



HAL
open science

Echange d'Information grâce à des caractérisations sémantiques

Anthony Ventresque, Philippe Lamarre, Sylvie Cazalens

► **To cite this version:**

Anthony Ventresque, Philippe Lamarre, Sylvie Cazalens. Echange d'Information grâce à des caractérisations sémantiques. Modèles Formels de l'Interaction - MFI 05, May 2005, Caen, France. pp.247-252. hal-00419637

HAL Id: hal-00419637

<https://nantes-universite.hal.science/hal-00419637>

Submitted on 24 Sep 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.

Echange d'information grâce à des caractérisations sémantiques

Anthony VENTRESQUE* Philippe LAMARRE* Sylvie CAZALENS*
ventresque@univ-nantes.fr lamarre@univ-nantes.fr cazalens@univ-nantes.fr

*Laboratoire d'Informatique de Nantes Atlantique (LINA)
Faculté des Sciences et Techniques de Nantes
2 rue de la Houssinière
BP 92208 – 44322 NANTES

Résumé :

L'utilisation d'ontologies permet à des agents de mettre du sens derrière les termes utilisés lors d'une interaction. Nous appuyant sur ce fait, nous proposons une entité permettant de qualifier sémantiquement différentes sortes d'informations : documents, bases de données, services, capacités d'agents, etc. Le *focus* consiste en une pondération des concepts d'une ontologie de façon à indiquer quels sens sont importants pour l'agent qui l'a mis en place.

Pour aider la saisie de focus, nous présentons alors une procédure de propagation des valeurs de pondérations dans le focus. Puis, pour comparer, nous définissons une mesure de pertinence d'un focus par rapport à un autre. L'originalité de notre approche réside dans le fait que le focus est une entité échangeable entre les agents et qu'elle ne nécessite pas de centralisation des données.

Mots-clés : Recherche d'information, ontologie, similarité, pertinence, qualification sémantique.

Abstract:

By using ontologies we allow agents to put meaning behind the terms used during interaction. So we propose an entity named *focus* that could represent various kinds of information contents : documents, data bases, services, agents' capabilities, etc. It consists of a weighting of the concepts of an ontology in order to indicate which meaning are important for the agent which has set it up.

To help the capture of the values in the focus we present a procedure to spread the weightings in the focus. Then for comparing we define a measure of relevance of a focus compared to another.

The originality of our approach lies in the fact that the focus is an exchangeable entity between the agents and does not require a centralization of the data collections.

Keywords: Information retrieval, ontology, similarity, relevance, semantic qualification.

1 Introduction

Tout système d'échange d'information, par exemple un système de recherche d'information, fait face à trois problèmes : 1 - taille et croissance des données ; 2 - expression des contenus ; et 3 - pertinence de la recherche.

Le premier problème ne fait que s'accroître. Il est ainsi estimé que la taille des informations stockées sur les disques durs à travers le monde en 2003 est comprise dans une fourchette de 403 à 1986 téraoctets et que la progression depuis

1999 est de 114%¹.

D'autre part, les informations sont décrites en utilisant différentes structures et langages. Les techniques de recherche doivent déployer des stratégies lourdes de traduction entre langages et structures sous peine de n'avoir accès qu'à un sous-ensemble des contenus.

Le troisième problème est double : il concerne la polysémie (plusieurs sens pour un terme) et la synonymie (plusieurs termes pour un sens). Cela engendre respectivement du bruit (des informations non pertinentes sont obtenues) et des silences (des informations pertinentes sont oubliées) lors de recherches d'information.

Utiliser des informations sémantiques doit permettre de dépasser les problèmes d'hétérogénéité syntaxique, de limiter bruit et silence et de rendre les échanges d'information plus pertinents et efficaces. Nos objectifs sont donc partiellement partagés avec ceux conduisant à la construction d'ontologies. Nous utilisons naturellement les ontologies comme point de départ en ajoutant la possibilité de pondérer chaque concept pour indiquer l'intérêt qui lui est porté localement. Nous appelons les structures ainsi obtenues des *focus*. Ils doivent permettre de qualifier sémantiquement les contenus, requêtes, mais aussi les préférences des agents, leur qualités, leur réputation... et ainsi de faciliter les échanges.

