

Distances sémantiques dans des applications de gestion d'information utilisant le web sémantique.

Fabien Gandon*, Olivier Corby*, Ibrahima Diop*, Moussa Lo**

*INRIA Edelweiss, 2004 rt des Lucioles BP93, 06902 Sophia Antipolis
Fabien.Gandon@sophia.inria.fr
<http://www-sop.inria.fr/edelweiss/>

**LANI, UFR SAT, Université Gaston Berger, Saint-Louis, Sénégal
lom@ugb.sn
<http://www.ugb.sn/>

Résumé. Nous résumons ici divers utilisations que nous avons faites des distances sémantiques et discutons notre position sur ce domaine. Nous présentons en suite un travail préliminaire sur l'extension de leur utilisation dans des applications reposant au minimum sur des modèles RDF/S.

1 Introduction : intuitivement proches

Intuitivement nous sommes tous portés à dire que le concept de *berline* est plus proche du concept de *monospace* que de celui d'*avion* ; cependant nous pensons aussi que le concept de *berline* est plus proche du concept d'*avion* que du concept de *livre*. Ces distances intuitives peuvent être simulées par exemple pour améliorer les moteurs de recherche du Web dans leurs algorithmes de filtrage et de tri des réponses.

En informatique, une ontologie est une théorie logique partielle rendant explicite une conception de la réalité (Gruber, 93) (Guarino et Giaretta, 1995). Les définitions en intension d'une ontologie sont donc naturellement traduites en des représentations logiques exploitées notamment dans des inférences de dérivation, par exemple pour améliorer le rappel en recherche d'information. Cependant, ces mêmes définitions et leurs relations peuvent être vues comme des espaces, notamment des graphes ou réseaux sémantiques, qui peuvent être dotés de métriques servant de base à toute une autre gamme d'inférences.

L'idée d'évaluer la proximité conceptuelle sur des réseaux sémantiques remonte aux travaux de (Quillian, 1968) et (Collins & Loftus, 1975) sur la mémoire sémantique humaine. La proximité de deux concepts peut venir d'une complémentarité fonctionnelle (ex: un clou et un marteau), d'une similarité fonctionnelle (ex: un marteau et un tournevis), etc. Ce dernier exemple appartient à la famille des similarités sémantiques dans laquelle la proximité est basée sur une caractéristique définitionnelle partagée (ex: être un outil).

Une structure supportant naturellement le raisonnement sur les similarités sémantiques est la hiérarchie des types telle que l'on peut la trouver dans un support en graphes conceptuels, dans la TBox des logiques de description, dans un schéma RDFS, etc. En effet, dans cette structure, les liens de subsomption groupent les types suivant les caractéristiques définitionnelles qu'ils partagent. Lorsqu'elle est appliquée au graphe d'une hiérarchie, une proximi-

Distances sémantiques dans des applications utilisant le web sémantique.

té calculée par propagation donne une distance sémantique, la première et la plus simple étant celle qui compte les arcs séparant deux sommets (Rada *et al.*, 1989).

Deux grandes familles d'approches peuvent être identifiées pour le calcul de telles distances: (1) celles qui incluent des informations externes à la hiérarchie, par exemple, des statistiques sur l'utilisation des types de concepts (Resnik, 1995) (Jiang & Conrath, 1997), et (2) les approches reposant uniquement sur les modèles en intensions comme par exemple une hiérarchie de types (Rada *et al.*, 1989)(Wu et Palmer, 1994). Dans le domaine des graphes conceptuels la deuxième approche est utilisée, en particulier pour proposer une projection ne donnant plus uniquement des valeurs booléennes *i.e.* une similarité $S:C^2 \rightarrow [0,1]$ où 1 correspond à la valeur vraie de la projection classique et toute autre valeur donne une idée de la similarité entre le graphe projeté et le graphe source. L'utilisation initiale faite par Sowa visait à permettre des déplacements latéraux dans le treillis des types. (Ralescu et Fadlalla, 1990) l'ont utilisée pour relaxer les contraintes de l'opérateur de jointure. Plus récemment, (Zhong *et al.*, 2002) ont utilisé une distance atténuée par la profondeur des types dans l'ontologie pour construire une mesure de similarité entre graphes conceptuels.

Dans la suite, nous donnons un aperçu des applications et des travaux que nous menons autour de cette notion de proximité ou distance sémantique à travers une opérationnalisation du web sémantique basée sur les graphes conceptuels.

