



# Ontologies et Recherche d'Information : une application au diagnostic automobile

Axel Reymonet, Jérôme Thomas, Nathalie Aussenac-Gilles

► **To cite this version:**

Axel Reymonet, Jérôme Thomas, Nathalie Aussenac-Gilles. Ontologies et Recherche d'Information : une application au diagnostic automobile. Sylvie Desprès. 21èmes Journées Francophones d'Ingénierie des Connaissances, Jun 2010, Nîmes, France. Ecole des Mines d'Alès, pp.283 à 294, 2010, Ingénierie des connaissances. <hal-00487737>

**HAL Id: hal-00487737**

**<https://hal.archives-ouvertes.fr/hal-00487737>**

Submitted on 30 May 2010

**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.

# Ontologies et Recherche d'Information : une application au diagnostic automobile <sup>\*</sup>

Axel Reymonet<sup>1</sup>, Jérôme Thomas<sup>1</sup>, Nathalie Aussenac-Gilles<sup>2</sup>

<sup>1</sup> Division Technologie ACTIA, 25 Chemin de Pouvoirville 31432 Toulouse Cedex 4  
axel.reymonet@actia.fr, jerome.thomas@actia.fr

<sup>2</sup> IRIT - Équipe IC3, UPS, 118 Route de Narbonne 31062 Toulouse Cedex 9  
aussenac@irit.fr

**Résumé** : Cet article décrit les principes fondateurs et le fonctionnement global de TextViz, outil de recherche d'information (RI) sémantique utilisé dans le domaine du diagnostic automobile. Les bases d'incidents (répertoriant un ensemble de pannes connues) ont toujours été prisées par les constructeurs automobiles et les garagistes : elles permettent d'abord de capitaliser de la connaissance afin de pouvoir y accéder par la suite. Toutefois, avec une architecture des véhicules toujours plus complexe, les causes possibles d'une panne se sont vite multipliées, ce qui rend crucial le processus de RI. Se fondant sur un modèle limité de connaissances du diagnostic automobile, notre logiciel a pour but de faciliter les tâches de stockage et de recherche sémantiques d'information parmi un grand nombre de cas de pannes connus.

**Mots-clés** : communication appliquée, ressource termino-ontologique, indexation et recherche d'information sémantique, dynamique/évolution des connaissances

## 1 Motivations

Au cours d'une session de diagnostic en garage, la méthode la plus directe et la moins coûteuse consiste à appliquer au véhicule en panne la solution technique - si elle existe - correspondant à un incident comparable déjà résolu sur un même modèle. En effet, cette approche évite de recourir à des modules de diagnostic plus complexes et plus longs. Elle est d'autant plus pertinente que la plupart des acteurs du monde automobile a déjà compris l'intérêt d'utiliser des bases d'incidents connus. Ces rapports de panne suivent généralement une structure bien définie et sont classés par modèle de véhicule. Chaque champ correspond à différentes étapes dans le processus de diagnostic : description de la panne, cause du problème, procédure de réparation. Par conséquent, la prise en

---

\*. Cette réflexion a été conduite dans le cadre du projet DYNAMO (ANR-07-TLOG-004 01), accessible à l'adresse <http://www.irit.fr/dynamo/>. A cet égard, les auteurs souhaitent remercier l'ANR ainsi que tous leurs partenaires scientifiques au sein du projet pour leur contribution.

compte des différents niveaux de structuration des documents peut aider à localiser de façon plus précise divers types d'information dans chaque texte.

La plupart du temps, le contenu de ces documents ne se conforme à aucune nomenclature existante. L'information est rédigée librement en langue naturelle, ce qui peut la rendre difficile à repérer selon sa formulation. Les moteurs de recherche actuels sont fondés sur des techniques statistiques appliquées au contenu lexical des textes, ainsi que, dans le cas de documents en ligne, sur le nombre et la popularité des sites y faisant référence (Baeza-Yates & Ribeiro-Neto (1999)). Ces moteurs n'exploitent donc pas directement le contenu sémantique des documents. Comme l'affirme Berners-Lee (1999), cette lacune entraîne un certain manque de précision dans les résultats de recherche, du fait de l'absence de prise en compte des relations de synonymies et de l'ambiguïté potentielle des requêtes.

L'indexation sémantique résout ces difficultés en établissant un lien entre les chaînes de caractères pondérées classiques et les idées qu'elles dénotent. De façon à pouvoir travailler à un niveau conceptuel, il est nécessaire de disposer d'un modèle préalable des connaissances mentionnées dans les documents. Pour une tâche d'indexation sémantique, il faut aussi pouvoir localiser leurs manifestations linguistiques, ce qui exige de manipuler à la fois des notions d'ontologie (Gruber (1995) : "*formalisation explicite d'une conceptualisation partagée*") et de terminologie. Pour cela, nous ferons appel à la structure de ressource termino-ontologique (RTO), définie dans Bourigault *et al.* (2004).

