

Un outil d'exploration pour le Défi EGC 2016

Olivier Parisot, Yoanne Didry, Thomas Tamisier

Luxembourg Institute of Science and Technology (LIST), Belvaux, Luxembourg
olivier.parisot@list.lu, <http://www.list.lu>

Résumé. Dans le cadre du défi EGC 2016, nous avons développé une application web pour explorer les données décrivant les articles publiés depuis 2004 lors des conférences EGC. L'outil permet de découvrir les thèmes importants qui ont été abordés dans ces papiers. De plus, il permet de déterminer automatiquement les articles sémantiquement similaires à des thèmes donnés.

1 Introduction

Dans le cadre du défi EGC, l'association EGC a mis à disposition les informations concernant les articles présentés lors des conférences entre 2004 et 2015¹. Ces données sont fournies via un fichier CSV dans lequel les caractéristiques de chaque article sont détaillées : *a)* L'année de publication. *b)* Le titre. *c)* Les auteurs. *d)* Le résumé. *e)* La première page de l'article.

Afin d'extraire des renseignements utiles sur les articles parus dans la conférence EGC, nous avons développé une plateforme permettant de charger ce fichier afin de l'analyser. La première partie de l'article décrit l'architecture et les technologies utilisées au sein de la plateforme, tandis que la seconde détaille le processus d'extraction des données. La dernière section donne un bref aperçu des différentes fonctionnalités du prototype.

2 Prototype

La plateforme développée est une application multi-tiers : les traitements des données et les calculs sont réalisés côté serveur ; côté client, les résultats sont présentés à l'utilisateur dans une interface web. De cette manière, les calculs lourds peuvent être réalisés sur une machine de puissance raisonnable, tandis que l'interface de l'outil est accessible via un simple navigateur.

Du point de vue technique, le développement a été réalisé avec Grails, un framework permettant de développer des applications client/serveur performantes (Ledbrook et Smith (2014)) : Java/Groovy pour la couche applicative, HTML5/Javascript pour l'interface web.

Des bibliothèques JAVA et des API externes ont été utilisées dans la plateforme (Table 1). Ces composants ont été intégrés de différentes manières : *a)* Les bibliothèques JAVA et open source ont été intégrées directement dans l'application. *b)* Les APIs externes sont invoquées lors de l'exécution en effectuant des appels distants (une connexion réseau est donc nécessaire).

1. http://editions-rnti.fr/files/RNTI_articles_export.txt.zip