un tel composant au sein des entreprises actuelles. Ensuite, nous allons expliciter le type de mémoire de projet, en présentant quelques méthodes de capitalisation qui y sont rattachées.

## **CHAPITRE 2**

### ***Mémoire d'Entreprise et capitalisation des connaissances***

## CHAPITRE 2

### **Mémoire d'Entreprise et capitalisation des connaissances**

*« le terme "mémoire d'entreprise" désigne l'ensemble des savoir et savoir-faire en action mobilisés par les employés d'une entreprise pour lui permettre d'atteindre ses objectifs (produire des biens ou des services) »*  
Dieng

#### **1. Introduction**

Le partage et la capitalisation transversale de la connaissance se concrétisent rapidement par des résultats tangibles. Notamment, pour ce qui concerne la conception de produit, l'entreprise voit augmenter sa capacité d'innovation, car elle sait créer le bon produit au bon moment. Elle anticipe mieux les changements, améliore sa réactivité, se montre capable de modifier son organisation et de faire évoluer ses produits. Les informations recueillies par une veille efficace sur les concurrents sont automatiquement intégrées. La formalisation et la numérisation de la capitale connaissance débouchent pratiquement sur une gestion en temps réel.

Forte de tous ces progrès l'entreprise augmente ainsi sa compétitivité sur son marché. La sensibilisation des salariés est primordiale, pour qu'ils mettent à disposition leur savoir comme pour aller chercher ce que d'autres ont fourni. Certaines entreprises choisissent même de motiver les salariés en les rémunérant en fonction de leur participation à l'enrichissement de la « mémoire de l'entreprise ».

Selon (Dieng et al., 1998), le terme "mémoire d'entreprise" désigne l'ensemble des savoir et savoir-faire en action, mobilisés par les employés d'une entreprise pour lui permettre d'atteindre ses objectifs (produire des biens ou des services). Le terme plus générique "mémoire d'organisation" indique que cette notion de mémoire peut s'appliquer à n'importe quel type d'organisation, qu'il s'agisse d'une entreprise, d'un service ou département au sein de l'entreprise, ou bien, à une échelle plus petite, d'un projet. Dans ce dernier cas, on parlera de "mémoire de projet". La nature de la mémoire d'entreprise et les efforts nécessaires pour sa construction dépendent de la taille de la société (grands groupes vs PME et PMI).

Une mémoire d'entreprise n'est pas nécessairement un système à base de connaissances. Les techniques adoptées pour établir une mémoire d'entreprise dépendent des sources disponibles : spécialistes humains, documents papiers ou électroniques (rapports, documentation technique, courrier électronique, bases de données existantes, bibliothèques de cas, dictionnaires, plans,...). Ces techniques dépendent également de la nature de la mémoire d'entreprise nécessaire pour les utilisateurs visés : elle peut se composer de documents écrits rendant explicites des connaissances implicites. La mémoire peut également être matérialisée par un système documentaire, une base de données, une base de connaissances, un système à base de cas, un système basé sur le Web, un collecticiel ou un système multi-agents...

## **2. Définitions**

Une mémoire d'entreprise a pour objectif de recueillir, afin de les préserver et de les transmettre, les savoirs acquis par une organisation, savoirs qui comprennent à la fois les connaissances détenues par les acteurs et les documents produits ou utilisés par l'entreprise.

La mémoire d'entreprise est donc un recueil des connaissances, « un ensemble structuré de connaissances attaché à l'expérience de l'entreprise dans un domaine donné » [Simon, 1996].

Les connaissances d'une société, et sa capacité à les gérer, sont considérées comme des atouts majeurs. L'objectif est donc bien d'améliorer la compétitivité de cette entreprise, en évitant par exemple la répétition d'erreurs déjà commises : « Une mémoire d'entreprise intègre des données du produit avec de la connaissance (à traduire : product legacy) afin d'amener les expériences accumulées auparavant à peser sur les nouvelles tâches. De cette façon une répétition des erreurs peut être évitée, le savoir-faire est amélioré systématiquement » [Kühn et Abecker, 1997], ou en permettant à l'entreprise de s'adapter dans de bonnes conditions au contexte: « Pour améliorer la puissance compétitive en continuant l'adaptation de l'organisation à l'environnement externe (le marché, le climat, les préférences clientes sociaux et politiques » [Van Heijst et al., 1996].

[Euzenat, 1996] Considère lui aussi la mémoire d'entreprise comme « une base de connaissances et du savoir-faire d'un ensemble d'individus travaillant dans une entreprise spécialisée ».

D'après [Pomian, 1996] la mémoire d'une entreprise inclut non seulement « une mémoire technique » obtenue par capitalisation du savoir-faire de ses employés mais également « une mémoire organisationnelle » (ou « mémoire managériale ») liée aux

structures organisationnelles passées et présentes de l'entreprise (ressources humaines, management, etc.) et des « mémoires de projet » pour capitaliser les leçons et l'expérience de certains projets.

De la même manière, [Van Heijst et al., 1996] définit ainsi la mémoire d'entreprise: « Une mémoire d'entreprise est la représentation de connaissances explicites, désincarnées et persistante dans une organisation. (...). N'importe quelle partie de la connaissance ou de l'information qui contribuent à la performance de l'organisation peut (et peut être doit) être stockée dans la mémoire d'entreprise. Ceci inclut les connaissances à propos des produits, des processus de production, clients, stratégies du marché, résultats financiers, plans stratégiques et les objectifs etc. ».

Le processus de capitalisation des connaissances permet de réutiliser, de façon pertinente, les connaissances d'un domaine donné, précédemment stockées et modélisées, afin d'accomplir de nouvelles tâches [Simon, 1996]. Le but est de "localiser et rendre visible les connaissances de l'entreprise, être capable de les conserver, y accéder et les actualiser, savoir comment les diffuser et mieux les utiliser, les mettre en synergie et les valoriser" [Grundstein, 1995].

Par conséquent, pour toute opération de capitalisation des connaissances, il est important d'identifier les connaissances stratégiques à capitaliser [Grundstein et Barthès, 1996]. Une mémoire d'entreprise devrait fournir "la bonne connaissance ou information à la bonne personne au bon moment et au bon niveau pour que cette personne puisse prendre la bonne décision".

Comme l'a noté [Nonaka, 1991], la chaîne des connaissances repose sur plusieurs étapes : faire le bilan des connaissances existantes, déterminer les connaissances requises, développer de nouvelles connaissances, assigner les nouvelles connaissances et celles existantes, les appliquer, les mettre à jour, retirer celles obsolètes. En étendant la définition proposée par [Van Heijst et al., 1996], [Dieng et al, 1998] ont considéré une mémoire d'entreprise comme la "représentation explicite et persistante des connaissances et des informations dans une organisation, afin de faciliter leur accès et leur réutilisation par les membres adéquats de l'organisation pour leur tâche". Nous proposons de considérer la gestion d'une mémoire d'entreprise comme reposant sur les étapes suivantes:

1. Détection des besoins en mémoire d'entreprise,
2. Construction de la mémoire d'entreprise
3. Diffusion de la mémoire d'entreprise,
4. Utilisation de la mémoire d'entreprise,

5. Evaluation de la mémoire d'entreprise,
6. Maintenance et évolution de la mémoire d'entreprise.

### **3. Typologies de mémoire d'Entreprise**

Comme pour les connaissances, il existe une multitude de typologies différentes de mémoire d'entreprise :

Ainsi [Tourtier, 1995] cité dans [Dieng et al., 1998] distingue quatre types de mémoire :

- la mémoire de la profession (métier), composée des références, documents, outils, méthodes employés dans une profession donnée;
- la mémoire de la société, liée à l'organisation, aux activités, produits, et acteurs comme les clients, sous-traitants, fournisseurs;
- la mémoire individuelle, contenant le statut, les compétences, savoir-faire et activités d'un salarié donné;
- la mémoire de projet, constituée de la définition d'un projet, les activités attenantes, l'historique et les résultats.

Pour [Pomian, 1996], la mémoire d'entreprise se décline en trois sous mémoires, la mémoire de projet, la mémoire organisationnelle et la mémoire technique :

- la mémoire de projet est directement liée à une mission particulière réalisée au sein de l'entreprise. Un groupe de personnes est formé pour un temps donné pour réaliser un projet bien défini. La mémoire de projet rassemble les connaissances, savoir-faire, compétences, documents qui ont été nécessaires à l'accomplissement de ce projet.
- la mémoire organisationnelle rassemble les connaissances pertinentes pour les activités de l'organisation à tous les niveaux. Elle peut ainsi inclure des informations sur les structures organisationnelles présentes et passées, sur les ressources humaines, etc. Par ailleurs, [Giboin, 1996] y distingue les connaissances sur les liens actions résultats et les conditions dans lesquelles une action donnée conduit à un résultat donné, les connaissances sur la technologie, sur l'environnement interne et l'environnement externe.
- la mémoire technique se rattache quant à elle à un métier. Elle s'intéresse à l'aspect opérationnel de l'entreprise, c'est-à-dire à l'expérience acquise liée au travail et permettant à l'entreprise de vivre. Elle est constituée des connaissances liées à un

métier et nécessaires à l'exécution des tâches des individus de l'entreprise en vue d'une activité particulière. Ces connaissances rassemblent les connaissances tacites inscrites dans le cerveau des hommes, leurs savoir-faire, leurs compétences, et les connaissances explicites, contenues dans les bases de données, manuels techniques, plans de fabrication, etc.

[Grundstein et Barthès, 1996] distinguent les connaissances critiques (i.e. utilisées quotidiennement à l'intérieur de l'entreprise, ses départements, ses services, ses filiales, par les employés pour réaliser leur travail quotidien) et les connaissances stratégiques définies par les managers ou les chefs d'entreprise.

[Dieng et al., 1998] proposent de distinguer la mémoire interne (correspondant aux connaissances et aux informations internes de l'entreprise) et la mémoire externe (correspondant aux connaissances et aux informations utiles pour l'entreprise mais provenant du monde extérieur).

Le fonctionnement de ces sous mémoires repose sur un principe commun un mécanisme d'acquisition, de rétention ou stockage et de restauration. Chaque sous-système utilise cependant des modes différents d'acquisition, de stockage ou de restauration selon ses spécificités et ses fonctions. C'est ainsi que l'on distingue généralement la mémoire de court terme qui stocke temporairement des informations et la mémoire de long terme qui stocke les informations de façon permanente et possède donc une capacité de stockage beaucoup plus importante que la mémoire de court terme.

#### **4. Besoins industriels en mémoire d'entreprise**

Les entreprises évoluent rapidement vers des services et des métiers qui s'ouvrent massivement à de nouveaux domaines. De ce fait, elles ne se limitent pas à des unités de production de biens ou de services conformes aux espérances des clients, mais c'est également une unité de production de connaissances (Grundstein, 1995). Les motivations de la capitalisation des connaissances peuvent être diverses:

- Eviter la perte de savoir-faire d'un spécialiste après sa retraite ou sa mutation;
- Exploiter l'expérience acquise des projets passés, et conserver les leçons du passé, afin d'éviter de reproduire certaines erreurs;
- Exploiter la cartographie des connaissances de la société à des fins stratégiques : un inventaire régulier du savoir-faire de l'entreprise devrait améliorer la capacité de l'entreprise à réagir et à s'adapter aux changements dans son environnement;

- Améliorer la circulation de l'information et la communication dans l'entreprise;
- Améliorer l'apprentissage des employés (nouveaux comme anciens) de l'entreprise;
- Intégrer les différents savoir-faire d'une organisation.

Dans le cadre de la conception de produits et de services, au moins trois défis majeurs se doivent d'être relevés par l'entreprise : i) la maîtrise de la connaissance, ii) l'innovation et iii) la pro-activité.

- La maîtrise de la connaissance possède au moins deux facettes : la veille technologique et le management des connaissances internes.
- La veille technologique consiste pour l'entreprise à se tenir informée de manière constante et organisée des activités de ses concurrents, connaître en temps réel les nouveaux brevets et les nouvelles normes pouvant impacter son action, être à l'affût des nouveaux composants, des articles de la presse spécialisée... et être en mesure d'exploiter ces informations après les avoir conceptualisées.
- Le management des connaissances internes est un processus de création, d'enrichissement, de capitalisation et de diffusion des connaissances impliquant tous les acteurs d'une entreprise, en tant que consommateurs et producteurs, qui débouche sur une meilleure qualité des produits et services, ainsi que sur une plus grande compétitivité de l'entreprise.
- L'innovation constitue un avantage concurrentiel majeur, à condition d'en maîtriser les risques, l'entreprise doit se doter d'une véritable ingénierie lui permettant de mettre en œuvre de manière systématique la recherche de solutions innovantes, ainsi que la sélection des meilleures d'entre-elles.
- La pro-activité s'applique à la maîtrise des connaissances et à l'innovation, il semble indispensable, pour préserver un avantage concurrentiel, non seulement de réagir aux événements, mais surtout de les anticiper [Prax, 2000]. Dans ce contexte, il devient nécessaire pour l'entreprise de se doter de nouveaux outils et modes d'organisation lui permettant de capter les informations opportunes (les rechercher, les filtrer), capitaliser les connaissances de l'entreprise, et les partager pour en générer de nouvelles connaissances.

## 5. Quelques méthodes pour la capitalisation

Dans le cadre de la capitalisation, plusieurs méthodes ont été définies. Quelques méthodes ont été empruntées à l'ingénierie des connaissances et adaptées pour aider à la définition de mémoires d'entreprise. Nous présentons dans ce chapitre un panorama de ces méthodes, afin de donner une vision globale sur les différentes sortes de méthodes de capitalisation de connaissances existantes.

### 5.1. La méthode REX

La méthode REX [Malvache et Prieur , 1993] a été définie au départ dans le but de capitaliser les expériences de conception de réacteurs nucléaires au sein du CEA. La méthode a été ensuite utilisée dans divers types d'applications tels que la spécification de systèmes de contrôle dans le domaine électrique, la conception de générateurs électriques, la conception aéronautique, etc.

Le principe de base de la méthode consiste à constituer des «éléments d'expériences», extraits d'une activité quelconque et à restituer ces éléments pour qu'un utilisateur puisse les valoriser. Les éléments d'expérience ainsi définis sont stockés dans une mémoire d'expérience appelée (CEMem) avant d'être restitués (Figure).

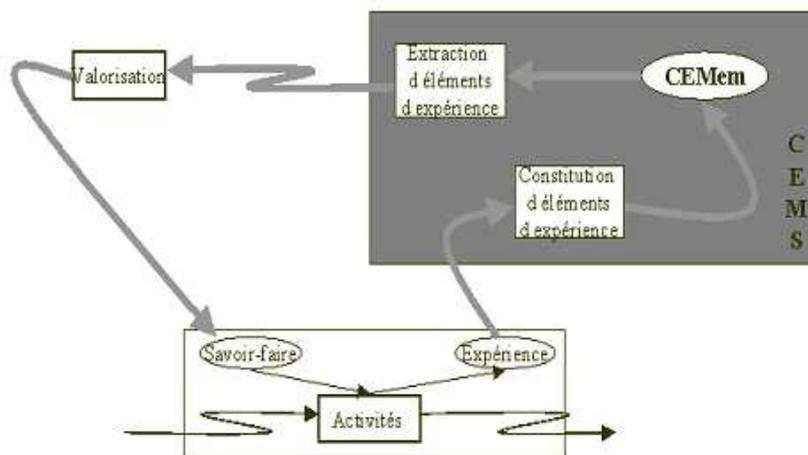


Figure.2.1. Principe de base de la méthode REX

Il s'agit d'un cycle de constitution d'éléments d'expérience à partir de l'expérience acquise pendant la réalisation des activités dans une organisation, ainsi que de l'extraction de ces éléments dans une optique de valorisation de l'expérience en un savoir-faire, utile dans la réalisation des activités.

## - Eléments d'expérience

Un élément d'expérience (Figure 2.1) est typiquement défini par: un entête, une description ou corps, une liste de références. Le corps est décomposé lui même en trois parties: une description neutre d'un fait, une opinion propre et des commentaires, des recommandations. Les éléments d'expérience sont construits principalement à l'issue des entretiens auprès d'experts et à partir des documents relatant une activité (i.e. documents de synthèse, bases de données).

Un questionnaire doit être élaboré. Il forme une base pour le cogniticien dans un entretien avec un expert. Il permet d'élucider les éléments d'expériences et leur description. Trois entretiens d'une demi-journée chacun sont recommandés à ce propos. Le premier entretien est mené d'une manière libre et vise à identifier les personnes concernées par un thème particulier et à collecter leur avis. Le texte recueilli à l'issue de cet entretien sert à identifier plusieurs éléments d'expérience correspondant aux différents faits cités.

Au cours du deuxième entretien, un ensemble provisoire d'éléments d'expérience est présenté à l'expert qui peut modifier leur contenu et les enrichir.

L'objectif du troisième entretien est de vérifier si toutes les modifications apportées aux éléments d'expérience ont été considérées. D'éventuelles corrections sont alors introduites.

Les éléments d'expérience peuvent être également extraits des documents techniques, des standards, des procédures, des codes de calcul, etc. Par exemple, chaque paragraphe dans un document peut être considéré par un élément d'expérience de la même façon qu'une partie d'un entretien. Les éléments d'expérience ainsi constitués sont ensuite organisés de façon à être facilement réutilisables. Cette organisation est spécifique à chaque domaine d'activité.

- Mémoire d'expérience : dans le but de considérer la diversité de vocabulaire et les points de vue utilisés dans une entreprise, un modèle descriptif et un réseau terminologique sont également définis pour constituer une mémoire d'expérience.

- Le modèle descriptif : il permet de représenter les différents points de vue identifiés dans une entreprise. Généralement, une douzaine de points de vue semble raisonnable, au delà de douze, le réseau défini par ces points de vue sera inexploitable. Par exemple, nous pouvons distinguer point de vue géographique, topologique, etc. dans une activité de conception.

Chaque point de vue est représenté par un réseau d'objets (définis sous forme de concepts), reliés entre eux, suivant un réseau sémantique. Un ensemble de catégories de liens est aussi défini comme: «ensemble/élément», «général/spécifique», «proximité», «Self évolution».

Le modèle descriptif peut ne pas être défini d'une manière exhaustive. Il sera enrichi au fur et à mesure.

## 5.2 La méthode MKSM

MKSM (Methodology for Knowledge System Management) [Ermine, 1996] est une méthode qui a été développée au CEA avec un objectif de gestion des connaissances. La méthode se base principalement sur un triangle sémiotique (Figure 2.2) où trois dimensions sont prises en compte: la syntaxe, la sémantique et la pragmatique. L'analyse des connaissances dans une organisation suivant ces trois dimensions consiste à considérer l'information (syntaxe), la signification (sémantique) et le contexte (pragmatique). L'analyse de ces dimensions est guidée par l'étude du traitement des données, des tâches et de l'activité du domaine.

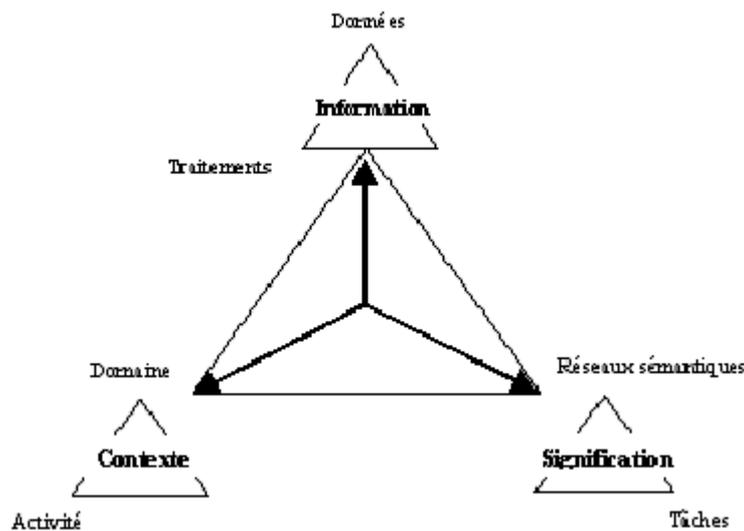


Figure 2.2. Triangle sémiotique de MKSM

Une modélisation du système des connaissances doit être aussi considérée. Cette modélisation met en évidence les flots de connaissances et d'informations, les acteurs (ou agents) producteurs et consommateurs de la connaissance. Chaque agent est défini par son rôle dans l'entreprise, les informations consommées, les informations produites, les connaissances consommées et les connaissances produites. Cette modélisation (Figure) peut être représentée sous forme d'un diagramme SADT dans lequel le système opérant représente des informations sur les agents, le système de décision définit l'environnement et

la capacité d'organisation et de structuration et le système d'information rassemble les documents et les bases de données de l'entreprise.

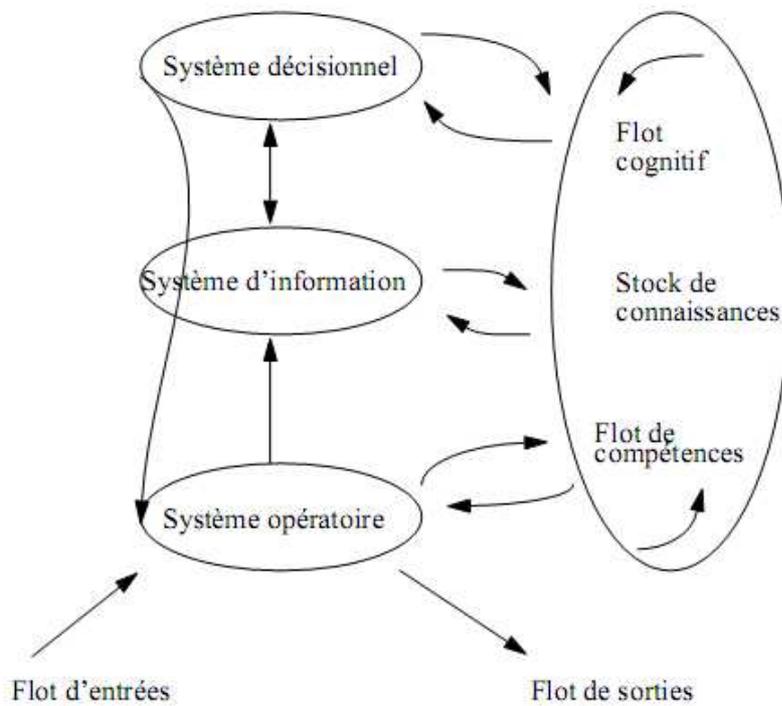


Figure 2.3. Modélisation du système de connaissances

Le patrimoine de connaissances comporte deux éléments principaux : le livre de connaissances et le système opérationnel de gestion de connaissances. Ce dernier peut être défini sous forme de systèmes de supervision, d'aide à la décision, d'information et de documentation, de formation, de veille technologique ou stratégique et de gestion de la qualité. Le livre de connaissances rassemble des modèles de connaissances en complément des documents de l'entreprise comme les fiches, les plans, les documents techniques, les images, les références, etc.

## 6. Cas spécial : Mémoire de projet

Dans ce qui suit, nous allons nous intéresser à un type particulier de mémoires qu'est la mémoire de projet et nous allons présenter ses principales caractéristiques.

### 6.1. Définition de la mémoire de projet

[Dieng et al., 1998] définissent une mémoire de projet comme une mémoire des

connaissances et des informations acquises et produites au cours de la réalisation des projets. Notre définition rejoint celles données dans la littérature comme: "Lessons and experiences from given projects" (Pomian, 1996) et "Project definition activities, history and results" (Tourtier, 1995).

Les travaux sur la mémoire de projet dans le domaine de la conception, se basent sur ceux de [Matta et al., 1999].

Une mémoire de projet doit fournir l'accès à des informations décrivant aussi bien les caractéristiques d'un projet que celles relatives à la résolution des problèmes rencontrés lors de la réalisation du projet.

## **6.2. Modèle de mémoire de projet**

[Matta et al., 1999] proposent un modèle de mémoire qui respecte la description d'un projet, tout en mettant à jour les connaissances et les informations dont les concepteurs ont besoin dans leurs activités. Notre modèle se décompose donc en deux parties :

- Mémoire de caractéristiques de projet.
- Mémoire de logique de conception (Design-rationale).

### **6.2.1. Mémoire de caractéristiques de projet**

Cette mémoire permet d'indexer les informations qui décrivent le contexte d'un projet, son organisation et ses résultats:

- Contexte: Directives et méthodes de conception, Exigences, Règlements
- Organisation: Tâches (définition et distribution), Participants (sous-groupes, tâches affectées).
- Résultats: Maquettes, Matériel, Logiciel, Documents techniques, Essais.

Chaque élément de cette mémoire donne accès aux informations extraites à partir des différentes ressources de l'entreprise: Bases de connaissances d'aide à la conception, Bases de données, données extraites des outils de conception, données extraites des outils de gestion, prototypes, maquettes, etc.

### **6.2.2. Mémoire de logique de conception**

Cette mémoire met en avant les connaissances investies dans la prise de décision dans la réalisation d'un projet ainsi que dans la gestion des incidents, à savoir les problèmes

rencontrés et leur résolution. Un problème peut être aussi bien un objectif à atteindre, un problème dans le processus de conception, dans l'organisation du projet, 'un problème du produit en conception, etc. La résolution de problèmes peut concerner les participants, les personnes impliquées, les méthodes de résolution et les choix potentiels.

L'évaluation des solutions concerne en général les solutions rejetées, les arguments de rejets, les avantages et les inconvénients. La décision sera prise en fonction de la solution retenue et des arguments qui portent sur les avantages et les inconvénients. Nous allons présenter dans la section suivante les différentes méthodes qui permettent de capitaliser les connaissances d'une mémoire de projet.

### **6.3. Méthodes de capitalisation**

Plusieurs approches de capitalisation des connaissances dans une mémoire de projet ont été proposées. Dans ce qui suit, nous citons un certain nombre de ces approches qui sont définies soit pour la conception soit pour d'autres domaines.

#### **6.3.1. La méthode IBIS**

L'objectif initial de la méthode était de fournir une structure pour le dialogue mené lors de la résolution de problèmes complexes de conception [Buckingham, 1997]. Depuis, la méthode a évolué et plusieurs outils support ont été développés. La méthode est actuellement utilisée pour représenter la logique de conception (Design Rationale) dans un projet.

La méthode IBIS se base sur une structuration de la prise de décision menée dans un projet de conception en trois éléments: Questions (Issues), Positions et Arguments. Dans une discussion, une question principale est généralement posée «Comment peut-on réaliser?». Un participant peut prendre une position, en proposant une réponse à la question posée. Il défend sa position avec des arguments. D'autres participants proposent d'autres manières de résoudre le problème posé en définissant des positions alternatives et en énonçant des arguments. Les arguments peuvent être des arguments de support pour la position ou des arguments d'objection pour d'autres positions. De nouvelles questions apparaissent dans la discussion. Ces questions peuvent être liées aux questions qui les ont suggérées. La figure suivante illustre le principe de cette méthode.

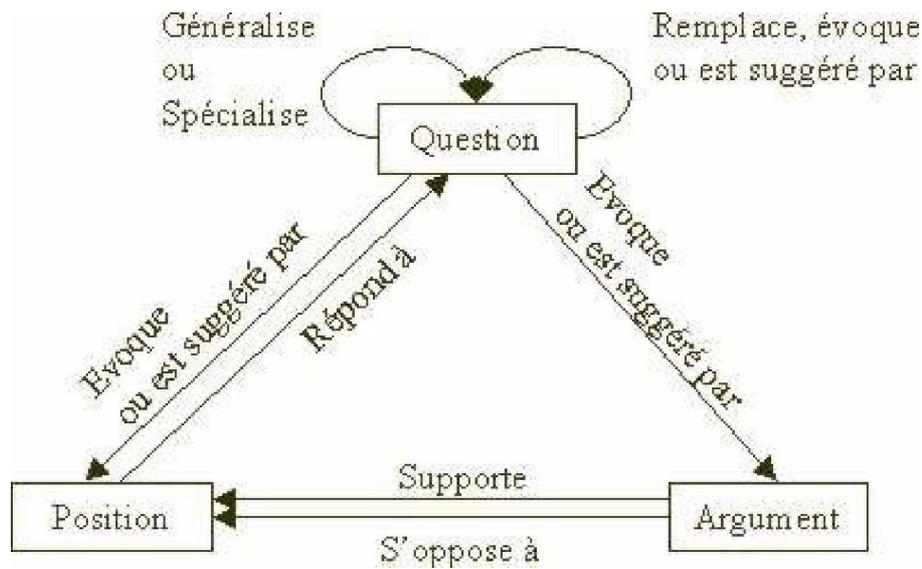


Figure 2.4. Principaux éléments de la méthode IBIS

Neuf liens ont été définis pour relier questions, positions et arguments. Il s'agit de:

- Position répond à Question
- Argument supporte Position
- Argument s'oppose à Position
- Question généralise Question
- Question spécialise Question
- Question évoque Question
- Question est suggérée par Question

### 6.3.2. L'approche QOC

L'approche QOC «Questions, Options and Criteria» [Buckingham, 1996] a pour objectif de représenter la logique de conception nommée aussi analyse de l'espace de conception. Cette approche recommande de représenter cette analyse sous forme de:

- Questions: les questions et problèmes posés lors d'une conception.
- Options: les différentes réponses données à ces questions.
- Critères: les critères qui permettent de discriminer telle ou telle option.

L'approche QOC ne fournit pas une représentation du processus de conception mais plutôt une analyse de la prise de décision dans la conception. Les auteurs de cette

approche considèrent que l'espace de conception peut être représenté par des choix de conception. Ces choix sont structurés comme réponse aux questions évoquées par les problèmes de conception. Des critères sont utilisés dans le choix de certaines options comme solution et réponse à une question. Des arguments peuvent justifier les choix d'une option suivant un critère donné. Un critère permet aussi de rendre explicite une ou plusieurs exigences dissimulées dans la question. Une structuration de l'espace de conception peut alors être obtenue sous forme d'association de questions, options et critères. Le principe de la méthode est montré par la figure suivante.

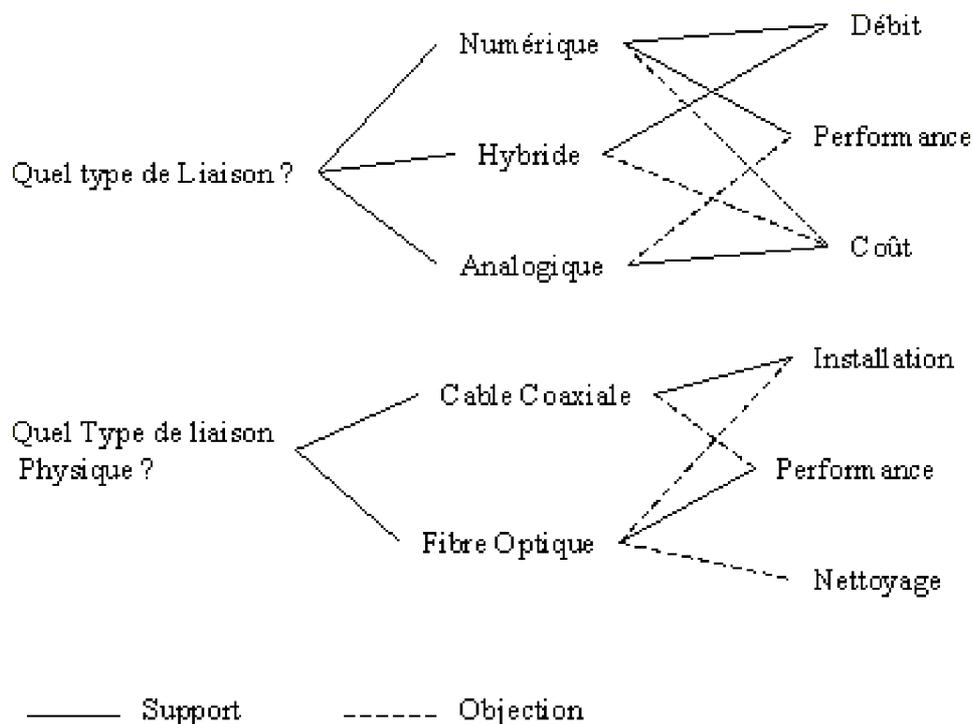


Figure 2.5. Représentation de la prise de décision selon QOC

### 6.3.3. Le système DRCS

DRCS (Design Rationale Capture System) [Klein, 1993] est un système qui permet de représenter la logique de conception dans un projet en ingénierie concurrente. Le système permet un travail collaboratif et utilise des liens hypertextes. DRCS se base sur un langage de représentation «DRCS language». Cinq types de modèles, dus aux différents types de relations utilisées entre les entités, ont été définis pour représenter la logique de conception dans un projet:

- Un modèle Synthèse permettant de mettre à jour des taxinomies,

- Un modèle Evaluation permettant de représenter comment les spécifications ont été atteintes.
- Un modèle Intention permettant de mettre en évidence les décisions prises lors de la résolution des problèmes ainsi que les stratégies qui en découlent.
- Un modèle Versions permettant de montrer les alternatives de conception pour trouver une solution à un conflit.
- Un modèle Argumentation permettant de montrer le type d'argumentation apportée à une alternative.

#### 6.3.4. Le système DRAMA

DRAMA [Brice, 1995] est un logiciel défini sur PC. Il permet de garder une trace de la logique de conception d'un projet de conception. DRAMA permet de représenter les buts, les options de solutions ainsi que le choix d'options, suivant un arbre de solution. Une table de critères permet aussi d'illustrer le choix d'une option comme solution à un but donné.

Des liens hypertextes sont offerts dans le logiciel. Ils fournissent des liens vers des documents, des dessins techniques, des simulations, etc. DRAMA permet aussi de générer d'une façon dynamique des rapports en format HTML, à partir des arbres de solution et des tables de décision. Il permet d'extraire des données de Bases de données et de les présenter en exploitant le langage XML.

CRITERE	OPTION 1	OPTION 2	OPTION 3
Performance	1	2	3
Débit	1	2	3
Coût	3	2	1
Installation	4	1	1
Score	0	-2	-1
Rang	3	2	1
Décision	Rejeté	Rejeté	Accepté

Table 2.1. Table de décision DRAMA

#### 6.3.5. L'approche EMMA

L'approche EMMA ("Evolution Memory Management Assistant") [McCullough et al., 1998] est une approche de représentation des connaissances utilisées dans un projet de

conception de logiciels. Cette approche fournit un modèle de représentation appelé "MetaModel" qui permet de décrire le processus de conception d'un logiciel sous forme de buts à atteindre. Le principe de base de l'approche consiste à distinguer dans un processus de conception :

- Les buts à atteindre.
- Les plans définis pour atteindre ces buts.
- Le contexte des plans définis
- Les changements apportés ainsi que l'évolution du processus.

### 6.3.6 La méthode SAGACE

La méthode SAGACE [Penalva, 1990], développée au sein du CEA a comme principe de base la modélisation des connaissances statiques décrivant un système de production. Cette modélisation fournit une base de dialogue entre les différents acteurs dans un projet et une aide au choix des moyens d'actions. Elle peut être ainsi exploitée pour mémoriser la réalisation d'un projet de production d'un système. Le système produit peut être à "forte automatisation", à "forte composante humaine" ou à "forte identité" comme par exemple, un système à base de connaissances.

	Activité	Fonctionnement	Evolution	
Vision Fonctionnelle	PROCESSUS	PROGRAMME	SCENARIOS	Ce que fait le système
Vision Organique	RESEAU OPERANT	RESEAU LOGISTIQUE	RESEAU AUXILIAIRE	Ce qu'est le système
Vision Stratégique	PILOTAGE	ADAPTATION	ANTICIPATION	Ce que décide le système
	Performance	Stabilité	Cohésion	

Table 2.2. Trois visions définies dans SAGACE

### 6.4. Discussion

Cette partie a présenté un ensemble de méthodes et de systèmes proposés pour aider à la capitalisation des connaissances dans une mémoire de projet. Certaines de ces méthodes sont définies pour représenter les connaissances relatives à la prise de décisions issues des réunions et des discussions menées dans un projet. D'autres méthodes fournissent des formalismes pour décrire les résultats intermédiaires et pour supporter la gestion d'un projet (l'organisation, le planning, etc.).

La plupart des méthodes sont dédiées à des projets de conception. D'autres sont plus générales. D'autres approches ont aussi été développées pour capitaliser les connaissances à partir de documents textuels, d'experts ou d'autres types de sources.

La méthode QOC permet de présenter l'analyse de l'espace de conception et non la représentation du processus de conception. Une conception étant une association de questions, options et critères. Nous nous sommes inspirés de cette logique pour la partie analyse de la prise de décision initiale dans la conception. Pour la partie structure de décision, la méthode IBIBS, nous a parue intéressante, mais elle ne permet pas d'étudier l'étape de gestion d'un projet. Pour la représentation de la gestion, nous utilisons une ontologie. Cette ontologie repose sur la définition d'un modèle de concepts et de relations entre concepts. Nous introduisons dans ce modèle comme dans celui de EMMA les concepts relatifs aux buts et aux plans. L'inconvénient dans EMMA est que cette approche se limite aux projets de conception de logiciels et elle n'est pas étendue pour tout type d'application. Nous pensons que l'utilisation d'une ontologie permet d'exploiter pour le mieux les connaissances dans le domaine d'étude ciblé.

## **6. Conclusion**

Dans une entreprise la mise en place d'une mémoire d'entreprise ne peut être que bénéfique que ça soit sur un plan organisationnel, humain, technologique, ou économique. L'aboutissement à une mémoire d'entreprise pourra être un des résultats attendus d'une démarche KM dans une entreprise. Cette mémoire d'entreprise doit assurer certaines fonctionnalités à savoir la gestion des documents, des métiers, des produits, la gestion des processus, la gestion des évolutions, la gestion des versions, la gestion des ressources, la gestion des organisations, et la gestion des justifications.

Une étude de l'ensemble des contributions pour la conception d'une mémoire d'entreprise montre qu'aucune d'elle ne couvre toutes ces fonctionnalités. C'est l'un des objectifs des travaux de recherche dans le domaine et dont l'objectif est d'avoir une mémoire d'entreprise remplissant un maximum de fonctionnalités. Nous allons tenter à contribuer dans cette lignée via le modèle de mémoire proposé et qui sera projeté pour le cas spécifique de la mémoire de projet. Ce modèle exploite une ontologie de domaine.

Dans le chapitre suivant, nous allons présenter les caractéristiques d'un web sémantique pour arriver à un web sémantique d'entreprise. Parmi les composantes du web sémantique, nous allons nous focaliser sur les notions d'ontologies, d'annotation classique et d'annotation sémantique

## ***CHAPITRE 3***

### ***Du Web sémantique au Web sémantique d'entreprise***

### ***Du Web sémantique au Web sémantique d'entreprise***

*“Semantic annotation plays both central roles of persistent repository of acquired and persistent indexing of heterogeneous information resources; together annotations and resources constitute the persistent memory”.*  
Fabien Gandon

#### **1. Introduction**

Pour de nombreux domaines, le Web et ses technologies associées sont devenues la plus grande source d'information actuelle. Mais la spécificité de telles sources d'informations les rend difficilement exploitables et leur évolution constante rend complexe les techniques de recherche d'informations. La raison principale est que les documents sont fragmentés, dispersés, hétérogènes et sont souvent très peu structurés. Il est donc, nécessaire de proposer des méthodes et des outils permettant de partager, manipuler et rechercher au sein de tels documents. L'annotation est actuellement la méthode la plus pertinente et la plus prometteuse pour pallier aux problèmes de volatilité et d'hétérogénéité des documents sur le Web. Les annotations permettent d'associer des informations complémentaires aux documents, de spécifier certaines parties de ceux-ci et enfin de les partager dans le cadre d'un groupe de travail. Cette annotation est donc très utile pour affiner les réponses aux requêtes des utilisateurs.

Dans ce chapitre, nous allons présenter le principe du web sémantique, les caractéristiques pertinentes des annotations classiques et sémantiques. Et finalement, nous allons montrer que la convergence vers un web sémantique d'entreprise est l'une des solutions cruciales dans le cadre de la matérialisation d'une mémoire d'entreprise.

#### **2. Principales composantes du web sémantiques**

Depuis sa création, le web a connu un succès gigantesque et est en train de devenir peu à peu le premier outil pour la production, la publication, la diffusion et le partage de l'information. Cependant la répartition à travers le monde d'un tel réseau d'informations, la croissance accrue du nombre de publications et la liberté totale d'y accéder ont révélé plusieurs limites et inconvénients. En effet, le web actuel ne dispose pas d'outils pour décrire et structurer ses ressources de manière satisfaisante afin de permettre un accès pertinent à l'information. Par exemple, les liens entre les pages web, bien que porteurs de sens pour les utilisateurs, n'ont

aucune signification exploitable par les machines. C'est pour pallier ces insuffisances que Tim Berners Lee a proposé dans [Berners-Lee et al., 2001] d'étendre le web actuel vers un web où l'information posséderait un sens bien défini permettant ainsi aux applications d'exploiter directement la sémantique des ressources et de coopérer avec l'utilisateur afin de lui faciliter ses tâches (recherche, commerce électronique...). Ce futur web baptisé web sémantique a été défini comme un « web intelligent » où les informations stockées dans les machines seraient en plus comprises par ces dernières afin de répondre efficacement aux requêtes lancées par les utilisateurs.

Le Web Sémantique a été proposé en se basant sur les critiques adressées au web actuel:

- l'absence de la sémantique des liens HTML, ce qui les rend pratiquement inexploitable par les machines,
- la non structuration des métadonnées utilisées et limitations dans leurs usages,
- la difficulté de faire des inférences et des raisonnements sur les connaissances décrites dans les documents publiés sur le web vu l'absence de modèles permettant la représentation sémantique de ces connaissances.

Tous ces problèmes ont fait l'objet de différents travaux de recherche qui ont convergé vers plusieurs solutions parmi lesquelles nous présentons celles qui nous semblent les plus essentielles :

- Proposer des langages et des formalismes de représentation et de structuration des connaissances. Ces langages permettent de modéliser et de représenter le contenu sémantique des ressources du web.

- Rendre disponibles des ressources conceptuelles (des modèles) représentées dans ces langages modélisant les connaissances et facilitant leur accès et leur partage : les ontologies.

- Proposer des métadonnées explicites, c'est-à-dire qui suivent un modèle et qui sont exprimées dans des langages définis formellement.

### **3. Ontologies**

Terme d'origine philosophique, ontologie désigne la théorie d'étude de la « nature de l'existant ». Cette notion a été reprise par les chercheurs dans le domaine de l'intelligence artificielle et utilisée dans le cadre de construction des systèmes à base de connaissances. L'idée était de séparer, d'un côté, la modélisation des connaissances d'un domaine, et d'un autre côté, l'utilisation de ces connaissances (i.e. le raisonnement). Dans ce contexte, plusieurs définitions des ontologies ont été proposées. La première a été proposée par [Neches et al., 1991]: « Une ontologie définit les termes et les relations de base du vocabulaire d'un domaine ainsi que les règles qui permettent de combiner les termes et les relations afin de pouvoir

étendre le vocabulaire ». Cette définition descriptive donne un premier aperçu sur la manière de construire une ontologie, à savoir l'identification des termes et des relations d'un domaine ainsi que les règles pouvant s'appliquer sur ces derniers. Deux années plus tard, [Gruber, 1993] donne la définition qui est devenue la plus utilisée dans la littérature :

*« Une ontologie est une spécification explicite d'une conceptualisation ».*

La conceptualisation se réfère ici à l'élaboration d'un modèle abstrait d'un domaine du monde réel en identifiant et en classant les concepts pertinents décrivant ce domaine. La formalisation consiste à rendre cette conceptualisation exploitable par des machines. Dans cette même logique [Guarino et Giaretta, 1995] proposent leur définition : « Une ontologie est une théorie logique proposant une vue explicite et partielle d'une conceptualisation ».

Depuis, de nombreuses définitions, à la fois complémentaires et précises, ont vu le jour. [Aussenac-Gilles et al., 2000] soulignent la dépendance entre la formalisation de l'ontologie et l'application dans laquelle elle va être utilisée : « Une ontologie organise dans un réseau des concepts représentant un domaine. Son contenu et son degré de formalisation sont choisis en fonction d'une application ». Nous retiendrons donc qu'une ontologie traduit un consensus explicite sur la formalisation des connaissances d'un domaine afin de faciliter le partage et la réutilisation de ces connaissances par les membres d'une communauté ou par des agents logiciels.

Dans l'ontologie, les connaissances du domaine sont représentées par une sémantique formelle. Les connaissances du domaine traduites par une ontologie sont véhiculées normalement par les éléments principaux suivants:

- Concepts (ou classes) : représentent les objets, abstraits ou concrets, élémentaires ou composites, du monde réel. Les concepts sont organisés en taxonomie par l'utilisation de la relation de subsomption.
- Relations : représentent des liens sémantiques de la connaissance du domaine. Une relation peut être distinguée comme une propriété ou un attribut. Les propriétés décrivent des interactions entre concepts, elles peuvent être fortement typées, i.e., associées un domaine et un codomaine précis qui permettent respectivement de spécifier les classes susceptibles d'initialiser une propriété et de contraindre les domaines de valeurs des propriétés.
- Attributs: correspondent à des caractéristiques, des spécificités particulières, attachées à un concept et qui permettent de le définir de manière unique dans le domaine. Leurs valeurs sont littérales, i.e. de type primitif, comme une chaîne de caractères ou un nombre entier.
- Axiomes : constituent des assertions (ou des prédicats) acceptées comme vraies qui s'appliquent sur les classes ou les instances des classes de l'ontologie et qui permettent de

restreindre les interprétations possibles d'une ontologie et/ou de déduire de nouveaux faits à partir des faits connus.

- Instances (ou objets ou individus) : représentent des éléments spécifiques d'un concept (ou d'une classe). Les instances peuvent être associées à un identifiant unique qui permet de distinguer une instance à une autre.

### **3.1. Types d'ontologies**

[Van Heijst et al., 1997] définissent deux grandes typologies d'ontologies : (i) une typologie fondée sur la structure de la conceptualisation et (ii) une typologie fondée sur le sujet de la conceptualisation.

Dans la première typologie, ils distinguent trois catégories à savoir

- les ontologies terminologiques (lexiques, glossaires...);
- les ontologies d'information (schéma d'une BD);
- les ontologies des modèles de connaissances.

Dans la deuxième typologie, qui est la plus citée, ils distinguent quatre catégories :

- les ontologies d'application : elles contiennent toutes les informations nécessaires pour modéliser les connaissances pour une application particulière.
- les ontologies de domaine : elles fournissent un ensemble de concepts et de relations décrivant les connaissances d'un domaine spécifique.
- les ontologies génériques (dites aussi de haut niveau) : elles sont similaires aux ontologies de domaine, mais les concepts qui y sont définis sont plus génériques et décrivent des connaissances tels que l'état, l'action, l'espace et les composants. Généralement, les concepts d'une ontologie de domaine sont des spécialisations des concepts d'une ontologie de haut niveau.
- les ontologies de représentation (dites aussi méta-ontologies) : elles fournissent des primitives de formalisation pour la représentation des connaissances. Elles sont généralement utilisées pour écrire les ontologies de domaine et les ontologies de haut niveau.

Dans cette même typologie, nous pouvons retrouver d'autres catégories telles que les ontologies de tâches et de méthodes définies dans [Mizoguchi et Vanwelkenhuysen, 1995].

### **3.2. Réutilisation des ontologies**

La réutilisation était considérée comme l'un des principaux apports de l'intégration des ontologies dans les systèmes de gestion des connaissances. Avec le développement de plusieurs applications basées sur des ontologies, cette question est de plus en plus mise en cause et alimente de longs débats. En effet, [Charlet et al., 1996] et [Aussenac-Gilles et al., 2000]

considèrent qu'une ontologie est difficilement réutilisable. D'après eux, une ontologie garde toujours la trace de la tâche pour laquelle elle a été développée. En d'autres termes, la construction d'une ontologie est toujours guidée par son application, ce qui rend le souhait d'avoir une ontologie universelle totalement utopique. D'un autre côté, [Uschold et Grüninger, 1996] proposent une solution originale à ce problème de réutilisation qui consiste à construire des bibliothèques d'ontologies, où ces dernières pourront ensuite être combinées pour pouvoir générer des nouvelles ontologies. Les auteurs proposent des recommandations pour faire ces combinaisons. La réutilisation d'une ontologie telle qu'elle dans deux tâches ou deux applications différentes semble donc difficile. Ainsi, comme le souligne [Gomez-Pérez et al., 2003], il existe une dépendance forte entre l'utilisation et la réutilisation des ontologies : « plus elle est réutilisable, moins elle est utilisable ».

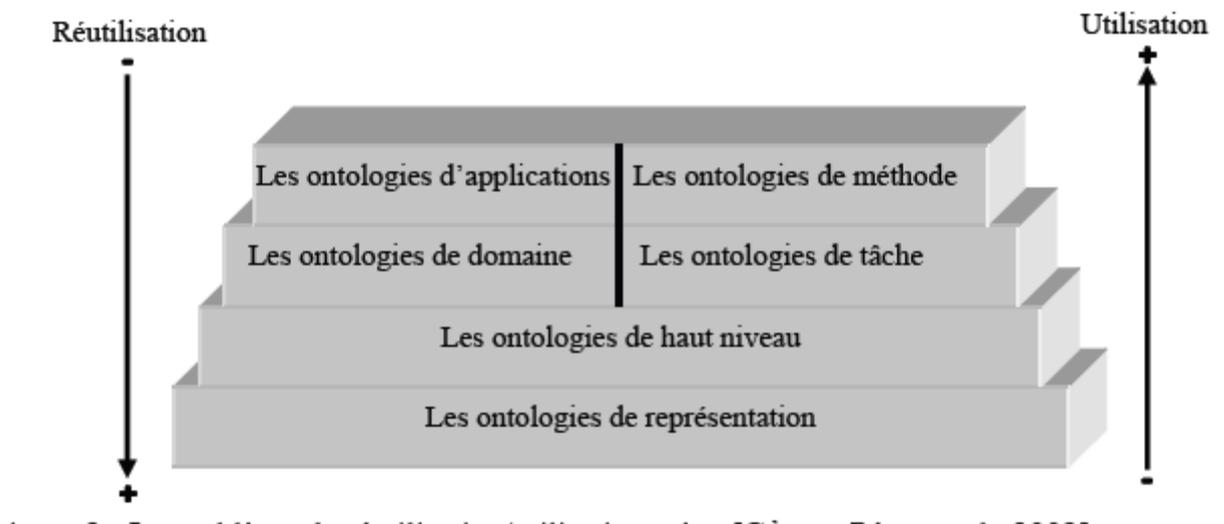


Figure 3.1 Problème de réutilisation/utilisation selon [Gomez-Pérez et al., 2003]

## 4. Annotation classique

Certes, dans notre travail nous nous intéressons aux annotations à base d'ontologies (i.e. les annotations sémantiques), mais les caractéristiques de quelques modèles d'annotations classiques nous ont paru intéressants de telle sorte, nous les avons exploités de manière structurelle, organisationnelle ou même technique dans la structuration de nos annotations.