Considérons par exemple que nous nous situons dans un système multi-agents composé de médiateurs, de fournisseurs de services et de demandeurs de services. Ces services peuvent être des calculs (par exemple trouver un itinéraire), de l'information (par exemple des documents), etc. Les médiateurs sont regroupés en communautés d'intérêt et hiérarchisés. Les fournisseurs de service s'inscrivent auprès du ou des médiateurs qui leur conviennent. Les médiateurs ont connaissance des qualités des agents fournisseurs de services et des intérêts de ces derniers. L'architecture globale de ce système est décrite en [5].

¹<http://www.sims.berkeley.edu/research/projects/how-much-info-2003/execsum.htm>

Les informations qui sont échangées (requêtes, intérêts, déclarations de capacité, de service, de qualité, etc.) peuvent être décrites de manière sémantique grâce aux focus. Par exemple, il peut s'agir d'un champ supplémentaire ou complémentaire dans un langage d'interaction comme FIPA-ACL. Les médiateurs indiquent l'intérêt de la communauté qu'ils représentent par un focus, les fournisseurs de service recherchent la communauté qui correspond le mieux à leur préférences, les requêtes sont routées par un calcul de pertinence entre leur focus et ceux des communautés, puis les médiateurs recherchent les fournisseurs de service les plus aptes et les plus intéressés pour répondre, etc.

Notre travail s'organise ainsi. Dans la section 2, nous présentons les approches semblables existant en pointant les propriétés que notre démarche d'échange d'information impose. Nous définissons ensuite (section 3) la notion de focus en présentant ce que nous entendons par ontologie et nous introduisons le processus de propagation de valeurs dans un focus. Puis nous étudions (section 4) la normalisation de focus afin de définir la pertinence relative entre deux focus. Les premiers tests (section 5) de notre approche montrent que la comparaison de focus telle que définie précédemment est un processus rapide.

2 Travaux liés

Mettre en place un thésaurus ou une ontologie et pondérer les mots-clés ou les concepts est une idée qui a déjà été explorée en recherche d'information ou en traitement du langage naturel. Nous nous situons dans le domaine de l'échange d'information : documents, requêtes, services, intérêts, capacités, etc. et nous demandons des propriétés auxquelles les approches existantes ne répondent pas ou partiellement.

2.1 Le modèle des vecteurs sémantiques

Les approches qui ressemblent aux focus utilisent la notion de vecteurs sémantiques. Il s'agit à chaque fois de construire un espace à nombreuses dimensions, chaque dimension étant soit un terme clé, soit un concept. Un contenu ou une requête prend une valeur pour chacune de ces dimensions. Nous obtenons alors un vecteur grâce aux pondérations sur chaque axe. Le *Vector Space Model* [4] utilise des mots-clés de domaine pour construire une matrice de

$M \times N$, avec M le nombre de mots-clés et N le nombre de documents. Les dimensions peuvent donc être très grandes : plusieurs centaines à plusieurs milliers de mots-clés doivent être envisagés.

Pour diminuer les dimensions et réduire bruit et silence dus à l'utilisation de mots-clés, le *Latent Semantic Indexing* [9] travaille sur des concepts et utilise des algorithmes permettant de diminuer les dimensions de l'espace de représentation en ne considérant que celles qui sont réellement porteuses de sens.

La limitation que nous voyons à cette approche est que la matrice contenant les termes ou les concepts d'un côté et les documents de l'autre a besoin d'être gérée d'une manière centralisée : pour la mise en place des valeurs, il est à chaque fois tenu compte de la fréquence d'apparition d'un terme ou d'un concept dans le document mais aussi dans tout le corpus (*i.e.* la matrice)² ; pour les algorithmes permettant de diminuer les dimensions, il est nécessaire d'avoir accès à toute la matrice. Or nous voulons que les agents puissent interagir en s'échangeant des informations sémantiques sans qu'il soit nécessaire de passer par une autorité centrale.

