

# Chapitre 20

## Ingénierie des connaissances

L'Ingénierie des connaissances consiste à articuler des travaux sur la nature et la représentation des connaissances dans des logiciels et chez les utilisateurs, sur l'analyse des usages, des sources de connaissances et des interactions homme-machine. Mais l'Ingénierie des connaissances consiste aussi à adapter des approches de génie logiciel, dans le but de mettre des connaissances à disposition des utilisateurs et au sein d'applications informatiques. Les innovations du domaine comprennent donc des méthodes, des logiciels et interfaces d'aide à la modélisation, ainsi que des représentations conceptuelles ou formelles. Nous dressons ici un panorama des questions et des avancées majeures de l'ingénierie des connaissances, dont les résultats récents sont marqués par l'essor du web sémantique, des ontologies et du web social.

### 20.1 Introduction

L'acquisition des connaissances est apparue comme une discipline avec un objet de recherche, les systèmes à bases de connaissances, au milieu des années 80. À la suite des développements des systèmes experts durant la décennie, la question de la modélisation et de l'acquisition des connaissances pour ces systèmes est devenue cruciale et problématique : la question du « goulet d'étranglement »<sup>1</sup> de l'acquisition des connaissances justifia de nombreux travaux de thèses, que ce soit avec des problématiques très cognitives (Aussenac, 1989) ou plus orientés vers la définition de représentations (Cordier et Reynaud, 1991; Charlet, 1992). A la fin des années 90, la perspective s'est élargie et a donné naissance à l'Ingénierie des Connaissances (IC). L'Ingénierie des connaissances se définit comme la partie de l'Intelligence Artificielle (IA) qui étudie et propose des concepts, méthodes, techniques et outils permettant d'acquérir, de modéliser et de formaliser des connaissances dans les organisations dans un but d'opérationnalisation, de structuration ou de gestion au sens large.

L'Ingénierie des connaissances trouve ainsi ses champs d'application dans les domaines où l'objectif est de modéliser les connaissances et les mettre à disposition comme support à

---

Auteurs : N. AUSSENAC-GILLES (IRIT - CNRS, Université de Toulouse), J. CHARLET (AP-HP, Paris & INSERM UMRS 872, Université Pierre et Marie Curie, Paris) et CH. REYNAUD-DELAÎTRE (LRI - Université Paris-Sud).

1. Ce goulet correspondait au fait que le système expert devait permettre de « capturer » l'expertise d'un expert et la représenter au sein du système expert de façon à ce que le système se comporte comme l'expert sollicité dans la même situation. Les difficultés à capturer précisément cette expertise ont amené à parler de goulet d'étranglement.

une activité ou à un raisonnement. Les applications concernées sont celles liées à la gestion des connaissances, à la recherche d'information (sémantique), à l'aide à la navigation, à l'aide à la décision. . Enfin, l'Ingénierie des connaissances entretient des relations étroites avec le Web sémantique qui fait l'objet d'un chapitre spécifique dans le tome 3 de ce livre ?? . Le web sémantique a un statut tout à fait particulier en raison de forts recouvrements avec l'Ingénierie des connaissances via le partage de nombreux outils et méthodes (ontologies, langages de représentation des connaissances, raisonnements, etc.).

Dans la suite de ce chapitre, nous allons, en section 20.2, présenter, selon un axe historique, les modélisations utilisées en Ingénierie des connaissances, puis, en section 20.3, développer les principaux problèmes de recherche qui se posent dans le domaine. La section 20.4 permettra de synthétiser les enjeux méthodologiques du domaine avant de conclure en section 20.5.

## 20.2 Modélisations utilisées

### 20.2.1 La notion de modèle conceptuel

Autour des années 1990, il a été proposé de construire un système à base de connaissances (SBC) en commençant par la description des connaissances du système indépendamment de leur implémentation. Cette représentation, abstraite et finalisée, devait tenir compte des multiples facettes et types de connaissances utiles pour que le système réponde aux besoins identifiés. Le support pour rendre compte de cette représentation a été appelé *modèle conceptuel*. Il s'agit de construire un modèle adapté à la nature des connaissances à décrire pour pouvoir ensuite le représenter dans des formalismes adéquats pour le SBC. Depuis, cette notion est devenue centrale en ingénierie des connaissances, et elle a largement évolué pour recouvrir des réalités différentes suivant les besoins auxquels elle devait répondre, s'adaptant aux nouvelles approches et aux recherches dans le domaine.

La façon dont sont décrites les connaissances conditionne la construction du SBC et surtout le fait que l'on puisse comprendre et s'approprier son fonctionnement. L'acquisition et l'ingénierie des connaissances ont repris à leur compte les travaux de Newell (1982) qui, le premier, a différencié les connaissances à représenter dans un système et l'implémentation de ce système. Newell a fait apparaître la nécessité d'un niveau de description des systèmes qui ne soit pas celui des symboles et langages informatiques, le *niveau des connaissances*. A ce niveau, le comportement du système observé est décrit indépendamment de son implémentation formelle. Ce système est de plus considéré comme un agent rationnel qui dispose de connaissances, doit atteindre des buts, sait effectuer des actions, et est rationnel, c'est-à-dire qu'il choisit (avec ses connaissances) l'enchaînement des actions qui va le mener le plus directement au but. Cette même proposition de Newell a également justifié l'organisation des modèles conceptuels en plusieurs parties. Il distingue les *connaissances du domaine* des *connaissances de raisonnement* à savoir des actions et des buts, modélisés via des *méthodes* et des *tâches*.

### 20.2.2 Les modèles de raisonnement

Les modèles de raisonnement décrivent de façon abstraite le processus de résolution à mettre en œuvre dans un SBC en termes de tâches et de méthodes, les tâches étant réalisées par des méthodes. Une tâche est une description de ce qui doit être fait dans l'application en

termes de buts et de sous-buts. Elle se définit par des connaissances de sortie obtenues à partir des connaissances d'entrée, et ce en fonction de contraintes et ressources disponibles. Pour décrire une résolution de problèmes, on peut mettre en évidence des tâches de différents niveaux, les tâches de plus bas niveau poursuivant des sous-buts pour les tâches plus générales. Les méthodes décrivent comment un but peut être atteint à l'aide d'une série d'opérations réalisées dans un certain ordre. Deux types de méthodes sont généralement distingués : celles qui consistent à décomposer une tâche en sous-tâches et celles qui mettent en œuvre une procédure élémentaire qui atteint directement un but. Progressivement, les concepts de tâches et de méthodes ont été explicitement distingués, suite à une proposition de L. Steels dans son approche « componentielle » (Steels, 1990). Cette distinction a été adoptée pour rendre compte de raisonnements (Klinker *et al.*, 1991; Puerta *et al.*, 1992; Schreiber *et al.*, 1994; Tu *et al.*, 1995) car elle présente l'avantage de décrire séparément le but visé de la façon de l'atteindre. Elle rend ainsi possible la définition explicite de diverses façons d'atteindre un même but en associant plusieurs méthodes à une même tâche. Ce type de modèle est à rapprocher de résultats établis en planification (Camilleri *et al.*, 2008), les planificateurs pouvant servir à rendre opérationnelle les tâches et méthodes ainsi définies, comme cela est proposé dans CommonKADS (Schreiber *et al.*, 1999).

### 20.2.3 Des modèles conceptuels aux ontologies

À la suite des travaux sur les raisonnements, il a semblé intéressant de construire le modèle conceptuel d'une application en combinant réutilisation de composants de modèles de raisonnement et abstraction de connaissances du domaine. Une analyse des connaissances du domaine devient alors nécessaire pour mettre en correspondance les connaissances du domaine et leurs rôles dans le raisonnement (Reynaud *et al.*, 1997). Les connaissances du domaine sont ainsi décrites en un noyau, l'*ontologie du domaine*, regroupant les classes d'entités du domaine, les concepts, et les relations entre ces entités, en leur associant éventuellement un certain nombre de propriétés, de contraintes ou de règles. A ce noyau, se rajoute une couche spécialisée, le modèle du domaine, décrivant des connaissances du domaine plus spécifiques, sous forme d'instances des entités de l'ontologie. Ainsi, l'ontologie définit le vocabulaire logique qui permet d'exprimer des faits et des connaissances du domaine sur lesquelles raisonner. Certains concepts, dits primitifs, sont définis par leur place dans la hiérarchie des classes et leurs propriétés, sous forme de conditions nécessaires. D'autres, dits concepts définis, s'expriment sous forme de conditions nécessaires et suffisantes, à partir des concepts primitifs et de leurs relations.

Le concept d'ontologie a été introduit en IA par le projet ARPA *Knowledge Sharing Effort* (Neches *et al.*, 1991) et c'est Gruber (1993) qui en a proposé une première définition en ingénierie des connaissances. Une définition actuelle, proposée dans (Studer *et al.*, 1998), fait consensus dans le domaine :

*An ontology is a formal, explicit specification of a shared conceptualisation.*

*Conceptualisation refers to an abstract model of some phenomenon in the world by having identified the relevant concepts of that phenomenon. Explicit means that the type of concepts used, and the constraints on their use are explicitly defined.*

*Formal refers to the fact that the ontology should be machine-readable.*

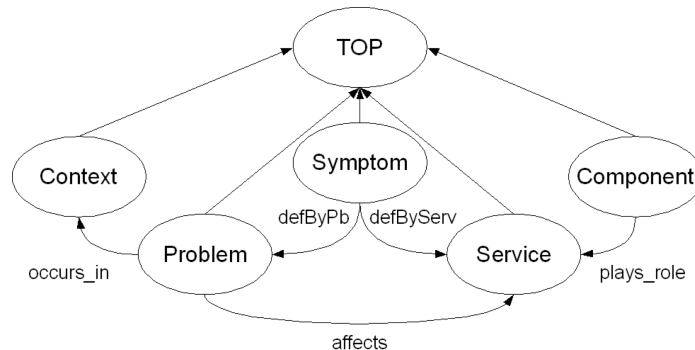


FIGURE 1: Concepts de haut niveau d'une ontologie utilisée dans le domaine du diagnostic.

*Shared reflects the notion that an ontology captures consensual knowledge, that is, it is not private of some individual, but accepted by a group.*

Ainsi, l'ontologie répond à des exigences complémentaires et symétriques : a) en tant que spécification, elle définit une *sémantique formelle* pour l'information permettant son exploitation par un ordinateur ; b) en tant que reflet d'un point de vue – partiel – sur un domaine, que l'on cherche le plus consensuel possible, elle fournit une *sémantique interprétative*. Le fait que l'ontologie soit formelle est à la fois sa qualité par les raisonnements qu'elle permet et son défaut, par la difficulté qu'il y a à la construire et la représenter formellement dans un langage de représentation des connaissances. La figure 1 présente les principaux concepts de haut niveau d'une ontologie construite pour un système de recherche d'information dans le domaine du diagnostic automobile. La notion de symptôme (*symptom*) suppose qu'un composant (*component*), rendant en principe un certain service à l'utilisateur d'un véhicule, soit touché par un problème (*problem*) dans un contexte (*context*) particulier. Dans la représentation formelle de ce modèle, les contraintes de cardinalité des relations *defByPb* et *defByServ* permettent d'exprimer qu'une instance de symptôme ne peut être identifiée que lorsqu'un service et un problème le sont aussi.

En fonction de l'utilisation de l'ontologie, celle-ci va être plus ou moins riche en relations et en concepts définis. Si par exemple, l'ontologie est utilisée dans une application de recherche d'information simple, l'ontologie servira avant tout à organiser hiérarchiquement les concepts du domaine associés à des termes de ce même domaine. On parle alors d'*ontologie légère*. On utilise *Ressource Terminologique-Ontologique* (RTO) lorsque les termes sont représentés par des structures à part entière. Une RTO comporte une taxonomie de concepts, souvent sans concept défini (Reymonet *et al.*, 2007). Si l'ontologie doit servir à décrire des domaines où l'on veut modéliser des raisonnements, elle sera en général plus volumineuse, comportera plus de relations et d'axiomes contribuant à construire des concepts définis et, de façon générale, à décrire les concepts sur lesquels on veut raisonner. On parle ici d'*ontologie lourde*.

Par rapport à leur construction, les ontologies ont suscité de nombreuses attentes fondées sur une idée de généralité. Plusieurs recherches ont contribué à définir ce qui pouvait être générique, et donc réutilisable, dans une ontologie et quelles pouvaient être les méthodes à mettre

en œuvre pour construire les parties non génériques. Ainsi, il est maintenant classiquement admis de distinguer dans une ontologie :

- l’ontologie de haut niveau ou *top*-ontologie, niveau le plus élevé, qui structure les connaissances avec des catégories générales – et qui se veulent universelles – dont l’organisation peut dépendre de réflexions philosophiques. Ainsi, elle définit des catégories qui ont pour vocation de décrire tout type de connaissances. Elles sont le point d’ancrage de niveaux plus spécifiques (Ontologie noyau et de domaine). Une *top*-ontologie est partagée, générique, et sa construction peut s’inspirer de travaux de référence comme les *top-level categories* de Sowa<sup>2</sup>, SUMO<sup>3</sup>, DOLCE<sup>4</sup>, . . . . En effet, les travaux du groupe de travail SUO<sup>5</sup> ou le communiqué du Upper Ontologies Summit<sup>6</sup> confirment qu’il est vain d’essayer d’élaborer une norme, puisqu’il existe plusieurs manières d’envisager les grandes catégories d’entités du monde ;
- l’ontologie noyau ou *core*-ontologie, qui fournit les concepts structurants d’un domaine et décrit les relations entre ces concepts – en médecine par exemple, on y trouve des concepts de *diagnostic*, *signe*, *structure anatomique* et des relations comme celles liées à la *localisation* d’une pathologie sur une structure anatomique (*cf.* GFO-Bio<sup>7</sup>) ; dans le droit, l’ontologie LKIF-Core<sup>8</sup> propose les notions de *norme*, *action légale* et *rôle légal* ;
- l’ontologie du domaine, qui décrit les concepts du domaine tels qu’ils sont manipulés par les professionnels. C’est la partie la plus spécifique de la modélisation. Il n’y a pas de frontière claire entre une *core*-ontologie et une ontologie du même domaine qui l’utilise ; elles sont construites de façon concomitante. Les approches à partir de corpus, complétées par la réutilisation de thésaurus ou de terminologies du domaine modélisé, sont souvent utilisées pour construire cette partie des ontologies (*cf.* § 20.4.1).

