

Nouvelle approche de contextualisation de tweets basée sur les règles d'association inter-termes

Meriem Amina Zingla*, Mohamed Ettaleb**
Chiraz Latiri** Yahia Slimani*

*Université de Carthage, INSAT, laboratoire de recherche LISI, Tunis, Tunisie

** Université de Tunis El Manar, Faculté des Sciences de Tunis,
Tunis, Tunisie

1 Introduction

Les tweets sont des messages courts ne dépassant pas 140 caractères. Cette contrainte impose l'utilisation d'un vocabulaire particulier pour les rédiger et donc elle rend indispensable de connaître leurs contextes pour les comprendre. Pour ces raisons, nous allons nous concentrer sur la tâche de contextualisation des tweets attribuée à INEX2014¹. Les participants devaient fournir un contexte, pour permettre aux lecteurs de bien comprendre le tweet en utilisant un système de recherche d'information SRI et système de résumé automatique SRA. Dans cet article, nous proposons une nouvelle approche de contextualisation de tweets basée sur les règles d'association inter-termes.

Cet article est organisé comme suit : Dans la section 2, nous détaillons notre nouvelle approche. la section 3 sera consacrée aux différentes expériences menées, finalement nous concluons dans la section 4.

2 Approche proposée

Les étapes suivantes décrivent le processus de contextualisation des tweets :

1. Sélection d'une sous collection d'articles à partir de la collection de documents.
2. Annotation des articles sélectionnés en utilisant TreeTagger².
3. Extraction des noms à partir des articles Wikipedia annotés.
4. Génération des règles d'association en utilisant l'algorithme CHARM³ Zaki et Hsiao (2002).
5. Obtention de l'espace thématique de chaque tweet en projetant les tweets sur l'ensemble des règles d'association ;
6. Création de la requête à partir des mots de tweet et l'espace thématique.
7. L'envoi de la requête au système Baseline pour extraire le contexte de tweet.

1. <https://inex.mmci.uni-saarland.de/>

2. <http://www.cis.uni-muenchen.de/~schmid/tools/TreeTagger/>

3. <http://www.cs.rpi.edu/~zaki/www-new/pmwiki.php/Software/Software>