
– **Conférence invitée 3** –

Automatic Query Expansion based on
Minimal Irredundant Association Rules

CHIRAZ LATIRI,
UNIVERSITÉ DE LA MANOUBA, TUNISIE

• Abstract

The steady growth in the size of textual document collections is a key progress driver for modern information retrieval techniques whose effectiveness and efficiency are constantly challenged. Given a user query, the number of retrieved documents can be overwhelmingly large, hampering their efficient exploitation by the user. In addition, retaining only relevant documents in a query answer is of paramount importance for an effective meeting of the user needs. In this situation, the query expansion technique offers an interesting solution for obtaining a complete answer while preserving the quality of retained documents, when the added terms to an initial query are accurately chosen.

Interestingly enough, query expansion takes advantage of large text volumes by extracting statistical information about index terms co-occurrences and using it to make user queries better fit the real information needs. In this respect, a promising track consists in the application of data mining methods to the extraction of dependencies between terms. In this talk, we present a novel approach for mining knowledge supporting query expansion that is based on association rules. The key feature of our approach is a better trade-off between the size of the mining result and the conveyed knowledge.

Thus, our association rules mining method implements results from Galois connection theory and compact representations of rules sets in order to reduce the huge number of potentially useful associations. An experimental study has examined the application of our approach to some real collections (Amaryllis 2002 and CLEF 2003), whereby automatic query expansion has been performed. We also study the case of association rules with high confidence values. The results of the study show a significant improvement in the performances of the information retrieval system, both in terms of recall and precision, as highlighted by the carried out significance testing using the Wilcoxon test.

Keywords Text mining ; Query expansion ; Information retrieval ; Association rule ; Generic bases ; Wilcoxon test