2 Expériences passées pour des distances sur web sémantique

Nous résumons ici les scénarios d'utilisation et les distances utilisées dans un échantillon représentatif de projets que nous avons menés. Pour nous, cette première série d'expériences a permis de démontrer l'intérêt et le potentiel des distances, et aussi de souligner l'importance du travail restant à faire pour identifier et caractériser les familles de distances existantes et leur adéquation respective aux tâches pour lesquelles nos utilisateurs souhaitent être assistés. Trois expériences sont présentées chronologiquement et discutées dans la quatrième section.

2.1 Distance et bases de connaissances distribuées

Dans un web sémantique d'entreprise les scénarios amènent souvent la contrainte de bases d'annotations distribuées (assertions RDF¹ à propos de ressources intranet). Pour gérer cette distribution nous avons proposé une architecture et des protocoles permettant en particulier de maintenir la spécialisation des bases d'annotations quant aux sujets abordés dans leurs assertions (Gandon, 2002).

Chaque archive de notre architecture maintient une structure appelée ABIS (Annotation Base Instances Statistics) décrivant des statistiques sur les types de triplets (relations binaires imposées par le modèle RDF) présents dans leur base d'annotations. Par exemple si, dans l'ontologie, il existe une propriété Auteur avec la signature:

[Document] → (Auteur) → [Personne]

L'ABIS pourra contenir des statistiques sur l'existence des instances de triplets suivantes:

[Article] → (Auteur) → [Etudiant]

[Livre] → (Auteur) → [Philosophe]

...

¹ <http://www.w3.org/RDF/>

L'ABIS est construit lors de la transformation des annotations RDF en graphes conceptuels et capture la contribution d'une archive à la mémoire globale en terme de types de connaissances contenues dans cette archive. L'ABIS fournit un moyen de comparer le contenu de deux bases et nous l'utilisons pour maintenir la spécialisation des bases d'annotations grâce à une distance sémantique définie entre un ABIS et une nouvelle annotation.

Pour comparer deux types primitifs, nous utilisons la distance de (Rada *et al.*, 1989) comptant le nombre d'arcs sur le chemin le plus court qui relie ces deux types à travers la hiérarchie; voir formule (1). En utilisant cette distance on peut définir une distance entre deux triplets RDF (ou deux instances d'une relation binaire), comme étant la somme des distances entre: les types des deux relations, les types des deux concepts en premier argument (domain) et les types des deux concepts en deuxième argument (range); voir formule (2). La distance entre un triplet et un ABIS est alors définie comme la distance minimale entre ce triplet et les triplets recensés par l'ABIS; voir formule (3). Et finalement, la distance entre une annotation et un ABIS est la somme des distances entre chaque triplet de l'annotation figurant dans l'ABIS; voir formule (4).

$$dist(t_1, t_2) = length(t_1, lcst(t_1, t_2)) + length(t_2, lcst(t_1, t_2)) \quad (1)$$

où, $lcst(t_1, t_2)$ est le plus proche supertype commun de t_1 et t_2 .

$$dist(triple1, triple2) = dist(domain(triple1), domain(triple2)) + dist(predicate(triple1), predicate(triple2)) + dist(range(triple1), range(triple2)) \quad (2)$$

$$dist(triple, ABIS) = \min_{triple_i \in ABIS} (dist(triple, triple_i)) \quad (3)$$

$$dist(An, ABIS) = \sum_{triple_j \in An} dist(triple_j, ABIS) \quad (4)$$

Cette distance donne une fonction d'évaluation / fonction de coût utilisée comme critère dans un protocole de mise aux enchères des nouvelles annotations à archiver : chaque nouvelle annotation est mise aux enchères entre les archives existantes; chaque archive fait une offre qui correspond à la distance entre son ABIS et l'annotation; l'archive avec l'offre la plus petite gagne l'annotation. Ce protocole permet de maintenir la spécialisation des bases et ainsi de faciliter l'optimisation de la résolution de requêtes distribuées en utilisant les ABIS pour la décomposition et le routage des projections.

Dans ce premier exemple, la définition d'une distance conceptuelle sur la hiérarchie des types permet de construire un consensus calculatoire (distance) au dessus du consensus ontologique (hiérarchie), et de l'utiliser dans un consensus protocolaire (enchères). Initialement utilisée pour un protocole de mémoire distribuée, la section suivante explique comment cette distance a ensuite été intégrée au moteur de recherche de chaque base pour proposer une nouvelle fonctionnalité: la recherche approchée de connaissances par relaxation des contraintes de typage.

