



HAL
open science

Similarité sémantique inter ontologies basée sur le contexte

Patrick Hoffmann

► **To cite this version:**

Patrick Hoffmann. Similarité sémantique inter ontologies basée sur le contexte. Informatique [cs].
Université Claude Bernard - Lyon I, 2008. Français. NNT: . tel-00363300v2

HAL Id: tel-00363300

<https://theses.hal.science/tel-00363300v2>

Submitted on 27 Apr 2009

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

N° d'ordre

Année **2008**

THÈSE

présentée

devant l'UNIVERSITÉ CLAUDE BERNARD – LYON 1

pour l'obtention

du DIPLÔME DE DOCTORAT
(Spécialité Informatique, CNU 27)

(arrêté du 7 août 2006 et arrêté du 6 janvier 2005)

présentée et soutenue publiquement le **16 décembre 2008**

par

M. Patrick HOFFMANN

TITRE :

**Similarité sémantique inter-ontologies basée sur le
contexte**

(RÉSUMÉ)

Directeurs de thèse : Parisa GHODOUS / Ram D. SRIRAM

JURY : M. Kokou YÉTONGNON, rapporteur
M. Aris M. OUKSEL, rapporteur
Mme Parisa GHODOUS, examinateur
M. Djamel BENSLIMANE, examinateur
M. Shaw C. FENG, examinateur

Table des matières

| | |
|---|-----------|
| Table des matières | 3 |
| Liste des figures | 5 |
| Liste des tables | 6 |
| INTRODUCTION GENERALE | 9 |
| Approche et contributions..... | 9 |
| Organisation du document | 11 |
| PARTIE I – INTRODUCTION | 13 |
| Chapitre 1 : Limitations de l’alignement d’ontologies comme un moyen d’atteindre l’interopérabilité sémantique | 14 |
| PARTIE II – ETAT DES LIEUX SUR LE CONTEXTE ET LA SIMILARITE SEMANTIQUE | 20 |
| Chapitre 2 : Revue de la littérature | 21 |
| 2.1 Une nouvelle classification pour les approches basées sur le contexte | 21 |
| 2.2 Évaluation de la relation entre concepts définis par différentes ontologie | 24 |
| 2.3 Comparaison de concepts basés sur le contexte..... | 26 |
| PARTIE III – METHODOLOGIE POUR UNE SOLUTION BASEE SUR LE CONTEXTE | 27 |
| Chapitre 3 : Méthodologie pour une contextualisation efficace | 28 |
| 11.1 Phase 1 : Directives pour contextualiser l’application | 29 |
| 11.2 Phase 2 : Connexion entre le contexte de la cible et la référence | 33 |
| 11.3 Phase 3 : construire une architecture qui atteint les objectifs fixés | 36 |
| 11.4 Conclusion | 36 |
| PARTIE IV – APPLICATION DE LA METHODOLOGIE A LA RECONCILIATION DES ONTOLOGIES | 38 |
| Chapitre 4 : Directives pour la contextualisation de l’alignement d’ontologies | 39 |
| 4.1 Objectifs..... | 39 |
| 4.2 Les usages du contexte..... | 40 |
| 4.3 Triplets (cible, référence, intention)..... | 41 |
| 4.4 Conclusion | 42 |
| Chapitre 5 : Connexion entre le contexte du concept origine et celui du concept enquis | 43 |
| 5.1 Caractéristiques pertinentes pour décrire le contexte des concepts..... | 43 |
| 5.2 Source des données et conditions pour les récupérer | 44 |

| | | |
|---|---|-----------|
| 5.3 | Modèle des perspectives..... | 45 |
| 5.4 | Modèle de la connexion entre perspectives..... | 48 |
| 5.5 | Évaluation des données pour la relation..... | 48 |
| 5.6 | Interprétation des données pour réaliser la désambiguïsation..... | 49 |
| 5.7 | Conclusion..... | 49 |
| Chapitre 6 : Connexion entre le contexte du demandeur et celui du concept origine. | | 51 |
| 6.1 | Caractéristiques pertinentes pour décrire le contexte du demandeur..... | 51 |
| 6.2 | Source des données et conditions pour les récupérer..... | 51 |
| 6.5 | Évaluation des données pour la connexion..... | 52 |
| 6.7 | Conclusion..... | 53 |
| Chapitre 7 : Connexion entre le contexte du besoin en interopérabilité et celui du concept enquis..... | | 54 |
| 7.1 | Caractéristiques pertinentes..... | 54 |
| 7.2 | Conditions pour récupérer des données concernant la tâche d'interopérabilité..... | 54 |
| 7.3 | Modèle de la tâche d'interopérabilité..... | 55 |
| 7.4 | Modèle de la connexion entre la tâche d'interopérabilité et les perspectives..... | 56 |
| 7.5 | Évaluation des données pour la connexion..... | 56 |
| 7.6 | Interprétation des données pour évaluer le concept enquis..... | 56 |
| 7.7 | Conclusion..... | 57 |
| Chapitre 8 : Une architecture basée sur le contexte qui prend en compte les implications pratiques pour une meilleure réconciliation des ontologies..... | | 58 |
| 8.1 | Architecture..... | 58 |
| 8.2 | Évaluation..... | 60 |
| 8.4 | Conclusion..... | 65 |
| CONCLUSION GENERALE..... | | 67 |
| | Résumé de la thèse et contributions..... | 67 |
| | Discussions et limites..... | 68 |
| | Travaux futurs..... | 69 |
| | Perspectives..... | 70 |
| BIBLIOGRAPHIE..... | | 71 |

Liste des figures

| | |
|--|----|
| FIGURE 1 : UTILISATION D'UNE ONTOLOGY POUR PERMETTRE A DEUX APPLICATION D'INTER-OPERER | 15 |
| FIGURE 2 : CLASSIFICATION OF ONTOLOGY MATCHING METHODS | 16 |
| FIGURE 3 : VARIOUS CLASSIFICATIONS OF CONCEPTS (KLEIN, 2001) | 17 |
| FIGURE 4 : ILLUSTRATION DE LA METHODOLOGIE..... | 28 |
| FIGURE 5 : PHASE 1 DE LA METHODOLOGIE..... | 29 |
| FIGURE 6 : DEUXIEME PHASE DE LA METHODOLOGIE | 33 |
| FIGURE 7 : TROISIEME PHASE DE LA METHODOLOGIE..... | 36 |
| FIGURE 9 : DEFINITIONS DE TERMES POUR NOTRE APPROCHE..... | 40 |
| FIGURE 10 : LES USAGES DU CONTEXT SELECTIONNES | 42 |
| FIGURE 11 : APPLICATION DE LA METHODOLOGIE (DEUXIEME PHASE, PREMIER USAGE DU CONTEXTE) | 43 |
| FIGURE 12 : SIGNES POUR RECONNAITRE LES DIFFERENTS TYPES DE PERSPECTIVE DANS UNE ONTOLOGIE | 45 |
| FIGURE 13 : CLASSIFICATION DES ROUTES FRANÇAISES | 46 |
| FIGURE 14 : EXEMPLE D'ONTOLOGIE DE PERSPECTIVES..... | 47 |
| FIGURE 15 : ILLUSTRATION DU PROCESSUS DE DESAMBIGUÏSATION | 49 |
| FIGURE 16 : APPLICATION DE LA METHODOLOGIE (DEUXIEME PHASE, LES SIENNES USAGE DU CONTEXTE)..... | 51 |
| FIGURE 17 : APPLICATION DE LA METHODOLOGIE (DEUXIEME PHASE, TROISIEME USAGE DU CONTEXTE)..... | 54 |
| FIGURE 18 : APPLICATION DE LA METHODOLOGIE (TROISIEME PHASE)..... | 58 |
| FIGURE 19 : ILLUSTRATION DE L'ARCHITECTURE AVEC UN EXEMPLE..... | 59 |
| FIGURE 20 : COMPARAISON DES NORMES : $-\log(x)$ ET NOTRE NORME ("OUR NORM")..... | 64 |

Liste des tables

TABLE 1 : COMMON FUNCTIONS FOR THE MEASURES OF SIMILARITY 64

Abbreviations et acronymes

| | |
|-------|---|
| CAD | COMPUTER-AIDED DESIGN |
| CAE | COMPUTER-AIDED ENGINEERING |
| CAM | COMPUTER-AIDED MANUFACTURING |
| CE | CONCURRENT ENGINEERING |
| DBF | DESIGN BY FEATURE |
| DL | DESCRIPTION LOGICS |
| FBS | FUNCTION-BEHAVIOUR-STRUCTURE FRAMEWORK |
| FTP | FILE TRANSFER PROTOCOL |
| IRI | INTERNATIONAL RESOURCE IDENTIFIER |
| IDE | INTEGRATED DEVELOPMENT ENVIRONMENT |
| ISO | INTERNATIONAL ORGANIZATION FOR STANDARDIZATION |
| NIST | NATIONAL INSTITUTE OF STANDARDS AND TECHNOLOGY |
| NLP | NATURAL LANGUAGE PROCESSING |
| OAEI | ONTOLOGY ALIGNMENT EVALUATION INITIATIVE |
| OWL | WEB ONTOLOGY LANGUAGE |
| PLM | PRODUCT LIFECYCLE MANAGEMENT |
| RDF | RESOURCE DESCRIPTION FRAMEWORK |
| SAT | PROPOSITIONAL SATISFIABILITY |
| STEP | STANDARD FOR THE EXCHANGE OF PRODUCT MODEL DATA |
| HTTP | HYPERTEXT TRANSFER PROTOCOL |
| URI | UNIVERSAL RESOURCE IDENTIFIER |
| XHTML | EXTENDED HYPERTEXT MARKUP LANGUAGE |
| XML | EXTENSIBLE MARKUP LANGUAGE |
| WWW | WORLD WIDE WEB |

Notes typographiques

Concepts : **Thing** ,

Relations sémantiques : *is-a*, *part-of*, etc.

Attribute : price

Introduction générale

La mondialisation des échanges et l'augmentation de la compétition rendent d'autant plus nécessaire pour les compagnies de collaborer les unes avec l'autres, afin de produire mieux à moindre coût. Dans de telles collaborations, elles échangent des données nécessaires à leurs applications logicielles, afin de leur permettre d'opérer de conserve (inter-opérer).

L'hypothèse de départ pour cette thèse, est que les compagnies modèlent par des ontologies les concepts qui décrivent la significations des données que leurs applications logicielles manipulent, et qu'ils permettent l'accès à cette ontologie aux organisations partenaires.

Il existe divers outils pour aider à réconcilier les ontologies, à l'aide de « mappings » entre leurs concepts. La plupart de ces outils se basent sur des méthodes terminologiques et structurelles pour déterminer le degré de similarité, ou même l'équivalence des concepts.

Lorsqu'il apparaît nécessaire de permettre à des applications logicielles d'opérer de conserve, c'est qu'il y a un besoin d'interopérabilité qui peut être exprimé à l'aide d'une requête. Le succès de la coopération et des échanges de données qui en découlent, sont à évaluer selon la pertinence de l'information ou du service obtenu pour répondre à ce besoin.

Mais différentes compagnies ont différents points de vue sur les données manipulées. Il n'est donc pas suffisant de s'assurer de ce que les concepts sont mappés correctement entre les ontologies. En effet, le concept peut être le concept approprié, mais ne pas permettre une réponse satisfaisante au besoin exprimé. Par exemple, il se pourrait que le type des données associées au concept n'ait rien à voir avec l'usage que la compagnie voulait faire des données.

Approche et contributions

Ce travail de recherche a été effectué dans le cadre d'une cotutelle entre le Laboratoire d'InfoRmatique en Image et Systèmes d'information (LIRIS) et le National Institute of Standards and Technology (NIST). Cette collaboration a permis de bénéficier des idées des chercheurs du NIST spécialisés sur l'interopérabilité des processus et de la conception de produit par standardisation.

L'approche entreprise a consisté à examiner si une approche basée sur le contexte est à même d'améliorer la réconciliation entre ontologies effectuée à l'aide de mappings. En effet, les méthodes et outils utilisés pour aligner les ontologies ne considèrent aucune information contextuelle renseignant sur l'utilisation effective des concepts. Nous estimons que par cette considération du contexte, il sera possible de fournir une évaluation de la pertinence d'un concept pour une requête donnée qui soit bien meilleure, car reliée à la réalité du terrain.

Nous avons donc passé en revue la littérature traitant du contexte en informatique, les diverses définitions, modélisation et utilisations, afin de découvrir les principes inhérents du contexte. Nous avons ensuite développé une méthodologie pour reconnaître ce qui est contextuel dans une certaine situation, comment le modéliser en vue de l'utiliser. Il s'agit ici de notre première contribution.

En appliquant cette méthodologie à la réconciliation d'ontologies dans le cadre de collaborations entre compagnies, nous avons trouvé trois contextes pertinents :

- Le contexte des concepts, que nous caractérisons par les « perspectives » qui ont guidé le développement de l'ontologie qui les définit, ou qui indique les variations dans le type de données associées avec les concepts de l'ontologie. C'est ici notre deuxième contribution.
- Le contexte de l'agent qui est à l'origine de la requête, modélisé par une sélection des domaines et tâches de la compagnie de l'agent.
- Le contexte du besoin en interopérabilité, constitué de la « tâche d'interopérabilité » qui décrit ce qui avait été projeté de faire avec le résultat de la requête.

Ces contextes seront employés selon les usages suivants :

- En comparant les perspectives associées aux concepts, on peut *désambiguïser* ce qu'ils impliquent en pratique.
- En comparant les domaines et tâches de l'agent avec les perspectives des concepts, on va pouvoir *personnaliser* la désambiguïsation effectuée, en reconnaissant la compréhension que l'agent a d'un concept.
- En comparant la tâche d'interopérabilité avec les perspectives associées au concept, on va pouvoir évaluer l'intérêt du concept de manière pratique.

Les mappings associent des concepts définis par des ontologies de la compagnie de l'agent avec des ontologies définis par d'autres compagnies. Comme l'agent peut exprimer une requête à propos de n'importe quel concept, ils peuvent aisément ne pas être en nombre suffisant.. Nous proposons par conséquent une mesure de similarité sémantique inter-ontologies. Nous réutilisons pour ce faire la mesure de [Leacock and Chodorow, 1998] qui est basée sur le comptage de relations *is-a*. Il s'agit ici de notre troisième contribution.

Finalement, nous proposons une mesure d'évaluation de la pertinence d'un concept. Cette mesure repose sur les trois modèles de contexte décrits précédemment. Comme le concept est évalué dans le cadre d'un besoin en interopérabilité, cette évaluation fait l'objet d'une requête. Nous proposons que cette requête soit composée des quatre éléments suivants : l'identifiant de l'agent qui la soumet, le concept « racine » défini par une des ontologies de la compagnie de l'agent, qui donne une approximation du résultat attendu ; la tâche d'interopérabilité ; enfin, le concept « demandé » défini dans une ontologie de l'organisation partenaire, et qui va être évalué. Cette mesure est notre dernière contribution.

Organisation du document

Ce résumé de thèse est organisé en quatre chapitres : le premier chapitre décrit la notion d'interopérabilité sémantique qui promet une mise en relation flexible de logiciels et données hétérogènes, mais montre ses limitations lorsque ceux-ci sont décrits par des ontologies différentes. Le deuxième chapitre passe en revue la littérature sur le contexte, l'évaluation de la relation de similarité et proximité sémantique de concepts et sa contextualisation. Le troisième chapitre décrit notre méthodologie pour déterminer ce qui est contextuel, modéliser le contexte et l'utiliser ; puis nous appliquons cette méthodologie à la réconciliation d'ontologies. Le quatrième chapitre expose l'architecture d'un système basé sur l'approche développée, et présente un exemple d'application.

PARTIE I – Introduction

Chapitre 1 : Limitations de l’alignement d’ontologies comme un moyen d’atteindre l’interopérabilité sémantique

Avec la mondialisation et l’augmentation de la compétition, les compagnies doivent collaborer plus que jamais avec d’autres organisations, afin de pouvoir mettre au point et produire de meilleurs produits pour un prix réduit. Dans de tels cas, ces organisations ont besoin d’échanger des données et d’utiliser les données échangées dans leurs applications logicielles, c’est-à-dire il y a un besoin d’interopérabilité.

L’interopérabilité est traditionnellement réalisée par le développement de formats standards qui servent de représentation neutre, ou par le développement de traducteurs et de mappings entre les schémas de bases de données, afin d’intégrer diverses applications logicielles les unes avec les autres. Mais l’intégration est coûteuse, demande du temps, et peut générer de nouvelles erreurs ; et le processus de standardisation requiert énormément de temps, trop pour pouvoir gérer les besoins croissants, vu l’accélération du rythme de création de nouvelles technologies. En raison de ces limitations, la communauté scientifique cherche un moyen de permettre à des systèmes développés de manière autonome d’échanger des informations avec le minimum d’interventions manuelles.

La direction prise actuellement est d’atteindre l’interopérabilité au niveau sémantique : quand les données sont annotées à l’aide d’annotations sémantiques qui décrivent leur signification, on peut établir des correspondances entre éléments de données qui ont la même signification. Alors ces mappings peuvent servir à relier les unes aux autres les applications logicielles qui manipulent les données, afin de pouvoir travailler ensemble. L’interopérabilité réalisée de cette manière est nommée « interopérabilité sémantique », qui est définie comme « la capacité de systèmes d’information d’échanger de l’information en se basant sur un ensemble de termes et expressions dont la signification fait l’objet d’un consensus, est établie à l’avance et négociée »¹ [Veltman, 2001].

La description de la signification de données établies dans ce qu’on appelle une ontologie. Ne les quitteront la définition suivante se d’une ontologie, adapter et de [Guarino, 1998] : un artefact d’ingénierie qui décrit un modèle traitable de manière informatique, est constitué d’un vocabulaire formel ainsi que d’un ensemble de conjonctures explicites qui définissent de manière précise et claire la signification donnée au vocabulaire; le modèle décrit les classes (représentants des concepts), des instances (représentant des individus de concepts), des attributs, et des relations exprimant des contraintes entre eux.

Les organisations vont établir la signification des données de leurs applications en se référant à des ontologies différentes. Il est donc nécessaire de réconcilier ces ontologies -- c’est-à-dire de les amener en agrément -- si l’on veut permettre à ces organisations de

¹ Ma traduction de “*the ability of information systems to exchange information on the basis of shared, pre-established and negotiated meanings of terms and expressions*”

collaborer. La réconciliation des ontologies est rendue difficile par les hétérogénéités des ontologies qui ont été développées de manière autonome ; ces hétérogénéités sont d'ordre terminologiques, conceptuelles, ou peuvent provenir de la modélisation. Trois sortes d'approches ont été proposées pour la réconciliation de deux ou plus ontologies :

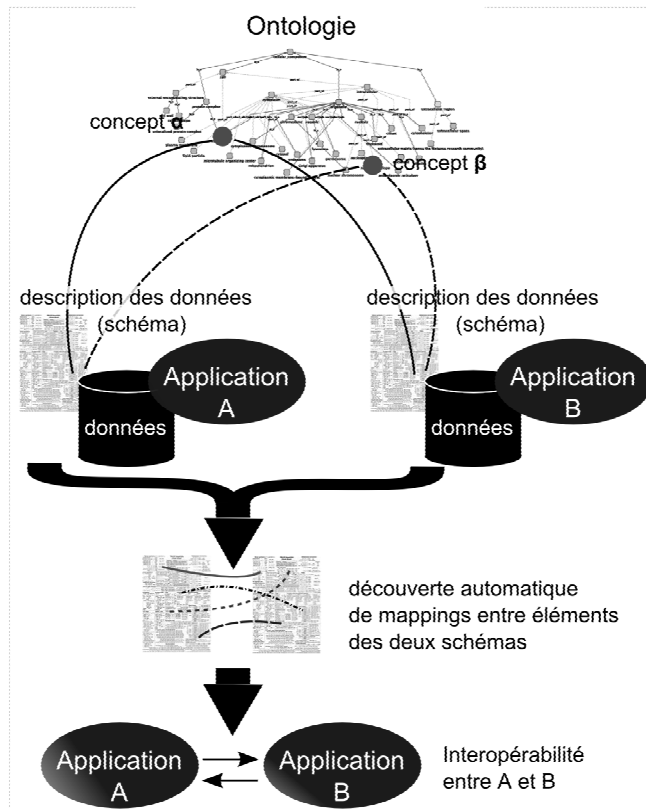


Figure 1 : Utilisation d'une ontology pour permettre à deux application d'inter-opérer

- Faire concorder les ontologies à l'aide d'un vocabulaire partagé ainsi que des règles définies dans une ontology de plus haut niveau (« upper ontology »). Cette approche nommée « *unification* » requiert qu'une telle ontology de plus haut niveau existe, que les organisations se mettent d'accord, et qu'elles développent ou modifient leurs ontologies en conséquence. Comme la plupart des collaborations ne seront pas connues avant que les ontologies ne soient développées, cette approche requiert de modifier les ontologies en accord avec l'ontology de plus haut niveau. Comme il peut y avoir plusieurs ontologies de plus haut niveau qui soient pertinentes, cette approche amènera à développer plusieurs versions d'ontology, et que les garder à jour.
- Fusionner les ontologies en une seule ontology. Cette approche nommée « *intégration* » requiert de développer une nouvelle ontology qui reflète un consensus sur le point de vue des différentes organisations qui collaborent, afin d'en limiter les inconsistances. En raison de ces compromis, la nouvelle ontology pourra n'avoir qu'une compatibilité limitée avec les ontologies source des différentes organisations.

- Trouver les correspondances entre les entités des deux ontologies (alignement ou mapping d'ontologies). Cette approche nommée fédération s'appuie habituellement sur des outils semi-automatiques, basés sur des méthodes heuristiques qui comparent principalement la terminologie et la structure des ontologies afin de détecter les couples de concepts qui sont reliés au niveau sémantique (similarité ou équivalence), et nommés mappings.

Comme le besoin en interopérabilité que nous considérons apparaît dans le contexte de collaboration, et il s'agit de réconcilier des ontologies développées de manière autonome, la flexibilité est le critère le plus important. Les approches fédératives permettent l'échange parmi des ressources développées dans des buts indépendants et qui évoluent de manière indépendante ; elles semblent donc les plus appropriées.

Cependant malgré tous les efforts investis dans les méthodes et outils pour l'alignement d'ontologies, les résultats sont toujours encore décevants. Les méthodes terminologiques et structurelles se basent sur des choix subjectifs des concepteurs de l'ontologie. Les méthodes sémantiques considèrent bien la sémantique décrite par les ontologies, mais comme les ontologies sont la plupart du temps incohérentes les unes avec les autres, le gain attendu est limité. La réelle signification des mappings est liée de manière directe à l'évaluation de la similarité entre les concepts. Si les mappings ont été générés par des outils automatiques d'alignement basés sur des méthodes terminologiques et structurelles uniquement, alors les concepts seront jugés équivalents s'ils sont nommés et de la même manière, ou ont des mêmes relations sémantiques avec des concepts similaires.

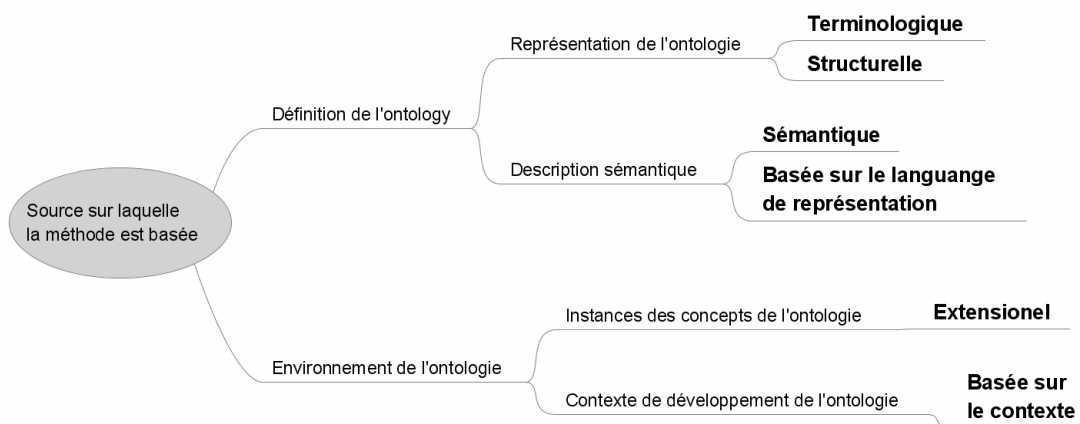


Figure 2 : Classification of ontology matching methods

Mais même lorsque l'alignement est vérifié par des experts, les concepts jugés équivalents pourront exprimer la même idée, mais selon les perspectives distinctes. En effet, les ontologies sont développées pour des applications différentes, parfois avec des points de vue incompatibles, selon diverses règles de conception et choix technique. Les décisions de conception sont prises en fonction des aspects du monde qui sont pertinents [stuckenschmidt, 2006] ; ceci comprend l'influence de la culture de l'organisation, l'expérience qu'a le développeur de l'ontologie dans le domaine, et ainsi de suite. Les ontologies sont des représentations partielles d'un monde complexe, et la conceptualisation qu'elles représentent dépend donc de la tâche pour laquelle elles ont

été développées. En comparant des ontologies de domaine dont l'intersection n'est pas nulle, ces différentes représentations sont comparées. Le paradigme utilisé pour le développement de l'ontologie n'est pas représenté dans l'ontologie, au moins pas formellement. La méthode utilisée pour convertir la conceptualisation ont une représentation n'est pas non plus incluse.