### 20.3 Problèmes considérés et résultats

Un des faits marquants de l’évolution de l’ingénierie des connaissances est d’avoir diversifié les sources de connaissances utilisées dans les systèmes de traitement d’information « intelligents », et ce pour tirer profit à la fois de leur complémentarité et de l’évolution des techniques disponibles pour les analyser. Nous faisons tout d’abord un panorama des sources de connaissances sur lesquelles les méthodes d’Ingénierie des connaissances se sont successivement focalisées et des questions de recherche que pose le passage des sources de connaissances aux modèles. Par modèle, nous faisons ici référence aux différents types de modèles cités en section 20.2 correspondant à des représentations des connaissances d’un SBC (modèle conceptuel), d’un processus de résolution (modèle de raisonnement) ou de connaissances du domaine spécifiques (modèle du domaine). Nous montrons ensuite comment des méthodes et techniques de modélisation proposées en Ingénierie des connaissances permettent de dépasser ces problèmes, et les questions de recherche associées. Enfin, nous présentons les résul-

2. <http://www.jfsowa.com/ontology/toplevel.htm>

3. <http://www.ontologyportal.org/>

4. <http://www.loa-cnr.it/DOLCE.html>

5. <http://suo.ieee.org/>

6. [http://ontology.cim3.net/cgi-bin/wiki.pl?UpperOntologySummit/](http://ontology.cim3.net/cgi-bin/wiki.pl?UpperOntologySummit/UosJointCommunique)

UosJointCommunique

7. <http://www.onto-med.de/ontologies/gfo-bio/index.jsp>

8. <http://www.estrellaproject.org/lkif-core/>

tats portant sur la réutilisation de modèles puis le lien avec les travaux en représentation des connaissances.

### 20.3.1 Les sources de connaissances

Historiquement, « la connaissance » a d'abord fait référence à une expertise humaine dont devait rendre compte la base de connaissances de *systèmes experts*. Ces connaissances étaient techniques et spécialisées. Elles correspondaient à des savoir-faire de haut niveau rarement verbalisés, qu'il s'agissait de pérenniser et de transmettre à un système informatique sous forme de règles de production<sup>9</sup>. Les systèmes experts ont évolué vers des systèmes à base de connaissances dont le rôle était d'aider les utilisateurs, en privilégiant l'efficacité et non la fidélité à un raisonnement humain. Les connaissances de référence sont alors devenues des connaissances partagées et les systèmes ont mis en œuvre des modes de raisonnement qui leur étaient propres. Aujourd'hui, de très nombreuses applications (correcteurs orthographiques, systèmes d'aide à la décision, systèmes d'aide à la facturation, etc.) intègrent des *modules de connaissances*. L'objectif est de réaliser des tâches, soit de manière autonome, soit de manière coopérative, en s'adaptant aux différents contextes d'utilisation et aux profils des utilisateurs. Les connaissances mises à disposition, comme support à un raisonnement ou à une activité, correspondent alors à des savoirs techniques, consensuels et partagés, sous forme de règles ou de plans d'action, ou à des descriptions structurées et finalisées d'un domaine.

L'évolution historique des systèmes informatiques « à base de connaissances » permet d'identifier les différentes dimensions des connaissances à prendre en compte : les connaissances individuelles expertes ; les connaissances liées aux pratiques, aux activités et aux usages individuels, les connaissances portant sur les organisations, les connaissances consensuelles et partagées relatives à des domaines d'application, les connaissances de sens commun, les connaissances provenant du recoupement de données ou d'informations réparties sur le web.

C'est pour accéder à ces différents types de connaissances que de nouvelles sources ont été prises en compte. Ainsi, la place occupée par les documents a augmenté avec la disponibilité croissante des documents numériques. Depuis les premières études sur l'acquisition des connaissances pour les systèmes experts, l'Ingénierie des connaissances s'intéresse aux documents, en particulier les documents textuels, comme porteurs de sens et révélateurs de connaissances depuis les premières études sur l'acquisition des connaissances pour les systèmes experts. Les documents sont exploités pour leur contenu, en complément ou en alternative aux experts et spécialistes du domaine. Des données peuvent également être la source de connaissances via des mécanismes d'extraction de connaissances à partir de données (ECD). Enfin, des composants de modèles de connaissances pré-existants peuvent être réutilisés lorsqu'ils portent sur des connaissances consensuelles et partagées. Il peut s'agir de méthodes de résolution de problèmes applicables à différents domaines (les bibliothèques de méthodes sont un des résultats majeurs des projets européens KADS et CommonKADS<sup>10</sup> (Schreiber *et al.*, 1999), de modèles du domaine ou ontologies correspondant à des représentations structurées définissant les concepts d'un domaine, ou de ressources telles que des bases de données lexicales (p. ex. WordNet<sup>11</sup> qui répertorie, classe et met en relation le contenu sémantique et

9. Pour un historique sur les systèmes à base de connaissances, lire (Aussenac, 1989), (Stefik, 1995), (Aussenac-Gilles *et al.*, 1996), (Charlet *et al.*, 2000)

10. <http://www.commonkads.uva.nl/>

11. <http://wordnet.princeton.edu/wordnet/>

lexical de la langue anglaise) ou des thesaurus correspondant à des vocabulaires normalisés sur un domaine, constitués d'ensembles structurés de termes.

### **20.3.2 Comment passer des sources de connaissances aux modèles : questions de recherche**

Une des problématiques centrales et originales de l'ingénierie des connaissances est d'outiller – techniquement et méthodologiquement – le passage des sources mentionnées en 20.3.1 aux modèles identifiés en partie 20.2. Ces techniques font souvent appel à des logiciels mais aussi à des cadres d'analyse ou à des grilles d'observation venant d'autres disciplines. La recherche en ingénierie des connaissances est bien celle d'une ingénierie au sens où il s'agit d'innover autant dans la création d'outils, langages, méthodes que dans leur sélection et leur adaptation lorsqu'ils existent, mais surtout dans leur agencement pertinent au sein de guides méthodologiques et de plateformes de travail intégrées. L'innovation porte autant sur la nature et le développement de ces outils que sur la définition des conditions de leur utilisation et de la complémentarité de leurs interactions, pour traiter des types particuliers de connaissances, à chaque étape du développement d'une application.

Depuis près de 20 ans, les recherches méthodologiques en Ingénierie des connaissances soulèvent des problématiques transverses, reformulées et renouvelées en fonction des sources de connaissances, de la nature des modèles construits ainsi que des utilisations et raisonnements prévus à partir de ces modèles.

#### **Comment construire un modèle ?**

Deux courants méthodologiques complémentaires ont défini d'abord des étapes et des techniques différentes (Aussenac-Gilles *et al.*, 1992). Les *méthodes ascendantes* privilégient les analyses de données, d'abord sur la base des besoins identifiés puis en fonction des parties de modèle à renseigner. Ces démarches mettent l'accent sur les logiciels et techniques de recueil, d'extraction, de fouille et d'identification de connaissances, puis sur l'aide à la caractérisation abstraite des connaissances (classification, structuration, identification des méthodes et raisonnements). A l'inverse, les *méthodes descendantes* privilégient la *réutilisation* pour construire des modèles par adaptation et assemblage de composants existants, le recueil et l'extraction étant au service du choix des composants puis de la spécialisation du modèle. Pour unifier l'ensemble, on peut considérer la modélisation comme un cycle alternant phases ascendantes et descendantes, progressant d'étapes essentiellement consacrées au recueil ou à la réutilisation à des phases de représentation de plus en plus formelle. La plupart des méthodes et outils présentés en 20.3.3 combinent ces deux courants, et nous insistons en 20.3.4 sur les résultats liés à la réutilisation des modèles.

#### **Comment exploiter la complémentarité entre différentes sources de connaissances ?**

La diversité des connaissances utilisées constitue un des moyens de parvenir à des modèles plus précis ou d'automatiser une partie de leur construction. Une méthode d'Ingénierie des connaissances doit identifier les sources qu'elle permet de traiter, les outils et techniques permettant de les exploiter au mieux puis guider l'*articulation* de ces outils au sein d'une dé-

marche, pour assurer l'exploitation complémentaire de leurs résultats et construire un modèle. C'est ce qu'illustrent les résultats présentés en 20.3.3.

### **De quoi les modèles sont-ils composés et quel est le niveau optimal de formalisation requis ?**

Les modèles eux-mêmes associent plusieurs types de connaissances : chaque méthode d'Ingénierie des connaissances s'interroge et fait des propositions sur la nature des modèles à élaborer, la manière de les construire ainsi que les langages de représentation (plus ou moins formels) associés (*cf.* section 20.3.5).

### **Comment l'ingénierie des modèles intègre-t-elle l'objectif de leur utilisation ?**

Plusieurs travaux ont montré que les modèles conceptuels étaient d'autant plus pertinents qu'ils étaient finalisés, produits pour des applications spécifiques. L'ingénierie des connaissances ne se limite donc pas à produire des modèles mais elle est concernée par leur utilisation dans la mesure où cette utilisation détermine en partie leur contenu, leur structure, et de ce fait la manière de les construire. La visée d'utilisation d'un modèle a donc un impact sur les choix méthodologiques et sur les formalismes de représentation retenus (Bourigault *et al.*, 2004).

### **Comment favoriser la réutilisation des modèles ?**

La réutilisation d'éléments de connaissances déjà structurés est souvent privilégiée pour réduire le coût de la modélisation de connaissances. Ceci n'est possible que si l'on connaît bien les principes selon lesquels ces modèles ont été construits, si on peut les comparer, le combiner, si le fait d'en reprendre des fragments et de les recomposer au sein de nouveaux modèles est techniquement possible et a du sens. Ces questions se retrouvent aujourd'hui dans les travaux sur l'alignement, la réutilisation et la composition de parties d'ontologies pour en construire de nouvelles.

### **Comment assurer l'évolution des modèles en lien avec leur contexte d'utilisation ?**

Les modèles de connaissances utilisés dans des applications trouvent leur place dans un cycle de vie qui intègre leur évolution, rendue nécessaire suite à l'évolution des sources de connaissances, des connaissances du domaine ou des besoins des utilisateurs. Depuis 2008, l'évolution des ontologies, posée comme un enjeu fort de leur utilisation, fait l'objet de recherches définissant un cycle d'évolution, des moyens d'identifier les connaissances à modifier tout en assurant la cohérence logique du modèle (Stojanovic, 2004), (Luong, 2007).

## **20.3.3 La construction de modèles : techniques, méthodes et outils**

Pour faire des propositions pratiques en matière d'accès aux connaissances à travers les personnes ou les documents qui sont supposés en fournir des indications, l'Ingénierie des connaissances a forgé des solutions qui lui sont propres : des techniques et des outils, qui peuvent être intégrés dans des méthodes et des plateformes. Elle s'est pour cela largement



inspirée de disciplines proches en fonction de la source de connaissances considérée : ces disciplines ont couvert successivement la psychologie cognitive puis l'ergonomie, la terminologie et enfin la linguistique de corpus.

La construction de modèles nécessite d'accéder aux connaissances fournies par les différentes sources. Les techniques d'accès diffèrent selon les types de ces sources, certaines techniques allant jusqu'à faire émerger des connaissances nouvelles, non explicitées. *Technique* fait ici référence à des « modes opératoires » préconisant des modes de choix ou de création de situations de production ou d'utilisation de connaissances, puis à la manière de repérer / recueillir / extraire ou analyser ces données et enfin des propositions pour interpréter, dépouiller, structurer les fruits de cette analyse. Nous nous intéresserons successivement à deux sources de connaissances, l'expertise humaine et les documents textuels.

### **L'expertise humaine comme source de connaissances**

Concernant l'expertise humaine, les approches ont évolué d'une vue *cognitiviste*, faisant l'hypothèse d'une passerelle possible entre des représentations chez les individus aux représentations informatiques, à des approches *constructivistes* puis de *cognition située*, prenant en compte la dimension contextuelle et parfois collective des savoirs. Dans le premier cas, il s'agit de localiser puis de rendre explicite des savoir-faire et de les représenter. Ce point de vue, qui correspond historiquement à celui des systèmes experts, considère que les connaissances sont accessibles chez un ou plusieurs experts et qu'il suffit de les expliciter pour construire un système produisant les mêmes raisonnements. La psychologie cognitive a fourni de nombreuses indications sur les modalités d'entretien, leur analyse, les avantages et limites de chacune pour sur l'étude des phénomènes cognitifs humains (Darses et Montmollin, 2006). Ces techniques ont été adaptées et mises en œuvre pour l'extraction de connaissances auprès d'experts, dont plusieurs sont disponibles dans (Aussenac, 1989), (Shadbolt *et al.*, 1999) ou (Dieng-Kuntz *et al.*, 2005). On y distingue les méthodes dites *directes*, qui consistent à interroger l'expert, à le faire s'exprimer oralement de manière plus ou moins guidée, des méthodes *indirectes*, comme l'analyse de *grilles répertoires*, qui se basent sur une interprétation des éléments recueillis alors que l'expert exécute des tâches faisant appel à son expertise mais incluses dans son activité habituelle.

Cette vue *cognitiviste* a été progressivement remise en question pour mieux répondre au caractère situé des connaissances. Les savoir-faire n'étant accessibles qu'à travers leur mise en œuvre en situation de résolution de problème ou de décision, l'Ingénierie des connaissances a repris à l'ergonomie des techniques d'analyse de la tâche et de l'activité.

Un résultat important de ces travaux a été de poser les bases de l'acquisition des connaissances comme champ disciplinaire s'intéressant aux connaissances pour *elles-mêmes* avant de considérer leur formalisation et leur exploitation dans un système donné. L'adoption d'un point de vue *constructiviste* et la prise en compte de méthodes établies en génie logiciel ont ensuite conduit à des propositions méthodologiques guidant l'ensemble du processus d'acquisition des connaissances. Plusieurs des méthodes ainsi définies dans des projets d'envergure, essentiellement européens, seront présentées en 20.3.3.