Par son incomplétude inhérente et l'apparition régulière de nouveaux modèles de véhicules, une base d'expérience est forcément amenée à évoluer pour s'adapter aux changements du domaine automobile. Au cours de nos travaux, nous nous sommes donc appliqués à trouver une manière de maintenir la cohérence entre la base d'incidents et le modèle de connaissances sur lequel s'appuient les processus d'indexation et de recherche sémantiques. Succinctement, nous avons d'abord mis en place un processus simultané et cyclique d'indexation sémantique et d'enrichissement de RTO. Une fois ce problème résolu, nous avons élaboré une mesure capable de comparer numériquement les annotations sémantiques d'une requête et celles d'un document. Les sections suivantes s'intéressent au lien entre ontologie et terminologie, au processus de construction d'une RTO de domaine, ainsi qu'au processus d'enrichissement de RTO par l'indexation. Nous détaillons ensuite le fonctionnement de l'outil de recherche, permettant aux garagistes de formuler librement une description des symptômes. Nous finirons en évoquant quelques perspectives à nos travaux.

## 2 Les ontologies pour la recherche d'information

Depuis environ 5 ans, la modélisation de connaissances à partir de textes est apparue comme un moyen prometteur pour réduire les coûts de construction et de maintenance des ontologies de domaine (Buitelaar *et al.* (2005)). Ce processus d'acquisition des connaissances permet de garder des traces des choix de modélisation et d'associer aux concepts de nombreuses entités lexicales (ou termes). Dans le cas de notre étude, cette caractéristique est d'autant plus intéressante que l'ontologie est utilisée comme langage formel permettant de représenter l'interprétation (d'une partie) du contenu textuel de chaque document. Comme la représentation des termes au travers d'un langage d'onto-

logie est un thème relativement peu étudié, nous nous sommes penchés sur la question de comment représenter à la fois des concepts et leur(s) manifestation(s) linguistique(s) dans un modèle standard unique.

## 2.1 Un nouveau méta-modèle en OWL

Après une analyse approfondie menée dans Reymonet *et al.* (2007a), nous avons conclu à l'insuffisance des modèles courants pour représenter conjointement une ontologie et sa terminologie associée : tout d'abord, la notion de terme n'est jamais réifiée dans la littérature, alors qu'il serait utile de pouvoir modéliser toutes les informations souhaitées sur les manifestations linguistiques d'un terme ; ensuite, la plupart des modèles existants (e.g. ceux utilisant GATE, décrit dans Bontcheva *et al.* (2004)) associe un concept et un terme à l'aide d'une relation d'instanciation, ce qui ne permet pas de reproduire fidèlement certains phénomènes linguistiques comme l'anaphore ou la polysémie.

Avec la volonté de respecter le standard ontologique le plus populaire, à savoir OWL<sup>1</sup>, nous avons proposé dans Reymonet *et al.* (2007a) un premier méta-modèle en OWL-DL, capable de représenter des termes sans altérer la syntaxe du sous-langage. Cependant, certains choix opérationnels avaient rendu notre modèle théoriquement incorrect (Reymonet *et al.* (2007b)). C'est pourquoi nous y avons apporté un certain nombre de modifications en nous fondant sur les primitives de méta-modélisation disponibles en OWL-Full. Nous sommes alors parvenus à la proposition schématisée sur la figure 1.