### 4.1. Définition d'une annotation

Dans cette section nous essayons de donner les différentes définitions du terme « annotation » mentionnées par plusieurs auteurs, chacun selon son domaine de recherche. Le Petit Robert définit le terme annotation comme une « note critique ou explicative qui accompagne un texte – une note de lecture qu'on inscrit sur un livre ». De son côté, le Dictionnaire de

l'Académie Française (9ème édition) 3, précise que le terme annotation est un dérivé du terme latin *annotare*, signifiant « noter ; annoter ».

L'annotation correspond donc à « l'action d'annoter, au résultat de cette action » comme dans la phrase « Cet exemplaire contient des annotations manuscrites de l'auteur ». Le terme annoter est quant à lui défini comme « accompagner un texte de notes, de remarques, de commentaires », par exemple « Montaigne annotait sans cesse les premières éditions des "Essais" ».

- Dans [Desmontils and Jacklin, 2002], l'auteur donne une définition qui concerne l'objet annotation et la façon dont cet objet est vu par l'auteur de l'annotation et le consultant de celle-ci : « une annotation est une information graphique ou textuelle attachée à un document et le plus souvent placée dans ce document ».

- Dans le contexte des recherches sur les IHM (Interface Homme Machine), [Baldonado et al. 2000] précise que « une annotation est un commentaire sur un objet tel que le commentateur veut qu'il soit perceptiblement distinguable de l'objet lui-même et le lecteur l'interprète comme perceptiblement distinguable de l'objet lui-même ».

A partir de ces deux définitions, nous pouvons retenir que l'annotation se rapporte à une cible, et que l'auteur et le consultant de celle-ci voient que l'annotation est distinguable de la cible.

- Pour les psycholinguistes et cognitivistes tels que [Veron, 1997] cité par [Azouaou, 2006], une annotation est une « trace de l'état mental du lecteur et une trace de ses réactions vis-à-vis du document. ». Lors de la lecture du document, l'annotateur se construit une représentation mentale du document qu'il annote, ses annotations sont des traces de sa compréhension du document.

- Dans [Desmontils et al. 2002] l'annotation est : « tout objet qu'une personne ajoute à un document avec un objectif spécifique. ». Les auteurs de cette définition considèrent que l'objet est plus général, il peut prendre plusieurs formes (du texte, image, lien hypertext, etc....), et que l'annotation est faite pour un objectif spécifique, mais il se peut que le lecteur annote selon un ou plusieurs objectifs.

- Enfin, Selon [Atilf, 1992] l'annotation est « l'action d'annoter, et du résultat de cette action », en définissant l'action d'annoter comme « Accompagner un texte de notes, de remarques, de commentaires ». L'annotation indique à la fois une action et un objet.

Pour notre travail, il est à noter que nous avons considéré l'annotation comme un objet « posé sur un document » et qui sera par la suite stocké dans une base d'annotation et une activité qui vise à créer ce même objet via une approche déterminée.

## 4.2. Méta-donnée ou Annotation

La méta-donnée est différente de l'annotation « c'est une donnée sur une donnée » utilisée généralement pour indexer les ressources du web [Prié, 2003]. Elle peut former une annotation particulière, dont le type est l'un des deux :

- des méta-données portant sur le document annoté (nom auteur, date production...).
- des méta-données portant sur l'annotation (nom de l'annotateur, date de production de l'annotation...).

Une méta-donnée a une pertinence a posteriori. Elle sera utile après la saisie ou au moment de la recherche, par exemple la recherche de documents qui ont pour auteur « Mr X ». La figure 3.2 illustre la relation entre une méta-donnée et une annotation [Bringay, 2003].

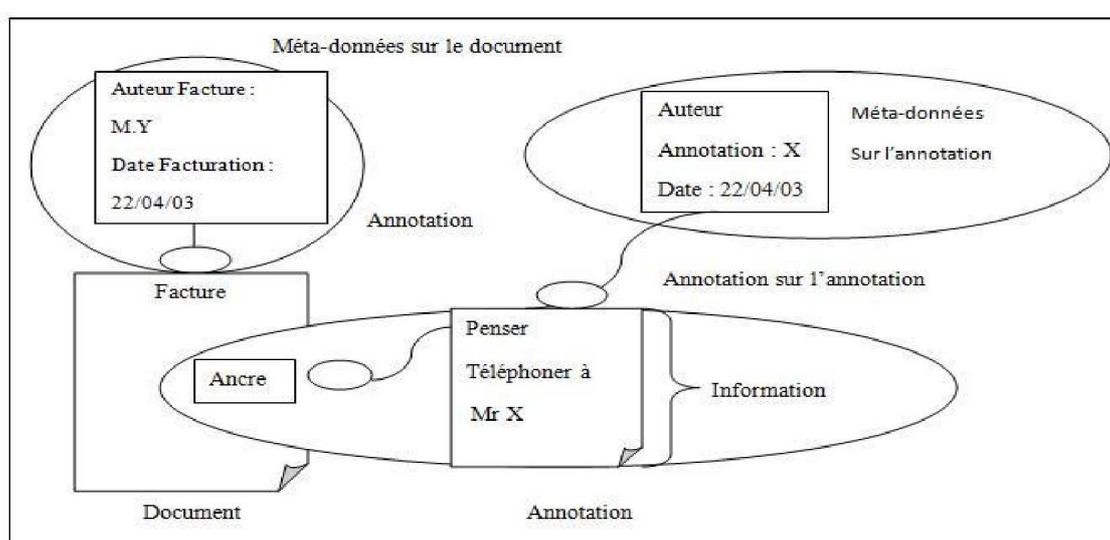


Figure 3.2. Relation entre l'annotation et la Méta-donnée

Nous avons considéré les métas données portant sur le document et sur l'annotation comme étant des rubriques attachées aux annotations dans la base d'annotation.

## 4.3. Quelques modèles pour l'annotation classique

Les annotations sont représentées de manière structurée, basée sur une formalisation bien définie, et sont caractérisées par des propriétés qui permettent de les identifier et de les décrire, nous présentons dans ce qui suit quatre modèles :

### 4.3.1. Modèle de Denoue [Denoue, 2000]

La structure proposée par Denoue est une structure à deux parties : une obligatoire qui est l'ancre de l'annotation qui permet d'attacher l'annotation à un endroit précis du document. La deuxième partie qui est le commentaire de l'annotation, c'est un ensemble d'attributs facultatifs tels que l'auteur, la date de création, le thème de document, un texte libre...

#### 4.3.2. Modèle de VERON [Veron, 1997]

L'auteur propose un modèle pour la représentation des annotations composé des propriétés suivantes :

- **La forme de l'annotation**, concerne l'aspect visuel de l'annotation et de son apparence graphique. Elle dépend des outils mis à disposition de l'utilisateur par le logiciel d'annotation. Exemples : souligné, surligné, marque, etc.
- **Le but de l'annotation**, c'est l'objectif pour lequel l'annotation est faite. Les buts annotatifs sont différents des buts de la lecture. Ces buts fixent la sémantique des annotations attachées sur le document.
- **Le lieu de l'annotation**, c'est l'endroit où est placée l'annotation. Il diffère selon le but de cette annotation.
- **L'auteur**, c'est la personne ou l'annotateur qui annote le document. L'annotateur est très important. Cette importance est vue dans le cas où plusieurs personnes peuvent annoter le même document.
- **L'histoire**, c'est la propriété qui décrit le moment où l'annotation a été placée sur le document. Elle comporte la date de création de l'annotation, mais aussi les dates de modification. La gestion de l'historique des annotations est comparable à une gestion des différentes Versions d'un document.
- **Le support de l'annotation**, c'est l'entité concernée par l'annotation. Elle peut être le document en entier, une phrase, un terme, un mot, une image...

#### 4.3.3. Modèle d'Annotea [Kahan et al., 2002]

Annotea est un système client-serveur collaboratif pour l'annotation de documents. Les annotations sont externes aux documents et peuvent être stockées dans un ou plusieurs serveurs d'annotation, afin de permettre aux utilisateurs connectés à un serveur d'annotation d'ajouter, modifier et consulter des annotations. Chaque annotation se compose de ces parties principales :

- **Type (*rdf:type*)**: une indication de l'intention du créateur de l'annotation lors d'une annotation.
- **Annote (*annotates*)**: la relation entre une annotation et la ressource à laquelle s'applique l'annotation.
- **Corps (*body*)**: le contenu de l'annotation.
- **Contexte (*context*)**: indique la partie du document où l'annotation est directement appliquée.

- **Auteur (dc:creator):** le créateur de l'annotation.
- **Crée (created):** la date et l'heure à laquelle l'annotation a été créée.
- **Modifié (dc:date):** la date et l'heure à laquelle l'annotation a été modifiée.
- **Lié (related):** la relation entre une annotation et une (collecte de) ressource (s) qui peuvent enrichir cette annotation. La figure suivante illustre la structure d'Annotea

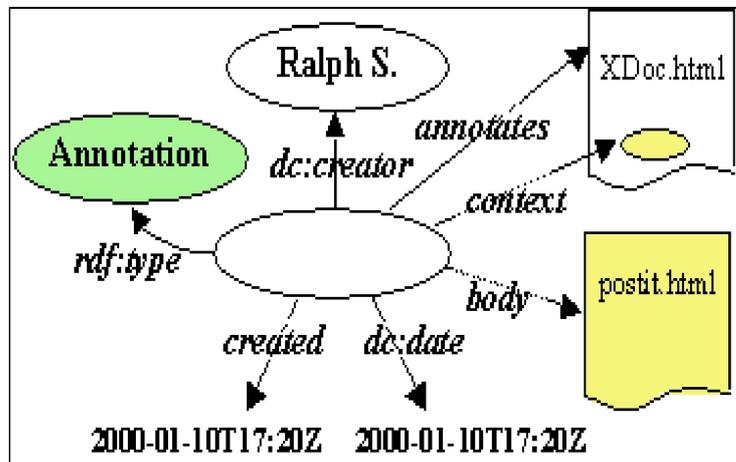


Figure 3.3. Modèle Annotea d'une annotation [Kahan et al. 2002]

D'un point de vue pratique, le modèle d'Annotea, nous semble plus conviviale à l'ensemble des acteurs ordinaires d'une entreprise. De ce fait, nous pouvons l'exploiter en proposant son utilisation via une intégration dans l'environnement développé.

#### 4.4. Acteurs interagissant avec l'annotation classique

Une annotation ayant été définie comme un document, nous retrouvons les acteurs suivant [Bringay, 2003], qui peuvent agir sur celle-ci, à différentes étapes de son cycle de vie:

- L'annotateur : auteur de l'annotation et peut être l'auteur du document annoté.
- Le consultant : celui qui perçoit l'annotation et peut être l'annotateur .il peut être visé (connu par l'annotateur au moment de la rédaction de l'annotation) ou inopiné.
- Le gestionnaire des annotations : en cas de documents numérique, c'est celui qui gère les annotations et les organise, pour pouvoir les retrouver ultérieurement avec les objets annotés. Il peut être l'annotateur, le consultant ou les deux, comme il peut être humain ou un agent logiciel. .

#### 4.5. Cycle de vie des annotations classiques

Ce cycle comprend généralement trois étapes à savoir:

- **La production de l'annotation (annotateur) :** consiste à choisir la cible, y attacher une ancre, puis inscrire des traces sur le support de l'annotation pour produire une information.

- **Le stockage et l'indexation de l'annotation (gestionnaire) :** consiste à classer les annotations (base d'annotations) et associer à cette classification des informations avec la possibilité d'identifier, de localiser et d'accéder à ces annotations.
- **La consultation d'une annotation (consultant) :** consiste à localiser une cible sur laquelle porte l'annotation et associer une signification à son information.

La figure 3 illustre le cycle de vie des annotations classiques [Bringay, 2003]

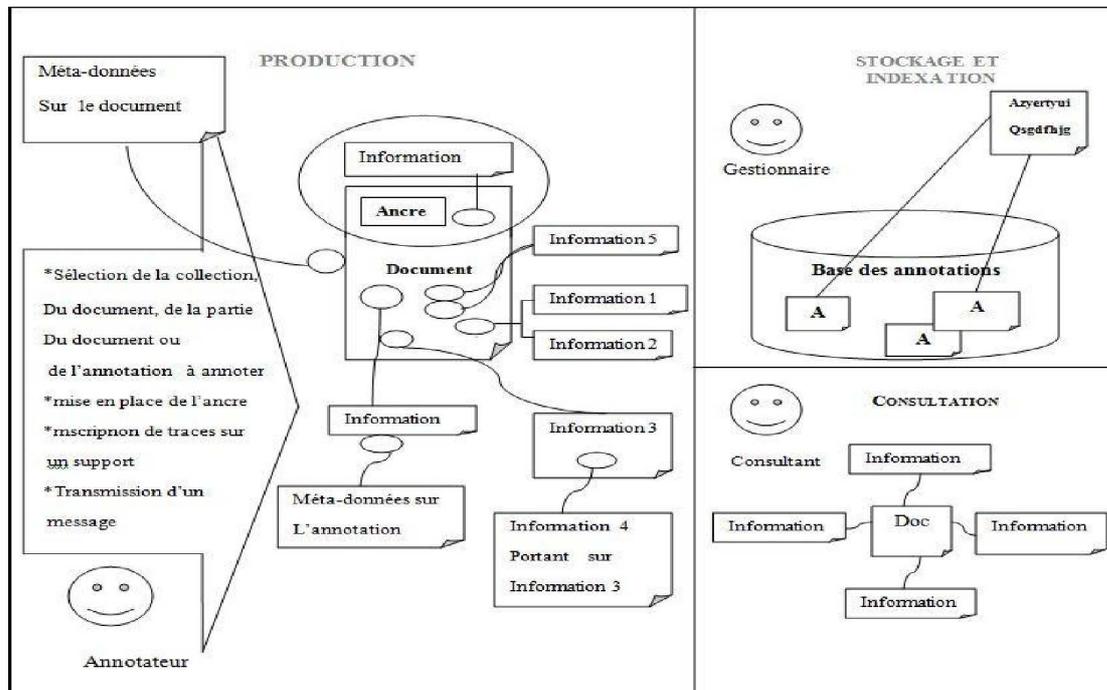


Figure 3.4. Cycle de vie des annotations [Bringay,2003]

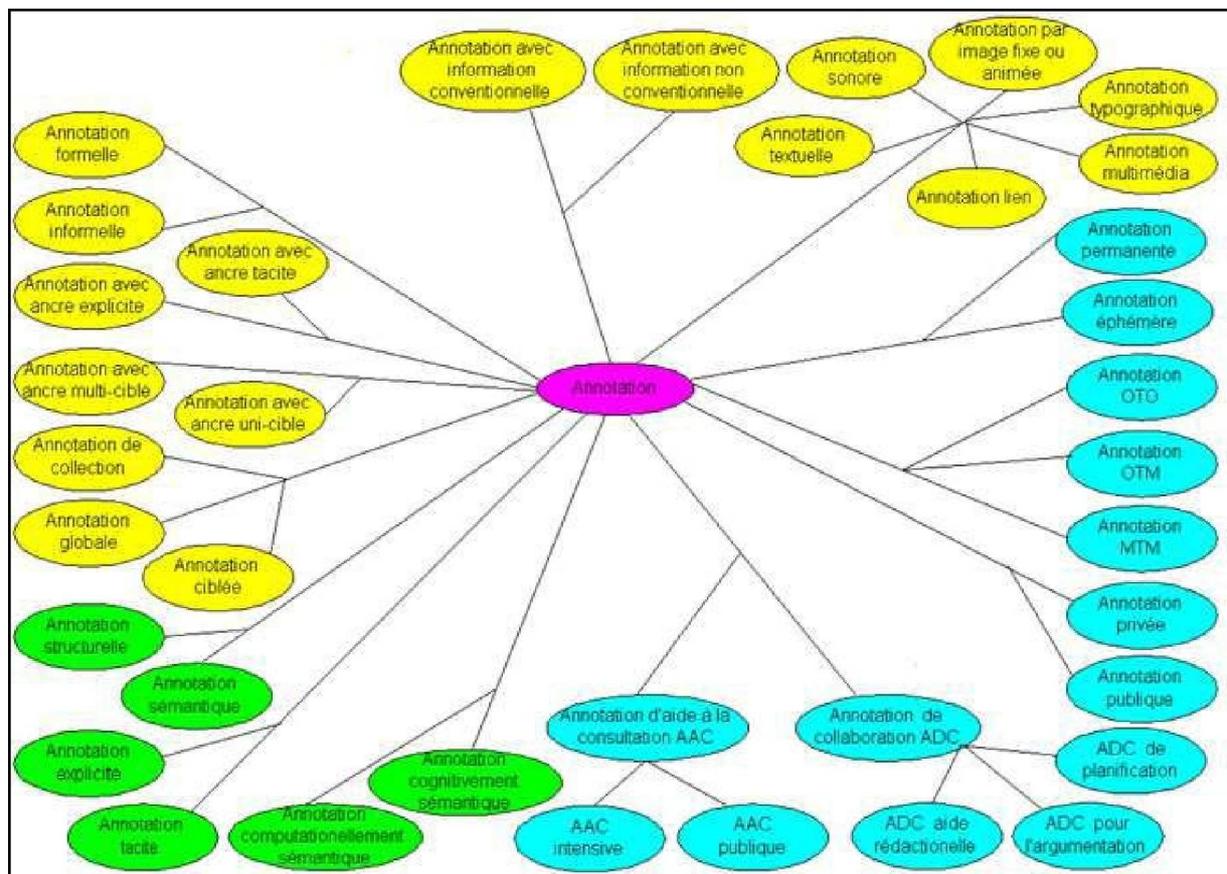
#### 4.6. Dimensions des annotations classiques

C. C. Marshall [Marshall, 1998] propose de caractériser les annotations selon différentes dimensions. Ces dimensions sont des espaces mono-dimensionnels continus. Elles décrivent les propriétés de l'annotation au niveau de sa structuration, de sa fonction et de son rôle dans la communication rédacteur/lecteur. Dans cette section, nous allons décrire ces dimensions en spécifiant les cas typiques, extrêmes, mais il est souvent possible de trouver des exemples d'annotations "intermédiaires".

- Dimensions liées à la structuration de l'annotation

- Dimensions concernant les fonctions de l'annotation

- Dimensions concernant le rôle de l'annotation dans la communication rédacteur/lecteur  
Le schéma suivant illustre les différentes dimensions utilisées pour décrire les catégories des annotations et qui sont : la forme, le signe et la relation.



- Une annotation considérée comme une forme
- Une annotation considérée comme une relation
- Une annotation considérée comme un signe

Figure 3.5. Schéma récapitulatif des types d'annotations selon les dimensions

#### 4.7. Quelques outils pour l'annotation classique

Plusieurs chercheurs et industriels se sont intéressés à la création des outils d'annotations afin de faciliter la tâche d'ajouter des notes sur des documents. On peut trouver de nombreux logiciels commerciaux et des prototypes de recherche créés pour annoter des documents consultés sur le Web, des documents personnels, pour construire des documents, etc. Les fonctionnalités et les interfaces utilisateur sont différentes et dépendent des contextes de création de ces outils et des matériels sur lesquels ces applications fonctionnent.

Bringay [Bringay 2003] classe les outils d'annotation en cinq catégories :

- Outils destinés à l'annotation de page Web : Imarkup, Yawas,
- Outils destinées à l'annotation de documents pouvant être utilisés pour des lectures personnelles ou pour la construction de documents : Adobe Acrobat et Microsoft Office (Word),

- Outils destinés à l'annotation de contenus multimédias : Dehora permet à des experts d'annoter des livres scannés,
- Outils destinés aux diapositifs mobiles, comme les PDA, les e-books, les tablettes PC : Adobe eBook Reader, Xlibris,
- Outils destinés à organiser un ensemble de ressources disponibles en ligne par la création de signets (marque-pages pouvant être vues comme des annotations) (exemple nestor).

#### 4.9. Synthèse

Les annotations sont multiples dans leurs formes et dans leurs fonctions. Par rapport au Web, les annotations les plus courantes sont des annotations en langage naturel (donc informelles) ou symboliques (surlignage), publiques, éphémères, explicites et le plus souvent associées à des termes. Elles sont généralement utilisées comme support de lecture (commentaires, recherche de passages, surlignage, cartes de navigation). Elles aident les annotateurs aussi bien que les consultants, à former une perception du document annoté.

Le stockage de l'annotation, en relation avec les éléments liés au contexte de sa production (nom de l'auteur, son point de vue sur le document annoté, etc) facilite une utilisation postérieure du document, tels que Yawas, Imarkup et nestor. Ces outils exploitent le fait que les annotations correspondent aux centres d'intérêt des consultants, alors il est possible de détourner les annotations de leur fonctions initiale, pour aider le lecteur lors de la consultation d'un document ,en proposant des outils d'aide à la navigation (carte nestor), de recherche d'informations (requêtes) et d'adaptation des documents à l'utilisateur (filtrage) .

Nous nous sommes influencés par des travaux comme celui de [Denoue, 2000], du moment où il propose d'exploiter les annotations classiques afin de capitaliser l'apport d'information issue de ces notes. Les annotations d'un document sont en effet exploitables pour améliorer la gestion du document. Une première utilisation des annotations peut être la constitution d'un résumé du document. En effet, les parties annotées peuvent être considérées comme importantes pour le lecteur. Aussi, construire un texte à partir de ces morceaux s'avère une méthode assez performante pour résumer le texte. Les annotations classiques sont aussi potentiellement intéressantes dans le cadre de la recherche d'information. En effet, les termes annotés peuvent être représentatifs du document. A cela, on peut ajouter l'exploitation du contenu des annotations dans l'indexation ou même au niveau de la recherche d'informations.

Après avoir présenté les annotations classiques, nous allons aborder dans la prochaine section, la problématique des annotations sémantique qui constituent un cas particulier des annotations.

## 5. Annotation sémantique

### 5.1 Définition de l'annotation sémantique

L'annotation sémantique consiste à assigner aux entités d'un document électronique, des liens à leurs descriptions sémantiques. Ce type de méta-données fournit, à propos des entités du document annoté, des informations de classes et d'instances liées à une ontologie décrivant le contenu de la ressource annotée, ce qui est illustré dans la figure 3.6.

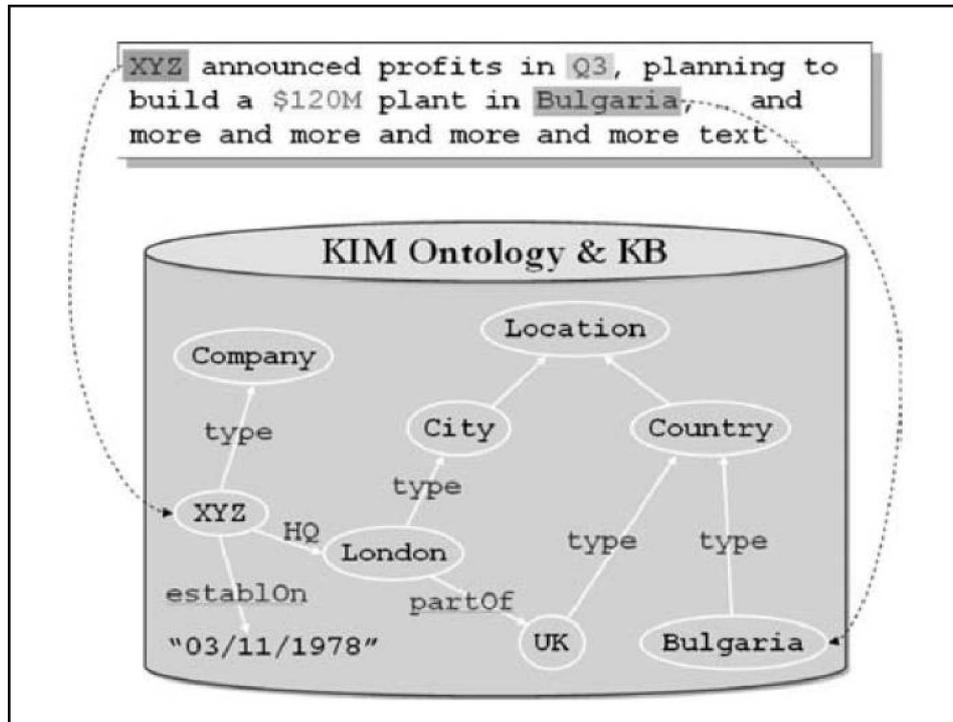


Figure 3.6. Annotation sémantique

les annotations sémantiques peuvent donc être définies comme : un ensemble d'instanciations attachées à un document (HTML ou XML) , qui peuvent être des instances de classes ( avec des URIs uniques), des instances d'attributs d'une classe ( propriétés instanciées d'une instance de classe vers une instance de type de données) ou une instance de relation (propriété instanciées d'une instance de classe vers une instance d'une autre classe). Par exemple en associant une notice comprenant des champs : Auteur, Date de création, Date de modification, Mots-clés, à une page Web, ceci permet de considérer ces informations non plus seulement comme comprenant du texte qui ne pourra qu'être traité statistiquement par un robot indexeur (« wrapper »), mais également des informations structurées dotées de sémantique connue et utilisable comme telle par un agent logiciel.

L'annotation se distingue de l'indexation automatique par l'utilisation d'une ou plusieurs ontologies qui définissent un domaine global de référence permettant de cadrer et de normaliser les annotations effectuées, par ailleurs une ressource annotée doit l'être non pas par une liste de mots clefs, mais par une ou plusieurs ontologies.

## 5.2. Annotation et méta-donnée sur le web sémantique

De façon générale, les termes « méta –donnée »et «annotation »prennent en compte la notion d'ajout d'information à une ressource. Nous pourrions à priori les utiliser indifféremment, pour décrire les informations que le web sémantique doit ajouter au web classique, afin de le rendre utilisable par des machines. Pour autant, si ces deux termes existent, c'est qu'ils n'ont pas le même sens [Garlatti and Prié, 2003]. La communauté anglophone du web sémantique considère que les annotations de pages web deviennent des méta- données dès qu'elles sont stockées dans une base sur un serveur.

D'un point de vue plus lié à la pratique des annotations et méta- données, une méta-donnée sera plutôt attachée à une ressource identifiée en tant que telle sur le web et aura une pertinence à priori. Elle est saisie suivant un schéma, ce qui permet de mettre en place des inférences. Tandis qu'une annotation est plus située au sein de cette ressource et écrite au cours d'un processus d'annotation (elle peut être complètement libre). Evidemment, cette distinction n'a rien de définitive, il s'agit simplement de mettre l'accent sur le caractère plus situé au sein de la ressource de l'annotation, par rapport à une méta-donnée plus indépendante. La figure 3.7 fournit une vision sur l'utilisation des méta-données sur le Web sémantique. Des pages Web sont annotées à partir de connaissances disponibles dans une ou plusieurs ontologies (qui ont pour objectif de normaliser la sémantique des annotations), et ces annotations, regroupées en entrepôts de méta-données deviennent utiles pour des agents de recherche d'information, faisant ou non appel à des moteurs d'inférence permettant de déduire de nouvelles connaissances formelles des annotations.

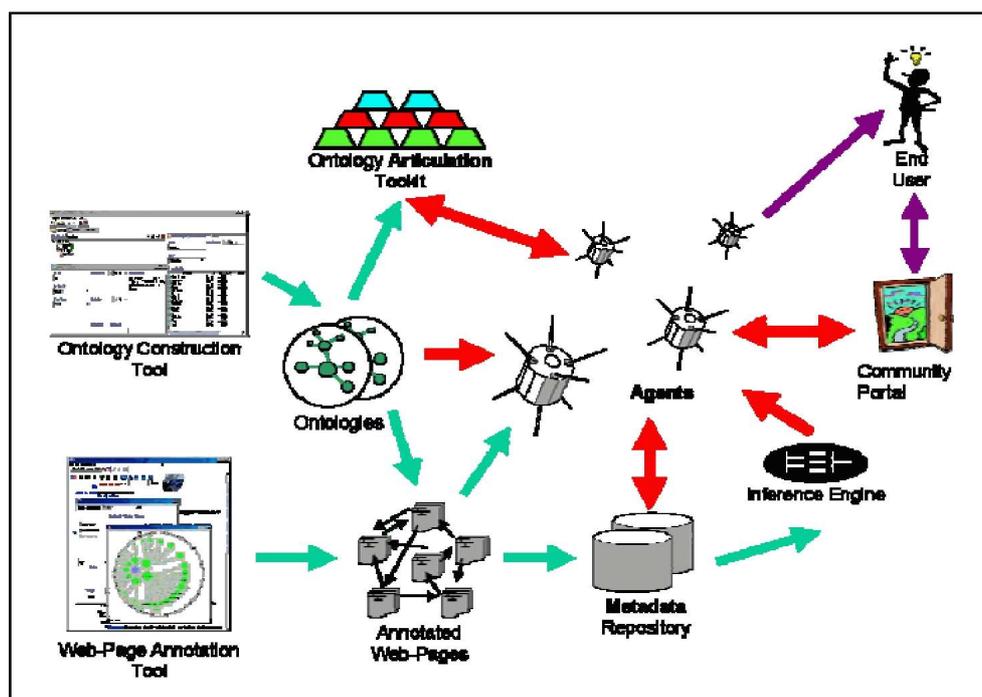


Figure 3.7. Utilisation des méta-données sur le Web sémantique

### 5.3. Caractéristiques des annotations sémantiques

Marshall a distingué sept dimensions permettant de caractériser les annotations de ressources documentaires [Marshall, 1998]. Ces dimensions sont applicables à l'annotation sémantique de ressources numériques, que ces dernières soient disponibles sur le Web ou sur un réseau d'entreprise. L'analyse de Marshall concluait que les annotations liées aux ressources documentaires traditionnelles, c'est-à-dire sur support « papier », sont généralement personnelles, informelles, intensives, transitoires et privées. A contrario, les annotations de ressources numériques sont créées dans un monde documentaire où ces ressources ont pour vocation d'être partagées et exploitées par des utilisateurs divers et variés.

A partir des travaux de Marshall, nous avons élaboré un document de type « inventaire » pour l'entreprise susceptible d'être exploité par les annotateurs. Dans ce document, les annotations sont prises en compte selon plusieurs aspects: formelles, hyper extensives, permanentes, globales et publiques (voir annexe 3).

### 5.4. Processus annotation sémantique

Derrière le processus général d'annotation sémantique de documents par des ontologies se cache plusieurs phases qui font référence à des annotations de natures un peu différentes. Ces phases sont au nombre de trois :

- **Repérer** : processus manuel [Handschuh, 2002] ou automatique [Desmontils and Jacquin, 2002] qui consiste à placer dans le document des références aux concepts de l'ontologie qu'il contient. Ces éléments sont considérés comme des méta-données.
- **Instancier** : processus manuel ou automatique permettant de valuer les attributs des concepts à l'aide des informations présentes dans le document (là encore, ce sont des méta-données).
- **Enrichir** : processus manuel visant à ajouter des informations par l'intermédiaire des attributs de concepts qui n'ont pas pu être évalués à la phase précédente.

Les deux premières étapes sont des étapes d'insertion de méta-données : il n'y a pas ajout d'information mais plutôt localisation et caractérisation de l'information déjà présente de manière intrinsèque. Par contre, la dernière est plutôt une étape d'annotation plus classique, car il y a ajout d'information : le document est enrichi d'information qui n'est pas explicitement présente dans le document. Cette annotation est directement "formalisée" par des méta-données.

### 5.5. Exploitation des annotations sémantiques

Les annotations sémantiques sont en effet exploitables pour améliorer la gestion des

documents annotés. Elles peuvent être utilisées dans le cadre de la recherche et la classification de documents, le résumé automatique et généralement, pour favoriser l'interopérabilité.

L'annotation sémantique peut être applicable pour toute sorte de texte : page web, documents réguliers, champs texte dans des bases de données, etc. Aussi, l'acquisition des connaissances peut se baser sur l'extraction des dépendances plus complexes et l'analyse des relations entre entités, événements et descriptions de situations. Les deux grandes tâches pour lesquelles les annotations sémantiques sont utilisées sont la recherche d'informations et la composition de services et de documents.

La composition est faite pour construire de nouveaux documents (cours à la carte, catalogues, réponses à une question, etc.) ou de nouveaux services. La construction de résumés automatiques de documents est une autre application qui prend en considération les parties annotés considérées comme importantes. Construire un texte à partir de ces morceaux s'avère une méthode assez performante pour résumer ce texte. Au delà de l'indexation, dans le même ordre d'idée que le résumé automatique, Denoue [Denoue, 2002] a proposé une méthode de classification de documents basée sur les termes annotés.

Dans notre contexte de travail, nous voulons de plus bénéficier des annotations sémantiques pour gérer une mémoire d'entreprise conçue à base de documents.

## 5.6. Méthodes et éléments de base pour l'annotation sémantique

[Garlatti and Prié, 2003] a défini les standards existants, et qui forment l'ossature technologique du web, pouvant servir dans le cadre de l'annotation sémantique et qui sont:

- **Le protocole HTTP** : pour la transmission.
- **Les langages HTML/XML** et les feuilles de style : pour la présentation des résultats.
- **Le langage RDF** (Ressource Description Framework) : pour l'expression et l'échange des méta-données, sous la forme de triplets, même si ceux-ci peuvent également être stockés dans des bases de données pour une gestion plus efficace.

Le stockage des annotations au début était dans les documents eux mêmes (balises ad hoc ou RDF), dans les URL (Uniform Ressource Locator) ou dans des bases sur des serveurs centralisés et les annotations étaient typées, plus ou moins partagées et souvent textuelles. Actuellement, il existe un ensemble d'éléments qui constituent, tous à la fois, un référentiel aussi bien technique que méthodologique sur lequel l'annotation sémantique de documents web peut s'appuyer. Il s'agit de :

- Méthodes de conception de schémas de méta-donnée (thésaurus, ontologies).
- Outils et principes d'utilisation de méta-données, d'annotation et de présentation et de présentation de celles-ci.

- Architectures de stockage, de requêtes et de diffusion d'annotation.
- Méthodes, outils et techniques d'extraction d'information des documents (fouille de données textuelles, méthodes statistiques, remplissage automatique de champs de méta-données, propositions aux utilisateurs, etc.).

Nous avons réalisé une étude comparative sur certains outils de l'annotation sémantique présentée en annexe 2 [Thabet et al., 2010].

### **5.7. Langages pour l'annotation sémantique dans le cadre du web sémantique**

Le Web Sémantique apparaît aujourd'hui comme la prochaine évolution importante du Web. Différents langages de représentation des connaissances dotés d'un niveau de complexité croissante et adaptés au Web ont été créés.

#### **- Le langage XML :( Définition de la structure d'un document)**

Le langage XML est aujourd'hui un langage qui est au cœur d'une réflexion sur le Web Sémantique. Le langage HTML précédemment défini était figé et restrictif. Il n'autorise pas l'utilisation de ses propres balises. D'autre part c'est un langage trop porté vers la présentation et non sur la description des données. C'est pourquoi le langage HTML est inefficace pour donner une dimension informationnelle aux données (i.e. pour exprimer des méta-données).

Le langage XML est un langage de balisage. Il permet de structurer l'information contenue dans un document grâce à des balises. Et contrairement à HTML, XML offre la possibilité de créer ses propres balises. La force de XML réside dans sa capacité à permettre une plus grande interopérabilité des applications à un niveau syntaxique. En effet, il standardise une représentation hiérarchique de l'information contenue dans les documents qu'elles peuvent s'échanger. La capacité d'échanger des informations de manière structurée et totalement indépendante des différentes plates-formes logicielles et matérielles est en effet un prérequis à tout système distribué comme le Web sémantique.

#### **- RDF : Langage d'assertions et d'annotations**

Les assertions affirment l'existence de relations entre des objets. Elles sont donc adaptées à l'expression des annotations que l'on veut associer aux ressources du Web]. Le W3C a adopté le langage RDF comme formalisme standard de représentations, utilisant la syntaxe XML qui constitue déjà un standard.

Le langage RDF fournit un moyen pour ajouter la sémantique à un document sans faire oppositions sur la structure du document. Le but de RDF est de fournir un cadre de description de données contenues dans des ressources du Web (principalement des pages Web), sans préjuger à priori sur des domaines d'applications particuliers ou sur la sémantique.

Les énoncés RDF sont des triplets (ressource –attribut- valeur), la valeur est une ressource ou une chaîne de caractères et la ressource doit disposer d’une URI telle que les triplets sont interprétables comme (sujet – prédicat – objet). Cela fait de RDF un langage relationnel de base. Bien que la syntaxe adoptée pour ce langage soit XML, son modèle de données est plutôt proche des réseaux sémantiques. La figure 3.8 illustre un exemple d’annotation en RDF.

### - Langage RDFS

Le langage RDFS a pour but d’étendre le langage RDF en décrivant plus précisément les ressources utilisées pour étiqueter les graphes. Pour cela, il fournit un mécanisme permettant de spécifier les classes dont les ressources étiquetées par ces classes seront des instances. L’objectif de RDFS, est la définition de modèles souvent appelés «vocabulaires». Ce vocabulaire sera spécifique à un domaine applicatif. Une application pourra alors, par simple analyse des types de données définis dans le modèle, connaître un certain nombre d’interprétations applicables aux annotations liées à ce modèle. Ce langage est basé sur RDF et permet de définir des notions de classes (en mettant en place des mécanismes d’héritage) et des contraintes sur des classes d’objets.

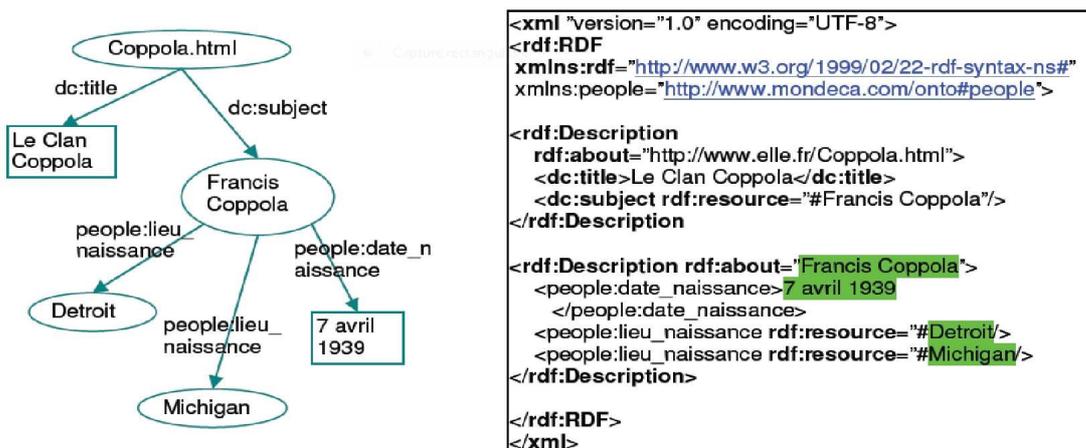


Figure 3.8. Exemple d’annotation sémantique en RDF (notation graphique à gauche et XML à droite)

### - RDFa

RDFa est un langage candidat à une recommandation du W3C. Son objectif principal est de rajouter les informations en RDF dans les documents HTML ou XHTML (i.e. (X)HTML). Le langage RDFa fournit une syntaxe et un ensemble de balises (tags) pour décrire les données structurées en (X)HTML. Avec les informations supplémentaires insérées dans les balises, les données structurées en (X)HTML sont rajoutées les “sémantiques” qui permettront l’échange d’informations par les applications automatiques ou par les agents informatiques.

## **- DAML+OIL**

Le langage de définition d'ontologies (DARPA Agent Markup Language). + (Ontology Interchange Language) est créée de la fusion de deux langages. Il a été introduit dans le but de surmonter les limites de RDFS en matière d'inférences. Ce langage est dédié aux définitions de classes et de types de propriétés de manière plus expressive.

## **- RIF (Rule Interchange Format)**

RIF est un formalisme proposé par le W3C permettant de fournir l'interopérabilité entre les langages de règles en général et ceux utilisés en particulier pour le Web. Le noyau de ce langage, RIF Core, correspond à la logique de Horn. Il offre certaines extensions inspirées des langages à objets et de frames, ainsi que les URIs, les types de données de XML Schema... La partie principale de RIF Core est le langage de condition (Condition Language). Ce langage définit la syntaxe et la sémantique des règles de RIF ainsi que la syntaxe pour des requêtes. D'autre part, la spécification de RIF repose sur certains types de règles : les règles de production, la programmation logique, les règles basées sur la logique du premier ordre, les règles réactives ou les règles.

## **- Langage OWL**

OWL est, tout comme RDF, un langage XML profitant de l'universalité syntaxique de XML. Fondé sur la syntaxe de RDF/XML, OWL offre un moyen d'écrire des ontologies web. OWL se différencie du couple RDF/RDFS en ceci que, contrairement à RDF, il est justement un langage d'ontologies. Si RDF et RDFS apportent à l'utilisateur la capacité de décrire des classes (avec des constructeurs) et des propriétés, OWL intègre, en plus, des outils de comparaison des propriétés et des classes : identité, équivalence, contraire, cardinalité, symétrie, transitivité, disjonction, etc. Ainsi, OWL offre aux machines une plus grande capacité d'interprétation du contenu web que RDF et RDFS, grâce à un vocabulaire plus large et à une vraie sémantique formelle. OWL a été fractionné en trois langages distincts :

- **OWL Lite**

- Convient aux utilisateurs qui ont principalement besoin d'une hiérarchie de classification et de contraintes simples.
- Reprend tous les constructeurs de RDF (c'est-à-dire fournit des mécanismes permettant de définir un individu comme instance d'une classe, et de mettre des individus en relation),
- Utilise les mots-clés de RDFS (rdfs:subClassOf, rdfs:Property, rdfs:subPropertyOf, rdfs:range, rdfs:domain), avec la même sémantique,
- Supporte les contraintes de cardinalité, mais ne permet d'utiliser que les valeurs 0 ou 1.

- **OWL LD**
  - Convient aux utilisateurs qui veulent le maximum d'expressivité,
  - Reprend tous les constructeurs d'OWL LITE,
  - Permet tout entier positif dans les contraintes de cardinalité,
  - Tire son nom de sa correspondance avec les logiques de descriptions.
- **OWL FULL**
  - Reprend tous les constructeurs d'OWL DL,
  - Reprend tout tous les constructeurs de RDF Schema,
  - Permet d'utiliser une classe en position d'individu dans les constructeurs.

## 6. Le web sémantique d'entreprise

Le Web sémantique d'entreprise (WSE) est utilisé pour la matérialisation d'une mémoire d'entreprise dans un environnement hétérogène et distribué qui peut entraîner des incohérences dans le système [Luong, 2009]. Le WSE est un cas particulier de WS adaptée pour les entreprises. De ce fait, les deux composants principaux du WSE demeurent l'ontologie et l'annotation sémantique. Dans la littérature existe quelques approches pour le WSE, parmi les quelles nous pouvons citer l'approche Acacia.

### - Approche Acacia

Dans cette approche [Dieng-Kuntz et al., 2005] les auteurs proposent une méthode particulière de la gestion des connaissances qui est la constitution d'une mémoire d'entreprise ou mémoire organisationnelle, matérialisant et indexant les connaissances et informations cruciales de l'organisation (i.e. une entreprise, une institution ou une communauté de pratique) afin d'améliorer leur accès, partage et réutilisation voire de permettre la création de nouvelles connaissances par les membres de l'organisation dans leurs tâches individuelles et collectives.

De ce fait, la mémoire d'entreprise est matérialisée sous forme d'un Web Sémantique d'entreprise. Parmi les projets réalisés au sein de cette équipe, nous pouvons citer : Samovar [Golebiowska et al., 2001] qui est un système/méthode de capitalisation de connaissances dans le domaine automobile. CoMMA (Corporate Memory Management through Agents) est un projet européen qui a permis de construire une mémoire d'entreprise matérialisée dans une base documentaire annotée par des annotations sémantiques en utilisant une société d'agents coopérant et guidés par l'ontologie O'CoMMA [Gandon, 2002]. CORESE (COncceptual REsource Search Engine) est un moteur de recherche sémantique dédié au langage RDF(S) et qui peut être utilisé pour interroger les différentes ressources d'un web sémantique d'entreprise [Corby et al., 2004]. MEAT (Mémoire d'Expériences pour l'Analyse du

Transcriptome) est un projet de construction d'une mémoire d'expériences, qui vise à faciliter la validation et l'interprétation des expériences puces à ADN [Khelif, 2006].

La Figure 3.9. présente l'architecture d'un Web sémantique d'entreprise dans lequel les technologies du Web sémantique sont utilisées pour réaliser les tâches du Web sémantique d'entreprise. Les ontologies communes décrivent d'une manière formelle des connaissances du domaine de l'entreprise. Les annotateurs annotent sémantiquement des ressources d'entreprise telles que la description du contenu des documents, les compétences des personnes, les caractéristiques des services... en utilisant le vocabulaire commun et partagé au sein de l'entreprise par des ontologies.

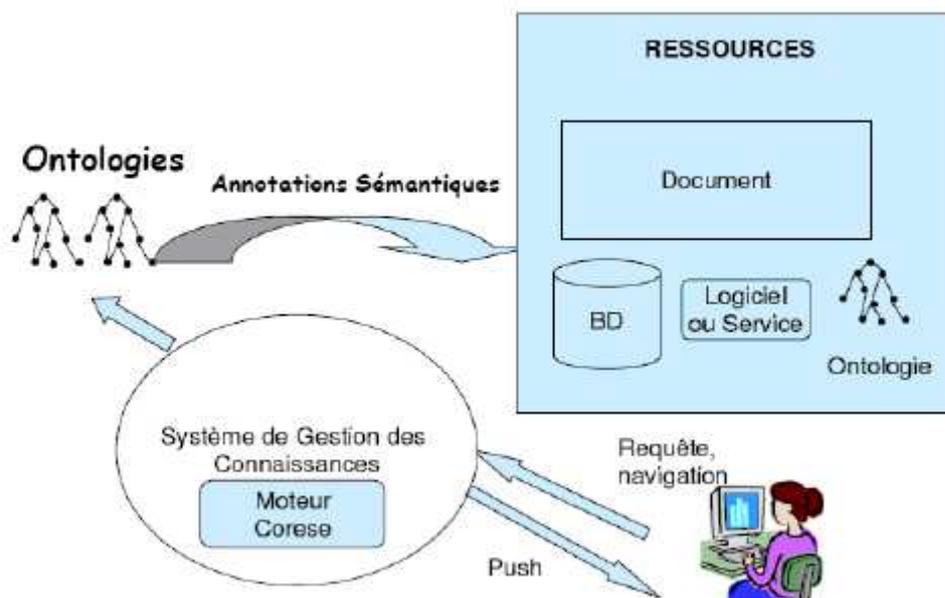


Figure 3.9. Architecture d'un web sémantique d'entreprise.

## 7. Conclusion

Dans ce chapitre, nous avons présenté les principales composantes du web sémantique à savoir les ontologies et les annotations. Sous cette optique les ontologies se présentent comme un pilier du web sémantique, car elles permettent de faire communiquer les hommes et les machines en utilisant la sémantique partagée par les différents acteurs du web et en décrivant ses ressources. Nous avons également, présenté quelques caractéristiques des annotations classiques puis nous nous sommes orientés vers l'annotation sémantique, pour enfin aboutir au web sémantique d'entreprise.

L'intégration d'un système de gestion des connaissances dans une entreprise est l'un des aspects les plus importants du Knowledge Management. En effet, un tel système offrant un accès global à l'ensemble des sources d'informations, peut jouer un rôle important de support pour le

transfert et le partage des connaissances dans l'entreprise. Au cours des dernières années, plusieurs systèmes ont été proposés pour gérer les informations et les connaissances de l'entreprise, comme la gestion électronique des documents, les bases de données, les intranets, etc. Ces systèmes, possédant chacun un modèle de données propre, ont posé des problèmes de coopération et d'intégration et présentent un coût non négligeable de développement et de maintenance. Actuellement, le web sémantique d'entreprise s'avère une solution de pointe.

De ce fait, Nous allons suivre pour la construction de la mémoire d'entreprise une approche issue de la gestion des connaissances permettant à l'entreprise de faciliter la préservation et l'utilisation de ses connaissances organisationnelles. Quant à la matérialisation, elle tend vers un Web sémantique d'entreprise (WSE) hétérogène et distribué.

Le prochain chapitre abordera les étapes choisies pour la conception d'une mémoire d'entreprise ainsi que les composantes de notre modèle de gestion dédié. Nous allons également, proposer une projection du modèle sur une mémoire de type projet. Nous avons basé l'étape de construction de la ME sur un cycle de développement d'ontologie de domaine. Le résultat sera une ontologie d'entreprise utilisée pour l'annotation.

## **CHAPITRE 4**

### ***Une approche de modélisation pour la gestion d'une mémoire d'entreprise***

### ***Un Modèle pour la Mémoire d'Entreprise***

*La connaissance est une affaire humaine; le knowledge management plonge donc ses racines dans les relations humaines; mais parallèlement il modifie profondément la contribution de ces ressources à l'économie*  
Jean Louis Ermine

#### **1. Introduction**

La gestion des connaissances désigne un ensemble de concepts et d'outils permettant aux membres d'une organisation de travailler ensemble de manière à permettre la production de connaissances et le développement des compétences individuelles, collectives et organisationnelles. Donc, Le KM tente de faire fonctionner ensemble, dans des processus cohérents, les connaissances critiques que sont les ressources essentielles de la production de biens et de services et celles qui sont issues de l'environnement économique et concurrentiel de plus en plus prégnant [Ermine, 96]. La gestion des connaissances est une problématique spécifique à chaque entreprise. Cette spécificité est due à plusieurs facteurs, tels que, sa culture et son organisation, son environnement technologique, son secteur d'activité, sa taille et son étendu géographique, etc. En revanche, l'approche de gestion des connaissances qu'il faudrait adopter, doit prendre en compte ces différentes caractéristiques.

Une approche de gestion des connaissances inclut une démarche, des processus et des outils dédiés à cette fin. Le problème majeur de la complexité se pose lors de la coordination entre ces éléments. La définition d'une telle approche repose sur l'étude des besoins de l'entreprise ainsi que l'analyse de l'existant. Mais quelque soit la démarche, son objectif majeur reste à identifier, à déterminer et à capitaliser les connaissances et les compétences utiles à l'exécution des activités opérationnelles de l'entreprise. Sous cette optique, afin d'intégrer les processus de gestion des connaissances dans les activités quotidiennes de l'entreprise, nous avons fixé pour notre démarche un ensemble d'objectifs, parmi eux, nous pouvons citer:

- Identifier les connaissances cruciales associées aux activités d'experts partant à la retraite ou d'équipes de spécialistes redéployées.
- Structurer le corpus d'informations et de documents en connaissances.
- Formaliser ces connaissances de manière à assurer leur accessibilité.
- Diffuser ces connaissances pour une meilleure exploitation.