Distances sémantiques dans des applications utilisant le web sémantique.

2.2 Distance et projection de graphes approchée

La plateforme CORESE (Corby et al, 2006) intègre une fonctionnalité de recherche approchée qui démontre une autre application des inférences simulant la proximité conceptuelle. CORESE utilise une extension de la distance atténuée par la profondeur (Zhong *et al.*, 2002) des types dans le treillis de l'ontologie; voir formules (5) et (6).

$$\forall (t_1, t_2) \in H_c^2; t_1 \leq t_2 \text{ on a } l_{H_c}(t_1, t_2) = \sum_{\{t \in \langle t_1, t_2 \rangle, t \neq t_1\}} \left[\frac{1}{2^{\text{depth}(t)}} \right] \quad (5)$$

avec H_c la hiérarchie des types de concepts, le chemin le plus court entre t_1 et t_2 , et $\text{depth}(t)$ la profondeur de t dans l'ontologie *i.e.* le nombre d'arcs sur le chemin le plus court entre t et la racine T

$$\forall (t_1, t_2) \in H_c^2 \text{ on a } \text{dist}(t_1, t_2) = \min_{\{t \geq t_1, t \geq t_2\}} (l_{H_c}(t_1, t) + l_{H_c}(t_2, t)) \quad (6)$$

En utilisant cette distance, on peut relaxer la contrainte d'égalité ou de spécialisation des types lors de la projection en la remplaçant par une contrainte de proximité utilisant une distance conceptuelle comme celle définie en (6). On obtient alors une projection approchée pour l'appariement de graphes requêtes (ex: un motif recherché par un utilisateur) et de graphes faits (ex: une annotation RDF) ; une telle projection préserve l'adjacence et l'ordre des arcs mais permet de relaxer les contraintes de typage. Dans ce deuxième exemple, la distance est utilisée pour remplacer une contrainte logique ($t_1(x) \Rightarrow t_2(x)$) par une contrainte numérique ($d(t_1, t_2) < \text{seuil}$). Cette transformation permet de relaxer une requête donnée par un utilisateur lorsqu'elle ne donne pas (suffisamment) de résultats. Pour accéder aux connaissances d'une base, une alternative aux requêtes est la navigation dans les annotations. La section suivante montre comment, là encore, les distances peuvent être utilisées.

2.3 Distance et ultra-métrie de clustering

Dans le cadre du projet KmP nous nous sommes intéressés à la construction d'un algorithme de regroupement (clustering) des compétences présentes sur la Télécom Valley de Sophia Antipolis et annotées en RDF (Gandon *et al.*, 2006). Le regroupement normalement effectué manuellement par les experts en gestion s'est révélé être un algorithme de regroupement monothétique (monothetic clustering). La représentation recherchée demandait de pouvoir fournir des moyens de contrôler simplement le niveau de détail et de granularité choisi pour générer le regroupement. En analyse de données (Jain *et al.*, 1999), une structure classique supportant le choix des niveaux de détail est le dendrogramme, un arbre qui, à chaque niveau de coupure, donne une solution de regroupement plus ou moins fin.

Un dendrogramme repose sur une ultra-métrie c'est-à-dire une distance avec une inégalité triangulaire sur-contrainte : $\forall t', t_1, t_2 \in H_c^3 \text{ dist}(t_1, t_2) \leq \max(\text{dist}(t_1, t'), \text{dist}(t_2, t'))$

Nous avons donc cherché à construire cette ultramétrie à partir de la distance sémantique de CORESE. Nous n'avons considéré pour cela que des hiérarchies en arbres simples, ce qui nous donne une distance exacte ayant pour formule (7).

$$\text{dist}(t_1, t_2) = \frac{1}{2^{\text{depth}(\text{lcs}(t_1, t_2)) - 2}} - \frac{1}{2^{\text{depth}(t_1) - 1}} - \frac{1}{2^{\text{depth}(t_2) - 1}} \quad (7)$$

Nous avons ensuite proposé une transformation produisant une ultramétrie et améliorant le nombre de niveaux de détail disponibles dans le dendrogramme obtenu en favorisant le regroupement des classes ayant une descendance peu profonde, formule (8) et figure 1.