Le modèle ontologique était conçu pour le raisonnement est dans le but de garantir la cohérence du modèle, mais pas pour comparer différents modèles les uns avec les autres pour le partage et la réutilisation. Il est donc adapté pour représenter une information qui est connue, mais pas – comme c'est souvent le cas – une information complexe et sujette à différents points de vue. Les ontologies ne définissent pas la signification des entités de manière formelle, mais uniquement textuelle, et elles identifient les entités par l'assignement d'une URI unique. Il n'y a donc aucun moyen formel pour comparer des entités similaires directement, sauf dans le cas où une relation sémantique est décrite de manière explicite entre elles. Il n'y a pas de librairie standard de concepts, pas de règle de construction non plus qui permette de définir une manière précise et formelle de nouveaux concepts à partir des concepts existants.

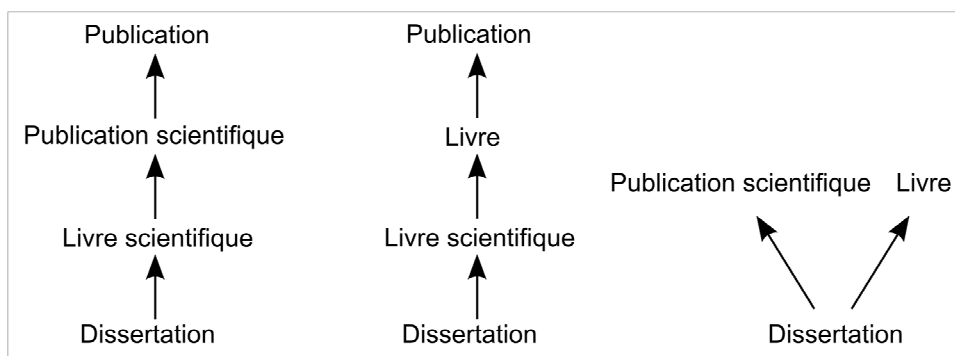


Figure 3 : Various classifications of concepts [Klein, 2001]

Les mappings auront donc un intérêt limité si les perspectives avec lesquels les ontologies ont été développées ne sont pas compatibles. Les entités jugées équivalentes ou similaires auront probablement été incluses pour des raisons distinctes, et ainsi la relation avec d'autres entités, le type de données associées à l'entité, peut être complètement différent. Musen et Day, dans leur présentation pour le BioPortal début 2008, ont donné l'exemple de mapping entre le concept « sang » d'une ontologie décrivant l'anatomie d'un poisson zèbre (un poisson d'aquarium populaire) et le concept « sang » d'une ontologie décrivant l'anatomie d'une souris adulte. Que veut dire un tel mapping ? Si du sang de poisson zèbre était transféré dans le sang d'une souris adulte, cela n'amènerait-il pas des complications ?

Nous nous intéressons à l'alignement d'ontologies pour l'interopérabilité sémantique. D'après [Giunchiglia et al, 2006], l'interopérabilité sémantique est hautement dépendante du contexte et de la tâche. Il est certain que toutes les données n'ont pas la même utilité pour chaque tâche, et que toute donnée n'est pas appropriée dans tout contexte. Pour réaliser l'interopérabilité sémantique, il est donc nécessaire de prendre en

considération le contexte et la tâche, lorsque les mappings sont établis, évalués, ou utilisés.

Lorsqu'un nouveau besoin en interopérabilité apparaît, c'est dans un contexte particulier, et le résultat attendu de la requête correspondante est censé être utilisé pour une tâche particulière. Les méthodes pour la réconciliation d'ontologies doivent donc prendre cette information en considération, afin de la comparer avec le contexte des sources de données interrogées, et juger ainsi de la pertinence de ces données. Cependant, le contexte de développement de l'oncologie, incluant la tâche pour laquelle l'ontologie a été développée, n'est habituellement explicité dans aucun document qui puisse servir de référence. Et l'information contextuelle qu'est le besoin en interopérabilité n'est pas connue encore quand les mappings sont établis.

Les mappings sont supposés exprimer un degré de similarité entre les concepts. Mais la notion de similarité dépend du contexte, et les propriétés qui sont utilisées pour comparer les concepts doivent donc être choisies selon ce contexte [Rada et al, 1989]. Les mappings générés par les outils d'aide à l'alignement d'ontologies établissent la similarité des concepts une fois pour toutes, et ne considèrent aucune information contextuelle. Ils ne peuvent donc satisfaire le besoin en interopérabilité sémantique, et lorsqu'ils sont utilisés, il faut donc considérer le contexte par d'autres moyens. En effet, même si le concept est le bon, il peut ne pas être une réponse pertinente pour le besoin exprimé. Il peut par exemple être associé avec des données complètement inutiles pour l'usage que la compagnie avait l'intention d'en faire.

Le but de cette étude est d'évaluer si la considération du contexte peut amener à une amélioration de la fiabilité des mapping pour l'interopérabilité. Nous proposons une méthodologie pour déterminer ce qui fait partie du contexte, et comment le collecter, le modéliser, et l'employer. En appliquant cette méthodologie, nous découvrons trois usages du contexte qui peuvent améliorer la réconciliation d'ontologies, ensemble ou pris séparément : nous proposons de *désambiguïser* parmi les significations pragmatiques possibles des concepts en comparant les « perspectives » avec lesquels les concepts ont été développés ; de *personnaliser* cet comparaison en considérant le contexte de l'agent, composé d'une sélection pertinente des domaines et tâches de sa compagnie ; d'*évaluer* la pertinence des données associées aux concepts pour la tâche qui a déclenché le besoin en interopérabilité.

Une telle proposition pourrait être appliquée par exemple à la collaboration d'une compagnie spécialisée dans la préparation de kits, leur assemblage, et leur emballage, avec ses fournisseurs. Un ingénieur en assemblage pourrait alors interagir avec une application logicielle de conception intégrée connectée à un système en ligne qui soit basé sur le contexte. Après avoir rempli l'information correspondant à son contexte, l'ingénieur pourrait entreprendre une recherche pour une pièce identifiée par un concept. L'application de conception pourrait alors interroger le système basé sur le contexte, afin de retrouver tous les concepts similaires dans les ontologies des fournisseurs, ordonnés selon leur pertinence pour le besoin présent. L'ingénieur pourrait alors trouver rapidement les instances les plus appropriées dans la base de données.

PARTIE II – Etat des lieux sur le contexte et la similarité sémantique

Chapitre 2 : Revue de la littérature

La notion de contexte est largement utilisée de nos jours en informatique, mais n'est pas bien comprise. J'ai donc proposé une nouvelle définition et classification, sur laquelle nous nous appuierons lorsque nous établirons la méthodologie basée sur le contexte.

2.1 Une nouvelle classification pour les approches basées sur le contexte

Historiquement, la notion de contexte a été appliquée tout d'abord à un mot ou passage, pour indiquer « *la partie d'un discours qu'il entoure et qui peut éclairer sa signification* ». Le terme a progressivement été utilisé dans d'autres circonstances avec la signification plus générale de « *conditions interdépendantes dans lesquelles quelque chose existe ou se produit* ». La notion de contexte a été appliquée à divers domaines en informatique, tels que l'intelligence artificielle (AI), le développement de logiciels, les bases de données, l'intégration de données, l'apprentissage automatique, la représentation de connaissance, la recherche d'informations (RI).

La définition du contexte que nous allons retenir pour cet article sera : « *toute information qui participe à caractériser une entité particulière impliquée dans l'événement qui a déclenché le besoin de contexte, incluant ses interactions avec d'autres entités, et où les caractéristiques distinctives qui composent le contexte sont jugées selon la pertinence avec laquelle elles permettent d'expliquer l'émergence de cette entité ou ses caractéristiques les plus remarquables.* »

Nous allons maintenant observer un panorama d'approches qui utilisent le contexte, classifiées en accord avec ce qui semble être la principale contribution de chaque approche, qu'il s'agisse de collecter, de modéliser, ou d'employer de l'information contextuelle. Bien que la plupart de ces approches ne sont pas pertinentes en soi pour notre étude, elles sont utiles pour mettre en évidence certains principes à propos de la notion de contexte.

2.1.1 Collecter l'information contextuelle

Certaines approches se focalisent principalement sur la collecte automatique des informations contextuelles. Les données récupérées peuvent être directement transmises à l'utilisateur final, ou servir comme provision pour d'autres applications. Quand la collecte d'informations est faite avant l'émergence d'un besoin, la quantité de données collectées est en fait supérieure à ce qui est à proprement parler contextuel. Des exemples de telles données sont les historiques des opérations, et l'enregistrement de métadonnées sur des artefacts.

L'historique des opérations est utile lorsqu'on veut caractériser l'émergence d'un événement particulier. Cela peut servir à sélectionner quelles actions sont appropriées [Brézillon, 2007] ; cela peut servir également à garder des traces des circonstances d'une découverte : c'est le cas pour l'outil SearchPad [Bharat, 2000] qui garde les traces des requêtes formées pendant une ou plusieurs sessions de recherche sur le Web et les associe avec les résultats respectifs.

[Paslaru-Bontas, 2007] a proposé un modèle de métadonnées en OWL-DL qui inclut les caractéristiques du développement de l'ontologie qui sont pertinentes pour guider le processus de réutilisation de l'ontologie. Ce modèle permet d'évaluer si deux ontologies ont été développées dans des contextes suffisamment similaires pour qu'il y ait un intérêt à les réconcilier. Les caractéristiques proposées se réfèrent au développement des ontologies comme un tout, et ne peuvent servir directement à la réconciliation ontologie ; il n'y a en effet pas de méthode évidente qui permette d'adapter la manière de relier les entités des ontologies les unes avec les autres en utilisant ces métadonnées.

2.1.2 Modéliser l'information contextuelle

Certains approches se focalisent sur la modélisation d'information contextuelle d'une telle manière que son utilisation en soit facilitée ; d'autres encore proposent un modèle qui permet d'interfacer des données collectées pour des applications qui ont besoin d'information contextuelle.

Certains auteurs essaient de représenter la connaissance de manière formelle afin de permettre des raisonnements logiques à partir d'ensemble de règles concernant les mêmes objets, mais incohérentes les uns avec les autres. Le but est de générer des conclusions dépendant du contexte. Selon les auteurs, les modèles sont destinés à représenter soit de l'information d'arrière plan qui varie selon l'endroit où l'on se trouve ou les circonstances [McCarthy, 1993], [Guha, 1992] ; soit différents points de vue sur une même réalité [Attardi et Simi, 1995], [Bouquet et al, 2003].

[Baldauf et al, 2007] ont rédigé un état de l'art sur les systèmes sensibles au contexte qui agissent comme une couche entre les capteurs – qui collectent les données correspondantes à la localisation géographique, le changement d'environnement, etc. – et les applications – qui vont sélectionner les données qui sont pertinentes dans le contexte d'une interaction particulière avec l'utilisateur.

[Firat et al, 2007] ont proposé un modèle de l'information contextuelle pour des données typiquement utilisées pour le commerce : unités (telles que les différentes monnaies), postulats (services ou taxes incluses ou non dans le prix), formats (par exemple, format de date). Les modèles de contexte sont instanciés pour définir les caractéristiques de l'information enregistrée pour chaque source de données représentée, aussi bien que les préférences de l'utilisateur. La liste des modificateurs de contexte possibles est fixée à l'avance, et associée à une liste de dimensions. Des fonctions de conversion transforment les données d'un contexte à un autre selon les besoins.

2.1.3 Employer l'information contextuelle

Certaines approches emploient l'information contextuelle pour un but particulier. Ces approches peuvent collecter et modéliser l'information contextuelle, ou bien se baser sur un modèle ou des métadonnées standards, afin d'atteindre leur objectif.

Le contexte est souvent un moyen de désambiguïser la recherche basée sur les mots-clés. Par exemple, [Finkelstein et al, 2001] ont proposé de désambiguïser la recherche basée sur les mots-clés en générant une requête augmentée à partir du paragraphe qui contient le mot sélectionné par utilisateur. [Budzick et Hammond, 2000] ont considéré

un contexte plus large dans ce même but ; ils assument que le besoin en information est probablement survenu alors que l'utilisateur était en train d'éditer un document ; ils utilisent comme contexte le contenu textuel d'artefacts tels que les documents Word ouverts au moment de la requête.

Le contexte est également utilisé dans un but de personnalisation. [Lawrence, 2000] a présenté un court état des lieux sur l'utilisation du contexte pour la recherche sur le Web. L'information considérée comme contextuelle peut être constituée de données personnelles sur les utilisateurs du moteur de recherche (en assumant que leur recherche concerne leurs propres besoins en information) pour personnaliser les résultats. Elle peut également être une information statistique basée sur des requêtes archivées d'un ensemble plus large d'utilisateurs, pour estimer le but de la recherche basée sur les associations de mots-clés (ceci est toujours une information contextuelle sur l'utilisateur, avec l'hypothèse que l'utilisateur est affecté par la culture environnante) ; cela peut aussi être la connaissance du moteur de recherche utilisé ou l'adresse IP de l'utilisateur (qui fournit la localisation géographique probable de l'utilisateur, et par là l'endroit où l'information retournée va probablement servir).

Le contexte va servir aussi à adapter le comportement de divers systèmes à une situation donnée. La « sensibilité au contexte » a pour objectif de permettre aux appareils électroniques d'être « sensible » à l'environnement dans lequel ils sont utilisés. Les systèmes sensibles au contexte s'appuient sur des capteurs divers (GPS, horloge, etc.) afin d'adapter l'interaction humain-ordinateur aux circonstances de l'utilisation. Le contexte est également compris comme une information « environnementale », essentiellement « situationnelle » [Baldauf et al, 2007].

Le contexte sert également comme un moyen pour mener une évaluation. [Paslaru-Bontas, 2007] fournit une méthodologie basée sur le contexte pour déterminer quelles ontologies peuvent être réutilisées pour un développement d'ontologies particuliers. Elle propose d'évaluer les ontologies candidates à la réutilisation à l'aide de critères tels que la pertinence estimée pour le domaine d'application, la qualité de la modélisation, le contexte technique, et les tâches pour lesquelles l'ontologie a été construite (qui va être comparée à la tâche pour laquelle la nouvelle ontologie va être utilisée). Le besoin de l'ontologie est en effet survenu dans un contexte particulier. La méthodologie s'appuie sur ce contexte pour raffiner et optimiser la stratégie d'intégration en conséquence. Le contexte de réutilisation de l'ontologie est donc représenté ici par la tâche et le rôle de l'ontologie, et le niveau de réutilisation.

De plus, l'information contextuelle est parfois employée pour récupérer des entités, à partir du contexte de l'interaction de l'utilisateur avec ces entités. Par exemple, la recherche d'images sur le Web est essentiellement effectuée par une recherche basée sur les mots-clés à partir du nom donné à l'image, ou des tags associés à l'image. La date de modification est aussi utilisée fréquemment comme un moyen de recherche. [Dumais et al, 2003] ont présenté un outil de recherche pour le bureau qui permet de récupérer facilement tout document que l'utilisateur a vu précédemment, en les classifiant par date. [Chirita et al, 2006] étendent l'infrastructure de recherche pour le bureau Beagle avec des capacités de recherche à partir d'information contextuelle, en se basant sur des métadonnées qui décrivent l'activité de l'utilisateur (annotations RDF).

Les caractéristiques qui composent le contexte pour une approche donnée ne seront habituellement pas pertinentes pour une autre approche. Il n'est donc pas possible de donner une définition du contexte qui serait directement utilisable, quel que soit le besoin. Nous allons maintenant étudier les mesures qui ont été proposées pour évaluer la relation entre les concepts, et nous intéresser à savoir si on peut étendre ces mesures pour l'évaluation de relation entre des concepts définis par des ontologies distinctes. Nous allons ensuite voir comment le contexte a été utilisé pour modifier de telles mesures.

2.2 Évaluation de la relation entre concepts définis par différentes ontologie

La plupart des mesures évaluent la relation entre les concepts en termes de similarité. Selon [Resnik, 1999], la similarité sémantique est une évaluation de la relation sémantique entre deux concepts avec l'objectif d'obtenir une estimation du degré de proximité entre la signification de ces deux concepts. Un autre type de mesures qui a été proposée consiste à mesurer la « connexité sémantique » (“semantic relatedness”), dont le but est d'évaluer la force de la relation sémantique entre deux concepts. Deux concepts peuvent ainsi être jugés fortement « connectés » sans être pour autant « similaires », comme par exemple les deux concepts de label « Table » et « Pot ».

2.2.1 Similarité et connexité des concepts

[Tversky, 1977] a étudié la perception humaine de la similarité d'un point de vue de psychologue. Il a proposé ce qu'on appelle des mesures basées sur les caractéristiques, qui déterminent la similarité de deux objets en comparant leurs caractéristiques communes et distinctives : plus les objets partagent de caractéristiques, et moins ils ont de caractéristique distinctive, plus ils sont similaires. [Rodriguez et Egenhofer, 1999] et [Rodriguez et Egenhofer, 2004] ont récemment appliqué ce type de mesure à la comparaison de classes d'entités² en considérant comme caractéristiques les propriétés des classes, qu'ils ont divisé en « parties », « fonctions », et « attributs ». Ils comparent les classes d'entités en se basant sur la terminologie (en supposant qu'un ensemble de synonymes est associé à chaque classe), sur la structure et sur les caractéristiques. Les auteurs construisent trois différentes mesures de similarité d'après le modèle ratio de Tversky, une pour chaque type de propriété (Equation 1). La mesure de similarité sémantique finale est une somme pondérée de ces trois mesures (Equation 2), où les poids normalisés sont déterminés par la « variabilité » du type de propriété dans le domaine d'application, afin de donner plus d'importance au type de propriété qui apparaît le plus pertinent pour caractériser les concepts localement. L'efficacité de cette approche est limitée à la comparaison de concepts, car la plupart des ontologies sont définies avec très peu de propriété, et ne font pas de distinction entre différents types de propriétés.

[Rodriguez et Egenhofer, 2003] ont proposé une version du même modèle, mais pour comparer des concepts définis par des ontologies distinctes, avec l'objectif d'améliorer la qualité de l'alignement d'ontologies. Les caractéristiques définies par les différentes

² Les “classes d'entité” sont des classes conçues pour faciliter l'accès à des bouts de données qui ont les mêmes particularités (entités) dans une base de données.

ontologies sont comparées à l'aide de fonctions de comparaison de chaînes. La mesure proposée est constituée d'une association entre une mesure de similarité basée sur les caractéristiques avec des méthodes terminologiques et structurelles utilisées traditionnellement pour l'alignement d'ontologies. Comme toutes les mesures de similarité basée sur les caractéristiques, cette approche n'est pas adaptable à la réconciliation des ontologies pour la collaboration, vu que des ontologies développées de manière autonome comportent habituellement une grande proportion de relations sémantiques spécifiques.

[Rada et al, 1989] ont adapté la mesure de similarité de Tversky à la comparaison de concepts dans une hiérarchie de relations « is-a ». Leur hypothèse était que les liens d'hyponymie indiquaient que deux concepts partageaient des caractéristiques communes. En considérant que ces relations sont suffisantes pour calculer une mesure de similarité sémantique, ils ont introduit ce qu'on appelle les mesures basées sur le nombre d'arcs, et qui considèrent le nombre minimum de relations d'hyponymie qui séparent deux concepts comme une bonne estimation de leur similarité sémantique. La mesure proposée a des propriétés d'une distance telle que symétrie et inégalité triangulaire (Equation 3).

[Resnik, 1995] a introduit les mesures de similarité basées sur le contenu en information, qui dépendent d'un corpus annoté afin de dériver une valeur de contenu en information pour les concepts (Equation 4). L'hypothèse de ces approches est que la mesure de similarité entre les termes d'une taxonomie peut être améliorée à l'aide d'un corpus qui permet de déterminer la spécificité des concepts (la mesure de similarité qui sert de base est généralement une mesure basée sur le nombre d'arcs entre les concepts). La notion de contenu en information vient du champ de la théorie de l'information, et son postulat est que les concepts qui sont plus spécifiques (moins présent dans le corpus) sont plus informatifs que les autres. La comparaison des concepts définis dans différentes ontologies par le moyen de mesures basées sur le contenu en information est cependant handicapée par la probable absence d'un corpus annoté commun aux ontologies développées par les organisations qui collaborent.

[Hirst et StOnge, 1998] ont proposé une mesure de connexité qui adapte l'approche de [Rada et al, 1989], afin de considérer tout type de relation sémantique, et non pas seulement les relations d'hyponymie. L'intuition des auteurs est que la connexité des concepts est plus élevée lorsque le chemin le plus court qui les relie est constitué d'un nombre inférieur de relations. Il n'y a cependant aucune raison de croire que deux concepts reliés par les chemins constitués d'au moins deux relations sémantiques soient liés d'une quelconque manière ; la comparaison de concepts définis dans des ontologies différentes utilisant une telle mesure est encore moins justifiable vu que les ontologies partagent ordinairement très peu de relations sémantiques communes.

| | |
|--|---|
| $C_f(c), C_p(c), C_a(c)$ | Ensemble respectifs des parties, fonctions et attributs pour le concept c |
| $lcs(c_1, c_2)$ | Moindre commun super-concept pour les concepts c_1 and c_2 |
| $\alpha(c_1, c_2) = \min_{c \in \{c_1, c_2\}} \frac{len(c, lcs(c_1, c_2))}{len(c_1, c_2)}$ | Poids normalisé pour tenir compte de la différence de profondeur dans la |

| | |
|---------------------------------------|--|
| with $0 \leq \alpha(c_1, c_2) \leq 1$ | hiérarchie <i>is-a</i> des concepts c_1 et c_2 |
| $len(c_1, c_2)$ | Nombre minimum d'arcs <i>is-a</i> qui séparent les concepts |
| $p(c)$ | Probabilité de rencontrer une instance du concept c dans le corpus associé à l'ontologie |

$$S_x(c_1, c_2) = \frac{|C_x(c_1) \cap C_x(c_2)|}{|C_x(c_1) \cap C_x(c_2)| + \alpha(c_1, c_2)|C_x(c_1) - C_x(c_2)| + (1 - \alpha(c_1, c_2))|C_x(c_2) - C_x(c_1)|} \quad (1)$$

$$Sim_{RE}(c_1, c_2) = \omega_{f,AD} \cdot S_f(c_1, c_2) + \omega_{p,AD} \cdot S_p(c_1, c_2) + \omega_{a,AD} \cdot S_a(c_1, c_2) \quad (2)$$

$$sim_{Rada}(c_1, c_2) = len(c_1, c_2) \quad (3)$$

$$sim_{Resnik}(c_1, c_2) = \max_{c \in S(c_1, c_2)} (-\log(p(c))) \quad (4)$$

2.3 Comparaison de concepts basés sur le contexte

La perception humaine de similarité est largement influencée par le contexte. Selon les expérimentations de [Goldstone et al, 1997], l'importance des caractéristiques varie avec le contexte et n'est pas fixée jusqu'au moment même de la prise de décision. Le contexte est donc nécessaire pour décider, parmi les divers critères qui permettent de comparer des concepts (fondés sur les attributs, les relations sémantiques, etc.), ceux qui sont les plus importants.

[Rodriguez et Egenhofer, 1999], suivi par [Rodriguez et Egenhofer, 2004], ont proposé de considérer la tâche de l'utilisateur pour améliorer le calcul de leur mesures de similarité MDSM basée sur les caractéristiques. Au moment de la requête, les utilisateurs sélectionnent l'action qu'ils ont l'intention d'exécuter parmi une liste d'opérations. L'ensemble des concepts possibles est constitué par les seuls concepts qui comprennent parmi leurs fonctions l'opération demandée par l'utilisateur. La mesure a les mêmes désavantages déjà cités pour les mesures des mêmes auteurs décrites dans la section précédente.

La tâche de l'utilisateur est seulement une partie du contexte de la réconciliation d'ontologies. Pour gagner en exactitude lorsqu'on évalue comment les concepts sont liés les uns aux autres, il est important de déterminer quelle information est effectivement contextuelle. Nous allons introduire dans la prochaine section notre méthodologie pour déterminer quelle information est contextuelle dans notre cas, et comment l'utiliser d'une manière appropriée lorsqu'on évalue la relation entre concepts.

PARTIE III – Méthodologie pour une solution basée sur le contexte

Chapitre 3 : Méthodologie pour une contextualisation efficace

Dans ce chapitre nous introduisons une méthodologie qui permette de développer une solution efficace basée sur le contexte. Dans ce but, nous proposons d'examiner quelle est l'application et quels sont les objectifs attendus de la contextualisation. Ces objectifs vont servir de guides pour juger comment caractériser les différents contextes adoptés. Les modèles et mesures développés pourront ainsi être combinés en une architecture globale que nous pourrions évaluer selon les objectifs fixés.

La méthodologie comporte trois phases, chacune d'elles étant composée d'un certain nombre d'étapes. La première phase consiste à déterminer les directives qui vont orienter l'application du reste de la méthodologie. La deuxième phase est effectuée autant de fois qu'il y a d'usages du contexte qui seront implantés pour réaliser les objectifs fixés : on cherchera à déterminer quelle information est contextuelle, est à la caractériser, la modéliser et l'utiliser. La troisième et dernière phase consiste à assembler les modèles correspondant aux divers usages du contexte pour former une architecture globale qui satisfasse les objectifs fixés lors de la première phase.

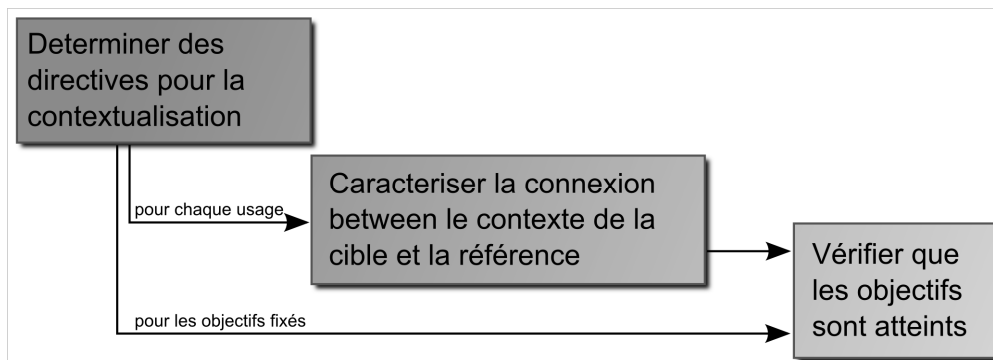


Figure 4 : Illustration de la méthodologie

Pour faire apparaître les points de la méthodologie de manière plus évidente, et ainsi faciliter son application, nous allons indiquer chaque point à l'aide du symbole ➤.

