

┌ Projet – Interrogation de graphes RDF(S) ┐

Resource Description Framework (RDF) est un modèle de graphe destiné à décrire de façon formelle et structurée les ressources Web. Développé par le W3C, RDF est la base du *Web sémantique*¹. RDF Schema (RDFS) fournit des éléments pour la définition du vocabulaire (on parle aussi de taxonomie voir d'ontologie) destiné à structurer des ressources RDF.

En d'autres termes, RDFS permet de modéliser les connaissances structurelles d'un domaine donné, tandis que RDF permet de modéliser les connaissances factuelles (en adéquation avec la structure) sur ce même domaine. Par analogie avec la modélisation objet, RDFS permet de décrire les classes et leurs relations, RDF permet de décrire les instances (on parle aussi d'assertions) de ces classes et relations.

Un document structuré en RDF est un ensemble de *triplets*, qui constitue un graphe RDF. Un triplet RDF est une association {sujet, prédicat, objet} :

- le *sujet* représente la ressource à décrire ;
- le *prédicat* représente une propriété applicable à cette ressource (on parle aussi de relation) ;
- l'*objet* représente une donnée ou une autre ressource : c'est la valeur de la propriété.

Notons que RDFS est défini en RDF, ainsi tout document RDFS est un document RDF. Notons aussi, que la plupart des documents RDF/RDFS actuellement sont codés en XML.

Le but de ce projet est, à partir d'un document structuré en RDF(S), d'une part de définir une représentation machine performante d'un tel graphe, d'autre part d'être capable d'interroger les connaissances contenues dans ce graphe (*graph morphism*). Les requêtes s'effectueront sur les connaissances factuelles et devront tenir compte des connaissances structurelles associées. Le résultat d'une requête pourra à son tour être interrogé, de sorte qu'il sera possible d'enchaîner un ensemble de requêtes.

Le langage de programmation utilisé est libre. L'élaboration d'un environnement d'interrogation graphique et/ou "à la SQL" serait un plus.

Afin de tester votre code sur des informations réelles, la sous-base "Films" de données DBpedia² sera utilisée.

Les descriptions faites dans d'autres catégories pourront, si le projet avance bien, être utilisées :

- Music album : Led Zeppelin III, Like a Virgin, Thriller, *etc.*
- Computer Game : Tetris, World of Warcraft, *etc.*

1. Le Web sémantique désigne un ensemble de technologies visant à rendre le contenu des ressources du World Wide Web accessible et utilisable par les programmes et agents logiciels.

2. wiki.dbpedia.org/Datasets, contient actuellement environ 274 millions de triplets RDF.