L'intégration de connaissances dans des logiciels a pour objectif de mieux aider l'utilisateur et a pour effet de modifier son mode de travail. Elle pose le problème de l'analyse de ses pratiques et de celles de collaborateurs, de l'étude de leurs activités et usages d'outils ou supports, de la prise en compte de leur contexte organisationnel, ce qui renvoie à des approches

ergonomiques, sociologiques ou managériales. Les résultats de ces analyses ont d'abord été restitués de manière figée, sous forme de modèles (modèles de la tâche, de l'interaction et de l'organisation par exemple dans CommonKADS) (Schreiber *et al.*, 1999). Ces modèles ont été rendus opérationnels par des langages de tâches et méthodes comme LISA, Task (Jacob-Delouis et Krivine, 1995) ou CML (Schreiber *et al.*, 1994). Pour aller plus finement dans la prise en compte des activités et la réinjection de ces connaissances au service d'une aide précise et contextuelle de l'utilisateur, la notion de trace d'activité a été approfondie. Ainsi, Laflaquière *et al.* (2008) définissent ce que peuvent être les traces d'une activité d'utilisation de logiciel ou de système documentaire, de manière à pouvoir récupérer, représenter et conserver ces traces puis les exploiter ou les réutiliser.

### **Les documents textuels comme sources de connaissances**

Concernant les documents textuels, deux problèmes se posent, celui de leur sélection et celui de leur analyse, qu'il s'agisse de documents techniques, de documents liés à une activité ou à tout un domaine d'application. L'analyse de documents peut porter sur le langage naturel formant le texte ou sur la structure de ces textes, explicitée sur papier ou écran par une mise en forme matérielle, et électroniquement par un étiquetage (Virbel et Luc, 2001). On parle dans ces derniers cas de documents structurés ou semi-structurés (documents XML). Nous décrivons dans un premier temps les atouts de l'analyse de documents textuels, puis les techniques et outils utilisés pour cette analyse.

**Atouts de l'analyse de documents textuels.** Les textes sont des ressources très riches en connaissances. L'analyse de textes a ainsi toujours été présente en ingénierie des connaissances mais la manière de l'aborder a radicalement changé après 1990. On ne cherche plus à reconstituer automatiquement les modes de compréhension d'un texte par un individu (Aussenac-Gilles *et al.*, 1995). L'analyse de textes doit son essor aux avancées du traitement automatique du langage naturel (TALN) écrit, qui ont débouché sur des logiciels spécialisés d'analyse robuste. Cette maturité du TAL a été concomitante au déploiement des ontologies. L'analyse de textes écrits a alors trouvé un champ d'application privilégié dans la construction d'ontologies et leur utilisation pour l'annotation sémantique de documents. L'hypothèse forte derrière l'exploitation automatique de textes est qu'ils fournissent les éléments stables, consensuels et partagés d'un domaine (Bourigault et Slodzian, 1999; Condamines, 2003). Or ce n'est pas toujours le cas, et deux éléments clés conditionnent l'obtention d'un modèle de qualité : tout d'abord, la constitution d'un corpus pertinent en amont du processus ; ensuite, une contribution régulière de spécialistes du domaine ou de modélisation pour assurer l'interprétation des résultats. L'analyse des textes est aussi utilisée pour la construction de ressources proches des ontologies telles que les thesaurus, index, lexiques ou bases de connaissances terminologiques.

**Techniques et outils pour l'analyse de textes.** La visée de l'analyse de textes en Ingénierie des connaissances est de repérer, de manière automatique ou coopérative, des éléments linguistiques dont l'interprétation permettra de construire des parties de modèles conceptuels.

*Les approches linguistiques* s'appuient sur les formulations présentes dans les textes pour repérer des contextes riches en connaissances (Barriere et Agbago, 2006). Il peut s'agir de groupes nominaux ou verbaux ayant une cohérence forte, et dont l'usage laisse penser qu'ils

sont des termes désignant des concepts du domaine, ou exprimant des relations entre concepts. Il peut s'agir aussi d'indices plus minces mettant en relation des éléments plus diffus et pour lesquels l'analyste humain doit reconstruire les liens de référence avant d'aboutir à des éléments de connaissance, des axiomes ou des règles. Les recherches en sémantique lexicale, terminologie et sémantique de corpus établissent des résultats en amont de la mise en œuvre informatique de ce type d'approche (Condamines, 2003; Constant *et al.*, 2008).

Les *approches statistiques* traitent le texte dans sa globalité et tirent profit des redondances, des régularités d'usage, des co-occurrences pour identifier des expressions figées et des termes, mais aussi des mots ou groupes de mots (clusters) ayant des comportements ou des contextes linguistiques analogues. Plusieurs de ces techniques sont présentées dans le livre *Foundations of Statistical Natural Language Processing* de Manning et Schütze (1999).

Dans les deux cas, des analyses préalables du texte, comme son découpage en phrases puis en mots (token) ou l'analyse grammaticale de ces mots, sont nécessaires. Le lecteur pourra trouver une description de ces travaux dans le chapitre « Traitement du langage » du tome 3 de ce livre ???. Plus ces traitements sont sophistiqués (comme l'analyse syntaxique complète de phrases), plus on peut définir des règles précises d'interprétation automatique. Malheureusement, la plupart des logiciels réalisant des analyses sophistiquées sont peu robustes et souvent disponibles pour quelques langues seulement, l'anglais étant très privilégié. De plus, ces analyses nécessitent parfois des ressources (lexiques, dictionnaires sémantiques, etc.) qui sont rarement disponibles dans toutes les langues.

Lorsque la structure des documents est accessible via des marqueurs explicites, les techniques d'analyse peuvent combiner des approches linguistiques à l'exploitation de la structure de façon à tirer parti à la fois de la structure du texte et de la langue elle-même (Kamel et Aussenac-Gilles, 2009). L'idée sous-jacente est que le découpage structurel du document contribue à une caractérisation sémantique de son contenu.

En matière de construction d'ontologies, l'analyse de texte répond à deux types de besoins (Maedche, 2002; Cimiano *et al.*, 2010) : l'identification des concepts et de leurs propriétés ou relations, tâche appelée *ontology learning* ; la recherche d'instances de concepts et de relations entre instances, ou *ontology population*. Des outils similaires peuvent être utilisés dans les deux cas : il s'agit de repérer en corpus des éléments linguistiques *riches en connaissances* (Meyer, 2000), des indices linguistiques qui peuvent être interprétés en tant que connaissances à modéliser.

La modélisation de vocabulaires a motivé la réalisation de logiciels spécialisés débouchant sur des résultats de plus haut niveau qui apportent des briques à intégrer dans un modèle : extracteurs de termes – Terminoweb (Barriere et Agbago, 2006), Syntex-Upéry (Bourigault, 2002), TermExtractor (Drouin, 2003) –, extracteurs de relations – Caméléon (Aussenac-Gilles et Jacques, 2008), Nooj<sup>12</sup>, Unitex<sup>13</sup>, RelExt (Schutz et Buitelaar, 2005) –, langages de patrons, extracteurs d'entités nommées (Poibeau et Kosseim, 2000) et de termes, recherche d'instances ou de relations entre instances (comme avec la plate-forme KIM<sup>14</sup>). La construction de modèles à partir de textes a également bénéficié de plateformes de TAL (GATE<sup>15</sup>, Linguas-

---

12. <http://www.nooj4nlp.net/>

13. <http://www-igm.univ-mlv.fr/~unitex/>

14. <http://www.ontotext.com/kim/>

15. <http://gate.ac.uk/>

tream<sup>16</sup>, UIMA<sup>17</sup>) qui permettent de développer des chaînes de traitement spécialisées pour la construction d'ontologies (par ex., Spart est un extracteur de relations sémantiques défini à partir de GATE). Enfin, des chaînes de traitements spécifiques, comme Text2Onto (Cimiano et Völker, 2005), et la version désormais intégrée par NeOn<sup>18</sup>, ont permis d'évaluer les points forts et limites de cette approche en poussant le plus possible l'automatisation et l'exploitation de l'apprentissage automatique. Les travaux actuels combinent analyse de textes, réutilisation de composants d'ontologie et interprétation humaine. On en trouvera un panorama assez complet dans (Cimiano *et al.*, 2010).

### Plateformes de modélisation

Les plateformes de modélisation intègrent l'accès à des sources de connaissances, ou à leurs traces, des techniques et logiciels des types que nous venons de présenter ainsi que des techniques et des langages de modélisation. Elles implémentent une méthodologie en définissant une chaîne de traitements qui guide le processus de modélisation. Dans la suite, nous présenterons en premier lieu les travaux les plus significatifs en matière de modélisation du raisonnement, résultats marquants des années 90, puis nous nous focaliserons sur les méthodes et plateformes de construction d'ontologies qui sont d'actualité depuis une dizaine d'années.

**Méthodes de modélisation du raisonnement.** Des guides méthodologiques ont été établis pour mieux maîtriser le développement de gros projets de SBC. Ces méthodes ressemblent dans leur principe à celles qui existent en génie logiciel car elles accordent beaucoup d'importance à la modélisation. Dans les deux cas, il s'agit de gérer les cycles de développement et de construire un ou des modèles du système à concevoir. La réalisation d'applications est vue comme un processus de transformation de modèles, au cœur desquels se situent les modèles conceptuels définis en section 20.2.1. Pour cela, ces méthodologies proposent un ensemble de primitives épistémologiques qui caractérisent à un haut niveau (en termes de connaissances) les capacités d'inférence du système devant être construit. Ces primitives définissent des structures génériques à instancier pour la représentation des connaissances.

Les travaux des années 80 et 90 ont fait évoluer la notion de modèle conceptuel, en accordant une place importante au modèle de raisonnement, et ont renouvelé les langages associés, en articulant les notions d'inférence et de tâche. Du point de vue méthodologique, ces recherches ont montré en quoi les primitives de modélisation fournissent une grille pour le recueil et l'interprétation de connaissances, et en cela guident leur modélisation. De ces travaux, en particulier les résultats sur les Tâches Génériques de Chandrasekaran (1983), a émergé l'intérêt de disposer d'éléments de modèles génériques, et réutilisables par instantiation à une application particulière. Plus tard, la méthode CommonKADS a montré l'intérêt que ces éléments soient adaptables et modulaires. Ces principes sont généraux dans la mesure où ils s'appliquent quels que soient la tâche, le domaine et le raisonnement mis en œuvre dans l'application développée. Les techniques de modélisation et les composants réutilisables sont intégrés au sein de plateformes qui offrent aussi des techniques de recueil d'expertise.

Faisant suite aux travaux sur les Tâches Génériques et sur les méthodes à limitation de rôles (Marcus et McDermott, 1989), et à partir des propositions de L. Steels dans l'approche

---

16. <http://linguastream.org/>

17. [http://domino.research.ibm.com/comm/research\\_projects.nsf/pages/uima.index.html](http://domino.research.ibm.com/comm/research_projects.nsf/pages/uima.index.html)

18. <http://www.neon-toolkit.org/>

composante COMMET et dans l'atelier KREST (1990), plusieurs travaux ont distingué explicitement les concepts de tâche et de méthode. Cette distinction présente l'avantage de décrire séparément le but visé de la manière de l'atteindre et rend possible la définition explicite de différentes façons d'atteindre un même but par l'association de plusieurs méthodes à une même tâche. Ces différents travaux ont été pris en compte par le projet européen KADS (Schreiber et Wielinga, 1992), pionnier en matière d'ingénierie des connaissances, pour déboucher sur la méthode et la plateforme la plus aboutie, CommonKADS (Schreiber *et al.*, 1999).

CommonKADS permet la construction de plusieurs modèles liés entre eux, nécessaires à la spécification d'un système à base de connaissances, dont le modèle organisationnel permet de prendre en compte l'utilisation des connaissances dans leur environnement. Le modèle d'expertise du système est désormais reconnu comme nettement différent d'un modèle cognitif de l'expert humain. Ce modèle d'expertise est décrit selon trois points de vue : les tâches, les modèles du domaine, les méthodes. Chaque méthode de résolution de problème est paramétrable et son adaptation est définie grâce à un questionnaire qui guide pas à pas, pour une application donnée, le choix d'une des méthodes possibles pour réaliser chacune des tâches principales du processus de raisonnement. Les tâches décrivent ce qui doit être réalisé par le SBC. Les modèles du domaine décrivent les connaissances mises en jeu dans la résolution de problèmes. Les méthodes décrivent comment les connaissances sont utilisées pour exécuter une tâche. Une méthode peut décomposer une tâche en sous-tâches ou réaliser une ou plusieurs tâches. La méthodologie propose de réaliser le modèle d'une application par itérations successives sur ces différents points de vue. Il s'agit de perspectives toutes nécessaires et complémentaires. Le choix d'un modèle du domaine est lié au choix d'une certaine méthode de résolution car la méthode définit les rôles de connaissances à pourvoir. Une méthode définit en grande partie la nature des sous-tâches qu'elle contrôle, etc. L'ambition de la méthodologie est donc d'identifier et de modéliser toutes les relations qui existent entre méthodes, tâches et modèles du domaine.

**Méthodes et plateformes de construction d'ontologies.** La construction d'ontologie a bénéficié de ces méthodologies et elle a vu le jour à partir du moment où la réutilisation des modèles du domaine a fait ressortir l'intérêt d'élaborer des modèles de qualité, consensuels et conçus selon de « bons » principes qui facilitent la réutilisation et l'adaptation. La construction d'ontologies soulève des problèmes variés :

1. Définir ce que doit être le contenu de l'ontologie et s'assurer de sa qualité ;
2. Exploiter efficacement les sources de connaissances disponibles via, par exemple, des processus d'analyse de textes ou de réutilisation d'ontologies existantes ;
3. Faciliter / assister par des outils la construction par le modélisateur ;
4. Définir un cadre méthodologique et la démarche à mettre en œuvre dans l'enchaînement des différentes tâches.

Des plateformes de gestion d'ontologies offrent des environnements uniformes et cohérents de développement d'ontologies, qui assurent une aide à ces différentes tâches, intègrent pour cela divers outils, et s'appuient sur une méthodologie qui assure l'enchaînement de ces tâches au cours du processus de développement.