11.1 Phase 1 : Directives pour contextualiser l'application

La première phase est composée de trois étapes (voir Figure 5) : fixer les objectifs que l'approche basée sur le contexte doit observer, sélectionner les usages du contexte qui seront implémentés pour atteindre ces objectifs et instancier ces usages pour l'application, à l'aide du triplet (cible, référence, but de la comparaison).

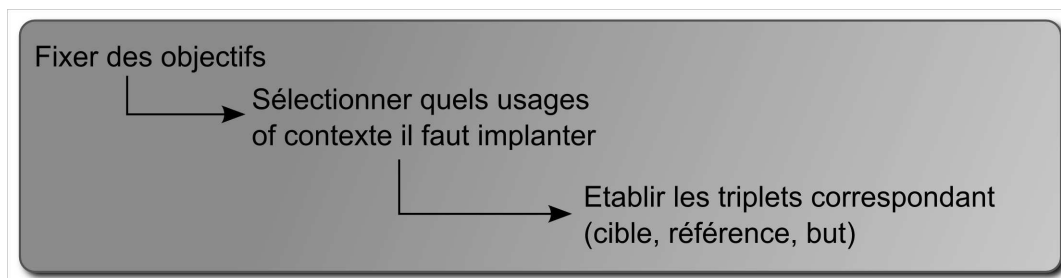


Figure 5 : Phase 1 de la méthodologie

① Fixer des objectifs

Une solution base sur le contexte sera jugée par l'effet qu'elle aura sur l'application, comme toute autre type de solution. Il est donc essentiel de ne pas perdre de vue l'application. On s'attend à ce que contextualiser l'application aura un effet positif : on attend de l'utilisation du contexte qu'elle permette d'ôter des déficiences de l'application, de l'améliorer d'une manière ou d'une autre, et de préserver les atouts qui lui sont inhérents.

Cette phase inclut donc les points suivants :

- Déterminer l'application à contextualiser
- Fixer les objectifs pour la contextualisation :
 - Quels sont les carences que l'on veut ôter ?
 - Quels sont les améliorations qu'on espère implanter ?
 - Quels sont les atouts que l'on veut conserver ?

② Sélectionner les usages du contexte à implanter

Nous proposons dans le Tableau 1 une catégorisation de différents usages du contexte, fondée sur la revue de la littérature sur le contexte : les approches dont la contribution comprend l'utilisation du contexte, se focalisent souvent sur la collecte, la modélisation ou l'utilisation du contexte.

Que ce soit pour éliminer une carence ou pour introduire une amélioration, cela peut se faire par l'implantation de différents usages de l'information contextuelle. Nous proposons de déterminer, pour chaque objectif, les usages du contexte qui seront à implanter, afin de guider la caractérisation ultérieure du contexte. Il s'agit donc de

- Déterminer les usages du contexte à implanter afin d'atteindre les objectifs fixes.

Tableau 1 : Principaux usages du contexte

| Ce qui est fait du contexte | Usages principaux du contexte |
|---|---|
| Collecter le contexte. | <p><i>Transmettre</i> à l'utilisateur : L'historique des opérations peut être tracé pendant une période déterminée. Quand on s'intéresse à un objet particulier, les opérations qui ont précédé sa « découverte » sont supposés la caractériser. Ces opérations font partie du contexte de cet objet, et peuvent permettre d'aider l'utilisateur à se rappeler les circonstances de la « découverte » de l'objet.</p> <p><i>Pourvoir</i> aux besoins d'autres applications : Des données qui décrivent l'objet selon des spécifications connues peuvent être utilisées par des applications, selon leurs besoins particuliers.</p> |
| Modéliser le contexte. | <p><i>Raisonnement</i> avec des ensembles de règles incohérents portant sur de même objets : différentes situations où les règles diffèrent peuvent être décrites de manière logique, de telle sorte que le raisonnement soit adapté à la situation.</p> |
| Employer le contexte (dans ce but, de telles approches pourront souvent également collecter et modéliser le contexte) | <p><i>Désambiguïser</i> entre différents sens : Quand l'information directement associée à un objet est insuffisante pour déterminer la procédure à suivre pour la suite, l'information contextuelle peut permettre de retrouver quel est le choix qui correspond le mieux à la nature de l'objet.</p> <p><i>Personnaliser</i> pour un individu ou un groupe : Le contexte associé à un individu peut permettre de ne pas tenir compte de résultats qui n'ont pas d'intérêt pour lui, ou mettre en avant ce qui est supposé être davantage pertinent.</p> <p><i>Adapter</i> pour un appareil particulier : Le contexte associé à un appareil peut être utilisé pour fournir des données à l'appareil dans un format qui lui soit adapté.</p> <p><i>Evaluer</i> la pertinence d'un objet pour une tâche donnée : Le contexte associé à l'objet peut permettre de mesurer dans quelle mesure l'objet est adéquat pour une tâche donnée.</p> <p><i>Chercher</i> un objet à l'aide d'informations connues sur l'interaction de l'utilisateur avec cet objet. L'information concernant un objet peut être retrouvée en recherchant ce qui est connu de l'interaction de l'utilisateur avec l'objet : par exemple, il peut s'agir de la date de modification de l'objet ou de son dernier accès, etc.</p> |

③ Etablir les triplets correspondant (cible, référence, but)

Nous proposons de déterminer les entités dont le contexte nous intéresse. Ceci doit être fait pour chaque usage reconnu lors de l'étape précédente.

Le contexte décrit la connexion d'une entité avec les autres. Il s'agit d'une information qui peut être considérée comme périphérique à l'entité qui se situe dans le focus. Une

telle information contextuelle est requise lorsque la connaissance intrinsèque de l'entité dans le focus n'est pas suffisante, ou pas récupérable de manière satisfaisante. Pour déterminer ce qui est contextuel, il est donc nécessaire de déterminer au préalable pour quelles entités ce contexte est nécessaire.

Quand le but est d'*employer* le contexte pour un usage donné, le contexte ne sera pas utilisé seul, mais il doit être comparé avec une autre information [Jouanot et al, 2003]. Nous définissons ainsi les éléments essentiels de cette comparaison :

- « *cible* » : l'entité dans le focus, pour laquelle le contexte est demandé
- « *référence* » : l'information à laquelle le contexte de la cible est comparé
- « *but* » : la raison pour laquelle on veut comparer le contexte de la cible avec la référence.

Le contexte doit révéler quelque chose à propos de la cible qui soit pertinent pour comparer la cible et la référence, et qui réponde au but établi pour la comparaison. Nous proposons en Tableau 2 une catégorisation de plusieurs usages du contexte, avec les cibles, références, et buts correspondant. Les points à suivre pour cette étape :

- Déterminer la cible, la référence, et le but de la comparaison

La méthodologie que nous proposons dans ce chapitre touche l'utilisation du contexte pour une application particulière ; nous limiterons notre méthodologie au type de contexte « connexion ».

Dans cette phase, nous avons montré comment l'application et les objectifs pour la contextualisation étaient utilisés pour déterminer les usages du contexte à implanter. Pour chacun de ces usages du contexte, nous proposons de déterminer le triplet (cible, référence, but) correspondant. Dans la prochaine phase, nous allons étudier comment caractériser chacun de ces triplets.

Tableau 2 : Usages du contexte et les cibles, references, et but correspondant

| Usage | Événement qui déclenche le besoin de contexte | Cible | Référence | But de la comparaison | Objectif | Types de contexte possible |
|----------------------|--|--|--|--|--|-----------------------------------|
| <i>Désambiguïser</i> | Situation où il est nécessaire de clarifier le sens associé à une entité | L'entité à désambiguïser | Les variantes possibles | Sélectionner la variante correcte | Utiliser l'entité de manière approprié à son sens | Linguistique, contexte |
| <i>Personnaliser</i> | L'interaction de l'utilisateur avec une application | L'utilisateur | Les options possibles | Sélectionner l'option la plus appropriée | Personnaliser l'interaction avec l'utilisateur | Connexion, contexte |
| <i>Adapter</i> | L'utilisation d'un appareil (à un instant et lieu donné) | L'environnement présent : heure, localisation (capteurs) | Liste de situations décrites à l'aide d'un ensemble de conditions | Sélectionner la situation qui correspond aux données captées | Changer le comportement de l'appareil lorsque la situation est reconnue | Situationnel |
| | Interaction avec un appareil pour un service particulier | L'appareil | Ensemble de configurations considérées | Sélectionner la configuration la plus appropriée | Utiliser l'appareil avec la configuration la plus adaptée | Connexion |
| <i>Évaluer</i> | Situation de choix entre diverses options | Une option à évaluer | Les critères et conditions requises | Noter les options selon ce qui sied, au vu des critères et conditions requises | Agir selon les résultats obtenus, ou transmettre cette information à l'utilisateur | Contexte |
| <i>Retrouver</i> | Besoin en information où les circonstances de l'interaction avec l'entité sont connues | Entité à chercher | Liste d'entités, et les informations par rapport à l'interaction de l'utilisateur avec elles | Retrouver l'entité décrite | Agir selon les résultats obtenus, ou transmettre cette information à l'utilisateur | Connexion |

11.2 Phase 2 : Connexion entre le contexte de la cible et la référence

La seconde phase doit être mise en œuvre pour chacun des triplets découverts lors de la première phase. Elle consiste à caractériser la connexion entre le contexte de la cible et la référence (Figure 6). Les étapes sont représentées dans la figure dans un ordre logique, mais la caractérisation du contexte va probablement demander des allers et retours, de manière similaire au modèle de développement en spirale. La phase 2 est composée de six étapes, incluant la caractérisation du contexte de la cible, la modélisation de la connexion, et les détails techniques concernant la récupération et la mesure des données.

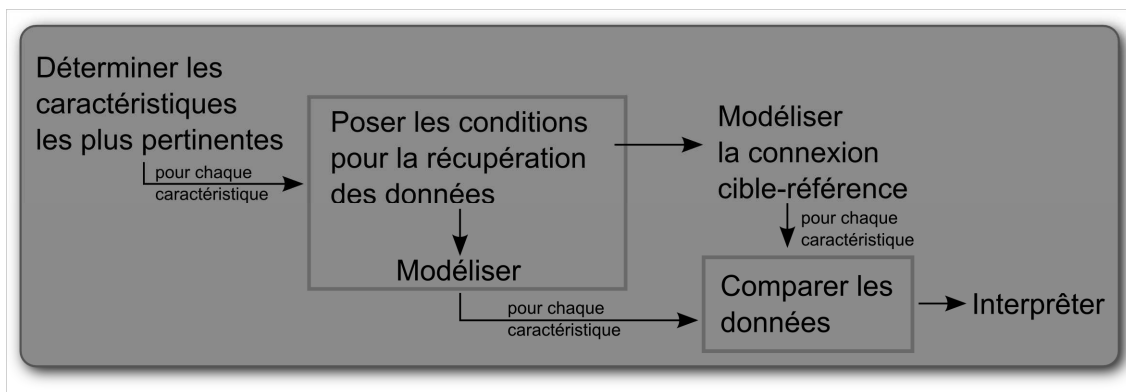


Figure 6 : Deuxième phase de la méthodologie

① Caractéristiques pertinentes

La première chose à faire est de rechercher les questions pertinentes concernant la cible, en s'aidant des questions « Qui, quoi, où, quand, comment, pourquoi (QOOQCP) ». D'habitude, seulement une partie de ces questions sont pertinentes lors de la comparaison avec la référence. La pertinence est jugée selon le but de la comparaison.

Nous définissons le terme suivant :

- « caractéristique » : tout trait particulier de la cible qui est jugée pertinent pour comparer la cible avec la référence, et peut être utilisée dans le cadre de cette comparaison.

Toutes les caractéristiques n'ont pas la même importance pour caractériser le contexte de la cible ; et il est nécessaire de sélectionner les caractéristiques les plus pertinentes, car le traitement du contexte est coûteux, et requiert d'être déterminé, récupéré, etc.

Il n'est pas nécessairement évident de déterminer les caractéristiques qui sont les plus pertinentes pour décrire le contexte de la cible, car ces caractéristiques sont souvent interconnectées. La recherche de caractéristiques peut être comparée à l'analyse de causes, et on peut se servir d'outils tels de diagramme de Pareto³. Il faut donc

- Déterminer les caractéristiques les plus pertinentes pour comparer cible et référence.

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

② Conditions pour récupérer les données

Les données contextuelles sont rarement disponibles directement. Pour employer le contexte, il est donc nécessaire de commencer par récupérer ces données. Cela peut être fait de manière automatique, ou manuellement quand les données sont difficiles à récupérer, ou n'existent pas sous forme numérique.

Selon les sources, la quantité et la qualité des données varie. Les sources ne sont pas toujours disponibles. Ces critères, et d'autres encore, peuvent être utilisés pour déterminer les sources de données les plus pertinentes.

Pour récupérer les données, il est nécessaire d'ajuster la précision de la méthode utilisée, de telle sorte que les données aient un grain approprié. Une mesure trop précise résultera en une série fluctuante de valeurs, au lieu d'une valeur unique. Si ces fluctuations ont un impact sur la comparaison de la cible avec la référence, elles sont contextuelles. Sinon, la précision doit être réduite, ou alors il faut calculer une valeur moyenne.

Les données ne sont pas toujours valides, mais la validité des données peut dépendre de conditions particulières, qui sont spécifiques à chaque type de données. Par exemple, lorsqu'on conduit des interviews, les questionnaires qui suggèrent qu'une réponse est préférée ne permettent pas d'obtenir des résultats fiables. Certaines mesures doivent être faites à certaines heures de la journée, parce que les conditions (par exemple météorologique, biologique, social) sont plus appropriées.

Cette étape inclut les points suivants :

- Déterminer les sources de données contextuelles disponibles les plus pertinentes
- Établir une précision et une échelle satisfaisante
- Fixer les conditions qui permettent d'obtenir des données valides

③ Modéliser la caractéristique

Les données contextuelles peuvent être utilisées directement, mais il peut sembler approprié dans la plupart des cas de structurer ces données dans un modèle, afin de simplifier l'utilisation ultérieure de ces données. Ce modèle joue alors le rôle d'une interface pour des applications logicielles externes : il permet de modifier les sources de données contextuelle, ou d'utiliser les données contextuelles pour d'autres applications encore, en limitant les modifications à faire à ce qui a déjà été développé.

La structure du modèle, la technologie utilisée pour l'implanter, etc. sont dépendant de la situation dans laquelle la contextualisation est effectuée. Nous proposons donc de commencer par spécifier les critères et conditions requises pour un tel modèle. Cette étape inclut donc les points suivants :

- Déterminer les critères et conditions requises pour représenter les caractéristiques d'une manière appropriée et efficace.
- Modéliser le contexte d'une manière conforme à ces critères et conditions

④ Modéliser la connexion

La connexion entre le contexte de la cible et la référence doit être connue pour les comparer avec succès. Nous proposons donc de modéliser la manière de mettre en relation les données présentes dans le contexte de la cible avec les données de la référence. Le modèle inclut

notamment une méthode pour comparer les différents types d'information, afin d'atteindre le but fixé pour la comparaison. Pour cette étape, il faut donc :

- Modéliser la connexion entre le contexte de la cible et la référence.
- Etablir une méthode pour comparer les caractéristiques de la cible avec l'information contenue dans la référence.

⑤ Evaluation des données pour la connexion

Les données associées à une caractéristique ont probablement été récupérées de sources diverses, et peuvent être de différente nature. Alors qu'on compare le contexte de la cible avec la référence, les données récupérées doivent être évaluées d'une manière appropriée pour chaque type de données. Ces mesures doivent être comparées avec l'information correspondante dans la référence.

Là où il y a plus d'une caractéristique, il semble également désirable de calculer une seule mesure par caractéristique, lorsque les données qui la décrivent sont multiples ou proviennent de plus d'une source. Il est préférable que ces mesures soient toutes normalisées quand c'est possible, afin de pouvoir les pondérer (éventuellement de manière dynamique) pour avoir une mesure spécifique à chaque caractéristique qui soit normalisée.

La norme utilisée pour ce faire peut être une somme pondérée, une somme statistique, etc. La valeur qui en résulte doit pouvoir constituer un indicateur significatif de la concordance entre les données contextuelles spécifiques à la caractéristique, et l'information correspondante dans la référence. L'étape comprend donc les points suivants :

- Déterminer des mesures normalisées pour évaluer chaque type de données contextuelles
- Attribuer des poids pour pondérer les différentes mesures selon l'impact de leur variabilité sur la caractéristique
- Etablir une norme pour obtenir une mesure unique pour chaque caractéristique

⑥ Interpretation des données

Quand le but de la comparaison entre le contexte de la cible et la référence est atteint, il faut généralement transformer les données pour accomplir le but de l'usage du contexte correspondant. Il faut donc développer une méthode appropriée, afin de transformer l'évaluation des mesures correspondant à chaque caractéristique.

Cette méthode doit donc « interpréter » les différentes évaluations effectuées et transposer le résultat en une action appropriée (par exemple, le déclenchement d'une alarme sous certaines conditions ou une décision), ou en une présentation parlante pour l'utilisateur (par exemple une valeur dans une échelle appropriée, un graphe), en une appréciation qualitative, ou autre. Cette étape inclut donc les points suivants :

- Déterminer une méthode pour transformer les résultats des mesures spécifiques à chaque caractéristique, pour atteindre l'usage recherché.

Nous avons décrit comment caractériser le contexte de la cible, modéliser la connexion entre cible et référence, et établir les mesures pour traiter les données correspondantes. Dans la phase suivante, nous allons assembler les différents modèles réalisés pour chacun des triplets (cible, référence, but) pour atteindre les objectifs fixés pour la contextualisation.

11.3 Phase 3 : construire une architecture qui atteint les objectifs fixés

La dernière phase consiste à rassembler les modèles et les mesures qui ont été mis en place lors de l'application de la seconde phase. Le but est donc de finaliser et de valiser une solution basée sur le contexte. Cette phase est constituée de trois étapes, incluant la construction de l'architecture globale, son évaluation à l'aide des objectifs fixés lors de la première phase, et la complétion du système avec les méthodes et modèles appropriés (Figure 7).

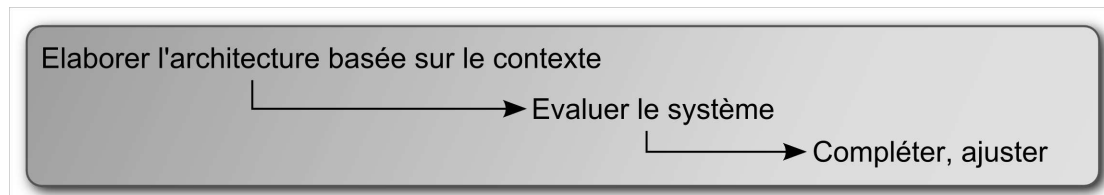


Figure 7 : Troisième phase de la méthodologie

① Architecture

Il est temps maintenant de développer une architecture qui combine tous les modèles et mesures en un système basé sur le contexte qui permette d'atteindre les objectifs fixés. Cette étape inclut le point suivant :

- Construire une architecture qui combine les modèles correspondant aux différents usages sélectionnés lors de la première phase

② Evaluation

Lors de l'application de la première phase de la méthodologie, des objectifs ont été déterminés. La solution basée sur le contexte doit répondre à ces objectifs. Nous allons donc évaluer la solution selon les objectifs fixés :

- Évaluer la solution basée sur le contexte selon les objectifs établis lors de la première phase.

③ Complétion et ajustement

A ce point, il peut apparaître que l'architecture ne répond pas à tous les objectifs fixés pendant l'application de la première phase de la méthodologie. Cela peut provenir d'objectifs trop ambitieux, de l'indisponibilité des données, ou de l'omission d'un usage du contexte.

Cela peut aussi provenir du fait que la solution basée sur le contexte est avant tout un système : se baser sur le contexte ne veut pas dire que seules des informations contextuelles doivent être utilisées pour atteindre les objectifs. Des méthodes, modèles, mesures, ... plus traditionnels peuvent être ajoutés au système, avec succès. Pour cette étape, il s'agit donc de :

Etablir des méthodes, modèles et mesures qui puissent contribuer à ce que la solution basée sur le contexte puisse atteindre les objectifs fixés.

11.4 Conclusion

Nous avons élaboré une méthodologie pour développer un système basé sur le contexte. Nous nous sommes servis pour cela de la situation particulière dans laquelle l'application doit être contextualisée. L'application et les objectifs pour cette solution aident à déterminer les différents usages du contexte qui sont désirables pour une solution basée sur le contexte.

Pour chacun de ces usages, on aura besoin de caractériser le contexte de la cible, et de déterminer comment les données récupérées pour décrire ce contexte pourront être comparées avec l'information de référence, de telle sorte que le but fixé pour la comparaison soit atteint. On aura ensuite à interpréter les mesures afin d'accomplir l'usage du contexte correspondant.

Finalement, les modèles et mesures développées doivent être assemblées dans une architecture globale, qui sera jugée selon les objectifs fixés précédemment, et ajustée selon ce qui est jugé nécessaire.

Dans les cinq chapitres qui vont suivre, nous allons appliquer cette méthodologie pour améliorer la réconciliation des ontologies, en prenant en considération les implications pratiques. Nous allons suivre la méthodologie élaborée dans ce chapitre pour déterminer les caractéristiques les plus pertinentes, les conditions pour récupérer les données associées, et construire un modèle du contexte qui soit approprié pour la réconciliation d'ontologies.

PARTIE IV – Application de la méthodologie à la réconciliation des ontologies

Chapitre 4 : Directives pour la contextualisation de l'alignement d'ontologies

Dans ce chapitre nous allons appliquer la première phase de la méthodologie (Figure 8) que nous venons d'élaborer. Ceci nous permettra de déterminer une approche basée sur le contexte qui permette d'améliorer la qualité de la réconciliation des ontologies, basée sur des mappings existants entre des concepts définis dans diverses ontologies.

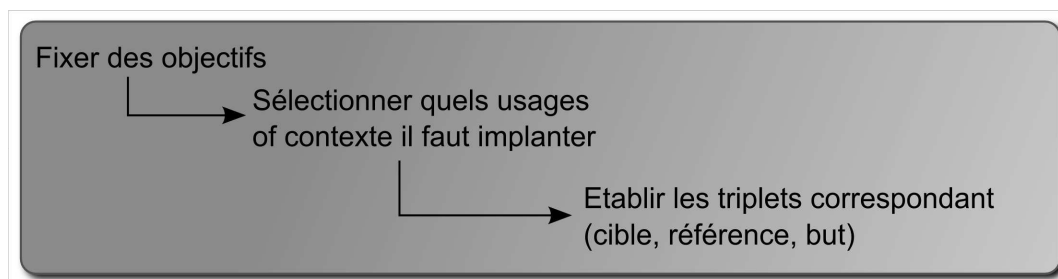


Figure 8 : Application de la méthodologie (phase 1)

4.1 Objectifs

Dans cette étape, nous allons tout d'abord préciser notre application, et fixer les objectifs que nous avons pour la contextualisation de cette application.

Notre application est la réconciliation d'ontologies développées de manière autonome, et évoluant de manière indépendante, dans le cadre de collaborations non anticipées entre des organisations diverses. Nos objectifs pour cette étude ont été introduits dans la première partie de cette étude, en particulier dans l'énoncé de la problématique :

- Les carences que nous avons signalées dans la problématique peuvent être résumées en : « la réconciliation d'ontologies indépendantes par l'alignement ne permet pas de lier les entités des ontologies d'une manière qui soit pertinente au regard des implications pratiques ». Nous espérons réduire ce problème en nous appuyant sur le contexte pour évaluer la connexion entre les concepts. Nous croyons que c'est seulement par la considération des informations contextuelles que l'on peut découvrir les implications pratiques associées originellement au concept [Porzel et al, 2006].
- Les améliorations que nous voulons effectuer sont :
 - Étendre l'utilisabilité de la réconciliation.
Quand les ontologies sont alignées, seuls quelques concepts sont reliés les uns aux autres par des mappings. Nous voulons proposer une mesure d'évaluation entre des concepts au choix.
 - Évaluer la réconciliation au niveau pragmatique.
Les mappings sont la plupart du temps limités à exprimer la similarité ou l'équivalence de deux concepts appartenant à des ontologies alignées. Cependant, comme la collaboration entraîne l'échange de données, il est important de connaître avec quelle pertinence un concept peut « remplacer » un autre concept (ce qui devrait être le cas si les concepts sont réellement « équivalents »). Nous voulons proposer une mesure d'évaluation de la pertinence d'un concept défini dans une ontologie d'une organisation

collaboratrice pour « remplacer » un concept défini dans une ontologie de la compagnie, et qui soit basée sur le contexte.

- L'atout que nous voulons préserver est la flexibilité des approches fédératives. En effet l'approche fédérative utilise des mappings pour réconcilier les ontologies, qui ne sont donc pas modifiées. C'est ce qui fait que ces approches sont adaptées à la réconciliation d'ontologies qui évoluent de manière indépendante. De la même manière nous voulons que notre approche ne modifie pas les ontologies alignées.