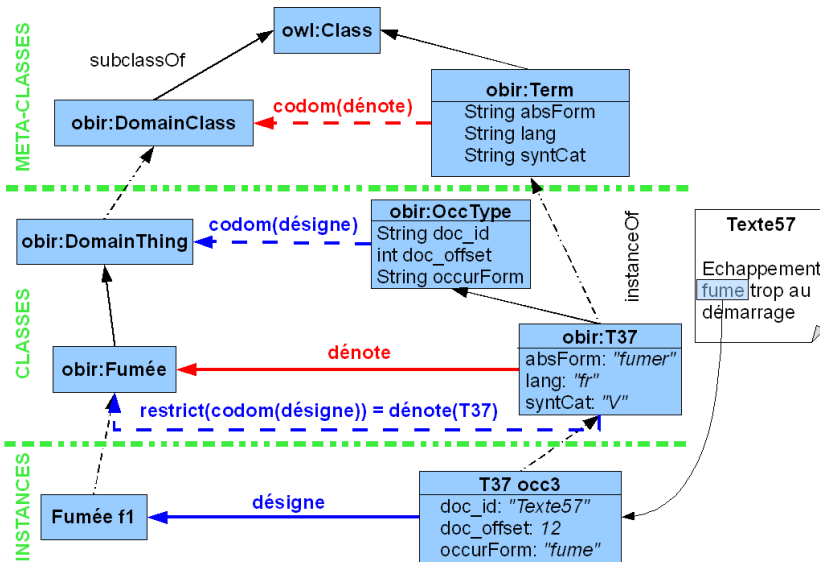


FIGURE 1 – Proposition de méta-modèle de RTO en OWL-Full

1. <http://www.w3.org/TR/owl-features/>

## 2.2 Construction de la RTO de domaine

En appliquant la méthode Terminae (détaillée dans Aussenac-Gilles *et al.* (2008)), il est possible d’obtenir des informations d’ordre terminologique et ontologique à partir d’une analyse syntaxique et distributionnelle du corpus de documents. Une seconde phase d’analyse recourant à l’aide d’un expert permet à l’ingénieur de la connaissance de construire une taxonomie du domaine, associée à une terminologie. Les étapes suivantes consistent à ajouter les relations dites ”transverses” entre concepts, à raffiner le modèle grâce aux critères de différenciation et à formaliser le modèle dans un certain langage, de façon à obtenir une première version de la RTO.

Après avoir suivi les premières étapes de construction, nous avons obtenu une ontologie dont nous représentons la partie haute dans la figure 2. Nous définissons un symptôme comme un concept défini à partir de quatres types primitifs :

- les problèmes, qui se décomposent en fautes génériques ou spécifiques, et qui peuvent apparaître sur une (ou plusieurs) prestation(s) donnée(s),
- les prestations, qui correspondent aux différents services rendus à l’utilisateur par le véhicule,
- les contextes, qui décrivent des états dans lesquels peut se trouver une prestation,
- les composants qui contrôlent le bon fonctionnement d’une prestation (tout composant inutile pour la définition d’un symptôme a été délibérément écarté)

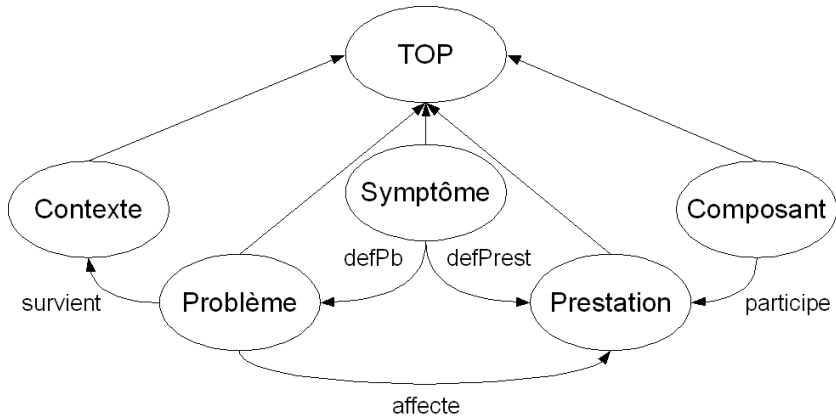


FIGURE 2 – Architecture globale de la RTO modélisée

Dans les sections suivantes, nous décrivons le logiciel TextViz, composé de deux outils ayant recours à une RTO pour effectuer une tâche de RI sémantique. Leur utilisation combinée est synthétisée sur la figure 3.