Nous proposons au cours de ce chapitre, un ensemble d'étapes. Ces étapes sont à la base du cycle de vie du KM. Elles permettent la description de l'approche proposée. Ces étapes s'inscrivent dans un but pragmatique qui vise à fournir un modèle pour la mémoire d'entreprise, ce qui permet une opérationnalisation de la démarche. Pour ce faire, dans une première partie, nous présentons une description des différentes étapes de la démarche de gestion des connaissances dédiés à la Mémoire d'Entreprise. Dans une deuxième partie, nous introduisons un modèle de gestion d'une mémoire d'entreprise qui sera exploité ultérieurement, lors de l'annotation sémantique des ressources documentaires de cette mémoire. Cette modélisation doit permettre d'intégrer différents niveaux de représentation des connaissances relatives aux activités opérationnelles de l'entreprise; modélisation des processus métier, formalisation des savoir-faire d'expert, indexation et élaboration d'une terminologie consensuelle.

## 2. Etapes de modélisation

En prenant en compte la chaîne de valorisation des connaissances de Nonaka [Nonaka, 91], nous proposons de considérer la gestion d'une mémoire d'entreprise comme reposant sur les étapes suivantes (Figure 4.1). Nous rappelons que cette chaîne se compose de sept étapes : faire le bilan des connaissances existantes, déterminer les connaissances requises, développer de nouvelles connaissances, allouer les connaissances, les appliquer, les maintenir et retirer les connaissances obsolètes.

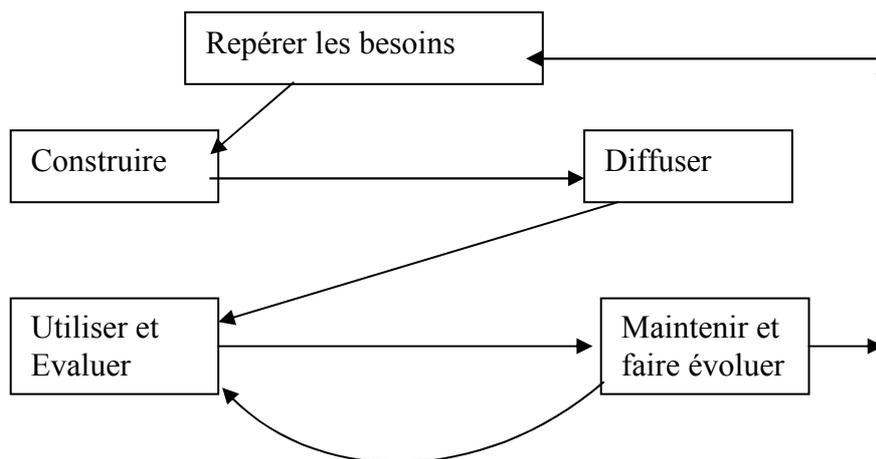


Figure 4.1 Etapes de Gestion d'une mémoire d'entreprise

## 2.1 Repérer les besoins

Cette étape consiste à détecter les besoins. C'est une sorte de diagnostic établi sur l'entreprise pour déterminer les besoins réels des utilisateurs et qui n'est pas une tâche simple. Les concepteurs de mémoire d'entreprise doivent apprendre autant que possible qui sont les utilisateurs, quelles sont les tâches à résoudre, dans quelles situations, quels types de connaissances ils doivent mémoriser pour réaliser leurs tâches, etc. Donc, ils doivent répondre à un certain nombre de questions relatives aux utilisateurs, aux tâches, aux situations, etc. Une première question se pose sur la nécessité d'une mémoire (l'entreprise a-t-elle réellement besoin d'une mémoire).

Nous avons utilisé quelques questions proposés par [Kuhn and Abecker 1997] telles que :

- quelles tâches doivent être assistées ?
- quelles sont les informations nécessaires pour réaliser ces tâches ?
- quel type d'assistance des utilisateurs ?
- quels sont les coûts et les bénéfices ?
- quels changements à prévoir dans l'avenir ?

Donc, dans notre cas, nous nous intéressons aux questions concernant les connaissances pour effectuer les tâches, les utilisateurs et l'environnement. [Buckingham, 1997] a identifié trois techniques : l'analyse des connaissances, les récits techniques entre membres et les discours centrés sur les documents. Nous avons essayé ces techniques au sein de l'entreprise de manière informelle. A un moment, nous nous sommes confrontés au besoin de définir le périmètre des connaissances à capitaliser. Ce qui nous permet de repérer les connaissances cruciales.

L'approche que nous avons adoptée pour identifier les connaissances cruciales se base sur l'analyse des processus d'activités de l'entreprise [Gueraich and Boufaïda, 2009]. La méthode de repérage des connaissances cruciales que nous appliquons s'inspire des principes de la méthode Gameth [Grundstein, 2002]. La méthode Gameth considère les processus sensibles de l'entreprise et les problèmes éventuels qui leur sont associés. Ce qui permet de modéliser les processus d'activités de l'entreprise. Et comme nous allons nous intéresser à la mémoire de projet par la suite, nous avons intégré la méthode DRCS (chapitre 3). En effet, cette méthode permet de représenter la prise de décision, les résultats d'un projet et sa gestion. Elle est jugée adéquate pour le domaine industriel.

Pour les utilisateurs, l'approche que nous avons suivie est l'approche « centrée sur l'utilisateur » [Eason and Olpherts, 1996]. Elle tend à définir la mémoire en fonction des

besoins des utilisateurs. Le mot « intéressé » renvoie à n'importe quelle personne appartenant à la communauté qui a un enjeu pouvant affecter par le système utilisé. Donc, les intéressés incluent les utilisateurs potentiels et d'autres (pas directement impliqués) comme des acheteurs, des clients, des agents, etc.,

Dans le cadre de la mémoire de projet, dans un projet d'entreprises les intéressés peuvent être [Dieng et al., 2005] :

- tous les membres de l'équipe
- les cadres dirigeants
- les autres membres de l'entreprise travaillant sur des projets similaires
- les membres de l'entreprise appartenant à d'autres services, comme les finances ou les achats
- les intéressés externes de l'organisation, tels que les clients, les investisseurs, les entreprises associées, les organismes de contrôle, les organismes de veillance située dans les pays étrangers, etc.

La technique que nous avons adoptée pour opérationnaliser cette approche est l'utilisation des scénarios. Il existe plusieurs variantes de la méthode des scénarios. L'usage des scénarios s'est développé avec les approches de conception qui impliquent directement les utilisateurs dans le processus de conception [Carroll, 1997]. Les scénarios sont des histoires problèmes ou exemple non résolus auxquels sont associés des solutions intuitivement possibles [Gruniger and Fox, 1995]. Les scénarios servent à motiver la conception de l'ontologie ou de l'application, à lui donner un sens.

## **2.2 Construire la mémoire d'entreprise**

La première étape de la construction d'une mémoire d'entreprise consiste à faire l'état des lieux [Dieng et al., 2005]. L'objectif de cet état est de préciser les membres de l'entreprise concernés par l'opération de capitalisation, les sources documentaires et les bases de données disponibles dans l'entreprise. Cet état des lieux doit permettre de répondre aux questions :

- ❖ Pour qui (quels sont les utilisateurs visés par la mémoire)?
- ❖ Pourquoi (quels sont les objectifs de la mémoire)?
- ❖ Quoi (quel sera le contenu de la mémoire) ?
- ❖ Comment (comment détecter les besoins, construire, utiliser, évaluer et faire évoluer la mémoire) ?
- ❖ Quand (dans quel contexte l'entreprise peut ou doit-elle se lancer dans une opération de capitalisation des connaissances.

Nous avons incorporé quatre composantes principales pour caractériser la mémoire à savoirs :

- le contexte (environnement)
- les documents
- l'ontologie
- l'annotation

Le contexte sera explicité dans la partie modèle, quant aux documents et annotation, nous allons les présenter dans le chapitre suivant. La section suivante est consacrée à la présentation de la composante ontologie.

Dons, dans cette partie du chapitre, nous allons décrire le processus que nous avons appliqué dans le développement de l'ontologie. Les grandes étapes de ce processus sont inspirées de la méthodologie de construction d'ontologie [Fernandez e al., 1997]. L'application de chacune des étapes de ce processus basée aussi sur l'exploitation d'autres travaux, tel que le processus proposé par [Hemmam 2005], et le guide de construction de [Natalya and Deborah, 2005] sera présentée en un ensemble d'étapes. La première étape est une étape de spécification de besoins dont le résultat est un document de spécification de l'ontologie. L'étape de la conceptualisation permet de modéliser le domaine sous forme d'un ensemble de concepts inter liés. L'étape de formalisation de l'ontologie en logique de description. L'étape de codification de l'ontologie en OWL-DL et son implémentation en protégée 3.1.1. L'étape de vérification et de validation de l'ontologie avec le raisonneur Racer.

### **2.2 .1 Spécification**

Le développement d'une ontologie commence par la définition du domaine et de la portée. Cela est basé sur la réponse de certaines questions : que est le domaine de l'ontologie que l'ontologie va couvrir ? à quoi cette ontologie va servir ? qui va utiliser l'ontologie etc. Les réponses à ces questions peuvent changer durant le processus de développement et elles permettent de délimiter la portée de l'ontologie qui consiste à définir une liste de questions. Ces questions peuvent servir à un test ultérieur (est ce que l'ontologie contient les informations suffisantes pour répondre à ces questions) mais elles ne doivent pas être exhaustives.

L'ontologie à construire peut être décrite à travers les aspects suivants :

- le domaine de connaissance
- l'objectif opérationnel
- les utilisateurs

- les scénarios d'usage de l'ontologie
- la portée
- le degré de formalisme
- le niveau de détail requis

Le résultat de cette phase est un document de spécification exprimé en langage naturel.

L'application de cette activité dans le cadre de notre travail, nous a conduit à répondre à certaines questions importantes sur le domaine, le but et la portée de l'ontologie. Le domaine que va couvrir l'ontologie est la gestion des documents de projets reliés à un domaine industriel projetée sur l'entreprise ENMTP. L'objectif opérationnel de celle-ci est la modélisation du contenu sémantique de ces documents en termes de leurs objectifs et termes à base d'annotation tout en assurant la modélisation des utilisateurs, leurs expériences avec leurs différents types d'invocations. Parmi les termes importants que va contenir l'ontologie, nous citons : projet, technique, organisation, problème, historique, etc.,

Quant aux scénarios d'usage les plus importants (les types de questions auxquelles l'ontologie est sensée fournir de réponses) nous citons :

- quels sont les documents de projets avec leurs différents types ?
- quels sont les contraintes techniques d'un projet particulier dans un document avec leurs différents types ?
- quels sont les utilisateurs potentiels ?
- quels sont les différents problèmes qu'on peut détecter ?
- quels sont les archives concernant un projet donné traité dans une période donnée ?

La description des principaux scénarios d'usage dépend surtout des utilisateurs futurs de l'ontologie. Dans notre cas il s'agit des acteurs invoqués dans un projet donné dans un secteur industriel. Quant à la maintenance de celle-ci elle reste des responsabilités du constructeur (expert). Nous avons une granularité fine correspondante à une ontologie très détaillée possédant ainsi un vocabulaire riche. Le degré de formalisation de celle-ci est sensé être complètement formel afin d'avoir une ontologie opérationnelle.

L'acquisition des connaissances faite durant la phase de spécification est basée sur le résultat des diagrammes uml et l'analyse des documents techniques. Le résultat de tout cela est un document technique qui porte la forme d'un élément d'expérience Rex (chapitre 3), illustré dans la figure suivante :

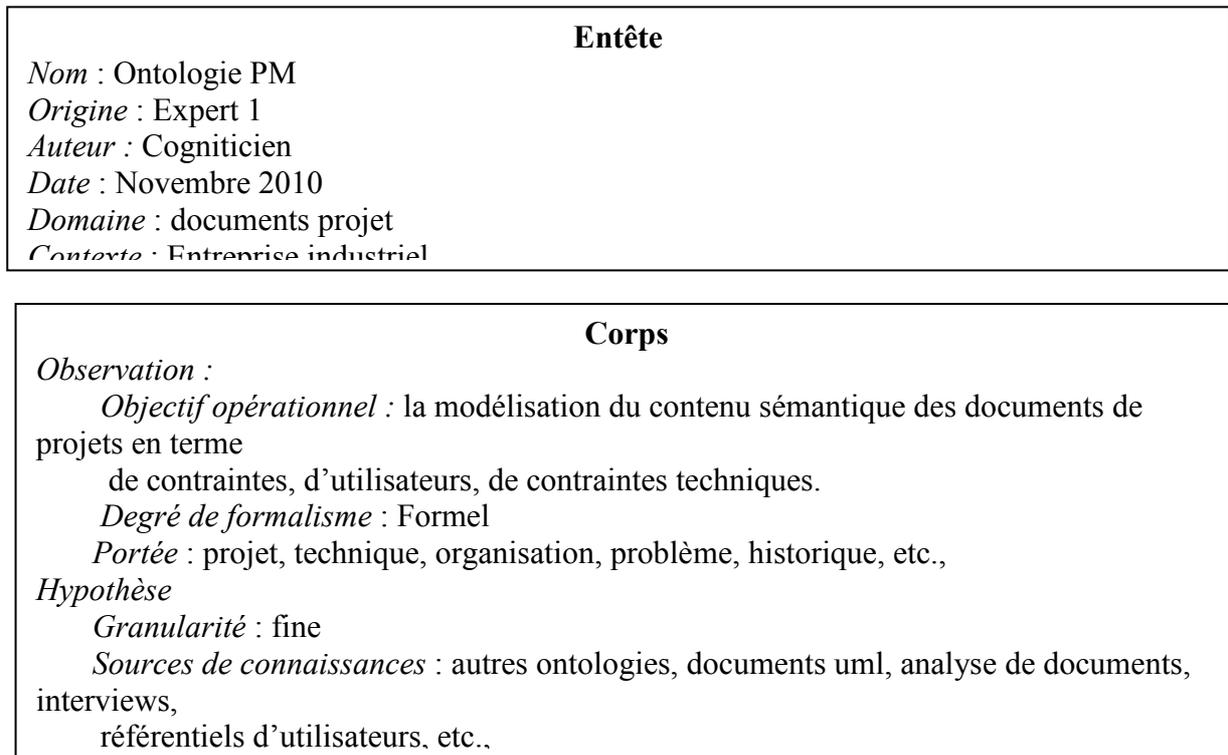


Figure 4.2 Document résultat de la spécification

### 2.2.2. Conceptualisation

Elle consiste à identifier et à structurer les connaissances du domaine à partir des sources d'information. Nous avons divisé cette étape en deux sous étapes : 1- activité du modèle conceptuel, 2- l'utilisation de représentation conceptuelle (diagrammes de classes uml). Cette idée est présentée en figure 4.3.

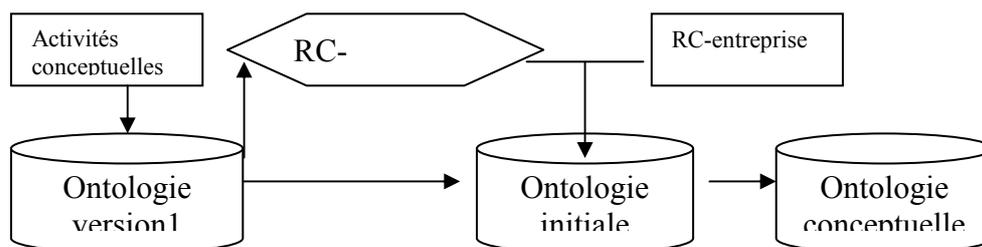


Figure 4.3 Etapes de conceptualisation

Donc, l'acquisition de ces connaissances peut s'appuyer à la fois sur l'analyse des diagrammes uml, l'analyse des documents et des interviews des experts du domaine. Une fois que les concepts sont identifiés par leurs termes, leur sémantique est décrite dans un langage semi-formel (tables et graphes) à travers leurs propriétés, leurs instances connues et les relations qui les lient entre eux. Cela peut se faire à travers la suite d'activités suivante dont le résultat est un modèle conceptuel :

- construction du glossaire de termes importants : pour chaque terme identifier sa description et sa référence sans s'inquiéter à propos de son type (concept, relation ou propriété)
- classification des concepts dans la hiérarchie des concepts : la hiérarchie qui est une sorte de taxonomie basée sur la relation « is-a », sa construction peut se faire selon une approche ascendante, descendante ou combinée. En cas d'apparition de nouveaux concepts, il faut les ajouter au glossaire sans oublier de s'assurer de l'absence de cycles dans ces hiérarchies.
- Représentation des relations entre concepts par un diagramme de relations binaires : la relation binaire lie un concept source à un concept cible qu'ils appartiennent ou non à la même hiérarchie. Les relations sont représentées par des arcs et les concepts par des rectangles. Si le nom d'une relation ne figure pas dans le glossaire des termes alors il faut le rajouter.
- Construction d'un dictionnaire de concepts pour chaque hiérarchie : identifier les concepts par leurs instances, leurs attributs ainsi que leurs concepts synonymes tout en garantissant une expressivité maximum de la connaissance attachée à chaque concept.
- Description des relations entre concepts dans une table de relations binaires : pour chaque relation qui apparaît dans le diagramme de relations binaires, décrire son nom, les noms de ses concepts sources et cibles, le nom de la relation inverse (si elle existe) et les cardinalités sources et cibles.
- Description des attributs des concepts dans une table d'attributs : pour chaque attribut identifié pour un concept, il faut définir le nom n de cet attribut, son concept et sa description, son type, sa cardinalité, sa valeur et son domaine de valeurs.
- Spécification des axiomes sur les concepts dans une table d'axiomes logiques : ce qui permet de définir certains concepts au moyen des expressions logiques.
- Description des instances des concepts dans une table d'instances : cela inclut toutes les instances du dictionnaire de concepts.
- Description des assertions dans une table des assertions : les champs de cette table sont le nom de la relation, le nom de l'instance source, la liste des instances cibles et la référence

L'application de l'activité de conceptualisation nous a permis la construction d'un ensemble de tables et de graphes formant le modèle conceptuel de l'ontologie de domaine à construire.

### **1. Glossaire des termes importants**

Il contient les termes importants du domaine, indifféremment de leurs types (concept, relation ou propriété) puisque leur nombre est grand et ils vont figurer dans d'autres tables, nous passons directement à la construction des hiérarchies de concepts.

## 2. Construction des hiérarchies de concepts

Un concept C1 est subsumé par un concept C2 si et seulement si toute instance de C1 est une instance de C2. La construction des hiérarchies de l'ontologie est basée essentiellement sur les diagrammes uml et les analyse de certains documents techniques ou l'analyse d'autres ontologies et des interviews avec des experts du domaine et les futurs utilisateurs.

Un concept universel « Thing » qui généralise tout les concepts racines des différentes hiérarchies de concepts peut être utilisé pour former une seule hiérarchie globale, afin d'éviter d'avoir des concepts isolés. La figure suivante représente les différentes hiérarchies de concepts existantes et la figure représente la hiérarchie globale de tous les concepts de l'ontologie. Celle-ci comporte 200 concepts. La stratégie de construction de cette ontologie est plutôt combinée. Les noms des concepts sont significatifs.

### - Etape de construction d'une ontologie à partir de représentations conceptuelles

La mémoire d'entreprise est considérée comme une partie intégrante d'un SGC (Système de Gestion de Connaissances). La construction d'une ontologie nécessite l'utilisation d'un processus permettant de passer des données brutes à l'ontologie. Si nous disposons de représentation conceptuelles (diagrammes uml) issues d'études effectuées au par avant dans l'entreprise. Nous considérons alors, que si notre ontologie sera construite à partir d'un ensemble de RC orientées objet traitant un domaine donné et le modèle issus de l'étape de conceptualisation peut être bénéfique d'un point de vue technique et sémantique [Gueraich and Boufaïda, 2011]. Nous supposons que les RC soient validées par leurs concepteurs ainsi que par des experts du domaine. Nous proposons les étapes suivantes :

- Utilisation de l'ontologie précédente (ontologie initiale).

- Alimentation de l'ontologie.

- Pour chaque RC considéré :

\* Extraction des concepts relatifs à RC.

\* Comparaison des concepts extraits avec ceux de l'ontologie initiale et détermination des relations sémantiques.

- Alimentation de l'ontologie avec les concepts de RC en respectant les liens sémantiques établis.

- Alimentation de l'ontologie avec les liens conceptuels de RC.

### - Construction de l'ontologie initiale

Pour la construction de l'ontologie initiale, nous optons à la prise en compte de la RC représentant l'ontologie et celles représentant le plus grand nombre de concepts pour un

domaine donné. Ce choix est justifié par le fait que cette RC contient le plus grand ensemble de classes, ainsi que leurs liens possibles, permettant ainsi de représenter un univers de discours (du domaine concerné) plus large que les autres RC. En plus cette RC, prise comme corpus, doit être validée par l'expert du domaine et bien documentée (dictionnaire de données et leur description). Cette ontologie (construite manuellement) contient un ensemble de concepts et ses liens conceptuels.

A ce stade, nous définissons un concept comme étant un terme caractérisé par une syntaxe et une sémantique représentées sous forme de chaînes de caractères. Il est caractérisé par une identité et contient un ensemble d'attributs et un ensemble d'opérations. Il peut être une classe ou bien une classe d'association (association porteuse de données).

La construction de l'ontologie initiale consiste à l'extraction des noms de classes, leurs propriétés (attributs et opérations) ainsi que leurs relations conceptuelles. Par relation conceptuelle nous sous-entendons tous les liens supportés par UML.

Afin d'illustrer notre démarche, on se basera dans la suite sur des RC relatives au domaine de conception de projet. L'entreprise veut lancer un appel d'offre pour réaliser un projet de construction de pièces de rechange.

Le diagramme de classe associé comporte les classes Usine et Fournisseur qui représentent respectivement l'usine qui fabrique les produits et le vendeur de la matière première. Chacune de ces deux classes possède une association avec la classe Transporteur, qui joue le rôle de courtier entre l'usine et le fournisseur. La classe Usine a une autre relation avec la classe Transporteur. Cette association admet une classe de lien `Cout_transport`. La classe `Cout_transport` concerne le cout de transport qui peut changer d'un moment à un autre. La classe Usine est en association avec la classe Produit. Un produit peut être une Pièce-légère ou Pièce-lourde. Cette relation est traduite par une relation de généralisation-spécialisation. La relation qui existe entre la classe Fournisseur et la classe Transporteur admet une classe d'association `Ordre`. Cet ordre concerne le paiement, représenté par la classe `Payment`, et la livraison représentée par la classe `Livraison`. Cette dernière peut être `Livraison_direct` ou `livraison Par_post`. Pour la classe `Payment`, il existe deux relations. La première est une relation d'agrégation avec la classe `Effet`. La deuxième est une relation de généralisation-spécialisation avec les classes `Carte` et `Cheque`. Les deux classes `Carte_crédit` et `Carte_débit` héritent de la classe `Carte`.

L'entreprise sera prise en compte pour la construction de l'ontologie initiale. Ainsi, le tableau suivant (table 1) représente une expression de la structure de notre ontologie initiale. Comme déjà mentionné, les relations conceptuelles qui peuvent exister entre les concepts. Ces relations peuvent être :



- L'alimentation d'une ontologie, consiste à l'ajout des nouveaux concepts à partir des RC introduites par les concepteurs. Elle contient les phases suivantes.

- Pour chaque RC : Extraction des concepts, leurs propriétés et leurs liens relatifs à la RC. Comparaison des concepts extraits avec ceux de l'ontologie initiale (Vi-1) : détermination des relations sémantiques. Alimentation de l'ontologie avec les concepts de RC en respectant les liens sémantiques établis.

- Détermination automatique des relations sémantiques entre les concepts de l'ontologie. Par relation sémantique, nous sous-entendons toute relation, autre que celles habituellement utilisées dans les RC UML (héritage, agrégation, ...) et qui est susceptible d'exister entre deux concepts. Ces relations peuvent être:

– Relation d'identité: c'est une relation sémantique qui existe entre deux concepts qui ont la même syntaxe, les attributs et les mêmes opérations. Exemple : Identité (Personne, Personne).

– Relation de Synonymie : c'est une relation sémantique qui existe entre deux concepts qui expriment le même sens. Exemple : Synonymie (Personne, Individu).

– Relation de classification sorte\_de, entre deux concepts exprimant que l'un est un cas particulier de l'autre. Exemple : Sorte\_de (Canari, Oiseaux).

– Relation d'Homonymie : un même concept peut avoir deux sens différents. Exemple : Homonymie (rose (couleur), rose (fleur)).

– Relation d'Equivalence : une relation sémantique qui existe entre deux concepts jouant le même rôle dans des RC différentes. Exemple : Equivalence (Usine, Fournisseur).

– Relation d'Antonymie : est utilisé entre deux concepts totalement disjoints. Exemple: Antonymie (Marié, célibataire).

L'ontologie conceptuelle relative à la mémoire d'entreprise est structurée en 5 sous ontologies.

- **Nature-ME Ontologie** qui spécifie la nature de la mémoire d'entreprise envisagée. C'est à dire le type de cette mémoire. Par exemple: mémoire de projet, de métier, d'expérience.... Elle contient les connaissances nécessaires pour un type particulier de mémoire,

- **Technique** Ontologie qui concerne les référentiels et les documents de travail essentiels pour l'élaboration d'une mémoire d'entreprise,

- **Problem** Ontologie qui concerne l'identification des typologies possibles de problèmes et les solutions envisageables,

- **Organizational** Ontologie qui concerne les aspects organisationnels tels que les acteurs, les taches, les methods de gestion, etc.
- **Historial** Ontologie qui correspond à l'archivage de connaissances jugées pertinentes à un moment donné.

La figure 4.4 illustre les sous ontologies de l'ontologie conceptuelle relative à la mémoire d'entreprise. La figure 4.5 est une partie de l'ontologie Project-memory (La nature de la ME est une mémoire de projet). Les aspects recensant les concepts, les relations et les attributs sont spécifiés dans l'annexe.

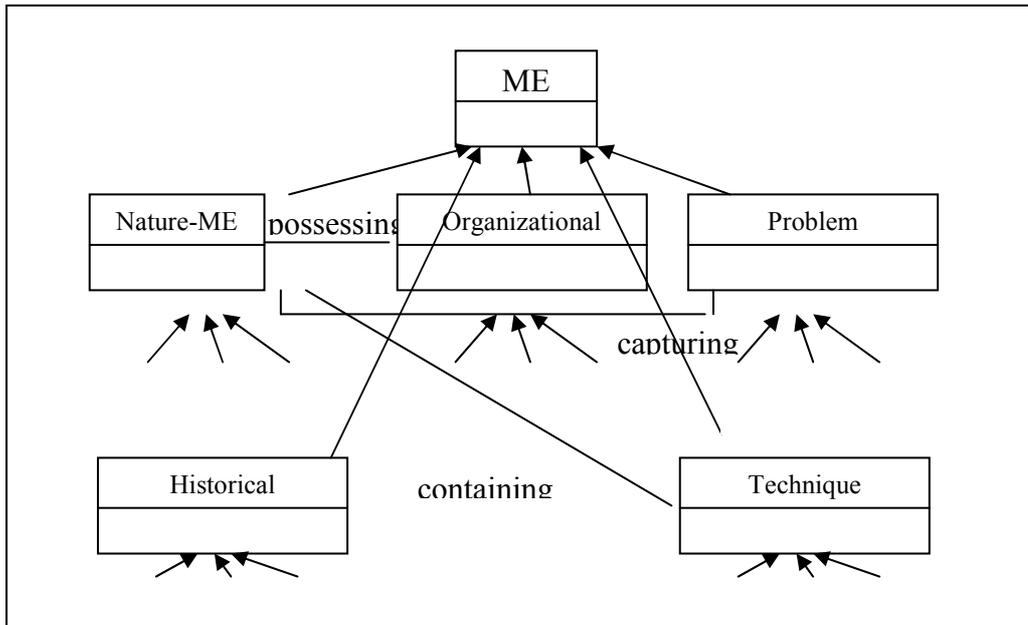


Figure 4.4 Ontologie ME

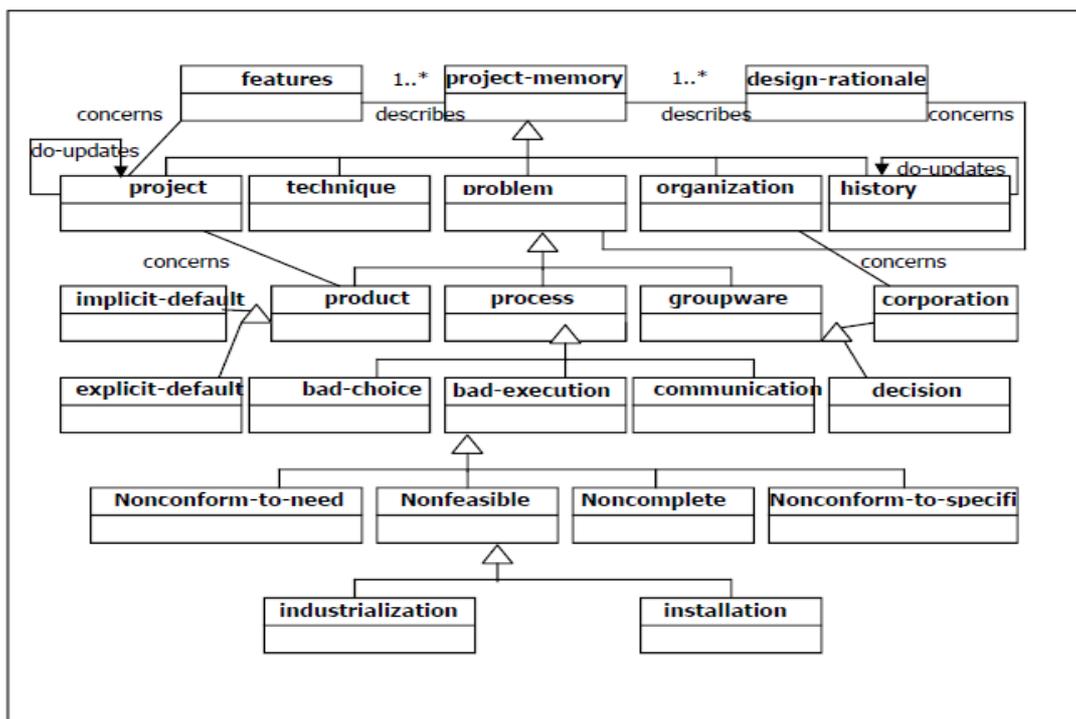


Figure 4.5 Ontologie Project Mémoire

### 2.2.3. Formalisation

Elle consiste à la représentation du modèle conceptuel obtenu par un formalisme de représentation des connaissances tels description [Nardi and Brachman, 2002]. La syntaxe est très expressive, généralement décidable et offre un certain nombre de constructeurs et d'axiomes, comme présenté dans la table 4.2. Pour exprimer le passage du modèle conceptuel au modèle formel, il suffit de prendre en considération:

- les concepts (leurs subsumants/ leurs subsumés)
- les rôles (domaine/codomaine)
- les individus
- les restrictions de cardinalité

La classification constitue un mécanisme d'inférence important des logiques de description. La classification est un processus permettant de déterminer la position d'un concept donné dans la hiérarchie de subsumption. Elle consiste à placer un concept dans la hiérarchie en spécifiant les concepts qui le subsument et les concepts qui sont subsumés par lui. La classification est mise en œuvre grâce à un programme spécialisé appelé le classifieur (Schmolze et al., 1983). Le classifieur prend une nouvelle description de concept et la place à l'endroit correct dans la hiérarchie.

Donc, le mécanisme de raisonnement est la classification de concepts. Elle est réalisée par un « classifieur » qui utilise la description d'un nouveau concept pour le placer à l'endroit correspondant de la hiérarchie. La recherche de la place correcte pour le nouveau concept comporte trois étapes :

- La recherche des subsumants les plus spécifiques SPS (concepts qui subsument le concept à classer et dont les fils ne le subsument pas).
- La recherche des subsumés les plus généraux SPG (concepts subsumés par le concept à classer et dont les pères ne sont pas subsumés par lui).
- Insertion du nouveau concept dans la hiérarchie.

Concepts		Axiomes de concepts (T_Box)	
Université/absurdité	$T/\perp$	Sous classes	$C \subseteq D$
Atomique	A,B	Equivalence	$C \equiv D$
Non	$\neg C$	<b>Axiomes assertionnels (R_Box)</b>	
Et	C.D	Sous role	$R \subseteq S$
Ou	C+D	Transitivite	Trans(S)
		<b>Axiomes assertionnels (A_Box)</b>	
Existe	$\exists R,C$	Instance	$C(a)$
Pour tous	$\forall R,C$	Role	$R(a,b)$
Au moins	$\geq n,R \{ \geq n,R \}$	Meme	$a=b$
Au plus	$\leq n,R,C \{ \leq n,R \}$	Différents	$a \neq b$
Enumération	$\{i1, \dots, in\}$		
<b>Roles</b>			
Atomique	R		
inverse	R-		

Table 4.2 Quelques opérateurs logique de description

La formalisation concerne aussi bien le niveau terminologique de l'ontologie. Ainsi que le niveau assertionnel. Elle comporte :

- Représentation de la partie terminologique (T\_box) : les concepts et les rôles sont exprimés de manière déclarative sous forme d'un ensemble de définitions (Figure 4.6).

ProjectMemory $\subseteq$ T $\cup$ ProjectMemory $\equiv$ Project $\cup$ Technique Organization $\cup$ Problem $\cup$ History Project $\cap$ Technique $\cap$ Organization $\cap$ Problem $\cap$ History $\equiv \perp$ Project $\subseteq$ ProjectMemory Project $\equiv$ LongProject $\cup$ ShortProject LongProject $\cap$ ShortProject $\equiv \perp$ $\geq 1$ concerns $\subseteq$ Project T $\subseteq \forall$ concerns (Product) Concerns $\equiv$ (concernedby-) .....
--

Figure 4.6 Exemple de la partie terminologique

- Représentation de la partie assertionnelle (A\_box) : cette partie est formée d'un ensemble d'axiomes décrivant les situations concrètes par rapport au domaine que l'on investit. Cela inclut l'instanciation des concepts et des rôles de l'ontologie.

### 3. Description du modèle

Le fondement théorique de notre modèle repose sur le modèle de la Marguerite de Jean-Louis Ermine [Ermine, 1996] (Figure 4.7).

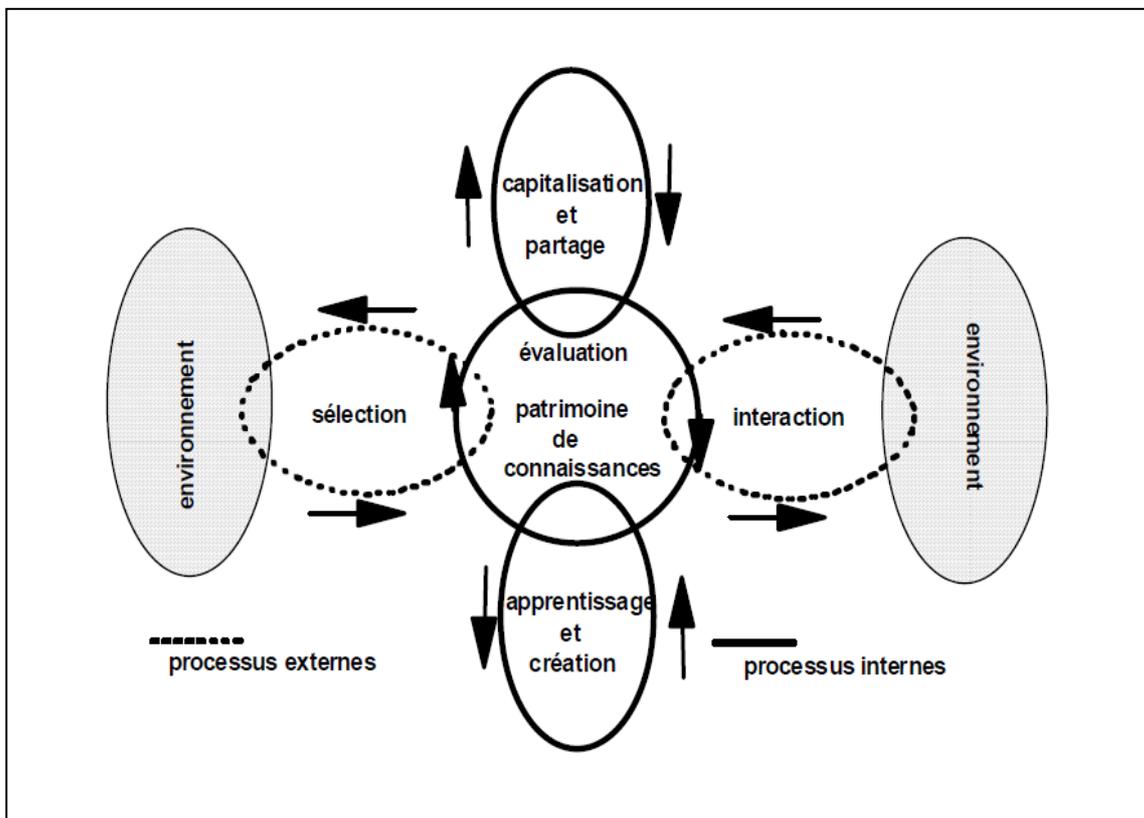


Figure 4.7 Modèle de la Marguerite originale

Dans le modèle de la marguerite initial, nous distinguons des processus internes comme la capitalisation et le partage, ou la créativité et l'apprentissage, et aussi externes, comme l'intelligence économique ou la veille. Les processus sont décrits en quatre grandes classes, qui correspondent aux " pétales " de la marguerite, et une classe qui correspond au cœur de ce modèle.

- Le processus de capitalisation et de partage des connaissances. C'est celui qui assure le partage (ou recyclage) de la ressource connaissance dans l'entreprise.
- Le processus d'interaction avec l'environnement. Comporte les connaissances qui se nourrissent des flux d'information qui proviennent de l'environnement de l'entreprise. Parmi les processus qui transforment ces flux d'information en capital de connaissances, on trouve le processus de veille ou d'intelligence économique ou stratégique.
- Le processus d'apprentissage et de création de connaissances, processus endogène et collectif qui est à la base de l'évolution des connaissances. Il inclut la problématique de l'organisation apprenante et de la créativité.

- Le processus de sélection par l'environnement, processus évolutionniste par excellence, de sélection des connaissances créées, par des critères de marché, d'acceptabilité..., à la fois économiques et socio-techniques. Il inclut des problématiques de marketing, de relation client...

Nous motivons notre choix d'adaptation du modèle de la marguerite est qu'il nous permet de guider l'analyse de la complexité du contexte. Cette complexité est due à son couplage entre le système de connaissances qui gère le patrimoine de connaissances et l'environnement interne et externe de l'organisation. Ce qui permet d'approcher dans une vision globale, les mécanismes d'adaptation et d'évolution entre eux. Cela permet de garantir le maintien intégral de l'organisation. Ce modèle est présenté par la figure 4.8.

Dans ce modèle, chaque pétale est un processus. Nous avons jugé qu'une approche centrée processus [Maurer and Holz, 1999] est intéressante. Car, en général, dans une entreprise, les membres des équipes travaillant sur des projets manipulent naturellement la notion de processus. Un processus est un regroupement cohérent d'activités, disposant de ressources, alimenté par des entrées, qui sont transformées en sorties en y apportant une valeur ajoutée pour le client. Le client peut être une personne ou groupe de personnes, interne ou externe à l'organisation. Il s'agit donc d'un ensemble ordonné d'actions ou activités, réalisées par des acteurs à l'aide des ressources ou moyens, destinées à produire un résultat. Ce résultat est orienté vers un bénéficiaire interne ou externe. Par exemple, la planification d'un projet se fait souvent autour de la notion de processus. De ce fait, nous allons bénéficier des caractéristiques d'un système centré processus. Du moment, où l'utilisateur peut à partir d'un processus trouver :

- Les méthodes décrivant comment réaliser ce processus pour atteindre ses objectifs.
- Les produits servant d'entrée et de sortie à ce processus.
- Les connaissances sur les qualifications nécessaires pour réaliser le processus.

La gestion des connaissances tente de faire fonctionner ensemble, dans des processus cohérents, les connaissances critiques que sont les ressources essentielles de la production de biens et de services et celles qui sont issues de l'environnement économique et concurrentiel de plus en plus prégnant. Donc, nous avons pensé à organiser ces processus selon un modèle.

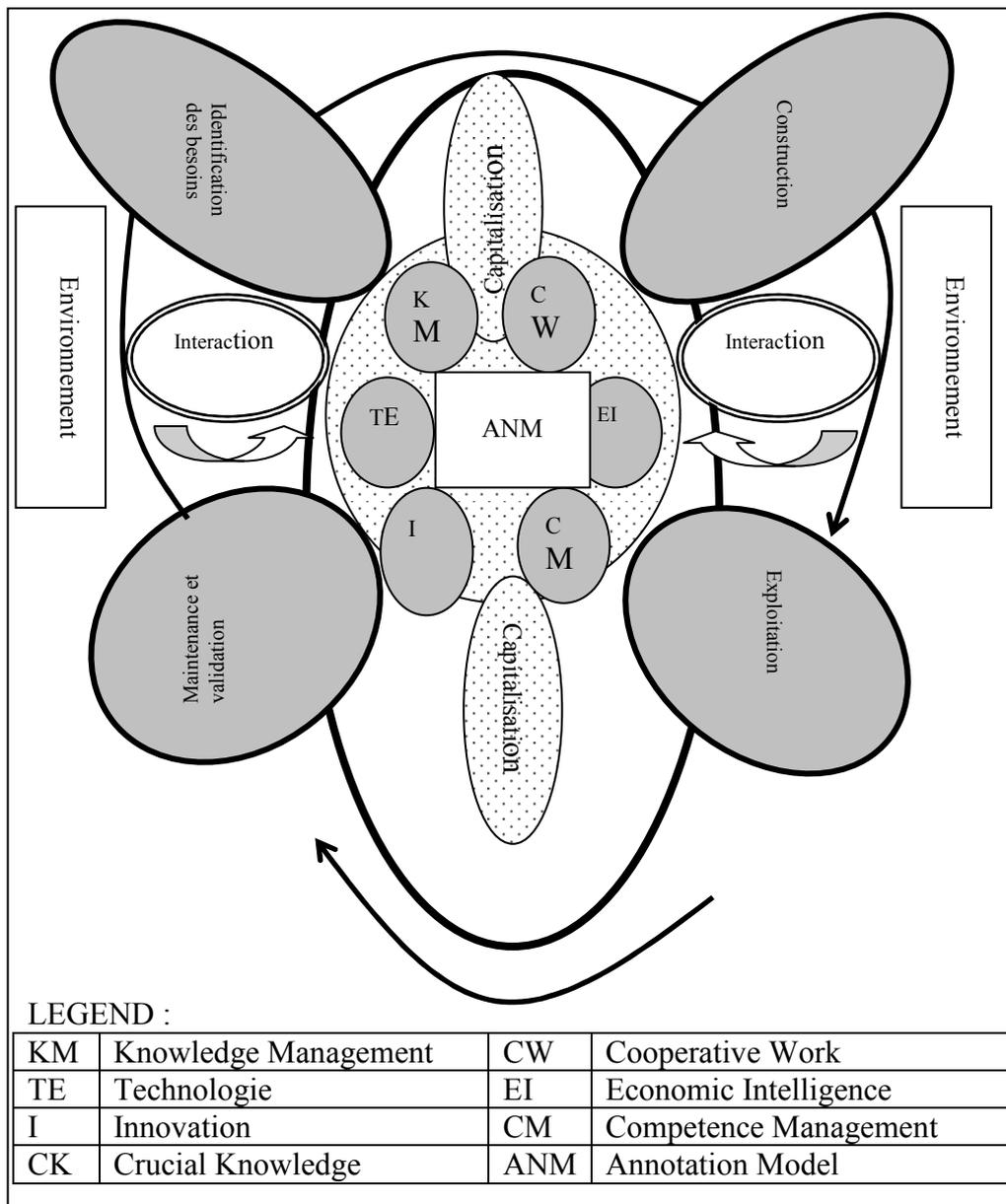


Figure 4.8 Modèle de la Marguerite étendu

Donc, nous avons modifié ce modèle pour deux raisons:

- La première consiste à l'adapter à la conception d'une mémoire d'entreprise qui sera annotée par la suite.
- La seconde est de doter la mémoire d'entreprise d'une description spéciale de ses composantes tout en considérant l'environnement externe et les composantes les plus internes (patrimoine de connaissances).
- Dans le modèle de la figure 4.8 nous avons proposé une classification des pétales en trois types plus le cœur composé de grains (l'ensemble prend la forme d'une fleur appelée la «Marguerite»).

- La première catégorie de pétales est composée des pétales du cycle de vie composé d'identification des besoins, construction, utilisation et entretien (que nous avons présenté au début du chapitre).
- La deuxième catégorie de pétales est constituée des pétales de capitalisation.
- La troisième catégorie est constituée des pétales d'interaction.

Le tableau 4.2 résumé l'essentiel du modèle.

<b>Life cycle petals: processes of constructing the original corporate memory</b>	
Needs	Identify users, tasks,
Identification petal	knowledge types they need to memorize and retrieve, ...
Needed Tools	EMK sheets+ UML diagrams
Building petal	Select the relevant documents + build Ontology
Tools	document Management + ontology +Information Extraction tools (IE)+UML diagrams
Use petal	Exploit and broadcast PM
Needed Tools	Semantic Web tools + Intranet/Internet technologies
Maintenance petal	Evaluate based on users' satisfaction.
Needed tools	Gain measurement techniques
<b>Capitalization petals: ensure the recycling of the shared knowledge in the enterprise</b>	
<b>Needed tools: capitalization project methods</b>	
<b>Interaction petals : ensure the interaction with the environment</b>	
<b>Needed tools: Interaction methods</b>	
<b>Heart and grains: set of the knowledge manipulated in the enterprise during its existence</b>	
KM	Knowledge Management
CW	Cooperative Work
TE	TEchnology
CM	Competence Management
EI	Economic Intelligence
CK	Crucial Knowledge
I	Innovation

Table 4.3. Composantes du modèle

### 3.1 Pétales d'interaction

Quand une entreprise doit se positionner dans son environnement, il est bon de remettre en cause ce point de vue purement topologique, et de considérer que l'entreprise et son environnement sont deux sous-systèmes distincts, mais en interaction forte, et que l'un des deux systèmes ne doit pas "dicter sa loi" à l'autre, ne serait ce que par une relation d'inclusion unilatérale. La pondération équivalente des interactions de l'environnement vers l'organisation et vice versa est un facteur fondamental de stabilité » [Toukara, 2002].

L'hypothèse de base du modèle de gestion des connaissances d'Ermine peut être expliquée à partir de la relation connaissance-environnement. A ce propos Toukara a dit « lorsque l'on cherche à positionner une organisation dans un contexte socio-économique, la vision qui s'impose est celle de l'entreprise qui est incluse dans son environnement. Cependant il ne s'agit que d'une vision spatiale, qu'on aurait tort d'utiliser comme référence.

Dans notre cas, pour construire et gérer une mémoire d'entreprise ( qui sera exploitée par la suite par les outils du web sémantique) il faudra étudier l'entreprise dans son contexte (que nous avons considéré comme composante d'une ME). En se basant sur la théorie de l'énaction de [Weick, 1979], dans l'identification de la relation connaissance-environnement, nous observons deux sortes de phénomènes, que nous appelons : processus d'interaction tacite et processus d'interaction explicite :

- le processus d'interaction tacite, représente le fait que la connaissance est incluse dans l'environnement. L'inclusion signifie que l'environnement est un seul domaine de connaissance, alors l'identification et la solution des problèmes doivent avoir lieu dans ce domaine de connaissance et pas dans un autre. En conséquence, le processus d'interaction par inclusion ne permet pas l'évolution de connaissances, mais la création de connaissances dans le même domaine. Et donc, il n'y a pas création d'autres domaines de connaissances ;

- le processus d'interaction explicite représente le fait que la connaissance n'est pas incluse dans l'environnement, autrement dit la connaissance à un domaine propre qui n'appartient pas à l'environnement. Dans ce contexte, l'explicite signifie que l'identification d'un problème est lié à son domaine de connaissance, mais sa solution peut se trouver dans l'environnement. La seule façon d'ouvrir un contact avec l'environnement se trouve à travers un "dialogue". Ce dialogue correspond à un processus d'interaction par parallélisme. Dans ce contexte, nous avons à la fois création et évolution aussi, bien pour le domaine de connaissance en question, que pour l'environnement.

Le processus d'interaction par parallélisme correspond au processus d'interaction (confrontation) chez Tounkara. Dans ce contexte il a dit « le processus de confrontation de l'entreprise à son environnement peut être modélisé en trois phases : la projection, le renseignement et la création de connaissances » [Tounkara, 2002].

Nous constatons donc, que le processus d'interaction est le fondement du processus de création et évolution de connaissances d'un modèle de gestion des connaissances. Ce qui renforce par exemple l'utilisation des ontologies (interagir pour arriver à un consensus) dans l'entreprise.

### **3.2 Pétales de capitalisation**

Afin de faire inclure dans le modèle une partie connaissances et une partie opérationnelle, la pétale capitalisation prend en charge ces deux parties. D'après Tounkara, l'organisation du modèle peut être décrite comme « l'ensemble des connaissances de l'organisation repose sur une "mémoire collective", un capital intangible que nous avons appelé le "patrimoine de connaissances". Ce patrimoine n'est pas réductible à des systèmes déjà existants comme le système d'information, de documentation, de ressources humaines (formation, compétences) etc. Le patrimoine de connaissances est un système spécifique, au même titre que le système cognitif chez l'être humain. Il a ses objectifs propres assignés par l'entreprise (Repérer, Préserver, Valoriser, Evoluer), son organisation et sa structure propre .

Nous constatons donc, deux traits caractéristiques du système de connaissances. L'un est que le système de connaissances est au cœur de l'entreprise. L'autre est que le système de connaissances fonctionne comme le système cognitif des individus. Cela signifie, que les mécanismes de création de connaissances et d'apprentissage individuel (mais aussi collectif) sont caractérisés par les quatre fonctions : (Repérer, Préserver, Valoriser, Evoluer).

A l'ensemble de ces processus, le processus liée au patrimoine de connaissances, permet de traiter plusieurs aspects tels que l'intelligence économique, la gestion des compétences, les technologies, le travail coopératif... Nous avons utilisé la méthode de l'annotation sémantique comme méthode pour évaluer le capital intangible dans l'entreprise. Cependant, il n'existe pas, à notre connaissance, une démarche complète permettant de valoriser, au sens large, un capital de connaissances.

### **3.4 Le fonctionnement du modèle**

Le fonctionnement de la structure du modèle est formé par l'interaction circulaire des trois processus autour du système de connaissances (ou patrimoine de connaissances), en suivant la démarche de la marguerite, à savoir : (1) le processus de gestion du cycle de vie. L'objectif de ce processus est de gérer les connaissances par le suivi des étapes de la gestion des

connaissances; (2) le processus de renseignement. L'objectif de ce processus est de préserver les connaissances par l'élaboration de corpus d'information, les outils privilégiés sont la capitalisation et le partage de connaissances ; et (3) le processus d'apprentissage et de création de connaissances. L'objectif de ce processus est de valoriser et évoluer les connaissances acquises avec l'environnement interne (interaction), et l'évolution du patrimoine de connaissances, les outils privilégiés sont le retour d'expérience et la créativité. Dans ce qui suit du chapitre, nous allons tester la validité de notre modèle sur un type particulier de mémoire d'entreprise qu'est la mémoire de projet.