Parmi les nombreuses méthodes de construction d'ontologies<sup>19</sup>, nous limiterons dans le cadre de cet article à la présentation des méthodes OntoClean, ARCHONTE et OntoSpec, qui toutes trois se focalisent sur la qualité du contenu de l'ontologie.

La méthode OntoClean a été conçue par Guarino et Welty (2004). Les premières idées sont décrites dans une série d'articles publiés en 2000, le terme OntoClean est apparu en 2002. Inspirée de la notion d'ontologie formelle et de principes de philosophie analytique, cette méthode constitue un apport important en tant que première méthode formelle de l'ingénierie ontologique. Elle propose d'analyser les ontologies et de justifier les choix ontologiques en exploitant des méta-propriétés de classes formelles indépendantes de tout domaine d'application, initialement au nombre de quatre (l'identité, l'unité, la rigidité et la dépendance).

La méthode ARCHONTE (ARCHitecture for ONTological Elaborating) élaborée par Bachimont *et al.* (2002) est une méthode ascendante de construction d'ontologies à partir des textes du domaine en trois étapes. Une première étape consiste à choisir les termes pertinents du domaine, et à les normaliser sémantiquement en précisant les relations de similarité et de différences que chacun des concepts entretient avec ses frères et son père (principe de la *sémantique différentielle*). La seconde étape consiste à formaliser les connaissances (*engagement ontologique*). Il s'agit de construire une ontologie différentielle en ajoutant des propriétés ou annotations, et en définissant les domaines et co-domaines des relations. Enfin, une troisième étape consiste à opérationnaliser l'ontologie dans un langage de représentation des connaissances. On obtient alors une *ontologie computationnelle*.

Enfin, OntoSpec (Kassel, 2002) est une méthode de spécification semi-informelle d'ontologies. Sa conception a été motivée par le fait que les définitions en langue naturelle associées aux entités conceptuelles permettent à des utilisateurs de l'ontologie de participer à son élaboration avec un ingénieur de la connaissance. Par ailleurs, cette méthodologie propose un cadre constitué d'une typologie de propriétés pouvant intervenir dans les définitions de concepts et de relations et de règles, pour paraphraser en langue naturelle ces types de propriétés. Ce cadre est un guide pour modéliser et facilite ensuite le passage à une ontologie formelle.

Le cœur des plates-formes de construction d'ontologie est en général un éditeur d'ontologies. C'est le cas de Protégé<sup>20</sup>, élaboré à l'université de Stanford, très utilisé pour créer ou modifier des ontologies RDFS ou OWL, ou de Swoop<sup>21</sup> développé à l'université de Maryland pour des ontologies OWL légères, ou encore de Hozo<sup>22</sup> conçu à l'Université d'Osaka, dont la particularité est de mettre l'accent sur la notion de rôle et de distinguer les concepts dépendant de contextes particuliers des concepts de base, ceci afin de faciliter la construction d'ontologies réutilisables. À côté de cette fonction d'édition, beaucoup d'autres fonctionnalités peuvent être intégrées dans des plateformes, parmi lesquelles on trouve : des fonctions de traduction de Schema XML, d'aide à l'accès à des modules d'ontologies ou d'aide au partitionnement d'ontologies, des modules de traduction de vocabulaires, des accès à des moteurs de recherche d'ontologies, à des étiqueteurs morpho-syntaxiques (p. ex. Tree-Tagger<sup>23</sup>), des modules d'aide à la personnalisation d'ontologies, de génération de documentation, d'aide à la gestion de leur évolution, de leur évaluation, des fonctionnalités d'alignement ontologies, des services de rai-

---

19. Pour une synthèse des principales méthodologies de construction d'ontologies, se reporter à (Fernández-López et Gómez-Pérez, 2002).

20. <http://protege.stanford.edu/>

21. <http://code.google.com/p/swoop/>

22. <http://www.hozo.jp/ckc07demo/>

23. <http://www.ims.uni-stuttgart.de/projekte/corplex/TreeTagger/>

sonnement et d'inférence, des services d'aide à la navigation, de visualisation ou encore des aides à la réutilisation d'ontologies. L'ensemble de ces fonctionnalités correspondent, à titre d'exemples, à des plug-ins de la plateforme Neon<sup>24</sup>.

Certaines des plateformes sont spécialisées pour traiter un type de données. Ainsi, Text2Onto, successeur de TextToOnto, et DaFOE4App sont plus spécialement conçues pour exploiter des textes comme sources de connaissances. Text2Onto (Cimiano et Völker, 2005) rassemble des logiciels de fouille de textes et de génération d'informations structurées à partir de documents faiblement structurés. Pour construire une ontologie, Text2Onto est associé à la plate-forme KAON (Karlsruhe Ontology Management Infrastructure) (Oberle *et al.*, 2004). DaFOE4App (Differential and Formal Ontology Editor for Applications) (Charlet *et al.*, 2010), dont la conception reprend, entre autres, les principes d'ARCHONTE, met en avant la dimension linguistique et couvre toutes les étapes allant de l'analyse d'un corpus (annoté au sein d'une plateforme de TALN) à la définition d'une ontologie formelle du domaine. Elle garantit la persistance, la traçabilité et le dimensionnement des modèles (plusieurs milliers de concepts). La plate-forme TERMINAE (Szulman, 2011; Aussenac-Gilles *et al.*, 2000), antérieure à DaFOE4App, a évolué avec les spécifications de cette dernière et propose un outil<sup>25</sup> testé à l'occasion de nombreux projets. Pour terminer cette liste, PlibEditor est plus particulièrement adapté au traitement des bases de données. Il permet de réaliser toutes les phases de création d'ontologies et d'importer ou d'exporter ontologies et données. Utilisé en complément d'OntoDB, un système de gestion de base de données à base ontologique, il permet la mise en œuvre d'une approche de conception de bases de données fondée sur des ontologies du domaine (Fankam *et al.*, 2009).

### 20.3.4 Réutilisation de modèles

Tout comme le génie logiciel vise la réutilisabilité de composants logiciels, l'acquisition des connaissances cherche à favoriser la réutilisation de composants de connaissances. Cette réutilisation peut se concevoir de différentes manières.

Initialement proposée dans le cadre du projet KADS, la réutilisation de modèles de raisonnement consiste à reprendre et adapter à la réalisation de tâches spécifiques des modèles de tâches exprimés dans une terminologie indépendante de tout domaine d'application. Cette approche est séduisante mais l'adaptation d'un modèle de raisonnement à un domaine pose deux types de problèmes. D'une part, une application met souvent en œuvre plusieurs raisonnements auxquels correspondent plusieurs modèles qu'il faut pouvoir distinguer et combiner. D'autre part, la réutilisation et l'adaptation de modèles génériques prédéfinis à une application donnée s'avère un travail long et difficile. Une des raisons en est que la tâche à réaliser et la base de connaissances du système doivent tous deux être exprimés dans les termes d'un domaine d'application donné, alors que les méthodes réutilisables, issues de bibliothèques, sont exprimées à l'aide d'un vocabulaire générique. Ainsi, adapter des éléments de résolution de problèmes à une application est d'abord un problème de mise en correspondance de termes. Ce constat a alors ouvert la voie à des approches plus flexibles, au sein desquelles les éléments réutilisés et adaptés sont de granularité plus fine. Il ne s'agit plus de réutiliser des modèles génériques complets de raisonnement mais des éléments de raisonnement.

24. [http://www.neon-toolkit.org/wiki/Neon\\_Plugins](http://www.neon-toolkit.org/wiki/Neon_Plugins)

25. <http://lipn.univ-paris13.fr/terminae/>

Certains environnements de développement ont alors mis en œuvre des approches par assemblage de composants génériques. Ils comportent de riches bibliothèques de composants ainsi que des éditeurs graphiques spécialisés pour formaliser les connaissances du domaine, représenter les tâches à résoudre, sélectionner et configurer les méthodes permettant la résolution de ces tâches (Musen *et al.*, 1994). Les solutions proposées pour adapter des éléments de modèles génériques à une application sont variées, allant de procédures manuelles d'aide à l'instanciation (Beys *et al.*, 1996) à des processus automatisés mettant en œuvre des mécanismes de vérification de la consistance des spécifications (Fensel *et al.*, 1996). Les résultats établis au sein du projet CommonKADS sont parmi les plus aboutis en matière d'aide à la modélisation du raisonnement. Le modèle d'expertise de CommonKADS peut être construit par abstraction de données ou par réutilisation de composants de modèles de raisonnement. Un des points forts de la méthode CommonKADS est d'offrir une bibliothèque de composants de granularités très différentes, et dont la réutilisation et l'adaptation sont guidées par une grille de questions qui assurent que le modèle produit est bien adapté à la tâche à réaliser.

Ce sont les besoins en réutilisation de modèles de connaissances du domaine qui ont conduit à définir la notion d'ontologie. Ces représentations structurées qui définissent les concepts d'un domaine se sont multipliées. Leur réutilisation d'une application à une autre est une des motivations de leur existence. La réutilisation d'ontologie vise à réduire les difficultés de leur développement ex-nihilo, apparues comme un élément de blocage pour certaines applications. Cette réutilisation pose différents types de problèmes : la sélection des ontologies à réutiliser, l'aide à la réutilisation d'ontologies volumineuses et difficilement compréhensibles, et enfin l'intégration de l'ontologie réutilisée au sein de l'ontologie en construction. Nous les abordons successivement ci-dessous.

La réutilisation d'ontologies a motivé la conception de moteurs de recherche d'ontologies : Watson<sup>26</sup>, Swoogle<sup>27</sup>, OntoSearch<sup>28</sup>. A partir de mots clés, ces moteurs retournent une liste d'ontologies contenant un concept, une relation ou un autre élément étiquetés par ces mots-clés. Une recherche doit donc être accompagnée d'un processus de sélection de l'ontologie la plus appropriée, ce qui suppose qu'il soit possible d'évaluer les ontologies en tant que telles ou par comparaison à d'autres, à l'aide de critères. Un des défis actuels porte ainsi sur l'évaluation d'une ontologie et sur la comparaison d'ontologies. Quels critères utiliser pour évaluer une ontologie ? Comment comprendre le point de vue modélisé dans une ontologie ? Comment intégrer une ontologie à une autre ? Dans quelle mesure deux ontologies reflètent-elles la même conceptualisation d'un domaine donné ? Les différences (niveau de détail, compatibilité, concepts centraux, couverture) peuvent-elles être décrites ? Sont-elles artificielles, conséquences de choix techniques ou terminologiques, ou plus profondes reflétant deux conceptualisations différentes ? Des travaux ont porté sur le développement d'algorithmes et d'outils pour identifier et résoudre des différences entre ontologies (analyse des différences entre termes, concepts, définitions). Aujourd'hui peu de travaux comparent globalement les ontologies afin de donner une sorte de vue résumée des points communs et des différences. Une direction intéressante à exploiter consiste à s'appuyer sur la visualisation d'ontologies. En effet, les logiciels de visualisation de grosses ontologies en fournissent des vues globales, et certains se basent sur l'identification des concepts importants de l'ontologie.

Pour réduire les difficultés d'adaptation d'ontologies réutilisées ou de construction d'on-

---

26. <http://kmi-web05.open.ac.uk/WatsonWUI/>

27. <http://swoogle.umbc.edu/>

28. <http://asaha.com/ebook/wNjE3MzI-/OntoSearch--An-Ontology-Search-Engine.pdf>



tologies volumineuses, la notion de patron de connaissances, directement issue des *design patterns* utilisés en génie logiciel, a été introduite dans le domaine de l'ingénierie ontologique par (Clark *et al.*, 2000) puis dans les travaux sur le Web sémantique par (Gangemi *et al.*, 2004), (Rector et Rogers, 2004) et (Svatek, 2004). Les patrons de connaissances sont des représentations récurrentes et partagées de connaissances explicitement représentées comme des modèles généraux, ré-utilisables après transformation (par renommage symbolique) pour créer des représentations spécifiques. Ils peuvent être utiles pour construire des ontologies plus rapidement<sup>29</sup> en aboutissant à un résultat de meilleure qualité, en résolvant, par exemple, des problèmes de conception indépendamment de toute conceptualisation comme l'a proposé le groupe de travail du W3C « Semantic Web Best Practices and Deployment » . Les patrons peuvent aussi faciliter l'application de bonnes pratiques (Pan *et al.*, 2007) ou encore guider la résolution de problèmes de contenu (Gangemi, 2005). Une bibliothèque de patrons de connaissances a été proposée dans le cadre du projet européen NeOn qui distingue les patterns structurels, des patterns de correspondance, de contenu, de raisonnement, de présentation et lexicosyntaxiques (Presutti *et al.*, 2008) ainsi qu'une méthodologie de conception d'ontologies à base de patrons, eXtreme Design (XD) (Daga *et al.*, 2010)<sup>30</sup>.

Réutiliser des modèles de connaissances nécessite par ailleurs de gérer leur intégration au sein du système en cours de construction de façon à ce que le modèle réutilisé puisse communiquer facilement avec les autres modèles.

Paradoxalement, alors que les ontologies devraient faciliter l'interopérabilité entre applications, l'utilisation faite aujourd'hui de ces ressources, issues de concepteurs différents, renvoie à des modélisations différentes. Leur utilisation au sein d'une même application soulève donc des problèmes d'hétérogénéité sémantique. Des termes identiques peuvent dénoter des concepts différents. Des concepts identiques peuvent être nommés différemment, ou encore, un même concept peut être caractérisé par des propriétés différentes. Face à cette hétérogénéité, des travaux sur la « réconciliation de modèles » ont vu le jour. La réconciliation peut s'opérer à deux niveaux différents. Au niveau schéma, elle consiste à identifier des mises en correspondance, alignement ou *mappings*, entre éléments d'ontologies distinctes. Dans les dernières années, beaucoup de travaux de recherche ont porté sur la conception d'outils d'alignement d'ontologies (Euzenat et Shvaiko, 2007). Chaque année, depuis 2004, une compétition internationale (OAEI) est organisée. L'objectif est de confronter les résultats produits par les outils d'alignement appliqués aux mêmes ontologies. Il existe de nombreux outils d'alignement disponibles sur internet comme OnAGUI<sup>31</sup> ou TAXOMAP (Hamdi *et al.*, 2009). Au niveau des données, il s'agit de déterminer si deux descriptions de données correspondent à la même entité du monde réel (même personne, même hôtel, ...). Ce problème est connu sous le nom de *réconciliation de références* (Saïs *et al.*, 2009).

