



PROJET INDIVIDUEL

PJI (54) - Mais que font nos députés ? Une sociologie informatique du travail parlementaire

Auteur : Quentin BAERT
Encadrant universitaire : Samuel HYM
Encadrant : Etienne OLLION

janvier à juin 2015

Sommaire

Remerciements	2
Introduction	3
1 Choix des outils	4
1.1 Le langage Scala	4
2 Récupération des comptes rendus intégraux de l'Assemblée nationale	5
2.1 Organisation des PDF sur le site de l'Assemblée nationale . .	5
2.2 Récupération des PDF	5
3 Filtrage des données	6
3.1 Conversion des PDF en fichiers textes	6
3.2 Analyse des fichiers textes pour isoler ceux contenant des scrutins	6
4 Mise en forme des informations récupérées	7
4.1 Représentation objet d'un scrutin	7
4.2 Extraction des données des scrutins	7
5 Construction des bases de données	8
5.1 Bases de données simples	8
5.2 Bases de données détaillées	8
Conclusion	9
Bibliographie	10

Remerciements

Introduction

1 Choix des outils

1.1 Le langage Scala

2 Récupération des comptes rendus intégraux de l'Assemblée nationale

2.1 Organisation des PDF sur le site de l'Assemblée nationale

2.2 Récupération des PDF

3 Filtrage des données

3.1 Conversion des PDF en fichiers textes

3.2 Analyse des fichiers textes pour isoler ceux contenant des scrutins

4 Mise en forme des informations récupérées

4.1 Représentation objet d'un scrutin

4.2 Extraction des données des scrutins

5 Construction des bases de données

5.1 Bases de données simples

5.2 Bases de données détaillées

Conclusion

Bibliographie