#### **4. Adaptation du modèle pour une mémoire de projet**

Une mémoire de projet est une représentation de l'expérience acquise pendant la réalisation de projets [Dieng *et al.* 2004]. Elle peut être obtenue à travers un processus de capitalisation. Ce processus de capitalisation permet de définir cette mémoire puis de réutiliser ses connaissances précédemment modélisées et stockées, afin de réaliser de nouvelles activités.

La gestion de connaissances est un processus d'explicitation, de modélisation, de partage et d'appropriation des connaissances. Les principales étapes de ce processus sont le recueil, la construction, la diffusion et l'utilisation de la mémoire, et enfin l'évaluation et l'évolution de la mémoire. Plusieurs travaux se sont intéressés aux techniques qui concrétisent ces différentes étapes [Matta *et al.*, 2000]. Cependant, de nombreux problèmes persistent comme le recueil des connaissances pendant la phase d'évolution de la mémoire ainsi que la diffusion et la réutilisation des connaissances de la mémoire.

En littérature, ces méthodes sont classées en deux grandes catégories [Bekhti, 2004]: les méthodes de capitalisation des connaissances et les méthodes d'extraction directe. Les méthodes de capitalisation des connaissances utilisent essentiellement des techniques de l'ingénierie de connaissances. Ces techniques consistent principalement en un recueil (entretiens avec les experts ou recueil à partir des documents) et une modélisation des connaissances.

Les méthodes d'extraction des connaissances visent à extraire les connaissances directement de l'activité de l'organisation. Plusieurs techniques existent comme la fouille de données, la fouille de texte, les techniques de traçabilité et de la logique de conception.

Nous jugeons que notre modèle pour le cas d'une mémoire de projet peut bénéficier des deux catégories de méthodes. Le modèle de la marguerite avec ses pétales de cycle de vie exploite parfaitement la première catégorie. Pour la deuxième catégorie, nous allons utiliser la logique de conception. La caractéristique de cette technique en la dotant des outils du web sémantique. De telle manière, nous pouvons l'exploiter pour l'annotation sémantique. L'utilisation des méthodes d'entretiens classiques est restrictive, puisqu'elles demandent une

contrepartie non négligeable des acteurs d'un projet qui ne disposent généralement pas du temps nécessaire aux différents entretiens avec l'ingénieur des connaissances tout le long du processus. Il est donc nécessaire d'étudier d'autres techniques de traçabilité indirecte de l'activité des acteurs dans un projet. La figure 4.9 illustre le modèle adapté.

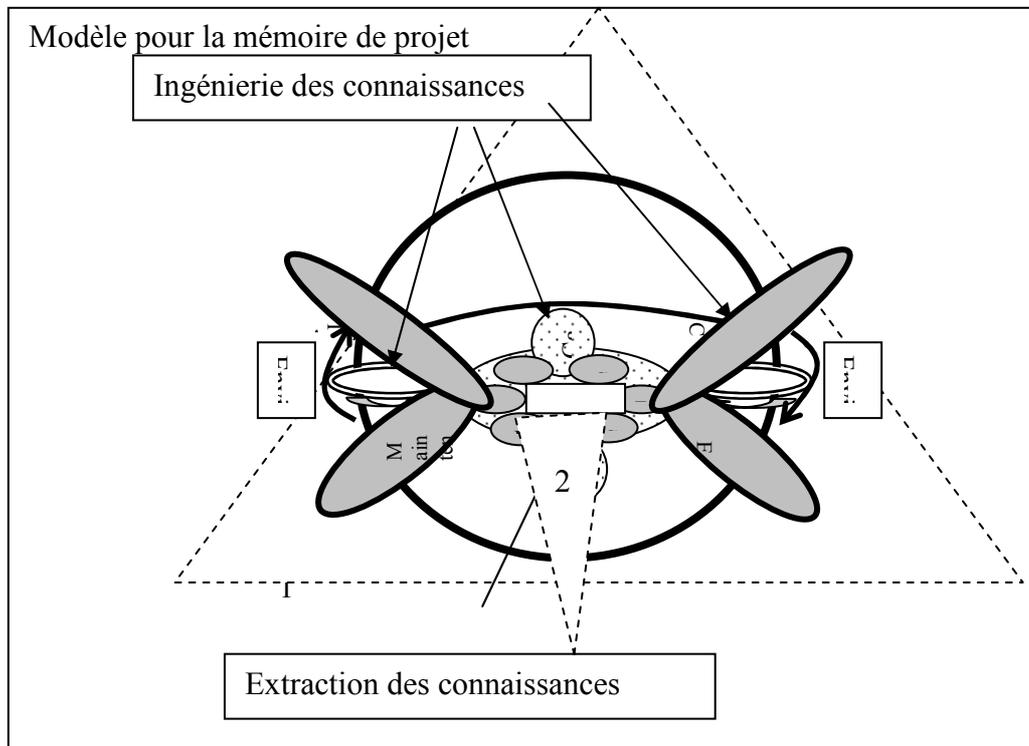


Figure 4.9 Modèle pour la mémoire de projet

Pour la traçabilité envisagée, les deux aspects à prendre en considération sont :

- Le recueil des connaissances qui constitue une étape importante dans la définition d'une mémoire de projet. Il s'agit d'acquérir les connaissances auprès d'experts et à partir des documents de l'entreprise.
- La réutilisation des connaissances qui permet de mettre la mémoire de projet à la disposition des différents acteurs de l'entreprise afin qu'ils y puisent des connaissances et les réutilisent pour résoudre de nouveaux problèmes. Il est alors possible d'exploiter les solutions apportées par les technologies internet via des connaissances structurées à l'aide du langage XML. Ce langage permet la structuration des données textuelles et la séparation des informations et des éléments de présentation. Donc, on peut présenter de plusieurs manières un même document XML à une communauté.

#### 4.1 Caractéristiques de la logique de conception

Une mémoire de projet doit contenir des éléments d'expérience provenant tant du contexte que de la résolution de problèmes dans un projet. Plusieurs méthodes ont été définies

pour représenter la logique de conception dans un projet. Ces méthodes peuvent être classées en deux principales catégories : représentation dirigée par la prise de décision [Buckingham, 1997] et représentation de la dynamique de résolution de problèmes.

- Dans la représentation dirigée par la prise de décision, l'espace de conception est représenté par des choix de conception. Ces choix sont structurés comme réponse aux questions évoquées par les problèmes de conception. Des arguments peuvent justifier les choix d'une option suivant un critère donné. Les options génèrent d'autres questions auxquelles les concepteurs répondent par des options. IBIS, DRAMA et QOC sont des exemples de ce type de méthodes.

- Dans la représentation de la dynamique de résolution de problèmes, la logique de conception est représentée sous forme d'une analyse cognitive d'une résolution de problème. Certains éléments du contexte comme l'organisation de l'activité, le rôle des acteurs ainsi que l'artefact sont représentés. DRCS et DIPA sont des exemples.

Nous proposons dans notre approche de logique de conception de décrire le contexte d'un projet en mettant en avant son influence sur la résolution de problèmes via l'exploitation de l'ontologie. Ceci va permettre d'une part d'identifier des éléments comme les compétences, les ressources, les contraintes, les modes de communication, etc. D'autre part, on va structurer dans l'ontologie une trace pour une prise de décision d'un projet de façon à ce qu'elle soit facilement réutilisable.

Communément avec [Bekhti, 2003], nous considérons que la traçabilité doit être assistée et guidée mais contrairement, nous considérons que le processus de traçabilité est constitué d'une seule étape qui englobe une retranscription directe structurée sous la forme d'une annotation. Au départ cette annotation est manuelle puis elle sera élaborée par un processus d'annotation sémantique.

## **4.2 Représentation de la mémoire de projet**

Dans notre travail, le contexte se base sur le modèle de la figure 4.6. Les éléments du contexte doivent être dans une mémoire de projet et les éléments de la logique de conception. Nous présentons alors cette mémoire en trois grandes parties, contexte, logique de conception et ontologie (construite selon la méthode proposée au début du chapitre) qui assure des influences mutuelles entre le contexte et la logique de conception et entre les éléments de la logique de conception elle-même (Figure 4.10).

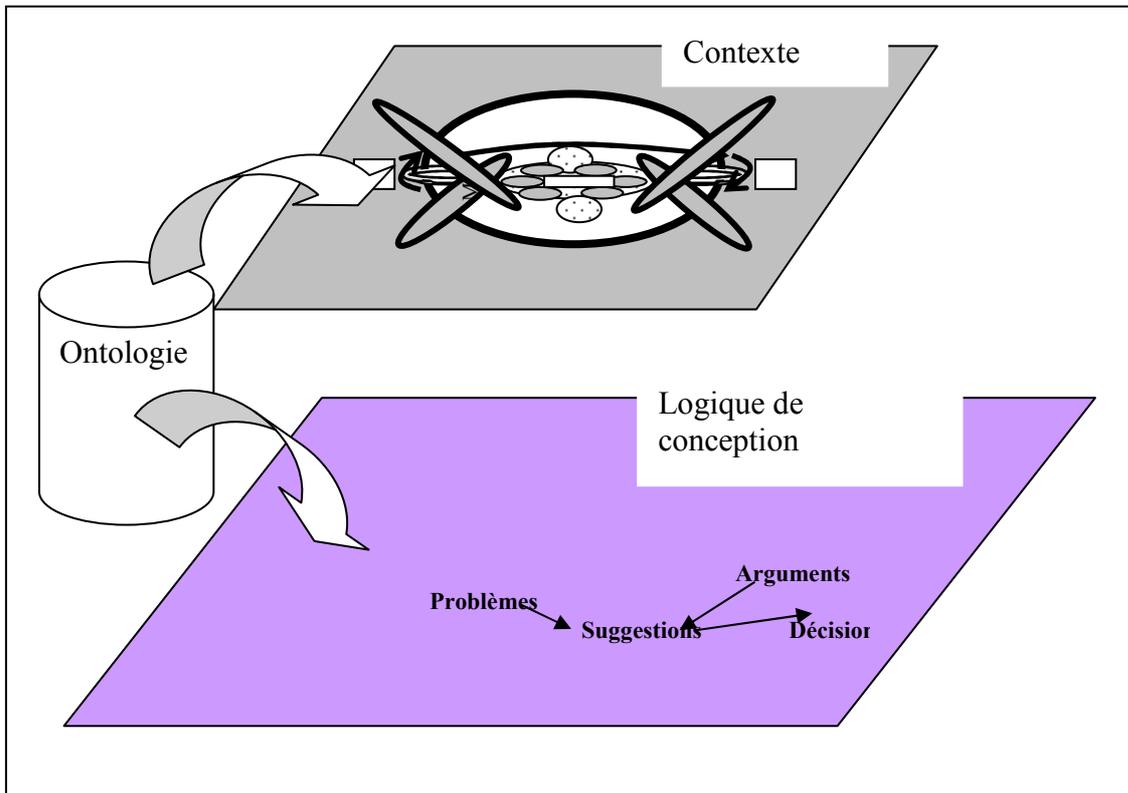


Figure 4.10 Composantes de la mémoire de projet

#### 4.2.1. Contexte

Nous représentons le contexte d'un projet comme une description du modèle avec ses différentes pétales, de l'environnement du travail (les moyens et techniques, les référentiels, les directives ainsi que les contraintes du projet) et de l'organisation du projet (les participants, leurs rôles et l'organisation des activités). Dans l'ontologie, nous pouvons le présenter comme illustré dans la figure 4.11.

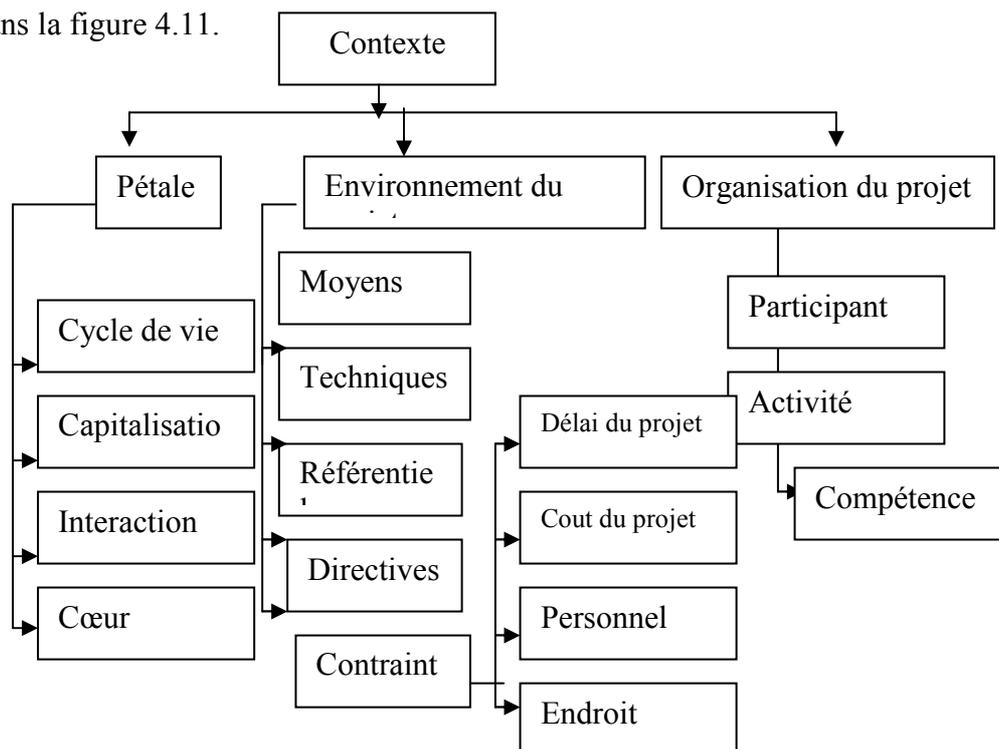


Figure 4.11 Contexte

#### 4.2.2. Logique de conception

La logique de conception représente l'espace de décision dans un projet. Nous proposons de décrire cet espace par le scénario suivant : le problème est réparti selon la vision de chacun des acteurs. Chaque vision est discutée par un acteur qui donne son opinion appuyées sur des arguments de plusieurs types. Après les acteurs proposent des suggestions concernant l'élément du problème. L'ensemble des arguments et suggestions permet au groupe de prendre une décision concernant cet élément du problème. L'élément du problème est donc résolu, sinon il sera rediscuté de la même manière et passera par le même cycle. Chaque cycle est une version de la solution.

Ceci constitue une évolution de la solution. La version finale est atteinte lorsqu'il y a un minimum de consensus. On ne s'intéresse pas à la formation des coalitions ou à la gestion des conflits entre les acteurs. La figure 4.12 montre ce scénario par un diagramme d'activité uml.

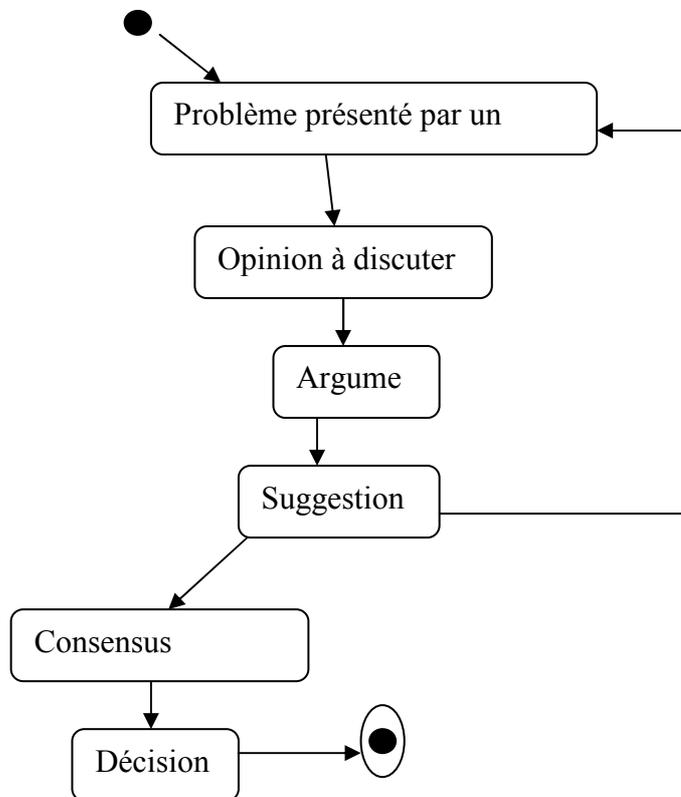


Figure 4.12 Scénario de la logique de conception

En pratique, nous allons enrichir les documents sous la forme d'éléments d'expériences REX (présentés dans la section ontologie) et plus particulièrement la partie corps par les éléments du scénario de la logique de conception à savoir :

- Le problème : le problème global discuté au cours des réunions est constitué d'un ensemble de sous-problèmes selon chaque participant.
- Arguments : elle présente la cause de la discussion du problème.

- Suggestions : représentent les solutions partiels relatives à chaque sous problème selon les arguments proposés.
- Consensus : chaque suggestion validée est un consensus.
- Décision : l'action à entreprendre face au problème initialement rencontré. C'est une façon de regrouper les consensus des acteurs.

### 4.2.3. Ontologie

Nous considérons que l'ontologie relative à une mémoire de projet est une sous ontologie de l'ontologie globale relative à une mémoire d'entreprise. Par exemple pour la partie conceptuelle liée aux problèmes, nous nous sommes inspirés des typologies de problème fournis par [Matta et al., 99] en plus des problème dégagés par notre expérience au sein d'une entreprise algérienne. Donc, Cette typologie a été validée avec les partenaires industriels. Nous notons quatre catégories de problèmes: Les problèmes concernant le produit, le projet, le processus de conception et ceux émanant du contexte. Le tableau 4.3 illustre le principe de cette classification.

Type des problèmes	Raffinements		
Problème lies aus produits	Défauts explicites	non faisabilité	installation industrialisation utilisation de catalogues périmés
		Non conforme aux besoins	specification
		incomplétude	exigences reglement specification
		Non conforme aux spécif des réglements	caractéristiques du produit
	Défauts implicites	incomplétude	reglement effets non considérés interaction non prévue comportement non prévu
		incompréhension	besoins environnement
		non conforme à la qualité de sécurité	produit
		mauvaise interaction	environnement produit
		mauvaise utilisation	fonctionnalités fonctions mal comprises interface inadéquate

Processus	Choix inappropriés	Méthodes inappropriées Méthode non utilisées	
	incomplétude	Limitation de techniques Activité incomplète Activité non accomplie Activité difficile à évaluer	
	Mauvaise exécution	Données erronées Techniques erronés	
projet	Mauvais délai	Mauvaise prévision	
	Mauvais cout	Mauvaise estimation	
	Mauvais acteurs	Pas de collaboration Pas de coopération Pas de négociation	
contexte	Mauvaise organisation	Mauvaises taches Mauvais groupes	
	Problémé relationnels	Mauvaises ralations internes Mauvaises relations entre équipes	Problèmes sociaux Erreurs de copréhension Erreurs de transmissions Mauvaises interprétations

Table 4.4. Typologies des problèmes

### 4.3. Modélisation de la mémoire de projet pour une entreprise

Les utilisateurs de la mémoire de projet souhaitent comprendre la logique de conception et le contexte des projets archivés. Pour ce faire ils vont essayer de comprendre ces éléments, leur contexte et leur logique de conception afin de les réutiliser dans d'autres projets. Un des objectifs de la construction et d'exploitation d'une mémoire de projet est sa réutilisation. L'entreprise peut ainsi gagner en productivité, mais également partager le savoir-faire de conception entre les acteurs dans la mesure où la mémoire est partagée. Pour atteindre l'objectif de réutilisation, l'accès aux éléments archivés doit être facile. Les utilisateurs doivent atteindre n'importe quel élément dans la mémoire de projet sans être obligés de parcourir toute la structure de la mémoire. Lors de la définition de la mémoire de projet, la structure de ces éléments doit être souple et facile à utiliser. Cette structure doit tenir compte de tous les éléments de la mémoire sans omettre aucune information utile dans la compréhension du déroulement du projet.

Dans une entreprise, on parle souvent d'un système d'information. L'idée est de proposer une modélisation pratique de la mémoire de projet et qui peut être vue comme un système d'information particulier. De ce fait, nous allons utiliser dans cette section des modèles développés en utilisant le formalisme UML [OMG, 2005]. Nous présentons un diagramme de cas d'utilisation pour les activités d'une mémoire de projet ainsi que des diagrammes de classes associés. Nous nous sommes intéressés au projet de conception, car ils sont les plus caractéristiques des entreprises industrielles.

UML est un langage qui permet de représenter des modèles, mais il ne définit pas le processus d'élaboration des modèles. Cependant dans le cadre de modélisation de l'application de notre projet, nous avons suivi une démarche :

- Itérative et incrémental.
- Guidée par les besoins de l'utilisateur du système.
- Centré sur l'architecture logicielle.
- Piloté par les risques.

Ces points sont les principales caractéristiques du Processus Unifié (UP). Ce processus de développement de logiciel est basé sur la notation d'UML. D'après les auteurs d'UML, un processus de développement qui possède ces qualités devrait favoriser la réussite d'un projet.

#### 4.3.1. Modèle des cas d'utilisation

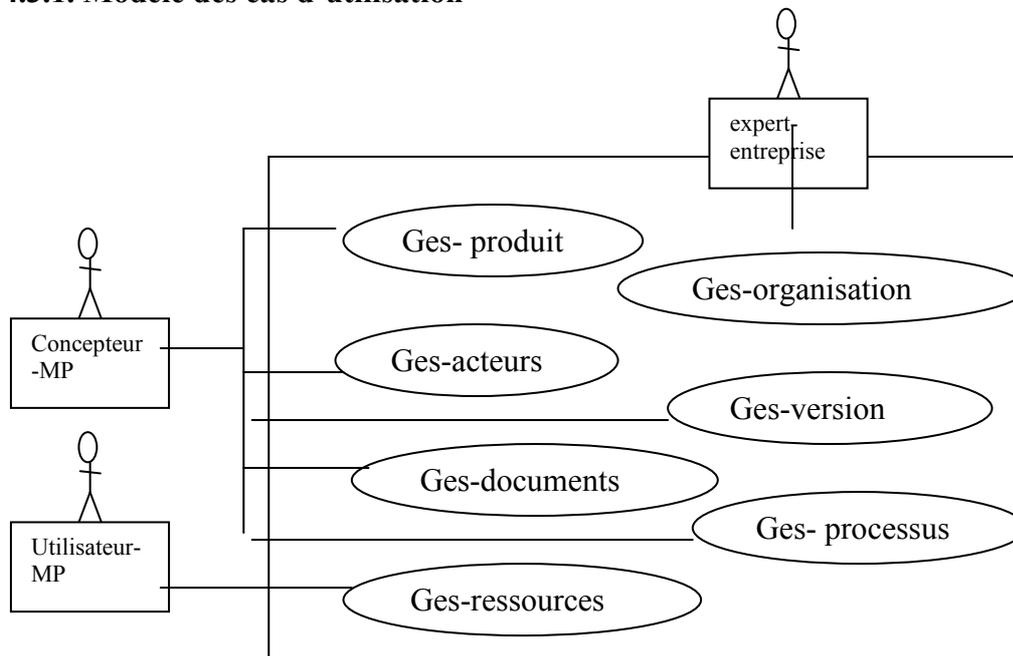


Figure 4.13 Diagramme de cas utilisation associé à la gestion des tâches d'une MP

#### - Les fiches Descriptives et les diagrammes d'activités

Les acteurs concernés sont :

- acteur concepteur: il s'occupe de définir les processus , les produits de la mémoire de projet, définir les évolutions et les gestions de produit, de limiter les documents

utilisés/générés.

- acteur utilisateur: c'est l'exploitant de la mémoire de projet. Il peut être interne ou externe (client).
- acteur responsable d'organisation : il définit la structure d'organisation il travaille en collaboration avec le concepteur. Il contribue par son savoir-faire en tant qu'agent de terrain, à décrire les problèmes rencontrés et à donner ses propositions et solutions de la mémoire de projet.

#### 4.3.1.1 Gestion des produits

<b>Nom</b>	<b>Gestion des produits</b>
<b>Acteur</b>	Concepteur-MP
<b>Type</b>	Primaire
<b>Description</b>	Permet la gestion des produits et leurs compositions en éléments de produits
<b>Actions</b>	<u>Au début</u> : Lire l'identifiant du produit <u>En cours</u> : vérifier la disponibilité du produit, s'il est disponible gérer produit sinon annuler. <u>En fin</u> : enregistrer le produit

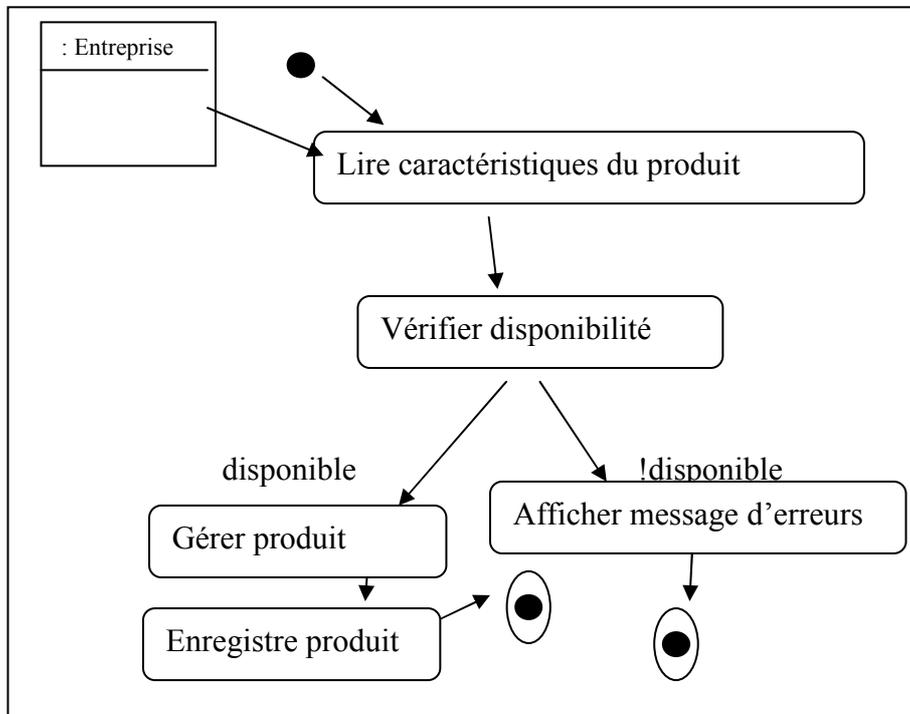


Figure 4.14 Diagramme d'activités associé à la gestion de produits

### 4.3.1.2 Gestion acteurs

<b>Nom</b>	<b>Gestion des acteurs</b>
<b>Acteur</b>	Concepteur-MP
<b>Type</b>	Primaire
<b>Description</b>	Permet la gestion de l'ensemble des intervenants dans le projet
<b>Actions</b>	<p><u>Au début</u> : Lire l'identifiant d'un acteur</p> <p><u>En cours</u> : vérifier l'affiliation d'un acteur à une équipe de projet de l'acteur, s'il est affilié alors gérer son rôle et identifier ses compétences sinon annuler.</p> <p><u>En fin</u> : enregistrer l'acteur</p>

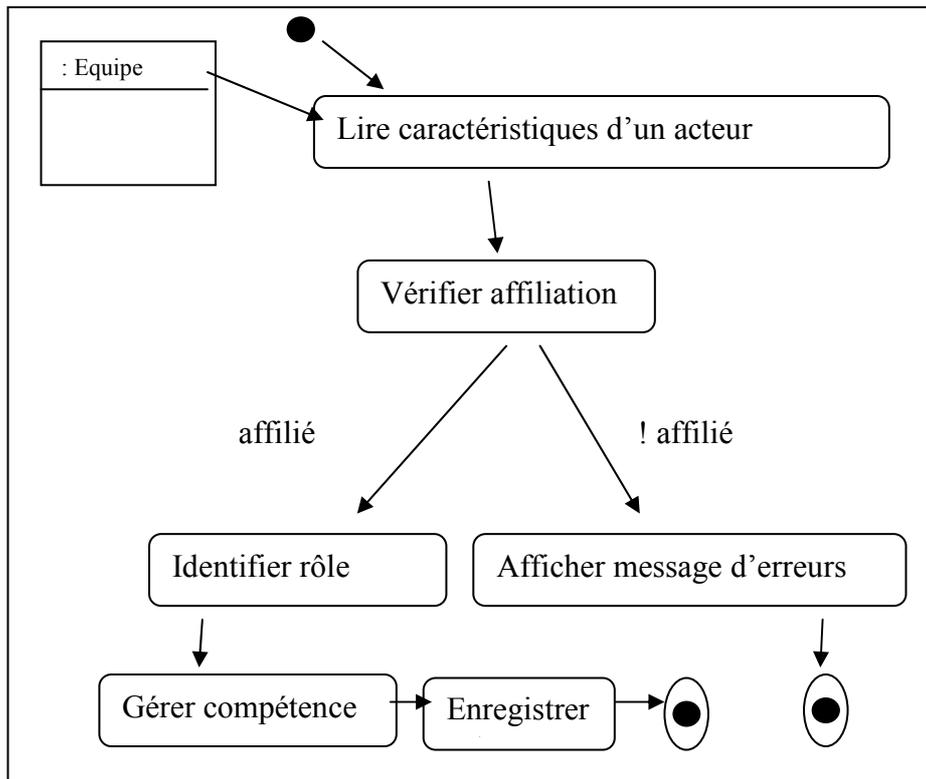


Figure 4.15 Diagramme d'activités associé à la gestion des acteurs

### 4.3.1.3 Gestion des documents

<b>Nom</b>	Gestion des documents
<b>Acteur</b>	Concepteur-MP
<b>Type</b>	Primaire
<b>Description</b>	Permet de gérer des documents utiles et des documents qui sont générés
<b>Actions</b>	<p><u>Au début</u> : Lire l'identifiant du document</p> <p><u>En cours</u> : vérifier la disponibilité du document, s'il est disponible gérer le document et générer autres sinon annuler.</p> <p><u>En fin</u> : enregistrer le document</p>

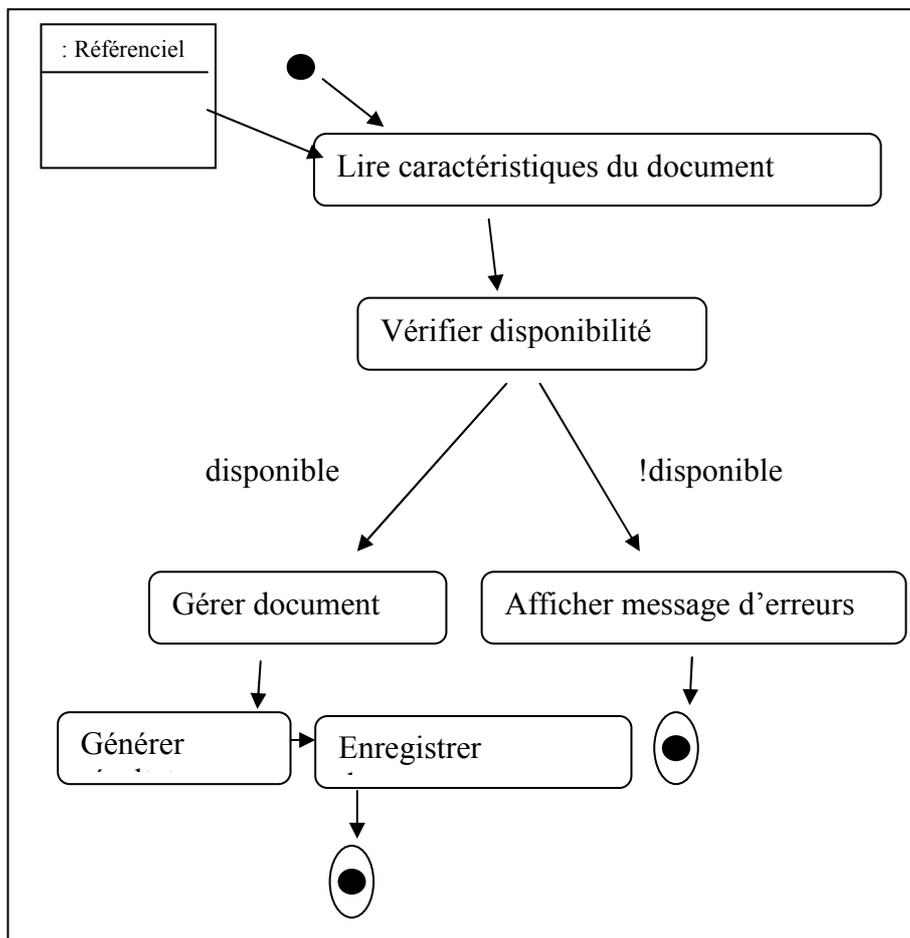


Figure 4.16 Diagramme d'activités associé à la gestion de documents

### 4.3.1.4 Gestion des processus

<b>Nom</b>	Gestion des processus
<b>Acteur</b>	Concepteur-MP
<b>Type</b>	Primaire
<b>Description</b>	Permet de gérer l'ensemble des processus et les enchaînements d'activités et gérer les justifications qui conduisent aux choix de décision lors des étapes de conception. Les éléments de justification peuvent être du type document, prototype, croquis, maquette, etc.
<b>Actions</b>	<p><u>Au début</u> : Lire la désignation du processus</p> <p><u>En cours</u> : vérifier l'existence du processus, s'il existe le gérer sinon annuler.</p> <p><u>En fin</u> : enregistrer le processus</p>

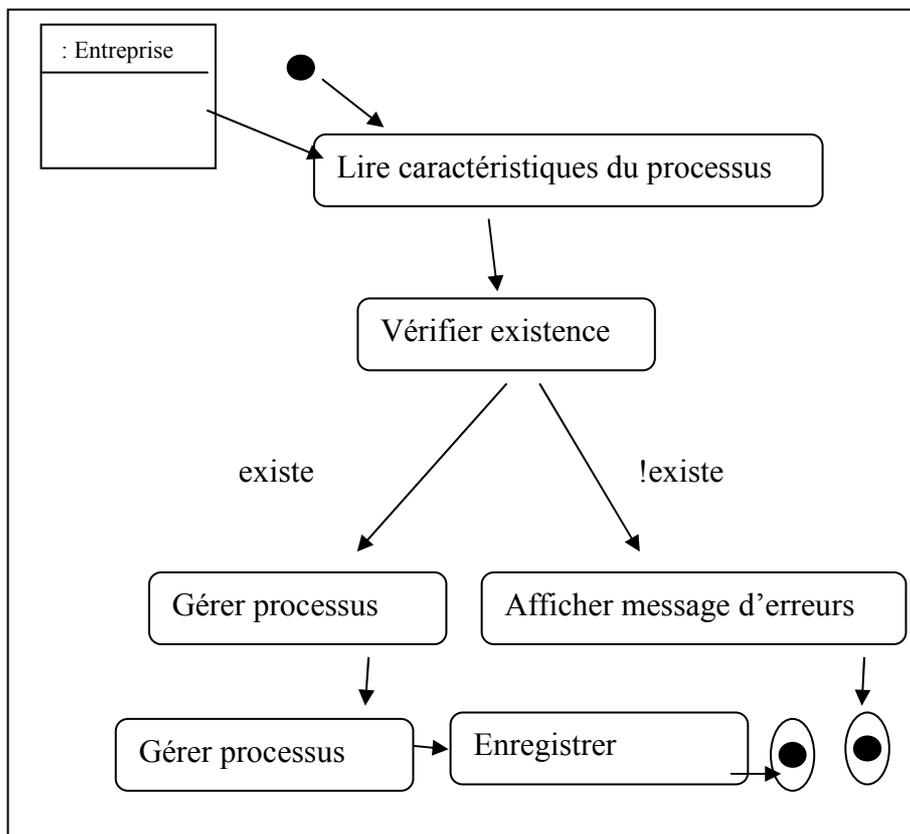


Figure 4.17 Diagramme d'activités associé à la gestion de processus

### 4.3.1.5 Gestion des ressources

<b>Nom</b>	Gestion des ressources
<b>Acteur</b>	Utilisateur-MP
<b>Type</b>	Primaire
<b>Description</b>	Permet la gestion des moyens mis en œuvre pour garantir en permanence à l'entreprise une adéquation entre ses ressources et ses besoins en personnel, sur le plan quantitatif comme sur le plan qualitatif
<b>Actions</b>	<p><u>Au début</u> : Lire l'identifiant de la ressource</p> <p><u>En cours</u> : vérifier la disponibilité de la ressource, si elle est disponible la gérer sinon annuler.</p> <p><u>En fin</u> : enregistrer la ressource</p>

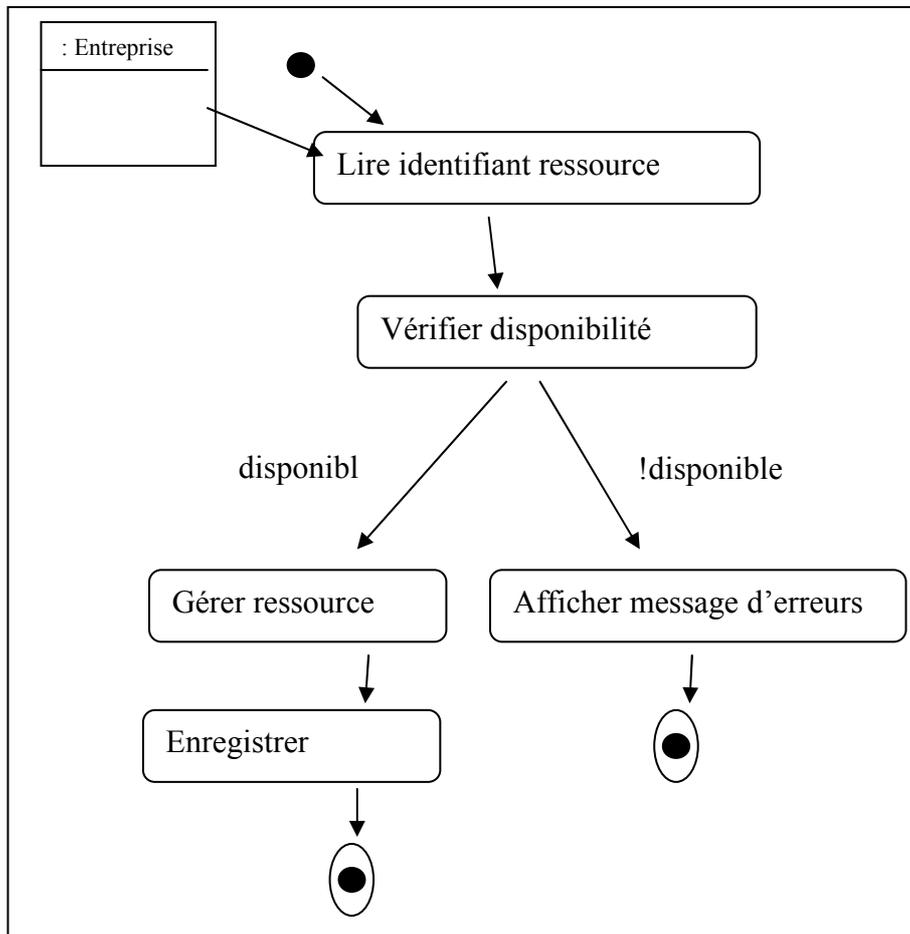


Figure 4.18 Diagramme d'activités associé à la gestion des ressources

### 4.3.1.6 Gestion de l'organisation

<b>Nom</b>	Gestion de l'organisation
<b>Acteur</b>	expert-entreprise
<b>Type</b>	Primaire
<b>Description</b>	gérer l'ensemble des organisations de l'entreprise et leurs décompositions en éléments d'organisation
<b>Actions</b>	<p><u>Au début</u> : Lire nature-organisation</p> <p><u>En cours</u> : vérifier la structure de l'organisation, s elle est structurée alors gérer organisation sinon annuler.</p> <p><u>En fin</u> : enregistrer l'organisation</p>

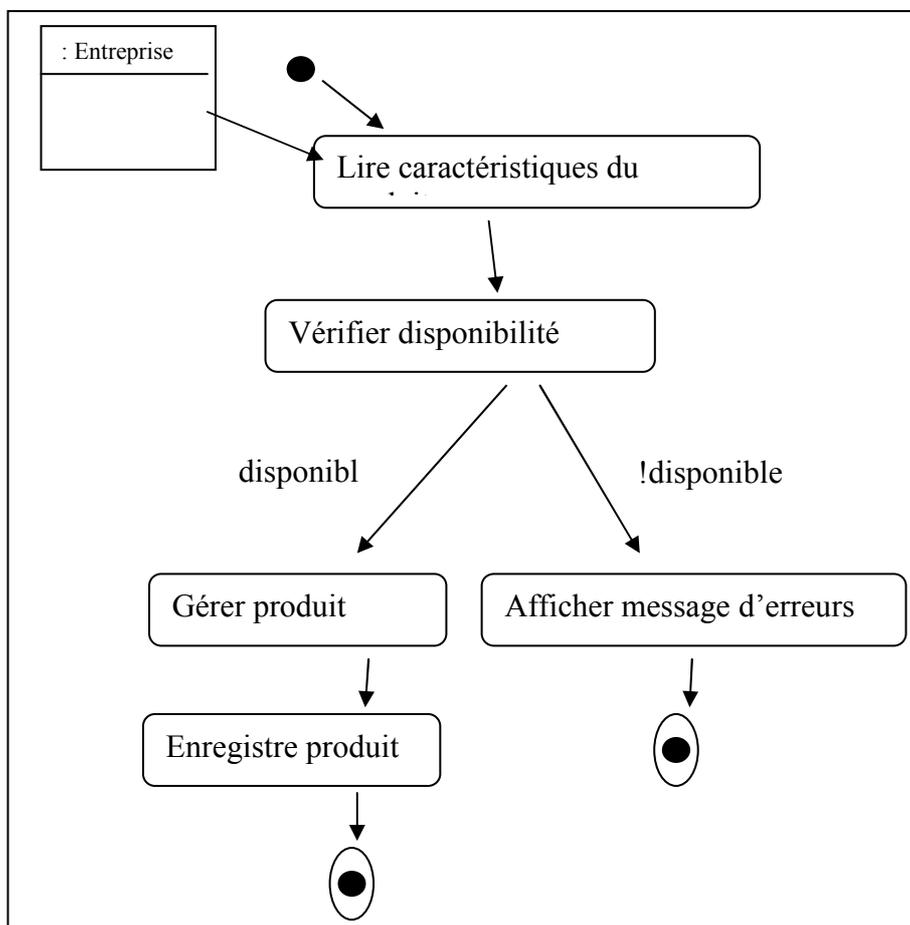


Figure 4.19 Diagramme d'activités associé à la gestion de l'organisation

### 4.3.1.7 Gestion des versions

<b>Nom</b>	Gestion des versions
<b>Acteur</b>	Concepteur-MP
<b>Type</b>	Primaire
<b>Description</b>	permet de conserver et de restituer les versions successives d'un ensemble de données relatifs a un produit, processus, documents, etc. La mémoire de projet doit gérer les changements dans la définition de produits. Ce changement peut être demandé ou résulté d'une non-conformité : « anomalie»
<b>Actions</b>	<u>Au début</u> : Lire le numéro de version <u>En cours</u> : vérifier l'objet des changements et valider les changements <u>En fin</u> : enregistrer la version

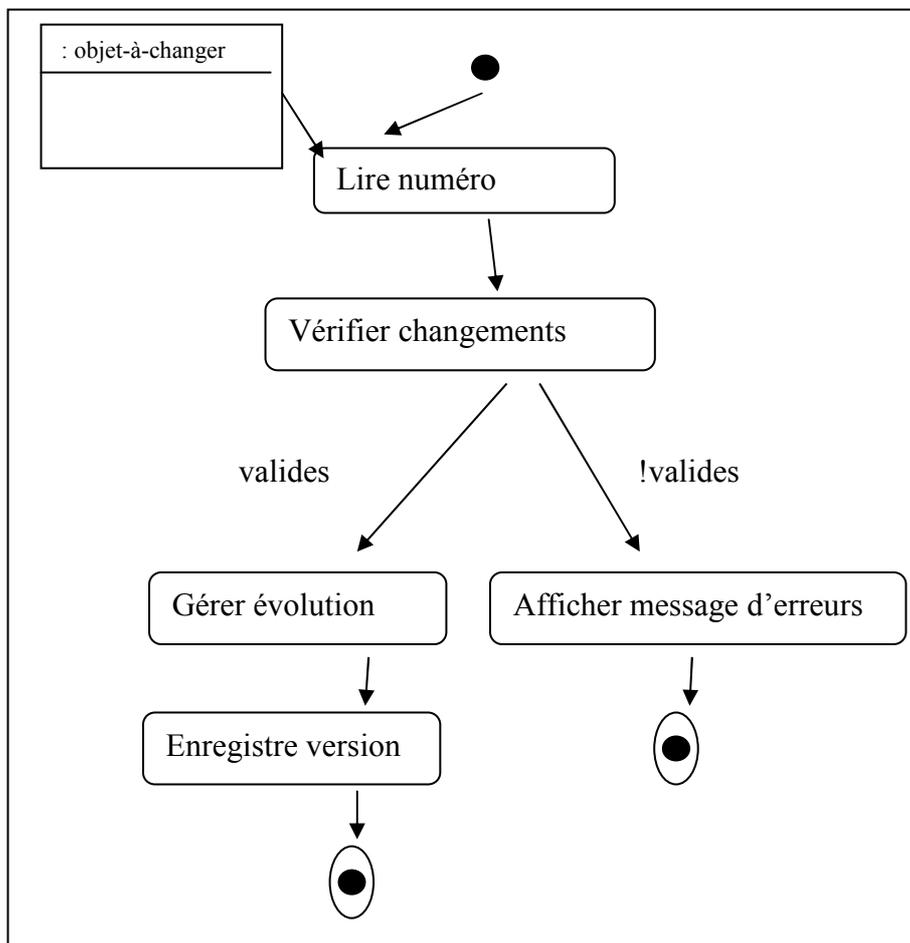


Figure 4.20 Diagramme d'activités associé à la gestion de versions

## Modèle de classes

Le modèle de classes spécifie le niveau concepts de la MP. Nous avons opté pour ce niveau comme dans [Ben staa, 2009] deux sous-niveaux :

- Sous-niveau paquetages : qui offre un mécanisme général pour la partition des modèles et le regroupement des éléments de modélisation. Il est composé d'un ensemble de paquetages (Figure 4.20). Chaque paquetage contient des classes et leurs associations. Notre décomposition en paquetages ne suit pas des critères fonctionnels; mais chaque paquetage est un groupement d'éléments selon un critère purement logique dans le but d'avoir une cohérence forte entre éléments d'un même paquetage et un couplage faible entre paquetages.

- Sous-niveau Classes : c'est un niveau de détail plus fin qui contient la description statique de la mémoire de projet en termes de classes et de relations entre ces classes.

La recherche de généricité dans la conception des classes conduit à utiliser un patron de conception. Nous avons opté pour le patron proposé par [Gamma et al, 1994] (figure 4.21) mais modifié. Dans le sens où la limitation de la non décomposition des feuilles sera annulée. Et de ce fait toutes les classes se présentent de la même manière. Les spécificités de chaque classe pourront être détaillées dans des diagrammes d'objets. Mais dans notre cas nous ne les avons pas abordés.

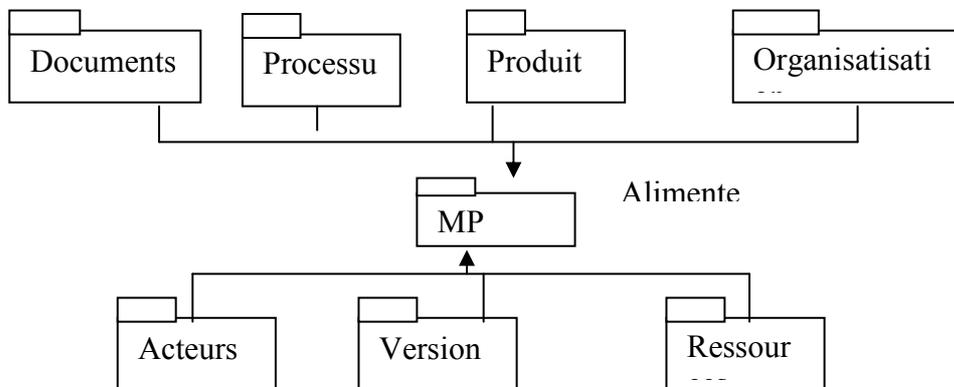


Figure 4.21 Diagramme de paquetages

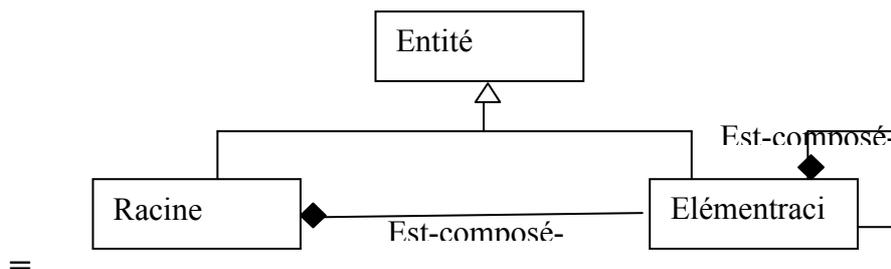


Figure 4.22 Patron décrivant une décomposition arborescente.

Par exemple pour le paquetage organisation, nous pouvons obtenir le diagramme suivant

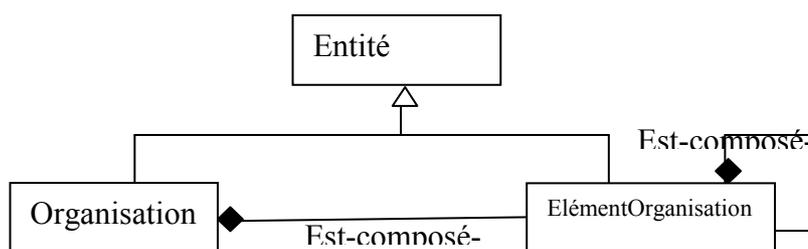


Figure 4.23 Diagramme de classes pour « Organisation »

Les diagrammes de classes seront incorporés dans l'ontologie selon le processus spécifié au début du chapitre

## 5. Conclusion

Dans ce chapitre, nous avons décrit un modèle qui supporte la gestion d'une mémoire d'entreprise et plus spécialement une mémoire de projet. Ce modèle se compose de trois catégories de pétales. La première pour prendre en considération le cycle de vie dans le cadre d'une gestion de connaissance. La seconde concerne l'interaction avec l'environnement et la troisième supporte la capitalisation. Au cœur, se situe le patrimoine des connaissances cruciales. Chaque pétale est un processus. Ce modèle est projeté sur un cas spécial de mémoires qu'est la mémoire de projet. Nous nous sommes intéressés à ce type de mémoire vu le terrain applicatif. Ajouté à cela, les bénéfices attendus de la mémoire de projet, tels que : l'amélioration continue des projets, l'évolution des connaissances et l'amélioration de l'organisation dans une entreprise.

