

Evaluation des outils d'extraction terminologique Quezao et Acabit

Edmond Lassalle*, Prem Kumar Casimir**
Emilie Guimier de Neef***

Orange Labs, 2 av. Pierre Marzin 22300 Lannion France
{edmond.lassalle, premkumar.casimir, emilie.guimierdeneef}@orange-ftgroup.com

Résumé. L'article décrit l'évaluation de deux outils d'extraction terminologique Acabit et Quezao. Si Acabit est plus connu car librement disponible, Quezao est issu des travaux d'Orange Labs sur la recherche d'informations. Après une comparaison sur les approches théoriques des deux systèmes, une évaluation concrète va porter sur un corpus d'actualité (2424Actu) pour l'aspect qualitatif et sur un corpus de presse pour l'aspect quantitatif (dont la tenue en charge).

1 Introduction

L'évaluation d'outils d'extraction terminologique (monolingue), en particulier pour le français, est à ce jour peu répandue [Nazarenko et al.]. La spécificité des langues, l'absence d'une fédération internationale sont en effet des freins à une campagne d'évaluation importante du même niveau que TREC [<http://trec.nist.gov/>].

Les rares campagnes d'évaluation comme ARC A3 [El Hadi et al.] ou CESART [Timimi] mettent en évidence la difficulté de comparer différents systèmes du fait de la diversité des fonctionnalités proposées. Des 9 systèmes participant à la campagne CESART (IDE/XTS, Lexter, WorldTrek, Termic, Termos, SeekJava, SynoTerm et Terminae, TermWatch), moins d'un système sur deux pouvait partager le même protocole d'évaluation. Enfin, le coût élevé de réalisation des corpus d'évaluation, en limitant leur nombre et le domaine traité, ne permet pas une estimation correcte de chaque outil en conditions réelles d'utilisation.

Le cadre de nos travaux sur Quezao et la présente évaluation sont dictés par des besoins internes à Orange. Il s'agit d'applications de Recherche d'Informations (RI), couvrant des domaines variées, de la documentation technique interne au contenu grand public (moteur de recherche généraliste ou dédié). L'évaluation doit porter sur des corpus d'applications réelles d'Orange. Elle a dû être menée en interne car les données propriétaires ne pouvaient être mises à disposition dans une campagne d'évaluation associant des partenaires externes. L'intervention humaine dans l'évaluation a donc été réduite, et consiste seulement à valider les termes produits en sortie de chaque système. L'absence de jeu de validation issu d'un corpus de test est cependant acceptable dans une évaluation de type «boîte transparente» où nous cherchons aussi à connaître le fonctionnement interne de chaque système.

2 Choix d'Acabit pour une comparaison avec Quezao

Les systèmes d'extraction terminologique (STE) récents sont basés sur la stratégie C/NC-