2.2 Le projet THESUS

Un autre type d'approche est celui qui est représenté par le projet THESUS [7] où les concepts d'une ontologie sont pondérés et où les comparaisons entre ontologies pondérées sont ensuite faites en utilisant la mesure de Jaccard³, qui est une mesure ensembliste qui ne tient pas compte des proximités sémantiques entre concepts, et la distance de Wu et Palmer [13] qui permet d'indiquer la proximité entre concepts. Ainsi, dans THESUS, si un document et une requête utilisent des concepts proches mais pas identiques, on regarde leur similarité pour donner une estimation de la proximité entre les deux ontologies pondérées.

Nous estimons que cette dépendance des caractérisations sémantiques vis-à-vis de la mesure de proximité choisie par un système est difficile à concevoir. Et ce, non seulement pour le choix de la mesure de proximité entre concepts, mais aussi parce que cette utilisation d'une mesure de proximité ne permet pas à un agent d'assumer sa saisie : ne pas pondérer un concept doit être por-

²Il s'agit de la fonction *term-frequency*inverse-document-frequency*.

$$S_{Jaccard} = \frac{A \cap B}{A \cup B}$$

teur de sens.

De plus, ces mesures ont une complexité de l'ordre de $O(n^2)$, puisque, pour chaque concept, il faut regarder sa proximité à l'ensemble de tous les autres concepts afin de pouvoir donner la mesure finale.

2.3 La mesure du cosinus

En ce qui concerne la mesure de proximité entre focus, il y a deux visions possibles. La première est celle que nous avons vue avec le projet THE-SUS précédent, dans lequel la proximité entre concepts est prise en compte et qui engendre un manque d'autonomie. La seconde utilise les pondérations présentes sur les concepts. Elle est représentée par la mesure de Jaccard ou par le cosinus. Dans les deux cas, les mesures sont des distances⁴. Or nous avons noté grâce aux travaux de [12] en psychologie qu'une mesure de similarité ne doit pas respecter symétrie et inégalité triangulaire. Ainsi il est plus aisé de juger que « la Corée du Nord est proche de la Chine » que l'inverse. De même, si nous considérons qu'un document est pertinent pour une requête donnée, il paraît difficile d'affirmer que cette requête est elle-même toujours aussi pertinente pour le document.

Nous cherchons donc à définir une mesure ne satisfaisant pas les propriétés de symétrie et d'inégalité triangulaire et permettant d'indiquer en quoi un focus est pertinent par rapport à un autre.

3 Focus

Intuitivement, un focus est simplement la pondération des concepts d'une ontologie dans le but de qualifier une requête, une description de capacité ou de préférence. Ceci nous amène à préciser le sens que nous donnons au terme *ontologie*.

3.1 Ontologie

Les ontologies servent à représenter des connaissances dans les systèmes informatiques [6]. Il s'agit avec elles de rendre explicites les conceptualisations partagées et tacitement admises dans différents domaines. L'ampleur de ce qui est formalisé n'est pas toujours identique : il peut s'agir de connaissances lexicographiques, conceptuelles ou même sémantiques,

⁴C'est-à-dire qu'elles respectent les propriétés de symétrie, d'inégalité triangulaire et de minimalité.

c'est-à-dire que nous pouvons être à un niveau non formel ou au niveau d'ontologies légères ou lourdes. FÜRST [6] présente différents niveaux d'engagement sémantique et de formalisme (cf. figure 1). Par engagement sémantique, il est entendu le degré d'intégration des connaissances du domaine dans l'ontologie. Notre positionne-

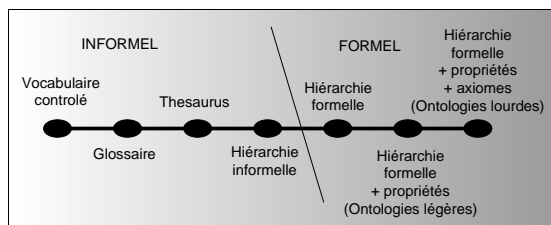


FIG. 1 – Les différents niveaux de formalisme et d'engagement sémantique en représentation des connaissances.

ment au niveau ontologique reste humble dans un premier temps : nous supposons avoir accès à des connaissances terminologiques ou conceptuelles, et à la relation hiérarchique *IS-A*. Word-Net [10] est ainsi pour une grande partie de la littérature du domaine de la recherche d'information un outil suffisant, bien que n'étant en fait qu'un thésaurus ; en effet il n'y est pas question de concepts mais de listes de synonymes.