### 20.3.5 Représentation des connaissances dans les modèles

L'ingénierie des connaissances n'a pas pour principal objectif de construire des langages de représentation des connaissances mais puisque les chercheurs spécifient des connaissances ou des modèles, ils ont souvent participé et participent encore à l'élaboration ou l'évolution de langages au sein de groupes de normalisation, comme le W3C. Comme pour la modélisation,

29. On parle d'*Ontology Design Pattern* ou ODP.

30. [http://ontologydesignpatterns.org/wiki/Main\\_Page](http://ontologydesignpatterns.org/wiki/Main_Page)

31. <http://sourceforge.net/projects/onagui/>

les langages de représentation des connaissances ont d'abord été liés aux modèles de raisonnement puis se sont intéressés aux ontologies (cf. 20.2, 20.2.1, 20.2.2) pour revenir maintenant aux raisonnements.

En matière de langages de représentation des ontologies, on peut noter, dans les années 1980, le succès de la logique d'une part, et du langage des graphes conceptuels d'autre part (Sowa, 1984). Les graphes conceptuels proposent à la fois une formalisation logique et une symbolique graphique (intéressante à l'époque où on n'avait pas d'IHM puissantes pour afficher des réseaux sémantiques ou des arbres à déployer et refermer à la demande). OWL s'est ensuite imposé. Ce langage est une évolution de DAML+OIL<sup>32</sup>, fusion des travaux des projets DAML<sup>33</sup> et OIL (Fensel *et al.*, 2001). Inspiré également des logiques de description ??, et défini comme une sur-couche de XML, il s'est stabilisé en 3 langages, OWL Lite, OWL-DL, OWL-full dans le cadre des travaux du W3C. La différence entre chacun de ces langages découle de choix de représentativité *versus* calculabilité.

Si l'on regarde l'empilement des langages du Web sémantique, le cake proposé par Tim B. Lee (cf. figure 2), on note que les langages qui intéressent plus spécifiquement l'ingénierie des connaissances sont ceux munis d'une sémantique formelle qui permettra d'inférer des propositions, donc à partir de *Ontology Vocabulary*. En dessous, en particulier avec RDF, on a un langage qui permet d'exprimer des triplets et donc potentiellement des connaissances mais l'interprétation est à la charge de l'utilisateur. Au-dessus, on a les langages liés à l'expression de règles d'inférences. Leur avantage, par rapport aux travaux antérieurs, est qu'ils font partie de cet ensemble de langages bâtis les uns sur les autres. Cela permet d'envisager plus facilement des applications où les ontologies et les règles associées sont utilisées de façon cohérente. La question du choix, en termes de standards, des langages pour exprimer ces règles n'est pas encore tranchée et il existe un certain nombre de candidats parmi les langages de règles permettant de prendre en compte les ontologies comme SWRL<sup>34</sup>, Description Logic Programs (DLP)<sup>35</sup> (Hitzler *et al.*, 2005), le *Rule Interchange Format* (RIF)<sup>36</sup> ou le langage de requêtes SPARQL<sup>37</sup>.

Il existe d'autres langages que ceux repérés ici. Récemment, le langage SKOS (pour *Simple Knowledge Organisation System*) a été défini pour représenter des modèles peu formels pour lesquels on ne veut justement pas inférer de conséquences logiques précises comme pour les ontologies. SKOS permet de représenter en particulier la relation de généralisation *versus* spécialisation (*broader-than* et *narrower-than* – BT/NT), très utilisée dans les thésaurus sans imposer les inférences logiques liées à la relation de subsomption dans OWL<sup>38</sup>. C'est d'autant plus nécessaire que les inférences logiques sur la relation de subsomption ne sont valides que sur les ontologies, construites pour respecter ces contraintes, et pas sur les thésaurus.

Par ailleurs, les applications utilisant des thésaurus et les ontologies sont arrivées à un état de maturité tel que ces différentes ressources – *i.e.* thésaurus et ontologies – sont elles-mêmes

32. <http://www.w3.org/TR/daml+oil-reference>

33. <http://www.daml.org/>

34. <http://www.w3.org/Submission/SWRL/>

35. <http://logic.aifb.uni-karlsruhe.de/wiki/DLP>

36. <http://www.w3.org/2005/rules/wiki/RIFWorkingGroup>

37. <http://www.w3.org/TR/2008/REC-rdf-sparql-query-20080115/>

38. Les demandes de descriptions des thésaurus et d'alignements de toutes sortes de terminologies sont telles que, après SKOS, une norme ISO (25964, tome 1 et 2) est sortie pour représenter plus finement les liens entre les termes et toutes les modalités d'alignement <<http://dossierdoc.typepad.com/descripteurs/2011/08/norme-iso-25964-1-thesaurus-publication-officielle.html>>.

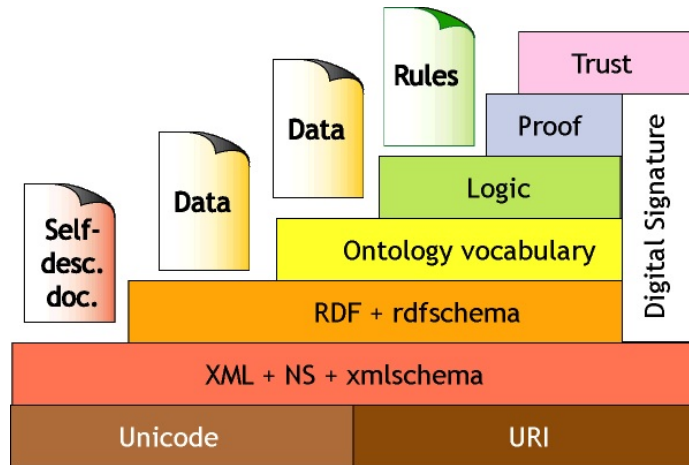


FIGURE 2: Le *cake* du Web sémantique.

impliquées dans des processus de développement qui amènent à utiliser les différents langages de représentation des connaissances à des étapes différentes du processus et pas toujours de la façon prévue par les concepteurs de ces langages. Ainsi un thésaurus et une ontologie utilisés conjointement dans une application vont être modélisés en OWL pour cette application, mais ils seront développés, l'un en SKOS, l'autre en OWL, et distribués, par exemple dans un format comme CTS2<sup>39</sup>.

## 20.4 Enjeux méthodologiques et applicatifs actuels

Les enjeux actuels de l'Ingénierie des connaissances sont à la fois d'ordre méthodologique et applicatif. Pour y répondre, un certain nombre de principes, qui ont fondé la discipline, fournissent un cadre :

- la nécessité d'une démarche pluridisciplinaire prenant en compte les analyses d'autres disciplines telles que la psychologie cognitive, l'ergonomie, les sciences des organisations, la linguistique, le traitement automatique des langues, la recherche d'information, la gestion documentaire ;
- le besoin de penser l'ergonomie des systèmes en amont de la construction du projet, en particulier, prendre en compte les usages des systèmes construits et leur intégration dans le système de traitement de l'information dans lequel ils s'insèrent ;
- l'intérêt d'une démarche de modélisation approfondie, faisant cohabiter des modèles différents à chaque temps du processus d'élaboration du système opérationnel.

Les applications liées à l'ingénierie des connaissances constituent alors un vaste champ de recherche, d'expérimentation et de valorisation dans lequel doivent se développer des méthodologies innovantes. C'est cette articulation entre méthodologie et applications qui sera le fil conducteur des enjeux décrits ci-dessous.

39. <http://www.3mtcs.com/resources/hl7cts>

### **20.4.1 Articuler la langue, les connaissances et leur support**

La langue naturelle est le vecteur privilégié de l'expression des connaissances. Il est donc logique que l'Ingénierie des connaissances se préoccupe de connaissances exprimées en langue naturelle (expression orale ou écrite) et qu'elle entretienne, de ce point de vue, des liens étroits avec les domaines de la linguistique, du TALN d'un côté et de la Recherche d'Information (R.I.) de l'autre. Des problématiques méthodologiques interdisciplinaires apparaissent alors en concordance avec des classes d'applicatifs que nous allons décrire ici.

#### **Construction de modèles de raisonnement et d'ontologies**

L'Ingénierie des connaissances s'intéresse aux documents comme porteurs de sens et révélateurs de connaissances depuis les premières études sur l'acquisition des connaissances pour les systèmes experts (années 90). Il s'agissait déjà de repérer des connaissances heuristiques, de rendre compte de raisonnements explicités plus ou moins dans des documents existants ou élaborés pour l'étude (retranscriptions d'entretiens). Les ontologies ont été définies comme des représentations d'un domaine, accentuant la dissociation (parfois provisoire) entre le raisonnement heuristique et la description des concepts manipulés par ces heuristiques. Les ontologies se focalisent donc sur l'essence d'un domaine (comme la médecine, ou un champ de la médecine par exemple), sur son vocabulaire et, au-delà, sur le sens dont il est porteur. Par la suite, sous l'impulsion du groupe TIA dans les années 90, on a pu mettre en avant l'utilité de corpus textuels générés durant une activité pour aider à construire l'ontologie de cette même activité, de ce domaine. Dans cette perspective, les documents d'un corpus sont considérés comme une source de connaissances complémentaire ou alternative aux experts et spécialistes du domaine. Le traitement de ces sources de connaissances passe d'abord par des outils de TALN mais aussi par des plates-formes aptes à utiliser le résultat de ces outils pour construire les ontologies ou plus largement des terminologies (*cf.* 20.3.3).

#### **Textes et documents en ingénierie des connaissances**

Dans cette perspective, le document en tant que tel est central, il est un élément support de connaissance à part entière. La gestion des documents produits et utilisés au sein de l'activité individuelle et collective étudiée, mais aussi, en tant que telle, la gestion de fonds documentaires (images, sons, vidéos) intéresse alors l'ingénierie des connaissances. Ces applications font appel aux technologies relevant de la gestion documentaire et permettant le partage, la diffusion, l'archivage, l'indexation, la structuration ou la classification de documents ou de flux de documents. Ces technologies sont propres à la nature des supports et la difficulté est d'appliquer ces solutions technologiques aux bons documents de manière à répondre au mieux aux besoins des utilisateurs, et qu'ils y trouvent les supports (entre autres les connaissances) utiles à la réalisation de leurs tâches.

Parce que de plus en plus de projets d'Ingénierie des connaissances intègrent la gestion de documents sous des formes très variées, les chercheurs du domaine ne peuvent s'affranchir d'une réflexion approfondie sur la notion de document, et particulièrement de document numérique. Ainsi, plusieurs chercheurs contribuent aux travaux du réseau thématique pluridisciplinaire sur le document (RTP-DOC) et à ses productions (Pédauque, 2003, 2005).

## Recherche d'information avec des ontologies

Sous l'impulsion du Web sémantique qui, dans ses premiers attendus, utilisait les ontologies comme source de métadonnées pour indexer des documents, les ontologies sont arrivées au cœur d'applications de R.I. Elles y facilitent l'accès aux ressources pertinentes en permettant l'intégration de sources distribuées et hétérogènes tant au niveau schéma que données, l'interrogation de sources multiples à l'aide d'un vocabulaire unifié, l'enrichissement des requêtes posées, le filtrage et le classement des résultats retournés. Ces travaux ont amené évidemment à des rapprochements avec les thésaurus ou les terminologies pour mieux marquer les différences et les apports des uns et des autres. Ces rapprochements amènent des réflexions sur des ontologies utilisées spécifiquement pour la R.I. et ayant une composante linguistique forte, a minima des termes associés aux concepts. Une dernière problématique de ce champ est la mise en œuvre d'environnements applicatifs où l'on fait cohabiter les ontologies avec les thésaurus (Vandenbussche et Charlet, 2009).

### 20.4.2 Faire face à l'explosion des données

Depuis près de 20 ans, nous assistons à une explosion des données disponibles. Les applications traitent des données de plus en plus nombreuses et diverses suscitant des besoins nouveaux pour les décrire et les intégrer.

La description de ces données très nombreuses passe par l'élaboration de modèles au sein desquels la quantité d'information à prendre en compte peut être très importante, suffisamment pour ouvrir de nouvelles perspectives aux approches et modèles statistiques. Si l'on veut maintenir l'usage et la gestion de modèles symboliques, l'enjeu est alors d'être capable de construire des modèles de très grande taille, en diminuant par exemple la quantité d'informations à prendre en compte simultanément. C'est ainsi que des travaux sur la modularité ont vu le jour, qui, appliqués à l'ingénierie des ontologies, visent la construction d'ontologies de très grande taille nécessaires aux applications. Citons, à titre d'exemple, les travaux sur la construction d'ontologies modulaires (Stuckenschmidt *et al.*, 2009) réalisés dans le cadre du projet Knowledge Web<sup>40</sup> (2004-2007). La modularité, au sens général du terme, correspond au fait de percevoir un grand entrepôt de connaissances (que ce soit une ontologie, une base de connaissances ou de données) comme un ensemble de modules, c'est-à-dire des entrepôts plus petits pouvant être considérés comme des parties d'un tout. Bien que le concept de modularité soit très utilisé en informatique, c'est une idée relativement nouvelle en Ingénierie des connaissances. Le projet Knowledge Web a montré la diversité de points de vue sur ce problème et pointé les directions de recherche importantes à développer : la construction de modules (déterminer l'ensemble cohérent de concepts, relations, axiomes et instances composant un module pour qu'il fasse sens), leur description, la façon dont ils peuvent être liés les uns aux autres, leur composition et leur réutilisation dans différents contextes.

