Projet Tutoré

M1 Technologies de l'Internet

Conception et développement d'une application d'annotation thématique dans l'environnement Gate

---

Auteurs :
Roland Bary
Charles Follet

Tuteurs :
Marie-Noëlle Bessagnet
Annig Lacayrelle
Albert Royer
Christian Sallaberry

Remerciements

Nous tenons à remercier nos tuteurs pour leur pédagogie et leur encadrement. Monsieur Royer pour sa précision et sa connaissance pointue du domaine. Madame Lacayrelle pour son soutien et sa clarté. Monsieur Sallaberry pour nous avoir remis sur de bonnes pistes quand nous nous égarions. Et enfin, madame Bessagnet pour avoir assuré la coordination et le suivi de ce projet.

Table des matières

Première partie

Introduction

Avec l'évolution de manière significative des volumes d'informations sur internet, on peut observer une évolution du web vers une approche dans laquelle chaque donnée acquiert un sens afin de rendre possible une interprétation du contenu des pages web par des machines. Cette extension constitue le web sémantique.L'une des principales motivations du web sémantique est la recherche d'information sémantique.
C'est donc dans ce cadre que nous sommes intervenus pour répondre à l'appel d'offre de nos encadrants. L'objectif est l'annotation sémantique d'un document texte spécifique, qui constitue effectivement la première étape dans un processus d'indexation et de recherche d'information sémantique.

Au regard de ce qui a été exprimé en amont, se pose les problématiques suivantes :

- Existe-t-il des outils qui se prêtent aisément à l'annotation sémantique ?
- Quelle approche de conception peut nous permettre de réaliser cette étape d'annotation sur un document texte non-structuré ?

La résolution de ces différentes problématiques, nous à donc amené à organiser ce document comme suit : /*Il faut caller notre plan ici */ :) Une première partie dans laquelle nous présenterons le cahier des charges. Ensuite une seconde partie décrira quelques connaissances existantes sur le sujet avec les technologies utilisées au sein du projet.

Deuxième partie

Cahier des charges

I Contexte

A partir des travaux de Georges DEPEYROT sur les monnaies carolingiennes, nous avons travaillé pour une équipe parisienne de numismates sur l'annotation du Numéraire Carolingien [1].
Sur celui-ci, l'équipe a besoin d'effectuer des recherches :

Temporelles : Quelles étaient les pièces en circulation de l'an 859 à l'an 865 ?

Spatiales : Dans quels ateliers, les pièces de type Obole de Charlemagne ont été produites ?

Thématiques : Combien d'exemplaires de la monaie d'or de Charles le Chauve ont été étudiés ?

Répondre à cette demande implique de définir puis d'explorer les dimensions temporelles, spatiales et thématiques de l'ouvrage.
Pour celà, il est nécessaire de connaitre le domaine et l'ouvrage afin de savoir quelle information correspond à quelle dimension.
Une fois cet apprentissage fait, nous pouvons construire des règles dans une chaine de traitement permettant d'annoter chaque information en fonction de sa dimension.

Les monnaies carolingiennes sont le domaine central pour la réalisation du projet. Les ressources nécessaires à l'annotation (ici sous forme de gazetiers) ont été construites à partir des données de l'ouvrage.

Forte de son expérience dans le domaine, la maitrise d'ouvrage nous a demandé d'utiliser la boîte à outils logicielle GATE qui sera utile pour le traitement du langage naturel.

En résumé, les carractéristiques du projet sont :
- l'apprentissage et la compréhension du domaine considéré,
- l'étude des principes d'annotation de documents,
- le dévellopement d'une chaine d'annotation dans GATE,
- la mise en place d'une visualisation des résultats.

II Description

III Diagramme de Gantt prévisionnel

1. `http://www.cgb.fr/le-numeraire-carolingien-moneta-77-3e-edition-depeyrot-georges,Ln71,a.html`

Troisième partie

Cadre d'analyse

Quatrième partie

Développement

Cinquième partie

Conclusion