3.2 Définition d'un focus en général

Un focus est une application à valeurs réelles définie sur une ontologie Ω . Soit \mathcal{C}_Ω l'ensemble des concepts de l'ontologie :

$$\forall c_i \in \mathcal{C}_\Omega, \vec{f}[c_i] \in [\min, \max] \quad (1)$$

min et *max* sont les bornes des valeurs réelles que peut prendre chaque concept de l'ontologie. Nous avons choisi une notation vectorielle : $\vec{f}[i]$ est la valeur du *i*-ème concept du focus \vec{f} ⁶.

3.3 Propagation

Nous avons vu (cf. 2.2) que nous laissons les agents assumer leur caractérisations. Néanmoins, la mise en œuvre de cette dernière approche nécessite de développer une technique

⁵Pour un exposé des raisons pratiques et théoriques sur le fait que la relation de subsomption est suffisante, voir [3].

⁶La question qui se pose est celle de l'ordre des concepts dans l'ontologie, permettant d'identifier de façon uniforme un concept à son numéro. Nous supposons un parcours de l'ontologie permettant d'obtenir les numéros de chaque concept, par exemple un parcours en profondeur à partir de la racine.

de propagation des valeurs, pour faciliter la saisie des focus. Intuitivement, les concepts considérés comme proches des concepts initialement pondérés doivent eux aussi recevoir une pondération dépendant à la fois de leur proximité et de la valeur affectée au concept initial (cf. figure 2). Par exemple, si un utilisateur pondère le concept de *désert* dans sa requête, alors il est possible que le concept de *Sahara* ait aussi à l'être. Ainsi, nous avons besoin :

- d'une mesure nous permettant de classer les différents concepts par rapport à leur proximité vis-à-vis d'un concept initialement pondéré, pris comme concept « pivot »
- d'une fonction de propagation qui permette d'indiquer la valeur des concepts de l'ontologie en fonction de leur proximité par rapport au concept pivot.

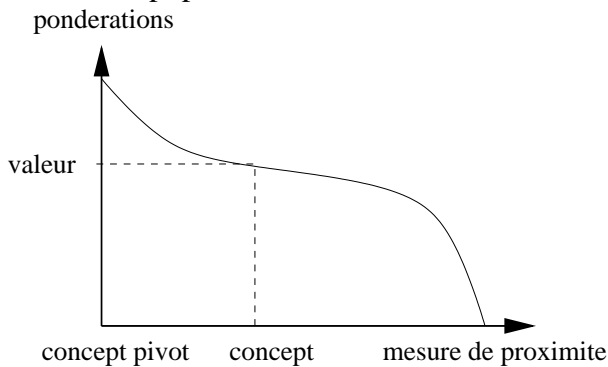


FIG. 2 – Problème de la propagation des valeurs dans une ontologie.

Dans [11] nous présentons une mesure de proximité entre concepts dans une ontologie inspirée de [1]. Contrairement aux mesures que nous pouvons trouver dans la littérature, elles ne respectent pas les propriétés de symétrie et d'inégalité triangulaire : voir table 1. Or c'est ce que nous voulons obtenir si nous suivons les réflexions de [12] en psychologie (cf. 2.3).

| Propriété \ Mesure | 1 | 2 | 3 | 4 | 5 | 6 |
|-----------------------------------|-----|-----|-----|-----|-----|-----|
| augmentation avec les similarités | oui | non | oui | oui | oui | oui |
| diminution avec les différences | non | oui | oui | oui | oui | oui |
| symétrie | oui | oui | oui | oui | non | non |
| inégalité triangulaire | non | oui | non | non | non | non |

TAB. 1 – Comparaison entre différentes mesures de similarité, inspiré de [2] : 1-RESNIK, 2-RADA, 3-WU et PALMER, 4-LIN, 5-BIDAULT et 6-notre mesure.