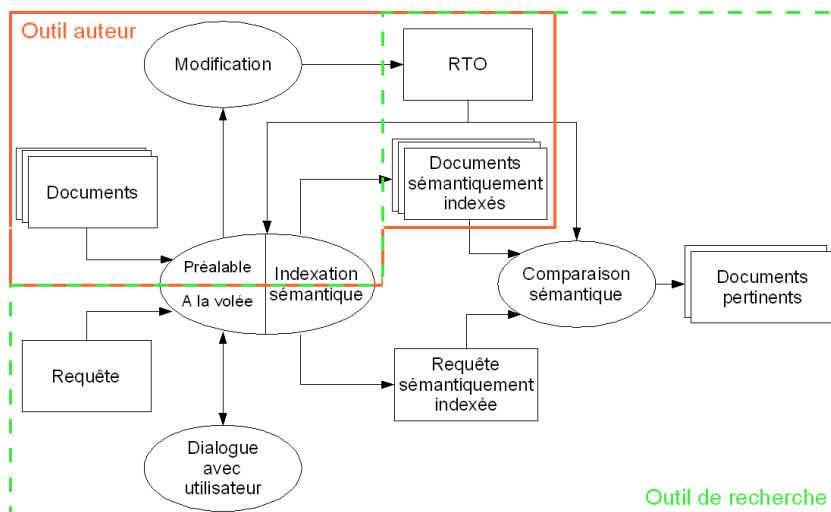


FIGURE 3 – Schéma de fonctionnement de TextViz

### 3 L’outil auteur, pour indexer le corpus et enrichir la RTO

Au cours de la construction d’une RTO, une étape importante consiste à vérifier qu’aucun concept ou aucune relation sémantique n’a été oubliée, et que la ressource ne contient pas d’assertion contradictoire. Comme le montre la figure 4, nous avons décidé de mener en parallèle les phases de complétion de la RTO et d’indexation sémantique du corpus. Ceci nous permet d’une part d’espérer un gain de temps (en utilisant l’ontologie avant même que celle-ci n’acquière un statut finalisé) ; d’autre part, en examinant les résultats d’indexation, l’ontologie sera en quelque sorte guidé par l’application et pourra faire évoluer le modèle de telle sorte qu’il atteigne un état satisfaisant pour annoter le corpus.

Les documents commencent donc par être automatiquement indexés avec le lexique déjà disponible dans la RTO. L’ontologie analyse ensuite les résultats pour déterminer quels termes et/ou concepts supplémentaires permettraient une meilleure interprétation des textes. Entre autres, il peut également décider de modifier la dénotation d’un terme, ou de réorganiser la taxonomie. Dans ce cas, certaines annotations sémantiques peuvent être remises en question si elles contiennent un des éléments remaniés.

La taille des corpus à indexer est généralement un facteur limitant et ne permet pas de présenter à l’utilisateur les annotations pour chaque texte. Nous avons donc défini plusieurs critères qui permettent d’évaluer automatiquement le niveau de ”compréhension” de chaque document par le système. Si pour un texte donné, un de ces critères n’est pas satisfait, le résultat est alors stocké pour être consulté ultérieurement par l’utilisateur, qui pourra décider de la marche à suivre : corriger manuellement les annotations proposées, éditer directement le document (s’il en a l’autorisation), ou modifier la RTO.

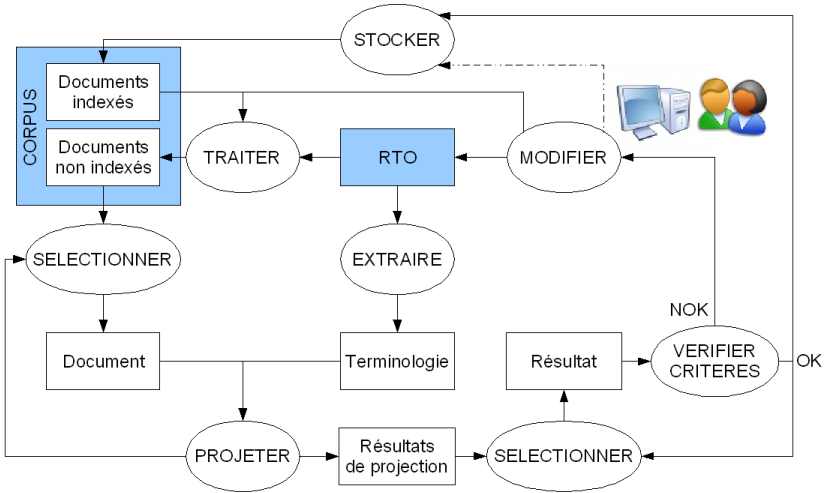


FIGURE 4 – Processus d’indexation sémantique

Dans le dernier cas, tout changement apporté à la ressource sera répercuté sur le processus d’indexation au cours de l’itération suivante, ce qui permettra au système de retrouver plus d’informations, ou de mieux interpréter les documents. Cette commutation cyclique entre étape automatique d’indexation et étape manuelle d’évaluation / modification prend fin lorsque tous les critères sont satisfaits sur la totalité du corpus (ou, à défaut, lorsque l’ontologie est globalement satisfait de la qualité de l’indexation).

Du fait de certaines propriétés spécifiques à notre base d’incidents (documents courts et concis), nous avons défini un premier critère de couverture minimale du champ contenant les symptômes. Nous cherchons ensuite à retrouver au moins un élément de type symptôme par document. Enfin, nous vérifions que les restrictions de cardinalité sont respectées, car elles peuvent être le signe d’une omission au cours de l’indexation. Prenons l’exemple de la relation sémantique *affecte* qui associe à un concept de type *problème* un concept de type *prestation* avec une cardinalité minimale de 1 (autrement dit, tout problème affecte nécessairement au moins une prestation). Si un terme dénotant un problème est repéré dans un document, le système s’attendra alors à y retrouver également au moins un terme dénotant une prestation compatible avec le problème.