4.2 Les usages du contexte

Différents usages du contexte peuvent servir nos objectifs pour cette étude. Nous voulons :

- Désambiguïser les significations pragmatiques des concepts.
Nous voulons que la mesure permette d'évaluer si les concepts comparés sont associés avec des données qui sont comparables en termes d'utilité pratique.
- Personnaliser le résultat.
Chaque agent peut avoir une autre compréhension des concepts des ontologies de la compagnie. Connaître l'identité de l'agent qui émet la requête est donc nécessaire pour évaluer de manière correcte la pertinence des concepts définis dans des ontologies des organisations collaboratrices.
- Évaluer la pertinence d'un concept pour la tâche d'interopérabilité.
Quand une requête est soumise pour l'évaluation de deux concepts, c'est dans la circonstance d'un besoin en interopérabilité. Comme l'interopérabilité est orientée tâche, la mesure de la pertinence d'un concept (défini par une ontologie de l'organisation collaboratrice) doit dépendre de la tâche d'interopérabilité qui a déclenché le besoin en information.

Nous définissons dans le Figure 9 une liste de terme qui va simplifier la description de notre approche et que nous allons utiliser à partir de maintenant.

| Terme | Définition |
|------------------|---|
| Demandeur | l'agent qui soumet la requête pour l'interopérabilité |
| Agent | utilisateur ou composant logiciel qui réalise des tâches spécifiées de manière autonome. |
| Compagnie | l'organisation dont le demandeur est membre |
| Partenaire | l'organisation collaboratrice |
| Concept origine | le concept sélectionné dans une ontologie de la compagnie qui illustre le type du concept que le demandeur recherche. |
| Sens pragmatique | Implications pratiques d'un concept en termes de données qui lui sont effectivement associées. |

Figure 9 : Définitions de termes pour notre approche

| Concept | Livre | Livre scolaire | Livre |
|-----------------------------|-------------------------------------|---|--------------------------------|
| Organisation | Livres sur demande | Editions Nathan | Amazon.com |
| Site internet | www.bod.fr | www.nathan.fr | www.amazon.com |
| Données (pour chaque livre) | Maquette digitale, ISBN, code barre | Titre, niveau, obligatoire ou optionnel | Titre, auteur, avis de lecteur |
| | Liste de librairies où le | Année de publication, | Liste de prix, prix, en stock |

| | | | |
|---------------------|-------------------|-------------------------------|-------------------|
| | livre est vendu | éditeurs, auteurs, collection | |
| Point de vue | Édition de livres | Vendeur de livres spécialisé | Vendeur de livres |

4.3 Triplets (cible, référence, intention)

Pour chacun des trois usages reconnus, nous avons en différents triplets (cible, référence, intention) :

☞ Désambiguïser

Nous voulons désambiguïser le sens pragmatique du concept enquis en le comparant avec le sens pragmatique du concept origine. Pour cela nous devons comparer le contexte des concepts des deux organisations :

- Cible : concept enquis
- Référence : contexte du concept origine
- Intention : déterminer si le concept enquis est pertinent pour « remplacer » le concept origine

Les collaborations requièrent d'ordinaire que les deux organisations aient accès aux données de l'une et de l'autre. Comme le processus de désambiguïisation doit comparer les sens pragmatiques possibles des deux concepts, il semble approprié de ne pas faire de distinction entre le concept enquis et le concept origine pour le modèle du contexte, mais de construire un modèle symétrique.

☞ Personnaliser

Nous voulons évaluer si le concept enquis peut être utilisé pour répondre au besoin en information exprimé par le demandeur. Ce besoin en information est décrit à l'aide du concept origine, sur lequel le demandeur a un point de vue particulier. Pour avoir accès au sens pragmatique que le demandeur associe au concept, il nous faut connaître le contexte du demandeur. Nous avons donc à comparer le contexte du demandeur avec le contexte du concept origine :

- Cible : concept origine
- Référence : contexte du demandeur
- Intention : sélectionner parmi les sens pragmatiques possibles du concept origine celui que le demandeur peut avoir effectivement associé au concept enquis.

Le contexte du concept origine est décrit ici de la même manière que le contexte des concepts pour l'usage de la désambiguïisation décrit dans le paragraphe précédent. Pour connecter les différents usages, il semble approprié d'utiliser le même modèle du contexte que pour le concept origine. Comme le contexte du demandeur doit également être déterminé, nous allons poursuivre la méthodologie en invertissant cible et référence :

- Cible : un demandeur
- Référence : contexte du concept origine

☞ Évaluer

Nous voulons déterminer si le concept enquis peut être une réponse satisfaisante au besoin en interopérabilité. Il nous faut donc comparer le contexte du concept enquis avec le contexte du besoin en interopérabilité :

- Cible : concept enquis
- Référence : contexte du besoin en interopérabilité

- Intention : déterminer dans quelle mesure les données associées avec le concept enquis peuvent être pertinentes pour répondre au besoin en interopérabilité qui a été à l'origine de la requête.

Pour la même raison que dans le paragraphe précédent nous allons intervertir et les références :

- Cible : un besoin en interopérabilité
- Référence : contexte du concept enquis.

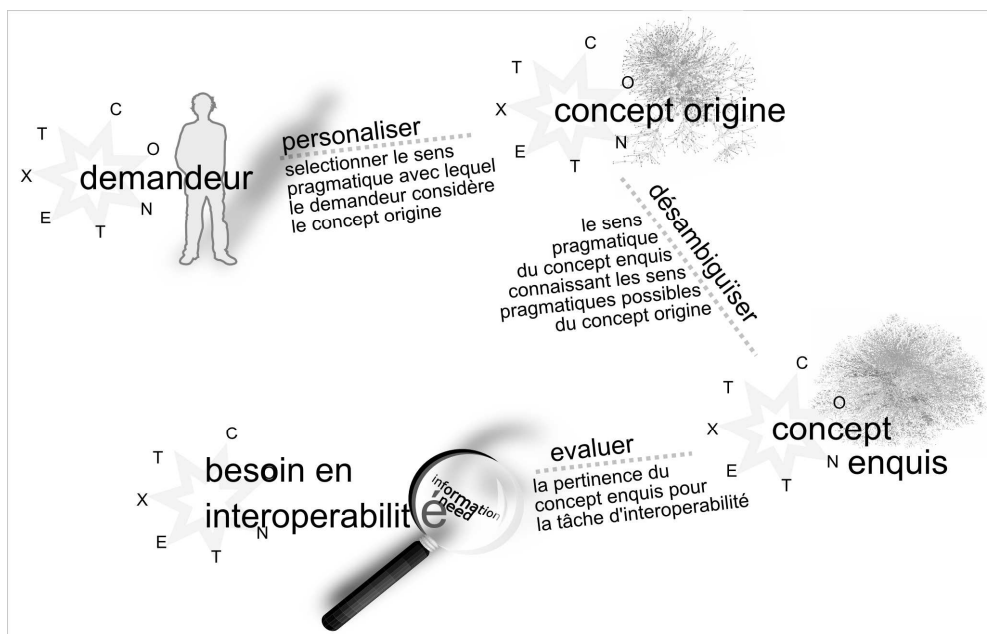


Figure 10 : Les usages du contexte sélectionnés

4.4 Conclusion

Nous avons déterminé dans ce chapitre l'application que nous voulons contextualiser, les objectifs de cette contextualisation, et les différents usages du contexte que nous allons employer pour atteindre ces objectifs. Nous voulons améliorer la réconciliation d'ontologies indépendantes effectuée par une approche fédérative en proposant une évaluation de la pertinence d'un concept défini par une ontologie d'une organisation partenaire.

Pour atteindre ces objectifs, nous avons déterminé pour chacun de ces usages la cible, référence, et l'intention. Nous avons trois différents contextes à considérer : le contexte des concepts, le contexte du demandeur, et le contexte du besoin en interopérabilité (Figure 10).

Maintenant il nous faut déterminer quelle information concernant ces entités est contextuelle, disponible, et peut être utilisée pour les comparaisons entre cible et référence, de telle manière à atteindre les objectifs fixés. Nous allons donc étudier, dans les trois chapitres suivants, comment mettre en relation le contexte du concept origine avec le contexte du concept enquis pour la désambigüisation, comment relier le contexte du concept origine avec le contexte de l'application pour la personnalisation, et comment relier le contexte du concept enquis avec le contexte du besoin en interopérabilité pour l'évaluation.

Chapitre 5 : Connexion entre le contexte du concept origine et celui du concept enquis

Dans ce chapitre nous allons appliquer la seconde phase de la méthodologie décrite dans le chapitre 3 afin de caractériser la connexion entre le contexte du concept enquis et celui du concept origine pour l'usage de désambiguïsation (Figure 11). Le but est d'améliorer la connexion entre les ontologies réconciliées en considérant leurs implications pratiques. Nous allons désambiguïser le sens pragmatique du concept enquis en le comparant avec les sens pragmatiques possibles du concept origine.

Comme le concept origine et le concept enquis jouent des rôles symétriques dans le processus de désambiguïsation, nous déterminons le contexte des concepts sans faire de distinction entre les deux rôles.

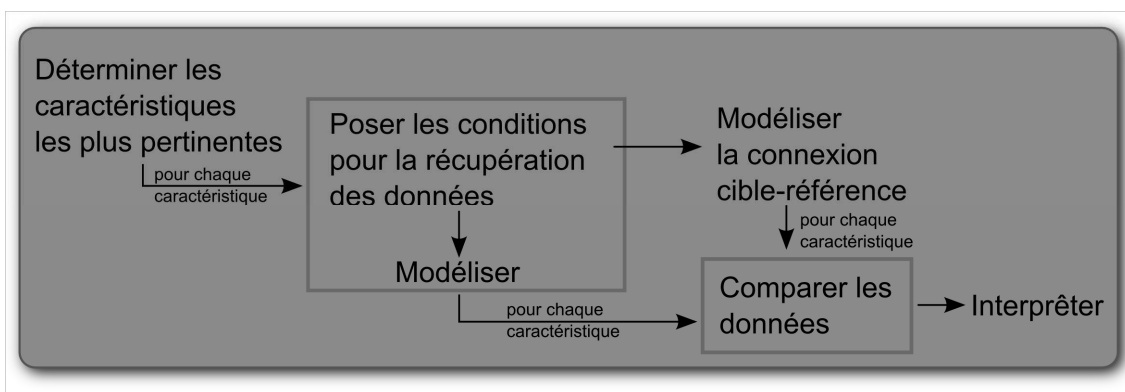


Figure 11 : Application de la méthodologie (deuxième phase, premier usage du contexte)

5.1 Caractéristiques pertinentes pour décrire le contexte des concepts

Dans cette étape, nous déterminons les caractéristiques qui sont les plus pertinentes pour décrire les concepts dans le cadre de la réconciliation d'ontologies.

[Ehrig et al, 2004] considèrent que le contexte d'utilisation des entités des ontologies réconciliées est le contexte le plus pertinent pour évaluer la réelle signification des entités. Le contexte d'utilisation pourrait en effet être utilisé pour désambiguïser le sens des entités dans les ontologies, mais il s'agit aussi très souvent d'une information confidentielle, difficilement disponible, et implémentée de manière hétérogène. Il semble donc impossible à nos yeux de comparer des contextes des ontologies les uns avec les autres sur la base d'une information si peu prédictible.

[Paslaru-Bontas, 2007] propose d'utiliser le contexte du développement de l'ontologie en tant qu'artefact afin d'évaluer si l'ontologie peut être réutilisée dans le cadre d'une application donnée. Dans notre cas, la question n'est pas de décider si une ontologie doit être gardée ou rejetée ; au contraire, la désambiguïsation est nécessaire au niveau des entités de l'ontologie. En raison de cela, nous ne sommes pas concernés par la plupart des critères qu'elle a établis.

Nous proposons de suivre néanmoins les traces de [Paslaru-Bontas, 2007] et de garder comme caractéristique le contexte du développement de l'ontologie. Le développement de l'ontologie

peut être caractérisé par la méthodologie employée pour la développer, les outils utilisés pour l'éditer et pour opérer la fusion et l'alignement, enfin les ingénieurs qui ont développé l'ontologie et les perspectives avec lesquels l'ontologie a été développée. Parmi toutes ces caractéristiques possibles, nous croyons que ces dernières sont les informations contextuelles qui permettent de décrire au mieux le contenu de l'ontologie.

Nous considérons donc que le contexte de l'ontologie est composé des perspectives avec lesquelles les experts ont développé et affiné l'ontologie. Dans notre compréhension de la notion de perspective, nous incluons aussi les données associées au concept : certaines perspectives peuvent marquer une différence dans le type de données associées au concept. Nous voulons détecter si le concept origine et le concept enquis ont été développés avec des perspectives qui sont compatibles.

Nous définissons le terme « *perspectives* » par « *les manières individuelles de regarder une situation, par exemple influencées par ses expériences ou considérations personnelles* ».

5.2 Source des données et conditions pour les récupérer

Dans cette étape nous évaluons comment obtenir des données relatives au contexte du développement de l'ontologie, la précision nécessaire, et les conditions à suivre pour s'assurer de leur validité.

Les sources de données les plus pertinentes pour caractériser les perspectives sont certainement les applications, les bases de données, et d'autres sources de données que les ontologies ont servi à intégrer, ainsi que les ontologies elle-même. Il semble impossible d'élaborer une méthode pour récupérer des données de telles ressources, vu que leur pertinence respective et leurs caractéristiques varient selon les conditions dans lesquelles est opéré le développement des ontologies. Les seules ressources disponibles pour lesquelles nous pouvons élaborer une méthode sont les ontologies elles-mêmes.

Les perspectives qui influencent le développement de l'ontologie ne restreignent pas les ingénieurs de telle manière qu'il y aurait une seule structure et un seul contenu possible. Au contraire, la construction de l'ontologie est subjective, et l'impact perceptible des perspectives dans l'ontologie n'est pas évident mais doit être analysé. L'information associée à une perspective peut être omniprésente dans l'ontologie, limitée à une portion spécifique, ou encore peut être constituée d'un seul type d'entité.

Ce n'est pas notre objectif de développer une méthode automatique pour aider les experts à déterminer les perspectives avec lesquelles une ontologie peut avoir été développée, ou pour découvrir les entités affectées par une perspective donnée. Nous discutons néanmoins dans la Figure 12 des moyens de reconnaître des perspectives de manière manuelle : bien que connaître le type d'une perspective ne permet pas de déterminer son effet sur l'ontologie, nous énumérons quelques signes qui peuvent aider à reconnaître les différents types de perspective dans une ontologie.

| Type des perspectives | La perspective peut être reconnue par : |
|-----------------------------------|---|
| Domaine d'application | Granularité plus élevée dans une portion de la hiérarchie is-a |
| Application ou objectif technique | Présence de certaines entités (en particulier des attributs) qui non pas beaucoup de sens pour le domaine d'application, ou sont détaillés avec une granularité plus fine que le reste de l'ontologie |
| Point de vue ou rôle | Présence de relations is-a avec différents critères de catégorisation. Ceci |

| | |
|--|--|
| | <p>peut être illustré par un exemple de classification partielle des routes françaises (figure 23) : le concept « Route » est classifié en différents concepts qui sont associés avec des perspectives diverses, telles que l'environnement dans lequel la route est située, et la responsabilité administrative. Une telle différence dans le catégorisation n'est cependant pas suffisante pour déterminer avec certitude des perspectives différentes, vu que les différentes catégorisations peuvent provenir d'un arrière-plan historique, mais avoir une même fin.</p> |
|--|--|

Figure 12 : Signes pour reconnaître les différents types de perspective dans une ontologie

Les ontologies doit être analysées de manière suffisamment précise pour découvrir la plupart des perspectives qui ont guidé son développement ou qui traduisent des différences dans le type de données associées à leurs entités. Le succès de l'approche proposée dépend de cette identification. Idéalement, la recherche des perspectives avec lesquelles une ontologie a été développée doit être faite par un expert du domaine qui a participé au développement de l'ontologie, vu qu'il est à la bonne place pour connaître les raisons du choix de chaque concept, son placement dans la classification is-a, son nom, ses attributs, et les relations sémantiques qui le relie aux autres entités.

5.3 Modèle des perspectives

Dans cette étape, nous cherchons à savoir quelles sont les conditions requises pour le modèle des perspectives, dans l'optique de la réconciliation d'ontologie. Ensuite, nous proposons un modèle de représentation des perspectives.

Conditions requises pour le modèle

Les trois conditions suivantes sont essentielles pour satisfaire les besoins d'une réconciliation d'ontologie flexible plus fiable. Pour que les organisations soient intéressées dans une approche qui connecte leurs données avec celles de leurs partenaires pour des collaborations non anticipées, le modèle des perspectives devrait être :

- Simple, et utiliser les standards, afin de limiter l'énergie et le temps passé à ces développements
- Flexible, afin de permettre la réutilisation de ce qui a déjà été fait pour de nouvelles collaborations
- Sécurisé, afin de limiter l'accès aux données confidentielles de la compagnie.

Nous allons maintenant discuter ces conditions afin de déterminer les directives correspondantes pour le modèle :

☞ Comme le but de cette approche est d'améliorer la réconciliation d'ontologie, il semble approprié de représenter les perspectives dans une ontologie développée avec le langage OWL-DL. Cela nous permet d'utiliser des relations sémantiques ou des mappings pour relier les perspectives les unes avec les autres ; nous pouvons également tirer profit du mécanisme de nommage d'OWL, afin de s'assurer de l'unicité des noms des perspectives. De manière alternative, une simple liste d'identifiants peut suffire, avec le risque de non unicité du nommage, lorsque deux listes de perspectives de différentes organisations sont comparées.

☞ De la même manière que l'alignement d'ontologies ne modifie pas les ontologies en présence, la représentation des perspectives ne devrait pas requérir de modifier les ontologies non plus. Comme le but du modèle est d'aider les organisations à collaborer sur des projets non anticipés, le modèle devrait permettre la réutilisation de la plupart des travaux manuels

effectués. Le modèle doit donc garder séparé ce qui est spécifique à l'organisation, et ce qui est spécifique aux collaborations.

☞ Les collaborations ne devraient pas être des brèches dans la sécurité de l'organisation. Les données devraient être filtrées avant d'être rendues disponibles, afin de ne pas permettre l'accès à des données confidentielles. Associer des perspectives spécifiques aux données confidentielles et gérer ces perspectives ainsi que l'accès aux données devrait permettre de restreindre l'accès aux données confidentielles. De manière alternative, une version réduite des ontologies de l'organisation pourrait être rendue disponible pour la collaboration, où tous les concepts et relations confidentiels seraient retirés.

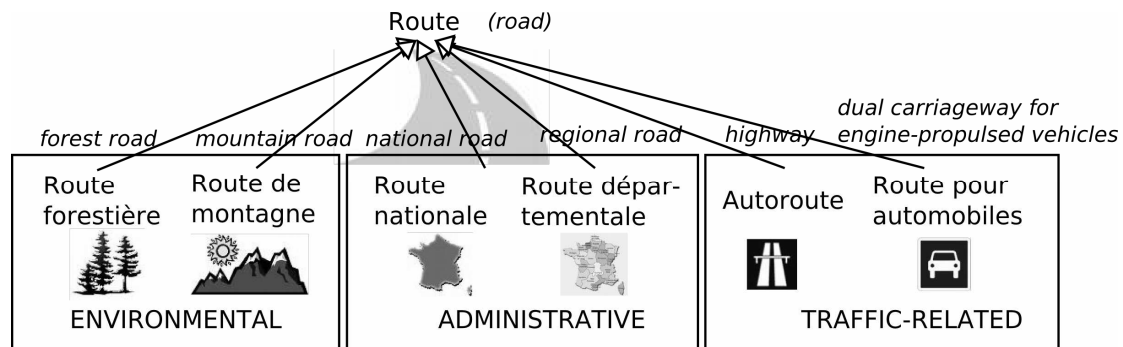


Figure 13 : Classification des routes françaises

Définition du modèle de perspectives

En ligne avec les directives déterminées dans les paragraphes précédents, nous proposons de modéliser les perspectives dans une « ontologie de perspectives ». Nous modélisons les perspectives par des classes, et les relier les unes aux autres avec des relations sémantiques dont le label correspond au sens du changement de perspective correspondant. Comme les changements de perspective ont une signification complètement différente lorsqu'ils sont pris dans une direction ou dans une autre, ces relations sémantiques ne sont pas symétriques.

Il est également possible d'utiliser les relations de subsumption pour utiliser les perspectives, bien qu'il n'y ait pas d'intérêt à définir plus de perspectives que celles qui ont déjà été reconnues comme impactant réellement le développement des ontologies.

La Figure 14 montre un exemple d'ontologies de perspectives, pour la représentation de produit. Cinq perspectives sont présentées : la conception, la géométrie, la fabrication et l'usinage. Ces perspectives sont reliées par des relations sémantiques définies pour l'occasion. La perspective d'usinage est définie comme une sous-classe de la perspective fabrication.

Préoccupations technique pour l'annotation des entités des ontologies avec des perspectives

Pour annoter les entités des ontologies avec les perspectives appropriées, sans modifier les ontologies, l'idée qui semble la plus naturelle est de placer cette information dans l'ontologie de perspectives. Cela peut cependant nous obliger à partager l'accès avec plus d'informations sur les ontologies, même les informations associées avec des perspectives jugées confidentielles.

Cela a aussi le désavantage d'inclure beaucoup de relations dans l'ontologie de perspectives (même s'il est possible de limiter le nombre de relations, par exemple en associant par une relation tous les concepts subsumés par un concept), avec le résultat de rendre cette ontologie

difficile à comprendre, et plus lourde. Finalement, OWL-DL ne permet pas de relier un concept à une propriété. Il serait donc impossible d'annoter directement des relations sémantiques ou des attributs.

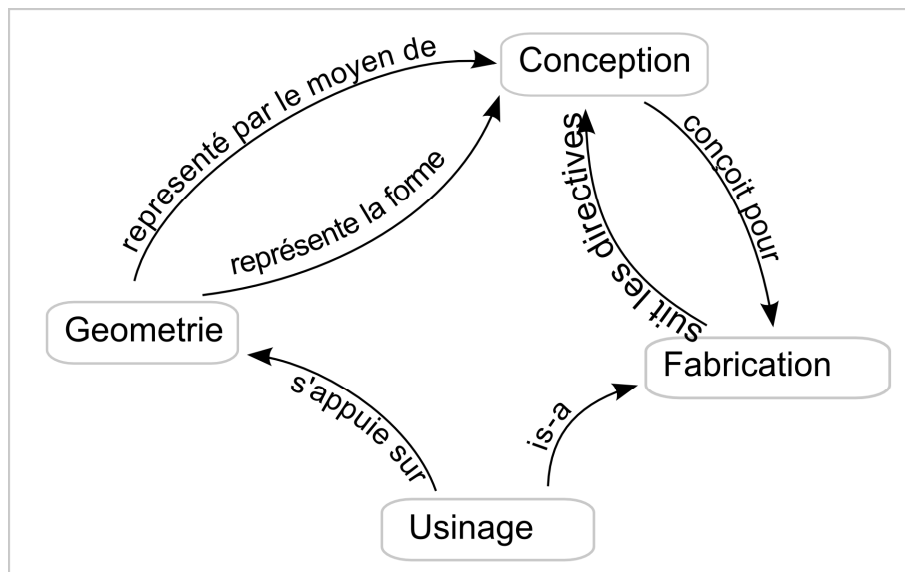


Figure 14 : Exemple d'ontologie de perspectives

Nous proposons donc de connecter les ontologies et les données dans une autre ressource, par exemple dans une base de données relationnelle. Cette solution a le désavantage d'inclure une nouvelle technologie ; cependant la plupart des organisations ont une base de données relationnelle, celle-ci est facile à installer, et performante. Il existe de nombreuses solutions techniques pour relier les bases de données relationnelles à des formulaires sur le Web (pour permettre à des personnes qui ne sont pas expertes en ontologies d'insérer les relations entre les perspectives et les ontologies), et pour utiliser les données enregistrées dans les applications. Cette solution est probablement la meilleure si la performance est une priorité.

Annoter les entités des d'ontologies avec des perspectives

Quand nous identifions qu'un concept a été défini selon une perspective donnée, alors les concepts qu'il subsume peuvent être vus selon la même perspective (ou par une perspectives qui la spécialise). Dans de tels cas, le concept qui subsume les concepts associés avec différentes perspectives possède aussi ces perspectives : dans l'exemple Figure 13, le concept « route » a les trois perspectives « environnement », « administratif », et « ayant trait au trafic », bien que ce ne soit pas apparent.

Dans OWL-DL, les relations sémantiques et les attributs sont définis de manière globale, bien que leur domaine et domaine de définition puissent être restreints. Nous proposons que lorsqu'ils sont annotés par une perspective, les concepts du domaine soient par défaut également annoté par la perspective; de manière réciproque, si un concept est annoté par une perspective, alors par défaut tous les attributs qui incluent le concept dans leur domaine sont annotés par la perspective également ; à moins qu'ils ne soient ôtés de manière explicite de la perspective.

Nous proposons un ensemble d'opérations basiques qui peuvent servir comme interface pour ajouter les entités d'une ontologie avec une perspective :

- Ajouter ou ôter un concepts
- Ajouter ou ôter un concept avec tous les concepts qu'il subsume

- Ajouter ou ôter une relation sémantique
- Ajouter ou ôter une relation sémantique avec toutes les relations sémantique qu'elle subsume
- Ajouter ou ôter un attribut
- Oter tous les attribus qui incluent un concept donné dans leur domaine

5.4 Modèle de la connexion entre perspectives

Dans cette étape, nous proposons un modèle pour comparer les perspectives définies dans différents organisations, les méthodes pour associer des perspectives aux ontologies qu'elles ont influencées.

Définition du modèle

Pour comparer le contexte de développement des ontologies de différentes organisations, il est nécessaire de pouvoir établir des relations entre leurs perspectives. Cela peut être fait en établissant des mappings entre les classes qui représentent les diverses perspectives. Mais en pratique, la compatibilité entre les perspectives peut varier. Nous proposons donc de relier des ontologies de perspective dans une nouvelle ontologie, dans laquelle nous importons les ontologies des perspectives des deux organisations. Dans le cas où les perspectives sont modélisées avec une simple liste d'identifiants, la relation entre les perspectives peut être faite, par exemple, dans une base de données relationnelle.