La question est maintenant de donner des valeurs plus ou moins fortes aux concepts voisins

du concept pivot en s'appuyant sur le classement des concepts du plus similaire au plus éloigné que nous venons de mettre en place.

Nous avons utilisé dans nos expérimentations des fonctions linéaires par morceaux type logique floue. Graphiquement nous avons donc la représentation de la figure 3, avec comme exemple une fonction de propagation pour $f_{0.2,0.5,1}$.

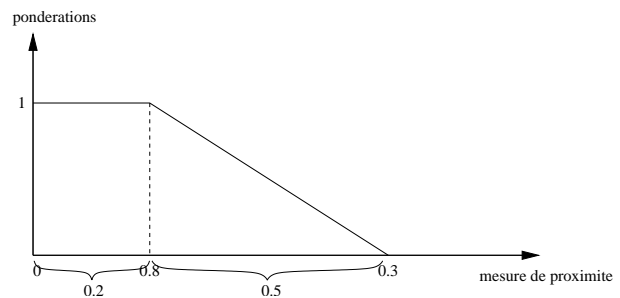


FIG. 3 – Représentation graphique pour $f_{0.2,0.5,1}$.

Si nous propageons à partir de plusieurs concepts pivots, nous pouvons arriver dans une situation où un même concept peut recevoir plusieurs valeurs. Les deux possibilités d'attribution d'une valeur unique sont la proximité la plus forte aux concepts pivots et la valeur maximum des différentes fonctions de propagation. Notre choix⁷ s'est porté sur la première procédure.

4 Normalisation et comparaison

Les focus bruts que nous avons définis précédemment sont difficilement comparables, en particulier à cause de problèmes d'échelle. C'est pourquoi nous devons normaliser avant de comparer.

4.1 Normalisation de focus

Dans [11], nous avons extrait deux manières de normaliser différentes : soit en considérant la valeur maximale d'un concept puis en ramenant à une constante et en ramenant proportionnellement toutes les valeurs du focus ; soit en considérant la somme des valeurs du focus, en la normalisant à une constante et en ramenant les valeurs de chaque concept proportionnellement.

⁷Cf. [11] et [8].

Cela doit nous permettre de comparer des focus entre eux, quelles que soient les manières de les saisir, que la valeur maximale soit de 10.000 pour l'un et de 50 pour l'autre, que la somme des pondérations soit de 10 pour l'un et de 300 pour l'autre. Les notations que nous avons introduites sont les suivantes :

- f^{max} pour la normalisation par le maximum d'un focus \vec{f} ;
- f^{Σ} pour la normalisation par la somme d'un focus \vec{f} .

Pour normaliser par le maximum (soit k_{max} ce maximum) un focus quelconque, nous avons la relation suivante :

$$\vec{f}^{max}[i] = \begin{cases} \frac{k_{max}}{\max_i \vec{f}} \vec{f}[i] & \text{si } \vec{f}[i] \neq 0 \\ 0 & \text{sinon} \end{cases}$$

De même en ce qui concerne la normalisation par la somme (soit k_{Σ} la somme de valeurs) :

$$\vec{f}^{\Sigma}[i] = \begin{cases} \frac{k_{\Sigma}}{\sum_i \vec{f}[i]} \vec{f}[i] & \text{si } \vec{f}[i] \neq 0 \\ 0 & \text{sinon} \end{cases}$$

Il est de plus trivial que nous avons les équivalences suivantes : $(\vec{f}^{\Sigma})^{\Sigma} = \vec{f}^{\Sigma}$, $(\vec{f}^{max})^{max} = \vec{f}^{max}$, $(\vec{f}^{\Sigma})^{max} = \vec{f}^{max}$, $(\vec{f}^{max})^{\Sigma} = \vec{f}^{\Sigma}$.

