

Règles d'association inter-langues au service de la recherche d'information multilingue

Belhaj Rhouma Sourour, Asma Ben Achour, Malek Hajjem, Chiraz Latiri

Laboratoire de recherche LIPAH, Faculté des Sciences de Tunis
Campus Universitaire Tunis El Manar, 1060 Tunis, Tunisie.
sourour.bhr@gmail.com, asmabenachour@gmail.com,
malek.hajjem@gmail.com, chiraz.latiri@gnet.tn

Résumé. Dans cet article, nous proposons de montrer l'intérêt et l'utilité de déploiement des règles d'association inter-langues (RAILs) dans le domaine de la Recherche d'Information Multilingue (RIM). Ces règles sont des connaissances additionnelles résultantes d'un processus de fouille de grands corpus parallèles alignés au niveau de la phrase. En effet, leurs conclusions exprimées dans une langue cible représentent des traductions potentielles de leurs prémisses, exprimées dans une langue source. Nous illustrons l'utilisation des RAILs dans le contexte de la RIM à travers deux propositions, à savoir : (i) la traduction des requêtes et (ii) la traduction des termes de l'index. L'évaluation expérimentale a été menée sur la collection de documents MUCHMORE. Les résultats ont montré une amélioration significative de la pertinence système.

1 Introduction et motivations

L'abondance des documents, notamment sur le Web, dans de nombreuses langues a rendu nécessaire l'existence d'une Recherche d'Information Multilingue. La RIM consiste ainsi à formuler une requête dans une langue source et à rechercher des documents pertinents dans des langues cibles (RIM) Nie (2010). Dans un contexte multilingue, la requête ainsi que les documents ne sont pas représentés dans un même espace d'indexation étant exprimés dans des langues différentes. Par conséquent, la mise en correspondance de leurs descripteurs sera impossible. Ainsi, pour permettre une recherche multilingue, l'enjeu consiste à représenter les documents et la requête dans un même espace d'indexation.

Dans cet article, notre motivation est double : en premier lieu, il s'agit de déployer les techniques d'ECT qui permettent d'inférer des relations de traduction entre des unités linguistiques pour l'identification de lexiques bilingues à partir des corpus parallèles, et, en deuxième lieu, exploiter ces lexiques bilingues dans le cadre de la RIM Lavecchia et al. (2008); Latiri et al. (2010).