Méthodes pour comparer les perspectives de différentes ontologies

Nous voulons comparer les perspective d'une telle manière qu'il soit possible de déterminer quelles perspectives sont compatibles. Pour cela nous nous appuyons sur des relations sémantiques qui relient les perspectives définies dans les ontologies de perspectives des deux organisations qui collaborent.

Nous jugeons que les perspectives sont compatibles, dans notre cas, lorsqu'elles sont reliées les unes aux autres par une relation sémantique dont le domaine est dans l'ontologie de la compagnie, et le domaine de définition dans l'ontologie partenaire. S'il n'y a pas une telle relation sémantique, alors les perspectives sont considérées comme étant non reliées.

Il n'y a pas de méthode particulière pour comparer le contexte des concepts l'un avec l'autre, vu que ce contexte est défini à l'aide de perspectives. Le moyen de comparer les perspectives est donc similaire à l'alignement des ontologies (ou de mettre en correspondance des éléments de listes, si les perspectives sont représentées par une liste d'identifiants).

5.5 Évaluation des données pour la relation

Il est difficilement possible de s'assister d'outils d'alignement d'ontologies pour comparer les ontologies de perspective. En effet, les méthodes terminologiques ne seraient pas efficaces car les perspectives sont des concepts abstraits, dont la terminologie est plus difficile à comparer ; les méthodes structurelles seraient limitées par le nombre réduit de perspective, et par le peu de relations entre elles ; les mesures sémantiques ne peuvent pas être utilisées, vu que les perspectives définies dans une organisation ne peuvent pas être strictement équivalentes aux perspectives définies dans une autre organisation, et il n'y a pas de règles qui permettraient de trouver davantage de mappings par inférence.

Cette limitation ne doit cependant pas être un obstacle à l'adoption de l'approche, car le nombre de perspectives à aligner ne devrait pas être élevé. Les experts pourraient tirer avantage d'informations statistiques sur les mappings reliant les entités des ontologies des

organisations. Si plusieurs concepts annotés par une perspective sont mappés avec des concepts définis dans une ontologie d'une organisation partenaire, et donc associés avec une autre perspective, alors il est très probable que ces perspectives soient reliées d'une manière ou d'une autre.

5.6 Interprétation des données pour réaliser la désambiguïsation

Le concept origine et le concept enquis sont associés tous les deux à des perspectives qui ont été reconnues comme influençant les ontologies qui les définissent. Les perspectives de la compagnie et de l'organisation partenaires sont reliées les unes aux autres par des relations sémantiques asymétriques définies par des experts au moment de la collaboration.

Le processus de désambiguïsation est accompli en recherchant les couples de perspective telles que :

- la première perspective est associée au concept origine
- la seconde perspective est associée au concept enquis, et
- les deux perspectives sont reliées les unes aux autres par une relation sémantique dont le domaine inclut la première perspective, et dont le domaine de définition inclut le second.

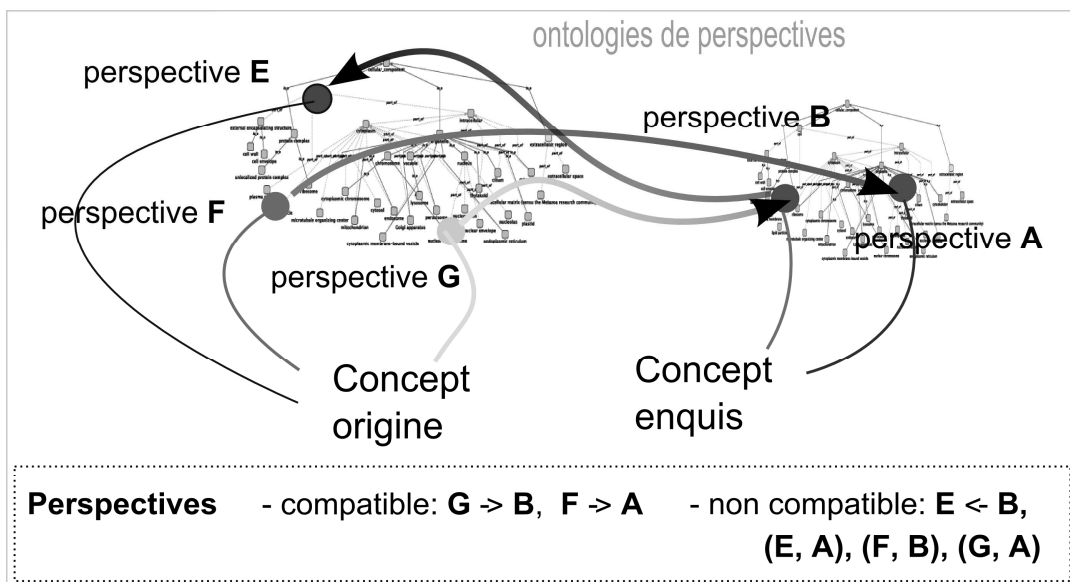


Figure 15 : Illustration du processus de désambiguïsation

5.7 Conclusion

Nous choisissons de représenter le contexte de l'ontologie par le moyen des perspectives avec lesquelles l'ontologie a été initialement développée et a évolué. Les experts représentent les perspectives de toutes les ontologies de l'organisation par des classes dans une ontologie en OWL-DL (ou alternativement, par une simple liste d'identifiants).

Quand un besoin de collaboration apparaît, ces perspectives sont reliées avec les perspectives de l'organisation partenaires, à l'aide de relations sémantiques asymétriques. Les perspectives du concept origine sont comparées avec les perspectives du concept enquis, afin de juger si elles sont compatibles. Quand les agents soumettent une requête, les perspectives associées au concept origine sont comparées avec les perspectives associées au concept enquis. Les couples de perspectives qui sont compatibles sont retournés (Tableau 3).

Mais les perspectives définies pour la compagnie ne sont pas tous pertinents pour chaque demandeur. Cela peut changer dramatiquement le résultat si l'évaluation de la relation entre les concepts retourne une valeur de relevance basée principalement sur ces perspectives. Nous allons étudier le contexte de demandeurs dans le chapitre suivant, et voir comment il peut être comparé aux perspectives de la compagnie, afin de trouver avec qu'elle perspective le concept origine est associé, afin d'obtenir une évaluation de relevance plus fidèle.

Tableau 3 : Réponses à la méthodologie

| | | | |
|-------------------------------------|--|---|---|
| Caractéristiques pertinentes | Caractéristiques : les perspectives avec lesquels le concept a été développé | | |
| Source des données | Source des données contextuelles disponibles : les ontologies elles-mêmes | Précision : aussi précis que possible | Conditions pour obtenir des mesures valides : si possible, les perspectives devraient être enregistrées par un expert qui a participé au développement des ontologies. |
| Modèle des perspectives | Critères et conditions requises : simple, flexible et sécurisé | Modèle du contexte : une ontologie OWL-DL, dans laquelle les perspectives sont représentées comme des classes, et les des changements de perspective comme des relations sémantique. De manière alternative, une liste simple d'identifiant pour les perspectives devrait suffire. | |
| Modèle de la connexion | Modèle de la relation entre les perspectives : une ontologie importe les ontologies des perspectives des deux organisations, et relie les perspectives avec des relations sémantiques asymétriques qui représentent des changements de perspective. | Méthode pour comparer les perspectives : l'alignement d'ontologies | |
| Evaluation | Mesures : il n'y a pas moyen d'évaluer de manière automatique la relation entre des perspectives. Les experts et les organisations devraient donc relier les perspectives les unes aux autres manuellement | | |
| Interpretation | Méthode pour désambiguïser le sens pragmatique des concepts comparés : trouver les perspectives à partir du concept origine qui sont reliées à une perspective du concept enquis. | | |

Chapitre 6 : Connexion entre le contexte du demandeur et celui du concept origine

Dans ce chapitre nous appliquons la deuxième phase de notre méthodologie pour déterminer les caractéristiques du contexte du demandeur et de sa connexion avec le contexte du concept origine pour l'usage de personnalisation (Figure 16). Le but visé est d'améliorer la réconciliation des ontologies, en restreignant les sens pragmatiques possibles du concept origine à ceux qui correspondent à la perception qu'a le demandeur de ce concept.

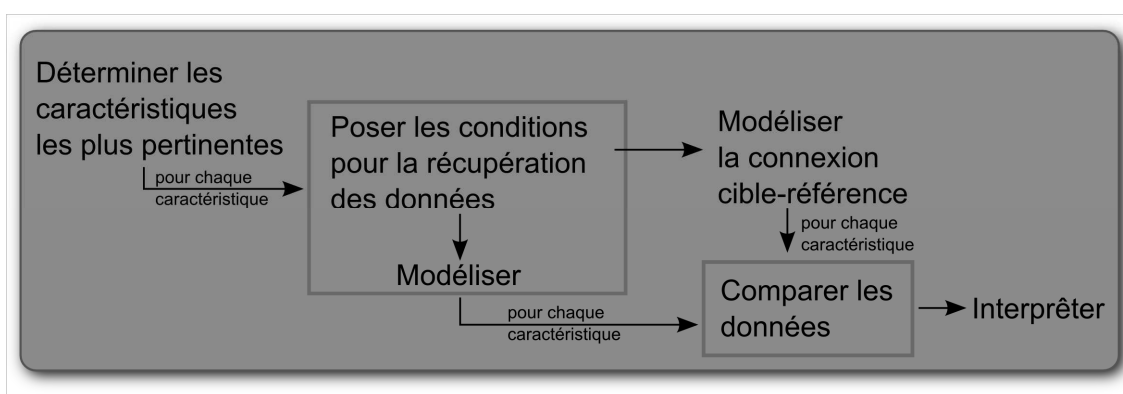


Figure 16 : Application de la méthodologie (deuxième phase, les siennes usage du contexte)

6.1 Caractéristiques pertinentes pour décrire le contexte du demandeur

Dans cette étape, nous déterminons les caractéristiques les plus pertinentes pour décrire le contexte du demandeur pour la réconciliation des ontologies.

Le contexte du demandeur est limité dans notre cas à son activité dans la compagnie. Nous décrivons donc le contexte du demandeur par une portion des domaines et tâches de la compagnie. Comme les domaines et les tâches sont complémentaires, nous les utilisons ensemble comme une seule caractéristique. La personnalisation est menée en regard avec la désambiguïsation, en restreignant le nombre de possibilités pour les sens pragmatiques du concept origine. L'information de référence doit donc être la même que dans le chapitre précédent, c'est-à-dire, le modèle de perspectives des ontologies de la compagnie.

6.2 Source des données et conditions pour les récupérer

Dans cet étape, nous avons à trouver les sources les plus appropriées pour récupérer les domaines et tâche du demandeur, ainsi que la précision requise.

Nous n'avons pas d'information explicite sur les agents qui peuvent soumettre une requête et leur utilisation d'une application logicielle particulière. Nous n'avons aucune information pour élaborer une méthode qui permette de déterminer de manière automatique les domaines et tâches qui sont pertinents pour eux. Leur contexte doit donc être saisi manuellement.

La granularité des données stockées dans la représentation des domaines et tâches doit être suffisamment fine pour permettre de distinguer les demandeurs l'un de l'autre : les utilisateurs ayant différents rôles dans l'organisation devraient obtenir des résultats différents. Mais elle ne doit pas être plus fine que nécessaire.

6.3 Modèle de domaines et tâches

Dans cette étape, nous établissons des critères pour qu'un modèle du contexte de l'agent soit efficace, et nous proposons un tel modèle.

Les critères que nous avons trouvés pour le contexte du demandeur sont :

- Saisie simple et sans effort
- fréquences minimales des mises à jour nécessaires.

Nous proposons donc que les agents décrivent leur contexte par une sélection simple des domaines et tâches de leur organisation, dans une liste, selon leur besoin d'informations potentiels.

Nous choisissons de représenter les domaines et tâches comme une liste d'identifiants qui peut être enregistrée dans une base de données relationnelle. Dans le cas où les domaines et tâches sont plus nombreux, on peut ajouter un niveau de domaine et tâches plus général. À moins qu'une classification complexe des domaines et tâches soit requise pour d'autres raisons, il semble raisonnable de privilégier la simplicité des bases de données relationnelles. Cela permet la génération facile de formulaires Web pour mettre à jour l'information et pour permettre aux agents de saisir leur information contextuelle.

6.4 Modèle de la connexion entre le contexte du demandeur et celui du concept origine

Dans cette étape, nous modélisons la connexion entre le contexte du demandeur et le contexte du concept origine, et déterminons une méthode pour comparer ces deux contextes.

Les domaines et tâches de la compagnie doivent être comparés avec les perspectives de la compagnie. Nous proposons donc ici d'importer les perspectives dans une table de la base de données relationnelle et d'associer une valeur normalisée de la pertinence entre chaque perspective et couple (domaine, tâche).

Pour simplifier l'administration d'un tel modèle pour la connexion entre domaines&tâches et perspectives, nous proposons que des valeurs par défaut soient introduites. Cela permet d'optimiser la connexion entre domaines&tâches et perspectives de manière progressive, selon les besoins. Par exemple, une valeur par défaut pourrait être associée à une perspective donnée, pour tout couple (domaine, tâche). La valeur finale associée à un triplet (domaines, tâche, perspective) est la valeur maximale parmi toutes les valeurs associées au triplet.

6.5 Évaluation des données pour la connexion

Dans cette étape, nous cherchons des mesures pour évaluer domaines&tâches, et établissons la connexion avec les perspectives.

Comme il est difficile pour les utilisateurs d'établir un profil directement en sélectionnant les perspectives qui les concernent, la relation entre domaines&tâche et perspectives doit être déterminée par un expert de la compagnie. Il serait donc préférable que ce soit quelqu'un qui a participé au développement des ontologies, et est impliqué dans la compagnie depuis longtemps, et connaisse donc la plupart des domaines et tâches de la compagnie.

6.6 Interprétation des données pour la personnalisation

Le processus de personnalisation est effectué en associant une valeur de pertinence à chaque perspective associée aux concept origine :

- Pour chaque perspective, l'agent peut sélectionner différents couples (domaine, tâche)
- A chaque triplet (domaine, tâche, perspective) est associée une valeur qui indique la pertinence de la perspective pour le couple (domaine, tâche).
- Pour chaque demandeur, une valeur de pertinence est associée à chaque perspective, c'est la valeur maximale de pertinence pour la perspective (qui peut être calculée off-line).

6.7 Conclusion

Nous proposons que des experts élaborent une liste des domaines et tâches de la compagnie, de telle sorte que les utilisateurs puissent facilement définir leur contexte, constitué d'une sélection des domaines et tâches de l'organisation. Les domaines et tâches sont reliés aux perspectives, et des experts peuvent affiner l'évaluation de pertinence des perspectives pour les couples (domaine, tâche) progressivement, selon les besoins. Basé sur ces valeurs de pertinence, il est possible de calculer une valeur de pertinence spécifique au demandeur pour chaque perspective.

Nous avons exposé dans ce chapitre comment déterminer les perspectives du concept origine qui sont pertinentes, selon les domaines et tâches du demandeur (Tableau 4). Mais quand un demandeur soumet une requête, il est possible que les données associées au concept enquis ne soient pas du tout adaptées à ce que le demandeur veut faire avec ces données, même si les perspectives associées au concept origine et au concept enquis sont compatibles. Nous allons étudier le contexte du besoin en interopérabilité dans le chapitre suivant, et voir comment il peut être comparé aux perspectives de la compagnie, afin d'évaluer la pertinence du concept enquis selon le besoin en interopérabilité.

| | | | |
|-------------------------------------|--|--|---|
| Caractéristiques pertinentes | Caractéristiques : domaine et tâche (complémentaires, à utiliser comme une seule caractéristique) du demandeur | | |
| Source des données | Source : l'agent doit fournir l'information | Précision et échelle : les domaines et tâches peuvent être décrits suffisamment précisément pour faire la différence entre les agents, et que manière suffisamment grossière pour rendre la sélection facile. | Conditions pour s'assurer la pertinence des données : aucune |
| Modèle des perspectives | Critères : la saisie du contexte doit être simple et sans effort pour les utilisateurs, avec un nombre minimum de mise à jour | modèle du contexte du demandeur : il est fait d'une simple sélection parmi les domaines et tâches de la compagnie, qui sont décrits comme une liste identifiant. | |
| Model de la connexion | Modèle de la relation entre domaine et tâches avec perspectives : une valeur de pertinence pour chaque triplé (domaines les plus tâches de perspective). | Méthode pour relier domaines et tâches avec perspective : les perspectives doivent être marquées avec des couples (domaines les plus tâches). | |
| Evaluation | Mesures : il n'y a pas moyen de déterminer un algorithme pour mesurer de manière automatique la relation entre domaine et tâches et les perspectives ; cela doit être évalué de manière manuelle. | | |
| Interpretation : | Méthodes d'interprétation pour la personnalisation : associer une valeur de pertinence spécifique aux demandeurs pour chaque perspective associée au concept origine. | | |

Tableau 4 : Résumé de la réponse à la méthodologie

Chapitre 7 : Connexion entre le contexte du besoin en interopérabilité et celui du concept enquis

Dans ce chapitre nous allons appliquer la deuxième phase de notre méthodologie pour déterminer les caractéristiques du besoin en interopérabilité et de sa connexion avec le contexte du concept enquis (Figure 17). Le but est d'améliorer la réconciliation des ontologies, en évaluant la pertinence des données associées au concept enquis pour le besoin en interopérabilité.

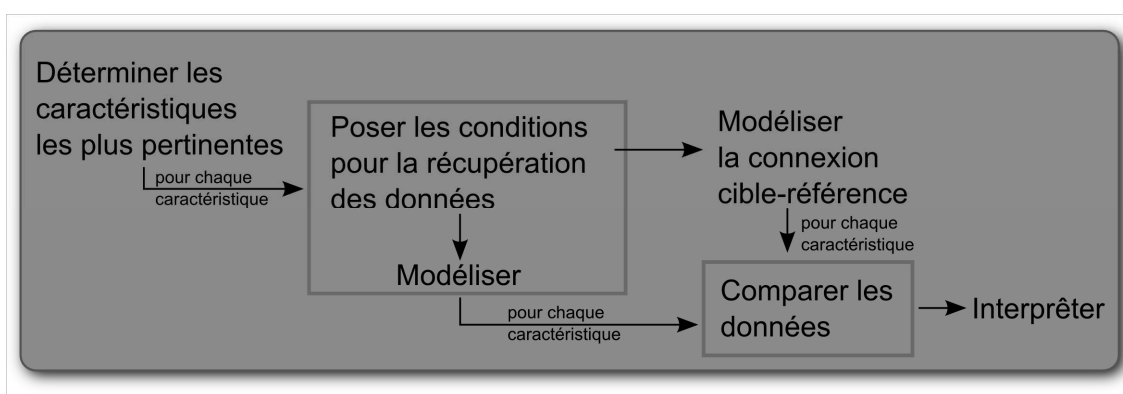


Figure 17 : Application de la méthodologie (deuxième phase, troisième usage du contexte)

7.1 Caractéristiques pertinentes

Dans cette étape, nous allons déterminer les caractéristiques qui sont les plus pertinentes pour caractériser le contexte du besoin en interopérabilité pour la réconciliation d'ontologies.

L'interopérabilité est « orientée tâche », ce qui signifie que la requête est effectuée dans une situation de collaboration : la compagnie s'intéresse à des données de l'organisation partenaire, avec l'objectif d'utiliser ces données pour une tâche particulière. L'évaluation des données récupérées doit donc logiquement dépendre de cette même tâche [Byström et Hansen, 2002]. Nous définissons la « tâche d'interopérabilité » comme la tâche pour laquelle les données récupérées au travers de la requête sont destinées. Un exemple de tâche d'interopérabilité est l'incorporation dans les designs élaborées par la compagnie de l'agent des designs des produits de l'organisation partenaires. D'autres caractéristiques possibles, telles que l'emplacement où le besoin est apparu, quand et pourquoi, ne sont pas relevant pour évaluer le concept enquis et le confronter au besoin en interopérabilité.

L'évaluation de la pertinence est associée à la désambiguïsation, en ce qu'elle restreint les sens pragmatiques possibles du concept enquis, en ne retenant que ceux qui sont pertinents pour le besoin en interopérabilité. Il est donc préférable que l'information de référence soit décrite de la même manière que dans le chapitre précédent, c'est-à-dire, par un modèle de perspectives.

7.2 Conditions pour récupérer des données concernant la tâche d'interopérabilité

Nous allons maintenant chercher à savoir comment déterminer la tâche d'interopérabilité correspondant à un besoin en information.

Le demandeur est l'agent qui connaît le mieux la tâche d'interopérabilité pour laquelle la requête a été soumise. Cependant, le contexte du besoin en interopérabilité doit être comparé avec une information de référence, c'est-à-dire, avec le contexte du concept enquîs. Pour que cette connexion puisse être effectuée à l'avance, les tâches d'interopérabilité possibles doivent avoir été catégorisées. Au moment de la requête, le demandeur sélectionne ainsi la tâche d'interopérabilité qui correspond le mieux à ce qu'il/elle a l'intention de faire avec les données récupérées.

Les tâches d'interopérabilité possibles dépendent de la spécialité de l'organisation partenaire et des données que cette organisation manipule et rend disponibles pour la collaboration. La liste des tâches d'interopérabilité possibles peut être développée au moment de la collaboration avec l'organisation partenaire, ou être réutilisée de collaborations précédentes avec des organisations similaires.

De la même manière que pour les domaines et tâches de la compagnie, un compromis doit être trouvé concernant la précision de la catégorisation des tâches d'interopérabilité. On peut remarquer que les tâches d'interopérabilité peuvent faire partie des sous-tâches d'une tâche de la compagnie. Les tâches d'interopérabilité sont plus spécifiques que les domaines et tâches catégorisés précédemment, qui ne sont pas directement associés avec des données particulières. Pour catégoriser les tâches d'interopérabilité, les experts qui ont relié les perspectives de la compagnie avec les perspectives de l'organisation partenaire, et qui ont eu un premier accès aux données du partenaire, sont les mieux placés pour identifier l'usage que la compagnie veut faire de ces données.

7.3 Modèle de la tâche d'interopérabilité

Dans cette étape, nous allons chercher des critères pour que le modèle du contexte du besoin d'interopérabilité soit efficace, et nous proposons un tel modèle.

Critères pour le modèle de la tâche d'interopérabilité

Les critères pour le modèle de la tâche d'interopérabilité découlent de ceux définis pour le modèle des perspectives :

- Simple, et utiliser des standards, pour faciliter l'adoption de l'approche
- Réutiliser la connaissance accumulée lors de collaborations précédentes
- Sécurisé, pour limiter l'accès aux données confidentielles de la compagnie

Nous allons maintenant discuter ces conditions afin de déterminer les directives correspondantes et les choix pour le modèle :

☞ Comme les tâches d'interopérabilité doivent être sélectionnées par le demandeur, et qu'une simple catégorisation est suffisante, nous proposons de les modéliser à l'aide d'une simple liste ou d'un arbre.

☞ Lors d'une nouvelle collaboration, les tâches d'interopérabilité déterminées pour des collaborations précédentes peuvent être utilisées comme un patron pour construire une nouvelle catégorisation des tâches d'interopérabilité. Une alternative est de catégoriser les tâches d'interopérabilité de manière générique pour toutes les collaborations, et de sélectionner pour chaque collaboration les tâches d'interopérabilité les plus pertinentes.

☞ Il n'y a pas de raison pour que le partenaire sache pour quelles tâches d'interopérabilité la compagnie va utiliser les données que l'organisation partenaire rend disponible pour la

collaboration. Le modèle doit donc garder séparés la catégorisation des tâches de l'information partagée avec l'organisation partenaire.

Définition du modèle

On peut modéliser la tâche d'interopérabilité par une liste simple d'identifiants, par exemple dans une base de données relationnelle, ou par une taxonomie.

Les tâches d'interopérabilité peuvent être reliées aux tâches de la compagnie. Cela peut simplifier la sélection des tâches d'interopérabilité au moment de la requête, et restreindre le nombre des tâches de la compagnie considérées pour le calcul de la valeur de pertinence spécifique au demandeur. Dans ce cas, le calcul d'une valeur de pertinence spécifique au demandeur pour chaque perspective annotant le concept origine devrait être retardé jusqu'au moment de la requête. En effet, cette valeur de pertinence est la valeur maximale de pertinence pour chaque triplet (domaines, tâches, perspectives), elle dépend donc des tâches d'interopérabilité effectivement considérées.

7.4 Modèle de la connexion entre la tâche d'interopérabilité et les perspectives

Dans cette étape, nous modélisons la connexion entre le contexte du besoin en interopérabilité et le contexte du concept enquis.

La pertinence d'une perspective pour une tâche d'interopérabilité ne change pas avec la requête. Elle peut donc être déterminée une fois pour toutes pour la collaboration, et être représentée par une valeur unique. Nous proposons donc de modéliser la connexion entre perspectives et tâches d'interopérabilité en associant une valeur normalisée de pertinence à chaque couple (perspective, tâches d'interopérabilité). Cela peut être implémenté à l'aide d'une base de données relationnelle.

Comme dans le chapitre précédent, nous proposons d'utiliser des valeurs par défaut pour que les experts puissent affiner l'évaluation de la pertinence des perspectives pour les tâches d'interopérabilité de manière progressive, selon les besoins. Alors qu'ils affinent la connexion entre les perspectives et les tâches d'interopérabilité, les experts peuvent se rendre compte que la connexion avec les données n'est pas bien représentée par les perspectives de l'organisation partenaire. Ils peuvent alors demander aux experts de l'organisation partenaire d'incorporer des perspectives plus détaillées qui puissent refléter mieux les caractéristiques des données associées aux ontologies de l'organisation partenaires.