$$dist_{CH}(t_1, t_2) = \max_{\forall st \leq lcst(t_1, t_2)} (dist(st, lcst(t_1, t_2))) \text{ quand } t_1 \neq t_2 \quad (8)$$

$$dist_{CH}(t_1, t_2) = 0 \text{ quand } t_1 = t_2$$

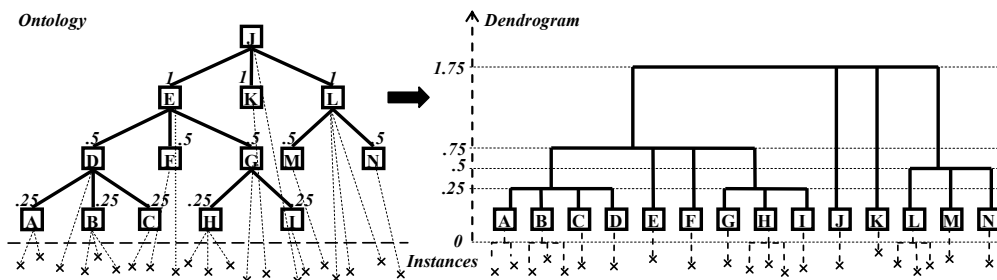


Figure 1. Transformation de la distance en ultramétrique.

Le regroupement suivant la hiérarchie des types, on peut nommer chaque regroupement. Un exemple de regroupement des compétences sur la Télécom Valley est donné en Figure 2.

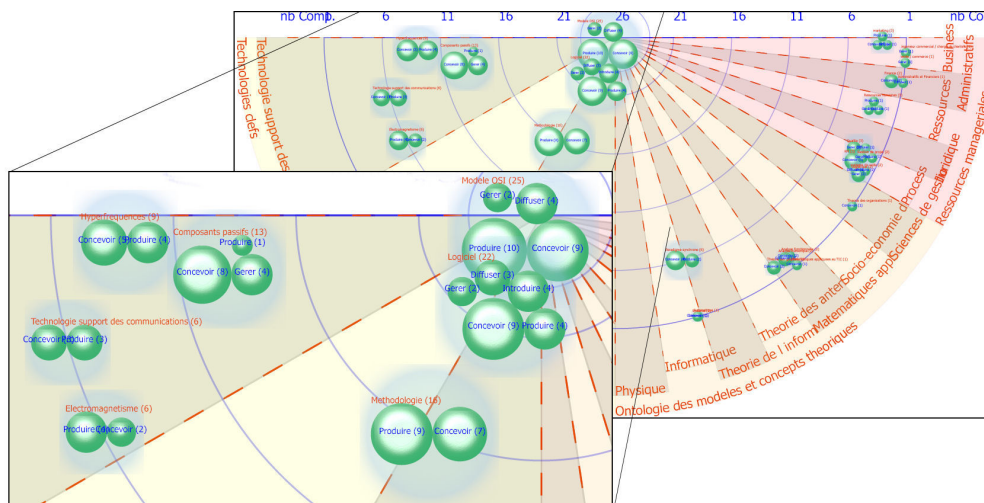


Figure 2. Vue en Radar sur 180° des regroupements de compétences.

2.4 Une première discussion sur les distances et leur utilisation

En parallèle avec ces trois premières explorations des caractéristiques, interprétations et applications des distances conceptuelles, nous avons commencé à questionner la valeur de ces distances et leur fidélité par rapport aux proximités naturellement ressenties par les humains. Pour cela nous avons commencé une étude empirique et statistique. La première hypothèse testée fut *"Est-il juste de considérer que les frères sont à égale distance du père et à égale distance les uns des autres ou est-ce un effet secondaire du fait que l'on repose sur la structure des chemins de subsumption?"*.

Afin d'étudier ces distances dans leur milieu naturel et de les comparer avec leurs simulations informatiques, nous avons conçu une plateforme permettant de réaliser, gérer, et d'ana-

Distances sémantiques dans des applications utilisant le web sémantique.

lyser des expériences où les participants organisent et regroupent spatialement des concepts selon leur proximité intuitive (Boutet *et al.*, 2005); voir figure 3.

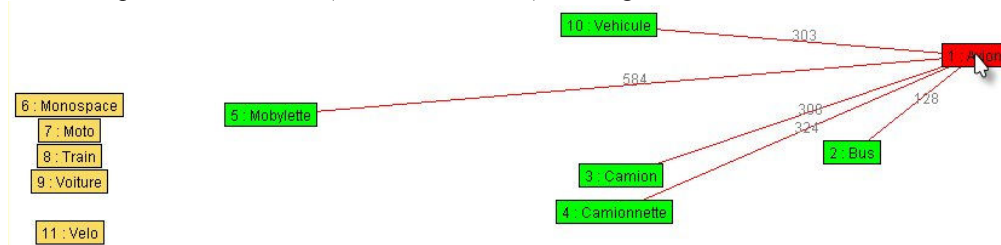


Figure 3. Applet de l'exercice de placement

A partir de ces exercices, des analyses statistiques sont faites. Prenons l'exercice de la figure 3 effectué par 30 participants de 13 à 50 ans. Les distances capturées ont été normalisées avant d'en calculer la moyenne, l'écart type et la variance. La figure 4 montre les distances entre le concept *camion* et la liste des autres concepts à placer.