4.2 Pertinence d'un focus par rapport à un autre

Prenant en compte les réflexions de [12] citées plus haut (2.3), nous avons cherché à mettre en place une mesure de pertinence relative. Nous avons donc un focus pivot, celui d'une requête par exemple, et des focus cibles, ceux de documents, que nous allons comparer. Soit \vec{f}_1 le focus pivot et \vec{f}_2 le focus cible :

$$PR(\vec{f}_1, \vec{f}_2) = \sum_i (comm(\vec{f}_1[i], \vec{f}_2[i]) - diff(\vec{f}_1[i], \vec{f}_2[i]))$$

avec :

$$comm(\vec{f}_1[i], \vec{f}_2[i]) = \min(\vec{f}_1^{max}[i], \vec{f}_2^{max}[i]) \times \vec{f}_2^{\Sigma}[i]$$

et :

$$diff(\vec{f}_1[i], \vec{f}_2[i]) = \begin{cases} (\vec{f}_1^{max}[i] - \vec{f}_2^{max}[i]) \times \vec{f}_1^{\Sigma}[i] \\ \text{si } \vec{f}_1^{max}[i] > \vec{f}_2^{max}[i] \\ 0 & \text{sinon} \end{cases}$$

Pour plus de détails, voir [11].

5 Premières expérimentations

Nos premiers tests ont cherché à mettre en valeur la rapidité de la comparaison de focus. En effet, une série d'approches utilisent des fonctions ayant une complexité de l'ordre de $O(n^2)$. Nous sommes évidemment en $O(2n)$ et nous montrons par l'exemple la vitesse de nos algorithmes.

| | 100 | 100000 | 500000 |
|---------|------|--------|--------|
| test1 | 19 | 1641 | 15787 |
| test2 | 13 | 1335 | 7299 |
| test3 | 26 | 914 | 3775 |
| test4 | 20 | 847 | 2051 |
| test5 | 77 | 1657 | 1657 |
| test6 | 12 | 1049 | 1049 |
| moyenne | 27.8 | 1240.5 | 5263.6 |

TAB. 2 – Tests de comparaisons de focus. Les temps sont en millisecondes, les colonnes correspondent à 100, 100000 et 500000 focus.

Nous avons mis en place une librairie Java et une interface de saisie et de gestion de focus, FMP⁸. Elle nous offre la possibilité de lire les ontologies sauvegardées en XML par Protégé, de générer des ontologies à partir de Wordnet situées entre deux synsets, de saisir des focus sur ces ontologies, de faire des comparaisons entre focus, etc. A partir de cette librairie, nous avons mesuré le temps qu'il nous fallait pour comparer un focus à une série de 100, 100000 ou 500000 autres focus. Chacun des focus est basé sur une même ontologie à 384 concepts et nous choisissons au hasard 2 concepts pivots à partir desquels nous mettons en place une propagation $f_{0.15,0.3,1}$. La machine sur laquelle les tests ont eu lieu est un Pentium III à 1 Ghz, avec 256 Mo de mémoire vive (SDRAM). Le tableau 2 et la figure 4 récapitulent ces résultats. Nous pouvons voir que la progression est linéaire et les temps de traitement faibles (5 secondes en moyenne pour comparer 500000 focus).

6 Conclusion

Pour rendre l'échange d'information entre agents pertinente, nous avons mis en place des pondérations de concepts d'une ontologie : les focus. Partant du principe que ce sont les agents qui doivent assumer les valeurs placées sur les

⁸Focus Management Program : voir <http://www.sciences.univ-nantes.fr/info/perso/permanents/lamarre-/recherche/focus/focus/>