La construction de la mémoire d'entreprise se base sur un processus de développement d'ontologies. Ce processus est une combinaison entre des représentations conceptuelle et une méthodologie de construction (inspirée de « methontology ») [Fernandez et al., 1997]. Cette méthodologie est constituée de cinq phases successives avec la possibilité de retour en arrière. Ces phases commencent par la spécification jusqu'à l'implémentation.

L'implémentation de l'ontologie qui consiste à sa codification dans un langage du web sémantique est faite en Owl-dl.

Pour être exploitée, l'approche d'annotation sémantique que nous avons adoptée sera présentée dans le chapitre suivant.

## **CHAPITRE 5**

### *Une Approche pour l'annotation d'une Mémoire d'Entreprise*

# **Une Approche pour l'Annotation d'une Mémoire d'Entreprise**

*Semantic annotation is a knowledge-level indexing structure that allows systems to use several indexing sources and to manage heterogeneous indexed resources. They play both central roles of persistent repository of acquired and persistent indexing of heterogeneous information resources; together annotations and resources constitute the persistent memory.*

*Fabien Gandon*

## **1. Introduction**

L'annotation d'une mémoire d'entreprise via ses documents constitue le moyen qui permet d'associer une sémantique au contenu de la mémoire. Enrichir le partage des connaissances et améliorer les échanges entre les différents partenaires sont parmi les principaux objectifs. En effet, avec la grande masse de données gérées à travers le monde et surtout avec l'avènement du web, l'annotation manuelle de ces documents est impossible. Dans ce chapitre, nous présentons une approche d'annotation sémantique de mémoire d'entreprise basée sur l'utilisation d'une ontologie. Nous nous intéressons dans notre travail à l'annotation semi-automatique. Notre approche consiste à relier dans un premier temps, les documents clés représentant la mémoire à annoter aux concepts de l'ontologie de documents. Puis, à relier dans un deuxième temps, les mots clés représentant le document à annoter aux concepts de l'ontologie de domaine pour aider l'auteur à réaliser l'annotation.

## **2. Hypothèses**

Avant de présenter le principe de l'approche d'annotation sémantique, nous avons proposé un ensemble de critères. Cette proposition est faite pour deux raisons : la première est de trouver un lien entre le modèle proposé pour la mémoire d'entreprise et la deuxième est de pouvoir la faire marcher sur un système via un outil adéquat.

En se basant sur les travaux présentés par [Handschuh , 2005] et ceux d'[Uren, 2006], nous avons cerné un ensemble de pré requis, que nous avons spécifié par le terme « critères ». Comme dans la plupart des travaux sur le web sémantique, nous avons considéré des critères qui portent sur les quatre aspects et qui sont : les ontologies, les documents, les annotations qui lient les ontologies aux documents, et les utilisateurs du système qui supporte les trois premiers aspects. De ce fait, chaque aspect suggère un ou plusieurs critères.

Par exemple, l'aspect ontologique suggère le critère outil pour soutenir des ontologies multiples en évolution et l'aspect document suggère la nécessité de soutenir la réutilisation et la gestion des versions des documents.

## **2.1. Critère 1 : Besoins en Mémoire d'Entreprise**

Le besoin de construire une ME dépend fortement de la taille de l'entreprise (grande, moyenne ou petit projet). Nous avons jugé comme dans [Dieng et al., 2005] que ces besoins ressortent de l'organisation elle-même, de ces connaissances et ces employés. Nous pouvons citer par exemple, (1) éviter la perte de savoir-faire d'un spécialiste après sa mutation, décès ou la retraite, (2) exploiter l'expérience acquise des projets passés afin d'éviter de reproduire certaines erreurs, (3) exploiter la carte des connaissances l'entreprise pour la stratégie d'entreprise, (4) améliorer la circulation de l'information et de communication dans l'entreprise, (5) améliorer l'apprentissage des employés dans l'entreprise, (7) savoir différencier les savoirs faires d'une organisation.

## **2.2. Critère 2 : Classification des Mémoire d'Entreprise**

Nous supposons que le choix d'un type de ME est nécessaire et il doit dépendre du besoin réel de gestion au sein de l'entreprise. Par exemple dans notre cas, une mémoire de projet dépend fortement du type du projet (industriel, commercial ...), néanmoins les actions ordinaires de la définition du projet telles que : les activités d'un projet, son historique et ses résultats restent communes pour tout types de mémoire.

## **2.3. Critère 3 : Modèle pour la mémoire d'entreprise**

La prise en compte d'un modèle pour la mémoire d'entreprise est une tâche cruciale. Dans notre travail, nous proposons le modèle de la Marguerite Etendu pour tout type de mémoire d'entreprise et par la suite, nous avons étendu ce modèle pour prendre en considération les spécificités d'une mémoire de projet.

## **2.4. Critère 4 : Support de documents hétérogènes**

Dans la construction d'une ME, nous trouvons les connaissances tacites et les connaissances explicites dans une entreprise. Quelque soit la nature de ces connaissances, la meilleure façon de les exploiter et de les formaliser sous forme de documents (hétérogènes selon le contenu et homogènes plus ou moins selon le format). Les standards du web sémantique pour l'annotation tendent à supposer que les documents étant annotés sont dans des formats du web-native, tels que HTML et XML. Par exemple, l'approche d'Annotea localise chaque annotation dans un point particulier du document en utilisant la technologie XPointers.

Les documents seront dans la majorité des cas dans de différents formats comprenant des fichiers de traitements, des bilans, des fichiers graphiques et des mélanges complexes de différents formats. Cela présente un défi technique plutôt qu'un défi de recherches, mais en traitant des documents de formats multiples consiste un préalable pour l'annotation de s'intégrer dans la pratiques comme matière existante du travail.

## **2.5. Critère 5 : Support de documents intelligents**

[Uren, 2006] définit un document intelligent comme étant un document qui "connaît" son propre contenu afin que les processus automatisés peuvent «savoir quoi faire". Nous projetons cette définition sur la ME et de ce fait, nous pouvons obtenir une ME intelligente comme étant une ME qui contient des documents intelligents.

## **2.6. Critère 6 : Support de l'évolution des documents**

Nous considérons que la cohérence et la flexibilité d'une ME dépend de la création dynamique et l'évolution des documents et des ontologies en relation. Dans la pratique, les ontologies changent parfois, mais certains documents changent assez fréquemment. Ces changements peuvent mener à plusieurs problèmes lors de l'annotation de ces documents. Donc, il faudra considérer plusieurs versions d'un document.

## **2.7. Critère 7 : Support de l'évolution des documents**

L'existence d'un ou de plusieurs ontologies est utile d'un point de vue sémantique. Nous considérons que la sémantique d'une EM ressorte de la sémantique obtenue à partir de ses ontologies.

## **2.8. Critère 8: Centrée sur l'utilisateur**

La conception d'une ME exige la collaboration entre les différents acteurs: les experts, les annotateurs, les employeurs, etc. Ce critère, pousse à exploiter les outils d'annotation utilisant des interfaces partagées entre les acteurs de l'entreprise. La simplicité et la flexibilité des interfaces rendent la tâche d'annotation un travail quotidien de collaboration. L'annotation peut potentiellement devenir un goulot si elle est réalisée par des annotateurs entourés par beaucoup d'exigences vis-à-vis de leur temps. Depuis peu d'organismes ont la capacité d'employer le professionnel des annotateurs, il devient, donc crucial de fournir des interfaces faciles à utiliser qui simplifient le procédé d'annotation et le placent dans le cadre du travail quotidien. Une meilleure façon est d'avoir une interface d'entrée unique de sorte que l'environnement dans lequel les utilisateurs annotent des documents soit intégré avec celui dans lesquels ils les créent, lisent, partagent et éditent.

## **2.9. Critère 9 : Stockage des annotations**

La conception d'une ME est une partie d'un système de gestion des connaissances de l'organisation. Le modèle du Web sémantique suppose que les annotations soient stockées séparément du document original, tandis que le modèle « d'unité de traitement de texte » suppose que des commentaires soient stockés en tant que partie intégrante du document, qui peut être consulté ou pas en tant que le lecteur préfère. Le modèle du Web sémantique, qui découple le contenu et la sémantique, fonctionne

particulièrement bien pour l'environnement du Web dans lequel les auteurs des annotations ne font aucun contrôle sur les documents qu'ils annotent. Dans un environnement de gestion de connaissances, la pratique a montré que beaucoup d'annotateurs sont plus familiers avec le document-central, c-a-d le modèle d'unité de traitement de texte. Nous supposons que les documents de la ME sont séparés de leur stockage des annotations, mais avec quelques restrictions pour des cas particuliers pour des raisons de contrôle de versions des documents.

## 2.10. Critère 10 : Automatisation

Pour réaliser l'automatisation, l'intégration des technologies d'extraction de connaissances dans l'environnement de l'annotation est essentielle. La principale motivation étant le fait de faciliter d'identifier automatiquement les entités. De ce fait, nous considérons que l'utilisation des outils d'annotation manuelle n'est pas suffisante. Donc, la combinaison de l'automatisation et les outils manuels semble une idée rationnelle (pour le projet, l'ontologie et l'annotation sémantique).

## 3. Modèle d'annotation

Après avoir présenté un ensemble de conditions, nous allons présenter dans cette section un modèle pour l'annotation inspiré du modèle formel Euzenat [Euzenat, 2002]. Ce modèle est fourni dans le cadre du web sémantique. Il formalise l'annotation comme étant une relation entre deux ensembles d'objets :

1. L'ensemble D de documents
2. l'ensemble C des représentations formelles.

Cette relation définit deux fonctions :

1. la fonction « annotation » dans le sens :  $D \rightarrow C$ . Telle que :  $\text{Annotation}(d) = c / D \ni d \text{ et } C \ni c$
2. La fonction inverse « index » dans le sens  $C \rightarrow D$ . Telle que :  $\text{Index}(c) = d$

Dans notre cas, nous allons modifier les fonctions  $\text{Annotation}()$  et  $\text{Index}()$  de la manière suivante :

**Annotation : (Cri, ME)  $\rightarrow$  Rep-fo**

**Index (Rep-fo)  $\rightarrow$  (Cri, ME)**, telle que :

**Cri**: Ensemble des critères définis précédemment.  $\text{Cri} = \sum_{i=1}^{10} (\text{Cri}_i)$

**ME**: Ensemble des documents et des ontologies informelles ainsi que l'ensemble des utilisateurs potentiels.  $\text{ME} = \sum (\text{DP}, \text{OI}, \text{UP})$  où :

DP: Document Potentiel

OI: Ontologie Informelle

UP: Utilisateur Potentiel

**Rep-fo** : représentation formelle (ontologie)

Les documents potentiels décrivent l'ensemble des documents dont on a besoin. Il pourra s'agir de :

- documents potentiels : notes, archives personnelles
- documents liés à un projet : comptes rendus de réunion de projet, documents de spécification de produits, documents de conception, documents contractuels
- des fiches d'expertise
- des bibles de référence sur un métier.
- Des rapports techniques
- etc.,

Ces documents peuvent avoir divers statuts : ils peuvent être confidentiel, approuvés, être des documents de référence ou des documents de travail, être obsolète, faire partie d'archives [Fondin 1998].

Différents supports sont possibles pour les documents de l'entreprise, en particulier ils peuvent être sous forme papier ou sous forme électronique. Selon la culture de l'entreprise, le papier joue un rôle plus ou moins important. Afin d'être intégré dans la mémoire d'entreprise à base de documents, un certain nombre de documents papier pourront être transformés en documents électroniques, grâce à l'utilisation de scanners ou d'outils de reconnaissance optique de caractères (OCR).

Dans certains cas, l'entreprise dispose en outre des formats internes de documents que doivent respecter tous les documents d'une certaine nature : par exemple, les rapports contractuels suivront un certain format, les rapports annuels d'activité, les fiches d'expertise ou les dossiers de veille technologique.

Le texte peut être non structuré comme par exemple lorsqu'il contient des marqueurs spécifiques permettant d'identifier de façon logique certains éléments du texte. Ou fortement structuré comme une base de données.

La structuration et la mise en forme d'un texte permettant d'y retrouver plus facilement des informations.

Comme la mémoire documentaire vise à permettre aux utilisateurs d'accéder aux documents pertinents, en cas de besoins, cela peut être facilité par un descripteur du contenu des documents, soit par indexation, soit par résumé. L'index permet d'associer au texte un ensemble de mots clés précisant les informations essentielles de ce texte.

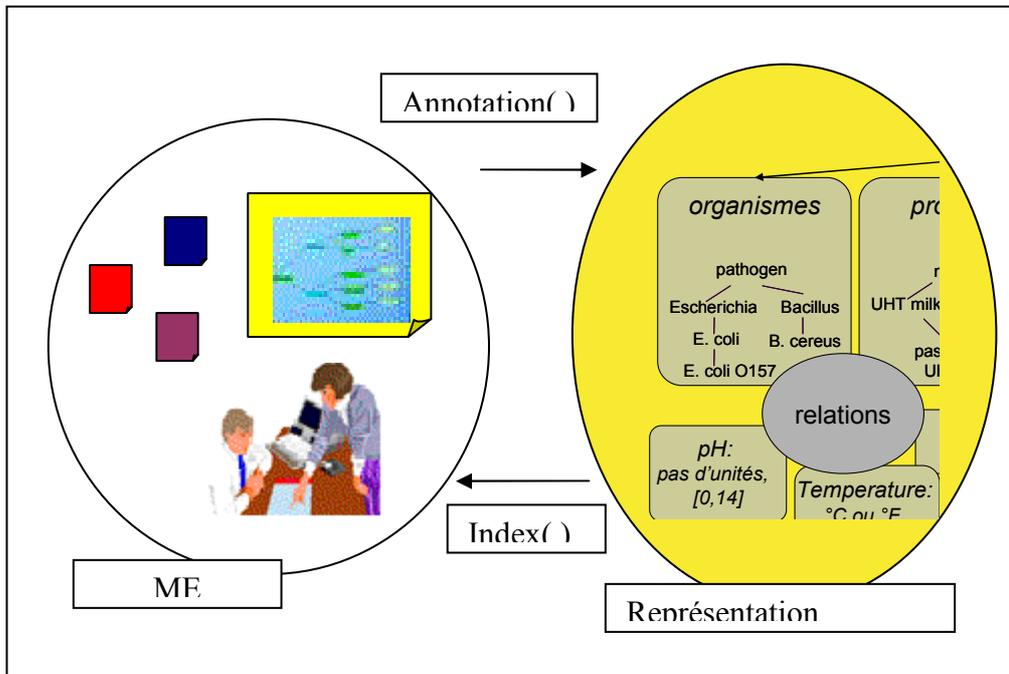


Figure 5.1 La fonction d'annotation

#### 4. Approche proposée

Annoter c'est accompagner un document textuel de notes, de remarques, des explications, de commentaires pour aider le lecteur à le comprendre. Actuellement et avec ce grand volume d'information, il est difficile d'annoter manuellement des millions de ressources mises à la disposition des utilisateurs. L'indexation d'un document textuel, consiste à repérer dans son contenu certains mots ou expressions particulièrement significatifs (Appelés termes d'indexation) dans un contexte donné, et à créer un lien entre ces termes et le document d'origine. Les mots qui entrent dans la composition de l'annotation sont déterminés à partir d'une analyse mixte : le calcul du degré de similarité et le calcul de la fréquence.

La tâche de notre système consiste à prendre en entrée une mémoire d'entreprise et fournir en sortie le même contenu enrichi par des annotations sémantiques basées sur des représentations de la connaissance plus ou moins formelles. Mais bien que ces annotations permettent de décrire rigoureusement le contenu sémantique des documents, leur création reste un processus difficile et coûteux pour les acteurs de l'entreprise (temps, personnes ...). L'objectif de cette partie est de proposer une méthodologie et un système pour la génération (semi-) automatique d'annotations sémantiques.

Comme nous l'avons décrit dans le chapitre 3, nous avons opté pour l'utilisation des techniques du Web Sémantique et en particulier pour l'utilisation des annotations sémantiques basées sur les ontologies pour faciliter l'accès aux connaissances contenues dans les documents. Toutes ces informations nous ont permis d'élaborer une méthodologie pour la génération semi-automatique

d'annotations sémantiques. Ces annotations décrivent les relations entre les termes contenus dans le texte et qui sont jugés intéressants pour La mémoire d'entreprise, tout en se basant sur les concepts et les relations de l'ontologie.

#### 4.1. Description de l'approche

L'idée est de faire intégrer les intéressés ou les acteurs dans tout le processus de l'annotation sémantique. Or l'objectif de l'annotation dans notre cas est de favoriser l'exploitation de la mémoire d'entreprise. Donc, nous pensons que les relations humaines peuvent être exploitées, telle que la négociation.

Une négociation est un espace de discussion entre plusieurs acteurs où différents objectifs sont confrontés, des alliances et des conflits sont constitués. De même, une négociation a une histoire et est influencée par les alliances et les décisions prises lors des négociations passées.

Notre approche permet de garder en mémoire cette dynamique de négociation en sorte que sa restitution soit facile pour faire apparaître les différents éléments inclus dans une résolution de problème à travers l'annotation sémantique.

Cette négociation repose sur le protocole d'appel d'offre [Davis and Smith, 1983]. Si les acteurs de l'entreprise veulent annoter une ressource de la mémoire d'entreprise et cette ressource est présentée soit par un diagramme de classe uml ou par une forme textuelle, cela peut guider vers un conflit, dont on souhaite résoudre. On suppose que le concepteur de la ME est chargé de montrer la forme du document à annoter et les utilisateurs quant à eux ils réalisent la fonction d'annotation. Donc, le concepteur propose la ressource (éventuellement avec ses présentations) à annoter et les utilisateurs choisissent la forme adéquate dotée des argumentations possibles. A la base de ces argumentations le concepteur prendra sa décision et donne le feu vert aux utilisateurs pour commencer l'annotation (figure.5.2).

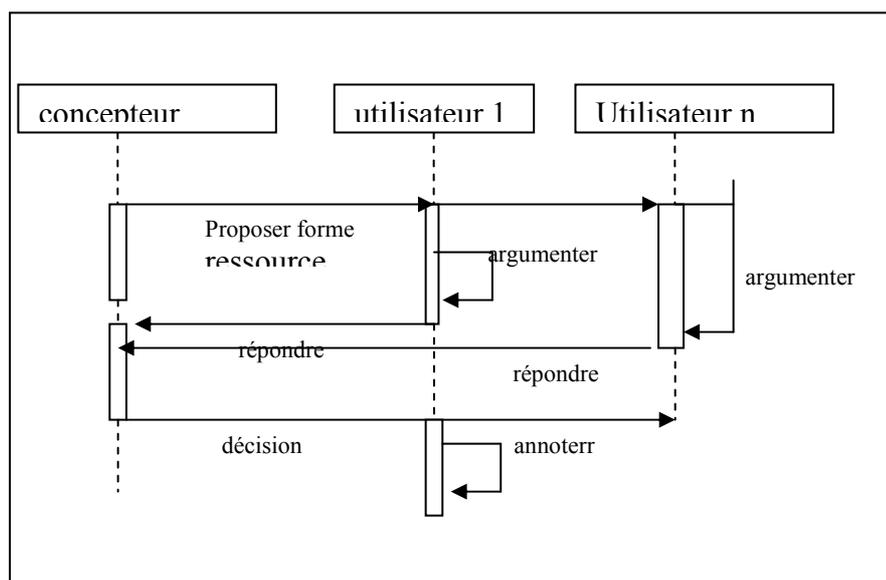


Figure 5.2 protocole d'appel d'offre pour annoter

#### 4.1.1 Annotation à partir des représentations conceptuelles

Le choix d'UML pour décrire des ontologies pourrait paraître comme une alternative beaucoup plus simple. Il existe entre les langages de description d'ontologies et UML un certain nombre de notions communes : classes, relations, propriétés, héritage,.. Les profils UML existant dans ODM (ODM (Ontology Definition Metamodel) [IBM et al, 2006] fournissent une passerelle entre UML et les différents langages de représentation de connaissances. Nous proposons d'utiliser des stéréotypes UML. Les stéréotypes peuvent s'appliquer sur la plupart des éléments du modèle de la représentation conceptuelle tels que les classes, attributs ou encore associations. Ces stéréotypes permettent de distinguer et de spécialiser les éléments du modèle en leur donnant une sémantique particulière. Les stéréotypes sont associés à des valeurs étiquetées qui se composent d'un nom de propriété et d'une valeur associée. L'application des stéréotypes sur la représentation conceptuelle est considérée comme annotation manuelle.

Pour l'annotation sémantique si nous considérons les nom des classes comme concepts et les objets comme étant des instances. Alors, il suffit de les regrouper dans un glossaire de termes, puis chercher dans la hiérarchie de l'ontologie l'existence. A ce stade l'annotateur peut utiliser un des outils d'annotations existant (chapitre2). Le tableau suivant présente des correspondances entre UML et RDF.

<b>UML</b>	<b>RDFS/RDF</b>
Class	rdfs: Class
Generalization	rdfs:subClassOf
Association	rdf:Property
Attribute	rdf:Property
InstanceOf	rdf :type
Attribute type : String	rdfs :Literal
Attribute value	Rdf : value

Table5.1. Correspondances entre UML et RDF/RDFS

Nous avons exploité les stéréotypes introduit par [Faucher et al., 2007] liés à la génération de l'ontologie et utilisés pour l'annotation. Ces stéréotypes sont présentés dans la figure suivante et interprétés par les tableaux ci dessous :

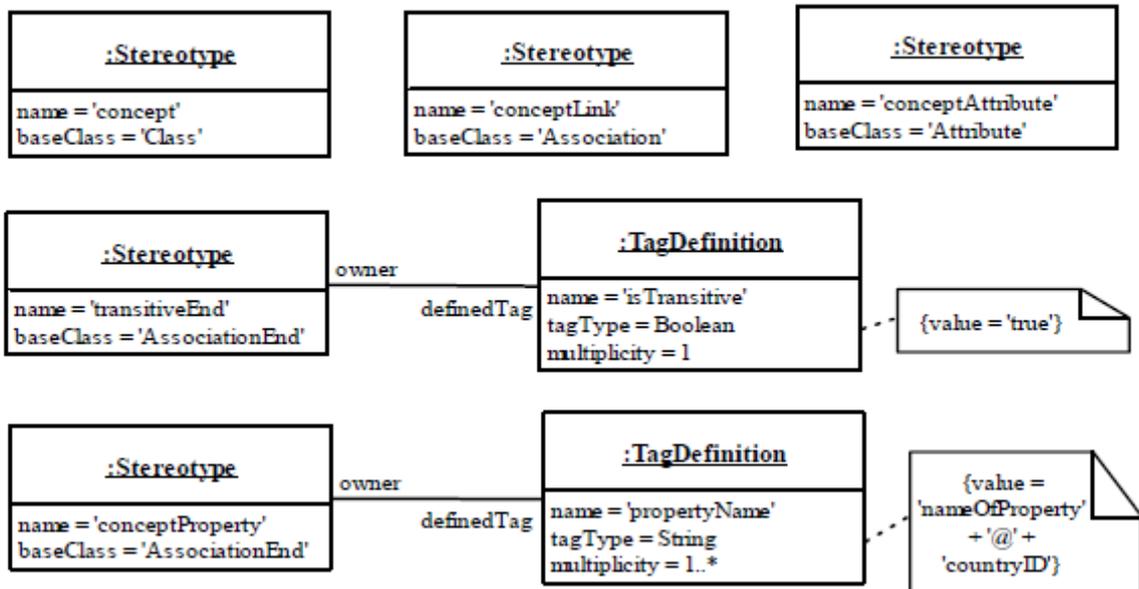


Figure 5.3 Définitions des différents stéréotypes utilisés pour l'annotation

Nom	S'applique à	Tag (valeur étiquetée)	Description
concept	Class		Classe représentant un concept que l'on souhaite intégrer dans l'ontologie
conceptAttribute	Property		Propriété représentant un attribut d'une classe stéréotypée concept
conceptLink	Association		Association, entre deux classes concept, sémantiquement intéressante pour intégrer l'ontologie
conceptProperty	AssociationEnd	propertyName de type String avec une cardinalité multiple (1..*). Les valeurs étiquetées correspondantes vont contenir les noms des rôles dans une langue donnée sous la forme : nameOfTheProperty@en	Extrémité d'association liée à une classe concept et à une association conceptLink, la valeur du nom de l'extrémité d'association contient un code. conceptProperty possède des valeurs étiquetées.

Table5.2 Stéréotypes liés à la génération de l'ontologie

transitiveEnd	AssociationEnd	Extrémité d'association conceptProperty possédant la particularité d'être transitive : lors de la génération de la base de connaissance une inférence de transitivité va être appliquée, ex. : relation « broader ».
---------------	----------------	--

Table5.3 Stéréotype lié à la génération de la base de connaissance

Le dernier tableau permet de décrire le stéréotype lié à la génération de la base de connaissance. Les « conceptLink » ne sont pas directement liés aux concepts. La notion de propriété est utilisée pour modéliser les rôles d'un concept par les « conceptProperty ». Les conceptLink sont liés aux concepts par l'intermédiaire des conceptProperty. Un concept peut avoir plusieurs propriétés elles mêmes liées à un conceptLink.

Si on reprend l'exemple présenté dans le chapitre précédent concernant l'entreprise qui veut lancer un appel d'offre pour réaliser un projet de construction des pièces de rechange (Figure 5.4).

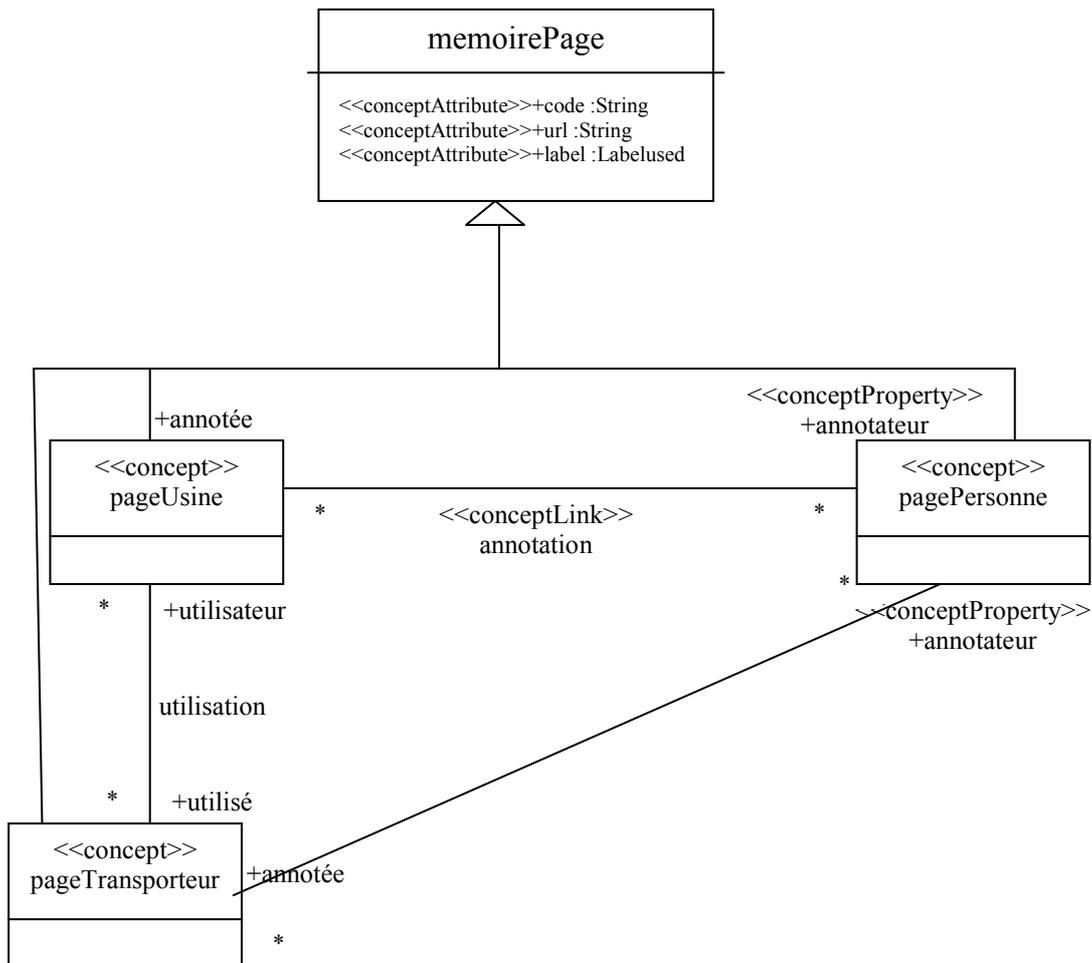


Figure 5.4. Exemple de stéréotypes utilisés pour l'annotation

Parmi les classes importantes, nous avons les classes Usine et Fournisseur qui représentent respectivement l'usine qui fabrique les produits et le vendeur de la matière première et la classe Transporteur qui joue le rôle de l'intermédiaire entre les deux. Pour la mise en oeuvre de l'annotation avec les stéréotypes définis précédemment, nous supposons que le concepteur de la mémoire d'entreprise souhaite naviguer d'une pageUsine vers un pageTransporteur en utilisant le rôle annotateur. Ceci se traduit par l'application aux classes pageUsine, pageTransporteur et pagePersonne du

stéréotype « concept ». L'association annotation doit être stéréotypée « conceptLink » afin de prendre en compte le rôle annotateur. Le concepteur souhaitant naviguer suivant le rôle annotateur, ainsi il est stéréotypé « conceptProperty ». Les classes pageUsine, pageTransporteur et pagePersonne stéréotypées par « concept » seront introduites dans l'ontologie comme de type rdfs:Class et comme des sous-classes de la classe memoirePage. Dans cet exemple d'annotation on peut montrer également la possibilité de filtrer les attributs utilisables pour la navigation et la recherche d'objets dans la base de connaissance. Par exemple, l'attribut label de la classe memoirePage est stéréotypé « conceptAttribute ». Ainsi les labels seront disponibles pour l'affichage de métadonnées et la recherche de concepts dans la base.

#### 4.1.2 Annotation à partir du texte

L'approche standard de la modélisation d'un domaine à partir des textes consiste à (i) identifier les termes caractérisant le domaine, pour ensuite (ii) extraire les relations sémantiques qui les unissent. Nous avons exploité cette approche en incorporant l'ontologie. Les différentes étapes de cette démarche sont schématisées dans la figure 5.5. et la figure 5.6.

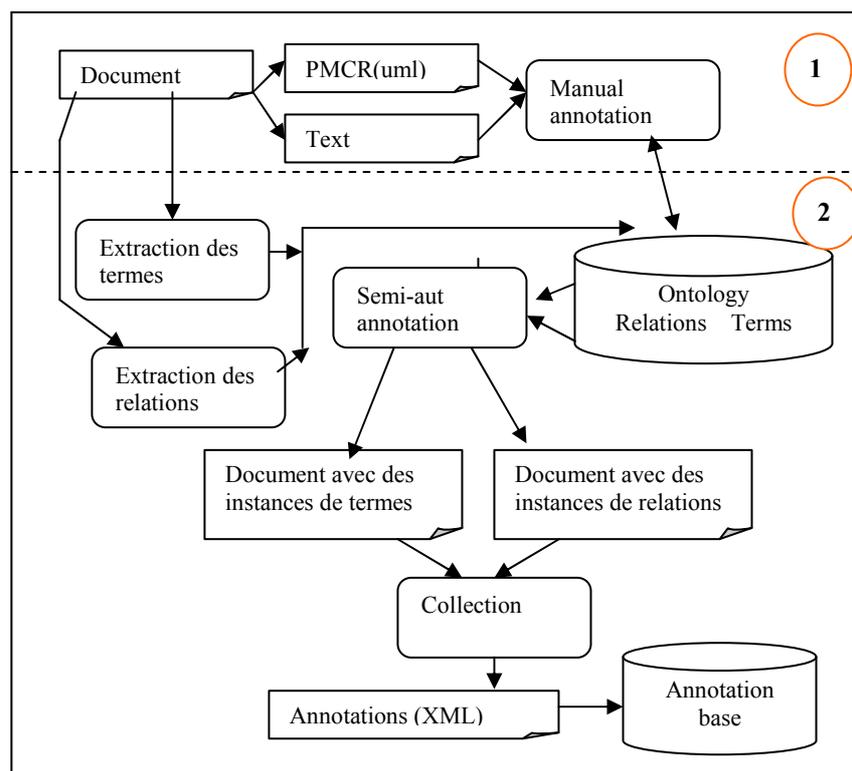


Figure 5.5. Approche d'annotation

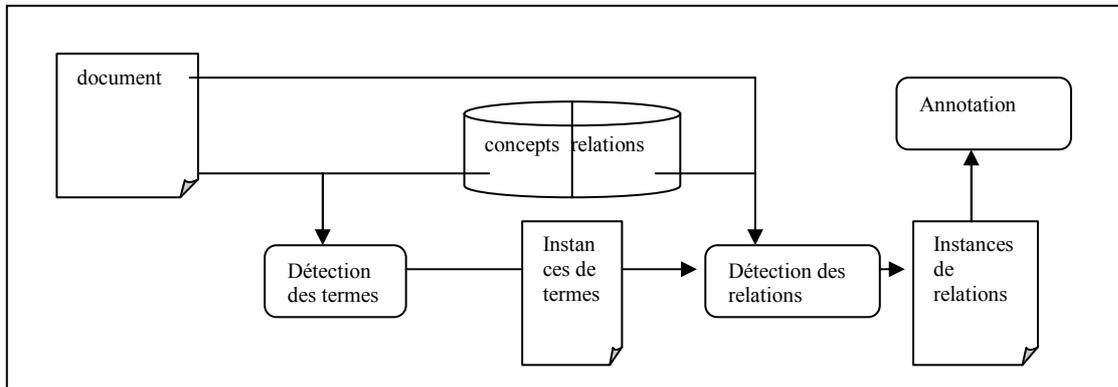


Figure 5.6. Méthode de détection

Afin de réaliser cette tâche, nous avons réparti la démarche selon deux étapes :

#### 4.1.2.1 L'étape manuelle

Dans cette étape, nous avons considéré les résultats de la modélisation formalisés par des diagrammes uml. Les acteurs de l'entreprise nous ont tout d'abord fourni plusieurs documents de projets dans lesquels nous leur avons demandé de surligner les informations essentielles, nous avons considéré celles-ci comme constituant donc les annotations de ces personnes sur les documents, annotations que nous devons reproduire automatiquement si possible. À partir de ces annotations manuelles, nous avons pu déterminer les points importants pour caractériser le contenu sémantique des documents et pour guider par la suite leurs opérations de recherches. Rappelons que le but de la mémoire de projet que nous proposons est de rendre facilement accessibles certaines connaissances cruciales aux acteurs concernés de l'entreprise. Ces connaissances peuvent provenir de plusieurs sources (i.e. humaines, BD...). En effet, les connaissances contenues dans les différents documents permettent aux employés à la fois, d'émettre des hypothèses sur leurs expériences dans un projet, de valider les résultats obtenus et enfin de pouvoir interpréter ces derniers. Nous nous posons donc la question : comment faciliter le plus possible l'accès à ces connaissances ?

#### 4.1.2.2 L'étape semi-automatique

Dans cette étape Nous nous appuyons sur le contenu qui se traduit par les mots clés qui représentent le mieux le contenu de la mémoire. Ces mots peuvent être des termes ou des relations qui doivent correspondre aux concepts et aux relations de l'ontologie.

L'étape de l'annotation consiste à associer à chaque mot clé des concepts ou des relations de l'ontologie. En se basant sur cette idée, nous avons proposé une méthodologie pour le déroulement de l'annotation sémantique à partir des documents de la mémoire. Cette méthodologie peut se décomposer en quatre grandes phases :

- La première phase consiste à effectuer une série d'analyses linguistiques et statistiques sur le document afin de le préparer pour les phases d'extraction.
- La deuxième phase consiste à analyser les phrases contenant une éventuelle instance d'une relation afin d'en extraire les instances des concepts de l'ontologie reliées par cette relation.
- La troisième vise à repérer les instances des relations de l'ontologie. Nous considérons que chaque relation est caractérisée par un ensemble de verbes et de syntagmes verbaux.
- Enfin, la quatrième et dernière phase consiste à collecter toutes les informations issues des phases précédentes pour obtenir une annotation structurée basée sur l'ontologie. Pour chaque document, une annotation globale décrivant ainsi son contenu est obtenue. Une phase préliminaire consiste à convertir les documents depuis leurs différents formats de publication (pdf, ps, etc.) vers un format textuel est indispensable pour commencer le traitement linguistique. Pour ce faire, nous avons proposé une architecture qui sera supportée par le système d'annotation. Nous avons appelé cette architecture CMOCA (Corporate Memory by Ontology and Collaborative semantic Annotation), elle est illustrée par la figure suivante.

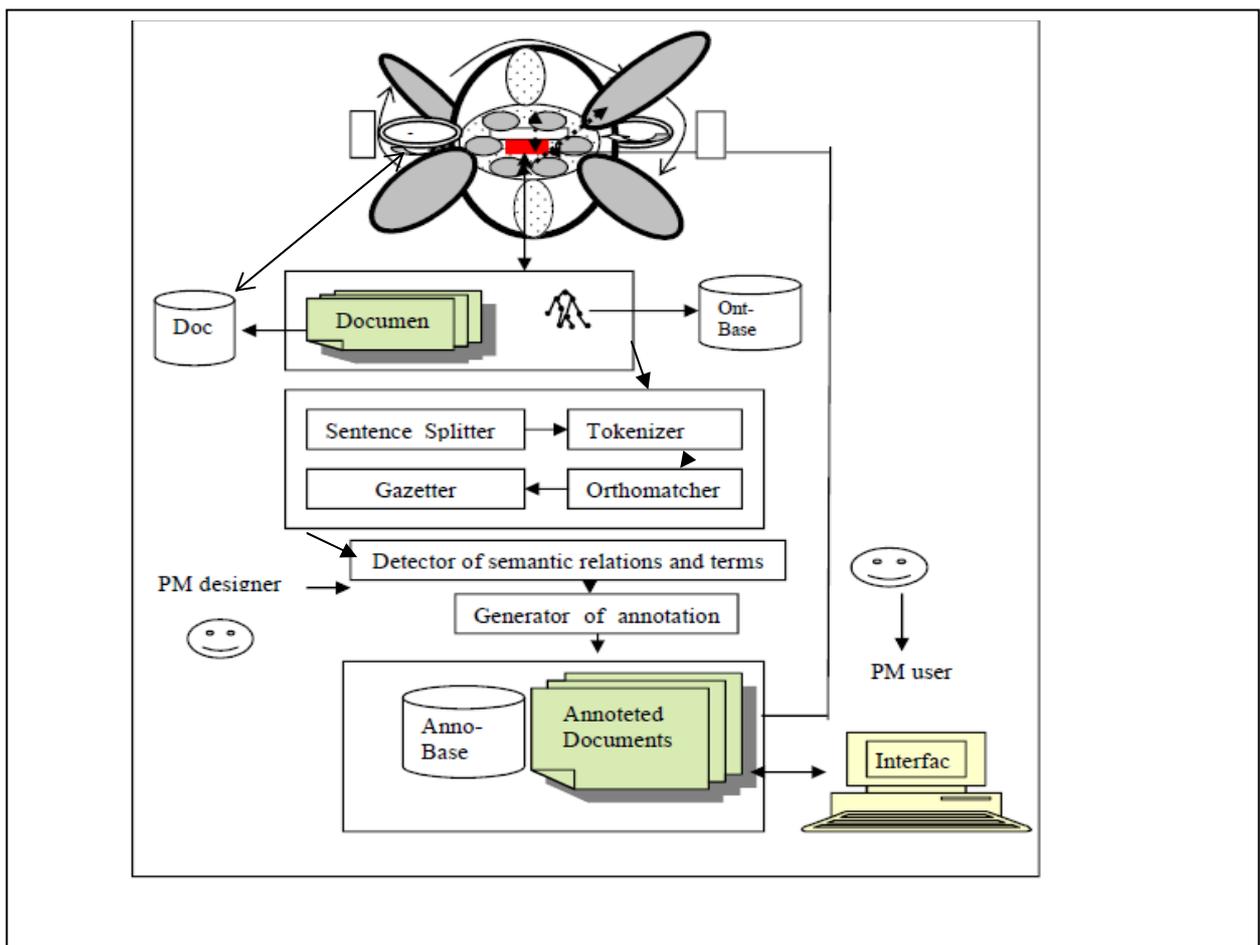


Figure 5.7 L'architecture CMOCA de base.

## 4.2 Description de l'architecture CMOCA

Cette section est consacrée à la matérialisation de la démarche proposée, à travers la conception d'un système de gestion des connaissances organisé autour d'une mémoire de projet collective. Nous proposons alors une architecture adaptée à l'environnement technologique de l'entreprise.

La matérialisation de notre démarche de gestion des connaissances s'est concrétisée par la conception d'une mémoire d'entreprise. C'est à travers cette mémoire, que nous tentons de rendre opérationnels la démarche de gestion des connaissances. La mémoire d'entreprise à concevoir est une mémoire «projet» dédiée à la capitalisation des connaissances relatives à un projet de conception d'une entreprise à caractère industriel. Cela signifie que les connaissances considérées concernent l'ensemble de documents, référentiels, outils et méthodes habituellement employés. Donc, la mémoire projet inclut non seulement les connaissances techniques, mais aussi, les connaissances opérationnelles utilisées par tous les collaborateurs pour réaliser leur travail.

L'étude de l'existant a montré le besoin d'exploiter simultanément plusieurs sources de connaissances. Ce besoin est dû aux caractéristiques des connaissances qui sont d'une part localisées dans des sources hétérogènes d'autre part, elles sont reliées entre elles sémantiquement. Une solution est de proposer une approche de médiation sémantique exploitant l'ontologie du domaine.

Cette approche [Wiederhold, 1992] consiste à définir une interface entre l'agent (humain ou logiciel) qui pose une requête et l'ensemble des sources de l'entreprise et qui sont dans notre cas des documents (utilisés pour un projet) potentiellement pertinentes. Dans une approche à base de médiation, il y a des logiciels qui exploitent des connaissances codifiées, des fournisseurs de méthodes d'accès et d'intégration de données provenant de bases de données différentes et des facilitateurs d'interopérabilité et d'échange d'information provenant de sources disparates. Cette approche permet la résolution des conflits sémantiques grâce à la spécification explicite des concepts et de leurs relations qui faciliter l'échange et la réutilisation des connaissances (ontologies).

L'objectif est de donner l'impression d'interroger un système centralisé et homogène alors que les sources interrogées sont réparties, autonomes et hétérogènes. Dans cette approche, les données semi-structurées et non-structurées peuvent être prises en charge par des modèles d'échanges adaptés. La médiation repose sur deux composants essentiels : le médiateur et le wrapper. Le médiateur identifie, localise, transforme et intègre les informations pertinentes vis-à-vis du contexte associé à une requête ; le wrapper permet à un (ou plusieurs) médiateur (s) d'accéder au contenu d'une source d'informations [Jouanot, 2001].

Nous avons incorporé l'approche médiateur dans notre architecture CMOCA de base (figure 5.8). Une architecture médiateur / wrapper est une architecture à trois niveaux : le premier niveau concerne les sources de connaissances de l'entreprise, le deuxième niveau englobe tous les outils et les connaissances nécessaires à l'interopération des sources (ontologie, médiateur, wrapper,...) et le troisième niveau est dédié aux applications.

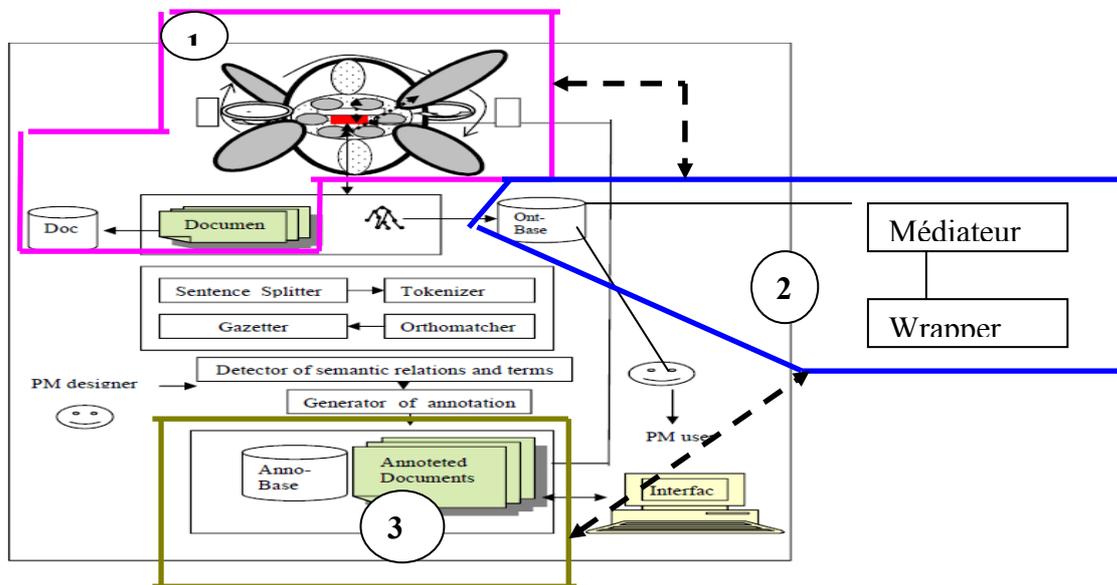


Figure 5.8 L'architecture CMOCA de base avec médiateur.

Partant de l'hypothèse que chaque concepteur possède la faculté d'intégrer de nouvelles connaissances dans la mémoire d'entreprise, le processus de capitalisation est alors assuré par les concepteurs eux mêmes. Ces derniers, formalisent, explicitent et structurent les connaissances selon le modèle proposé et en utilisant l'ontologie du domaine préalablement développée et mise à la disposition des concepteurs. Le concepteur peut être ainsi guidé dans son travail de formalisation et de structuration des connaissances. Il peut décrire de nouvelles connaissances. Ces derniers vont assurer la structuration des connaissances en tenant compte des structures : physique et logique des documents. Pour cela, nous utilisons XML comme un langage pour décrire la structure logique des documents à l'aide des balises.

Concrètement, le moyen de formalisation et d'explicitation mis à disposition des concepteurs peut être un formulaire html. Les éléments de connaissances introduits dans le système via ce formulaire sont ensuite générés automatiquement en format XML. Les connaissances ainsi formalisées et explicitées sous forme de documents exploitables, sont ensuite annotées par les concepts qui constituent l'ontologie. Nous avons choisis d'utiliser l'ontologie comme vocabulaire normalisé pour annoter les ressources électroniques et ce en raison de la fonction de représentation sémantique fournie par les ontologies.

Nous avons également, intégrer dans l'architecture de base, un serveur de connaissances. (figure 3.9). Ce ci permet de rapprocher l'architecture de base à un WSE. Le serveur de connaissances est chargé de mettre à la disposition des utilisateurs, en particulier les experts du domaine, des fonctionnalités nécessaires à l'intégration de nouvelles connaissances dans le système. Grâce à ces fonctionnalités l'expert doit pouvoir saisir les connaissances implicites selon le formalisme de l'ontologie préalablement prédéfini et transformer ainsi les savoir-faire en une matière utilisable. Une solution est de générer les éléments de connaissances introduits par l'expert, en documents XML selon des structures prédéfinies.

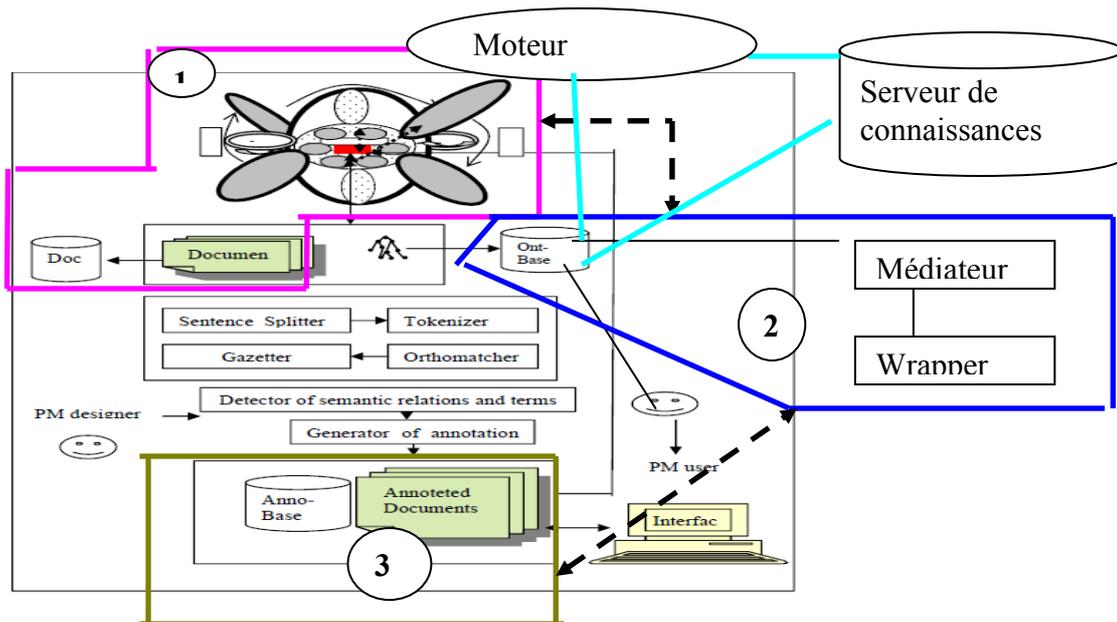


Figure 5.9 L'architecture CMOCA de base avec serveur de connaissances.

Les principales fonctionnalités de ce serveur :

- L'intégration et la validation des connaissances sous forme de ressources documentaires générées en format XML ;
- L'association des annotations via des métadonnées liées aux ressources documentaires;
- La recherche des ressources documentaires selon des métadonnées.

Nous suggérons son implémentation conformément aux standards actuels reconnus par la communauté de l'ingénierie documentaire à savoir :

- le standard XML pour l'échange et l'encodage des documents ;
- le DOM (Document Object Model) pour la manipulation et le traitement des ressources documentaires ;
- XPath et XSL/XSLT pour l'adressage et la transformation structurelle des documents XML

La base de connaissances est le noyau de la mémoire d'entreprise car, elle contient les connaissances sémantiques et conceptuelles modélisant le domaine de connaissances. Nous

considérons que cette base contient l'ontologie formelle écrite en OWL avec l'ensemble des raisonnements permis par ce formalisme basé sur les logiques de description (moteur d'inférence). Ainsi, la base de connaissances doit offrir l'ensemble des services fondamentaux nécessaires à la gestion de l'ontologie: ajout, suppression, modification, recherche, etc.

