

Etude comparative des langages de requêtes sémantiques pour l'extraction des liens complexes dans une base de connaissances

Thabet Slimani*, Boutheina Ben Yaghlane**

*LARODEC, ISG de Tunis BP 41 - Rue de la liberté- Bardo 2000, Tunisia
thabet.slimani@issatm.rnu.tn

**LARODEC, IHEC Carthage, Carthage Présidence 2016, Tunisia
boutheina.yaghlane@ihcc.rnu.tn

Dans le cadre de la recherche de l'information sémantique, les techniques existantes d'extraction et de recherche sémantique se concentrent principalement sur l'identification des entités, des relations directes ou indirectes. Une association sémantique se rapporte à une relation (ou un lien) indirecte qui relie deux entités contenues dans une base de connaissances (Aleman-Meza et al., 2003). L'extraction des liens complexes (chemins) à partir d'une base de connaissances est une méthodologie de recherche sémantique qui a été traitée par divers langages de requêtes des métadonnées.

Néanmoins, les langages d'interrogation RDF existants n'offrent pas de mécanisme adéquat pour formuler des requêtes destinées à l'extraction des relations complexes ou chemins dans un graphe RDF. Afin de pallier à cette défaillance, nous avons proposé un nouveau langage (PmSPARQL) (Slimani et al., 2008) permettant d'étendre le langage d'interrogation actuel de W3C (SPARQL). Les principaux objectifs de PmSPARQL est d'inclure des possibilités d'extraction d'associations sémantiques et de pouvoir les évaluer.

Dans cet article, nous comparons notre langage PmSPARQL avec d'autres langages d'interrogation. Cette comparaison prend en compte plusieurs propriétés de graphes évoquées dans le travail de (Angles et al., 2004) et sept langages d'interrogation. La possibilité des requêtes appliquées sur des graphes flexibles pour des bases de données RDF est envisageable par certains langages tels que RxPath, Versa, PPARQL, SPARQLer et SPARQ2L. Néanmoins, les langages de requêtes qui ont des buts plus similaires à PmSPARQL et qui sont seulement destinés à l'extraction des chemins (associations sémantiques) se limitent à SPARQLer et SPARQ2L.

D'une manière similaire à SPARQLer, l'expression des requêtes pour l'extraction des chemins indirects est actuellement possible avec PmSPARQL, mais cette caractéristique n'est pas permise dans le langage SPARQ2L. En plus, l'expression des requêtes d'extraction des chemins qui incluent des contraintes d'inclusion (motifs impliquant des entités et des propriétés) n'est pas permise par SPARQLer qui présente seulement des contraintes sur les propriétés, mais qui est offerte par PmSPARQL et SPARQ2L. Les résultats de notre étude comparative sont présentés dans le tableau récapitulatif suivant (Tableau 1) :

Propriétés	G	G+	Gram	RXPath	SPARQ2L	SPARQLer	PmSPARQL
Noeuds adjacents	±	*	*	×	×	×	*
Arcs adjacents	±	*	*	×	×	×	*
Degré du noeud	×	*	×	×	×	×	*
Chemin	*	*	*	±	*	*	*
Distance entre 2 noeuds	×	*	×	×	*	×	*
Diamètre	×	*	×	×	*	×	×
Chemin avec entité unique	?	?	?	?	×	*	*
Chemin indirect	×	×	×	×	×	*	±
Degré de chemin	×	×	×	×	×	×	*
Chemin rangé	×	×	×	×	×	×	*
Chemin de longueur fixée	*	*	*	±	*	*	*

TAB. 1 – Capacité de certains langages de requêtes sémantiques vis à vis du support pour certaines propriétés (* support complet, ± support partiel, × aucun support, et ? indique la non disponibilité de l'information).

Références

- Aleman-Meza, B., C. Halaschek-Wiener, B. Arpinar, et A. Sheth (2003). Context-aware semantic association ranking. In *First International Workshop on Semantic Web and Databases.*, Berlin, Germany, pp. 33–50.
- Angles, R., C. Gutierrez, et J. Hayes (2004). Rdf query languages need support for graph properties (technical report tr/dcc-2004-3). Technical report, Department of Computer Science, University of Chile.
- Slimani, T., B. Ben Yaghlane, et K. Mellouli (2008). Pmsparql : Extended sparql for multi-paradigm path extraction. *International Journal of Computer, Information, and Systems Science, and Engineering.* 2 (3), pp. 179–190.

Summary

In the context of Semantic Association extraction, we define path extraction queries as paths in RDF graph between two entities which represents a kind of associations between them. According to the needs of path extraction queries, such request can extract paths which connect a source entity with target entity. Several paradigms of path extractions can be formulated. Nevertheless, the current RDF query languages have the lack to formulate requests intended for the extraction of semantic associations. In this paper, we propose a comparative study of our PmSPARQL query language with some other query languages which have similar goals.