Nous faisons la synthèse de notre proposition par l’exemple suivant : l’ontologie cherche à annoter la phrase *“Le moteur tire mal à l’accélération.”*. Supposons que la RTO contient les concepts de *motorisation* (prestation) et d’*accélération* (contexte), ainsi que leurs termes associés. La première tentative d’indexation échoue puisque le deuxième critère (présence d’un symptôme) n’est pas vérifié. La nécessité de repérer au moins un problème (troisième critère appliqué à la relation *defPb*, de cardinalité minimale 1) nous pousse à créer le concept *manque de puissance*, relié dans la RTO au nouveau terme *“tire mal”*. Le même raisonnement sur la relation

affecte nous incite à relier le nouveau problème à la prestation motorisation. Comme tous les critères sont vérifiés au bout de 2 itérations (la couverture est maximale et dépasse forcément le seuil défini au préalable), la phrase est considérée comme correctement interprétée par le système. Cet exemple met bien en avant la nécessité et les avantages de conduire en parallèle les étapes d'indexation sémantique et de complétion/évolution de la RTO.

En guise d'implémentation, nous avons développé l'outil TextViz comme extension au logiciel Protégé-OWL<sup>2</sup>. TextViz exploite les techniques de recherche disponibles dans Lucene<sup>3</sup> pour projeter sur le corpus la partie terminologique de la RTO en cours de construction. Il crée automatiquement les occurrences de termes et instances de concepts adéquates pour chaque document, puis vérifie les critères de bonne adéquation pour proposer à l'utilisateur une liste de documents potentiellement mal annotés. Pour chacun de ces textes, l'ontologue peut visualiser en contexte les éléments de la RTO retrouvés ainsi que l'interprétation du système (voir fig 5). Il peut alors effectuer un ensemble d'opérations visant à enrichir / corriger l'annotation sémantique et, par là même, la RTO.

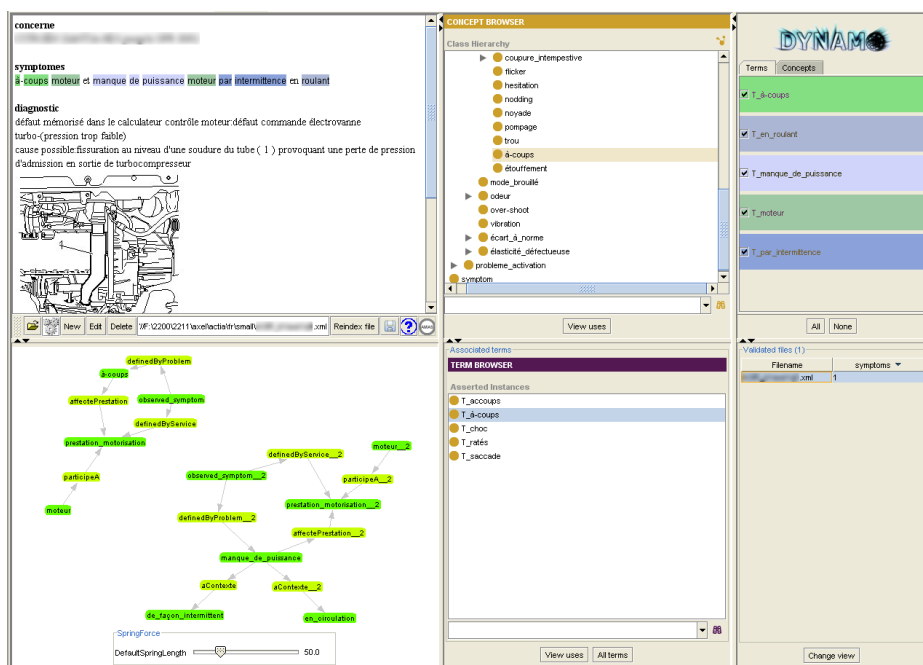


FIGURE 5 – Copie d'écran de l'outil auteur

2. <http://protege.stanford.edu/overview/protege-owl.html>  
 3. <http://lucene.apache.org/java/docs/>



## 4 L'outil de recherche, pour retrouver les documents sémantiquement proches d'une requête

Au cours de ce processus, nous supposons que le corpus sur lequel vont porter nos requêtes a été au préalable indexé sémantiquement. En guise de point de départ, l'outil permet à l'utilisateur de saisir une requête (dans notre cas, la description d'un ensemble de symptômes). Plutôt que d'obliger l'utilisateur à construire directement sa requête à partir de l'ontologie, l'IHM (Interface Homme-Machine) permet aussi d'entrer librement sa requête en s'affranchissant de l'ontologie. Dans ce cas, l'outil va d'abord automatiquement traduire la requête de l'utilisateur au travers de la RTO ; il va ensuite lui afficher sa proposition d'interprétation sous forme graphique, afin que l'utilisateur puisse la corriger (si nécessaire), avant de lancer le processus de recherche sur l'ensemble des documents (voir fig. 6).