Gérer une grande masse de données dans un contexte distribué peut aussi amener à prendre appui sur un ensemble d'ontologies existantes nécessitant d'être remaniées, alignées, transformées sous forme de modules ou intégrées avec des ressources non ontologiques telles que des bases de données, des *folksonomies* ou des thesaurii. La méthode de construction d'ontologie en réseau définie par le projet NeOn<sup>41</sup> (2006-2010) que nous avons déjà présenté, inclut un

40. [http://cordis.europa.eu/ist/kct/knowledgeweb\\_synopsis.htm](http://cordis.europa.eu/ist/kct/knowledgeweb_synopsis.htm)

41. <http://www.neon-project.org/>

support au développement collaboratif et la prise en compte de l'aspect dynamique et évolutif de ces ontologies (Gómez-Pérez et Suárez-Figueroa, 2009). Il s'agit d'un enjeu important pour le développement de grosses applications à base d'ontologies.

### 20.4.3 Gérer l'intégration des connaissances par les ontologies

Une approche de plus en plus privilégiée, autant dans le domaine des bases de données que dans celui de la recherche d'information, est d'appuyer l'intégration des données sur une ontologie du domaine concerné. Lors de l'intégration de sources multiples et hétérogènes, les ontologies peuvent en effet aider à comprendre et interpréter des descriptions hétérogènes de contenus relatifs à un même domaine, puis à les mettre en relation plus facilement (Assele Kama *et al.*, 2010).

Dans certains domaines, comme dans le domaine géographique, peu d'ontologies ont été réalisées en pratique (Laurens, 2006) ou celles-ci décrivent des domaines ciblés, comme Townology dans le domaine de l'aménagement et de l'urbanisme (Roussey *et al.*, 2004) ou FoDo-MuSt dans le domaine du traitement d'images (Brisson *et al.*, 2007). L'enjeu consiste alors à construire les ontologies utiles.

Dans d'autres domaines, des ontologies existent mais sont de très grande taille et donc difficiles à exploiter. C'est le cas, par exemple, des domaines de l'agronomie ou de la médecine pour lesquels existent des ontologies de plusieurs milliers de concepts. Dans ce cas, l'enjeu consiste à permettre de comprendre le contenu de ces ontologies pour aider à en extraire le sous-ensemble pertinent pour une application. Dans le domaine médical, où il existe des classifications de plusieurs dizaines de milliers de concepts et une ontologie de plusieurs centaines de milliers de concepts, le problème atteint un niveau de complexité supplémentaire. En suivant ces travaux (Rosenbloom *et al.*, 2006), on parle d'*ontologies de référence* et d'*ontologies d'interface* : 1) en médecine, les ontologies de référence ont des visées de représentations larges et de référentiel pour des futures études épidémiologiques. La plus connue est la SNOMED-CT ; 2) les ontologies d'interface sont, elles, développées pour des applications spécifiques. Entre les deux types d'ontologies, on a besoin de services d'alignements et, comme précédemment, de la possibilité d'extraire le sous-ensemble de l'ontologie de référence pertinent pour une application. C'est ce que permet un standard comme CTS2 (*cf.* section 20.3.5).

Ce contexte, renforcé par le besoin d'exploiter des connaissances diversifiées ou plusieurs modèles (ou modules) de connaissances partiels, nécessite de faire face au problème d'hétérogénéité entre modèles/ontologies/connaissances, d'où l'intérêt que suscitent actuellement les travaux portant sur l'interopérabilité sémantique. Leur objectif est d'établir des mises en correspondance de la façon la plus automatique possible entre des éléments de schémas, d'ontologies ou de sources sémantiquement hétérogènes (en utilisant des techniques d'appariement de schémas, d'alignement d'ontologies (*cf.* section 20.3.4), ou de réconciliation de données).

### 20.4.4 Gérer les données distribuées

Le Web et les standards du web ont changé la donne en matière de données distribuées. On dispose en particulier maintenant d'un protocole de communication entre machines, les services web qui ont transformé le Web en un dispositif distribué de calcul où les programmes (services) peuvent interagir de manière intelligente en étant capables de se découvrir automatiquement, de négocier entre eux et de se composer en des services plus complexes. De

nombreuses connaissances sont indispensables pour l'obtention de services intelligibles pour des machines. C'est le concept de *service web sémantique* : il s'agit de créer un Web sémantique de services dont les propriétés, les capacités, les interfaces et les effets sont décrits de manière non ambiguë et exploitables par des machines. La sémantique ainsi exprimée doit faciliter l'exploitation automatique des services. Les services Web sémantiques apparaissent indispensables pour une utilisation effective des services web dans des applications industrielles. Ils posent aujourd'hui toutefois un certain nombre de problèmes qui interpellent entre autres la communauté des chercheurs en Ingénierie des connaissances. Par exemple, ils font appel aux ontologies pour expliciter la sémantique des services facilitant ainsi les communications hommes-machines, d'une part, et machines-machines, d'autre part. La modélisation d'aspects sémantiques peut également permettre d'identifier des critères de qualité d'un service web à retenir dans le processus de découverte ou de composition de services.

Les systèmes pair à pair P2P se sont également beaucoup développés et de nombreux travaux récents ont visé à améliorer la fonction de recherche dans les systèmes non structurés en remplaçant le routage aléatoire par un routage guidé par la sémantique. Plusieurs dimensions du problème sont analysées : Quelle sémantique retenir ? Quelle représentation adopter ? Comment la construire ? Qu'est ce qui est partagé entre les pairs ? Comment utiliser la sémantique ? comment la diffuser ? Ces problèmes restent aujourd'hui des problèmes ouverts et l'Ingénierie des connaissances peut contribuer à apporter des solutions.

#### 20.4.5 Tirer parti des nouvelles sources de connaissances

Nous discuterons deux sources qui constituent aujourd'hui des enjeux majeurs : le Web 2.0 et le Web des données liées.

Le Web 2.0 ou Web social (O'Reilly, 2007) accorde une place plus importante aux utilisateurs que le Web dans sa version initiale en leur permettant de devenir actif. A la fois auteurs et acteurs, les internautes peuvent utiliser le web pour élaborer leurs propres contenus et les partager. Ceci est rendu possible par des outils tels que les blogs, les réseaux sociaux, les sites collaboratifs, les plates-formes de mise en relation, les services de partage en ligne. Ces outils et ces services sont de plus en plus utilisés dans les organisations. Mais les applications qui gèrent ces contenus ont leur propre format de données et celles-ci sont de plus en plus distribuées et hétérogènes, ce qui pose des problèmes importants d'intégration d'information, d'identification de la fiabilité et des auteurs des sources, de la trace de leur historique, ... De la même façon, l'*étiquetage*<sup>42</sup>, une pratique souvent utilisée pour regrouper des contenus jugés similaires et en faciliter la recherche, présente des limites du fait de l'ambiguïté et de l'hétérogénéité des étiquettes, couramment appelées *tags*. Ainsi, les applications Entreprise 2.0 (McAfee, 2006), qui tendent à se développer, constituent un terrain d'expérimentation et de valorisation des techniques de l'Ingénierie des connaissances, tout en permettant au domaine de se renouveler en faisant des propositions nouvelles pour faciliter la navigation, l'interrogation ou l'extraction. Il s'agit, pour l'Ingénierie des connaissances, d'un enjeu applicatif d'une importance majeure.

Egalement proposé par Tim Berners-Lee, le Web des données liées<sup>43</sup> réfère à un style de publication et d'interconnexion des données structurées sur le Web fondé sur le modèle RDF.

42. On décrit ici le fait d'indexer un contenu avec des méta-données choisies par l'utilisateur. Les ensembles d'étiquettes forment alors des *folksonomies*.

43. Tim Berners-Lee parle de Web of Data. Il promeut un projet qui va dans ce sens, le Linking Open Data (LOD).

Les données liées ont l'avantage de fournir un mécanisme simple d'accès unique et normalisé au lieu de s'appuyer sur différents formats d'interface et de résultat. Les sources de données peuvent ainsi être plus facilement explorées par les moteurs de recherche, elles peuvent être accessibles à l'aide de navigateurs génériques de données, elles peuvent avoir des liens avec des sources de données différentes. Le nombre de données publiées selon les principes des données liées croît rapidement (on parle de milliards de triplets RDF disponibles sur Internet). Du fait de l'utilisation d'un vocabulaire non ambigu et relié, cette masse de données représente une source prometteuse que l'Ingénierie des connaissances doit absolument considérer.

#### **20.4.6 Faire face à l'évolution des connaissances**

Les modèles de représentation des connaissances existants sont inadaptés pour représenter les données liées à l'évolution et surtout, ils ne bénéficient d'aucun mécanisme d'adaptation qui leur permettrait de suivre de manière rigoureuse les évolutions d'un domaine. Les travaux récents sur l'évolution d'ontologies (Noy et Klein, 2004; Stojanovic, 2004) montrent l'intérêt des communautés du Web sémantique et de l'Ingénierie des connaissances pour trouver des solutions adaptées aux problèmes qui en découlent comme la spécification des changements (Plessers *et al.*, 2007), la vérification de la cohérence de l'ontologie modifiée (Haase et Stojanovic, 2005) et la propagation des changements dans le cadre d'ontologies distribuées (Stuckenschmidt et Klein, 2003). La caractérisation et la représentation de l'évolution d'un domaine demeurent toutefois des problèmes ouverts même si quelques propositions commencent à émerger (Guelfi *et al.*, 2010).

#### **20.4.7 Dimension collective versus personnelle des connaissances**

La plupart des approches précédentes accordent peu de place à la dimension sociale de la gestion des connaissances. Or cette dimension est suffisamment forte pour considérer certaines communautés professionnelles comme des communautés d'intérêt ou de pratique.

Les communautés de pratique désignent des processus d'apprentissage qui émergent par le partage de connaissances échangées en réseau. Les modèles de l'Ingénierie des connaissances doivent capturer ces processus d'apprentissage ou les intégrer dans leur processus de gestion des connaissances. Ainsi dans (Lewkowicz et Zacklad, 2001), une nouvelle forme de gestion de connaissances fondée sur la structuration des interactions collectives est proposée. L'objectif est de permettre une meilleure mise en valeur des savoirs échangés de façon à faciliter leur réutilisation, le savoir d'une organisation étant considéré dans cette approche comme relevant avant tout d'une compétence collective.

L'identification de communautés d'intérêt du fait du développement du Web 2.0 ou de l'analyse des traces des utilisateurs recherchant des informations thématiques proches suppose la représentation de connaissances relatives aux centres d'intérêts et aux activités de leurs membres, introduisant une sorte de dimension collective des connaissances prise en compte par certains travaux en l'Ingénierie des connaissances. Cette dimension collective fait référence aux travaux de la communauté du *Computer Supported Cooperative Work (CSCW)*. Ainsi, M. Zacklad propose, entre les thésaurus et les ontologies formelles, des *ontologies sémiotiques* mieux à même d'être partagées par une communauté de travail dans un cadre de recherche d'informations (Zacklad, 2007). A l'opposé, nous assistons aujourd'hui à la conception de logiciels, des interfaces Web par exemple, sensibles au contexte d'interaction (profil utilisateur,



environnement, plateforme d'interaction, ...). L'acquisition, la modélisation et le traitement de ces contextes d'interaction jouent un rôle fondamental pour le développement de tels systèmes adaptatifs et personnalisés (Garlatti et Prié, 2004).

#### 20.4.8 Évaluer la qualité des modèles

Enfin, une question fondamentale pour l'Ingénierie des connaissances concerne l'évaluation de la qualité des modèles utilisés et des résultats produits. L'exploitation de connaissances de qualité médiocre conduit à des erreurs, des doublons, des incohérences qu'il faut éviter. Au delà de son intérêt dans le domaine de la recherche, le thème de la qualité est devenu critique avec le déploiement d'applications dans les entreprises.

La qualité des modèles/ontologies peut être assurée de manière méthodologique, lorsque l'ontologie a été construite selon une méthode rigoureuse s'appuyant sur les fondements théoriques et philosophiques de ce qu'est une ontologie (Cf. les méthodes présentées en partie 20.5).

D'autres travaux méthodologiques visent à passer de démarches « artisanales », essentiellement manuelles, dont le coût et la durée sont difficiles à estimer, à des approches plus systématiques, outillées et mieux maîtrisées. Ils mettent bien sûr l'accent sur la réutilisation comme Methontology (Gómez-Pérez *et al.*, 2007) puis NEON dans (Suárez-Figueroa *et al.*, 2012), sur des guides pratiques (Noy et Hafner, 1997) ou sur l'analyse systématique de textes à l'aide de logiciels de TALN comme Terminae (Aussenac-Gilles *et al.*, 2000) et les méthodes répertoriées dans (Maedche, 2002) ou DaFOE4App (Charlet *et al.*, 2010).

Dans (Brank *et al.*, 2005), un état de l'art classe les différentes techniques d'évaluation d'ontologie en quatre catégories : (1) celles effectuant une comparaison syntaxique par rapport à un langage de référence (RDF, OWL, ...), par exemple (Maedche et Staab, 2002), (2) celles consistant à utiliser l'ontologie dans un contexte applicatif, par exemple (Porzel et Malaka, 2004) (3) celles consistant à effectuer une comparaison avec une source de données du domaine (une collection de documents) devant couvrir l'ontologie, comme (Brewster *et al.*, 2004) (4) les validations manuelles par rapport à un ensemble de critères prédéfinis, de standards ou des besoins, par exemple (Lozano-Tello et Gomez-Perez, 2004). De plus, dans (Brank *et al.*, 2005), les approches de validation sont organisées en six niveaux d'évaluation : le niveau lexical, le niveau des relations taxonomiques, des autres relations sémantiques, le niveau application (impact de l'ontologie sur les résultats de l'application qui l'utilise) ou contexte (prise en compte du fait qu'une ontologie peut être une partie d'une autre ou peut référencer ou être référencée par d'autres), le niveau syntaxique ou encore le niveau des principes de conception. Il peut en effet être plus facile d'évaluer une ontologie par niveau du fait de sa complexité.

### 20.5 Conclusion

L'ingénierie des connaissances a vécu des changements d'orientation successifs. C'est un domaine en constante évolution de l'intérieur (nouvelles analyses, nouvelles perspectives, manière originale de poser les problèmes, nouveaux concepts théoriques) et de l'extérieur (les types d'applications ciblés mais aussi les données disponibles, en particulier avec les bouleversements du web, ont changé au fil des années, les contributions d'autres disciplines viennent apporter des méthodes et des concepts nouveaux). Ces évolutions élargissent progressivement

le champ de l'Ingénierie des connaissances. Les nouveaux cadres théoriques proposés englobent les travaux précédents. Même si certains changements de point de vue correspondent à des ruptures, les résultats du domaine se complètent au fil du temps et peuvent être repris sous un nouvel angle.