7.5 Évaluation des données pour la connexion

La connexion entre les tâches d'interopérabilité et les perspectives de l'organisation partenaire doit être déterminée par un expert de la compagnie. Il serait préférable qu'il s'agisse de quelqu'un qui était impliqué dans la connexion entre les perspectives des deux organisations collaboratrices, et connaît donc les perspectives de l'organisation partenaire.

7.6 Interprétation des données pour évaluer le concept enquis

Le contexte du besoin en interopérabilité est constitué de la tâche d'interopérabilité pour laquelle les données récupérées au travers de la requête sont destinées. Le processus d'évaluation est simplement la récupération des valeurs de pertinence qui correspondent aux perspectives associées au concept enquis :

- À chaque perspective associée au concept enquis, correspond une valeur unique de pertinence pour la tâche d'interopérabilité sélectionnée au moment de la requête.

- S'il y a plus d'une valeur de pertinence, la valeur maximale est gardée.

7.7 Conclusion

Nous proposons que les experts élaborent une liste de tâches d'interopérabilité que la compagnie est à même d'accomplir avec les données de l'organisation partenaire. Les experts évaluent la pertinence des données liés aux perspectives de l'organisation partenaires pour les différentes tâches d'interopérabilité, et enregistrent la valeur de pertinence associée, par exemple dans une base de données relationnelle. Ils peuvent ensuite affiner l'évaluation de relevance des perspectives pour les couples (perspectives, tâche d'interopérabilité), progressivement, selon les besoins.

Au moment de la requête, le demandeur sélectionne la tâche d'interopérabilité associée au besoin en interopérabilité qui a déclenché la requête. La pertinence du concept enquis est évaluée en confrontant ses perspectives à la tâche d'interopérabilité, et est la valeur maximale de pertinence qui associe la tâche d'interopérabilité sélectionnée à l'une des perspectives du concept enquis.

Nous avons déterminé dans ce chapitre comment évaluer la pertinence du concept enquis, selon une tâche d'interopérabilité donnée. Dans le chapitre suivant, nous allons considérer les différents modèles que nous avons construits pour l'usage de personnalisation, désambiguïsation, évaluation, et concevoir une architecture pour les utiliser ensemble.

| | | | |
|-------------------------------------|--|--|--|
| Caractéristiques pertinentes | Caractéristiques : la tâche d'interopérabilité | | |
| Source des données | Source des données contextuelles disponibles : les demandeurs sont ceux qui préparent la requête, et qui devrait sélectionner la catégorie de la tâche d'interopérabilité et la plus proche dit ce qu'ils ont l'intention de faire avec les données. | Précision : les experts doivent trouver un compromis pour la précision de la catégorisation la tâche d'interopérabilité, de la même manière pour la définition des domaines et tâche de la compagnie. | Conditions pour la mesure : la catégorisation devrait être accomplie par des experts qui ont une bonne connaissance des affaires menées par la compagnie, et qui ont été impliqué dans la connexion entre la perspective pour la collaboration. |
| Modèle des perspectives | Critères : simple, réutilisables, sécurisé | Modèle de la tâche d'interopérabilité : une liste d'identifiants | |
| Model de la connexion | Modèle de la collection : une valeur de relevant est associé à chaque couple (perspective, tâches d'interopérabilité) | Méthode pour relier perspectives et tâches d'interopérabilité : la tâche d'interopérabilité doive être marquée avec des perspectives qui correspondent aux données pertinentes pour les accomplir. | |
| Evaluation | Mesures : il n'y a pas moyen de mesurer de manière automatique la connexion entre le latin est interopérabilité des perspectives ; elle doit être évaluée manuellement. | | |
| Interpretation : | Méthodes d'interprétation pour évaluer le concept enquis pour une tâche d'interopérabilité : c'est la valeur maximale de pertinence pour tous les couples (perspectives, tâches d'interopérabilité sélectionnée), où la perspective est associée au concept enquis. | | |

Chapitre 8 : Une architecture basée sur le contexte qui prend en compte les implications pratiques pour une meilleure réconciliation des ontologies

Dans ce chapitre nous allons appliquer la dernière phase de notre méthodologie pour concevoir une architecture qui regroupe les modèles développés, et s'assure de remplir tous les objectifs que nous avons établis dans la première phase (Figure 18). Le but est d'améliorer la réconciliation des ontologies par le moyen d'une approche basé sur le contexte.

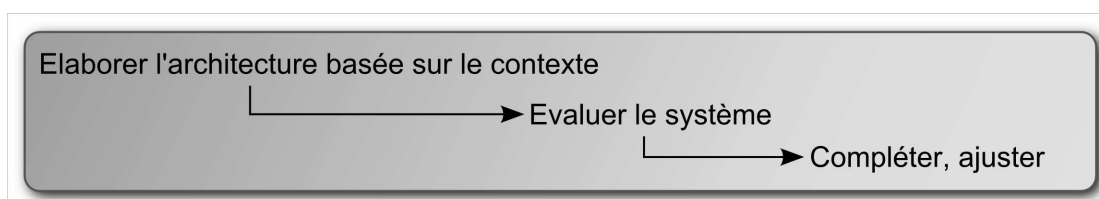


Figure 18 : Application de la méthodologie (troisième phase)

Les objectifs que nous avons établis dans la première phase de la méthodologie étaient :

- Améliorer l'utilisabilité de la réconciliation des ontologies en fournissant une mesure d'évaluation entre des concepts au choix
- Réduire le manque de considération des implications pratiques dans la réconciliation d'ontologie en considérant le contexte
- Améliorer la connexion entre concepts en considérant la pertinence des données qui leur sont associées
- Préserver la flexibilité procurée par l'alignement d'ontologie à la réconciliation des ontologies.

8.1 Architecture

Dans cette étape, nous développons une architecture afin d'utiliser les modèles décrits dans les trois chapitres précédents, selon les objectifs que nous venons de rappeler.

L'architecture est très simple : il suffit de placer les usages les uns après les autres de manière linéaire : personnalisation, désambiguïsation, et évaluation (Figure 19, le haut de la figure). Il faut noter que l'ordre de préparation des données contextuelles est toujours le même : (1) personnalisation (basée sur le modèle de perspectives élaboré par les experts), puis, quand une collaboration est planifiée, (2) la désambiguïsation et (3) l'évaluation d'un concept d'une ontologie de l'organisation partenaire.

La requête est composée d'informations telles que l'identifiant du demandeur, du concept origine et du concept enquis, et enfin de la tâche d'interopérabilité :

- Le concept origine décrit le type de données que le demandeur recherche. Le demandeur peut sélectionner ce concept dans une ontologie de sa compagnie.
- Le concept enquis est le concept défini dans une ontologie partenaire, et qui doit être évalué.
- La tâche d'interopérabilité est la tâche que le demandeur a l'intention d'effectuer, et pour laquelle il/elle compte utiliser les données associées aux concepts retournés par la requêtes.

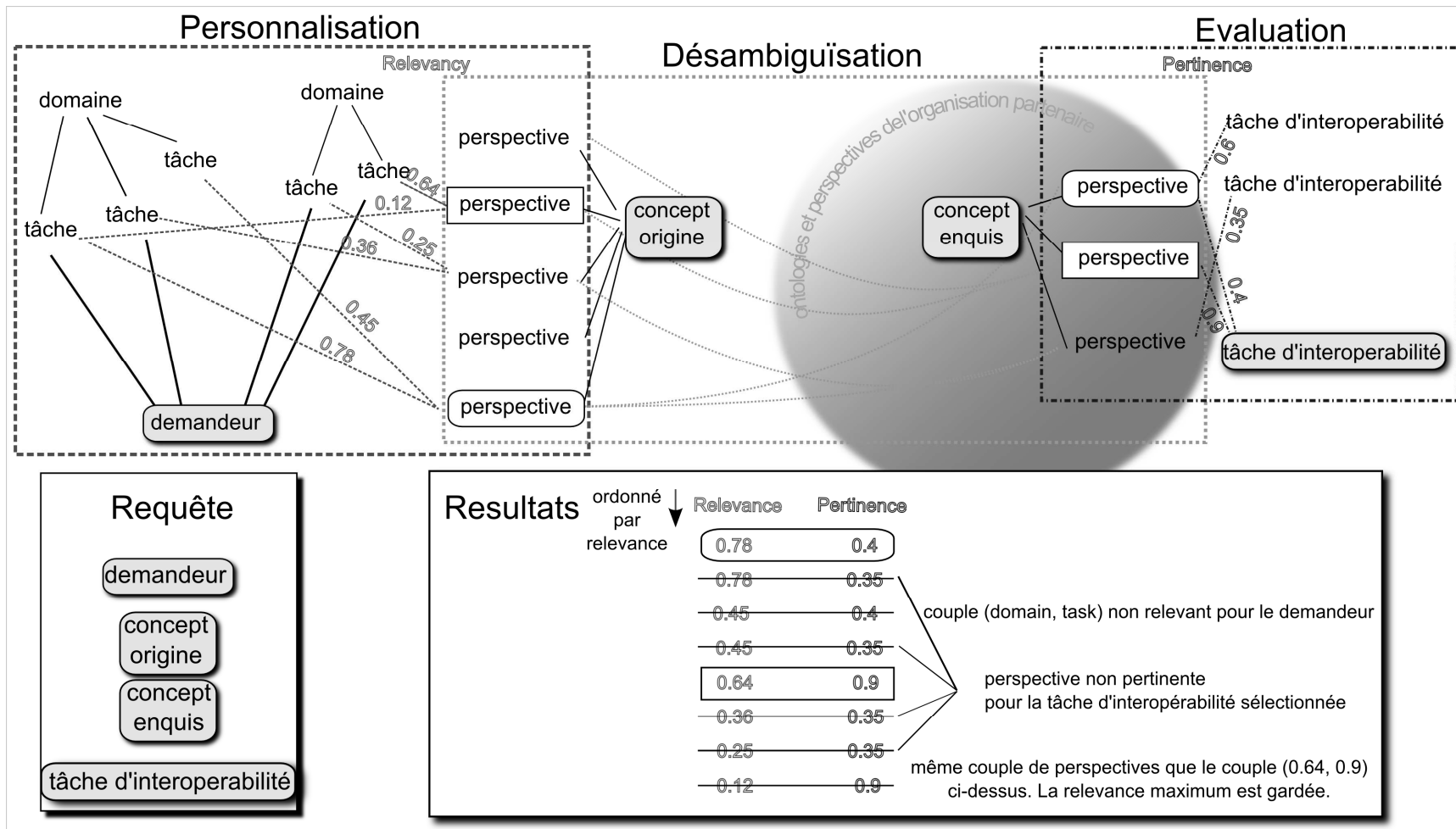


Figure 19 : illustration de l'architecture avec un exemple

☞ L'usage de la personnalisation restreint les perspectives associées au concept origine, afin de ne garder que celles qui sont pertinentes pour le demandeur. À chacune de ces perspectives, une valeur de pertinence est associée, qui est la valeur de pertinence la plus élevée entre la perspective et un couple (domaine, tâche) présent dans le contexte du demandeur.

☞ L'usage de la désambiguïsation limite les perspectives associées au concept enquis à celles qui sont compatibles avec des perspectives associées au concept origine (et sont donc pertinentes pour le demandeur).

☞ L'usage de l'évaluation associe une valeur de pertinence aux perspectives associées au concept enquis, que les données qui leurs sont associées sont utiles pour accomplir la tâche d'interopérabilité sélectionnée au moment de la requête.

Quand une requête est exécutée et que les calculs sont effectués, le résultat peut être composé de l'information suivante :

- Une liste (relevance, pertinence) pour chaque couple de perspective possible (une perspective annotant le concept origine et l'autre le concept enquis). Chaque rang de cette liste peut être complété par une liste des domaines&tâches de la compagnie qui sont pertinents pour la perspective correspondante. La liste peut être ordonnée par valeur décroissante de pertinence.
- Plus de détails peuvent être donnés selon les besoins (en particulier lorsqu'il y a peu ou pas de résultats, pour savoir comment élargir la requête) :
 - Pour chaque rang, le couple de perspective
 - Proportion des perspectives du concept origine qui sont pertinentes pour le contexte du demandeur
 - Proportion des perspectives du concept enquis qui sont compatibles avec une ou davantage des perspectives du concept origine
 - Proportion des perspectives du concept enquis qui sont pertinentes pour la tâche d'interopérabilité

Les trois usages pris ensemble permettent d'effectuer une évaluation du concept enquis basée sur le contexte. L'ordre des calculs qui semble le plus intuitif est le contraire de l'ordre de préparation des données contextuelles : comme la tâche d'interopérabilité est donnée au moment de la requête, on peut facilement trouver les différents couples (relevance, pertinence) non nuls (voir les résultats au bas de la figure 28).

8.2 Évaluation

Dans cette étape, nous allons examiner la solution basée sur le contexte que nous venons de proposer à la lumière des objectifs que nous avons rappelés au début du chapitre.

☞ Notre solution améliore l'utilisabilité réelle de la réconciliation d'ontologies en fournissant une mesure d'évaluation entre des concepts au choix.

☞ Notre solution est basée sur des informations contextuelles, qui incluent le demandeur, le développement de l'ontologie (incluant les données que l'ontologie sert à intégrer), et le besoin en interopérabilité. Cette information contextuelle place la

réconciliation d'ontologie dans le champ de référence plus large de la collaboration entre deux organisations, et de l'utilisation pratique des données récupérées au travers de la collaboration.

☞ Notre solution considère la pertinence des données associées au concept enquis. Bien qu'elles ne tiennent pas compte directement des données associées au concept origine, les perspectives associées à ce concept devraient montrer la connexion aux données. L'évaluation de la connexion entre les concepts est basée sur la comparaison des perspectives avec lesquelles les concepts ont été annotés, prenant ainsi en considération les données qui leur sont associées.

☞ Notre solution ne nécessite pas de modification dans les ontologies, ni dans les sources de données, et préserve donc la flexibilité des approches fédératives. De plus, si une ontologie évolue et une entité est modifiée, par exemple lorsque son label est changé, la solution va toujours fonctionner (même si on ne pourra peut-être pas comparer l'entité modifiée ou ajoutée avec une autre).

La solution que nous avons proposée répond donc presque complètement aux objectifs posés au début de l'application de notre méthodologie. Il y a cependant des points qui requièrent davantage de discussions :

- L'information contextuelle considérée ne fournit aucune aide pour réconcilier le format des données associées au concept. Mais l'approche vise à améliorer la connexion entre les ontologies. Or les ontologies définissent le sens des données et non pas leur format (niveau plus élevé d'abstraction). Le lecteur intéressé à réconcilier des formats de données à l'aide du contexte pourra lire [Firat et al, 2007].
- L'évaluation de la connexion entre les concepts est limitée à la comparaison des perspectives. Elle ne considère ni les mappings ni les relations sémantiques. Nous allons donc tâcher de fournir une mesure générique qui se base sur ces informations. Combinée à la mesure d'évaluation, elle va permettre une meilleure évaluation de la connexion entre les concepts.

8.3 Complétion par une mesure générique

Dans cette étape, nous allons chercher à compléter notre solution basée sur le contexte avec une évaluation de la connexion entre les concepts, qui utilise l'information décrite dans les ontologies et les mappings.

Types de mesures recherchées

Nous avons discuté dans l'état des lieux trois types de mesures entre les concepts. Ces mesures déterminent la similarité ou la connexité sémantique entre les concepts dans une ontologie, ou encore la similarité entre des concepts définis dans différentes ontologies.

Nous avons établi précédemment que l'évaluation de pertinence du concept enquis et sa comparaison avec le concept origine ne devraient pas être limitées à une notion stricte de similarité. Les mesures de connexité semblent donc une alternative intéressante, car elles considèrent différents types de relations sémantiques entre les concepts. Mais les ontologies définissent des relations sémantiques diverses et variées, dont la signification

n'est pas accessible directement. Dans ces conditions, il semble presque impossible de définir dans quels cas une relation sémantique peut indiquer une connexité entre les concepts qui soit pertinente pour répondre au besoin en interopérabilité. Ceci est particulièrement vrai lorsqu'il faut combiner deux relations sémantiques ou davantage : il n'est pas évident que des concepts reliés à travers au moins deux relations sémantiques soient effectivement proche sémantiquement.

En considérant alors les mesures similarité entre les concepts, il semble évident que les mesures les plus appropriées sont celles qui comparent des concepts définis dans différentes ontologie. Mais ces mesures se basent sur des méthodes terminologiques et structurelles, qui nécessitent un long temps de calcul. La mesure que nous proposons doit être effectuée entre les concepts au choix, et donc déterminés au moment de la requête. De plus, nous voulons améliorer l'alignement d'ontologie, et non pas le refaire. Nous voulons donc nous baser sur le travail effectué, et nous appuyer sur des mappings existants, et sur des mesures de similarité qui ont déjà fait leurs preuves.

Ce qui est important pour cette approche, ce n'est pas d'améliorer la mesure de similarité, mais de combiner une mesure de similarité avec notre solution basée sur le contexte. Nous proposons donc de sélectionner une mesure de similarité intra-ontologie connue, et de l'adapter pour qu'elle puisse utiliser les mappings, afin de comparer des concepts définis dans différentes ontologies.

Choix d'une mesure de similarité sémantique existante

Les mesures basées sur les caractéristiques considèrent comme caractéristique les attributs des concepts, ou parfois la relation sémantique. Mais quand les concepts à comparer sont définis dans des ontologies distinctes, il est quasiment impossible qu'ils aient des attributs ou des relations sémantiques en commun. Même s'il existait des mappings entre des attributs ou relations sémantiques de ces ontologies, il est probable que cela ne concernerait qu'une minorité d'entre eux, résultant en une mesure non efficiente.

Les mesures basées sur le contenu en information se basent sur des corpus de documents associés aux ontologies. Lorsqu'on compare un concept défini dans une ontologie avec un concept défini dans une ontologie d'une autre organisation, il est très probable que les corpus associés aux ontologies soient très différents l'un de l'autre. Une option serait de définir un corpus commun au moment de la collaboration, mais comme les documents sont certainement spécialisés, ils viendraient d'une organisation ou de l'autre, et le résultat ne serait pas équilibré, tout en nécessitant beaucoup de travail. Des arguments similaires peuvent être évoqués contre les mesures extensionnelles.

Les mesures basées sur le nombre d'arcs utilisent la classification d'hyponymie qui est presque toujours présente dans les ontologies. Bien que cette relation n'est pas toujours utilisée à bon escient, c'est un moyen d'évaluer si deux concepts ont un degré d'abstraction similaire, et dans quelles proportions ils peuvent être dits avoir un type similaire. Nous ne nous intéressons pas seulement à l'évaluation de la similarité stricte entre le concept origine et le concept enquis, mais il est intéressant d'être capable de comparer le degré d'abstraction des concepts, et de savoir s'ils peuvent être catégorisés

d'une manière similaire. Nous choisissons donc d'utiliser une mesure basée sur le nombre d'arcs.

[Petrakis et al, 2006] fournissent une évaluation de 12 mesures des quatre types, pour les ontologies WordNet et MeSH. La mesure basée sur le nombre d'arcs de [Leacock et Chodorow, 1998] est une mesure de similarité parmi les plus simples, cependant elle est soit première soit deuxième dans leur test (Equation 5). Dans [Budanitsky et Hirst, 2001], cette même mesure obtient des résultats homogènes et plutôt bons. Nous allons donc considérer cette mesure comme notre mesure de similarité de base.

Adaptation de la mesure pour comparer les concepts de différents d'ontologie

Nous voulons adapter une mesure qui compare les concepts définis par une ontologie, et soit basée sur les relations de subsomption, d'une telle manière qu'elles puissent être utilisées pour comparer les concepts d'ontologies distinctes, en utilisant des mappings entre les concepts. Les deux ontologies à comparer ont probablement une profondeur différente, et donc il n'est pas possible de garder exactement la même formule.

Soit un concept origine c_1 défini par une ontologie O_1 , comparé à un concept enquis c_2 défini par une ontologie O_2 . Supposons qu'il existe un mapping qui établisse l'équivalence de ces deux concepts (par exemple, un mapping avec une valeur de similarité entre 90% et 100%) qui permette de relier le concept c_{m1} défini par O_1 avec le concept c_{m2} défini par O_2 , de telle sorte que les concepts c_1 et c_2 puissent être reliés l'un à l'autre par une succession de relations *is-a*, et le mapping $c_{m1} - c_{m2}$.

La mesure basée sur le nombre d'arcs, sur laquelle la mesure de similarité de [Leacock and Chodorow, 1998] est fondée (Equation 3), $\text{len}(c_1, c_2)/2D$, est linéaire (Table 1). Nous proposons donc de remplacer la distance de c_1 à c_2 par la somme des distances de c_1 à c_{m1} et de c_{m2} à c_2 . Si le mapping qui relie c_{m1} à c_{m2} est un mapping de subsomption, il est équivalent d'ajouter une relation de subsomption entre c_{m1} et un concept temporaire mappé comme un équivalent de c_{m2} ; donc, dans ce cas, nous ajoutons 1 à la longueur calculée.

Nous proposons de modifier la formule (Equation 6) de telle sorte que le résultat soit normalisé. De plus, par la sommation des distances $\text{len}(c_1, c_{m1})/D_1$ et $\text{len}(c_2, c_{m2})/D_2$, nous augmentons le domaine de définition de la distance à $[0, 2]$. La norme devrait donc alors retourner des valeurs entre 0 et 1, et être décroissante sur l'intervalle $[0, 2]$. Nous augmentons un peu le domaine de définition, pour prendre en considération la valeur ajoutée à la distance dans le cas d'un mapping de subsomption. La Figure 20 montre la norme choisie (en vert) comparée au logarithme choisi par [Leacock and Chodorow, 1998] pour transformer la distance en une valeur de similarité.

Il se peut que différents mappings puissent être utilisés pour relier c_1 et c_2 . Le mapping choisi est celui qui maximise la valeur de similarité (Equation 7).

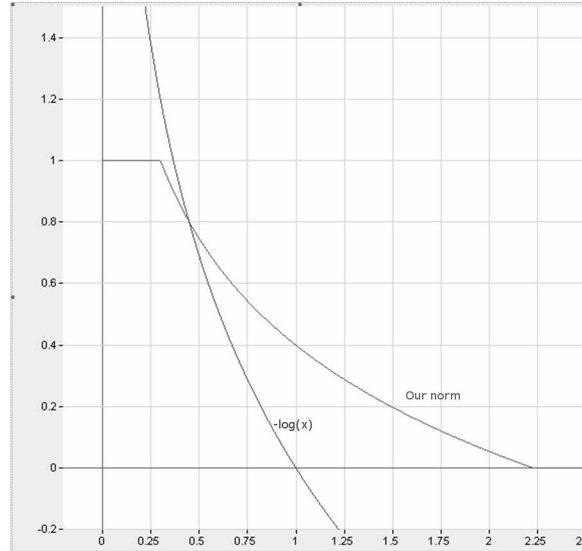


Figure 20 : Comparaison des norms : $-\log(x)$ et notre norme (“our norm”)

Table 1 : Common functions for the measures of similarity

| | |
|--|---|
| $c_1 \in O_1, c_2 \in O_2$ | Concepts définis dans une ontologie de la compagnie ou dans une ontologie de l'organisation partenaire, respectivement. |
| $len(c_1, c_2)$ | nombre minimum de relations <i>is-a</i> pour relier les concepts |
| $\lambda_{subs}(c_{m1}, c_{m2})$ | 1 si le mapping déterminé par (c_{m1}, c_{m2}) est un mapping de subsumption, 0 sinon |
| $p_1 \triangleleft c_1, p_2 \triangleleft c_2$ | Perspectives p_1 et p_2 associées respectivement aux concepts c_1 et c_2 |
| $rel_{applicant}(p_1)$ | Relevance spécifique au demandeur |
| $pert_{ITask}(p_2)$ | Pertinence spécifique à la tâche d'interopérabilité |

$$sim_{LC}(c_1, c_2) = -\log \frac{len(c_1, c_2)}{2D} \quad [5]$$

$$norm(x) = \min(\max(-\log(x^2) + 0.4, 1, 0)) \quad [6]$$

$$sim(c_1, c_2) = \max_{\{m, m \equiv c_{m1} - c_{m2}\}} \left[norm \left(\frac{len(c_1, c_{m1})}{2D_{O_1}} + \frac{len(c_{m2}, c_2)}{2D_{O_2}} + \frac{\lambda_{subs}(c_{m1}, c_{m2})}{D_{O_1} + D_{O_2}} \right) \right] \quad [7]$$

$$sim_{applicant,ITask}(c_1, c_2) = \max_{p_1 \triangleleft c_1, p_2 \triangleleft c_2} \sqrt{rel_{applicant}(p_1) \cdot (\lambda \cdot sim(c_1, c_2) + (1 - \lambda) \cdot pert_{ITask}(p_2))} \quad [8]$$

with $0 \leq \lambda \leq 1$

Intégration de la mesure de similarité générique à l'approche

La mesure de similarité générique vise à compléter la mesure d'évaluation basée sur le contexte que nous avons déterminée au fil des trois chapitres précédents.

En observant l'illustration de l'architecture, Figure 19, nous voyons que la mesure de similarité permet de relier le concept origine et le concept enquis, qui étaient précédemment connectés seulement de manière indirecte par leurs perspectives.

La mesure de similarité générique doit être calculée, au moins de manière partielle, au moment de la requête : la mesure globale doit pouvoir évaluer tout couple de concept. Le calcul de la mesure nécessite cependant un long temps de calcul, ce qui ne peut pas être fait au moment de la requête. Nous proposons donc d'adapter la méthode d'indexation de [Bidault, 2002] à notre configuration. Nous allons donc associer à chaque concept une chaîne de caractères alphanumérique unique pour l'ontologie indexée. Cette chaîne va permettre de comparer facilement les concepts en fonction des chemins constitués des relations *is-a* qui les relie à la racine de l'ontologie.