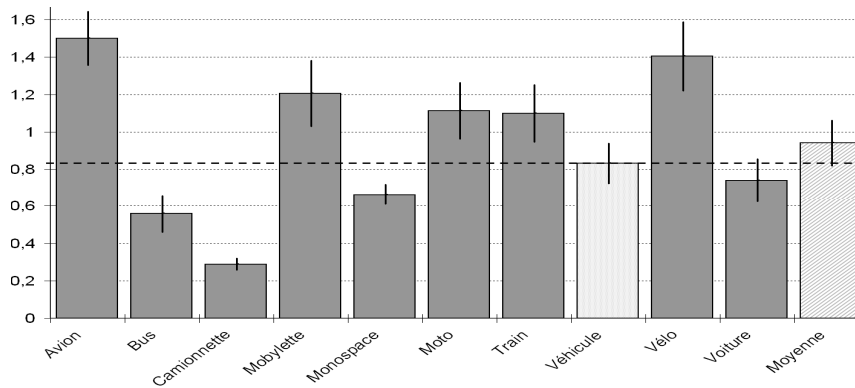


Figure 4. Distances entre Camion et d'autres véhicules

On peut lire sur ce graphique que le concept *camionnette* est en moyenne très proche de *camion* et, étant donné que la variance est très faible, qu'il s'agit d'un consensus. Le concept *véhicule* est particulier dans cette liste puisque dans une ontologie il serait naturellement placé comme père (ou ancêtre) des autres concepts et que ces autres concepts seraient entre eux des frères (ou des cousins). On voit notamment que la distance entre *camion* et ses frères est parfois plus petite (4 cas) et parfois plus grande (5 cas) que la distance à son père ou ancêtre *véhicule* ce qui serait impossible à faire dans un hiérarchie de types classique comme celles utilisées dans la majorité des distances puisque l'on passe forcément par le père pour aller à un frère. Il reste beaucoup à faire dans cette étude mais si ces résultats se confirmaient, ils montreraient qu'une structure de subsomption seule ne permet pas de simuler de tels comportements.

3 Expériences en cours d'extension des distances

Nous sommes convaincus que la tâche devant être assistée par un système conditionne la ou les distances employées et leur combinaison éventuelle. Notre démarche actuelle est donc

d'essayer les différents espaces métriques disponibles. Dans le cadre de plusieurs projets en cours, nous explorons actuellement de nouvelles distances ou extensions de distances. Les sections suivantes donnent des descriptions préliminaires de ces travaux en cours. Nous l'avons vu en introduction, il existe deux grandes approches pour calculer les distances: utiliser des informations extérieures aux modèles des connaissances (ex: statistiques sur des corpus de textes) ou reposer uniquement sur la structure des modèles (ex: la hiérarchie des types). Nous observons actuellement comment ces approches se déclinent sur l'utilisation d'informations extérieures (extraites d'une base d'annotations RDF) et sur des extensions de la structure habituellement exploitée dans les modèles (ici la hiérarchie de types en RDFS).

3.1 Distance de cooccurrence et contexte en extension

Notre premier essai consiste à considérer la proximité d'usage de deux types *i.e.* la fréquence avec laquelle ces deux types sont employés ensemble dans des descriptions. Soient deux types de concepts $t_x, t_y \in H_c^2$ on définit le comptage de cooccurrences comme le nombre de triplets impliquant ces deux types:

$$\text{count}(t_x, t_y) = || \{t \text{ triplets RDF} \mid t = (x, T_p, y) \wedge (x, \text{rdf:type}, T_x) \wedge (y, \text{rdf:type}, T_y)\} ||$$

On définit alors une distance inverse:

$$\text{dist}_{\text{count}}(T_x, T_y) = \frac{1}{1 + \text{count}(T_x, T_y)} \in [0,1]$$

Cette distance capture une proximité d'usage des types. En effet, comme nous le disions en section 2.1, une même signature de relation peut engendrer bien des familles de triplets par spécialisation des types spécifiés. De plus, détail important, la signature en RDFS est utilisée pour de l'inférence de type (ajout supplémentaire de types) et non de la validation de type ; par conséquent la variété effective des types sur les instances desquels une propriété est utilisée peut être bien plus grande que la liste des types obtenue par la fermeture transitive de la subsumption de ses domaine et co-domaine.