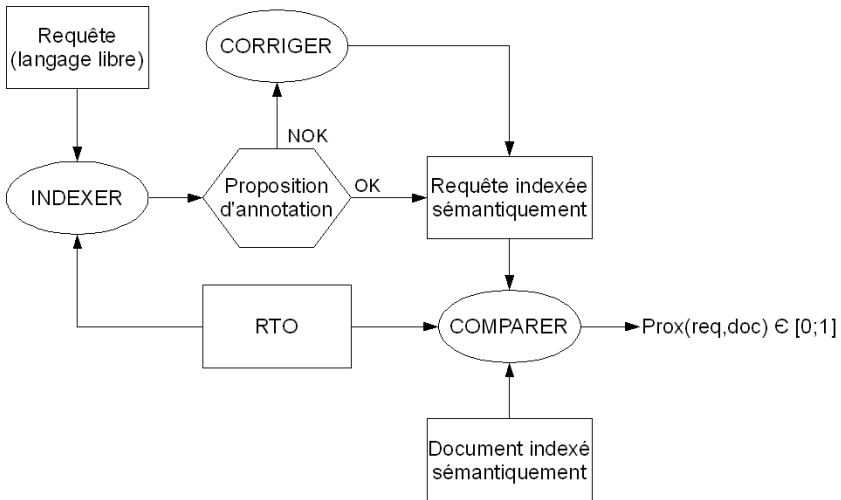


FIGURE 6 – Processus de recherche sémantique

L'étape de comparaison sémantique est sans nul doute la plus difficile. Dans notre cas d'étude, la recherche porte sur les concepts de type symptôme. Toutefois, ceux-ci ne sont définis qu'à travers les concepts primitifs auxquels ils sont reliés, ce qui empêche de calculer directement la similarité entre deux symptômes. Nous devons donc utiliser une moyenne de similarités entre les concepts primitifs adéquats (ie problème, prestation et contexte). La plupart des proximités de la littérature comme Mili *et al.* (1989); Wu & Palmer (1994); Resnik (1995) suit l'intuition selon laquelle deux concepts sont d'autant plus similaires que leur ancêtre commun le plus spécifique (en termes taxonomiques) est proche. Sur l'exemple de la figure 7, la mesure de Wu et Palmer permet de déduire de la structure taxonomique de l'ontologie qu'un problème de calage est plus proche d'un à-coups que d'un manque de puissance ( $d$  correspond à la profondeur du concept dans la taxonomie).

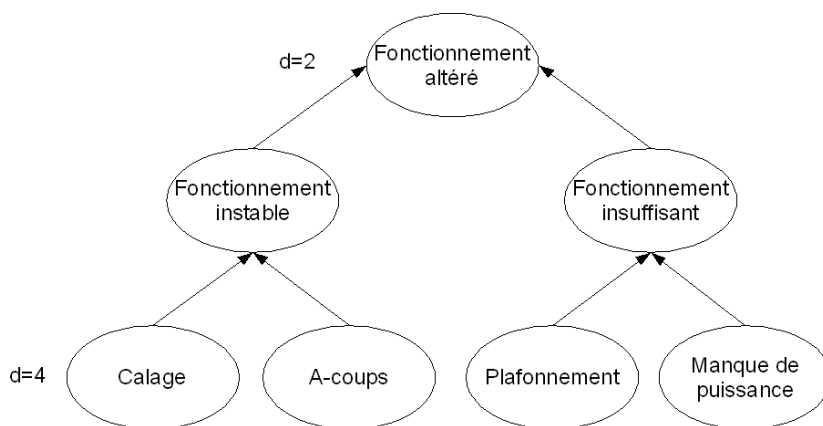


FIGURE 7 – Extrait d'ontologie

$$\begin{aligned}
 \text{sim}(\text{calage}, \text{a\_coups}) &= \frac{2 * d(\text{fct\_instable})}{d(\text{calage}) + d(\text{a\_coups})} \\
 &= \frac{2 * 3}{4 + 4} = 0.75
 \end{aligned}$$

$$\begin{aligned}
 \text{sim}(\text{calage}, \text{m\_de\_p}) &= \frac{2 * d(\text{fct\_altéré})}{d(\text{calage}) + d(\text{m\_de\_p})} \\
 &= \frac{2 * 2}{4 + 4} = 0.5
 \end{aligned}$$

De façon à prendre en compte les spécificités lexicales du corpus, nous avons choisi la mesure décrite dans Lin (1998) pour calculer la similarité entre deux concepts. En effet, en plus de suivre le même principe que Wu et Palmer, cette mesure pondère le résultat avec la proportion d'occurrences de termes qui correspondent aux deux concepts. Au final, seuls les documents les plus proches de la requête (ie dont la proximité est supérieure à un certain seuil) sont affichés dans l'IHM, par ordre décroissant de similarité. Fort logiquement, plus un incident comportera de symptômes en commun avec la requête, mieux il sera classé parmi les résultats.