Au moment de la requête, les mappings qui relient les ontologies dans lesquelles le concept origine et le concepts requis sont définis, sont sélectionnés. Pour chaque mapping $c_{m1}-c_{m2}$, la distance $len(c_1, c_{m1})$ et $len(c_2, c_{m2})$, où c_1 est le concept origine et c_2 le concept enquis, sont calculées. La comparaison requise de milliers de chaînes de caractères (deux fois le nombre possible de mappings) de longueur inférieure à 100 caractères (profondeur moyenne du concept dans d'ontologie) peut être effectuée en moins de 100 ms sur un ordinateur récent.

Nous avons proposé, précédemment dans le chapitre, que la comparaison des concepts soit retournée essentiellement à l'aide d'un couple de valeur (relevance, pertinence). Il y a une seule valeur de similarité pour tout couple possible de (relevance, similarité). Il peut être intéressant de proposer une valeur unique de pertinence pour le concept enquis. Cela pourrait aider, par exemple, lorsqu'on compare la pertinence des différents concepts pour un même besoin en interopérabilité. Nous proposons donc de calculer une valeur globale de pertinence basée sur le produit de la relevance avec la moyenne paramétrique des valeurs de similarité et de pertinence (Equation 8).

8.4 Conclusion

Les modèles construits pour nos trois usages du contexte forment ensemble une mesure d'évaluation de la pertinence basée sur le contexte. Ils sont utilisés ensemble, et forment une séquence : personnalisation, désambiguïsation, et évaluation. Une solution basée sur le contexte, et fondée sur cette architecture pourrait ainsi retourner un couple (relevance, pertinence) pour chaque couple de perspective possible.

L'architecture répond à tous les objectifs fixés dans l'application de la première phase de la méthodologie. Cependant, il y a une limitation sérieuse, c'est-à-dire, le concept origine et le concept enquis ne sont comparés par aucune des relations sémantiques définies par les ontologies, ni par des mapping ; ils peuvent donc être très peu similaires, alors que le concept origine est supposé guider l'évaluation du concept enquis.

Nous avons donc proposé d'inclure une mesure générique de similarité sémantique dans l'architecture. Nous adaptons la mesure de similarité sémantique de [Leacock et Chodorow, 1998] à la comparaison de concepts définis par des ontologies distinctes. Nous utilisons la méthode d'indexation proposée par [Bidault, 2002] afin de permettre un calcul rapide de similarité entre les concepts au moment de l'enquête.

Nous avons maintenant appliqué la méthodologie à la réconciliation des ontologies, et avons conçu une architecture composée des modèles du contexte que nous avons

caractérisés. Par cette architecture, nous avons proposé une évaluation de pertinence d'un concept défini par l'ontologie d'une organisation partenaire, guidé par un concept défini par l'ontologie de la compagnie de l'agent, et par la tâche d'interopérabilité pour laquelle les données récupérées doivent être utilisées. La mesure retourne une valeur globale de pertinence et éventuellement une valeur de similarité générique, ainsi qu'une liste de couples (redevance, pertinence) pour chaque couple de perspective, ordonnée par pertinence.

Pour voir comment un système peut être implémenté en suivant l'approche proposée, le lecteur pourra se reporter à la version originale (anglaise) du document. La partie correspondante comprend une illustration à l'aide d'un exemple montrant comment un tel système peut être utilisé, et des détails techniques de l'implantation.

Tableau 5 : Résumé de la réponse à la méthodologie

| |
|--|
| <p>Architecture basée sur le contexte. L'architecture est constituée d'un simple ré-ordonnement des usages pour former la séquence : personnalisation, désambiguïsation, et évaluation. La connexion entre les domaines&tâches et les perspectives indique la pertinence des perspectives pour les domaines&tâches ; la connexion entre les perspectives compatibles indique la compatibilité entre les perspectives des diverses organisations ; la connexion entre la perspective et la tâche d'interopérabilité indique la pertinence de la perspective pour la tâche. Les couples (relevance, pertinence) sont le résultat d'une évaluation globale du concept enquis basée sur le contexte.</p> |
| <p>Evaluation : La solution basée sur le contexte que nous venons de proposer fournit une réponse à l'ensemble des d'objectifs fixés. Il y a cependant une limitation sérieuse : l'évaluation ne considère même pas les relations sémantiques et les mappings qui relient le concept origine avec le concept enquis. Ceux-ci peuvent donc être très dissemblable. Nous proposons donc d'inclure une mesure sémantique générique à notre approche.</p> |
| <p>Complétion : Nous proposons d'intégrer une mesure de similarité générique dans l'architecture basée sur le contexte, et des mappings : la mesure de connectivité sémantique se base sur des relations sémantiques qui sont difficilement comparables d'une ontologie à l'autre, et les mesures de similarité sémantique entre concepts de diverses ontologies utilisent les mêmes méthodes sur lesquelles les outils d'alignement d'ontologie sont basés. Nous choisissons donc une mesure basée sur le nombre d'arcs, car ces mesures s'appuient sur des relations hyponymie qui, malgré leur ambiguïté, sont présentes dans la plupart des ontologies. Nous sélectionnons la mesure de [Leacock et Chodorow , 1998] pour sa performance et sa simplicité. Nous adaptons la mesure pour la comparaison de concept définis par des ontologies distinctes, et pour l'évaluation rapide de concept au choix. Nous proposons finalement une mesure globale pour évaluer la pertinence d'un concept.</p> |

Conclusion générale

Résumé de la thèse et contributions

Cette thèse rapporte l'étude d'impact d'une approche basée sur le contexte pour améliorer l'interopérabilité entre des ontologies hétérogènes développées et évoluant de manière autonome. L'hypothèse sur laquelle cette étude est basée est que les organisations devraient avoir des ontologies qui représentent les concepts correspondant aux données qu'elles manipulent, et qu'elles peuvent être amenées à collaborer avec d'autres organisations, sans que cela ait été prévu au préalable. Des données sont échangées dans le cadre de ces projets, et ces organisations ont à réconcilier leurs ontologies afin de faciliter cet échange.

Comme ces ontologies ont été développées de manière indépendante et évoluent de manière autonome, l'alignement d'ontologies est l'approche la plus appropriée pour les réconcilier. Des outils existent pour assister les ingénieurs de développement d'ontologies à trouver des mappings possibles entre les entités définies dans les ontologies. Cependant comme les ontologies ont été développées dans des circonstances complètement différentes, les entités mappées ont différentes portées, sont associées avec des types de données différents, et ainsi de suite (voir le chapitre 1). En supposant que le besoin d'un agent est décrit d'une manière satisfaisante par un concept défini dans une ontologie de sa compagnie, le concept défini dans l'ontologie d'une organisation partenaire qui correspond le mieux au besoin de l'agent pourrait bien ne pas être le concept qui est en moyenne le plus similaire au premier concept ; ce pourrait être plutôt un concept qui est davantage similaire à ce concept dans le contexte particulier du besoin en interopérabilité.

Nous proposons dans cette étude d'examiner le contexte de la réconciliation des ontologies par des mapping. Traditionnellement, les ontologies sont alignées sans considération d'aucune information contextuelle qui puisse informer sur l'utilisation effective des concepts récupérés. Notre approche est basée sur l'intuition que par la considération du contexte, on devrait pouvoir fournir une meilleure évaluation de la pertinence d'un concept pour une requête d'interopérabilité donnée.

Ceci nous amène à notre première contribution : nous avons revu l'état des lieux qui a trait à l'utilisation du contexte en informatique, afin de mettre en évidence les principes de la collecte, de la modélisation et de l'utilisation du contexte. Nous avons proposé une définition générale du contexte, comme étant « toute information qui participe à la caractérisation d'une entité particulière (...) où les caractéristiques distinctives qui composent le contexte sont jugées selon un but particulier ». Suivant cette définition, nous avons développé une méthodologie pour déterminer ce qui est contextuel, comment le modéliser, et l'utiliser.

Nous avons appliqué cette méthodologie à la réconciliation d'ontologies, dans la situation d'une collaboration concurrente entre l'organisation et deux collaborateurs ou

davantage, où une certaine flexibilité est nécessaire. Nous avons trouvé qu'une solution raisonnable pour désambiguïser parmi les différents sens pragmatiques des concepts était d'annoter les concepts avec des perspectives. Ces perspectives ont été reconnues comme ayant guidé le développement de l'ontologie, ou comme des indicateurs des différents types de données associées à l'ontologie. Ceci est notre seconde contribution.

La réconciliation d'ontologies par l'alignement est limitée en ce que seulement certains concepts sont liés les uns aux autres par des mappings. Nous proposons donc une mesure de similarité sémantique générique inter-ontologies qui permet de relier n'importe quel concept défini dans une des ontologies de la compagnie de l'agent avec des concepts définis dans les ontologies d'autres organisations. Nous basons cette mesure sur la mesure intra-ontologie basée sur le nombre d'arcs de [Leacock et Chodorow, 1998], et sur des mappings existant entre les ontologies des organisations qui collaborent. Ceci est notre troisième contribution.

Finalement, nous avons proposé une évaluation de la pertinence d'un « concept entier », basée sur trois sortes d'information contextuelle. L'évaluation requiert l'information suivante : identité de l'agent qui prépare la requête, la description du besoin par un « concept origine », et la « tâche d'interopérabilité » pour laquelle la requête est préparée. À l'aide de la méthodologie, nous avons en effet reconnu et motivé trois contextes :

- Le contexte des concepts de l'ontologie, constitué des perspectives, comme décrit précédemment. Les perspectives du concept origine sont comparées avec les perspectives du concept enquis, afin de désambiguïser les sens pragmatiques des concepts en gardant seulement les couples de perspectives qui sont compatibles les unes avec les autres.
- Le contexte de l'agent qui lance la requête, et qui consiste en une sélection de domaines et tâches de sa compagnie ; ce contexte est introduit afin de personnaliser la mesure, en reconnaissant la compréhension que l'agent a du concept origine, et en retenant seulement les perspectives qui sont pertinentes pour lui.
- Le contexte du besoin en interopérabilité, constitué de la tâche d'interopérabilité pour laquelle l'entité recherchée est amenée à être utilisée. Ce contexte est introduit afin d'évaluer la valeur pratique du concept enquis. Les perspectives du concept enquis sont évaluées en les comparant avec la tâche d'interopérabilité, afin de juger de leur pertinence.

L'évaluation de la pertinence basée sur le contexte est composée d'une valeur de similarité générique, et du couple de valeur (pertinence, relevance) correspondant au couple de perspectives compatibles parmi toutes celles qui décrivent le sens pragmatique du concept origine et du concept enquis. Cette mesure basée sur le contexte est notre quatrième et dernière contribution.

Discussions et limites

Cette approche permet une évaluation de la pertinence pratique d'un concept pour un besoin en interopérabilité donnée, avec des hypothèses de départ limitées. Cette étude ne fournit pas une réponse directe aux attentes que l'on peut avoir pour les approches basées sur le contexte appliquées à la réconciliation d'ontologies. Notre approche est

basée sur des données que nous estimons être disponibles dans la plupart des cas. Il semble que dans l'État de l'art du développement des ontologies, la plupart de l'information qui est contextuelle pour la réconciliation d'ontologie ne peut pas être récupérée de manière automatique. Afin d'être traitable informatiquement de manière automatique, cette information contextuelle devrait en effet être décrite complètement par une liste de métadonnées établie pour chaque usage du contexte qu'il est prévu d'implémenter. Si l'on veut généraliser la contextualisation de la réconciliation d'ontologies, il faudrait généraliser l'annotation des ontologies avec ces métadonnées.

Bien que l'approche basée sur le contexte pourrait être menée plus loin que ce que nous avons proposé dans cette étude, cela demanderait certainement encore plus de travail manuel pour annoter les entités avec les métadonnées appropriées, ou pour récupérer d'autres types d'information contextuelle manuellement. La quantité d'efforts à fournir en préparant les données contextuelles doit être déterminée en relations avec le gain attendu.

Notre approche requiert une certaine quantité de travail préliminaire accompli par certains membres de l'organisation ayant une bonne connaissance des domaines et tâches de la compagnie, et qui sont capables d'identifier des perspectives avec lesquels les ontologies ont été développées. Les experts ont non seulement à vérifier les mappings retournés par les outils d'alignement, mais aussi à déterminer des perspectives, afin de les comparer avec des perspectives des organisations avec lesquels la compagnie veut collaborer, et de comparer ces dernières avec les tâches d'interopérabilité de la compagnie.

Est-ce que ce travail préliminaire en vaut la peine ? Certainement, si les mappings peuvent être utilisés, et si l'on peut faire confiance à la comparaison entre les concepts. Cependant la fiabilité de cette approche est limitée par le même problème qui affecte la modélisation d'ontologies et leur développement : la reconnaissance des perspectives et l'attribution de la pertinence et de la pertinence va varier d'un expert à un autre.

Travaux futurs

Cette approche pourrait bénéficier des améliorations suivantes :

- Découvrir des concepts enquis potentiels parmi les concepts des ontologies des autres organisations, par exemple en cherchant les concepts dont le label s'approche du label du concept origine. En effet l'approche présente dans cette thèse requiert que la requête soit constituée de deux concepts, alors qu'il est probable que dans la plupart des cas, seul le concept décrivant le besoin (concept origine) soit connu.
- Donner la possibilité de décrire la requête, non pas avec un seul concept origine, mais avec deux concepts ou davantage. Le système implémentant une telle approche pourrait par exemple tirer avantage de relations sémantiques standard afin de récupérer les concepts qui sont reliés de manière sémantique à tous les concepts indiqués dans la requête.
- Améliorer la mesure de similarité sémantique générique en considérant le niveau de confiance associée aux mappings, afin de présenter différentes valeurs de

similarités en fonction du niveau minimal de confiance requis par le demandeur pour un besoin en interopérabilité donné.

Perspectives

Nous proposons les directions de recherche suivante pour la réconciliation d'ontologies :

- Développer des micro-ontologies standards, qui seraient limitées à un domaine très spécifique et restreint :
 - Développer des outils pour le développement d'ontologies (tel qu'un plug-in dans Protégé) qui se connecte à ces micro-ontologies afin de proposer de référer des concepts existants quand c'est possible, par exemple en opérant une recherche basée sur les mots-clés parmi toutes les micro-ontologies disponibles.
 - Ces micro-ontologies devraient être « orientées tâches ». Par exemple, une ontologie pourrait définir l'information incluse dans des adresses postales. En se référant à cette ontologie, on saurait que la donnée associée au concept pourra servir pour des opérations relatives à l'envoi de courrier.
- Standardiser des relations sémantiques entre les concepts, et obliger celui qui développe une ontologie à spécialiser des relations sémantiques existantes quand il/elle veut définir des nouvelles relations sémantiques. En faisant cela, la comparaison d'ontologies et la recherche dans les ontologies seraient simplifiées, est rendues bien plus fiables. Cela permettrait d'ouvrir la porte à l'utilisation des relations sémantiques pour déterminer l'intérêt d'un concept. Les exemples de relations standard pourraient inclure « médium » (inverse de la relation « contenu ») et « source du contenu ».
- Modifier le modèle conceptuel des ontologies, de telle sorte que la signification de la relation « is-a » soit associée à un contexte, et définie de manière explicite : la catégorisation dépendrait du point de vue que l'on a sur les concepts. En effet les catégorisations sont seulement valides pour un paradigme donné, qui devrait être inclus dans leur représentation.

Bibliographie

- Albertoni, Riccardo, et Monica De Martino. "Asymmetric and Context-Dependent Semantic Similarity among Ontology Instances." Dans *Journal on Data Semantics X*, 1-30, 2008. http://dx.doi.org/10.1007/978-3-540-77688-8_1.
- , "Semantic Similarity of Ontology Instances Tailored on the Application Context." Dans *On the Move to Meaningful Internet Systems 2006 : CoopIS, DOA, GADA, and ODBASE*, 1020-1038, 2006. http://dx.doi.org/10.1007/11914853_66.
- Aleman-Meza, Boanerges. "Searching and Ranking Documents based on Semantic Relationships," 2006. [http://csdl2.computer.org/persagen/DLAbstoc.jsp?resourcePath=/dl/proceedings/s/&toc=comp/proceedings/icdew/2006/2571/00/2571toc.xml&DOI=10.1109/ICDEW.2006.131](http://csdl2.computer.org/persagen/DLAbstoc.jsp?resourcePath=/dl/proceedings/&toc=comp/proceedings/icdew/2006/2571/00/2571toc.xml&DOI=10.1109/ICDEW.2006.131).
- Aleman-Meza, Boanerges, Christian Halaschek-Wiener, I. Budak Arpinar, Cartic Ramakrishnan, et Amit P. Sheth. "Ranking Complex Relationships on the Semantic Web." *IEEE Internet Computing* 9, no. 3 (2005) : 37-44. doi : <http://doi.ieeecomputersociety.org/10.1109/MIC.2005.63>.
- Analyti, Anastasia, Manos Theodorakis, Nicolas Spyrtos, et Panos Constantopoulos. "Contextualization as an independent abstraction mechanism for conceptual modeling" (2007) : 24-60.
- Asperti, A., et G. Longo. *Categories, types, and structures*. MIT Press Cambridge, Mass, 1991.
- Attardi, G., et M. Simi. "A Formalization of Viewpoints." *Fundamenta Informaticae* 23 (1995) : 149-173.
- Baeza-Yates, Ricardo, et Berthier Ribeiro-Neto. *Modern Information Retrieval*. 1er éd. Addison Wesley, 1999.
- Baldauf, Matthias, Schahram Dustdar, et Florian Rosenberg. "A survey on context-aware systems." *IJAHUC* 2 (2007) : 263-277.
- Bao, Jie, Doina Caragea, et Vasant Honavar. "Modular Ontologies - A Formal Investigation of Semantics and Expressivity.." Dans , édité par Riichiro Mizoguchi, Zhongzhi Shi, et Fausto Giunchiglia, 4185 :616-631. Lecture Notes in Computer Science. Springer, 2006. <http://dblp.uni-trier.de/db/conf/aswc/aswc2006.html#BaoCH06>.
- Bartholomew, Doug Duvall. *PLM : Boeing's Dream, Airbus' nightmare*, 2007.
- Benslimane, Djamel, Ahmed Arara, Gilles Falquet, Zakaria Maamar, Philippe Thiran, et Faiez Gargouri. "Contextual Ontologies." Dans *Advances in Information Systems*, 168-176, 2006. http://dx.doi.org/10.1007/11890393_18.
- Benslimane, Djamel, Christelle Vangenot, Catherine Roussey, et Ahmed Arara. "Multirepresentation in Ontologies." Dans *Advances in Databases and Information Systems*, 4-15, 2003. <http://www.springerlink.com/content/5papg0dn59ycyj9r>.
- Berg, E. van den, R. Bidarra, et W. F. Bronsvort. "Web-based interaction on feature models." *From Geometric Modeling to Shape Modeling : Ifip Tc5 Wg5*. 2

- Seventh Workshop on Geometric Modeling : Fundamentals and Applications, October 2-4, 2000, Parma, Italy* (2001).
- Berners-lee, Tim. *Weaving the Web : The Original Design and Ultimate Destiny of the World Wide Web*. 1er éd. Collins, 2000.
- Bernstein, Abraham, Esther Kaufmann, Christoph Bürki, et Mark Klein. "How Similar Is It? Towards Personalized Similarity Measures in Ontologies." Dans *Wirtschaftsinformatik 2005*, 1347-1366, 2005. http://dx.doi.org/10.1007/3-7908-1624-8_71.
- Bharat, Krishna. "SearchPad : explicit capture of search context to support Web search." *Computer Networks* 33 (2000) : 493-501. doi:10.1016/S1389-1286(00)00047-5.
- Bidault, A. "Affinement de requêtes posées à un médiateur." Université Paris XI, Orsay, Paris, France, 2002.
- Blanchard, E., M. Harzallah, P. K. H. Briand, et P. Kuntz. "A typology of ontology-based semantic measures." Porto, Portugal, 2005.
- Blanchard, Emmanuel, Pascale Kuntz, Mounira Harzallah, et Henri Briand. "A Tree-Based Similarity for Evaluating Concept Proximities in an Ontology." Dans *Data Science and Classification*, 3-11, 2006. http://dx.doi.org/10.1007/3-540-34416-0_1.
- Bodenreider, Olivier, Tom Gruber, Nicola Guarino, Ivan Herman, Deborah McGuinness, Mark Musen, et coll. *Ontology, Taxonomy, Folksonomy : Understanding the distinctions*. NIST, Gaithersburg, Maryland, USA. : *Ontology Summit 2007*, 2007. http://ontolog.cim3.net/cgi-bin/wiki.pl?OntologySummit2007_Communique.
- Bouquet, Paolo, Heiko Stoermer, et Xin Liu. "Okkam4P : A Protégé Plugin for Supporting the Re-use of Globally Unique Identifiers for Individuals in OWL/RDF Knowledge Bases." Dans , 314 : Bari, Italy, 2007.
- Bradley, N.A., et M.D. Dunlop. "Toward a multidisciplinary model of context to support context-aware computing," 2005. ScientificCommons. <http://eprints.cdlr.strath.ac.uk/2927/>.
- Brézillon, Patrick. "Context Modeling : Task Model and Practice Model." Dans , 122-135. Springer Berlin / Heidelberg, 2007. http://dx.doi.org/10.1007/978-3-540-74255-5_10.
- Brunnermeier, Smita, et Sheila A. Martin. *Interoperability Cost Analysis of the U.S. Automotive Supply Chain : Final Report*. DIANE Publishing, 1999.
- Budanitsky, A., et G. Hirst. "Semantic distance in WordNet : An experimental, application-oriented evaluation of five measures." Dans , 2 ;, 2001.
- Budzik, J., et K. J. Hammond. "User interactions with everyday applications as context for just-in-time information access." *Proceedings of the 5th international conference on Intelligent user interfaces* (2000) : 44-51.
- Byström, K., et P. Hansen. "Work Tasks As Units for Analysis in Information Seeking and Retrieval Studies." *Emerging Frameworks and Methods : Proceedings of the Fourth International Conference on Conceptions of Library and Information Science (Colis 4)* (2002).
- Castano, Ferrara, et Montanelli. "Matching Ontologies in Open Networked Systems : Techniques and Applications." *Journal on Data Semantics V* 3870/2006. Lecture Notes in Computer Science (2006) : 25-63.

- Chalmers, M. "A Historical View of Context." *Computer Supported Cooperative Work (CSCW)* 13 (2004) : 223-247.
- Chalupsky, H. "OntoMorph : A Translation System for Symbolic Knowledge." Dans , 471-482. Colorado, USA, 2000.
- CHEN, David, et Guy DOUMEINGTS. "Basic Concepts and Approaches to Develop Interoperability of Enterprise Applications." Dans , 323-330. Lugano, Switzerland : Kluwer Academic Publisher, 2003.
- Clark, Nicola. "The Airbus saga : Hubris and haste snarled the A380." *International Herald Tribune* (2006). <http://www.iht.com/bin/print.php?id=3854623>.
- Costello, Roger L. *Enhancing Data Interoperability with Ontologies, Canonical Forms, and Include Files*, 2003. <http://www.xfront.com/interoperability/CanonicalForms.html>.
- Curino, Carlo, Giorgio Orsi, et Letizia Tanca. "X-SOM : A Flexible Ontology Mapper." Dans , 424-428. IEEE Computer Society, 2007.
- Dartigues, Christel, Parisa Ghodous, Michael Gruninger, Denis Pallez, et Sriram Ram. "CAD/CAPP Integration using Feature Ontology." *Concurrent Engineering* (2007) : 237-249.
- Dey, Anind K. "Understanding and Using Context." *Personal and Ubiquitous Computing* 5, no. 1 (Février 20, 2001) : 4-7. doi :10.1007/s007790170019.
- Ding, Kolari, Ding, et Avancha. "Using Ontologies in the Semantic Web : A Survey." Dans *Ontologies*, 14 :79-113. Integrated Series in Information Systems. Springer US, 2007. http://dx.doi.org/10.1007/978-0-387-37022-4_4.
- Do, Melnik, et Rahm. *Comparison of Schema Matching Evaluations*, 2003. <http://www.springerlink.com/content/4wbxb1jbw8w4ek9v>.
- Doan, AnHai, Jayant Madhavan, Robin Dhamankar, Pedro Domingos, et Alon Halevy. "Learning to match ontologies on the Semantic Web." *The VLDB Journal* 12 (2003) : 303-319.
- Ehrig, M., P. Haase, M. Hefke, et N. Stojanovic. "Similarity for Ontologies-a Comprehensive Framework." Dans , 13-24, 2004.
- Ekbia, Hamid R., et Ana Gabriela Maguitman. "Context and Relevance : A Pragmatic Approach." Dans , 156-169. Springer-Verlag, 2001.
- Euzenat, Jérôme, Thanh Le Bach, Jesús Barrasa, Paolo Bouquet, Jan De Bo, Rose Dieng-Kuntz, Marc Ehrig, et al. *State of the art on ontology alignment*. Deliverable. Knowledge web, 2004. <ftp://ftp.inrialpes.fr/pub/exmo/reports/kweb-223.pdf>.
- Euzenat, Jérôme. *An API for ontology alignment*, 2006. <http://gforge.inria.fr/docman/view.php/117/251/align.pdf>.
- Euzenat, Jérôme, Marc Ehrig, Anja Jentsch, Malgorzata Mochol, et Pavel Shvaiko. *Case-based recommendation of matching tools and techniques*. Knowledge Web, 2007. <ftp://ftp.inrialpes.fr/pub/exmo/reports/kweb-126.pdf>.
- Euzenat, Jérôme, et Pavel Shvaiko. *Ontology Matching*. Berlin Heidelberg (DE) : Springer-Verlag, 2007.
- Finkelstein, L., E. Gabrilovich, Y. Matias, E. Rivlin, Z. Solan, G. Wolfman, et coll. "Placing search in context : the concept revisited." *Proceedings of the 10th international conference on World Wide Web* (2001) : 406-414.
- Firat, A., S. Madnick, et F. Manola. "Multi-dimensional Ontology Views via Contexts in the ECOIN Semantic Interoperability Framework" (2005).