Pendant longtemps, l'Ingénierie des connaissances s'est intéressée à la production de modèles de connaissances selon un processus bien structuré sous le contrôle d'ingénieurs de la connaissance. Les modèles élaborés, généralement complexes, étaient utilisés dans des applications spécifiques. Aujourd'hui, les applications dans lesquelles des connaissances sont utilisées comme support à un raisonnement ou une activité se sont beaucoup diversifiées ces dernières années. Depuis 2000, elles concernent la gestion des connaissances au sens large, la recherche d'information sémantique, l'aide à la navigation, l'aide à la décision, beaucoup d'applications relevant également du domaine du Web sémantique. Cet élargissement se poursuit et de nouveaux champs d'application émergent encore, posant les problèmes de l'Ingénierie des connaissances en des termes nouveaux.

Ainsi, à l'ère de l'informatique ubiquitaire, ce sont le salon, le train, l'automobile, l'atelier, la salle de classe ou de réunion, le moindre instrument de cuisine qui deviennent des outils « intelligents ». L'exploitation de connaissances au sein de ces outils nécessite un processus *dynamique* d'acquisition des connaissances de contexte de façon à les *acquérir en continu* auprès d'une grande variété de sources (des capteurs, des bases de données, internet, mais aussi des types d'utilisateurs variés). Par ailleurs, ces outils intelligents doivent être dotés d'un *comportement pro-actif* leur permettant d'initier une communication ou une action sur la base de leur compréhension de la situation courante et de leurs objectifs. Ainsi, par exemple, les téléphones savent localiser où nous sommes à un moment donné et deviennent capables d'automatiser certaines opérations comme, lors de la prise de photos, leur étiquetage par des métadonnées.

La dernière décennie a connu une transformation majeure dans la façon dont les individus interagissent et échangent. L'information est dorénavant coproduite, partagée, classée et évaluée sur le Web par des milliers de personnes. Ces usages et les technologies sous-jacentes sont connus sous le nom de Web 2.0. L'ingénierie et la gestion des connaissances dans le contexte de communautés d'intérêt ou de pratiques dont l'émergence spontanée et l'activité sont permises par ces évolutions du Web (Web 2.0, Web social) sont des enjeux majeurs de la future décennie.

Enfin, à l'ère du Web sémantique, un nouveau paradigme prometteur apparaît, celui du Web des données. Le Web des données, qui fait suite au Web des documents, entend faire face au déluge informationnel en connectant les données. La finesse des analyses, la qualité des modèles et la maîtrise de leur contenu sont abandonnées au profit d'analyses automatiques, de la production automatique, systématique et volumineuse de représentations à partir des documents mais surtout des traces des usages du web. Ces données se prêtent bien à un traitement massif, quantitatif mais leur pertinence sémantique est encore à évaluer. Pour favoriser l'émergence de ce Web des données, un projet du W3C (Linked Data) s'est donné pour objectif de construire un réseau global des données en liant des données provenant d'horizons divers, permettant ainsi, à partir d'une donnée, d'obtenir l'ensemble des données qui y sont liées. Une première hypothèse à valider est de vérifier que l'interconnexion des données contribue réellement à leur donner du sens. A partir de là, il faudra évaluer dans quelle mesure ces données constituent une source importante d'innovation : on attend d'elles qu'elles augmentent la pertinence des contenus et aboutissent à la constitution d'une base de données à l'échelle du web..

De nombreuses sources de données existent déjà, parmi lesquelles trône DBpedia qui structure le contenu de Wikipedia en triplets RDF de façon à rendre les informations de l'encyclopédie réutilisables. DPpedia est une source très puissante car elle est interconnectée avec d'autres sources de données, Geonames, MusicBrainz, etc..

L'Ingénierie des connaissances se doit d'alimenter et d'évaluer tous ces développements nouveaux, de les mettre en regard avec les modèles précédents (modèles de raisonnement, bases de règles), d'estimer la nécessité ou non d'avoir recours à des ontologies et à leur alignement pour typer ou organiser les données, de définir si besoin de nouvelles techniques et langages, de justifier le recours à des métadonnées pour enrichir ces données et mieux les réutiliser, etc. La rapidité des évolutions de web peut être vue comme un accélérateur fou du rythme des recherches ou comme une alarme qui invite à prendre du recul et à poser les problématiques à un meilleur niveau d'abstraction, toujours nécessairement interdisciplinaire, qui se détache des technologies mises en œuvre pour mieux qualifier l'essence des connaissances, de leur diffusion et de leur formalisation en vue de traitements informatiques.

## Références

- ASSELE KAMA, A., MELS, G., CHOQUET, R., CHARLET, J. et JAULENT, M.-C. (2010). Une approche ontologique pour l'exploitation de données cliniques. In DESPRES, S., éditeur : *Acte des 21èmes Journées Francophones d'Ingénierie des Connaissances*, pages 183–194, Nîmes, France. Ecole des Mines d'Alès.
- AUSSENAC, N. (1989). *Conception d'une méthoologie et d'un outil d'acquisition de connaissances expertes*. Thèse de doctorat, Université Paul Sabatier, Toulouse, France.
- AUSSENAC-GILLES, N., BIÉBOW, B. et SZULMAN, S. (2000). Revisiting Ontology Design : a method based on corpus analysis. In *12th International Conference on Knowledge Engineering and Knowledge Management, Juans-Les-Pins (F), 03-06/10/2000*, pages 172–188, Heidelberg (G). Springer Verlag.
- AUSSENAC-GILLES, N., BOURIGAULT, D. et CONDAMINES, A. (1995). How can knowledge acquisition benefit from terminology ? In *Proceedings of the 9th Knowledge Acquisition Workshop*, Banff, University of Calgary (CA).
- AUSSENAC-GILLES, N. et JACQUES, M.-P. (2008). Designing and evaluating patterns for relation acquisition from texts with caméléon. *Terminology, special issue on Pattern-Based approaches to Semantic Relations*, 14(1):45 – 73.
- AUSSENAC-GILLES, N., KRIVINE, J. et SALLANTIN, J. (1992). Editorial du numéro spécial Acquisition des connaissances. *Revue d'Intelligence Artificielle*, 6(2):7–18.
- AUSSENAC-GILLES, N., LAUBLET, P. et REYNAUD, C., éditeurs (1996). *Acquisition et ingénierie des connaissances : tendances actuelles*. Cepadues Editions, Toulouse.
- BACHIMONT, B., ISAAC, A. et TRONCY, R. (2002). Semantic Commitment for Designing Ontologies : A Proposal. In GOMEZ-PÉREZ, A. et BENJAMINS, V., éditeurs : *13<sup>th</sup> International Conference on Knowledge Engineering and Knowledge Management (EKAW'02)*, volume 2473 de *Lecture Notes in Artificial Intelligence*, pages 114–121, Sigüenza, Espagne. Springer Verlag.
- BARRIERE, C. et AGBAGO, A. (2006). Terminoweb : a software environment for term study in rich contexts. In *International Conference on Terminology, Standardization and technology transfer*, pages 103–113, Beijing.
- BEYS, B., BENJAMINS, V. et VAN HEIJST, G. (1996). Remedying the Reusability - Usability trade-off for Problem-Solving Methods. In GAINES, B. et MUSEN, M., éditeurs : *Proceedings of the 10th Knowledge Acquisition Workshop (KAW)*, pages 2–1/2–20, Banff, Canada.
- BOURIGAULT, D. (2002). Upery : un outil d'analyse distributionnelle étendue pour la construction d'ontologies à partir de corpus. In *Actes de la 9ème conférence annuelle sur le Traitement Automatique des Langues (TALN 2002)*, pages 75 – 84, Nancy, France.
- BOURIGAULT, D., AUSSENAC-GILLES, N. et CHARLET, J. (2004). Construction de ressources terminologiques ou ontologiques à partir de textes : un cadre unificateur pour trois études de cas . *Revue d'Intelligence Artificielle*, 18(1/2004):87–110.
- BOURIGAULT, D. et SLODZIAN, M. (1999). Pour une terminologie textuelle . *Terminologies Nouvelles*, 19:29–32.
- BRANK, J., GROBELNIK, M. et MLADENIC, D. (2005). A survey of ontology evaluation techniques. In *Data Mining and Data Warehouses Conference (SIKDD)*, Lubiana, Slovénie.

- BREWSTER, C., ALANI, H., DASMAHAPATRA, S. et WILKS, Y. (2004). Data driven ontology evaluation. *In LREC*.
- BRISSON, R., BOUSSAID, O., GANÇARSKI, P., PUISSANT, A. et DURAND, N. (2007). Navigation et appariement d'objets géographiques dans une ontologie. *In EGC*, pages 391–396.
- CAMILLERI, G., SOUBIE, J.-L. et ZARATÉ, P. (2008). A replanning support for critical decision making situations : A modelling approach. *In Intelligent Decision Making : An AI-Based Approach*, pages 173–192.
- CHANDRASEKARAN, B. (1983). Towards a taxonomy of problem solving types. *AI Magazine*, 4(1):9–17.
- CHARLET, J. (1992). Acte : Acquisition des connaissances par interprétation d'un modèle causal. *Revue d'Intelligence Artificielle*, 6(1–2):99–129.
- CHARLET, J., SZULMAN, S., AUSSENAC-GILLES, N., NAZARENKO, A., HERNANDEZ, N., NADAH, N., SARDET, E., DELAHOUSSE, J., TEGUIAK, V. et BANEYX, A. (2010). Dafoe : une plateforme pour construire des ontologies à partir de textes et de thésaurus. *In EGC 2010*, pages 631–632. Paris : Hermès.
- CHARLET, J., ZACKLAD, M., KASSEL, G. et BOURIGAULT, D., éditeurs (2000). *Ingénierie des connaissances : Evolutions récentes et nouveaux défis*. Eyrolles, Paris.
- CIMIANO, P., BUITELAAR, P. et VÖLKER, J. (2010). Ontology construction. *In INDURKHYA, N. et DAMERAU, F. J., éditeurs : Handbook of Natural Language Processing, Second Edition*. CRC Press, Taylor and Francis Group, Boca Raton, FL. ISBN 978-1420085921.
- CIMIANO, P. et VÖLKER, J. (2005). Text2onto. *In NLDB*, pages 227–238.
- CLARK, P., THOMPSON, J. A. et PORTER, B. W. (2000). Knowledge patterns. *In KR*, pages 591–600.
- CONDAMINES, A. (2003). *Sémantique et corpus spécialisés : Constitution de bases de connaissances terminologiques*. Habilitation à diriger des recherches, Université de Toulouse 2, Toulouse, France.
- CONSTANT, M., DISTER, A., ERMIKIANIAN, L. et PIRON, S. (2008). Description linguistique pour le traitement automatique du français. *Cahier du CENTAL*.
- CORDIER, M.-O. et REYNAUD, C. (1991). Knowledge acquisition techniques and second-generation expert systems. *Applied Artificial Intelligence*, 5(3):209–226.
- DAGA, E., BLOMQVIST, E., GANGEMI, A., MONTIEL, E., NIKITINA, N., PRESUTTI, V. et VILLAZON-TERRAZAS, B. (2010). NeOn project : NeOn D2.5.2. Pattern-based ontology design : methodology and software report. Rapport de contrat.
- DARSES, F. et MONTMOLLIN, M., éditeurs (2006). *L'ergonomie*. La Découverte - Col. Repères, Paris.
- DIENG-KUNTZ, R., CORBY, O., GANDON, F., GIBOUIN, A., GOLEBIOWSKA, J., N., M. et RIBIÈRE, M., éditeurs (2005). *Knowledge management : Méthodes et outils pour la gestion des connaissances*. Dunod.
- DROUIN, P. (2003). Term extraction using non-technical corpora as a point of leverage. *Terminology*, 9:99–117.
- EUZENAT, J. et SHVAIKO, P. (2007). *Ontology matching*. Springer.
- FANKAM, C., BELLATRECHE, L., HONDJACK, D., AMEUR, Y. A. et PIERRA, G. (2009). Sistro,

- conception de bases de données à partir d'ontologies de domaine. *Technique et Science Informatiques*, 28(10):1233–1261.
- FENSEL, D., SCHÖNEGGE, R. et WIELINGA, B. (1996). Specification and verification of knowledge-based systems. *In Proceedings of the 10th Knowledge Acquisition Workshop (KAW)*, Banff (Can). University of Calgary (Can).
- FENSEL, D., van HARMELEN, F., HORROCKS, I., MCGUINNESS, D. L. et PATEL-SCHNEIDER, P. F. (2001). Oil : An ontology infrastructure for the semantic web. *IEEE Intelligent Systems*, 16(2):38–45.
- FERNÁNDEZ-LÓPEZ, M. et GÓMEZ-PÉREZ, A. (2002). Overview and analysis of methodologies for building ontologies. *Knowledge Eng. Review*, 17(2):129–156.
- GANGEMI, A. (2005). Ontology design patterns for semantic web content. *In International Semantic Web Conference*, pages 262–276.
- GANGEMI, A., CATANACCI, C. et BATTAGLIA, M. (2004). Inflammation ontology design pattern : an exercise in building a core biomedical ontology with descriptions and situations. *In MARIA, P. D., éditeur : Ontologies in medicine*. IOS Press, Amsterdam.
- GARLATTI, S. et PRIÉ, Y. (2004). Adaptation et personnalisation dans le web sémantique. *Revue I3 - Numéro hors série Web Sémantique*.
- GÓMEZ-PÉREZ, A., FERNÁNDEZ-LÓPEZ, M. et CORCHO, O. (2007). *Ontological Engineering : with examples from the areas of Knowledge Management, e-Commerce and the Semantic Web. (Advanced Information and Knowledge Processing)*. Springer-Verlag New York, Inc., Secaucus, NJ, USA.
- GÓMEZ-PÉREZ, A. et SUÁREZ-FIGUEROA, M. (2009). Scenarios for building ontology networks within the neon methodology. *In K-CAP*, pages 183–184.
- GRUBER, T. R. (1993). A translation approach to portable ontology specifications. *Knowledge Acquisition*, 5:199–220.
- GUARINO, N. et WELTY, C. A. (2004). An overview of ontoclean. *In Handbook on Ontologies*, pages 151–172.
- GUELFI, N., PRUSKI, C. et REYNAUD, C. (2010). Experimental assessment of the target adaptive ontology-based web search framework. *In NOTERE*, pages 297–302.
- HAASE, P. et STOJANOVIC, L. (2005). Consistent evolution of owl ontologies. *In ESWC*, pages 182–197.
- HAMDI, F., SAFAR, B., NIRLAULA, N. et REYNAUD, C. (2009). TaxoMap in the OAEI 2009 alignment contest. *In The Fourth International Workshop on Ontology Matching*, Chantilly, Washington DC., États-Unis.
- HITZLER, P., SURE, Y. et STUDER, R. (2005). Description logic programs : A practical choice for the modelling of ontologies. *In Principles and Practices of Semantic Web Reasoning*.
- JACOB-DELOUIS, I. et KRIVINE, J. (1995). Lisa : un langage réflexif pour opérationnaliser les modèles d'expertise. *revue d'Intelligence Artificielle*, 9(1):53–88.
- KAMEL, M. et AUSSÉNAC-GILLES, N. (2009). Utiliser la Structure du Document dans le Processus de Construction d' Ontologies (regular paper). *In L' HOMME, M.-C. et SZULMAN, S., éditeurs : Conférence Internationale sur la Terminologie et l'Intelligence Artificielle (TIA), Toulouse (France), 18-20/11/2009*, page (on line), <http://www.irit.fr/>. IRIT.
- KASSEL, G. (2002). Ontospec : une méthode de spécification semi-informelle d'ontologies.

