

RICSH : Recherche d'information contextuelle par segmentation thématique de documents

Rachid Aknouche, Omar Boussaid, Fadila Bentayeb

Laboratoire ERIC, Université Lumière Lyon2
5 avenue Pierre Mendès-France, 69676 Bron Cedex, France

{Rachid.Aknouche, Omar.Boussaid, Fadila.Bentayeb }@univ-lyon2.fr

Résumé

Le but principal des systèmes de recherche d'informations (SRI) classiques est de retrouver dans un corpus de documents l'information considérée comme pertinente pour une requête utilisateur. Cette pertinence est souvent liée à la fréquence d'apparition des termes dans le texte par rapport au corpus sans tenir compte du contexte de la recherche. Partant de ce constat, nous proposons dans cet article une approche pour la recherche d'information contextuelle par segmentation thématique de documents (RICSH). Cette approche s'appuie sur la méthode de pondération *tf-idf* que nous avons adaptée dans notre cas pour indexer le corpus. Cette adaptation se situe au niveau de l'importance du terme et de son pouvoir de discrimination par rapport aux fragments de textes et non au corpus. Ces fragments sont obtenus grâce à un processus d'identification des unités thématiques les plus pertinentes pour chaque document.

Summary

The principal goal of traditional Information Retrieval Systems (IRS) is to find relevant information for a user requests in a corpus of documents. This relevance, often relates to frequency appearance terms in the text without taking into account of their contex. We propose in this article an approach for a contextual information retrieval by topic segmentation of documents (*RICSH*). In this approach, we adapat *tf-idf* method to index the corpus. This adaptation focuses on term importance considering the fragment of texts rather than corpus. The fragmentation is obtain by an identification process of most relevant thematic units for each document.