

# SyRQuS - Recherche par combinaison de graphes RDF.

Adrian Tanasescu\*

\*Université Lyon 1, Villeurbanne, F-69622, France, LIRIS CNRS UMR 5205 43,  
Bat. Nautibus, 43 Bld. du 11 Novembre 1918, 69622 Villeurbanne  
atanases@liris.cnrs.fr,  
<http://bat710.univ-lyon1.fr/atanases/>

**Résumé.** Nous nous intéressons à un mécanisme permettant la construction de réponses combinés à partir de plusieurs graphes RDF. Nous imposons, par souci de cohérence, que cette combinaison soit réalisée uniquement si les graphes RDF ne se contredisent pas. Pour déterminer la non-contradiction entre deux graphes RDF nous utilisons une mesure de similarité, calculée au moment de l'ajout de documents RDF dans la base de documents.

## 1 La plateforme SyRQuS

Même si cela fait plusieurs années que RDF est devenu un standard recommandé par W3C, le développement des langages de requête RDF a été plus long. Après l'apparition de RDF, des langages permettant d'accéder aux triplets RDF ont émergé, comme TRIPLE (Sintek et al., 2002) ou encore Squish (SquishQL, 2002). De ces premiers sont inspirés d'autres langages comme RQL, RDQL - langage d'origine de la plateforme Jena (Jen) - ou encore SeRQL langage de base de Sesame (Kampman et Broekstra). Tous ces efforts convergent aujourd'hui vers un langage SQL-like qui est en train de devenir la future recommandation W3C : SPARQL (Seaborne et Prud'hommeaux, 2006). Déjà en statut de recommandation candidate dans sa version d'avril 2006, ce langage est petit à petit adopté par les plateformes orientées vers le Web sémantique utilisant RDF.

Pour cette raison nous avons orienté notre effort vers le développement d'un outil permettant d'interpréter les requêtes formulées à l'aide de ce langage. SyRQuS (Syntetizing RDF Query System) a été développé dans un environnement PHP/MySQL afin d'assurer un déploiement indépendant par rapport aux systèmes d'exploitation. Il utilise le parseur ARC RDF/XML afin d'extraire les triplets des documents RDF et l'analyseur de requêtes SPARQL de RAP (RDF API for PHP).

Les fonctionnalités de la plateforme SyRQuS se décomposent en deux parties :

1. **Ajout de nouveaux documents RDF.** Cette fonctionnalité réalise, d'une part, l'insertion des documents RDF dans la base de données et, d'autre part, la mise à jour de la matrice de similarité pour chaque nouveau document RDF ajouté.
2. **Interrogation de la base de données.** Après la formulation d'une requête en SPARQL, le moteur de recherche suit les étapes suivantes :
  - (a) Décomposition de la requête et récupération des triplets de la clause WHERE ;

- (b) Pour chaque triplet de la requête on cherche les document RDF y répondant et on construit ainsi la matrice document-but ;
- (c) Tri de la matrice document-but par ordre croissant du nombre de but répondus par chaque document ;
- (d) Affichage des réponses complètes ne nécessitant pas de combinaison de graphes ;
- (e) Pour chaque document  $D_b$  qui fournit une réponse partielle et dans la limite du nombre de réponses attendus :
  - Construction de l'ensemble  $E$  de documents qui ne contredisent pas  $D_b$  et qui répondent aux sous-requêtes auxquelles  $D_b$  ne répond pas ;
  - Ajout des triplets réponses appartenant aux documents dans  $E$  aux triplets réponses de  $D_b$ , sous la condition d'équivalence des sujets des premiers à au moins une ressource des derniers.
- (f) Présentation du résultat dans un format HTML et un fichier XML

## 2 Conclusion et perspectives

SyRQuS est une application destinée à la recherche d'informations dans les documents RDF. Son implémentation permet d'obtenir des réponses plus complètes par la combinaison de plusieurs graphes RDF fournissant des réponses partielles. Développée dans un souci de compatibilité avec le langage de requête SPARQL, elle permet le traitement des requêtes de type SELECT, ce type de requête étant le seul à pouvoir profiter de notre algorithme de recherche par combinaison de graphes.

Actuellement, nous travaillons sur l'intégration des ontologies référencées dans les documents RDF, ce qui nous permettra de procéder à des approximations de réponses au moment de l'interrogation de la base de documents.

## Références

- HP Labs Semantic Web Research. Jena - A Semantic Web Framework for Java. <http://jena.sourceforge.net/>.
- Kampman, A. et J. Broekstra. Sesame - Open Source RDF Database. <http://www.openrdf.org/>.
- Seaborne, A. et E. Prud'hommeaux (2006). SPARQL query language for RDF. W3C candidate recommendation, W3C. <http://www.w3.org/TR/2006/WD-rdf-sparql-query-20061004/>.
- Sintek, M., S. Decker, et A. Harth (2002). TRIPLE. <http://triple.semanticweb.org/>.
- SquishQL (2002). Squish Query Language. <http://swordfish.rdfweb.org/rdfquery/ql>.

## Summary

Nowadays, RDF is a W3C recommended standard for describing knowledge about resources over the Web. In this context, we explore the RDF query answering field and take a closer look at the possibility of combining several RDF graphs while searching for query answers. We propose a prototype tool that generates virtual complete answers by combining partially answering RDF graphs. The generation of such answers is performed under a non-contradiction condition calculated through a similarity measure between the RDF graphs to be combined.