Hybrid Method for Personalized Search in Scientific Digital Libraries

Thanh-Trung Van and Michel Beigbeder

Centre G2I/Département RIM
Ecole Nationale Supérieure des Mines de Saint Etienne
158 Cours Fauriel, 42023 Saint Etienne, France
{van,mbeig}@emse.fr

Abstract. Users of information retrieval systems usually have to repeat the tedious process of searching, browsing, and refining queries until they find relevant documents. This is because different users have different information needs, but user queries are often short and, hence, ambiguous. In this paper we study personalized search in digital libraries using user profile. The search results could be re-ranked by taking into account specific information needs of different people. We study many methods for this purpose: citation-based method, content-based method and hybrid method. We conducted experiments to compare performances of these methods. Experimental results show that our approaches are promising and applicable in digital libraries.

1 Introduction

Search in digital libraries is usually a boring task. Users have to repeat the tedious process of searching, browsing, and refining queries until they find relevant documents. This is because different users have different information needs, but user queries are often short and, hence, ambiguous. For example, the same query "java" could be issued by a person who is interested in geographical information about the Java island or by another person who is interested in the Java programming language. Even with a longer query like "java programming language", we still do not know which kind of document this user wants to find. If she/he is a programmer, perhaps she/he is interested in technical documents about the Java language; however, if she/he is a teacher, perhaps she/he wants to find tutorials about Java programming for her/his course.

From these examples, we can see that different users of an information retrieval system have different information needs. Furthermore, a person can have different interests at different times. A good information retrieval system have to take into account these differences to satisfy its users. This problem could be solved if the system can learn some information about the interests and the preferences of the users and use this information to improve its search results. This information is gathered in *user profiles*. Generally, a user profile is a set of information that represent interests and/or preferences of a user. This information could be collected by implicitly monitoring the user's activities [1, 2] or by