

TP L017 : Indexation sémantique

Objectifs

Découvrir les notions d'ontologie et de base de connaissance RDF.

Maîtriser un langage de requêtes sur une base de connaissance RDF.

Utiliser un langage de recherche plein texte pour exploiter une base de connaissance.

Documents (accessibles sur le site lo17)

| | |
|-----------------------------|---|
| cours/lo17-owl-2013.pdf : | support de cours |
| documents/lo17.n3 : | ontologie lo17 au format N3 |
| documents/lo17.owl : | ontologie lo17 au format OWL |
| documents/kblo17-utf8.owl : | base de connaissance documentaire lo17 au format OWL |
| documents/kblo17-utf8.n3 : | base de connaissance documentaire lo17 au format N3 |
| src/main/tplo17main.java : | quelques méthodes java utiles pour le tp ; à compléter durant le tp |

Librairies java nécessaires pour le tp. Elles sont aussi téléchargeables à partir du site JENA :

<http://jena.apache.org/>

On utilise la version 2.7.4

jena-arq-2.9.4.jar ; jena-core-2.7.4.jar ; jena-iri-0.9.4.jar ;

jena-larq-1.0.0-incubating.jar; lucene-core-3.3.0.jar ;

log4j-1.2.16.jar ; slf4j-api-1.6.4.jar ; slf4j-log4j12-1.6.4.jar; xercesImpl-2.10.0.jar; xml-apis-1.4.01.jar

Ces librairies se trouvent dans le répertoire lib.

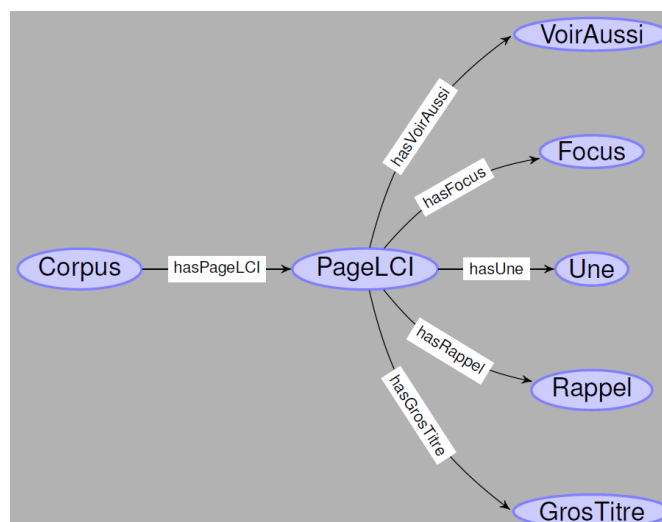
Aide

Cours LO17 : <cours/lo17-owl-2013.pdf>

Jena Apache : <http://jena.apache.org/>

Documentation : <http://jena.apache.org/documentation/>

Langage SPARQL : <http://www.w3.org/TR/rdf-sparql-query/>



Question 1

1. Le transparent 51 du cours présente l'ontologie des éléments indexés. A partir de l'écriture N3 de l'ontologie (fichier lo17.n3), indiquer comment retrouver les concepts, les relations et les attributs.
2. Représenter les 7 concepts (voir figure précédente) et indiquer à côté le noms des attributs qu'ils possèdent.

Question 2

Créer un répertoire servant de workspace pour Eclipse et importer le projet java (ex lo17). Il possède un sous-répertoire documents contenant l'ontologie et la base de connaissance et un sous-répertoire lib contenant les librairies. Le fichier source main/tplo17main.java est compléter.

Compléter la méthode q2() qui implémente le parsing de l'ontologie (fichier lo17.owl) pour exécuter une requête SPARQL SELECT à partir d'un fichier ou d'une chaîne (utiliser readFileQuery ou readStringQuery selon le cas). Ecrire quatre requêtes pour retrouver :

- a) Les concepts de l'ontologie
- b) Les relations de l'ontologie
- c) Les attributs de l'ontologie
- d) Les concepts domaines à la fois de « hasTitreArticle » et de « hasAuteur »

Question 3

1. Créer une méthode implémentant le parsing de la base de connaissance `kblo17-utf8.owl`. Ecrire une requête SPARQL pour rechercher les auteurs des unes.
2. Rechercher les thèmes des gros titres (thèmes distincts, triés par ordre alphabétique).
3. Rechercher les résumés qui contiennent le mot « Canada ».
4. Rechercher les résumés contenant une date comprise entre à 2003 et 2009 (inclus).

Question 4

1. Créer un index LARQ des chaînes de la base de connaissance à l'aide de la méthode `getWholeStringIndex`. En utilisant cet index, rechercher les résumés dont les auteurs sont A. GUILLET et F. LEFEBVRE.
2. Rechercher les résumés qui parlent de problèmes aériens :
 - a. Déterminer les termes utilisés dans les résumés à ce propos.
 - b. Construire plusieurs requêtes permettant d'affiner la recherche.
 - c. Comment éliminer les résumés qui parlent de vol d'oiseau ?
 - d. Vérifier l'effet du ranking sur les résultats.
 - e. Comparer avec la même recherche mais effectuée sur le titre.
3. Rechercher les résumés qui parlent de tempêtes ayant un rapport avec la Floride mais pas Cuba. Indiquer leur auteurs éventuellement. Analyser la localisation de la tempête par rapport à la Floride.