E-Applications

Compte Rendu TP1, Première Partie, Traduction d’un texte en RDF

# Lucas LAFON & Théo CHAMBON

## 2014/2015



Une première étape afin de traduire ce texte au format RDF a été d’extraire les principales informations données dans ce texte sous formes de triples RPO, comme par exemple les informations sur les différentes personnes (Francis, Stanley Kubrick…) comme leurs dates de naissances, leurs métiers, et les différents films auxquels ils ont participés.

Une première modélisation du texte nous a permis d’obtenir les triplets suivants :

S P O

Realisateur est une Personne

Producteur est une Personne

Scénariste est une Personne

Acteur est une Personne

Compositeur est une Personne

Francis est réalisateur

Francis est Scénariste

Francis est producteur

Marlon Brando est Acteur

Francis est né le 7 avril 1939

Francis est né a Detroit

Detroit est situé dans le Michigan

Francis a réalisé Le Parrain

Francis a réalisé Apocalypse Now

Francis a réalisé Dracula

Apocalypse Now est un film

Apocalypse Now a pour thème La Guerre Du Vietnam

Marlon Brando est acteur de Apocalypse Now

Full Metal Jacket a pour thème La Guerre Du Vietnam

Vivian Kubric est une compositrice

Vivian Kubrick a pour père Stanley Kubrick

Vivian Kubrick a composé la musique de Full Metal Jacket

Une fois toutes les informations extraites sous forme de triplets RDF, nous avons spécifié et structuré nos ressources RDF à l’aide de RDFS.

Ainsi, nous avons pu déterminer que Personne et Films serait des classes (de type rdfs:class). De plus, les classes Acteur, Réalisateur, Compositeur, Scénariste et Producteur sont donc maintenant des sous classes de Personnes (rdfs:subClassOf).

Il était également important d’identifier quelles parties serait utilisée comme ressources, car importantes, et posséderaient donc une URI.

Une fois cette structure de base obtenue (les classes et leurs sous-classes), nous avons donc pu mettre en place les différentes propriétés (le thème d’un film, la date de naissance d’une personne par exemple).

Ensuite, nous avons implémenté les relations entre les différentes ressources (par exemple, Francis à réalisé le film Apocalypse Now).

Avec tout cela, nous avons donc réalisé un schéma, représentant la structure de notre RDF, et les instances présentes dans le texte.

Nous avons ensuite définis plusieurs namespace, permettant de bien délimiter quelles seraient les propriétés des Personnes et les propriétés des Films, par exemple, nous avons donc défini les namespace film et personne, comme ceci :

*xmlns:film="http://www.films.fr/film#"  
xmlns:personne=*[*http://www.films.fr/personne#*](http://www.films.fr/personne#)

A l’aide de ce schéma, nous avons pu créer manuellement un fichier RDF représentant les données que nous avons extraites du texte et formalisées.

La validation de ce fichier RDF nous a donné le schéma que vous pourrez trouver joint à ce document, et nous a permis de vérifier que le code RDF que nous avions écris était bien correct.

La structure du graphe obtenu à l’aide du validateur W3C correspond à la structure que nous avions prévu à l’aide notre schéma. La traduction du texte en RDF semble donc correcte.

La dernière étape de ce TP fut d’utiliser la librairie JENA, et de produire en Java du code permettant la génération automatique d’un fichier RDF représentant le texte. Le fichier RDF obtenu s’avère similaire à celui que nous avons créé manuellement, cela confirmant à nouveau que notre traduction en RDF est correcte.