Au sein de cette base, les connaissances sont représentées en deux parties principales qui sont la T-Box et la A-Box (voir chapitre 4). La TBox contient les concepts qui constituent l'ontologie: Project, Problem, etc., tandis que la A-Box contient les déclarations sur les instances en utilisant les termes déclarés dans le vocabulaire ; les individus sont déclarés en leur donnant des noms, et en leur associant des assertions. Les liens entre le serveur de connaissance et la base est un lien directe.

### **4.3 Description du fonctionnement de l'architecture CMOCA**

Avant de présenter les différentes étapes du fonctionnement de l'architecture, nous allons illustrer dans la section suivante, les différentes analyses utilisées.

#### **4.3.1 Analyses utilisées**

Nous supposons que chaque document est présenté initialement sous une forme textuelle qui sera par la suite transformée sous une forme électronique (pages web par exemple). Nous avons utilisé une combinaison d'analyses. Dans notre travail, nous supposons que le concepteur de la mémoire a la possibilité à désigner les mots importants du texte qui seront considérés comme des mots clés. Dans ce cas les différentes étapes d'analyses utilisées sont illustrées dans la figure 5.10. Egalement, le concepteur peut suivre une démarche qui tend à réaliser une analyse morpho-syntaxique du texte (section 4.2.2).

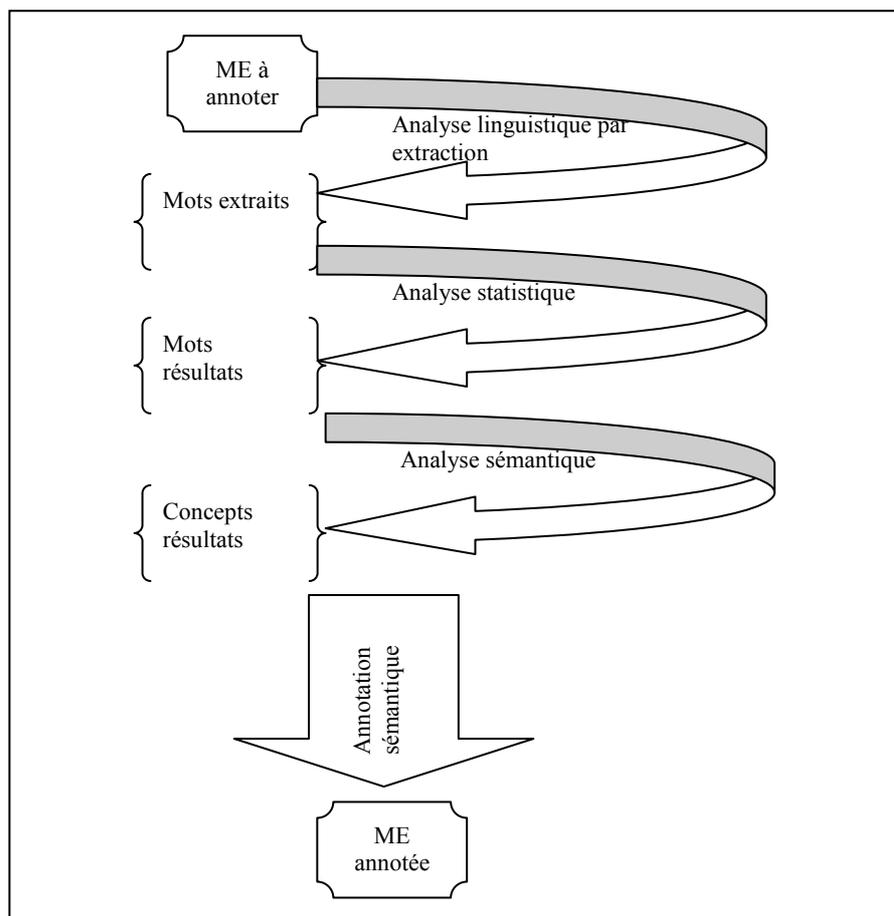


Figure 5.10 Etapes d'analyses

#### 4.3.1.1 L'analyse linguistique

Cette phase vise essentiellement à extraire les termes composants le document potentiel. L'analyse est conduite par un traitement linguistique qui représente le document potentiel par un ensemble de termes jugés importants.

#### 4.3.1.2 L'analyse sémantique

Dans cette phase, des mesures de similarités sont utilisées. Ces mesures à base de calculs visent à déterminer le poids d'un mot dans le document. Dans la littérature, Des mesures de proximité sémantique ont été proposées en utilisant des structures de réseaux sémantiques ou hiérarchiques, nous pouvons citer les mesures suivantes :

- Mesure basée sur le chemin (path based measures) entre les deux concepts à comparer telles que définies [Rada et al., 1997].
- Mesure basée sur la notion de contenu d'information (Information Content ou IC) telle que celle définie par [Wu and Palmer, 1997].
- Mesure basée sur une combinaison du chemin et du contenu d'information par [Loke and Davison, 1998].

- Mesure basée sur l'algorithme de Lesk adapté à WordNet par [Patwardhan et al, 2003].

Nous avons utilisé la mesure de Lin qui a été utilisée dans d'autres travaux, elle se base sur celle de Wu-Palmer, cette mesure a l'avantage d'être simple à implémenter et d'avoir d'aussi de bonnes performances que les autres mesures de similarité. Son principe est le suivant :

Dans un domaine de concepts, la similarité est définie par rapport à la distance qui sépare deux concepts dans la hiérarchie et également par leur position par rapport à la racine. La similarité entre C1 et C2 est :

$$Consim(C1, C2) = \frac{2 * depth(C)}{depth_c(C1) + depth_c(C2)} \quad (1)$$

Où C est le PPG de C1 et C2 (en nombre d'arcs), depth (C) est le nombre d'arcs qui sépare C de la racine et depth<sub>c</sub> (Ci) avec i le nombre d'arcs qui séparent Ci de la racine en passant par C.

Dans cette phase, un mot sera accepté si et seulement s'il est fortement en relation avec d'autres mots de cette page. Cette décision dépend du choix d'un seuil défini par l'utilisateur.

#### 4.3.1.3 L'analyse statistique

Cette phase a pour but de déterminer l'importance d'un terme dans un document, dans cette analyse nous avons utilisé la technique de pondération des termes car elle permet d'affecter aux termes d'un document, un poids pour traduire son importance dans le document, donc son degré de « porteur de connaissances ». Dans cette technique on s'intéresse à la pondération locale qui mesure la représentativité locale d'un terme. La fonction utilisée est la fonction normalisée qui permet de réduire les différences entre les valeurs associées aux termes du document. Elle est donnée par la formule suivante :

$$0.5 + 0.5 \frac{tf_{ij}}{\max_{t \in D_j} (tf_{ij})} \quad (2)$$

Où  $\max_{t \in D_j} (tf_{ij})$  est la plus grande valeur de  $tf_{ij}$  des termes du document Dj. ).Le résultat de cette étape est un ensemble de mots nommé degré\_significatif

- L'annotation

L'annotation consiste à associer à chaque mot clé des concepts de l'ontologie (noeuds). Dans cette étape on utilise une Ontologie de domaine, nous avons fait un passage des mots clés candidats à l'ontologie pour définir les concepts correspondants. A chaque passage d'un terme à l'ontologie, un ensemble de concepts sera présenté aux auteurs pour choisir les concepts à utiliser dans l'étape de l'annotation.

Cette étape est semi-automatique, la recherche et la proposition se fait à l'aide de wrappers qui sont implémentés dans le système et le choix des concepts les plus significatifs reste aux experts via le concepteur. Une association entre ces mots et ces concepts élus sera stockée dans un fichier RDF qui correspond au document.

#### **4.3.2 Détection des connaissances cruciales**

Dans cette partie, nous détaillons les différentes étapes de notre méthodologie. Ces étapes, ont été implémentées dans le système CMOCA. Ce dernier utilise la plate-forme GATE pour intégrer plusieurs modules et les applique en cascade sur le texte brut. Ces différents modules sont décrits dans les sections suivantes.

##### **4.3.2.1 Analyse morpho-syntaxique du texte**

L'objectif de cette étape est de récupérer toute information permettant de caractériser le comportement d'un mot dans son contexte d'énonciation. Dans ce cas, nous supposons que le concepteur suit les étapes classiques du traitement sur un texte.

- Le découpage du texte en phrases permet de repérer les frontières de chaque phrase dans le texte afin de pouvoir la séparer pour un éventuel traitement spécifique.
- Le repérage des entités linguistiques de base (tokenisation), à savoir les mots (tokens). Cette étape permet aussi de distinguer la morphologie de ces entités (ponctuation, nombres, etc.) et de retrouver la racine de chaque entité (lemmatisation).

L'étiquetage grammatical du texte permet d'associer à chaque token une catégorie d'ordre grammatical (Nom, Verbe, Adjectif, etc.) en tenant compte du contexte de leur occurrence.

CMOCA intègre deux modules de GATE pour effectuer les deux premières étapes (à savoir le Sentence\_Splitter et le Tokenizer) et utilise le TreeTagger pour l'étiquetage grammatical.

Cette phase fournit des informations de base sur les textes qui sont nécessaires pour les phases suivantes.

##### **4.3.2.2 Extraction des termes**

Cette phase consiste à extraire les termes susceptibles d'être des instances des concepts de l'ontologie. Lorsqu'une instance de concepts est détectée, ce ci consiste à extraire des termes candidats. CMOCA procède à l'identification des termes représentant des instances des concepts de l'ontologie qui sont dans la même phrase. Cette identification permet de déterminer les termes de l'ontologie pouvant avoir un lien ontologique. Cette extraction se décompose en deux étapes :

- La première consiste à traiter le texte et constituer une liste de candidats termes. Cette étape peut être réalisée à l'aide de n'importe quel extracteur de termes, à savoir par exemple Lexter, Nomino, Syntex, etc. (aucun outil particulier n'est imposé).

- La deuxième consiste à faire la correspondance entre la liste des termes candidats et les termes de l'ontologie afin de filtrer la première et de ne laisser que les instances des concepts de l'ontologie.

Afin d'alléger cette phase d'extraction, l'extracteur de termes exploite le résultat de l'analyse morpho-syntaxique pour optimiser le traitement et intègre quelques restrictions, telles que : un terme ne peut ni commencer ni se terminer par une préposition, un terme ne peut être un nombre, etc.

#### **4.3.2.3 Repérage des relations sémantiques dans les textes**

L'objectif de cette phase est de repérer dans le texte, les relations sémantiques déjà modélisées dans l'ontologie afin de générer des annotations basées sur ces relations. Ce repérage consiste à essayer d'identifier les différentes formes syntaxiques d'apparition des relations dans le texte. Ces formes syntaxiques peuvent ainsi être considérées comme des instances possibles de relations formalisées dans l'ontologie.

Les relations sémantiques se décomposent généralement en deux grandes familles :

- Les relations syntagmatiques : qui sont identifiables par une étude directe des formes syntaxiques : chaque relation est caractérisée par un ou plusieurs syntagmes (verbaux, adjectivaux...). Ces relations sont liées à leurs arguments au niveau syntaxique.

- Les relations paradigmatisées : qui sont l'opposé des relations syntagmatiques car elles n'apparaissent pas sous la forme d'un lien syntaxique standard. C'est le cas, par exemple, de la synonymie ou l'hyponymie. La détection de ce type de relations nécessite généralement une validation humaine.

À des fins d'automatisation, nous nous sommes intéressés à la première famille de relations, à savoir, les relations syntagmatiques et en particulier les relations pouvant être caractérisées par des syntagmes verbaux en se basant sur les catégories grammaticales des termes ayant une dépendance syntaxique avec le verbe caractérisant la relation. Nous avons utilisé comme dans [Khelif, 2006] des grammaires de détection de relations

Une grammaire de détection est une grammaire qui couvre les différentes formes syntaxiques d'apparition d'une relation dans le texte. À chaque relation sont associées une ou plusieurs grammaires, l'application de ces dernières sur le texte permet d'identifier une éventuelle expression de cette relation. Le langage JAPE (Java Annotation Patterns Engine), une variante du standard CPSL adapté au langage de programmation Java [Cunningham et al., 2002], a été proposée pour écrire des grammaires qui, une fois appliquées sur le texte, permettent d'y ajouter des informations en forme d'annotations.

Une grammaire de JAPE comporte un ensemble de phases, chaque phase étant un ensemble de règles sous forme de patron et action. Dans une règle, si les patrons sont satisfaits, alors une action pourra être déclenchée. La partie droite d'une règle qui contient des patrons doit être écrite en JAPE mais la partie gauche, qui contient les actions, peut être écrite en JAPE ou en JAVA.

Une idée pour repérer une instance d'une relation consiste à détecter l'apparition de l'un des verbes ou l'un des syntagmes verbaux dans le texte. Pour ce faire, nous avons utilisé le langage JAPE décrit précédemment pour écrire des grammaires permettant d'effectuer cette tâche. Etant donné que GATE dispose d'un transducteur permettant l'application des grammaires JAPE sur le texte, nous avons utilisé ce langage pour écrire nos grammaires de détection de relations.

## **5. Validation**

Après la présentation des différentes étapes de l'approche d'annotation proposée. Nous allons présenter dans la section suivante une étude de cas pour valider cette approche sur une mémoire de projet. Dans cette partie, nous allons fournir un aperçu sur l'entreprise d'accueil, puis nous allons présenter l'environnement de programmation utilisé (environnements et les langages de développement, outils, captures d'écran).

### **5.1. Organisme d'accueil**

**L'ENMTP** est une entreprise spécialisée dans le développement de tout processus de fabrication et de distribution entrant dans la filière travaux publics de l'industrie mécanique.

Le domaine d'activité de L'ENMTP est spécialisé dans le développement de tout processus de fabrication et distribution entrant dans la filière travaux publics de l'industrie mécanique L'activité actuelle porte sur la conception, la production et la commercialisation de matériels de terrassement, de levage, de manutention et de compactage, ainsi que sur les matériels pour air comprimé et enrobés; et ceux pour la préparation de béton et autres matériaux.

L'ENMTP travaille sous les licences de marques internationales reconnues:

LIEBHERR ( ALLEMAGNE ) pour les pelles et grues

O & K ( ALLEMAGNE ) pour les chargeurs sur pneus

LIEBHERR ( ALLEMAGNE ) pour les bulldozers et chargeurs 4m3

INGERSOLL RAND CIE ( USA ) dans le domaine du compresseur et compacteur

POTAIN ( FRANCE ) pour la fabrication de grues de bâtiment

Le potentiel industriel de L'ENMTP est considéré comme l'un des plus importants d'Afrique. Il est structuré selon une spécialisation par lignes de produits, chaque unité disposant de la plus grande autonomie industrielle. Le taux d'intégration moyen des engins produits est estimé à l'heure actuelle à 70% les unités de production de L'ENMTP sont les suivantes :

- La Filiale Complexe Pelles et Grues- CPG- (CONSTANTINE )
- Le complexe compresseurs et compacteurs (CONSTANTINE )
- L'unité grues bâtiment (BEJAIA )
- La Filiale SOMABE pour le matériel à Béton (ALGER )

### **5.1.1. Description du problème**

Nous nous sommes intéressés à la filiale Complexe Pelles et Grues- CPG- de Ain Smara à Constantine. Nous avons observé de nombreux problèmes qui augmentent les coûts et les délais dans les projets de conception. Durant les dernières années, l'entreprise a connu une masses importantes de départs ce qui a causé une perte importante de savoirs et de savoirs faire irremplaçables. Ainsi, une solution peut être, en l'installation d'un système de gestion de la mémoire d'entreprise qui rend le patrimoine de connaissances disponible pour de futurs projets.

Pour la conception du projet, une fois que la stratégie de planification a été définie, cela implique la réalisation des étapes suivantes :

- Définir plus en détail l'objectif.
- Définir les résultats nécessaires afin d'atteindre l'objectif.
- Décider ce dont nous avons besoin afin d'obtenir les résultats, par exemple 'Quelles activités devons nous réellement entreprendre ?'
- Considérer ce qui pourrait mal se passer (hypothèses).
- Définir quand 'le projet est un succès' (critères et indicateurs).

Nous avons utilisé un outil nommé matrice de cadre logique (ou matrice de planification de projet)[Blackler, 1995] (voir Table 5.4). Elle permet de transformer une stratégie de projet en une matrice de projet. La matrice de cadre logique est une feuille de travail qui résume le schéma de conception du projet, qui autrement serait trop complexe. La matrice ne couvre que les éléments essentiels. Ces éléments-clés doivent être détaillés dans la description de projet. Dans la matrice, le projet sera décrit par les éléments-clés qui définissent son contexte (objectifs, résultats, activités, données de départ, facteurs extérieurs, et critères de mesure du succès). Elle est utile dans le cas où nous aurons besoin d'un aperçu du projet et lorsque vous aurons besoin de savoir où nous sommes dans le processus de sa conception.

	Description	Descripteurs et apports extérieurs	Moyens/sources de vérification	vérification Hypothèses de travail et préconditions
Objectif d'ensemble	L'objectif auquel le projet contribuera. L'objectif peut être fixé par le pays, un bailleur de fonds, une ONG, etc			
Objectif du projet	L'objectif qui sera atteint à la fin du projet.	Description de l'objectif du projet en termes de qualité, quantité, durée, et localisation possible.	La source dans laquelle les données seront trouvées pour vérifier si les indicateurs ont été réalisés. Ceci s'obtient généralement de rapports et autres documents.	Toutes les conditions qui doivent être remplies pour que l'objectif de projet contribue à l'objectif d'ensemble, mais qui doivent être réalisées par d'autres parties que les partenaires du projet. Celles-ci sont hors de l'influence du projet.
Résultats du projet	Les objectifs qui seront atteints pendant la durée du projet. Il y en aura presque toujours plus d'un.	Description des résultats du projet en termes de qualité, quantité, durée, et localisation possible.	La source dans laquelle les données seront trouvées pour vérifier si les indicateurs ont été réalisés. Ceci s'obtient généralement de rapports et autres documents.	Toutes les conditions qui doivent être remplies pour que l'objectif de projet contribue à l'objectif d'ensemble, mais qui doivent être réalisées par d'autres parties que les partenaires du projet. Celles-ci sont hors de l'influence du projet.
Activités	Ce qui sera fait pour atteindre les résultats – à détailler pour chacun d'eux.	Moyens ou ressources nécessaires pour entreprendre les activités.	Hypothèses de travail	
			Conditions préalables	

Table 5.4 Matrice de cadre logique

Pour décrire l'interaction entre le projet et les facteurs extérieurs (hypothèses et conditions préalables) nous avons utilisé la logique de zigzag

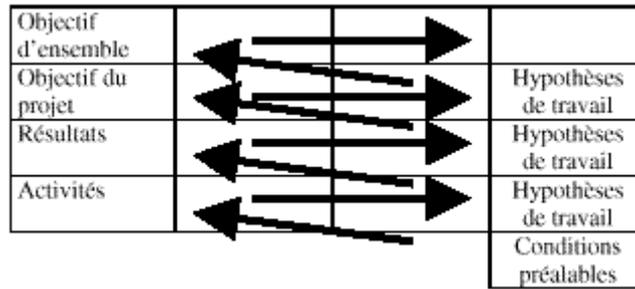


Table 5.5 Logique de zigzag

- Dans cette logique avant d'entreprendre les **activités**, certaines **conditions préalables** doivent être remplies. Les parties en dehors du projet rempliront ces conditions préalables.

- Si les **activités** se déroulent avec succès au cours du projet, et si les parties extérieures au projet assurent que certaines **hypothèses** importantes sont confirmées, alors le résultat sera réalisé.

- Si le projet atteint les **résultats**, et si les autres parties extérieures au projet assurent que certaines **hypothèses** importantes sont confirmées, alors **l'objectif du projet** sera atteint.

- Si le but du projet est atteint à la fin de la durée du projet, et que des **hypothèses** importantes sont confirmées, alors l'ensemble des objectifs sera atteint.

Cela peut être formulé par les équations suivantes :

Activités + Hypothèses = Résultats

Résultats + Hypothèses = Objectif du projet

Objectif du projet + Hypothèses = Objectif d'ensemble

Objectif d'ensemble = nouvel génération de machine chargeur+ réalisation en algerie

Objectif du projet = nouvel génération de machine chargeur+ réalisation à constantine

Réalisation en quantité et en qualité de machines= Production + à constantine en algérie

Nous avons proposé également aux concepteurs d'annoter (classiquement) cette représentation par des couleurs (Bleu pour l'objectif d'ensemble, etc.) comme illustré dans la table suivante.

	Quality descriptors/ indicators	Quantity descriptors/ indicators	Time descriptors/ indicators	Target descriptor/ indicators	Location descriptors/ indicators
Overall objective	New generation of loader	2000	2 years	Algeria	Constantine factories
Project objective	Extend recent loader	3000	1 year	Algeria	Ain smara factory
Results	Amelioration of roads design				
Activities	Assumptions and preconditions work				
	Assumptions and preconditions work				Resources / Budgets

Table 5.6 matrice annotée classiquement

Ce type d'annotation a été bien accepté par l'ensemble des utilisateurs, ce qui nous a poussé à entamer les étapes de l'annotation sémantique.

Le cycle de développement de production d'un véhicule est fait selon des sous-cycles (conception / développement / validation / d'entretien) à répétition. Lors d'un projet de conception du véhicule, le département de conception vérifie que les parties fournies répondent aux exigences des spécifications du produit. Le projet est considéré des éléments les plus élémentaires jusqu'à ce que le produit final soit conçu. Nous avons établi des inventaires de questionnements avec 35 acteurs impliqués dans le projet. Ces inventaires concernent le jugement de l'importance des 10 critères. Nous avons résumé les résultats dans le tableau suivant.

criteria	percents	Criteria	percents
Criterion 1	25%	Criterion 6	10%
Criterion 2	5%	Criterion 7	5%
Criterion 3	5%	Criterion 8	15%
Criterion 4	15%	Criterion 9	5%
Criterion 5	5%	Criterion 10	10%

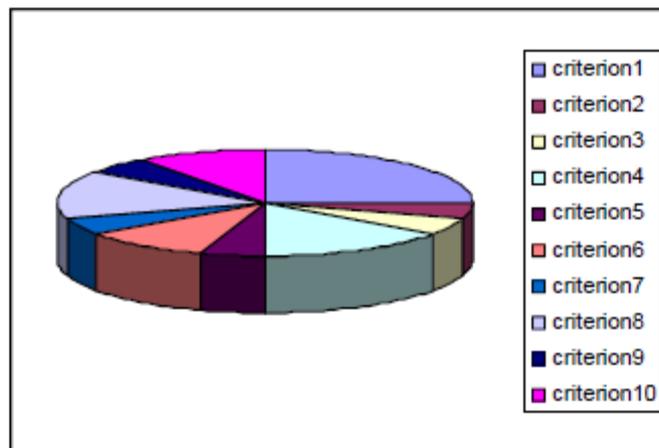


Table 5.6 Evaluation1 à partir des critères

Pour chaque critère, nous avons numéroté les acteurs intéressés et puis nous avons établi des pourcentages correspondants. Nous avons également, observé que ces personnes sont intéressées en particulier par le critère 1, le critère 4, le critère 6, 8 et critère 10.

Et c'est ainsi, que les utilisateurs sont intéressés par les connaissances et le fait de savoir s'il y a un réel besoin de mettre en œuvre une mémoire de projet, ils sont influencés par l'évolution et la nature des documents porteurs de savoir. Les utilisateurs veulent également obtenir des logiciels et des matériels adéquats pour faciliter leurs tâches. Pour les critères qui restent on a constaté qu'ils ne sont pas vraiment intéressants pour eux. Cette conclusion nous a motivé à essayé les techniques du web sémantique et notamment l'annotation sémantique.

### -5.1.2. Description du scénario

Dans l'entreprise, le scénario commence avec le choix d'un projet. Dans notre cas, nous avons choisi un projet appelé «Extension de grues et des pelles". C'est un projet enregistré il ya deux ans. Dans ce projet, les concepteurs ont besoin de comprennent résoudre les problèmes liés aux grues et pelles. Ils ont observé que ces problèmes peuvent être similaires à d'autres rencontrés dans d'autres projets ayant le même objectif, mais dans des contextes différents. La modélisation des documents commence par la définition des documents cruciaux contenant les connaissances explicites ou implicites (cas de l'expérience). Donc, nous avons proposé de structurer chaque type de documents sous la forme

d'éléments d'expérience inspirés de REX. Ces structures ont été modélisés par des diagrammes de classes UML (Unified Modeling Language). Ces documents sont par la suite transformés en fichiers XML pour être exploités dans le cadre du web sémantique. Les fichiers xml sont initialement vides puis remplis par les experts (annotation manuelle). Les documents peuvent ainsi être stockés dans des bases de données XML. Dans notre cas, nous avons généré 30 fichiers XML (12 dossiers pour les grues et 18 pour les pelles).

### 5.1.3. Approche d'annotation sémantique

Pour illustrer notre approche, nous proposons qu'après qu'on a construit l'ontologie. Nous disposons d'un document dont on veut annoter la partie suivante :

- *The EXTCL problem concerns projects of crane products*

Le concepteur basé sur l'éditeur d'ontologies, navigue dans l'ontologie conçue (Figure 5. et tente de vérifier dans la phrase l'existence d'instances de l'ontologie. Nous avons proposé l'utilisation de l'outil ontomat pour aider le concepteur dans sa recherche. Les résultats de cette tâche sont considérés comme une annotation semi-automatique. En appliquant, les phases morpho syntaxiques pour la détection des instances de concepts et de relations comme illustré dans la figure 5. suivante.

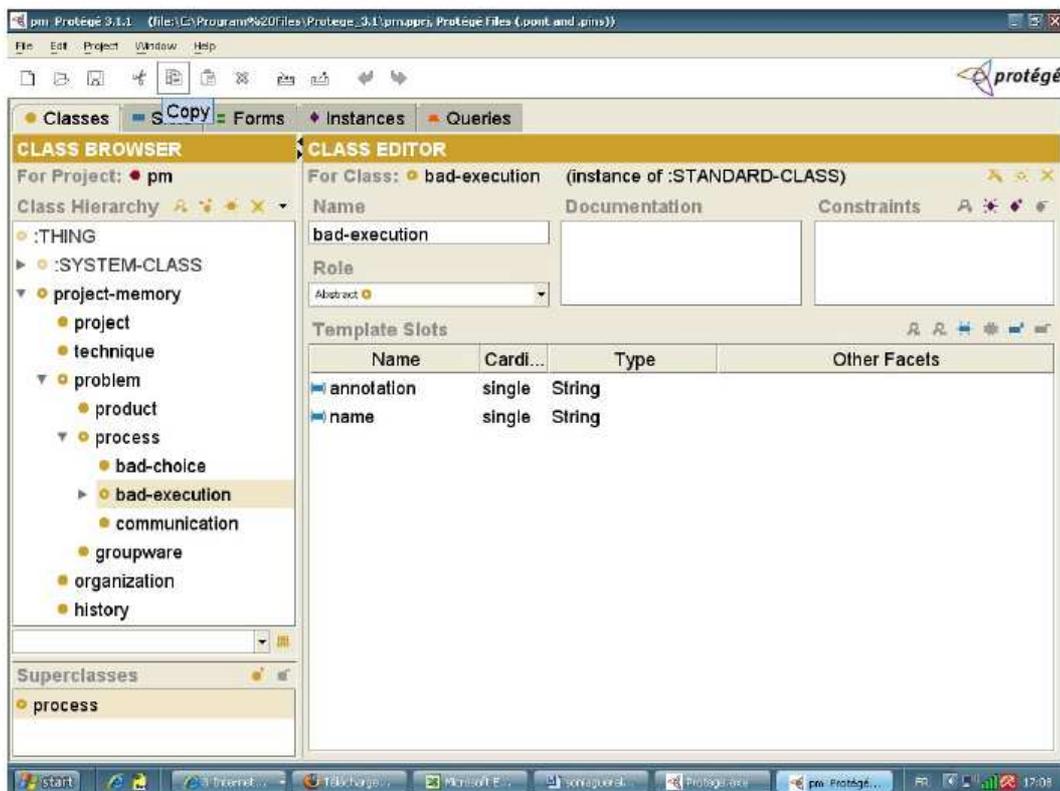


Figure 5.11 Ontologie conçue

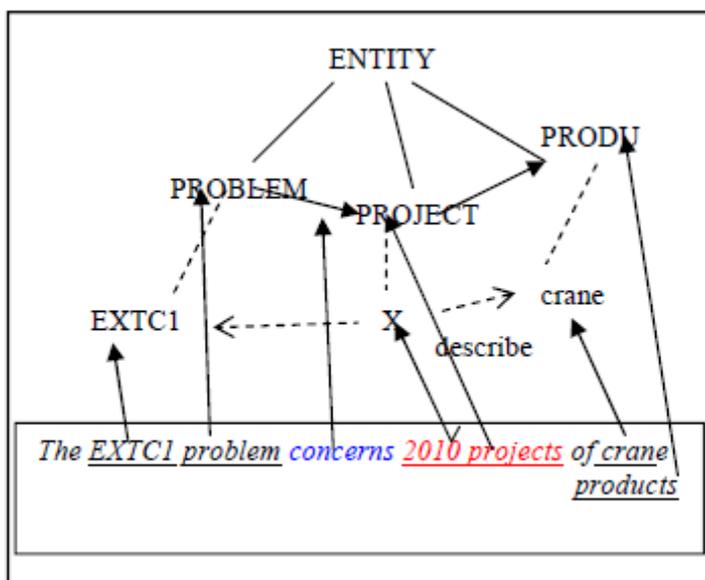


Figure 5.12 Exemple de phrase annotée

Nous avons observé après la vulgarisation de ces techniques entre les acteurs de l'entreprise, un changement concernant l'acceptation des critères ayant lien avec l'impact du web sémantique sur une entreprise. Cette remarque est illustrée par la table suivante.

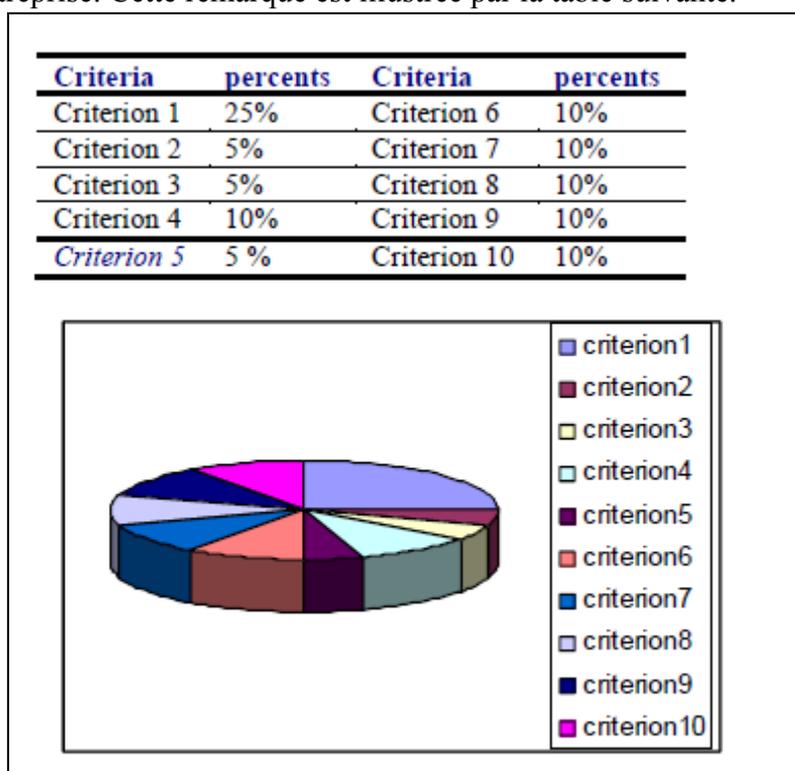


Table 5.7 Evaluation2 à partir des critères

La qualité d'une annotation générée automatiquement dépend de la qualité de la méthode d'extraction d'informations utilisée. Dans notre travail, ceci consiste à évaluer :

- La capacité de l'extracteur des termes à reconnaître toutes les instances de concepts,

- La capacité des grammaires de détection à détecter toutes les instances possibles d'une relation sémantique entre concepts (relier les bons termes par la bonne relation)

Lors du processus de validation, nous avons utilisé un corpus de 30 fiches de projets annotés (1500 mots) manuellement. Cette étape de validation nous a permis de calculer le taux d'erreurs ainsi que le taux de couverture de l'extracteur de termes. Pour ce faire, nous avons utilisé deux mesures fréquemment utilisées dans l'évaluation des systèmes d'extraction d'informations, à savoir la précision et le rappel :

$$\text{Précision} = \frac{\text{Nombre de termes corrects}}{\text{Nombre total des termes extraits}}$$

$$\text{Rappel} = \frac{\text{Nombre de termes corrects}}{\text{Nombre des termes corrects qui auraient dus être extraits}}$$

La précision (P) est le pourcentage des termes correctement extraits, cela mesure donc l'absence de bruit dans l'extraction. Le rappel (R) est le pourcentage des termes extraits par rapport aux termes qui auraient dû être extraits, cela mesure l'absence de silence dans l'extraction. Une autre métrique peut être calculée en combinant les deux mesures, il s'agit de la F-mesure ;

$$\text{F-mesure} = \frac{2PR}{P + R}$$

Si l'annotation porte sur le choix de décision sur les critères, alors ce choix n'est valide que si (i) la décision désigne bien une relation sémantique, (ii) les deux termes (critères) sont bien liés par cette relation.

Donc, nous pouvons utiliser la précision et le rappel pour calculer la qualité des décisions :

La précision est le rapport entre le nombre de décisions évaluées correctes et le nombre total des décisions extraites.

Le rappel est le rapport entre le nombre de décisions évaluées correctes et le nombre des décisions qui auraient du être extraites

Au cours de cette phase, nous avons remarqué que quelques décisions ont été considérées correctes mais inutiles pour les utilisateurs car elles décrivaient soit des connaissances de base, soit des connaissances vagues. Nous avons donc adapté la mesure de qualité nommée utilité pour mesurer le taux des suggestions utiles. L'utilité est le rapport entre le nombre de décisions jugées utiles et le nombre de décisions évaluées correctes.

La valeur de cette mesure peut être subjective car elle est relative à un point de vue d'un utilisateur ou d'un groupe d'utilisateurs. Les annotations jugées inutiles par un utilisateur sont gardées dans la base d'annotations.

Sur le corpus des 30 fiches de projets annotés (1500 mots) manuellement relatifs au projet des pelles et des grues.

Pour la validation des décisions proposées, nous avons choisi au hasard un corpus de 10 documents (1500 mots 100 phrases) annotés manuellement relatifs au projets des pelles et des grues. Parmi les documents proposés par les utilisateurs et nous avons présenté les décisions aux utilisateurs à travers une interface de validation pour qu'ils évaluent leur qualité. Cette interface a été conçue de manière à présenter les annotations dans un format compréhensible (textuel) pour les utilisateurs, qui ne sont pas spécialistes des langages de représentation des connaissances tel que RDF/S .

Le tableau suivant montre les résultats pour calculer la précision, le rappel et la F-mesure

	Crane sheet	Loader sheet
Manual-Annot	25	15
Extracted	18	14
Correctly-extracted	14	8
Recall	0.77	0.80
Precision	0.56	0.53
F-measure	0.64	0.63

Table 5.8 précision, rappel et mesure

Le rappel montre que les termes correctement extraits couvrent 77% des décisions sur les pelles et 80% des grues. Cependant, la colonne Précision montre que la qualité des termes extraits est assez bonne car elle arrive à extraire 56% des décisions sur les pelles et 80% des grues. Les valeurs de la F-mesure montrent aussi que nous avons un compromis assez acceptable entre le rappel et la précision.

## 5.2. Environnements de développement

### 5.2.1. Outils du TALN utilisés

Cette section présente les outils de traitement automatique de la langue naturelle que nous avons utilisés pour la réalisation de ce travail.

## **- GATE**

GATE est une plate-forme d'ingénierie linguistique [Cunningam et al., 2002] qui repose sur l'application successive de transducteurs<sup>35</sup> aux textes. Conformément aux termes employés par ses concepteurs, nous parlons ici de ressources de traitement (Processing Resources : PR). Ces ressources de traitement utilisent le texte<sup>36</sup> modifié par les ressources précédemment appliquées pour ajouter de la structure au texte. Les ressources de traitement les plus courantes sont les segmenteurs (Tokenizers), les analyseurs morpho-syntaxiques (Part Of Speech ou POS Taggers), les lexiques (Gazetteers), les transducteurs (JAPE transducers), et les patrons d'extraction (Templates). Ils sont appliqués au texte au sein d'une cascade (chaîne de traitement ou pipeline).

GATE peut être utilisé de deux façons différentes : environnement de développement ou bibliothèque.

L'utilisation la plus simple est comme environnement de développement au travers des ressources développées par ses concepteurs. L'environnement de développement est constitué d'une interface graphique permettant aux utilisateurs de créer de nouvelles ressources ou paramétrer des ressources disponibles et de les appliquer aux textes au sein d'une chaîne de traitement. Nous avons utilisé cette interface pour tester nos modules d'extraction de termes et de relations et pour visualiser les résultats.

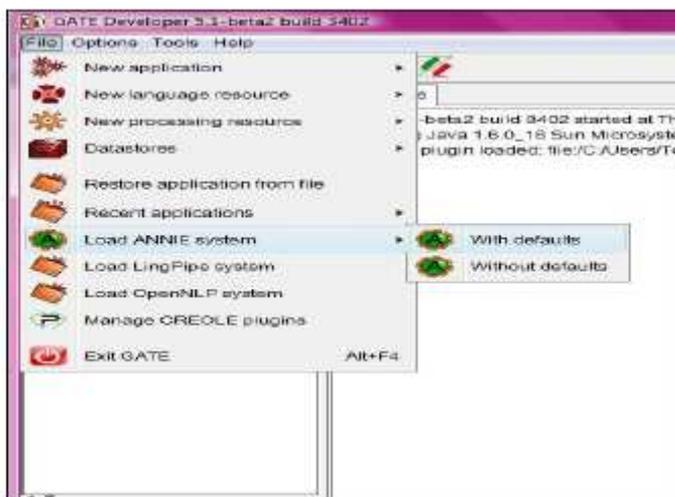
Le second niveau d'utilisation consiste à tirer parti des ressources disponibles dans le système. Nous pouvons les utiliser au sein de l'environnement de développement ou de façon embarquée dans des applications autonomes. De cette manière, il est possible de se passer de l'interface graphique de GATE et de traiter un texte dans un programme autonome hors de l'environnement de développement. C'est l'utilisation pour laquelle nous avons opté, nous avons ainsi bénéficié de l'architecture et de la bibliothèque de GATE afin d'intégrer nos différents modules et afin de les appliquer en chaîne sur les textes.

L'architecture Gate comporte un système d'extraction d'information qui s'appelle ANNIE que nous avons utilisé.

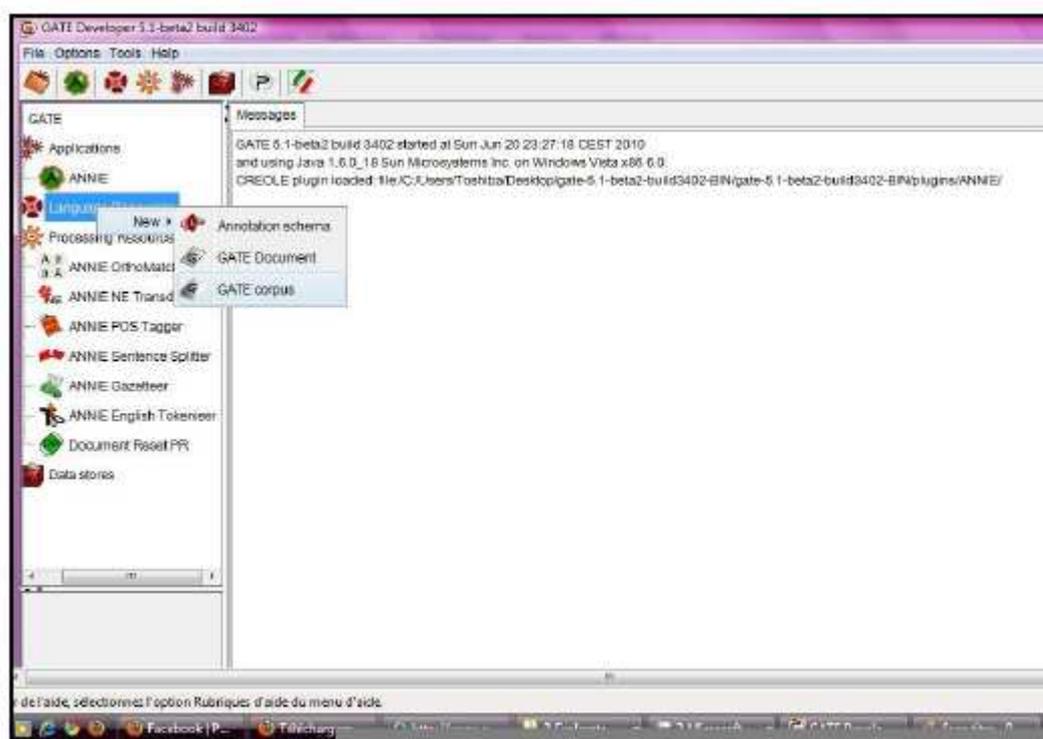
## **- ANNIE**

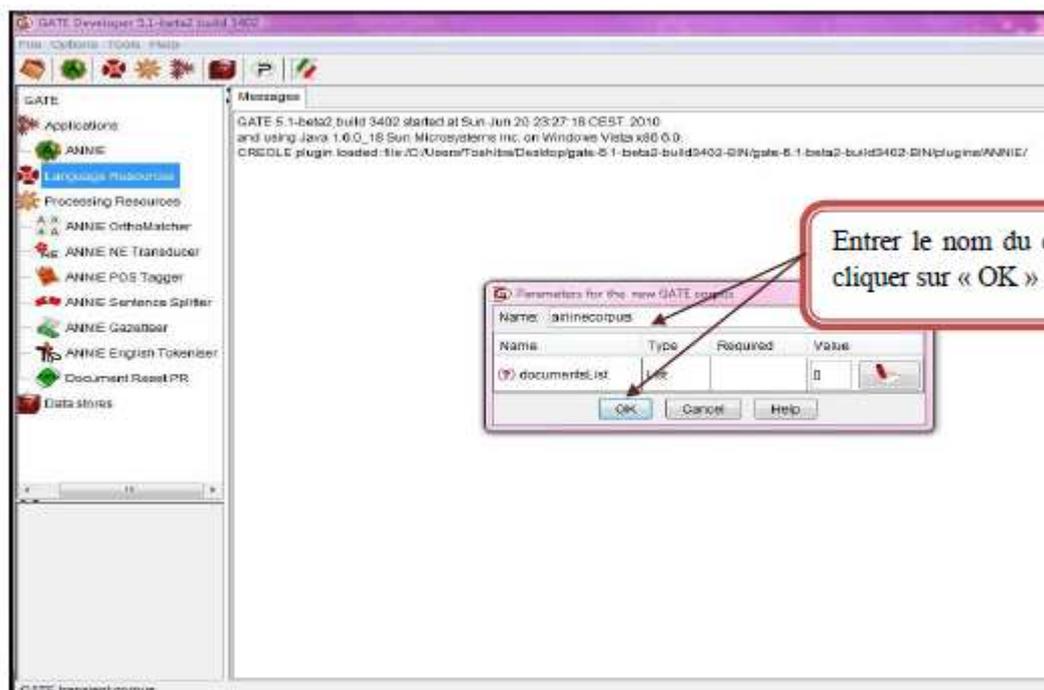
ANNIE (A Nearly-New Information Extraction System, pour système quasi nouveau pour l'extraction d'information), est un composant de Gate, lui-même formé de modules parmi lesquels un analyseur lexical, un gazetteer, un segmenteur de phrases (avec désambiguïsation), un étiqueteur, un module d'extraction d'entités nommées et un module de détection de coréférences. ANNIE offre toute la gamme de Processing Ressources nécessaires au dépistage d'information sur les textes (Information Extraction). Et il offre aussi entre autre les outils pour le traitement de phrases, pour la détection des entités et pour la détection de références entre les sections d'un texte. Les étapes de la création d'une annotation avec ANNIE : sont résumées dans les points suivants :

## 1. Chargement de l'outil ANNIE

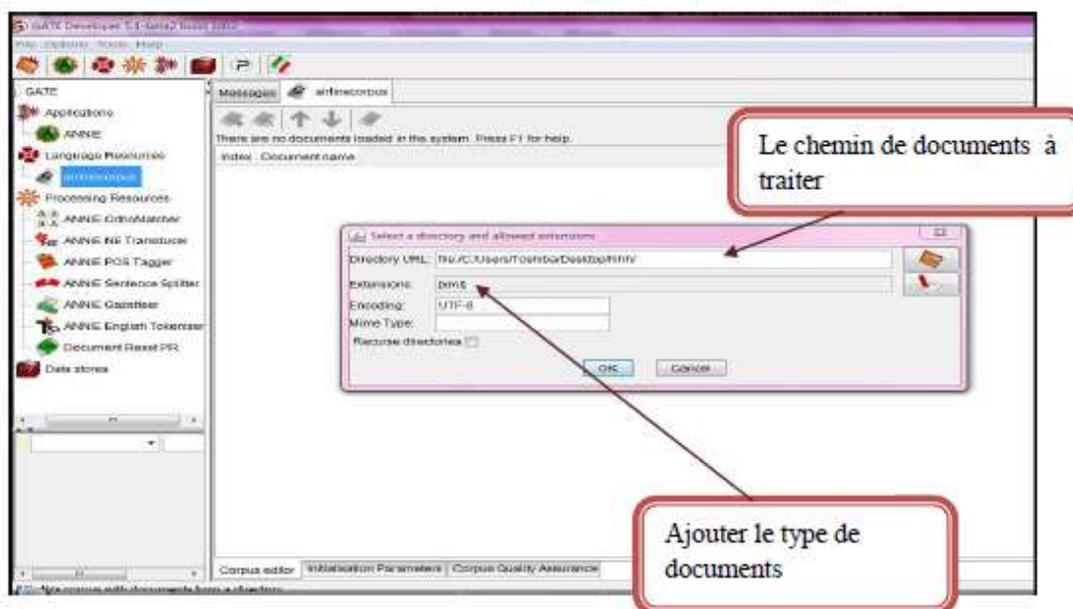


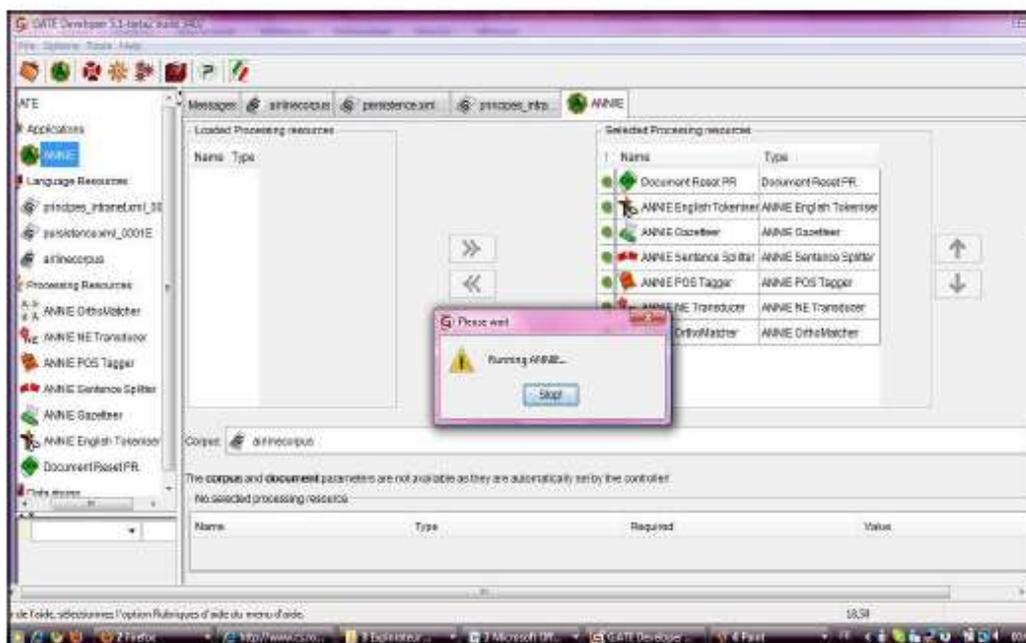
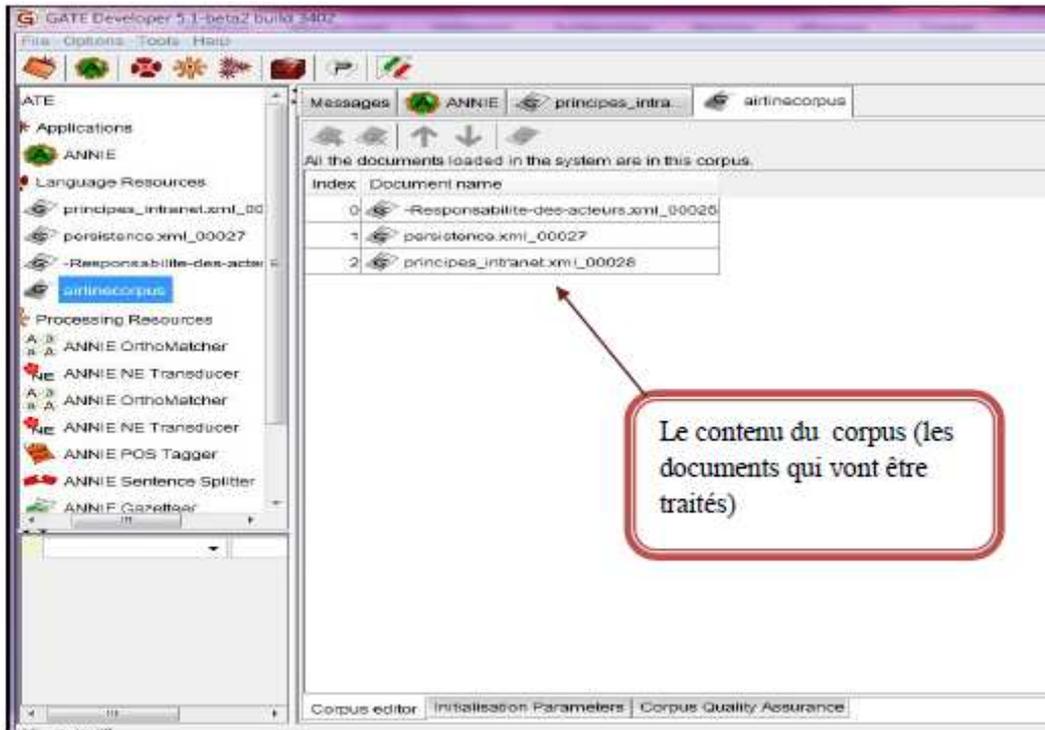
## 2. Création d'un corpus de documents (.xml)

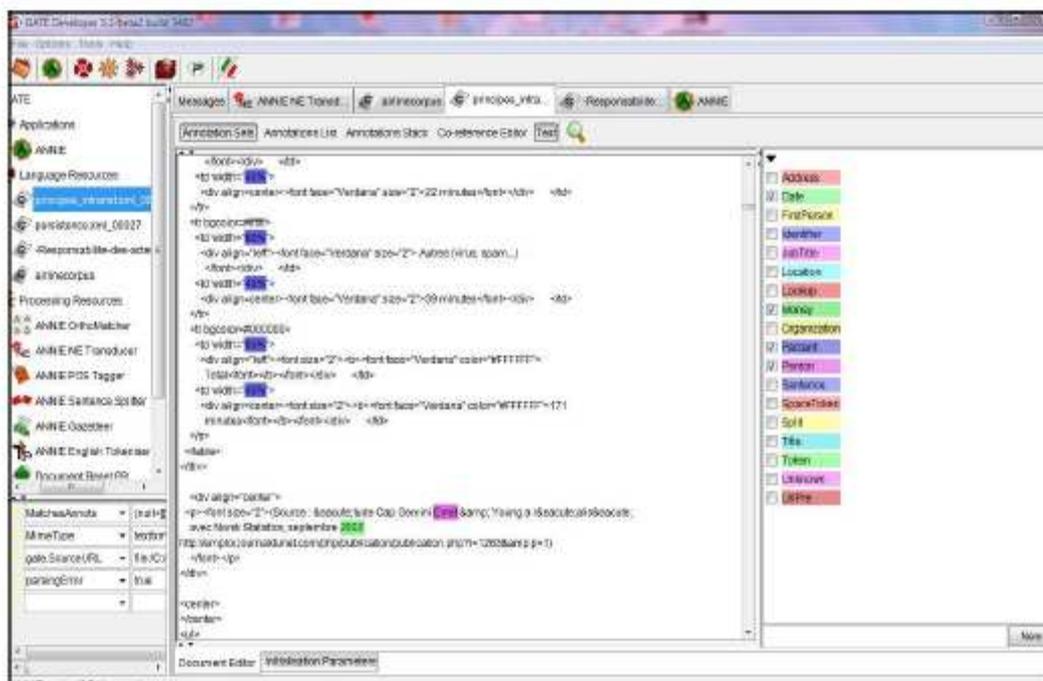
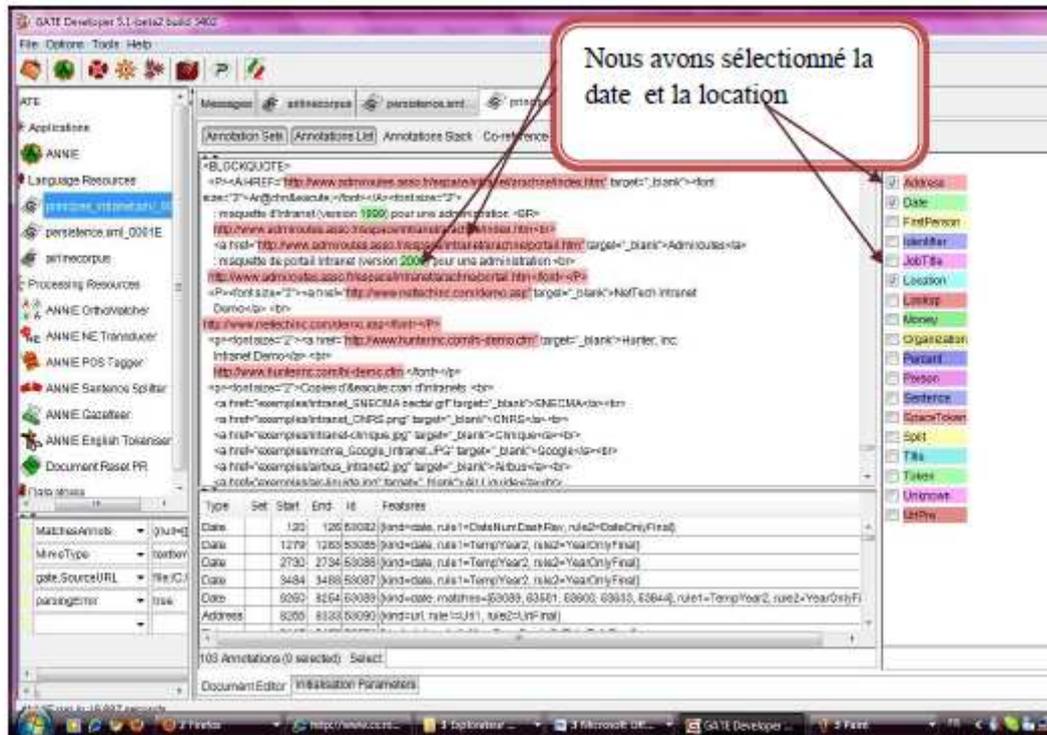




### 3. Peupler le corpus







## - TreeTagger

TreeTagger est un outil développé au sein l'institut de linguistique computationnelle de l'université de Stuttgart [Schmid, 1994]. Cet outil, utilisé pour l'étiquetage des textes en français et en anglais, spécifie pour chaque mot sa catégorie syntaxique et indique son lemme. L'estimation de la catégorie grammaticale d'un mot se base sur la construction récursive d'arbres de décisions binaires et un

calcul de probabilité. Le Tableau 6 représente un exemple de sortie pour l'entrée suivante "The project is difficult to realize." :

Mot	Catégorie	Signification	Lemme
The	DT	Article	The
TreeTagger	NP	Nom propre	TreeTagger
Is	VBZ	Verbe conjugué au présent	Be
Easy	JJ	Adjectif	Easy
to	TO	TO : catégorie spéciale	to
Use	VB	Verbe à l'infinitif	Use
.	SENT	Ponctuation	.