Une utilisation immédiate de cette distance est d'assister les utilisateurs dans des interfaces de navigation (ex: tag cloud figure 5), de requête ou d'annotation en suggérant des types co-occurents au dernier type sélectionné (figure 6).



Figure 5. Nuages de termes obtenus avec la distance

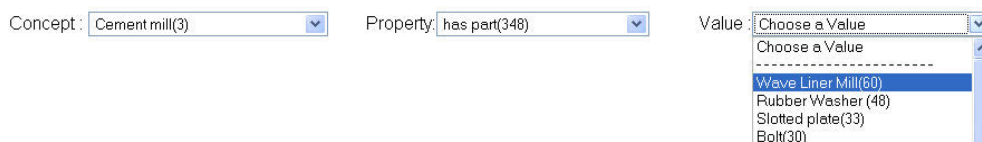


Figure 6. Suggestion dans l'interface en utilisant la distance pour ordonner les options

Distances sémantiques dans des applications utilisant le web sémantique.

3.2 Distance de signatures et contexte en intension

L'ontologie capture des caractérisations en intension des concepts et des relations du domaine, et en particulier des relations en intension entre ces concepts. Nous l'avons vu, l'espace le plus utilisé par les distances à base d'ontologies est le graphe de la hiérarchie des types de concepts. Nous nous intéressons ici à augmenter ce graphe avec les chemins de la hiérarchie des types de relations et les chemins des signatures des relations.

Formellement, on repose sur une première extension du méta-modèle RDFS introduisant une propriété symétrique parente directe du domaine et co-domaine et appelée signature (définition 1).

Définition 1: la propriété `cos:signature` est telle que

$$\begin{aligned} \text{rdfs:domain}(T_p, T_x) &\Rightarrow \text{cos:signature}(T_p, T_x) \\ \text{rdfs:range}(T_p, T_x) &\Rightarrow \text{cos:signature}(T_p, T_x) \\ \text{cos:signature}(T_x, T_y) &\Leftrightarrow \text{cos:signature}(T_x, T_y) \end{aligned}$$

Définition 2: un chemin de signatures $C_S(T_x, T_y)$ entre les types $T_x, T_y \in H_C^2$ est un chemin de T_x à T_y composé exclusivement d'arcs de type `cos:signature` inférés à partir des signatures de $H_R: C_S(T_x, T_y) := \langle T_x, \text{signature}, T_1, \text{signature}, T_2, \text{signature}, \dots, \text{signature}, T_n, \text{signature}, T_y \rangle$

Définition 3: une distance de signature $d_S(T_x, T_y)$ entre les types $T_x, T_y \in H_C^2$ est définie par $d_S(T_x, T_x) := 0$ et $d_S(T_x, T_y) := \min_{\{C_i \in \{C_S(T_x, T_y)\}\}} \text{long}(C_i)$ avec $\text{long}(\langle T_x, \text{signature}, T_1, \text{signature}, T_2, \text{signature}, \dots, \text{signature}, T_n, \text{signature}, T_y \rangle) := n$

Intuitivement, avec cette distance, deux types sont proches s'il y a (en intension) une possibilité de les impliquer dans un graphe d'annotation concis. Par exemple *document* et *pays* peuvent être proches dans une ontologie s'il existe les déclarations suivantes (et les implications par subsomption et symétrie):

$$\begin{aligned} \text{rdfs:domain}(\text{auteur}, \text{document}) &\Rightarrow \text{cos:signature}(\text{document}, \text{auteur}) \\ \text{rdfs:range}(\text{auteur}, \text{personne}) &\Rightarrow \text{cos:signature}(\text{auteur}, \text{personne}) \\ \text{rdfs:domain}(\text{nationalité}, \text{personne}) &\Rightarrow \text{cos:signature}(\text{personne}, \text{nationalité}) \\ \text{rdfs:range}(\text{nationalité}, \text{pays}) &\Rightarrow \text{cos:signature}(\text{nationalité}, \text{pays}) \\ \text{soit } d_S(\text{document}, \text{pays}) &= 3 \end{aligned}$$

Cette première version permet de comprendre le principe mais ne rend pas compte de la subsomption des types de relations et des types de leur signature; en d'autres termes si une relation (ex: titre) spécifie un type (ex: document) dans sa signature, elle inclut les sous-types de ce type (ex: livre, roman, etc.) et cette spécification se propage au sous-type de la relation (ex: sous-titre, titre court, etc). Pour rendre compte de ce point nous avons défini et nous expérimentons la distance suivante (définition 4).