- In Actes d'IC*, pages 75–87.
- KLINKER, G., BHOLA, G., DALLEMAGNE, G., MARQUÈS, D. et DERMOTT, M. (1991). Usable and reusable programming constructs. *Knowledge Acquisition*, 3:117–136.
- LAFLAQUIÈRE, J., PRIÉ, Y. et MILLE, A. (2008). Ingénierie des traces numériques d'interaction comme inscriptions de connaissances. *In Actes d'IC*, pages 183–195.
- LAURENS, F. (2006). Création d'une ontologie à partir de textes en langage naturel. Rapport de master 1 linguistique - informatique, Paris 7.
- LEWKOWICZ, M. et ZACKLAD, M. (2001). Une nouvelle forme d'gestion des connaissances basée sur la structuration des interactions collectives. *In GRUNDSTEIN, M. et ZACKLAD, M., éditeurs : Ingénierie et Capitalisation des connaissances*, pages 49–64. Hermes Sciences Europe LTD.
- LOZANO-TELLO, A. et GOMEZ-PEREZ, A. (2004). ONTOMETRIC : A method to choose the appropriate ontology. *journal of database management*, 15(2):1–18.
- LUONG, P. H. (2007). *Gestion de l'évolution d'un web sémantique d'entreprise*. Thèse de doctorat, Ecole des Mines de Paris, Paris, France.
- MAEDCHE, A. (2002). *Ontology learning for the semantic Web*. Kluwer Academic Publisher.
- MAEDCHE, A. et STAAB, S. (2002). Measuring similarity between ontologies. *In EKAW*, pages 251–263.
- MANNING, C. et SCHÜTZE, H. (1999). *Foundations of Statistical Natural Language Processing*. MIT Press, Cambridge.
- MARCUS, S. et MCDERMOTT, J. (1989). Salt : a knowledge acquisition language for propose and revise systems. *Artificial Intelligence*, 39(1):1–38.
- MCAFEE, A. (2006). Enterprise 2.0 : The dawn of emergent collaboration. *MIT Sloan Management Review*, 47(3):21–28.
- MEYER, I. (2000). Extracting knowledge-rich contexts for terminography : A conceptual and methodological framework. *In BOURIGAULT, D., L'HOMME, M.-C. et JACQUEMIN, C., éditeurs : Recent Advances in Computational terminology*.
- MUSEN, M. A., ERIKSSON, H., GENNARI, J. H., TU, S. W. et PUERT, A. R. (1994). Protege ii : a suite of tools for development of intelligent systems for reusable components. *In Proc. Annu Symp Comput Appl Med Care*.
- NECHES, R., FIKES, R., FININ, T. W., GRUBER, T. R., PATIL, R. S., SENATOR, T. E. et SWARTOUT, W. R. (1991). Enabling technology for knowledge sharing. *AI Magazine*, 12(3):36–56.
- NEWELL, A. (1982). The knowledge level. *Artif. Intell.*, 18(1):87–127.
- NOY, N. F. et HAFNER, C. D. (1997). The state of the art in ontology design : A survey and comparative review. *AI Magazine*, 18(3):53–74.
- NOY, N. F. et KLEIN, M. C. A. (2004). Ontology evolution : Not the same as schema evolution. *Knowl. Inf. Syst.*, 6(4):428–440.
- OBERLE, D., VOLZ, R., STAAB, S. et MOTIK, B. (2004). An extensible ontology software environment. *In Handbook on Ontologies*, pages 299–320.
- OREILLY, T. (2007). What is web 2.0 : Design patterns and business models for the next generation of software. *Communications & strategies*, (1):17.

- PAN, J., LANCIERI, L., MAYNARD, D., GANDON, F., CUEL, R. et LEGER, A. (2007). Success stories and Best practices. Knowledge web deliverable d.1.4.2.v2.
- PLESSERS, P., TROYER, O. D. et CASTELEYN, S. (2007). Understanding ontology evolution : A change detection approach. *J. Web Sem.*, 5(1):39–49.
- POIBEAU, T. et KOSSEIM, L. (2000). Proper name extraction from non-journalistic texts. *In CLIN*, pages 144–157.
- PORZEL, R. et MALAKA, R. (2004). A task-based approach for ontology evaluation. *In ECAI - Workshop on Ontology Learning and Population*.
- PRESUTTI, V., GANGEMI, A., DAVID, S., DE CEA, G. A. et SUÁREZ-FIGUEROA, M. C. (2008). NeOn project : NeOn D2.5.1. a library of Ontology Design patterns : reusable solutions for collaborative design of networked ontologies - NeOn project. Rapport de contrat.
- PUERTA, A., EGAR, J. W., TU, S. W. et MUSEN, M. (1992). Method knowledge-acquisition shell for the automatic generation of knowledge-acquisition tools. *Knowledge Acquisition*, 4(2):171–196.
- PÉDAUQUE, R. T., éditeur (2003). *Le document : forme, signe et medium les re-formulations du numérique*. STIC-CNRS.
- PÉDAUQUE, R. T., éditeur (2005). *Le texte en jeu, permanence et transformations du document*. STIC-SHS-CNRS.
- RECTOR, A. et ROGERS, J. (2004). Patterns, properties and minimizing commitment : Reconstruction of the galen upper ontology in owl. *In EKAW*.
- REYMONET, A., THOMAS, J. et AUSSENAC-GILLES, N. (2007). Modélisation de ressources termino-ontologiques en owl. *In Actes d'IC*, pages 169–181.
- REYNAUD, C., AUSSENAC-GILLES, N., TCHOUNIKINE, P. et TRICHET, F. (1997). The notion of role in conceptual modeling. *In EKAW*, pages 221–236.
- ROSENBLOOM, S., MILLER, R. A. et JOHNSON, K. B. (2006). Interface terminologies : facilitating direct entry data into electronic health record systems. *Journal of the American Medical Informatics*, 13(3):277–288.
- ROUSSEY, C., LAURINI, R., BEAULIEU, C., TARDY, Y. et ZIMMERMANN, M. (2004). Le projet towntology : un retour d'expérience pour la construction d'une ontologie urbaine. *Revue Internationale de Géomatique*, 14(2):217–237.
- SAÏS, F., PERNELLE, N. et ROUSSET, M.-C. (2009). Combining a Logical and a Numerical Method for Data Reconciliation. *Journal on Data Semantics*, (12):66–94.
- SCHREIBER, G., AKKERMANS, A., ANJEWIERDEN, A., DEHOOG, R., SHADBOLT, N., Van de VELDE, W. et WIELINGA, B., éditeurs (1999). *Knowledge Engineering and Management : The CommonKADS Methodology*. MA : MIT Press, Cambridge, London.
- SCHREIBER, G. et WIELINGA, B., éditeurs (1992). *KADS : a principled approach to knowledge-based system development*. Academic Press, London.
- SCHREIBER, G., WIELINGA, B. J., AKKERMANS, H., de VELDE, W. V. et ANJEWIERDEN, A. (1994). Cml : The commonkads conceptual modelling language. *In EKAW*, pages 1–25.
- SCHUTZ, A. et BUITELAAR, P. (2005). Relext : A tool for relation extraction from text in ontology extension. *In International Semantic Web Conference*, pages 593–606.
- SHADBOLT, N., O'HARA, K. et CROW, L. (1999). The experimental evaluation of knowledge



- acquisition techniques and methods : history, problems and new directions. *Int. Journal of Human-Computer Study*, 51(4):729–755.
- SOWA, J. F. (1984). *Conceptual Structures : Information Processing in Mind and Machine*. Addison-Wesley, London.
- STEELS, L. (1990). Components of expertise. *AI Magazine*, 11(2):28–49.
- STEFIK, M. (1995). *Introduction to knowledge systems*. Morgan Kaufmann.
- STOJANOVIC, L. (2004). *Methods and tools for ontology evolution*. Thèse de doctorat.
- STUCKENSCHMIDT, H. et KLEIN, M. C. A. (2003). Integrity and change in modular ontologies. *In IJCAI*, pages 900–908.
- STUCKENSCHMIDT, H., PARENT, C. et SPACCAPIETRA, S., éditeurs (2009). *Modular Ontologies : Concepts, Theories and Techniques for Knowledge Modularization*, volume 5445 de *Lecture Notes in Computer Science*. Springer.
- STUDER, R., BENJAMINS, V. R. et FENSEL, D. (1998). Knowledge engineering : Principles and methods. *Data Knowl. Eng.*, 25(1-2):161–197.
- SUÁREZ-FIGUEROA, M.-C., GÓMEZ-PÉREZ, A., MOTTA, E. et GANGEMI, A., éditeurs (2012). *Ontology Engineering in a Networked World*. Springer, Berlin.
- SVATEK, V. (2004). Design patterns for semantic web ontologies : Motivation and discussion. *In Conference on Business Information Systems*.
- SZULMAN, S. (2011). Une nouvelle version de l’outil Terminae de construction de ressources termino-ontologiques. *In 22èmes Journées francophones d’Ingénierie des Connaissances. (IC 2011)*, page démonstration, Chambéry, France.
- TU, S. W., ERIKSSON, H., GENNARI, J. H., SHAHAR, Y. et MUSEN, M. A. (1995). Ontology-based configuration of problem-solving methods and generation of knowledge-acquisition tools : application of PROTÉGÉ-II to protocol-based decision support. *Artificial Intelligence in Medicine*, 7:257–289.
- VANDEBUSSCHE, P.-Y. et CHARLET, J. (2009). Méta-modèle général de description de ressources terminologiques et ontologiques. *In Actes d’IC*, pages 193–204.
- VIRBEL, J. et LUC, C. (2001). Le modèle d’architecture textuelle : fonements et expérimentation. XXIII(1):103–123.
- ZACKLAD, M. (2007). Classification, thesaurus, ontologies, folksonomies : comparaisons du point de vue de la recherche ouverte d’information (roi). *In Conférence CAIS/ACSI*.

# Index

- alignement
  - d'ontologies, 17
  - de concept, 17
- approche constructiviste, 9
- cognition située, 9
- communautés de pratiques, 24
- connaissances
  - acquisition des, 1
  - de raisonnement, 2
  - du domaine, 2
  - ingénierie des, 1
  - représentation de, 17
  - source de, 5
  - systèmes à base de, 2
- construction de modèle, *voir* modèle
- document, 20
- document, 6
  - textuel, 6
- grille répertoire, 9
- intégration de données, 22
- méthode
  - ascendante, 7
  - descendante, 7
- moèle
  - qualité, 25
- modèle
  - conceptuel, 2
  - construction de, 9
  - de raisonnement, 2, 12
  - réutilisation de, 13
  - construction de, 11, 13
  - de haut niveau, 5
  - du domaine, 3, 5
  - langage de représentation, 18
  - modulaire, 21
  - noyau, 5
  - réutilisation de, 13, 16
- patron de connaissances, 16
- plateforme
  - de modélisation, 12
  - de construction d'ontologie, 12
  - de TAL, 11
- représentation des connaissances, *voir* connaissances
- sémantique
  - formelle, 4
  - interprétative, 4
  - service web, *voir* service web sémantique
- service web, 22
  - sémantique, 23
- tâche, 12
- tâche générique, 12
- Web
  - 2.0, 23
  - des données liées, 23
  - sémantique, 18
- ontologie, 3
  - évolution d', 24

# Table des matières

<b>20 Ingénierie des connaissances</b>	<b>1</b>
20.1 Introduction	1
20.2 Modélisations utilisées	2
20.2.1 La notion de modèle conceptuel	2
20.2.2 Les modèles de raisonnement	2
20.2.3 Des modèles conceptuels aux ontologies	3
20.3 Problèmes considérés et résultats	5
20.3.1 Les sources de connaissances	6
20.3.2 Comment passer des sources de connaissances aux modèles : questions de recherche	7
20.3.3 La construction de modèles : techniques, méthodes et outils	8
20.3.4 Réutilisation de modèles	15
20.3.5 Représentation des connaissances dans les modèles	17
20.4 Enjeux méthodologiques et applicatifs actuels	19
20.4.1 Articuler la langue, les connaissances et leur support	20
20.4.2 Faire face à l'explosion des données	21
20.4.3 Gérer l'intégration des connaissances par les ontologies	22
20.4.4 Gérer les données distribuées	22
20.4.5 Tirer parti des nouvelles sources de connaissances	23
20.4.6 Faire face à l'évolution des connaissances	24
20.4.7 Dimension collective <i>versus</i> personnelle des connaissances	24
20.4.8 Évaluer la qualité des modèles	25
20.5 Conclusion	25