Nous avons intégré cet outil dans la plate-forme GATE par le biais d'un traducteur (wrapper) que nous avons développé, et nous l'avons utilisé pour l'étiquetage grammatical et la lemmatisation des textes. Les informations concernant chaque mot sont intégrées dans la structure XML du document texte. Notons ici que GATE propose son propre étiqueteur grammatical mais ce dernier ne calcule pas la forme lemmatisée des mots ( information indispensable dans notre processus).

### - RASP

RASP (Robust Accurate Statistical Parsing) [Briscoe et Carroll, 2002] est un analyseur probabiliste pour l'anglais, il permet par des calculs statistiques de prédire la relation grammaticale (sujet, objet...) de chaque mot dans le texte. Il fournit comme résultat une forêt de mots interconnectés par des probabilités.

Considérons la phrase « RASP predicts grammatical relations. », le résultat de RASP sur cette phrase est le suivant :

```

("RASP" "predicts" "grammatical" "relations.")
(|nsubj| |predict+s:2_VVZ| |RASP:1_NN1| |_)
(|dobj| |predict+s:2_VVZ| |relation.+s:4_NN2| |_)
(|ncmod| |_ |relation.+s:4_NN2| |grammatical:3_JJ|)

```

Un exemple de grammaire est donné en annexe2.

### 5.3.2. Outils de développement

Dans la partie programmation du système, nous avons utilisé le langage JAVA version JDK 5.0. Ce langage est le plus adéquat, vu que les outils que nous utilisons sont entièrement développés en JAVA. De plus, la plupart de ces outils ont des problèmes de compatibilité avec les autres langages dans leur version actuelle.

De plus, nous avons choisi ce langage pour ses qualités, à savoir :

- C'est un langage orienté objet simple,
- Java est indépendant de toute plate-forme : grâce à sa machine virtuelle « Java Virtuelle Machine JVM »,
- Il fournit aux programmeurs de nombreuses bibliothèques pour la gestion des interfaces et des programmes et il permet la gestion des exceptions,
- Il est compatible avec L'API JENA version 2.5.5, ce qui nous a permis la manipulation, le parcours et la modélisation des documents OWL.

Java est exploité par Eclipse. Nous avons opté pour Eclipse version 3.3.1.1 qui est un environnement de développement intégré (Integrated Development Environment), développé par IBM, dont le but est de fournir une plate-forme modulaire pour permettre de réaliser des développements informatiques. Eclipse utilise énormément le concept de modules nommés "plug-ins" dans son architecture (hormis le noyau de la plate-forme nommé "Runtime", tout le reste de la plate-forme est développé sous la forme de plug-ins). Ce concept permet de fournir un mécanisme pour l'extension de la plate-forme, ainsi il fournit la possibilité à des tiers de développer des fonctionnalités personnalisées.

Les principaux modules de base d'Eclipse fournis concernent le langage de programmation Java. Les modules agissent sur des fichiers qui sont inclus dans l'espace de travail (appelée workspace). L'espace de travail regroupe les projets qui contiennent une ou plusieurs arborescences de fichiers. Bien que développé en Java, Eclipse présente de très bonnes performances d'exécution, car il n'utilise pas Swing pour l'interface homme-machine, mais il utilise un toolkit particulier nommé SWT (Standard Widget Toolkit) associé à la bibliothèque JFace. SWT est développé en Java, en utilisant au maximum les composants natifs fournis par le système d'exploitation sous-jacent. JFace utilise SWT et propose une API pour faciliter le développement des applications nécessitant des interfaces graphiques.

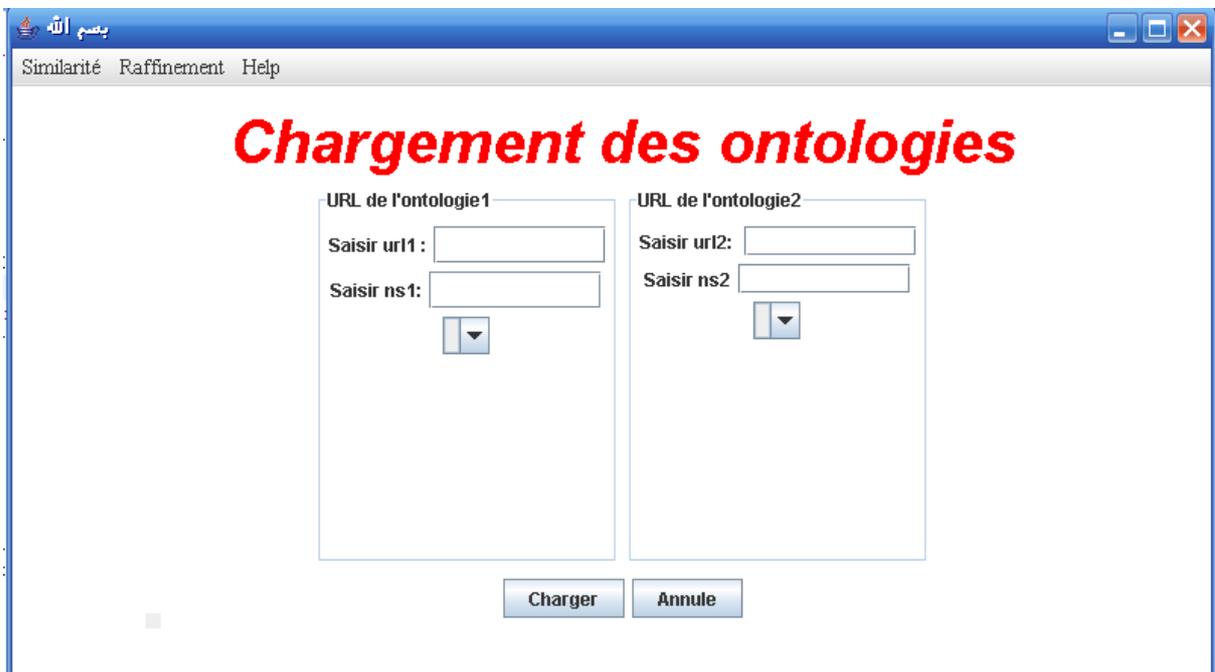
### 5.3.2. Outil de création des ontologies

Protégé-2000 version 3.4.4 est un éditeur qui permet de construire une ontologie pour un domaine donné, de définir des formulaires d'entrée de données, et d'acquies des données à

l'aide de ses formulaires sous forme d'instances de cette ontologie. Protégé est également une librairie Java qui peut être étendue pour créer de véritables applications à bases de connaissances en utilisant un moteur d'inférence pour raisonner et déduire de nouveaux faits par application de règles d'inférence aux instances de l'ontologie et à l'ontologie elle-même (méta-raisonnement).

L'ontologie est implémentée en langage OWL-DL. Ce langage a été choisi pour son pouvoir d'expression, sa décidabilité et ses mécanismes d'inférences (raisonneur racer) Pour le passage à l'ontologie opérationnelle, il existe une correspondance directe entre la logique de description et le langage owl-dl. Nous avons utilisé le raisonneur racer pour détecter les inconsistances, les redondances, les dépendances cachées et les erreurs de classification. Les fichiers OWL sont inexploitable en état brut, c'est-à-dire la structure du fichier OWL est très complexe. Donc, pour pouvoir l'exploiter il nous a fallu un « traducteur » capable de traduire les balises la sémantique véhiculée par le fichier OWL en objet manipulable par des programmes. L'outil disponible qu'on a peut avoir c'est l'API (Application Programming Interface) JENA. Cet outil est développé par une équipe de la firme HP (Hewlett Packard) dans le cadre du Projet HP « Labs Semantic Web Programme» qui a pour but de réaliser un outil d'exploitation des fichiers OWL. JENA est développé entièrement en Java, elle donne aux programmes la possibilité d'exploiter du contenu des fichiers RDF et OWL (extraction du contenu sémantique de ces derniers).

Pour exécuter l'application l'utilisateur doit lancer en premier temps le serveur web apache Tomcat version 5.5 pour pouvoir se connecter à l'ontologie.



Pour charger les ontologies nous allons remplir les ontologie. Après le chargement des ontologies nous allons calculer les mesures de similarité. Cette étape comprend cinq sous étapes : similarité des concepts, similarité des attributs, similarité formulaires par l'URI et le Namespace de chaqu'une des relations, similarité structurelle et la similarité globale.



Similarité de concepts : La figure suivante montre les valeurs de similarités entre les concepts.

théorique	domaine_travail	-0.7212840975020888
théorique	grade	0.3640945276635167
théorique	sexe	-0.7192695522088766
role_fonctionnelle	role	0.7218133048757093
role_fonctionnelle	taille_equipe	0.12782532580983488
role_fonctionnelle	role_fonctionnelle	0.608573620475813
role_fonctionnelle	domaine_travail	0.12782532580983488
role_fonctionnelle	grade	0.36069716352201686
role_fonctionnelle	sexe	0.12787453226005857
domaine_travail	role	0.7216304643362598
domaine_travail	taille_equipe	0.12787453226005857
domaine_travail	role_fonctionnelle	0.1085878369679821
domaine_travail	domaine_travail	0.6278111093176657
domaine_travail	grade	0.3608456363961255
domaine_travail	sexe	0.12791551665780512
grade	role	0.7464699786106903
grade	taille_equipe	-0.7208817360132543
grade	role_fonctionnelle	2.3403299775953965E-5
grade	domaine_travail	-0.721175644270597
grade	grade	0.860673760222241
grade	sexe	-0.7060867139253748
sexe	role	0.025122458116620827
sexe	taille_equipe	0.1561005115866207
sexe	role_fonctionnelle	0.12787453226005857
sexe	domaine_travail	0.15543814098004236
sexe	grade	0.7366083269635888
sexe	sexe	0.655333733639903
domaine_recherche	role	0.7214519277846214
domaine_recherche	taille_equipe	0.1278345126174417
domaine_recherche	role_fonctionnelle	0.10857886641388276

NB : l'ensemble des valeurs  $< 0$  sont obtenus à cause du manque des instances dans les concepts. Similarité des attributs : Les résultats de la similarité entre les attributs sont montrés dans la figure suivante.

théorique	domaine_travail	9.315688624544426E-4
théorique	grade	4.744203352035881E-4
théorique	sexe	0.006841534882751672
role_fonctionnelle	role	0.005694716860882406
role_fonctionnelle	taille_equipe	3.43752347769167E-4
role_fonctionnelle	role_fonctionnelle	1.0
role_fonctionnelle	domaine_travail	0.011262587836979141
role_fonctionnelle	grade	0.005669820612267945
role_fonctionnelle	sexe	0.011262587836979141
domaine_travail	role	0.5000385268937784
domaine_travail	taille_equipe	0.0015335558934354298
domaine_travail	role_fonctionnelle	0.011262587836979141
domaine_travail	domaine_travail	1.0
domaine_travail	grade	0.5000234032997759
domaine_travail	sexe	0.01854055953225703
grade	role	0.500828707940272
grade	taille_equipe	0.0016106092809921773
grade	role_fonctionnelle	0.01133964122453599
grade	domaine_travail	1.0000468085995518
grade	grade	1.0000468085995518
grade	sexe	0.0186674054170427
sexe	role	0.009374687106267909

- Similarité des relations : les valeurs de similarité des relations sont montrés par la figure suivante.

bilal			
calculer		Retour	Quitter
			Similarité de Relation
théorique	domaine_travail		0.0013151940020787146
théorique	grade		0.0013151940020787146
théorique	sexe		0.0
role_fonctionnelle	role	0.0	
role_fonctionnelle	taille_équipe	0.0	
role_fonctionnelle	role_fonctionnelle		0.0
role_fonctionnelle	domaine_travail		0.0
role_fonctionnelle	grade	0.0	
role_fonctionnelle	sexe	0.0	
domaine_travail	role	0.0	
domaine_travail	taille_équipe	0.0010346862084581975	
domaine_travail	role_fonctionnelle		0.0
domaine_travail	domaine_travail		0.0
domaine_travail	grade	0.0	
domaine_travail	sexe	0.0	
grade	role	0.0	
grade	taille_équipe	0.0010346862084581975	
grade	role_fonctionnelle		0.0
grade	domaine_travail		0.0
grade	grade	0.0	
grade	sexe	0.0	
sexe	role	0.0	

## 6. Conclusion

Dans ce chapitre, nous avons décrit un processus d'annotation sémantique. Cette approche est une continuité du modèle qui supporte la gestion d'une mémoire d'entreprise et plus spécialement une mémoire de projet. Ce modèle se compose de trois catégories de pétales. La première pour prendre en considération le cycle de vie dans le cadre d'une gestion de connaissance. La seconde concerne l'interaction avec l'environnement et la troisième supporte la capitalisation. Au cœur, se situe le patrimoine des connaissances cruciales. Chaque pétale est un processus.

Cette approche utilise son propre modèle et suppose un ensemble de critères qui permettent de cerner un ensemble de condition qui conduisent l'annotation.

Pour être exploitée, une partie étude de cas a été présentée, où nous avons présenté l'ensemble des outils et environnements utilisés pour décrire l'annotation sémantique dans un contexte réel. Le mariage de ces outils montre qu'il est possible d'arriver à des fins encourageantes dans des entreprises à caractère industriels.

## ***Conclusion générale***

# Conclusion générale

## 1. Bilan

Le contexte et la problématique de gestion des connaissances peuvent être abordés de divers points de vue. Il est possible de faire une étude précise des besoins, au travers d'une analyse du système de connaissance ou bien de mettre directement en place des solutions génériques qui seront adaptées aux besoins par la suite. Nous avons opté pour le premier choix, où nous avons procédé à une étape de capitalisation des connaissances, c'est-à-dire tout d'abord d'une identification des connaissances à gérer, puis une explicitation et une modélisation (plus ou moins formelle) et enfin la constitution d'une mémoire d'entreprise (en structurant les connaissances modélisées).

En termes de solutions opérationnelles de gestion des connaissances, il existe des solutions orientées information et des solutions orientées connaissances [Caussanel et Choraqui, 99]. La première catégorie ne s'intéresse pas à la formalisation des connaissances à traiter, mais se charge de gérer les informations support à la connaissance. La majeure partie des outils du marché relèvent de ce type de solution. Cependant, pour des besoins spécifiques, il peut être nécessaire d'adapter ces outils ou de leur adjoindre des fonctions supplémentaires. La seconde catégorie se concentre sur la représentation et la formalisation des connaissances dans un but d'exploitation plus complexe que pour les solutions orientées information. Il peut s'agir d'offrir des outils de recherche sémantique sur les connaissances, mais également d'outils permettant l'appropriation de ces connaissances par les collaborateurs (Ingénierie des Connaissances, l'Extraction de Connaissance, etc.). De ce fait, les outils opérationnels utilisés sont en général plus spécifiques aux besoins de l'entreprise et plus adaptés à ses caractéristiques.

L'approche proposée, facilite le développement d'une mémoire d'entreprise via des moyens offerts par le web sémantique, tels que l'annotation sémantique et les ontologies. Elle fait appel à un concept organisationnel traditionnel qui est le processus. Elle met en œuvre les étapes nécessaires à la modélisation de la mémoire. Elle offre un moyen de construire une ontologie pour l'entreprise et elle permet enfin d'exécuter un ensemble d'étapes d'annotations en fonction des composants de la mémoire qui se produisent.

La gestion de la mémoire d'entreprise que nous proposons, au niveau de cette thèse, est basée sur trois principes :

Le premier consiste à définir la mémoire d'entreprise comme une collection dynamique de processus définis en phase de conception et pouvant être dynamiquement ré-orchestrés en fonction du contexte en cours, au moment de l'exécution. Avec un avantage majeur de réutilisation de connaissances dans l'organisation [Gueraich and Boufaïda, 2008].

Le second principe est le fondement de la gestion sur la notion de mémoire de projet. L'utilisation de la notion de caractéristiques apporte une abstraction du processus de modélisation tout en ignorant totalement les considérations du niveau opérationnel [Guerich and Boufaïda, 2009]. Ceci va permettre une projection directe des objectifs organisationnels de l'organisation.

Le troisième principe est le choix de l'approche d'annotation pour exploiter la mémoire d'entreprise. L'annotation sémantique utilise une ontologie de domaine dédiée à l'entreprise. Cette ontologie est considérée comme un modèle qui permet, en plus des représentations formelles et consensuelles des connaissances du domaine, un accès unique aux ressources de connaissances via une terminologie partageable et non ambiguë tout en offrant des mécanismes de raisonnement sur les connaissances modélisées. L'approche d'annotation proposée fournit une représentation de la démarche basée sur des acteurs et des extractions. Il va nous permettre de capturer la notion d'annoter utilisée dans la définition de la gestion d'une ME. L'architecture CMOCA supportant la démarche s'applique aussi bien aux entreprises de production de biens manufacturés (production continue ou discrète) qu'à celles fournissant des services. Les projets modélisés peuvent correspondre à des procédures administratives, techniques (processus productifs) ou de support de l'entreprise. Ces projets peuvent être bien ou mal définis.

Déployée efficacement, la gestion d'une ME à la base d'une approche annotation offre des avantages considérables, tels que :

- La création plus rapide et à moindre frais des projets via la réutilisation des autres,
- L'implication des utilisateurs fonctionnels dans la conception de la mémoire,
- Un contexte plus favorable pour le contrôle des projets,
- Une souplesse permettant de résoudre les problèmes au moment où ils se produisent, plutôt qu'après,
- La possibilité d'adaptation dynamique aux nouvelles conditions de fonctionnement.

L'architecture CMOCA tente de tenir l'une des promesses les plus importantes des technologies de WSE.

## 2. Perspectives

Au terme de ce travail qui peut être vu comme un début et non comme une fin, nous pouvons proposer un nombre de perspectives. Ces dernières peuvent venir compléter, améliorer, voir étendre ce modeste travail. Parmi ces perspectives, on peut citer :

- Etudier plus amplement les aspects d'évolution des connaissances et à travailler la partie restitution des connaissances et interfaces utilisateurs. Dans ce travail, nous avons axé nos recherches sur les aspects de modélisation et de représentation des connaissances actuelles.
- Mettre en place un dispositif qui permettrait l'intégration des connaissances cruciales par les employés eux-mêmes d'une façon systématique. Nos premières réflexions à ce sujet s'orientent vers l'élaboration d'un modèle de signalement des connaissances réutilisables. La notion de connaissances réutilisables signifie que certaines connaissances possèdent une importance particulière pour une tâche donnée par exemple, et que par conséquent, ces connaissances sont susceptibles d'être réutilisées.
- Intégrer d'autres mécanismes de raisonnement, tels que le mécanisme des règles par exemple. En effet, le raisonnement à base de règles offre un cadre déclaratif pour exprimer des connaissances procédurales en permettant de voir clairement les conditions dans lesquelles une règle est applicable. De plus, le fait que la connaissance soit exprimée de façon uniforme par des règles et des faits, ces règles et faits étant indépendants, la base est modulaire et facilement modifiable.
- Intégrer la notion de contexte afin de représenter par exemple les aspects primordiaux du contexte d'un repérage des connaissances tels que les tâches professionnelles et les thèmes choisis par l'utilisateur. Dans le contexte de l'entreprise un thème peut être : une méthode de conception, un langage de programmation etc. Le modèle devrait permettre la description d'un nouveau contexte d'utilisation ainsi que les conditions dans lesquelles seront exploitées les connaissances repérées.

# ***Références bibliographiques***

# Références bibliographiques

[Abecker et al., 1998] Abecker. A, Bernardi. A, Hinkelmann. K, Kuhn. O, and Sintek. M. “ Toward a technology for organizational memories”. IEEE Intelligent Systems, May/June 1998.

[Abowd et al., 1999] Abowd. G. “Towards a better understanding of context and context-awareness”. International Symposium on Handheld and Ubiquitous Computing, Karlsruhe, Germany. 27-29 Septembre 1999.

[Adler and Van Doren, 1972] Adler, M. et Van Doren. C. “How to read a book”. New York, Simon and Schuster. 1972.

[Adobe-Systems-Incorporated, 2004] Adobe-Systems-Incorporated. Adobe® Acrobat® 6.0 Professional software. San Jose, California USA. Année Url du document: www.adobe.com

[Alavi et al., 1999] Alavi. M et D. Leidner. “Knowledge Management Systems: Issues, Challenges, and Benefits”. Communications of the Association for Information Systems. 1 (7). 1999.

[Alexander et al., 1977] Alexander. C, Ishikawa. S et Silverstein. M. , ‘A Pattern Language: Towns, Buildings, Construction’. Oxford, Oxford University Press. 1977.

[Amardeilh et al., 2009] Amardeilh. F, Damljanovic.D et Bontcheva. K , “ CA Manager: a Framework for Creating Customised Workflows for Ontology Population and Semantic Annotation”. Semantic Authoring, Annotation and Knowledge Markup Workshop, LosAngeles. 2009.

[Ambler, 1999] Ambler, S. More Process Patterns: Delivering Large-Scale Systems Using *Object Technology*, Cambridge University Press. 1999.

[Atilf, 1992] Atilf “ Le trésor de la langue française”. Paris, France, 1992.

[Aussenac-Gilles, 2000] Aussenac-Gilles. N, Biebow. B et Szulman. S “ Revisiting Ontology Design: a Method Based on Corpus Analysis”. In Proc. EKAW'2000, Juan-les-Pins, 2000, pp.172-188.

[Austin, 1962] Austin, J. L. How To Do Things With Words, Oxford University Press. 1962.

[Azouaou, 2004] Azouaou, F. Spécification de l'état initial, Document interne. Projet MemoNote. Grenoble, CLIPS-IMAG. 2004

[Azouaou et al., 2004] Azouaou, F., Chen, W. et Desmoulins, C. Semantic Annotation Tools for Learning Material. Semantic Web and e-learning Workshop. Adaptive Hypermedia (AH'04), pp. 359-364, Eindhoven University of Technology, Pays-bas. 2004.

[Azouaou et al., 2005] Azouaou, F., Chen, W. et Desmoulins, C. Semantic Annotation for Learning Material: specification and categorization. CAiSE'05 (The 17th Conference on Advanced Information Systems Engineering), pp. 645-660, Porto, Portugal, FEUP edicoes. 2005.

[Azouaou, 2006] Azouaou, F., Desmoulins, C. "Teachers' Document Annotating: Models for a Digital Memory Tool." Int. J. Cont. Engineering Education and Lifelong Learning (IJCEELL). 16 (1/2). pp.18-34. 2006.

[Bachimont et al., 2000] Bachimont, B. et Charlet, J. Ingénierie des connaissances - Gestion des connaissances (cours). Année d'accès: 2006. Url du document: [http://formation.enst.fr/SIMAN/supports\\_siman/IC\\_GestiondC04.pdf](http://formation.enst.fr/SIMAN/supports_siman/IC_GestiondC04.pdf)

[Baget et al., 2004] Baget, J.-F., Canaud, É., Euzenat, J. et Saïd-Hacid, M. "Les langages du Web sémantique." Revue I3 (Information - Interaction - Intelligence), Hors Série: Web Sémantique. 5 (1). Cépaduès-Éditions. pp.9-24. 2004.

[Baker et al., 1995] Baker, M. et K., B. Missed opportunities for learning in collaborative problem-solving interactions. Artificial Intelligence in Education, pp. 210-217, Washington, D.C. August 1995.

[Bahloul, 2006] Bahloul.D , "Une approche hybride de gestion des sur les ontologies connaissances basée: application aux incidents informatiques," Thesis In Informatics, Ecole doctorale : Informatique et Information pour la société (EDIIS), Lyon, 2006.

[Baldonado et al., 2000] Baldonado. M, Cousins. S, Gwizdka. J. et Paepcke. A. "Notable: At the Intersection of Annotations and Handheld Technologies". Bristol. 2000, HUC, pp. 100-113.

[Barth, 2000] Barth, S. The power of one. Knowledge Management. Année d'accès: 2005. Url du document: <http://www.destinationkm.com/print/default.asp?ArticleID=615>

[Barthès, 1996] Barthès. J.-P. "ISMICK and Knowledge Management. In J. F. Schreinemakers ed, Knowledge Management: Organization, Competence and Methodology". Proceedings of ISMICK'96, Rotterdam, the Neth., Wurzburg:Ergon Verlag, 21-22 Octobre 1996, pp. 9-13.

[Barthès et al., 1999] Barthès, J., P, Dieng, R. et Kassel, G. "Dossier Mémoire d'entreprise." *Le bulletin de l'AFIA. Association Française pour l'Intelligence Artificielle.* n°36.1999.

[Baumard, 1999] Baumard, P. Tacit Knowledge in Organizations. London, UK, Sage Publications Ltd. 1999.

[Bekhti, 2003] S. Bekhti, "DYPKM : Un Processus Dynamique de Définition et de Réutilisation de Mémoire de Projet," Thèse de l'UTT, spécialité Réseaux, Connaissances et organisations, Décembre 2003.

[Bensta et al., 2005] BenSta. H, BenSaid. L, K. Ghédirak, Bigand. M and Bourey. J.P, "Cartographies of ontology concepts," International Conference on Enterprise Information System, Miami, 2005.

[Berners-Lee et al., 2001] Berners-Lee, T., Hendler, J. et Lassila, O. "The Semantic Web." Scientific American. 5. pp.34-43. 2001.

[Blackler, 1995] Blackler . K. " Building a large Knowledge Base for Machine Translation". Proceedings. Amer. Assoc. Artificial Intelligence Conf. (AAAI-94). AAAI Press, , Melon Park, California. 1994, pp. 773-778.

**[Bouvin et al., 2002]** Bouvin, N. O., Zellweger, P. T., Grønbaek, K. et Mackinlay, J. D. Fluid annotations through open hypermedia: using and extending emerging web standards. eleventh international conference on World Wide Web, SESSION: Hypermedia in the Small table of contents, pp. 160-171, Honolulu, Hawaii, USA, ACM Press. New York, NY, USA. 2002.

**[Brince, 1995]** Brice. A “Design rationale management (DRAMA), 1995, <http://www.enviros.com/drama>.

**[Bringay et al., 2003]** Bringay, S., Barry, C. et Charlet, J. "Les documents et les annotations du dossier patient hospitalier." Information - Interaction - Intelligence. 4 (1). pp.191-211. 2003.

**[Briscoe and Carroll, 2002]** Briscoe. E. et Carroll. J. “Robust accurate statistical annotation of general text”. In Proceedings of the Third IC LR E, Las Palmas, Gran Canaria, 2002, pp. 1499-1504.

**[Buckingham, 1996]** Buckingham .S. “Analysing the usability of a design rationale notation”. Design rationale, concepts, techniques and use, Laurence Erlbaum, Hillsdale, 1996, pp. 185-215.

**[Buckingham, 1997]** [Buckingham .S. “Representing hard to formalize contextualized multidisciplinary knowledge”. Proceedings of AAAI Spring symposium on artificial intelligence in knowledge management, 1997, pp. 9-16.

**[Caroll, 1997]** Caroll. J.M “Scenario-Based Design”. Chapter 17 "Handbook of Human-Computer Interaction" Second, completely revised edition, M. Helander, T.K. Landauer, P. Prabhu (eds), Elsevier Science B.V, 1997.

**[Caussanel et al., 2002]** Caussanel, J., Cahier, J.-P., Zacklad, M. et Charlet, J. Cognitive Interactions in the Semantic Web. SemantivWeb, Hawai. 2002.

**[Chabert-Ranwez, 2000]** Chabert-Ranwez, S. Composition Automatique de Documents Hypermédia Adaptatifs à partir d'Ontologies et de Requêtes Intentionnelles de l'Utilisateur. Ecole Doctorale : Information, Structures, Systèmes, Montpellier, France.p. 189. Sous la direction de M. Crampes. 2000.

**[Charlet et al., 1996]** Charlet. J, Aussenac-Gilles. N, Bachimont. B et Philippe. L. Cours de DEA/DESS sur l'ingénierie des connaissances développé par Jean Charlet et Nathalie Aussenac-Gilles avec la participation de Bruno Bachimont et de Philippe Laublet.

**[Champin et al., 2003]** Champin, P.-A. et Prié, Y. Semantic Web. Annotation for the Semantic Web. S. Handschuh, IOS Press, Amsterdam. pp.180-190 2003. MUNETTE: uses-based annotation for the

**[Charlet et al., 2003]** Charlet, J., Laublet, P. et Reynaud, C. Web sémantique. Rapport final. Action spécifique 32 CNRS / STIC. J. Charlet, P. Laublet et C. Reynaud. p. 143. 2003.

**[Ciravegna et al., 2002]** Ciravegna, F., Dingli, A., Petrelli, D. et Wilks, Y. Timely and Non-ECAI 2002. Lyon, France. 2002.

**[Corby et al., 2004]** Corby, O, Dieng-Kuntz. R et Faron-Zucker. C. “Querying the Semantic Web with the CORESE search engine”. Proceedings. of the 16th European Conference on Artificial Intelligence (ECAI’2004), In R. Lopez de Mantaras and L. Saitta eds, Valencia, 22-27 August 2004, IOS Press, pp. 705-709.

**[Corcho et al., 2005]** Corcho, O., et al. Combination of DROOL rules and Protégé knowledge bases in the ONTO-H annotation tool. Workshop Protégé with Rules in the 8th Intl. Protégé Conference, Madrid, Spain. July 18-21 2005.

**[Cunningham et al., 2002]** H. Cunningham., D. Maynard, K. Bontcheva and V. Tablan, “GATE: A Framework and Graphical Development Environment for Robust NLP Tools and Applications,” ACL'02.F, 2002.

**[Davis and Smith, 1983]** Davis. R et Smith. R.G. “Negotiation as a Metaphor for Distributed Problem Solving. Artificial Intelligence” 20(1), pp. 63–100.

**[Davis et al., 1995]** Davis, J.-R. et Huttenlocher, D.-P. Shared Annotation for Cooperative Learning. The first international conference on Computer support for collaborative learning (CSCL '95), pp. 84-88, Mahwah, NJ, USA, Lawrence Erlbaum Associates, Inc. 1995.

**[Desmontils and Jacquin, 2002]** Desmontils. E, Jacquin. C. “Indexing a Web Site with a Terminology Oriented Ontology”. In The Emerging Semantic Web Amsterdam : IOS Press, 2002, pp. 181-197.

**[Decker et al., 1999]** Decker, S., Erdmann, M., Fensel, D. et Studer, R. “Ontobroker: Ontology Based Access to Distributed and Semi-Structured Information”. Database Semantics: Semantic Issues in Multimedia Systems. pp.351–369. 1999.

**[Denoue, 2000]** Denoue, L. De la création à la capitalisation des annotations dans un espace personnel d'informations. Informatique. Annecy, Université de Savoie.p. 159. Sous la direction de L. Vignollet. 2000.

**[Denoue et al., 2000]** Denoue, L. et Vignollet, L. An annotation tool for Web browsers and its applications to information retrieval. Content-Based Multimedia Information Access (RIAO 2000), Paris, France. 2000.

**[Dieng et al., 1998]** R. Dieng, O. Corby, A. Giboin and M. Ribière, “Methods and Tools for Corporate Knowledge Management,” International Journal of Human-Computer Studies, ” 1999, pp. 567-598.

**[Dieng, 2002]** Dieng, R. Panorama: Méthodes et outils pour la gestion des connaissances. Projet ACACIA, INRIA, Sophia-Antinopolis. Journée Lorient sur la gestion des connaissances., pp. 36 pages., Nancy. 28 Novembre 2002.

**[Dieng et al., 1998]** R. Dieng, O. Corby, A. Giboin and M. Ribière, “Methods and Tools for Corporate Knowledge Management,” International Journal of Human-Computer Studies, ” 1999, pp. 567-598.

**[Dieng et al., 2004]** Dieng, L. et Finin, T. Swoogle: A Search and Metadata Engine for the Semantic Web. Thirteenth ACM Conference on Information and Knowledge Management, Washington D.C., U.S.A. November 8-13, 2004, pp. 58-61.

**[Dieng-Kuntz, 2005]** Dieng-Kuntz. R. “Corporate Semantic Webs”. In Encyclopaedia of Knowledge Management, D. Schwartz ed, Idea Publishing Group, September 2005. 154

**[Dieng et al., 2006]** R. Dieng, O. Corby, F. Gandon, N. Matta and M. Ribière, “Knowledge Management: Méthodes et Outils pour la Gestion des Connaissances,” 3ème édition, Dunod, 2006.

**[Dorsey, 2005]** Dorsey, P. A. What is PKM, Overview of Personal Knowledge Management. Année d'accès: 2006. Url du document: <http://www.millikin.edu/webmaster/seminar/pkm.html>

**[Eason and Olphert, 1996]** Eason. K, Olphert. “ Organizational learning: implications of CSCW systems” In P . Thomas, Springer Verlag, Londre, 1996, pp. 75-89.

**[Elofson and Konsynski, 1991]** Elofson. G et Konsynski. B. “ Delegation ontologies”. Journal of management information systems, 1991, 37-62. In P . Thomas, Springer Verlag, Londre, 1996, pp. 75-89.

**[Ermine, 1996]** Ermine. J.L. “ Les systèmes de connaissances”. Hermès, 1996

**[Ermine et al., 1996]** J.L. Ermine, M., Chaillot, P. Bigeon, B. Charreton, and D. Malavielle, “MKSM: Méthode pour la gestion des connaissances”. Ingénierie des systèmes d'Information, AFCET-Hermès, Vol. 4, No. 4, 1996, pp. 540-575.

**[Ermine, 1998]** J. L., Ermine “Capter et créer le capital savoir”. Annales des Mines, Novembre 1998, pp. 82-86.

**[Ermine, 2000]** Ermine. J.L. “ Challenges and Approaches for Knowledge Management in Companies”. Workshop Knowledge Management: Theory and Applications, Fourth European Conference on Principles and Practice of Knowledge Discovery in Databases, September 13-16, 2000, Lyon, France.

**[Ermine, 2003]** Ermine. J.L. “ La gestion des connaissances”. Science Publication, Hermes, Paris, 2003.

**[Ermine and Tounkara, 200\_]** Ermine. J.L, Tounkara. T. “ Méthodes de cartographie pour l’alignement stratégique de la gestion des connaissances”. Chapitre 4, Lavoisier (edit), France, 2008.

**[Euzenat, 1996]** Euzenat, J. “Corporate memory through creation of knowledge bases and hyper documents”. In Proceedings of KAW’96, Gains and Musen (ed), Banff, Canada, Novemeber, 1996, pp.18-36.

**[Euzenat, 2002]** Euzenat, J. "Eight questions about Semantic Web annotations." IEEE Intelligent Systems, pp.55-62. 2002.

**[Euzenat, 2005]** Euzenat, J. L'annotation formelle de documents en (8) questions. Ingénierie des connaissances. R. Teulier, J. Charlet et P. Tchounikine. Paris, France, L'Harmattan. pp.251-271 2005.

**[Evrard, 2002]** Evrard F. L'annotation : un outil de travail collaboratif. Journées Daniel Vanderveken, Toulouse, France. 07/2002 2002.

**[Fernandez et al., 1997]** Fernandez. M., Gomez-Perez. A et Juristo. N. “METHONTOLOGY : from ontological art towards ontological engineering”. In Proceedings of the Spring Symposium Series on Ontological Engineering (AAAI’97), AAAI Press,, 1997.

**[Fondin, 1998]** Fondin. H. “ Traitement numérique des documents”. Hermès, 1998.

**[Gandon, 2002]** F. Gandon, “ Distributed Artificial Intelligence and Knowledge Management: ontologies and multi-agent systems for a corporate semantic web,” Scientific. Philosopher Doctorate Thesis In Informatics, INRIA and University of Nice - Sophia Antipolis, 2002.

**[Giboin, 1996]** Giboin. A. “How a new writers cooperate in referring”. In Proceedings of COOP’96, INRIA Sophia Antipolis, 1996, pp.37-56.

**[Girod, 1995]** Girod. M. “La mémoire”. Revue française de gestion, 1995, n° 150, pp. 30-42.

[Gómez-Pérez et al., 1996] Gómez-Pérez. A, Fernandez. M, De Vicente. A. “Towards a Method to Conceptualize Domain Ontologies”. Workshop on Ontological Engineering. ECAI'96, 1996, pp. 41-51.

[Golebiowska, 2001] Golebiowska. J , “ Exploitation des ontologies pour la mémoire d'un projet-véhicule: Méthode et outil SAMOVAR,” Ph. D. Computer Science Thesis, University of Nice Sophia Antipolis, 4th of February 2001.

[Gruber, 1993] Gruber.T.R, “Toward Principles for the Design of Ontologies Used for Knowledge Sharing, In Formal Ontology in Conceptual Analysis and Knowledge Representation,” edited by Nicola Guarino and Roberto Poli, Kluwer Academic, 1993.

[Gruninger and Fox, 1995] Gruninger. M et Fox. M. S. “Methodology for the design and evaluation of ontologies”. In Proceedings of the Workshop on Basic Ontological Issues on Knowledge Sharing, IJCAI'95, 1995.

[Grundstein, 1995] Grundstein. M. “La capitalisation des connaissances dans l'entreprise, systèmes de production de connaissances”. L'entreprise apprenante et les sciences de la complexité, Aix en Provence, mai, 1995.

[Grundstein, 2002] Grundstein M. “De la capitalisation des connaissances au renforcement des compétences dans l'entreprise étendue”. Conférence invitée au 1er colloque du groupe de travail Gestion des Compétences et des Connaissances en Génie Industriel Nantes, 12-13 décembre 2002.

[Grunstein and Barthès 2006] M. Grunstein and J.-P. Barthès, “An Industrial View of the Process of Capitalizing Knowledge,” Proceedings of ISMICK'96, Rotterdam, vol. 1, October 2006, pp. 258-264.

[Guarino and Giaretta, 1995] Guarino. N, Giaretta. P. “ Ontologies and Knowledge Bases: Towards a Terminological Clarification”. In N. J. I. Mars (ed.), Towards Very Large Knowledge Bases, IOS Press 1995.

[Gueraich and Boufaïda 2008] Gueraich, S., Boufaïda, Z A Semantic Annotation Method for Managing a Corporate Memory ACIT'2008, Hammamat, Tunisie, Decembre, 2008, pp. 13-24.

[Gueraich and Boufaïda 2009] Gueraich, S., Boufaïda, Z. Conception d'une mémoire de projet basée sur l'annotation sémantique JSIA'09, Mars 2009, Guelma, Algerie.

[Gueraich and Boufaïda 2010] Gueraich, S., Boufaïda, Z. Acquisition de connaissances basée annotation sémantique pour la construction d'une mémoire de projet. In conférence magrébine sur l'extraction et la gestion des connaissances (EGC-M), 14-15 décembre 2010, Alger.

[Gueraich and Boufaïda 2011] Gueraich, S., Boufaïda, Z., Building a Project Memory using Semantic Design Rationale Process, in the Journal of Software Engineering and Applications (JSEA), Vol. 4, No. 6 (June), 2011, pp.356-370. ISSN: 1945-3116

[Gueraich and Boufaïda 2010] S. Gueraich and Z. Boufaïda, “ Building a corporate memory through collaborative semantic annotation and ontology population,” International Conference on Web and Information Technologies (ICWIT'2010), Maroc, June 2010, pp. 523-529.

- [Handschuh et al., 2002]** Handschuh. S, Staab. S et Ciravegna. F. “S-CREAM – Semiautomatic CREAtion of Metadata” .The 13th International Conference on Knowledge Engineering and Management (EKAW 2002), ed Gomez-Perez, A., Springer Verlag, 2002.
- [Handschuh, 2005]** Handschuh. S “ Creating Ontology-based Metadata by Annotation for the Semantic Web. Thèse de doctorat, Université de Karlsruhe, 2005.,
- [Harani, 1997]** Harani. Y. “ Une Approche Multi-modèles pour la Capitalisation des Connaissances dans le Domaine de la Conception”. Thèse de l'INPG, spécialité en Génie Industriel, 19 Novembre 1997.
- [Hemma, 2005]** Hemmam. M. “ Un processus de développements d'ontologies dans le cadre du web sémantique”. Mémoire de magister en informatique, université Oum el bouaghi, 2005.
- [Huhns et al., 1989]** Huhns. M.N, Mukhopadhyay. U, Stevens. L.M, Bonnell. R.D. “ DAI for document retrieval” Distributed artificial intelligence, Morgan Kaufman, San Mateo, 1989, pp. 249-283.
- [Hu, 2000]** Hu. X, “ a survey on design rationale: representation, capture and retrieval,” Proceedings of DETC'00, 2000 ASME Design Engineering Technical Conferences, September 10-3, Balmore, Maryland,2000.
- [IBM et al., 2006]** IBM, and Sandpiper Software. “ Ontology Definition Metamodel”. Sixth Revised Submission to OMG/RFP ad/2003-03-40, 2006.
- [Iglesias et al., 1997]** C. A. Iglesias, M. Garijo, J. C. González and J. R. Velasco, “Analysis and design of multiagent systems using MAS-CommonKADS,” In AAI'97 Workshop on Agent Theories, Architectures and Languages, Providence (ATAL), Springer Verlag, July 1997.
- [Jouanot, 2001]** Jouanot F. “DILEMMA : vers une coopération de systèmes d'informations basée sur la médiation sémantique et la fusion d'objets”. Thèse. Dijon : Université de Bourgogne.
- [Karl, 1972]** Karl.S “Organization Design. An information processing view”. Sloan working paper, Cambridge, Mass. M.I.T, 1972.
- [Kahan et al., 2002]** Kahan.J, Koivunen.M, Prudhommeaux.E et Swick.R. “ Annotea: an open RDF infrastructure for shared web annotations”. Computer networks, Elsevier, 2002, 39(5), pp. 598-608.
- [Klein, 1993]** Klein. M. “ Capturing design rationale in concurrent engineering teams”. IEEE Computer special issue on computer support for concurrent engineering, 1993, 26(1): pp. 39-47.
- [Khelif and Dieng-Kuntz 2004]** Khelif.K et Dieng-Kuntz.R, “ Ontology-Based Semantic Annotations for Biochip Domain,” Proceeding of EKAW 2004 Workshop on the Application of Language and Semantic Technologies to support KM Processes, U.K., 2004.
- [Khelif et al., 2006]** Khelif.K, Dieng-Kuntz.R et P. Barbry, “Semantic web technologies for interpreting DNA microarray analyses: the MEAT system,” Proceedings of WISE'05, New York, 2006.
- [Khelif, 2006]** K. Khelif, “ Web sémantique et mémoire d'expériences pour l'analyse du transcriptome,” Doctorate Thesis at University of Nice - Sophia Antipolis, 2006.
- [Konda et al., 1992]** Konda. S, Monarch. I, Sargent. P.M. et Subrahmanian. E. “ Sharedmemory in design”. Research in engineering design, vol 4, n° 1, 1992, pp. 23-42.

- [Kuhn and Abecker, 1997]** Kuhn. O et Abecker. A. “ Corporate theories for knowledge management in industrial practice: prospects and challenges”. *Journal of Universal Computer Science*, 1997, 3(8) : pp. 929-954.
- [Labrousse et al., 2004]** Labrousse. M., “proposition d’un Modèle Conceptuel unifié pour la Gestion Dynamique des Connaissances d’Entreprise,” Thèse de l’Ecole Centrale de Nantes, spécialité Génie Mécanique, 2004.
- [Loke and Davison , 1998]** Loke S. W. et Davison A. “ LogicWeb: Enhancing the Web with Logic Programming.” *Journal of Logic Programming*, 1998, pp 195-240.
- [Loke and Davison , 1998]** Loke S. W. et Davison A. “ LogicWeb: Enhancing the Web with Logic Programming.” *Journal of Logic Programming*, 1998, pp 195-240.
- [Lynch et Chen , 1992]** Lynch. K.J et Chen. H. “Knowledge discovery from historical data” *Proceedings of the 25th AHICSS’92*, vol 3, 1992, pp. 70-79.
- [Malvache and Prieur, 1993]** Malvache. P et Prieur. P. “ Mastering corporate experience with the rex method”. In *Proceedings of ISMICK’93*, Compiègne, 27-28 Octobre, Paris, pp. 33-41.
- [Marshall, 1998]** Marshall.C. “ Toward an ecology of hypertext annotation” *Proceedings of the 9th ACM on hypertext*, ACM, Press New york, USA, 1998, pp. 40-49.
- [Matta et al., 2000]** Matta. N, M. Ribiere, O. Corby, M. Lewkowicz, M. Zaklad. “ Project Memory in Design Industrial Knowledge Management -A Micro Level Approach”. Springer-Verlag, 2000.
- [Matta et al., 1999]** Matta. N, Corby. O, Ribiere. M. “ Méthodes de capitalisation de mémoire de projet”. *Projet ACACIA*, Rapport de recherche, 1999, No. 3819.
- [Maurer and Holz, 1999]** Maurer, F et Holz. H. “ Process oriented knowledge management for learning software organizations”. In *proceedings of the 13th KAW’99 Workshop*, Banff, Canada, 1999.
- [Mhiri et al., 2005]** M. Mhiri, A. Mtibaa and F. Gargouri ,” Towards an approach for building information systems’ontologies ,” 1st workshop Formal Ontologies Meet Industry, June, Verona, Italy, 2005.
- [McCullough et al., 1998]** McCullough. D. Korelsky. et Whiter. “Information management for release based software evolution using Emma” *Proceedings of the 10th conference on software engineering and knowledge engineering*, 1998, pp. 219-226.
- [Morrison, 1993]** Morrison. J. “ Team memory: information management for business teams *Proceedings of HICSS’93*, 1993, pp. 122-131.
- [Nahapiet et al., 1998]** Nahapiet. M, Callot et. Oldham. M. “ Methododology and tools oriented to knowledge based engineering applications. *Rapport public 2.0*, 1998.
- [Nardi and Brachman, 2002]** Nardi. D et Brachman. R.J. “ An introduction to description logic”. *Description logic handbook*, Cambridge university Press, 2002, pp. 5-44.
- [Natalya and Deborah, 2005]** Natalya. F.N et Deborah. L.M. “ Ontology development 101: a guide to creating your first ontology” 2005, <http://ksl.sanford.edu/people/papers/ontology>.