- Firat, Aykut, Stuart Madnick, et Benjamin Grosf. "Contextual alignment of ontologies in the eCOIN semantic interoperability framework" (2007) : 47-63.
- FRIEDLAND, NS, PG ALLEN, G. MATTHEWS, M. WITBROCK, D. BAXTER, J. CURTIS, et coll. "Project Halo towards a Digital aristotle." *The AI magazine* 25 (2004) : 29-47.
- Gagnon, Michel. "Ontology-based integration of data sources." Dans , 1-8, 2007. doi :10.1109/ICIF.2007.4408086.
- Gal, Avigdor, Giovanni Modica, Hasan Jamil, et Ami Eyal. "Automatic ontology matching using application semantics" (2005) : 21-31.
- Gallaher, Michael P., Alan C. O'Connor, et Thomas Phelps. *Economic Impact Assessment of the International Standard for the Exchange of Product Model Data (STEP) in Transportation Equipment Industries*. RTI International, 2002. www.nist.gov/director/prog-ofc/report02-5.pdf.
- Gärdenfors, P. *Conceptual Spaces : The Geometry of Thought*. Mit Pr, 2000.
- Geryville, H. M., A. Bouras, Y. Ouzrout, et N. S. Sapidis. "A solution for actors? viewpoints representation with collaborative product development." *Proceedings of Virtual Concept* (2006).
- Ghidini, et Serafini. "Reconciling Concepts and Relations in Heterogeneous Ontologies," 2006. http://dx.doi.org/10.1007/11762256_7.
- Ghidini, Serafini, et Tessaris. "On Relating Heterogeneous Elements from Different Ontologies," 2007. http://dx.doi.org/10.1007/978-3-540-74255-5_18.
- Ghidini, Chiara, et Fausto Giunchiglia. "Local Models Semantics, or contextual reasoning=locality+compatibility" (2001) : 221-259.
- Giunchiglia, F., M. Schorlemmer, F. McNeill, A. Bundy, M. Marchese, M. Yatskevich, et coll. *OpenKnowledge Deliverable 3.1. : Dynamic Ontology Matching : a Survey*. Trento, Italy, 2006.
- Goldstone, R. L., D. L. Medin, et J. Halberstadt. "Similarity in context." *Memory and Cognition* 25, no. 2 (1997) : 237-255.
- Gracia, Jorge, Vanesa Lopez, Mathieu D'Aquin, Marta Sabou, Enrico Motta, et Eduardo Mena. "Solving Semantic Ambiguity to Improve Semantic Web based Ontology Matching." Dans , édité par Pavel Shvaiko, Jérôme Euzenat, Fausto Giunchiglia, et Bin He, 2007.
- Gruber, Thomas R. "A translation approach to portable ontology specifications." *Knowl. Acquis.* 5 (1993) : 199-220. doi :<http://dx.doi.org/10.1006/knac.1993.1008>.
- Guarino, N. *Formal Ontology in Information Systems : Proceedings of the 1st International Conference June 6-8, 1998, Trento, Italy*. IOS Press, 1998.
- Guarino, N., et Giaretta, P. "Ontologies and Knowledge Bases : Towards a Terminological Clarification." Dans *Towards Very Large Knowledge Bases : Knowledge Building and Knowledge Sharing*, 25-32. IOS Press, Amsterdam, 1995.
- Guenther, R., et J. Radebaugh. "Understanding Metadata." *National Information Standard Organization (NISO) Press, Bethesda, USA* (2004).
- Guha, Ramanathan. "Contexts : a formalization and some applications." Stanford University, 1992.
- Haeri, S. H., H. Abolhassani, V. Qazvinian, et B. B. Hariri. "Coincidence-Based Scoring of Mappings in Ontology Alignment." *Journal of Advanced Computational Intelligence Vol 11* (2007).

- Hirst, Graeme, et David St-Onge. "Lexical chains as representations of context for the detection and correction of malapropisms.." Dans *WordNet : an electronic lexical database*, 305–332. I NetLibrary, 1998.
- Horrige, M., S. Bechhofer, et O. Noppens. "Igniting the OWL 1.1 Touch Paper : The OWL API." Innsbruck, Austria, 2007. http://www.webont.org/owled/2007/PapersPDF/submission_32.pdf.
- Hyvönen, Eero, Eetu Mäkelä, Mirva Salminen, Arttu Valo, Kim Viljanen, Sampa Saarela, et coll. "MuseumFinland--Finnish museums on the semantic web." *Web Semantics : Science, Services and Agents on the World Wide Web 3* (2005) : 224-241.
- IEEE Standard Computer Dictionary : A Compilation of IEEE Standard Computer Glossaries : 610*. Inst of Elect & Electronic, 1991.
- Jacobs, J., et A. Linden. *Semantic Web technologies take middleware to the next level*. Gartner Group, 2002. http://www.gartner.com/DisplayDocument?doc_cd=109295.
- Janowicz, Krzysztof. "Extending Semantic Similarity Measurement with Thematic Roles." Dans *GeoSpatial Semantics*, 137-152, 2005. http://dx.doi.org/10.1007/11586180_10.
- , "Kinds of Contexts and their Impact on Semantic Similarity Measurement," 2008. http://ifgi.uni-muenster.de/%7Ejanowicz/wp-content/uploads/2007/12/janowicz_comorea2007_kinds_of_contexts_and_their_impact_on_semantic_similarity_measurement.pdf.
- Jiang, J. J., et D. W. Conrath. "Semantic similarity based on corpus statistics and lexical taxonomy." *Proceedings of International Conference on Research in Computational Linguistics 33* (1997).
- Jouanot, Fabrice, Nadine Cullot, et Kokou Yétongnon. "Context Comparison for Object Fusion." Dans *Advanced Information Systems Engineering*, 1031, 2003. http://dx.doi.org/10.1007/3-540-45017-3_36.
- Kalfoglou, Yannis, et Marco Schorlemmer. "Ontology mapping : the state of the art." *Knowl. Eng. Rev* 18 (2003) : 1-31.
- Keßler, Carsten. "Similarity Measurement in Context." Dans *Modeling and Using Context*, 277-290, 2007. http://dx.doi.org/10.1007/978-3-540-74255-5_21.
- Keßler, Carsten, Martin Raubal, et Krzysztof Janowicz. "The Effect of Context on Semantic Similarity Measurement." Dans *On the Move to Meaningful Internet Systems 2007 : OTM 2007 Workshops*, 1274-1284, 2007. http://dx.doi.org/10.1007/978-3-540-76890-6_55.
- Kim, Junhwan, Michael J. Pratt, Raj Iyer G, et Ram D. Sriram. "Standardized data exchange of CAD models with design intent." *Computer-Aided Design* (2007).
- Kim, Kyoung-Yun, David G. Manley, et Hyungjeong Yang. "Ontology-based assembly design and information sharing for collaborative product development" (2006) : 1233-1250.
- Kingston, John. "Multi-perspective ontologies : Resolving common ontology development problems" (2008) : 541-550.
- Klein, Michel, Asuncion Gomez-Perez, Michael Gruninger, Heiner Stuckenschmidt, et Michael Uschold. "Combining and relating ontologies : an analysis of problems and solutions." *Workshop on Ontologies and Information Sharing, IJCAI'01* (2001). <http://citeseer.ist.psu.edu/klein01combining.html>.

- Laamari, et Ben Yaghlane. "Uncertainty in Semantic Ontology Mapping : An Evidential Approach," 2007. http://dx.doi.org/10.1007/978-3-540-75256-1_38.
- Lassila, Ora, et Deepali Khushraj. "Contextualizing Applications via Semantic Middleware." Dans , 183-191. IEEE Computer Society, 2005.
- Lawrence, S. "Context in Web Search." *IEEE Data Engineering Bulletin* 23 (2000) : 25-32.
- Leacock, C., et M. Chodorow. "Combining local context and WordNet similarity for word sense identification." *WordNet : An Electronic Lexical Database* 49, no. 2 (1998) : 265-283.
- Li, W. D., W. F. Lu, J. Y. H. Fuh, et Y. S. Wong. "Collaborative computer-aided design?research and development status." *Computer-Aided Design* 37 (2005) : 931-940.
- Lin, D. "An information-theoretic definition of similarity." *Proceedings of the 15th International Conference on Machine Learning* (1998) : 296-304.
- Lord, P. W., R. D. Stevens, A. Brass, et C. A. Goble. "Investigating semantic similarity measures across the Gene Ontology : the relationship between sequence and annotation." *Bioinformatics* 19, no. 10 (Juillet 1, 2003) : 1275-1283. doi :10.1093/bioinformatics/btg153.
- Ma, Y. -S., et T. Tong. "Associative feature modeling for concurrent engineering integration." *Computers in Industry* 51 (2003) : 51-71. doi :10.1016/S0166-3615(03)00025-3.
- Madhavan, Jayant, Philip A. Bernstein, et Erhard Rahm. "Generic Schema Matching with Cupid." Dans , 49-58. Morgan Kaufmann Publishers Inc, 2001.
- Maguitman, Ana, Filippo Menczer, Fulya Erdinc, Heather Roinestad, et Alessandro Vespignani. "Algorithmic Computation and Approximation of Semantic Similarity." *World Wide Web* 9, no. 4 (Décembre 23, 2006) : 431-456. doi :10.1007/s11280-006-8562-2.
- Malgorzata Mochol, et Elena Paslaru-bontas. "A High-Level Architecture of a Metadata-based Ontology Matching Framework." Dans , 354-358, 2006. <http://csdl2.computer.org/persagen/DLAbsToc.jsp?resourcePath=/dl/proceedings/&toc=comp/proceedings/dexa/2006/2641/00/2641toc.xml&DOI=10.1109/DEXA.2006.9>.
- Mascardi, Viviana, Valentina Cordi, et Paolo Rosso. *A comparison of upper ontologies*, 2007.
- McCarthy, John L. "Notes on Formalizing Context.." Dans , 555-562, 1993.
- Meilicke, Christian, et Heiner Stuckenschmidt. "Analyzing Mapping Extraction Approaches." Dans , édité par Pavel Shvaiko, Jérôme Euzenat, Fausto Giunchiglia, et Bin He. Bexco, Busan, Korea, 2007.
- Mitra, P., et G. Wiederhold. "Resolving terminological heterogeneity in ontologies." Proc. ECAI-02 Workshop, CEUR-WS 64, Lyon, France, 2002.
- Mochol, Malgorzata, et Elena Paslaru-bontas. "A High-Level Architecture of a Metadata-based Ontology Matching Framework." Dans , 354-358, 2006. <http://csdl2.computer.org/persagen/DLAbsToc.jsp?resourcePath=/dl/proceedings/&toc=comp/proceedings/dexa/2006/2641/00/2641toc.xml&DOI=10.1109/DEXA.2006.9>.
- Moldovan, D., A. Badulescu, M. Tatu, D. Antohe, et R. Girju. "Models for the Semantic Classification of Noun Phrases." Boston, MA, 2004.

- Moon, Peter. *The future of the Web as seen by its creator*, 2007. <http://www.itworld.com/Tech/4535/070709future/>.
- Musen, Mark. *Ontology classification by purpose(?)*. National Institute of Technologies, Gaithersburg, Maryland, USA, 2007.
- Nagy, M., E. Motta, et M. Vargas-Vera. "Multi-Agent Ontology Mapping with Uncertainty on the Semantic Web." Dans , 49-56, 2007.
- Naiman, Channah F., et Aris M. Ouksel. "A classification of semantic conflicts in heterogeneous database systems." Dans , 167-193. Dallas, Texas, United States : Ablex Publishing Corp., 1995.
- Noy, Natalya Fridman, et Mark A. Musen. "PROMPT : Algorithm and Tool for Automated Ontology Merging and Alignment." Dans , 450-455. AAAI Press / The MIT Press, 2000.
- Ouksel, A. M., et A. Sheth. "Semantic interoperability in global information systems." *SIGMOD Rec.* 28, no. 1 (1999) : 5-12.
- Paslaru-Bontas, Elena. "A contextual approach to ontology reuse." University of Berlin, 2007. <http://www.diss.fu-berlin.de/2007/230/>.
- Patil, Lalit M. "Interoperability of formal semantics of product data among product development systems." University of Michigan, 2005.
- Paulheim, H., M. Rebstock, et J. Fengel. "Context-Sensitive Referencing for Ontology Mapping Disambiguation." Dans , 49-58, 2007.
- Petrakis, E. G. M., G. Varelas, A. Hliaoutakis, et P. Raftopoulou. "X-Similarity : Computing Semantic Similarity between Concepts from Different Ontologies." *Journal of Digital Information Management (JDIM)* 4, no. 4 (2006) : 233-238.
- Porzel, Robert, Hans-Peter Zorn, Berenike Loos, et Rainer Malaka. "Towards a separation of pragmatic knowledge and contextual information." Riva del Garda, Italy, 2006.
- Poslad, Stefan, et Landong Zuo. "An Adaptive Semantic Framework to Support Multiple User Viewpoints over Multiple Databases." Dans *Advances in Semantic Media Adaptation and Personalization*, 261-284, 2008. http://dx.doi.org/10.1007/978-3-540-76361_13.
- Pratt, M.J. "Extension of ISO 10303, the STEP standard, for the exchange of procedural shape models." Dans , 317-326, 2004.
- Pratt, Michael J. "Introduction to ISO 10303---the STEP Standard for Product Data Exchange." *Journal of Computing and Information Science in Engineering* 1 (2001) : 102-103.
- Qazvinian, V., H. Abolhassani, et S. H. Haeri. " Evolutionary Coincidence-Based Ontology Mapping Extraction." *Journal of Knowledge Engineering Special Issue on Advances in Ontologies Expert Systems* (2008). http://www.macs.hw.ac.uk/~hossein/expert_sys07.pdf.
- Rada, R., H. Mili, E. Bicknell, et M. Blettner. "Development and application of a metric on semantic nets." *Systems, Man and Cybernetics, IEEE Transactions on* 19 (1989) : 17-30.
- Rahm, Erhard, et Philip A. Bernstein. "A survey of approaches to automatic schema matching." *The VLDB Journal* 10 (2001) : 334-350.
- Ray, Steven R. "Interoperability Standards in the Semantic Web." *Journal of Computing and Information Science in Engineering* 2 (2002) : 65-69.
- Rector, Alan, Natasha Noy, Nick Drummond, et Mark Musen. *Ontology Design Patterns and Problems : Practical Ontology Engineering using Protege-OWL*.

- Galway, Ireland, 2005. <http://www.code.org/resources/tutorials/iswc2005/ISWC-Tutorial-Best-practice-2005.ppt>.
- Resnik, P. "Semantic Similarity in a Taxonomy : An Information-Based Measure and its Application to Problems of Ambiguity in Natural Language." *Journal of Artificial Intelligence* 11 (1999) : 95-130.
- , "Using information content to evaluate semantic similarity in a taxonomy." Dans , 1 : 448-453, 1995.
- Rodríguez, M., et Max Egenhofer. "Putting Similarity Assessments into Context : Matching Functions with the User's Intended Operations." Dans *Modeling and Using Context*, 833, 1999. http://dx.doi.org/10.1007/3-540-48315-2_24.
- Rodríguez, M. A., et M. J. Egenhofer. "Comparing geospatial entity classes : an asymmetric and context-dependent similarity measure." *International Journal of Geographical Information Science* 18 (2004) : 229-256.
- Rodríguez, M. Andrea, et Max J. Egenhofer. "Determining Semantic Similarity among Entity Classes from Different Ontologies." *IEEE Trans. on Knowl. and Data Eng.* 15 (2003) : 442-456.
- Rodríguez, M. Andrea, Max J. Egenhofer, et Robert D. Rugg. "Assessing Semantic Similarities among Geospatial Feature Class Definitions." Dans , 189-202. Springer-Verlag, 1999.
- Rosén, J. "FEDERATED THROUGH-LIFE SUPPORT, ENABLING ONLINE INTEGRATION OF SYSTEMS WITHIN THE PLM DOMAIN." *Proceedings of the 1st Nordic Conference on Product Lifecycle Management* (2006).
- Scanlan, J., A. Rao, C. Bru, P. Hale, et R. Marsh. "DATUM Project : Cost Estimating Environment for Support of Aerospace Design Decision Making." *Journal of Aircraft* 43 (2006) : 1022-1028.
- Schlenoff, Craig, Peter Denno, Rob Ivester, Don Libes, et Simon Szykman. "An analysis and approach to using existing ontological systems for applications in manufacturing" (2000) : 257-270.
- Schwering, Angela. "Hybrid Model for Semantic Similarity Measurement." Dans *On the Move to Meaningful Internet Systems 2005 : CoopIS, DOA, and ODBASE*, 1449-1465, 2005. http://dx.doi.org/10.1007/11575801_32.
- Seco, N., T. Veale, et J. Hayes. "An intrinsic information content metric for semantic similarity in WordNet." Dans , 1089--1090, 2004.
- Segev, et Gal. "Putting Things in Context : A Topological Approach to Mapping Contexts to Ontologies," 2007. http://dx.doi.org/10.1007/978-3-540-74987-5_4.
- Semy, Salim K., Mary K. Pulvermacher, et Leo Obrst. *Toward the Use of an Upper Ontology for U.S. Government and U.S. Military Domains : An Evaluation*. Bedford, Massachusetts : MITRE, 2004. <http://colab.cim3.net/file/work/SICoP/resources/UpperOntologyUseLong.doc>.
- Seydoux, Florian. "Exploitation de connaissances sémantiques externes dans les représentations vectorielles en recherche documentaire." Ecole Polytechnique Fédérale de Lausanne (EPFL), 2006. http://biblion.epfl.ch/EPFL/theses/2006/3654/EPFL_TH3564.pdf.
- Shadbolt, Nigel, Tim Berners-Lee, et Wendy Hall. "The Semantic Web Revisited" (2006). <http://eprints.ecs.soton.ac.uk/12614/>.
- Shvaiko, P., et J. Euzenat. "A Survey of Schema-Based Matching Approaches." *Journal on Data Semantics IV* 4 (2005) : 146-171.

- SILVA, Catarina E. FERREIRA DA. “Découverte de correspondances sémantiques entre ressources hétérogènes dans un environnement coopératif.” Université Claude Bernard Lyon I, 2007.
- de Souza, Kleber Xavier Sampaio, et Joseph Davis. “Aligning Ontologies and Evaluating Concept Similarities.” Dans *On the Move to Meaningful Internet Systems 2004 : CoopIS, DOA, and ODBASE*, 1012-1029, 2004. <http://www.springerlink.com/content/lvxrln7qmvv38728>.
- Sowa, J. F. “The Challenge of Knowledge Soup.” Homi Bhabha Centre, Mumbai : J. Ramadas and S. Chunawala., 2006. <http://www.jfsowa.com/pubs/challenge.pdf>.
- Sowa, John. “Processes and participants.” Dans *Conceptual Structures : Knowledge Representation as Interlingua*, 1-22, 1996. http://dx.doi.org/10.1007/3-540-61534-2_1.
- Standardization, ISO - International Organization for. *STEP*, 2006. http://www.iso.org/iso/iso_cafe_step.htm.
- Straccia, Umberto, et Raphaël Troncy. “oMAP : Combining Classifiers for Aligning Automatically OWL Ontologies.” Dans *Web Information Systems Engineering ? WISE 2005*, 3806/2005 :133-147. Lecture Notes in Computer Science. Springer Berlin / Heidelberg, 2005. http://dx.doi.org/10.1007/11581062_11.
- Strang, Thomas, et Claudia Linnhoff-Popien. *A Context Modeling Survey*, 2004. <http://elib.dlr.de/7444/>.
- Stuckenschmidt. “Toward Multi-viewpoint Reasoning with OWL Ontologies,” 2006. http://dx.doi.org/10.1007/11762256_21.
- Su, Xiaomeng. “Semantic enrichment for ontology mapping.” Norwegian University of Science and Technology, Trondheim, Norway, 2004.
- Su, Xiaomeng, et Jon Atle Gulla. “Semantic Enrichment for Ontology Mapping.” Dans *Natural Language Processing and Information Systems*, 217-228, 2004. <http://www.springerlink.com/content/0yrgqvpv34lxbe21>.
- Su, Xiaomeng, Sari Hakkarainen, et Terje Brasethvik. “Semantic enrichment for improving systems interoperability.” Dans , 1634-1641. Nicosia, Cyprus : ACM, 2004.
- Sunna, et Cruz. “Structure-Based Methods to Enhance Geospatial Ontology Alignment.” Dans *GeoSpatial Semantics*, 2007. http://dx.doi.org/10.1007/978-3-540-76876-0_6.
- Tang, M. X., et J. Frazer. “A representation of context for computer supported collaborative design.” *Automation in Construction* 10 (2001) : 715-729.
- The role of standards in engineering and technology*. IEEE Computer Society, 2005. http://www.ieee.org/portal/cms_docs_iportals/iportals/education/setf/tutorials/baseline/glossary.html.
- Turney, P. D. “Measuring semantic similarity by latent relational analysis.” *Proceedings of the Nineteenth International Joint Conference on Artificial Intelligence (IJCAI-05)* (2005) : 1136–1141.
- , “Similarity of Semantic Relations.” *Computational Linguistics* 32, no. 3 (2006) : 379-416.
- Tversky, A. “Features of similarity.” *Psychological Review* 84, no. 4 (1977) : 327-352.
- Veltman, Kim. H. “Syntactic and semantic interoperability : new approaches to knowledge and the semantic web.” *NEW REVIEW OF INFORMATION NETWORKING* (2001) : 159-184.

- Visser, P. R. S., D. M. Jones, T. J. M. Bench-Capon, et M. J. R. Shave. "Assessing heterogeneity by classifying ontology mismatches." *Formal Ontology in Information Systems* (1998) : 148-162.
- Wache, H., T. Vogele, U. Visser, H. Stuckenschmidt, G. Schuster, H. Neumann, et coll. "Ontology-Based Integration of Information? A Survey of Existing Approaches." Dans , 108-117. Seattle, WA, 2001.
- Wu, Z., et M. Palmer. "Verb semantics and lexical selection." *Proceedings of the 32nd Annual Meeting of the Association for Computational Linguistics* (1994) : 133-138.
- Yétongnon, Kokou, Seksun Suwanmanee, Djamel Benslimane, et Pierre-Antoine Champin. "A web-centric semantic mediation approach for spatial information systems." *Journal of Visual Languages & Computing* 17, no. 1 (Février 2006) : 1-24. doi :10.1016/j.jvlc.2005.04.002.
- ZACHMAN, JA. "A framework for information systems architecture." *IBM systems journal* 26 (1987) : 276-292.
- Zha, X. F., et H. Du. "Knowledge-intensive collaborative design modeling and support Part I : Review, distributed models and framework." *Computers in Industry* 57 (2006) : 39-55.
- Zhang, S., et O. Bodenreider. "Comparing associative relationships among equivalent concepts across ontologies." *Medinfo* 11, no. Pt 1 (2004) : 459-466.
- Zhdanova, Anna, et Pavel Shvaiko. "Community-Driven Ontology Matching." Dans *The Semantic Web : Research and Applications*, 34-49, 2006. http://dx.doi.org/10.1007/11762256_6.
- Zhu, et Madnick. *Scalable Interoperability Through the Use of COIN Lightweight Ontology*, 2007. http://dx.doi.org/10.1007/978-3-540-75474-9_3.
- "Scalable Interoperability Through the Use of COIN Lightweight Ontology," 2007. http://dx.doi.org/10.1007/978-3-540-75474-9_3.

TITRE en français

Similarité sémantique inter-ontologies basée sur le contexte

RÉSUMÉ en français

Cette thèse étudie l'intérêt du contexte pour améliorer l'interopérabilité entre ontologies hétérogènes, d'une manière qui permette leur évolution indépendante. Lors de collaborations, les organisations échangent leurs données, qui sont décrites par des concepts définis dans des ontologies. L'objectif est d'obtenir un service d'évaluation de tels concepts, basé sur le contexte.

Nous proposons une méthodologie pour déterminer, modeler et utiliser le contexte. En l'appliquant, nous découvrons trois usages du contexte qui contribuent à améliorer la réconciliation d'ontologies : Nous proposons de *désambiguïser* les sens pragmatiques possibles des concepts en comparant les "perspectives" avec lesquelles les concepts ont été développés ; de *personnaliser* en considérant le contexte des agents, constitué d'une sélection pertinente parmi les domaines et tâches de l'organisation ; d'*évaluer* la pertinence des données associées au concept pour la tâche qui a suscité le besoin en interopérabilité.

RÉSUMÉ en anglais

This thesis studies how a context-based approach could improve interoperability between heterogeneous ontologies developed and evolving autonomously. The hypothesis is that during collaborations, organizations exchange their data whose meaning is described by concepts defined by ontologies. The objective is to propose a context-based evaluation of such concepts.

We propose a methodology to determine, model and use the context. Applying this methodology, we uncover three usages of context that may be used together to improve ontology reconciliation : We propose to *disambiguate* among the possible pragmatic meanings of the concepts by comparing the "perspectives" with or for which the concepts have been developed; to *personalize* this comparison by considering the agent's context, made of a relevant selection of the agent's company domains and tasks; to *evaluate* the pertinence of the data associated with the concept for the task that triggered the interoperability need.

DISCIPLINE

Informatique (CNU 27)

MOTS-CLÉS

Ontologies, ontology matching, mapping, contexte, point de vue, similarité sémantique

INTITULÉ ET ADRESSE DE L'U.F.R. OU DU LABORATOIRE :

UFR Informatique
Université Claude Bernard Lyon 1
43, Boulevard du 11 novembre 1918
69622 Villeurbanne Cedex