Définition 4: la propriété `cos:sous-type-et-signature` est telle que

$$\begin{aligned} \text{cos:signature}(T_p, T_x) &\Rightarrow \text{cos:sous-type-et-signature}(T_p, T_x) \\ \text{rdfs:subClassOf}(T_p, T_x) &\Rightarrow \text{cos:sous-type-et-signature}(T_p, T_x) \\ \text{rdfs:subPropertyOf}(T_p, T_x) &\Rightarrow \text{cos:sous-type-et-signature}(T_p, T_x) \\ \text{cos:sous-type-et-signature}(T_x, T_y) &\Leftrightarrow \text{cos:sous-type-et-signature}(T_x, T_y) \end{aligned}$$

Les définitions du chemin et de la distance sont des adaptations des définitions 2 et 3 *mutatis mutandis*.

Une utilisation testée actuellement pour cette distance est le pilotage de l'analyse de la langue naturelle utilisée pour lever l'ambiguïté dans la génération automatique des annotations à partir de textes. Il s'agit de prédire les patrons d'annotation les plus probables pour l'extraction de connaissances d'un texte donné en fonction d'une ontologie donnée. La figure 7 montre une expérience dans laquelle on passe en neuf étapes par pondération d'une distance n'utilisant que les chemins de subsomption à une distance n'utilisant que des chemins de signatures. Pour chaque jeu de poids la courbe donne la précision de la désambiguïsation pour 10 termes apparaissant chacun dans 100 documents. Le point intéressant ici est que le meilleur résultat est obtenu pour un espace métrique mixte.

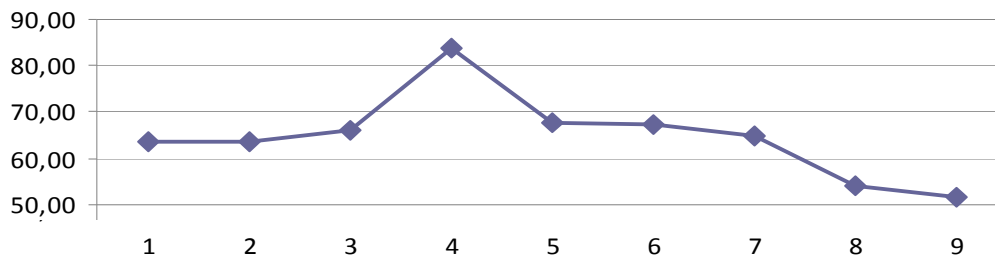


Figure 7. Précision de la désambiguïsation entre subsomption et signature

4 Discussion : intuitivement loin

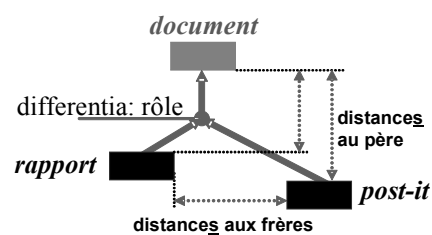
Les distances conceptuelles laissent de nombreuses questions de recherche nourries par un ensemble grandissant d'applications, en particulier:

Proximité naturelle vs. distance mathématique: la distance utilisée par CORESE dans sa recherche approchée est une semi-distance *i.e.* elle ne vérifie pas l'inégalité triangulaire. Quelle est la valeur des conditions nécessaires de la définition mathématique des distances? Devons nous à tout prix essayer de les respecter ou est-ce simplement une limite de la métaphore des distances? Quelles seraient sinon les caractéristiques définitionnelles d'une distance conceptuelle?

Distance conceptuelle vs. espaces métriques:

les structures de la hiérarchie de types sont le terrain favori pour la définition des distances conceptuelles. Faut-il considérer des représentations plus riches (Bachimont, 2000) incluant, par exemple, des liens frère-frère, qui permettraient de mieux simuler de telles distances? Comment mieux recenser et définir ces espaces métriques *i.e.* les espaces pertinents et les métriques adaptées? Comment étudier les différentes familles de distances qui semblent cohabiter dans nos inférences au quotidien? Comment les capturer, les apprendre, pour les utiliser dans des inférences de recherche d'information?