## 5 Evaluation

Bien que les deux outils présentés doivent être utilisés conjointement pour réaliser la tâche macroscopique de recherche d'information sémantique, la mise en place d'un processus global et "absolu" d'évaluation (ie reconnu à la fois par les communautés d'IC et de RI) s'est avérée difficile. En effet, chaque logiciel vise une catégorie différente d'utilisateurs, avec des courbes d'apprentissage bien distinctes : le moteur de recherche a été conçu dans un souci d'ergonomie, pour être utilisé par un garagiste ; à l'inverse, l'outil auteur exige que l'utilisateur possède des compétences dans des domaines aussi différents que l'IC, l'indexation sémantique et le domaine modélisé. Même si ce constat nous a amené à évaluer les deux logiciels séparément, nous avons choisi des critères d'évaluation d'ordre pratique afin de juger de l'intérêt d'utiliser un processus de RI sémantique sur un "simple" moteur de recherche (i.e. non fondé sur une ontologie). Nous avons donc retenu des critères de temps (*quel est le coût additionnel d'un processus de RI sémantique ?*) et d'efficacité (*Les propositions d'annotation sont-elles correctes ? Les documents retournés sont-ils pertinents vis-à-vis de ma requête ?*).

Pour tester l'outil d'indexation, nous avons pris aléatoirement 50 documents et les avons fait annoter sémantiquement par le système. Nous avons mesuré le temps global passé sur cette tâche et avons vérifié que la RTO et les annotations sémantiques convergeaient vers un état stable. Au final, les documents ont été indexés en 80 minutes, occasionnant la création dans la RTO de 20 nouveaux concepts et 40 nouveaux termes. Si l'on compare ce temps avec ce dont est capable un robot d'indexation généraliste (qui traite automatiquement un très grand nombre de pages Internet selon une approche "sac de mots"), on pourrait trouver le processus proposé long et fatigant. Toutefois, il faut bien réaliser que notre outil n'est pas destiné à être utilisé sur une même échelle de taille : une base d'incidents moyenne comporte de l'ordre de quelques milliers de cas et n'augmente que d'une cinquantaine de nouvelles pannes tous les 3 mois. Par conséquent, il est difficile d'interpréter ces résultats sans regarder en parallèle l'efficacité "finale" du moteur de recherche.

Pour évaluer celui-ci au regard des performances des moteurs dits "classiques", nous avons utilisé 20 requêtes choisies au hasard parmi un ensemble de documents non encore indexés. Nous avons ensuite jugé de l'efficacité des deux types d'outils au travers de deux mesures largement acceptées au sein de la communauté de RI, à savoir la précision (proportion de documents pertinents parmi ceux retrouvés par le système) et le rappel (proportion de documents retrouvés parmi tous les documents pertinents). Les résultats, résumés en figure 8, montrent une amélioration significative des résultats en utilisant le moteur sémantique. Ceci peut s'expliquer à travers le fait que notre solution gère non seulement la synonymie (ce qui améliore le rappel), mais aussi interprète formellement tout ou partie des documents et des requêtes (ce qui améliore la précision en levant toute ambiguïté de sens).