[Neches et al., 1991] Neches. R, Fikes. R.E, Finin. T, Gruber. T.R, Patil. R, Senator. T., et Swartout. W. R. “ Enabling technology for knowledge sharing”. AI Magazine, 12(3), pp. 16-36.

[Nonaka and Takeuchi, 1995] Nonaka. I, Takeuchi. H. “ Knowledge-Creating Company: How Japanese Companies Create the Dynamics of Innovation”. Oxford University Press, 1995.

[Nonaka, 1991] Nonaka.I. “ The knowledge creating company” Harvard business review, November-December, 1991, 69(6): pp. 96-104.

[Nomino] <http://www.ling.uqam.ca/nomino>

[Nunamaker et al., 1991] Nunamaker. J.F. Dennis. A.R, Valacich. J.S et Vogel. D.R “ Electronic meeting systems to support group work”. Communication of the ACM, 1991, pp. 40-61.

[O’Leary, 1998] O’Leary. D, “Enterprise Knowledge Management,” Computer, Vol. 31, No. 3, March 1998, pp. 54–61.

[OMG, 2004] OMG, “Unified Modeling Language: Superstructure, Version 2.0,” 2004.

[ontomat ] <http://www.acemedia.org/aceMedia/results/software/ m-ontomat-annotizer.html>

[Patwardhan et al, 2003] Patwardhan.C Levenshtein I. V et Minghong L. “ Binary Codes capable of correcting deletions, insertions, and reversals. Cybernetics and Control Theory”. 10(8):707–710, 2003.

[Penalva, 1990] Penalva. J.M “SAGACE : une représentation des connaissances pour la supervision de procédés”. EC2Ed, Système experts de deuxième génération, Avignon, 1990.

[Polanyi, 1966] Polanyi. M. “ The Tacit Dimension”. London: Routledge & Kegan Paul, 1966.

[Pomian, 1996] J. Pomian, “Mémoire d'entreprise techniques et outils de la gestion du savoir,” Ed. Sapiientia, 1996, ISBN 2911761006.

[Popov et al., 2003] Popov. B, Kiryakov. A., Manov. D, Kirilova. A, Ognyanoff. D and Goranov. M., “ Towards Semantic Web Information Extraction,” in Proceedings of the Human Language Technologies Workshop (ISWC’03), Sanibel, Floride, 2003, pp. 1-22.

[Prax, 2000] Prax.J.Y. “ Le guide du knowledge Management”. Dunond, 2000. ISBN 2-10-004701-9

[Prié, 2003] Pié.Y. “ Annotations et méta-données pour le web sémantique”. Dans les actes des journées scientifiques « web sémantique et SHS » mai, 2003.

[Rada et al., 1997] Rada. R. J. Tempich. C et Oezsu. M. “An axiomatic model of dynamic schema evolution in object-base management systems. ACM Transactions on Database Systems, 1997, Volume 22, Number 1, pp. 75-114.

[Ramon, 2001] Ramon. S “ Techniques for Plan Recognition”. User Modeling and User-Adapted Interaction, Kluwer Academic Publishers. 11: 31-48, 2001.

[Ryle, 1949] Ryle. H “ Dilemmas in a general theory of planning”. 1942.

- [**Saloff, 2000**] Saloff. T. “ Organizing Knowledge”. California Management Review. Spring 2000, Vol. 40, No. 3; pp. 90-111.
- [**Schmid, 1994**] Schmid. H. “ Probabilistic part-of-speech tagging using decision trees”. In proceedings of International Conference on New Methods in Language Processing, Manchester, 1994.
- [**Simon, 1996**] G. Simon, “ Knowledge acquisition and modelling for corporate memory,” Lessons learnt from experience. In Proceedings of the 10 th Banff knowledge acquisition for knowlsdzgz based systems workshop (KAW'96), 1996, pp. 18-41.
- [**Steels. 1993**] L. Steels, “ Corporate Knowledge Management,” In Barthès Proceedings (ISMICK'93), 1993, pp. 9-30.
- [**Swan et al., 1999**] Swan. M.P, Huhns. M. N et Grandbastien. M. “Social Abstraction for Information management”. In the book Intelligent Information Management on the Internet, Matthias Klusch Springer, 1999, pp37-52.
- [**Snowden, 2000**] Snowden.J. “Enterprise Modeling and Education”. Chapter 12 pages 183-199 of "Enterprise Modeling: Improving Global Industrial Competitiveness"; Edited by Asbjørn Rolstadås and Bjørn Andersen; Kluwer Academic Publisher; 2000; ISBN 0-7923-7874-1
- [**Thabet et al., 2010**] Thabet. M, Chemmaa. A et Gueraich. S “Etude synthétique de quelques outils d'Annotation sémantique”. Projet de fin d'étude pour l'obtention du diplôme de Master Recherche en Informatique LMD, SITW Université Mentouri, Constantine, Algérie, Juin 2010.
- [**Thepchai et al., 1999**] Thepchai, S., et al. Learning Goal Ontology Supported by Learning Theories for Opportunistic Group Formation. AIED'99, pp. 67-74, Le Mans France. 1999.
- [**Thepchai et al., 1999**] Thepchai, S., et al. Learning Goal Ontology Supported by Learning Theories for Opportunistic Group Formation. AIED'99, pp. 67-74, Le Mans France. 1999.
- [**Toukara, 2002**] Toukara.T “ Gestion des Connaissances et Veille : vers un guide méthodologique pour améliorer la collecte d'informations”. Thèse, Université de Paris IX Dauphine, Paris, France, 2002.
- [**Tourtier et al., 1995**] Tourtier, P.-A. et Amergé, C. Toward Human-Computer Collaboration The 6th International Conference on Human-Computer Interaction (HCI International'95), Tokyo, Japan. Juillet 1995.
- [**Tourtier, 1995**] P. Tourtier, “Analyse préliminaire des métiers et de leurs interactions,” Rapport intermédiaire du projet GENIE, INRIA-Dassault-Aviation, 1995.
- [**Tsui 2000**] Tsui, E. "Exploring the KM Toolbox." Knowledge Management. 4 (2). pp.11-14. 2000.
- [**Tsui, 2001**] Tsui, E. “Technologies for personal and peer-to-peer (p2p) knowledge management”. CSC Leading Edge Forum (LEF)? 2001.

[**Tzturk and Aamodt , 1997**] Tzturk, O. et Aamodt. A. “ Towards a model of context for case-based diagnostic problem solving”. Context-97; Proceedings of the interdisciplinary conference on modeling and using context, Rio de Janeiro, Brasil. February 1997, pp. 198–208.

[**Uren et al. 2006**] V. Uren. V, Cimiano. P, Iria. J, Handschuh. S, Vargas-Vera. M, Motta. E et Ciravegna. F, “ Semantic annotation for knowledge management: Requirements and a survey of the state of the art,” Web Semantics: Science, Services and Agents on the World Wide Web, Vol. 4, No. 1, 2006, pp.14-28.

[**Uschold and Gruninger, 1996**] Uschold. M et Gruninger. M. “ Ontologies: Principles, methods and applications”. Knowledge Engineering Review, 1996. Also available as AIAI-TR-191 from AIAI, The University of Edinburgh, Vol. 11:2, pp. 93-136.

[**Van Heijst and al., 1996**] Van Heijst . G, Van der Spek.R et Kruizinga. E. “ Organizing corporate memories”. In Proceedings of KAW’96, Gains and Musen (ed), Banff, Canada, Novemeber, 1996, pp.42-67.

[**Van Heijst and al., 1997**] Van Heijst . Shreiber. A et Wielinga.B. “ Using explicit ontologies in KBS development”. International journal of human computer studies, 1997, pp. 183-297.

[**Vargas-Vera et al., 2002**] Vargas-Vera. M, Motta. E, Domingue. J, Lanzoni. M., Stutt. A. et Ciravegna. F. “ Ontology Driven Semi-Automatic and Automatic Support for Semantic Markup”. In Proceedings. of the 13th International Conference on Knowledge Engineering and Knowledge Management EKAW 2002, Springer Verlag LNAI 2473, 2002, pp. 379-391.

[**Veron, 1997**] Veron. M “ Modélisation de la composante annotative dans les documents électroniques, Master de recherche, IRIT, Toulouse, 1997.

[**W3C 2011**] W3C. The World Wide Web Consortium. Url du document: <http://www.w3.org/>

[**Weick, 1979**] Weick. K. “The Social Psychology of Organizing” Addidon-Wesley, 1979.

[**Wenyin et al. 2001**] Wenyin, L., et al. "Semi-Automatic Image Annotation." Human-Computer Interaction--Interact '01. pp.326-333. 2001.

[**Wiederhold, 1992**] Wiederhold G. “Mediators in the architecture of future information Systems”, Computer, 1992, Vol. 25, n°3. pp.38-49.

[**Wiig, 1999**] Wiig. G. “ Mediators in the architecture of future information systems.” IEEE Computer 25, 3 (March), 1999, pp. 38-49.

[**WikipediaFr, 2005**] WikipediaFr, Wikipedia l'encyclopédie libre, Open Source. Année d'accès: 2011. Url du document: <http://fr.wikipedia.org/wiki/Accueil>

[**Wojahn et al. 1998**] Wojahn, P., Neuwirth, C. et Bullock, B. Effects of interfaces for annotation on communication in a collaborative task. CHI, pp. 456-463. 1998.

[**Wu and Palmer, 1994**] Wu. Z et .Palmer. M. “ Verb semantic and lexical selection” proceedings of the 32nd annual meeting of the associations for computational linguistics, 1994, Pages133-138.

# **ANNEXES**

# ANNEXE 1 : TABLES DES CONCEPTS ET DES RELATIONS

TERMS	CONCEPT ID	PARENT IDS	NATURAL LANGUAGE DEFINITION
Corporate memory	ME	KnowledgeEngineeringTopic	An explicit, disembodied and persistent representation of knowledge and information in an organization, in order to facilitate their access and reuse by members of the organization, for their tasks.
Project meory	PM	ME	It contains project definition activities, history and results. It is a knowledge and information acquired and produced during the realization of projects.
nature-ME	Nature-ME	ME	Defines the nature and the type of the corporate memory
organizational	Organizational	ME	Defines the organizational aspects
problem	Problem	ME	Identifies problems
historical	Historical	ME	Identifies archives
projetc	project	Feature	Project definition
industry	Industry	Industry fiels	Activity of industrialization
activity,	Activity	NonSpatialEntity	Entity representing a type of voluntary action.
activity able entity,	ActivityAbleEntity	RoleEntity	Entity that can have an activity.
addition,	Addition	CorporateMemoryEvent	Corporate Memory Event corresponding to content added to the memory.
additional topic,	AdditionalTopic	Entity	Entity representing subjects that hold attention and possibly something one wants to discover. These topics are additional in the sens that they were not introduced or used anywhere else in the ontology but were identified as relevant for document annotation - domain concepts, general subjects...-
business,	Business	Activity	Activity composed of commercial or industrial activities intended to make profits.
consultancy report,	ConsultancyReport	Report	Report on studies performed for clients. Most of them are confidential.
consultant,	Consultant	Professional	Professional who works with some business in a consulting capacity.
consultation,	Consultation	CorporateMemoryEvent	Corporate Memory Event corresponding to a user seeking information from the memory.
consultation trace,	ConsultationTrace	NonSpatialEntity	Such an element is created when the user visits a document, and is updated every time he returns to the document.
consulting,	Consulting	Service	Service of providing professional advice or expertise.
pmdesigner	Pmdesigner	designer	Person who design the project memory
pmuser	Pmusep	user	Person who use the the project memory
buy	Buy	Commerce	Commerce activity where a person or an organization obtains something by paying money for it.
anoteted	Anotated	Resource	Resource to be annotatde
catalog, catalogue,	Catalog	ReferenceDocument	Reference document containing an enumeration of things usually linked to a domain or an activity.
CD, C.D., compact disc,	CD	DocumentaryMedium	Digital Medium where the Documentary Element is recorded on a optical disc.
CD-ROM,	CD-ROM	CD	CD used for storing computer data.
chart,	Chart	Document	Document corresponding to a visual display of information often intended to show the information more clearly.
clock time, time,	ClockTime	TimeEntity	Time Entity corresponding to a time of day.

club,	Club	GroupOfIndividuals	Group of individuals with a common purpose or interest who meet regularly and take part in shared activities.
cluster,	Cluster	OrganizationPart	Organization part grouping projects according to their client type.
code of conduct,	CodeOfConduct	Manual	Manual that makes explicit the expectations governing the behavior of those agents subject to it in certain kinds of situations.
cognitive sciences,	CognitiveSciencesTopic	AdditionalTopic	Topic concerned with the sciences studying cognition.
commerce,	Commerce	Service	Activity concerned with buying or selling goods or services for a profit.
competitor,	Competitor	Organization	Organization having similar activities to an organization and targeting the same markets.
computer file format,	ComputerFileFormat	DigitalFormat	Digital format corresponding to a single collection of computer data or information that has a name.
computer science,	ComputerScienceTopic	AdditionalTopic	Study of automatic information and data processing methods, and of computers, including both hardware and software design.
conference, lecture,	Conference	FormalGathering	Formal Gathering of persons with common interests, esp. professional interests, for the purpose of sharing information and opinions especially through lectures and debates.
confidential document,	ConfidentialDocument	DiffusionRight	The diffusion is restricted to a defined community. There are different levels of confidentiality for internal or external diffusion. A document can be composed of confidential parts or not confidential parts (for example: the reference and the abstract can be freely diffused but not the integral text).
consortium,	Consortium	Organization	Organization of several Businesses joining together as a group for some shared definite purpose.
consultancy report,	ConsultancyReport	Report	Report on studies performed for clients. Most of them are confidential.
consultant,	Consultant	Professional	Professional who works with some business in a consulting capacity.
consultation team	ConsultationTeam	CorporateMemoryActivity	Corporate Memory activity corresponding to a team seeking information from the memory.
corporate memory,	CorporateMemoryTopic	KnowledgeEngineeringTopic	An explicit, disembodied and persistent representation of knowledge and information in an organization, in order to facilitate their access and reuse by members of the organization, for their tasks.
corporate memory event,	CorporateMemoryEvent	Event	Event corresponding to changes in the Corporate Memory.
course, training document,	CourseDocument	Document	Training document containing pedagogical material and usually used by lecturers or teachers as a support of their lessons.
customer, client,	Customer	Organization	Organization who pays for goods or services.
data file format,	DataFileFormat	ComputerFileFormat	Computer File format dedicated and formatted to be manipulated by a Program.
date,	Date	TimeEntity	Time Entity corresponding a numbered day in a month, often given with a combination of the name of the day, the month and the year.
deletion,	Deletion	CorporateMemoryEvent	Corporate Memory Event corresponding to content removed from the memory.
department,	Department	OrganizationPart	Thematic grouping of services.
development,	Development	Service	Service corresponding to analysis, design, implementation and testing of research solutions.
diagram,	Diagram	Chart	Chart intended to explain how something works a drawing showing the relation between the parts.

dictionary,	Standard	ReferenceDocument NumberableEntity	Reference used as a point of reference to compare and evaluate quality of products or systems.
dictionary,	Dictionary	ReferenceDocument	Reference Document in which words are listed alphabetically and their meanings.
diffusion right,	DiffusionRight	DocumentAttribute	Authorized scope for the diffusion of a document.
digital format, numeric format,	DigitalFormat	StorageFormat	Storage format based on discontinuous data.
direction,	Direction	OrganizationPart	Organization part with a special activity inside the company eg: HR, Project Planning.
distributed artificial intelligence, DAI, D.A.I.,	DistributedArtificialIntelligenceTopic	ArtificialIntelligenceTopic	Distributed approach of Artificial Intelligence.
division,	Division	OrganizationPart GroupOfIndividuals	Functional subdivision of a service. It has generally 10 to 25 people.
document,	Document	Entity EntityConcerningATopic NumberableEntity	Entity including elements serving as a representation of thinking.
documentary file,	DocumentaryFile	Document	Thematic document, regularly updated and made with heterogeneous material (articles, references, synthesis...).
documentary medium,	DocumentaryMedium	PhysicalEntity	Physical entity through which signals/messages travel as a means for communication.
document attribute,	DocumentAttribute	Attribute	Attribut characteristic of documents.
document origin,	DocumentOrigin	DocumentAttribute	Where the document was produced or written.
duration,	Duration	TimeEntity	Time Entity corresponding to the length of time that something lasts.
DVD, D.V.D., digital versatile disc, digital video disc,	DVD	DocumentaryMedium	Digital Medium where the Documentary Element is recorded on a optical disc called Digital Versatile Disc or Digital Video Disc.
DVD-ROM,	DVD-ROM	DVD	DVD used for storing computer data.
dynamic image,	DynamicImageRepresentation	ImageRepresentation	Image changing over time quickly enough to be noticed.
XML, X.M.L., X.M.L. Format, Extensible Markup Language Format, format du Langage de Marqueurs Extensible,	XMLFormat	TextFormat	Text Format of a Documentary Element written in eXtensible Markup Language.
XML, X.M.L., extensible markup language,	XMLTopic	WebTopic	Web Extensible Markup Language enabling the definition, transmission, validation, and interpretation of data between applications and between organizations.
zip, zip format,	ZIPFormat	DataFileFormat	Data File Format for files compressed in ZIP format.
facebook	facebook	WebTopic	Social network
travel,	Travel	Service	Service giving information about prices and schedules for a trip and arranging tickets and accommodation.

trend analysis,	TrendAnalysis	Document	Synthesis Document written by an expert on the trends of a technological area.
tutor,	Tutor	IntegrationProcessActor	A person who gives private advice and instruction to a newcomer.
UML, U.M.L., unified modeling language,	UML	ArtificialLanguage	General-purpose notational Artificial Language for specifying and visualizing complex software, especially large, object-oriented projects.
union,	Union	GroupOfIndividuals	Group of individuals formed to bargain with the employer.
unit,	Unit	GroupOfIndividuals	Group of individuals corresponding to a group of researchers focusing on a sub interest field e.g.: microwaves.
university,	University	Organization	Organization which does university-level teaching and/or research.
update, mise a jour,	Update	Modification	Modification corresponding to content changed into more recent one.
virtual reality,	VirtualReality	IconicRepresentationSystem	Iconic Representation System for simulating systems or environments with and within which people can interact.
virtual reality,	VirtualRealitySimulationTopic	ComputerGraphicsTopic SimulationTopic	A computer simulation of a real or imaginary system that enables a user to perform operations on the simulated system and shows the effects in real time.
visual,	VisualPerceptionMode	DocumentaryPerceptionMode	Documentary Perception Mode by sight.
Djezzy	Djezy	MobilePhone provider	that allows users to access information instantly via handheld wireless devices such as mobile phones, pagers, two-way radios, smartphones and communicators.
wav, WAV Format,	WAVFormat	DataFileFormat	Data File Format of a audio document encoded in the WAV format.
wave propagation,	WavePropagationTopic	PhysicsTopic	
Web,	WebTopic	NetworkTopic	Internet servers networks that support specially formatted documents.
web page, web site,	WebPage	Document	Document corresponding to a page on the World Wide Web.
web site,	WebSite	Document	Document made up of interconnected Web Pages, usually including a Homepage, generally located on the same server, and prepared and maintained as a collection of information by a person, group, or organization.
Word, Word format, Microsoft Word format,	MSWordFormat	DataFileFormat	Data File Format of a visual document encoded in the format of Microsoft Word.
worker,	Worker	Professional	Professional member of the working class with a specific job.

<b>TERMS</b>	<b>PROPERTY ID</b>	<b>PARENT IDS</b>	<b>DOMAIN</b>	<b>RANGE</b>	<b>NATURAL LANGUAGE DEFINITION</b>
address	Address	SomeRelation	Location	Text	Address of a location.
possessing	Possessing	SomeRelation	Nature-ME	Organizational	The task of possession
fax number,	FaxNumber	HasNumber	Location	Text	Fax number of a location.
first name, given name,	FirstName	Designation	Person	Text	The name that occurs first in a person's full name.
first visit,	FirstVisit	SomeRelation	ConsultationTrace	Text	Date of the first visit to a document.
geographical area,	Geographical Area	SomeRelation	IndexCard	Text	geographical area in which the specific service or technology are developed or exist.
beginning,	Beginning	SomeRelation	Gathering	Text	Starting date/hour of an gathering.
birth date,	BirthDate	SomeRelation	Person	Text	Date of birth.
colleague,	Colleague	SomeRelation	Person	Person	one of a group of people who work together.
has for activity,	HasForActivity	SomeRelation	ActivityAbleEntity	Activity	Relation denoting that an Entity is carrying out an activity.
has for diffusion right,	HasForDiffusionRight	SomeRelation	Document	DiffusionRight	Relation denoting the diffusion right of a document.
has for favorite annotation,	HasForFavoriteAnnotation	SomeRelation	Person	Pattern	Relation denoting a favorite pattern of a user for annotating resources.
concern,	Concern	SomeRelation	EntityConcerningATopic	AdditionalTopic	Relation denoting that an entity (e.g.: a document, a gathering...) concerns a topic.
contain,	Contain	SomeRelation	Document	Document	Relation denoting that a document includes another one.
created by,	CreatedBy	SomeRelation	Document	ActivityAbleEntity	Relation denoting that a Document has been created by an Entity.
service provider,	ServiceProvider	SomeRelation	IndexCard	Text	indicates the provider of telecommunication services or technologies or terminals.
service type,	ServiceType	SomeRelation	IndexCard	Text	
situated,	Situated	SomeRelation	SituableEntity	Location	Relation denoting that an Entity is located in a Location.
sleeping partner,	HasForSleepingPartner	SomeRelation	Report	ActivityAbleEntity	Relation designating a sleeping partner of the report.
designation ,	Designation	SomeRelation	Something	Text	Identifying word or words by which a thing is called and classified or distinguished from others.
domain, thesis research domain,	HasForThesisDomain	SomeRelation	Thesis	Text	Domain associated given by the national nomenclature.
employed by,	EmployedBy	SomeRelation	Employee	OrganizationalEntity	Relation denoting that an Organization has an Employee working or doing a job for it and pays this Employee for it.
employment contract,	EmploymentContract	SomeRelation	Employee	ContractTypeAttribute	Relation denoting the Type of the contract that link an

					Employee to an organization.
end,	End	SomeRelation	Gathering	Text	Ending date of an gathering
extracted from,	ExtractedFrom	SomeRelation	Article	ExtractedDocument	Relation designating the document from which the article was extracted.
had for participant,	HadForParticipant	SomeRelation	Gathering	GatheringEntity	Relation denoting that an Organizational Entity participates/participated to a Gathering.
has for favorite query,	HasForFavoriteQuery	SomeRelation	Person	Pattern	Relation denoting a favorite pattern of a user for querying the base.
has for internal report number,	HasForInternalReportNumber	HasNumber	Report	Text	Internal report number.
has for medium,	HasForMedium	SomeRelation	Document	DocumentaryMedium	Relation denoting that a document uses a medium.
has for number,	HasNumber	SomeRelation	NumberableEntity	Text	Number associated to a 'numberable' entity.
has for ontological entrance point,	HasForOntologicalEntrancePoint	SomeRelation	Person	Something	Relation denoting a preferred entrance point for browsing the ontology.
has for origin,	HasForOrigin	SomeRelation	Document	DocumentOrigin	Relation denoting the origin of a document.
has for perception mode,	HasForPerceptionMode	SomeRelation	Document	DocumentaryPerceptionMode	Relation denoting that a document uses a perception mode -audio, visual, tactile-.
has for personal interest,	HasForPersonalInterest	IsInterestedBy			Relation denoting that an entity has a personal interest
has for representation system,	HasForRepresentationSystem	SomeRelation	Document	DocumentaryRepresentationSystem	Relation denoting that a document uses a representation system.
has for storage format,	HasForStorageFormat	SomeRelation	Document	StorageFormat	Relation denoting that a document is stored in a given format.
has for work interest,	HasForWorkInterest	IsInterestedBy			Relation denoting that an Entity has a special work interest.
hiring date,	HireDate	SomeRelation	Employee	Text	The date when the employee was hired.
include,	Include	SomeRelation	Organizational Entity	GroupableEntity	Relation denoting that an Entity has as a part another Entity.
indication,	Indication	SomeRelation	Location	Text	Textual signs/clues pointing to the location e.g.: "in the cupboard of the rest room".
ISBN, International Standard Book Number,	HasForISBN	HasNumber	Book	Text	The International Standard Book Number (ISBN) is a system of numerical identification for books, pamphlets, educational kits, microforms, CD-ROM and braille publications.
is interested by,	IsInterestedBy	SomeRelation	InterestableEntity	AdditionalTopic	Relation denoting that an Entity is interested in a topic.
ISRN,	HasForISRN	HasNumber	Report	Text	Identification code for reports.
ISSN,	HasForISSN	HasNumber	ISSNHolderDocument	Text	Standardized international code which allows the identification

					of any serial publication.
issued on the occasion of,	IssuedOnTheOccasionOf	SomeRelation	Document	Gathering	Relation denoting that a document has been/is issued on the occasion of a gathering.
keyword,	Keyword	SomeRelation	Document	Text	Keyword representative of the content of the document.
last visit,	LastVisit	SomeRelation	ConsultationTrace	Text	Date of the last visit to a document.
manage, oversee, supervise, superintend,	Manage	SomeRelation	ManagementAbleEntity	ManageableEntity	Relation denoting that an entity is in charge/controls of another entity.
mobile number,	MobileNumber	HasNumber	Person	Text	Mobile phone number.
name, family name, surname, last name,	FamilyName	Designation	Person	Text	The name used to identify the members of a family.
network type,	NetworkType	SomeRelation	IndexCard	Text	ex: mobile, fixed, satellite
number, period number,	HasForPeriodNumber	HasNumber	ISSNHolderDocument	Text	(Period) number of a serial document.
organized by,	OrganizedBy	SomeRelation	Gathering	GatheringEntity	Relation denoting that an entity organizes / organized a gathering.
phone number,	PhoneNumber	HasNumber	Location	Text	Phone number of a location.
pushed document,	PushedDocument	SomeRelation	PushedDocumentTrace	Document	Document pushed by the CoMMA Push Mode.
rating given,	RatingGiven	SomeRelation	ConsultationTrace	RatingValue	User s feedback after a consultation. Can be like Good, Bad, etc. cf RatingValue.
RDF string,	RDFString	SomeRelation	Pattern	Text	String representing the RDF pattern.
refers monitoring to,	RefersMonitoringTo	SomeRelation	Observer	AreaReferent	Links an observer to the referent of his area.
related to,	RelatedTo	SomeRelation	AdditionalTopic	AdditionalTopic	Relation denoting that an Interest Field is linked to another.
reliability,	Reliability	SomeRelation	InformationForm	RatingValue	Reliability of the information.
see also,	SeeAlsoInformationSource	SomeRelation	InformationForm	Text	Other source of the information.
some relation,	SomeRelation		Something		An abstraction belonging to, linking, or characterising of two things.
source, information form source,	InformationFormSource	SomeRelation	InformationForm	Text	Source of the information.
summary,	Summary	SomeRelation	Document	Text	A account of the main points of a document.
supervised by,	HasForSupervisingOrganization	SomeRelation	Thesis	Organization	Organization (school, university, laboratory...) supervising the thesis.
target,	Target	SomeRelation	Document	InterestAbleEntity	Relation denoting that a Document is intended for some Entity.

technology,	Technology	SomeRelation	IndexCard	Text	
title,	Title	SomeRelation Designation	Document	Text	Designation of a document.
type, thesis research type,	HasForThesis Type	SomeRelation	Thesis	Text	Type associated given by the national nomenclature.
visit count,	VisitCount	SomeRelation	ConsultationTrace	Text	Number of visits to a document in the whole history of a profile.
visited projectdocument,	VisitedDocuProj	SomeRelation	Consultationfactory	Document	Document concerned by this element of the use history of a project
visitor,	Visitor	SomeRelation	ConsultationTrace	Employee	Person concerned by this element of the history of a profile.

## EXTRAIT DE L'ONTOLOGIE

```
<rdf:RDF
  xmlns:rss="http://enmtp.org/rss/1.0/"
  xmlns:jms="http://jena.hpl/jms#"
  xmlns:gms="http://gate.hpl/gms#"
  xmlns:ams="http://anni.hpl/ams#"
  xmlns:rms="http://rasp.hpl/rms#"
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
  xmlns:owl="http://www.w3.org/2003/07/owl#"
  xmlns:daml="http://www.daml.org/2001/03/daml+oil#"
  xmlns:dc="http://enmtp.org/dc/elements/1.1/" >

<rdf:Description rdf:about="#Concept">
  <rdf:type rdf:resource="http://www.w3.org/2003/07/owl#Class"/>
</rdf:Description>
<rdf:Description rdf:about="#">
  <rdf:type rdf:resource="http://www.w3.org/2003/07/owl#Class"/>
  <rdfs:subClassOf rdf:resource="#Concept"/>
</rdf:Description>
<rdf:Description rdf:about="#pm_designer">
  <rdf:type rdf:resource="http://www.w3.org/2003/07/owl#DatatypeProperty"/>
  <rdfs:range rdf:resource="http://www.w3.org/2001/XMLSchema#string"/>
  <rdfs:domain rdf:resource="#pm_designer"/>
</rdf:Description>
<rdf:Description rdf:about="#Project">
  <rdfs:subClassOf rdf:resource="#Concept"/>
  <rdf:type rdf:resource="http://www.w3.org/2003/07/owl#Class"/>
</rdf:Description>
<rdf:Description rdf:about="#Application">
  <rdfs:subClassOf rdf:resource="#Product"/>
  <rdf:type rdf:resource="http://www.w3.org/2003/07/owl#Class"/>
</rdf:Description>
<rdf:Description rdf:nodeID="A4">
  <rdf:first rdf:resource="#Produit"/>
  <rdf:rest rdf:nodeID="A5"/>
</rdf:Description>
<rdf:Description rdf:about="#Problem">
  <rdfs:subClassOf rdf:resource="#ME"/>
  <rdf:type rdf:resource="http://www.w3.org/2003/07/owl#Class"/>
</rdf:Description>
<rdf:Description rdf:about="#TechniqueProblem">
  <rdfs:subClassOf rdf:resource="#MemoryProblem"/>
  <rdf:type rdf:resource="http://www.w3.org/2003/07/owl#Class"/>
</rdf:Description>
<rdf:Description rdf:about="#Productprovider">
  <rdfs:subClassOf rdf:resource="#Product"/>
  <rdf:type rdf:resource="http://www.w3.org/2003/07/owl#Class"/>
</rdf:Description>
```

# EXEMPLE D'ANNOTATION CONCERNANT LE PROJET PIECE AU SEIN DU SERVICE ACHAT

```
<xml "version="1.0" encoding="UTF-8" xmlns:rdfs="http://www.w3.org/2000/01/rdf-  
schema#">
```

```
<rdf:Description rdf:about="http://www.enmtp.dz/serviceachat.html">  
  <rdf:type rdf:about="#Article"/>  
  <rdfs:label>Piece Noble</rdfs:label>  
  <dc:subject rdf:resourceRef="#FFpiecenoble"/>  
</rdf:Description>
```

```
<rdfs:Class rdf:ID="#moteur">  
<rdfs:Class rdf:ID="#piece ">  
<rdfs:Class rdf:ID="#piece_de_freinage">  
  <rdfs:subClassOf rdf:resource="#Piece"/>  
</rdfs:Class>
```

```
<rdf:Property rdf:about="moteur_de">  
  <rdfs:domain rdf:resource="#Piece"/>  
  <rdfs:range rdf:resource="#moteur"/>  
</rdf:Property>
```

```
<Piece rdf:about=" piece_de_freinage ">  
  <rdfs:label>PF1</rdfs:label>  
  < moteur_de rdf:resource="#mot1"/>  
  </ piece_de_freinage >  
</xml>
```

# ANNEXE 2 : TABLEAUX ET GRAPHES D'OUTILS D'ANNOTATIONS

Dans cette partie, nous présentons des tableaux et des graphes qui récapitulant quelques outils pour l'annotation sémantique.

## - Discussion 1

Le tableau suivant fournit une synthèse des outils d'annotation sémantique selon les critères évoqués dans le chapitre 5.

<i>Outil d'annotation</i>	<i>Les formats standards</i>	<i>conception centrée sur Utilisateur</i>	<i>Ontologie de Soutien</i>	<i>Formats de documents</i>	<i>Évolution de document</i>	<i>Stockage d'annotation</i>
Amaya 1996-2009	RDF(S) XLink, XPointer	Navigateur web et éditeur	Serveur d'annotation	HTML, XHTML and XML	XPointer	locale ou serveur d'annotation
Mangrove	RDF	Outil d'annotation graphique		HTML, email		Base de données RDF (Jena)
Vannotea	XML	Collaboration support		MPEG-2, JPEG2000, Direct3D		Serveur d'annotation
OntoMat	DAML+ OIL, OWL, SQL, XPointer	Drag & drop, créer annoter	OntoBroker serveur d'inference d'annotation	HTML, Deep Web	XPointer, pattern matching	Serveur d'annotation intégré dans des pages web, separate file
M-OntoMat- Annotizer	XML, RDF(S) DOLCE	extraction automatique des descripteurs visuels		MPEG-7		Serveur d'annotation
SHOE Knowledge annotator	SHOE	Prompting	Serveur d'ontologie	HTML		intégré dans des pages web

53

SMORE	RDF(S)	Navigateur web et éditeur	Serveur d'ontologie et édition d'ontologie	HTML, text, email Et images		
Open Ontology Forge	RDF(S), XML, Xlink XPointer, Dublin Core	Navigateur web + drag & supprimer, Créer & annoter	Locale, editable ontologies	HTML, text, images (SVG)	XPointer	RDF Locale ou fichier XML
COHSE annotator	DAML+OIL	Plug in pour Mozilla & IE	Serveur d'ontologie	HTML (via DOM)	XPointer	Serveur d'annotaion, DLS
Lixto	Wrappers					
MnM	RDF(S), DAML+OIL, OCML	Navigateur web	Serveur d'ontologie	HTML, text	Stores annotated page	Incorporé dans des pages web
Melita	RDF(S) DAML+ OIL	Controle d'intrusion d'IE	Local, editable ontologies	HTML, text	Les expressions reguliers	
Parmenides	XML (CAS)		Clustering to suggest additions			
Armadillo	RDF(S)			HTML		RDF triple store
KnowItAll	HTML					

SmartWeb	RDF, RDF(S), OWL					base de connaissance RDF
PANKOW	HTML	CREAM				
AeroSWARM (AeroDAML)	OWL	Service Web	Ontologies Locales	HTML		
SemTag	RDF(S)			HTML		Label bureau (PICS)
KIM	RDF(S), OWL	Différent plug-in front ends	KIMO	HTML		Base de connaissance RDF(S)
Rainbow Project	RDF WSDL/SOAP	Amphora XHTML database	Shared upper level ontology	HTML		RDF repository (Sesame)
h-TechSight	DAML+ OIL, RDF	KM Portal	Editeur d'ontologies, dynamics metrics	HTML		Tagged HTML, serveur web
WiCK Office	Microsoft Smart Documents	Office applications, support for form filling		Microsoft Office		Serveur d'Annotation (3 store)
AktivDoc	HTML, RDF	Integrated editing environment		HTML		RDF triple store

SemanticWord	DAML+OIL	Microsoft Word GUIs		Word		Mark-up tied to text
Maggie	HTML OCML	Web browser plug in		HTML		None, real time
Thresher	RDF	Navigateur web (Thresher)	Ontology personalization	HTML		None, real time
ANNIE( gate)				XML		None, real time

## - Discussion 2

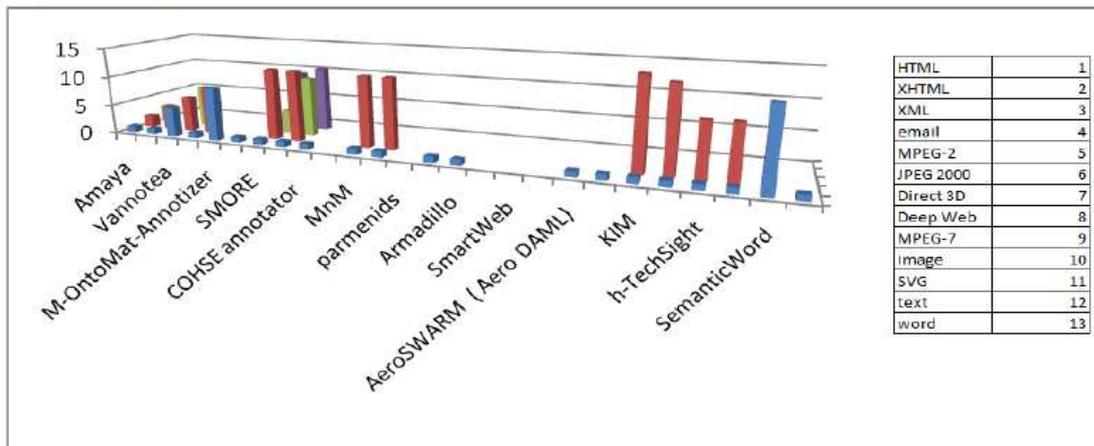
Le tableau suivant fournit un résumé de quelques aspects d'automation.

Outil d'annotation	Automation	Type d'analyse pour l'automatisation	Étude dans l'automatisation
Amaya	Non		
Mangrove	Non		
vannotea	Non		
OntoMat	Oui	PANKOW, Amilcare	Étude dirigée
M-OntoMat- Annotizer	Oui	Extraction des descripteurs spatiaux	Algorithme génétique
SHOE knowledge annotator	Oui	Running SHOE (wrappers)	Non
SMORE	Oui	Screen scraper (Racleur d'écran)	Non
Open Ontology Forge	Oui	String matching	Non
COHSE annotator	Oui	Ontology string matching	Non
Lixto	Oui	Wrappers	Non
MnM	Oui	POS tagging, Named Entity Recognition	Étude dirigée
Melita	Oui	String matching, POS tagging, Named Entity Recognition	Étude dirigée
Parmenides	Oui	Exploitation des textes avec des contraintes	Étude non surveillée

Armadillo	Oui	String matching, POS tagging, Named Entity Recognition	Étude non surveillée
KnowItAll	Oui	String matching, Hearst patterns	Étude non surveillée
SmartWeb	Oui	Shallow linguistic parsing	Étude non surveillée
PANKOW	Oui	Hearst patterns	Étude non surveillée
AeroSWARM (AeroDAML)	Oui	AeroText	Non
SemTag	Oui	Seeker, similarity, TBD	Étude non surveillée
KIM	Oui	String matching, POS tagging, Named Entity Recognition	Non
Rainbow project	Oui	Hidden Markov models, bit-map classification	Étude dirigée
h-TechSight	Oui	Shallow linguistic analysis (POS tagging, Named Entity Recognition)	Non
WiCKOffice	Oui	Named Entity Recognition	Non
AktiveDoc	Oui	String matching, POS tagging, Named Entity Recognition	Étude dirigée et non surveillée
SemanticWord	Oui	AeroDAML	Non
Magpie	Oui	String-matching, Named Entity Recognition	Non
Thresher	Oui	Screen scraping, wrappers	Étude dirigée

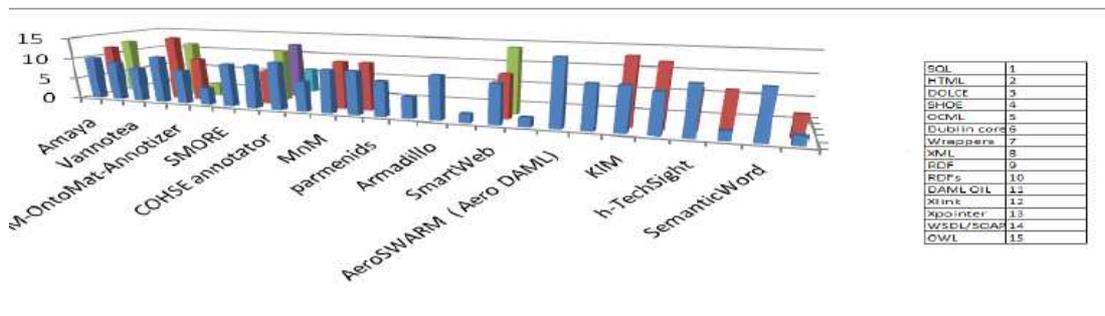
**Interprétation :** D’après le tableau 1 nous allons modéliser quelques colonnes sous la forme des histogrammes suivants :

## Formats de documents

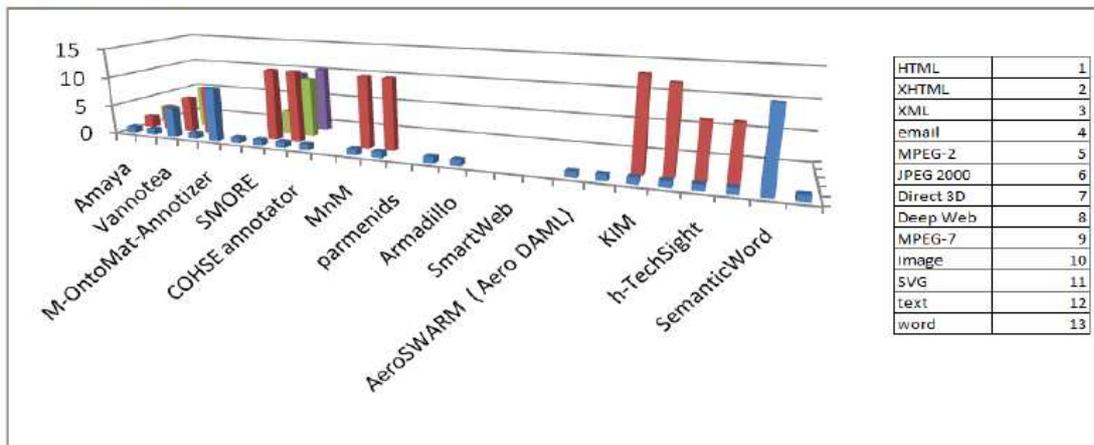


## Les formats standards

Dans cet Histogramme nous avons classifié les formats standards selon leurs dates d'apparitions.



## Formats de documents



### 3.3.3 Stockage d'annotation

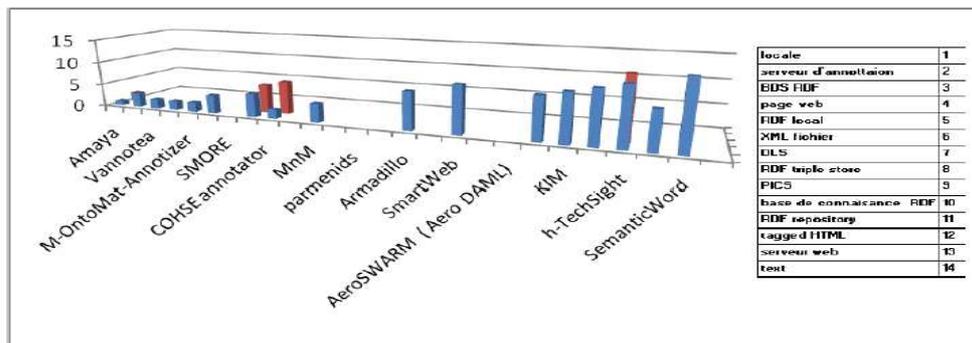


Figure 14: le stockage d'annotation

Dans cet histogramme nous pouvons déduire que chaque outil d'annotation contient son propre stockage d'annotation, c.-à-d. une annotation peut être enregistrée localement ou bien stocké sur des serveurs d'annotations ou web, des pages web, des bases de données RDF, des bases de connaissance RDF...etc.

Par exemple dans le premier histogramme, nous pouvons déduire qu'un outil d'annotation peut avoir plusieurs formats standards tels que Amaya, M-OntoMat-Annotizer, MnM, Open Ontology Forge, smart web ...etc. Nous remarquons aussi que certains outils utilisent les mêmes formats (langages) comme RDF, RDFs, OWL, DAML-OIL. Parmi ces formats nous constatons que le langage OWL est le langage le plus récent.

### **- Discussion 3**

Nous pouvons déduire quelques points de convergence et de divergence entre les différents outils cités dans les sections précédentes, on peut remarquer que :

1- Un des problèmes avec ces outils d'annotation sémantique basés sur les techniques d'Extraction d'Information est qu'ils ne fournissent pas à l'utilisateur une manière de s'adapter à la technologie : par exemple dans le cas où les utilisateurs n'ont pas besoin ou ne voudraient pas d'équipements hautement personnalisés.

2- Quelques outils ne se rendent pas compte de l'existence de l'ontologie des utilisateurs comme Amilcare. Ce qui freine ces derniers s'ils ont déjà des ontologies avec des données riches.

3- Il y a des fois un espace entre les annotations et leurs types produits et les classes et propriétés de l'ontologie de l'utilisateur. Dans ce cas, quelques travaux (Handsuh et al, 2005) proposent l'écriture d'un certain genre de règles, telles que des règles logiques.

## EXEMPLE DE GRAMMAIRE JAPE

Si nous reprenons l'exemple précédent concernant la relation 'playes\_a\_role\_in', la grammaire décrivant cette relation est présentée par:

**Rule:Plays\_role**

**Priority: 1**

```
(  
{Tag.lemme == "play"} |  
{Tag.lemme == "have"})  
{SpaceToken}  
{Tag.lemme == "a"} |  
{Tag.lemme == "an"})  
{SpaceToken}  
{Tag.cat == "JJ"} {SpaceToken})?  
{Tag.lemme == "role"}  
{Tag.lemme=="in"})  
):has_role -->  
  
:has_role.RelationShip = {kind = "playes_role", rule=Plays_role}
```

Dans la grammaire ci-dessus, Tag.lemme correspond à la forme lemmatisée du verbe et Tag.cat correspond à la catégorie grammaticale (JJ=adjectif) du terme qui peut être présent entre le verbe et le terme 'role' (le ? signifie qu'il est optionnel).

Ces grammaires sont appliquées sur le texte déjà analysé et exploitent les informations collectées lors de la phase de l'analyse morpho-syntaxique, à savoir, la tokenisation, la lemmatisation et l'étiquetage grammatical.

Le transducteur exécutant (nomino dans notre cas) les grammaires de détection, génère des informations concernant le type de la relation détectée, la phrase où la relation été détectée et l'emplacement exact dans le texte. Ces informations sont ensuite utilisées dans la phase de génération de l'annotation.

# ANNEXE 3 : INVENTAIRE D'ANNOTATION

Cocher les rubriques adéquates :

(à remplir par l'annotateur)

- Niveau de formalisme

formelle                       informelle

- Niveau de partage

tacite                       explicite

- Niveau de lecture

ecriture                       informelle

- Niveau d'extensivité

hyper-extensive       intensive                       informelle extensive

- Niveau de statut

publique                       privée

- Niveau d'utilité

permanente                       transitoire                       globale                       personnelle

Visa du chef de projet

Dimension	Description	Remarques
<b>Formelle versus informelle</b>	Les annotations informelles sont celles écrites en langage naturel dans la marge du document alors que les annotations formelles prennent la forme de métadonnées structurées par l'utilisation d'un langage standard définissant un ensemble de conventions de nommage et de valeurs par défaut. Ces annotations formelles permettent d'assurer l'interopérabilité entre les différentes annotations qui suivent ce standard et leur interprétation par des outils qui implémentent ce même standard.	Différents langages, plus ou moins formels selon les langages de représentation de la connaissance utilisés, permettent de représenter des annotations sémantiques.
<b>Tacite versus Explicite</b>	Les annotations personnelles sont très souvent tacites (un passage souligné ou une marque allusive comme un point d'exclamation par exemple). Elles posent des problèmes d'interprétation pour les autres utilisateurs que l'auteur de l'annotation. Plus les annotations ont pour but d'être partagées avec d'autres utilisateurs, plus elles doivent être explicites.	Les ressources numériques, tout comme leurs annotations, ont souvent pour but d'être partagées entre divers utilisateurs. Par conséquent, les annotations doivent être les plus explicites possibles et pour ce faire, s'appuyer sur des langages formels afin de pouvoir désambiguïser le contenu des documents.
<b>Ecriture versus lecture</b>	Les annotations oscillent entre représenter une aide ou une explication à la lecture du document ou bien constituer une nouvelle forme d'écriture en tant que telle, ajoutant du sens au texte écrit.	Les annotations de ressources numériques représentent non seulement une aide à la lecture de la ressource annotée, mais plus encore à la recherche de ces ressources. Elles permettent également de générer de nouvelles instances de connaissances pouvant être stockées et réutilisées par diverses applications informatiques.
<b>Hyper-extensive versus extensive versus intensive</b>	Une annotation hyper-extensive est une annotation de surface (structurée, un peu à la manière d'un lien hypertexte) alors qu'une annotation dite « intensive » est une annotation de fond (un commentaire descriptif par exemple). Les annotations dites « extensives » représentent un intermédiaire entre ces deux distinctions.	Les annotations sémantiques exploitent surtout l'hyper-extensivité, notamment par l'utilisation des liens hypertextes pour le référencement et l'adressage.
<b>Permanente versus transitoire</b>	Certaines annotations ne sont utiles qu'à son auteur à un moment donné alors que d'autres peuvent perdurer tout en gardant leur valeur ajoutée aussi bien pour l'auteur que pour d'autres utilisateurs	Comme l'objectif premier des annotations sémantiques est le partage et la réutilisation, elles sont donc préférablement permanentes plutôt que transitoires. Mais d'un autre côté les documents numériques sont plus sujets à modification que les documents papiers, surtout les pages Web. Les annotations sémantiques doivent alors évoluer en fonction du contenu modifié

---

<b>Publique versus privée</b>	Les annotations peuvent être destinées à rester dans l'intimité d'un auteur, qui y consigne ses impressions de lecture par exemple, ou au contraire à être divulguées à de multiples utilisateurs. Ces utilisateurs pourront à leur tour compléter les annotations produites par l'auteur initial.	Les annotations sémantiques ont principalement une visée publique, notamment lorsque les ressources documentaires associées sont mises en ligne sur le Web. Néanmoins, un utilisateur peut désirer créer des annotations sémantiques pour son usage personnel.
<b>Globale versus institutionnelle versus personnelle</b>	Les bénéfices attendus des annotations créées varient en fonction des groupes d'utilisateurs qui exploiteront ces annotations.	Dans le cadre du Web Sémantique, les utilisateurs ne sont plus seulement des humains, mais aussi des machines, des agents logiciels. La portée de ces annotations est donc plus institutionnelle, voire globale, que personnelle.