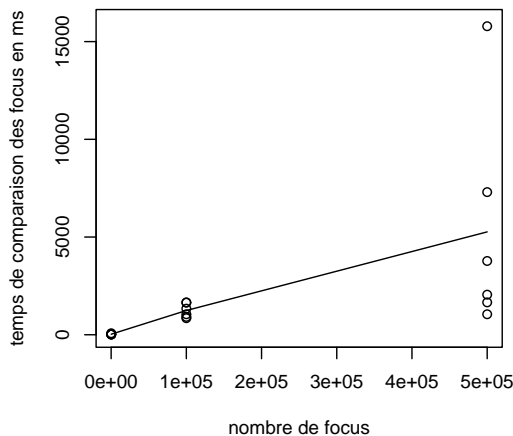


FIG. 4 – Valeurs de comparaisons des focus. Chaque point correspond à un test, et la courbe relie les moyennes.

concepts, il est nécessaire qu'ils mettent des valeurs sur tous les concepts qui sont pour eux intéressants. Dans ce cadre, nous avons défini une fonction de propagation des valeurs dans un focus.

Nous nous sommes ensuite attachés à proposer une mesure de pertinence relative d'un focus par rapport à un autre : par exemple un document par rapport à une requête, un service par rapport à un besoin, etc. Pour ce faire, nous avons défini des formes normales de focus que nous avons ensuite utilisées dans une mesure qui respecte les propriétés d'une mesure de similarité telle que définie en psychologie pour les humains.

Nous avons de plus développé une application de saisie, de gestion et de manipulation des focus sur des ontologies.

Références

- [1] Alain BIDAULT. *Affinement de requêtes posées à un médiateur*. PhD thesis, Université Paris XI, Orsay, Paris, France, jul 2002.
- [2] Dekang LIN. An information-theoretic definition of similarity. In *Proc. 15th International Conf. on Machine Learning*, pages 296–304. Morgan Kaufmann, San Francisco, CA, 1998.
- [3] Ray RICHARDSON, Alan F. SMEATON, and J. MURPHY. Using WordNet as a

knowledge base for measuring semantic similarity between words. Technical Report CA-1294, Dublin City University, Dublin, Ireland, 1994.

- [4] Michael W. BERRY, Zlatko DRMAC, and Elizabeth R. JESSUP. Matrices, vector spaces, and information retrieval. *SIAM Rev.*, 41(2) :335–362, 1999.
- [5] S. CAZALENS, E. DESMONTILS, C. JACQUIN, and P. LAMARRE. Sources d'informations et de connaissances : de la gestion locale la recherche distribuée. *RSTI, L'Objet*, 8(4), Nov. 2002.
- [6] Frédéric FÜRST. *Contribution à l'ingénierie des ontologies, une méthode et un outil d'opérationnalisation*. PhD thesis, Université de Nantes, France, 2004.
- [7] Maria HALKIDI, Benjamin NGUYEN, Iraklis VARLAMIS, and Michalis VAZIRGIANNIS. Thesus : Organizing web document collections based on link semantics. *The VLDB Journal*, 12(4) :320–332, 2003.
- [8] Géraldine JEAN and Jean-Marie FAVREAU. Expression de focus sur des ontologies pour traiter les demandes des utilisateurs dans un système multi-agent dédié à la recherche d'information. Mémoire de maîtrise d'informatique, Université de Nantes, France, 2004.
- [9] Thomas K. LANDAUER, Peter W. FOLTZ, and Darrell LAHAM. Introduction to latent semantic analysis. *Discourse Processes*, 25 :259–284, 1998.
- [10] Georges A. MILLER, Richard BECKWITH, Christiane FELBAUM, Derek GROSS, and Katherine MILLER. Introduction to wordnet : an on-line lexical database. *International journal of lexicography*, 3(4) :235–244, 1990.
- [11] Anthony VENTRESQUE. Focus et ontologie pour la recherche d'information. Mémoire de DEA d'informatique, Université de Nantes, France, 2004.
- [12] Amos TVERSKY. Features of similarity. *Psychological Review*, 84(4) :327–352, July 1977.
- [13] Zhibiao WU and Martha PALMER. Verb semantics and lexical selection. In *32nd. Annual Meeting of the Association for Computational Linguistics*, pages 133–138, New Mexico State University, Las Cruces, New Mexico, 1994.