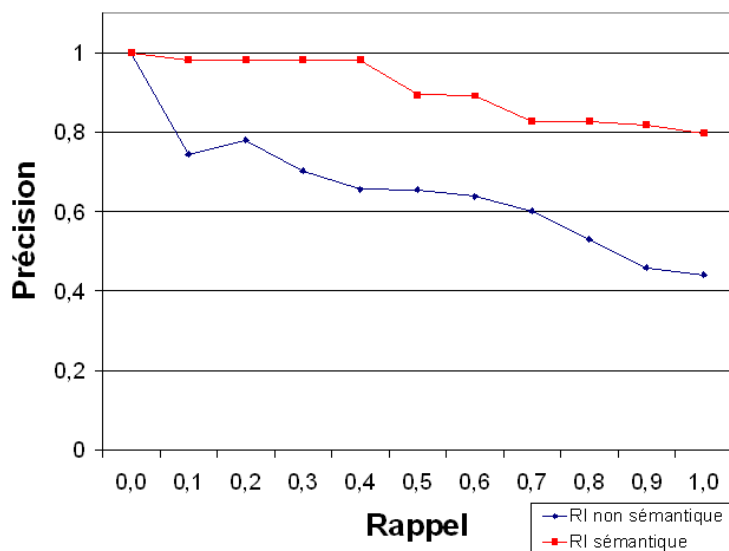


FIGURE 8 – Courbe de Rappel-Précision

## 6 Conclusion

Dans cet article, nous avons présenté un moteur de recherche sémantique capable de gérer au travers de deux outils séparés (mais complémentaires) les deux principales tâches d'un processus de RI : l'indexation et la comparaison. Bien que les idées et technologies utilisées puissent sembler peu originales, très peu de logiciels (voire aucun) permettent - à notre connaissance - de mener à bien un processus de recherche sémantique dans un contexte industriel. Actuellement en phase de validation, TextViz a été mis à disposition d'une équipe de téléassistance dans une filiale d'ACTIA, notamment chargée d'aider leurs clients garagistes en cas de diagnostic difficile : 2 techniciens ont été formés à l'utilisation de l'outil auteur, dont ils se servent pour enrichir progressivement une base d'incidents au fil des appels ; installé sur un parc de 6 ordinateurs, l'outil de recherche permet ensuite aux membres de l'équipe de consulter la base, située sur un serveur local. Les premiers retours sont très encourageants, seules quelques réserves (en cours de prise en compte) ont été émises sur l'ergonomie du logiciel.

D'un point de vue académique, ces travaux nous ont poussé à nous intéresser à plusieurs autres sujets de recherche : tout d'abord, avec un corpus de taille croissante, de nouvelles notions peuvent apparaître au cours du temps, ce qui rend indispensable une étude approfondie des conséquences croisées d'une évolution de la RTO et/ou des annotations sémantiques ; en outre, bien que notre moteur de recherche n'utilise pour l'instant que des connaissances explicitement mentionnées dans un document, il serait intéressant de pouvoir exploiter des liens de causalité existant dans la base de faits

de l'ontologie dans le but de rapprocher un document et une requête sur la base de connaissances dites "implicites" (ou de sens commun) ; enfin, si l'on considère la RTO comme un moyen de partager formellement une conceptualisation et un lexique d'un domaine, il est permis d'imaginer que dans un contexte international, elle pourrait facilement constituer un pivot entre un modèle de connaissances (indépendant de la langue) et plusieurs lexiques de langues différentes.

## Références

- AUSSENAC-GILLES N., SZULMAN S. & DESPRES S. (2008). *The TERMINAE Method and Platform for Ontology Engineering from Texts*, In *Bridging the Gap between Text and Knowledge : Selected Contributions to Ontology learning from Text*, p. 199–223. IOS Press.
- BAEZA-YATES R. A. & RIBEIRO-NETO B. (1999). *Modern Information Retrieval*. Boston, MA, USA : Addison-Wesley Longman Publishing Co., Inc.
- BERNERS-LEE T. (1999). *Weaving the Web : The Original Design and Ultimate Destiny of the World Wide Web by Its Inventor*. Harper San Francisco.
- BONTCHEVA K., TABLAN V., MAYNARD D. & CUNNINGHAM H. (2004). Evolving GATE to Meet New Challenges in Language Engineering. *Natural Language Engineering*, **10**(3/4), 349–373.
- BOURIGAUT D., AUSSENAC-GILLES N. & CHARLET J. (2004). Construction de ressources terminologiques ou ontologiques à partir de textes : un cadre unificateur pour trois études de cas. In J.-M. PIERREL & M. SLODZIAN, Eds., *Techniques informatiques et structuration de terminologies*, volume 18/1 of *Revue d'Intelligence Artificielle*. Hermes Sciences.
- P. BUITELAAR, P. CIMIANO & B. MAGNINI, Eds. (2005). *Ontology Learning from Text : Methods, Evaluation and Applications*. IOS Press.
- GRUBER T. R. (1995). Toward principles of the design of ontologies used for knowledge sharing. *International Journal of Human-Computer Studies*, **43**(5-6), 907–928.
- LIN D. (1998). An information-theoretic definition of similarity. In *ICML'98 : Proceedings of the Fifteenth International Conference on Machine Learning*, p. 296–304, San Francisco, CA, USA : Morgan Kaufmann Publishers Inc.
- MILI R. R. H., BICKNELL E. & BLETNER M. (1989). Development and application of a metric on semantic nets. *IEEE Transactions on Systems, Man and Cybernetics*, **19**(1), 17–30.
- RESNIK P. (1995). Using information content to evaluate semantic similarity in a taxonomy. In *IJCAI*, p. 448–453.
- REYMONET A., THOMAS J. & AUSSENAC-GILLES N. (2007a). Modélisation de Ressources Terminologiques en OWL. In *Actes des 18e journées francophones d'ingénierie des connaissances*.
- REYMONET A., THOMAS J. & AUSSENAC-GILLES N. (2007b). Modelling Ontological and Terminological Resources in OWL DL. In *Proceedings of ISWC'07 workshop "From Text to Knowledge : The Lexicon/Ontology Interface" (OntoLex'07)*.
- WU Z. & PALMER M. (1994). Verb semantics and lexical selection. In *32nd. Annual Meeting of the Association for Computational Linguistics*, p. 133–138, New Mexico State University, Las Cruces, New Mexico.