

Mysins : Make Your Semantic INformation System

Anthony Ventresque*, Thomas Cerqueus**, Louis-Alexandre Celton***, Gaëtan Hervouet***,
Damien Levin***, Philippe Lamarre*, Sylvie Cazalens**

*INRIA & LINA, Université de Nantes

prenom.nom@univ-nantes.fr,

**LINA, Université de Nantes

prenom.nom@univ-nantes.fr,

***Université de Nantes

prenom.nom@etu.univ-nantes.fr,

1 Introduction

La sémantique est de plus en plus utilisée dans différents domaines comme la recherche d'information (RI) et le Web sémantique. Dans le domaine de la RI, différents participants interviennent : des fournisseurs d'informations et des utilisateurs. Dans ce contexte, la mise en place d'un système distribué semble approprié car il permet d'éviter les problèmes liés à la centralisation de données (passage à l'échelle, confidentialité, etc.) et de garantir l'autonomie sémantique : chaque participant peut choisir son ontologie, la manière d'indexer ses données, etc. Par ailleurs, l'utilisation de la sémantique nécessite la mise en œuvre de mécanismes précis. En RI, il s'agit entre autre de l'utilisation d'ontologies, du calcul de similarité et de l'indexation. L'étude de chacun de ces axes nécessite un effort important de synthèse et d'intégration. Par exemple, lorsqu'une nouvelle mesure de similarité est proposée, il faut mettre en place les mécanismes d'accès à l'ontologie, pour pouvoir comparer ses résultats à ceux des autres mesures (Wu et Palmer (1994), Resnik (1995)). Lors de l'étude d'un processus complexe, comme l'indexation sémantique, cela devient particulièrement laborieux. Pour pallier le manque évident d'une architecture générique distribuée pour la conception de systèmes d'information sémantiques, nous proposons un framework : Mysins.

2 Le framework Mysins

Mysins définit un certain nombre de « briques sémantiques » (ie. d'unités logicielles) nécessaires à la conception d'un système de partage d'information : accès aux données, accès aux ontologies, calcul de similarité, requêtage, etc. Ces composants sont facilement intégrables et réutilisables au sein d'une application distribuée. L'aspect lié à la distribution nous semble primordial pour deux raisons principales. Premièrement, certaines tâches liées à la sémantique peuvent être contraignantes en ressources ou en compétences, il faut donc que chaque participant ait la possibilité de sous-traiter une tâche qu'il n'est pas capable de réaliser. Deuxièmement, cela nous permet d'aborder les problématiques liées à la personnalisation (chaque par-

ticipant du système peut proposer une nouvelle brique sémantique) et à l'autonomie (chaque participant choisit lui-même les services qu'il veut consommer/proposer). C'est pourquoi Mysins permet très facilement une implémentation sous forme de services.

3 Démonstration

Nous présentons les possibilités de Mysins avec deux rôles différents. D'un côté, un fournisseur d'information a la possibilité d'indexer un corpus de documents en choisissant les paramètres qui composent le processus d'indexation : choix du composant gérant l'accès à une ontologie, choix du composant chargé du calcul de similarité sémantique, etc. D'un autre côté, un utilisateur a la possibilité de formuler une requête et de demander au fournisseur d'information de lui envoyer les meilleurs documents correspondants à sa demande. Pour caractériser sémantiquement sa requête, l'utilisateur peut utiliser ses propres composants ou les services distants proposés par le fournisseur d'information.

Cette application intègre la méthode ExI²D proposée par Ventresque et al. (2008), dont l'objectif est d'améliorer l'interopérabilité sémantique entre l'initiateur de requête et le fournisseur d'information, par l'ajout de nouvelles briques d'analyse sémantiques des requêtes très adaptées à un cadre hétérogène sémantiquement. Nous mettons ainsi en valeur le fait que Mysins permet de construire ou de modifier un système d'information sémantique aisément.

4 Conclusion

Mysins définit une architecture générique distribuée simplifiant le développement de systèmes d'information sémantiques. En nous basant sur ce framework, nous proposons des composants de base essentiels à toute application utilisant la sémantique : accès aux ontologies, calcul de similarité, indexation des données. Nous proposons également des composants plus complexes utilisés par la méthode ExI²D (expansion et interprétation de requête, image de document, etc.).

Références

- Resnik, P. (1995). Using information content to evaluate semantic similarity in a taxonomy. In *In Proceedings of the 14th International Joint Conference on Artificial Intelligence*, pp. 448–453.
- Ventresque, A., S. Cazalens, P. Lamarre, et P. Valduriez (2008). Improving interoperability using query interpretation in semantic vector spaces. In *ESWC*.
- Wu, Z. et M. Palmer (1994). Verb semantics and lexical selection. In *32nd. Annual Meeting of the Association for Computational Linguistics*, pp. 133–138.