S'il est clair que depuis près de trente ans ces similarités ne cessent de resurgir dans l'exploitation de modèles et d'espaces de représentation, intuitivement, il nous semble qu'il reste encore beaucoup à faire pour identifier, étudier, caractériser et simuler ces similarités, dans la continuité des travaux de Blanchard *et al.*, 2005.



Distances sémantiques dans des applications utilisant le web sémantique.

Références

- Bachimont B., (2000) *Engagement sémantique et engagement ontologique: conception et réalisation d'ontologies en ingénierie des connaissances*, In "Ingénierie des connaissances Evolutions récentes et nouveaux défis", Jean Charlet, Manuel Zacklad, Gilles Kassel, Didier Bourigault; Eyrolles, ISBN 2-212-09110-9
- Blanchard E., Harzallah M., Briand H., Kuntz, P., (2005), *A Typology Of Ontology-Based Semantic Measures*, Workshop EMOI-INTEROP at CAISE'05.
- Boutet, M., Canto, A., Roux, E., (2005) *Plateforme d'étude et de comparaison de distances conceptuelles*, Rapport de Master, Ecole Supérieure En Sciences Informatiques,
- Collins, A., Loftus, E., (1975) *A Spreading Activation Theory of Semantic Processing*. Psychological Review, vol. 82, pp. 407-428,
- Corby O., Dieng-Kuntz R., Faron-Zucker C., Gandon F., (2006) *Searching the Semantic Web: Approximate Query Processing Based on Ontologies*, IEEE Intelligent Systems, January/February (Vol. 21, No. 1), pp. 20-27, ISSN: 1541-1672.
- Gandon F., (2002) *Distributed Artificial Intelligence and Knowledge Management: ontologies and multi-agent systems for a corporate semantic web*, PhD Thesis., INRIA
- Gandon F., Corby O., Giboin A., Gronnier N., Guigard C., (2005) *Graph-based inferences in a Semantic Web Server for the Cartography of Competencies in a Telecom Valley*, ISWC, Lecture Notes in Computer Science LNCS 3729, Galway,
- Gruber, T.R. (1993). *Toward Principles for the Design of Ontologies Used for Knowledge Sharing*, In Formal Ontology in Conceptual Analysis and Knowledge Representation, edited by Nicola Guarino and Roberto Poli, Kluwer Academic Publishers
- Guarino N., Giaretta P., (1995) *Ontologies and Knowledge Bases: Towards a Terminological Clarification*. In N. J. I. Mars (ed.), *Towards Very Large Knowledge Bases*, IOS Press.
- Jain, A.K., Murty, M.N., and Flynn, P.J. (1999) *Data Clustering: A Review*, ACM Computing Surveys, 31(3) 264-323.
- Jiang, J., Conrath, D., (1997) *Semantic Similarity based on Corpus Statistics and Lexical Taxonomy*. In Proc. of International Conference on Research in Computational Linguistics, Taiwan,
- Quillian, M.R., (1968) *Semantic Memory*, in: M. Minsky (Ed.), *Semantic Information Processing*, M.I.T. Press, Cambridge.
- Rada, R., Mili, H., Bicknell, E., Blettner, M., (1989) *Development and Application of a Metric on Semantic Nets*, IEEE Transaction on Systems, Man, and Cybernetics, vol. 19(1), pp. 17-30.
- A.L. Ralescu, A. Fadlalla, (1990) *The Issue of Semantic Distance in Knowledge Representation with Conceptual Graphs*, In Proc. Of AWOC90, pp. 141-142,
- Resnik, P., (1995) *Semantic Similarity in a Taxonomy: An Information-Based Measure and its Applications to Problems of Ambiguity in Natural Language*. In Journal of Artificial Intelligence Research, vol 11, pp. 95-130,
- Sowa, J.F., (1984) *Conceptual structures: Information Processing in Mind and Machine*. Addison-Wesley, Reading, Massachusetts,
- Wu Z, Palmer, M. (1994) *Verb Semantics and Lexical Selection*. In Proc. of the 32nd Annual Meeting of the Association for Computational Linguistics, Las Cruces, New Mexico.
- J. Zhong, H. Zhu, J. Li, Y. Yu. (2002) *Conceptual Graph Matching for Semantic Search*, In Proc. of 10th International Conference on Conceptual Structures, ICCS2002, LNCS 2393, Springer Verlag, pp. 92-106, Borovets, Bulgaria,