

Technologies du Web Sémantique pour l'exploitation de données lexicales en réseau (*Lexical Linked Data*)

David Rouquet

LIG-GETALP

david.rouquet@imag.fr

RÉSUMÉ

Nous présentons des technologies du Web Sémantique utiles pour la gestion, le développement et l'exploitation de données lexicales en réseau.

ABSTRACT

Semantic Web technologies for Lexical Linked Data management

We present Semantic Web technologies for *Lexical Linked Data* management.

MOTS CLÉS : LEXICAL LINKED DATA, LEXIQUE MULTILINGUE, PIVOT, AXIES, SPARQL, SPIN.

KEYWORDS : LEXICAL LINKED DATA, MULTILINGUAL LEXICON, PIVOT, AXIES, SPARQL, SPIN.

1 Introduction

Les ressources lexicales multilingues sous forme de données en réseau (*Linked Data*) reçoivent un intérêt croissant en TALN (Chiarcos et al. 2012). Nous imaginons les ressources lexicales en réseau (*Lexical Linked Data*, *LLD*) comme un nuage de ressources interopérables améliorant la couverture des ressources isolées.

Les LLD offrent des avantages théoriques mais leur utilisation opérationnelle dans des applications de TALN n'est pas triviale. Parmi les avantages, on retient en particulier : l'interopérabilité syntaxique garantie par le standard RDF, l'interopérabilité conceptuelle que l'on peut atteindre à l'aide de schémas partagés (SKOS, Lemon, etc.) ou d'alignements entre ces schémas et enfin la possibilité d'interroger simultanément les dernières versions de ressources distribuées (sorte de *rolling release* pour les ressources).

Nous démontrons des solutions concrètes pour l'exploitation de LLD. Les technologies utilisées sont dérivées de SPARQL¹ et supportées par une API ouverte. La démonstration est réalisée avec l'environnement de développement propriétaire *TopBraid Composer*.

2 Problèmes traités dans la démonstration

Le premier problème pour exploiter de façon unifiée des données distribuées est leur référencement. L'enjeu est non seulement d'inventorier les ressources utiles mais également de décrire leurs schémas internes (*microstructures*). De plus les données en réseau sont liées selon un modèle pair à pair, comme illustré par la figure 1. Aussi, les « chemins » possibles pour résoudre une requête de traduction entre deux langues ne sont pas connus *a priori* ce qui augmente la complexité de la requête.

¹SPIN (*SPARQL Inferencing Notation*) rules, SPIN map et SPARQL motion.

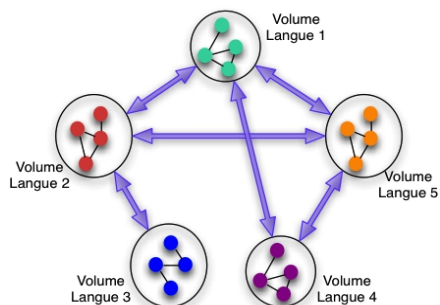


FIGURE 1 – Données en réseau "pair à pair"

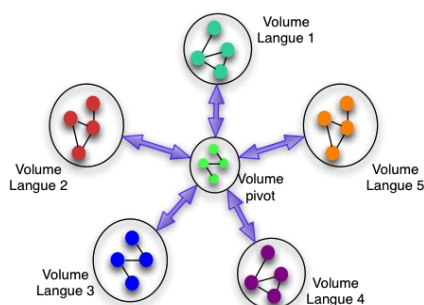


FIGURE 2 – Données en réseau avec pivot

Une solution pour résoudre à la fois le problème du référencement des LLD et leur utilisation efficace comme dictionnaire multilingue est de les organiser dans une architecture à pivot, comme illustré par la figure 2. Nous utilisons un pivot composé d'axies comme proposé dans la norme ISO *Lexical Markup Framework* (Francopoulo et al. 2006).

Afin d'intégrer une ressource dans l'architecture à pivot, nous commençons par aligner sa microstructure avec un schéma de référence. Nous proposons dans la démonstration un schéma simple pour des requêtes de traduction entre sens de mots mais ce schéma peut être modifié ou étendu pour des applications spécifiques. L'alignement entre la micro-structure d'une ressource et le schéma de référence joue le rôle de métadonnées qui décrivent les informations disponibles dans la ressource et les chemins pour y accéder. Nos alignements, supportés par la technologie SPINmap, peuvent être créés dans un outil graphique et « exécutés » pour passer effectivement du schéma source au schéma de référence. Ensuite, un ensemble de règles SPIN permettent de construire automatiquement les axies sous forme de nœuds RDF anonymes. Une axie représente un lien n-aire qui existe entre les entrées de différentes ressources dans le graphe multilingue. Les règles peuvent être exécutées pour initialiser la structure à pivot ou la mettre à jour avec de nouvelles ressources (ou de nouvelles versions des ressources).

Ainsi, nous obtenons un ensemble de LLD accessibles de façon unifiée via les alignements avec le modèle de référence et indexées par le volume d'axies. L'architecture à pivot permet une résolution optimisée des requêtes de traduction. Notre prototype inclut divers services pour l'import, l'export, la consultation et l'amélioration incrémentale des données à partir de la structure des ressources. Ces services peuvent être combinés dans un outil graphique et déployés sous forme de services Web à l'aide de la technologie SPARQL motion.

Références

CHIARCOS, C., NORDOFF, S. et HELLMANN, S. (2012). *Linked Data in Linguistics*. Springer, ISBN 978-3-642-28249-2.

FRANCOPOULO, G., NURIA B., MONTE G., CALZOLARI, N., MONACHINI, M., PET, M., et SORIA, C. (2006). « Lexical Markup Framework (LMF) for NLP multilingual resources ». In *Proc. Workshop on Multilingual Language Resources and Interoperability*, 1–8. MLRI '06. Stroudsburg, PA, USA.