Chapter III
Collaborative Classification for Group–Oriented Organization of Search Results

Keiichi Nakata
International University in Germany, Germany

Amrish Singh
International University in Germany, Germany

ABSTRACT

In this chapter the authors examine the use of collaborative classification to support social information retrieval by organizing search results. It subscribes to the view that the activity of collaborative classification can be characterized by top-down and bottom-up approaches, both in terms of the nature of concept classification and the process of classification development. Two approaches, collaborative indexing and search result classification based on shared classification schemes, are described and compared. It suggests that by allowing open access to classification development tools to generate shared classification schemes, which in turn become collaborative artifacts, cooperating user groups will generate their own coordination mechanisms that are not dependent on the system itself.

INTRODUCTION

It is widely understood that the task of information retrieval involves representation, storage, organization of, and access to information items (Baeza-Yates & Ribeiro-Neto, 1999). Moreover, the retrieved information should satisfy the user’s information needs. As such, the user carries the task of translating his or her information needs into a representation or language provided by a retrieval system. This in itself is often problematic; users may not be able to express their information needs adequately, or the results may not be satisfactory due to limitations in the expressivity
of the query language in articulating the information needs. Such problems that occur in individual information retrieval tasks are compounded when carried out in a group. The articulation of information needs of a group is far more complex than that of the individual.

The information retrieval task in the World Wide Web (WWW) is primarily assumed to be an individual task. This is reflected in the interaction design of most commercial search engines, in which users either enter search terms or browse through the categorized information space, often referred to as directories. Web-based search engines are engineered to provide links to pages that satisfy users' information needs. Most search engines return search results in a ranked ordered list. The ranks of the documents are determined by their relevance to the corresponding query. This relevance measure depends predominantly on the user's ability to suitably describe his or her information need as a query text. Since most queries are short, unconscious assumptions are made regarding the context of query terms, making the query less precise. This leads to a low precision in the retrieved results, and users are forced to manually sift through the list of returned pages to find relevant documents. This would be unproblematic if the users could easily separate irrelevant documents from the relevant ones. However, the presentation style of ranked lists used by most search engines does not make it easy to do so. Document ranking becomes virtually obsolete when documents lower in the list are more relevant to the user's information needs than the ones with higher ranks. Therefore, a major challenge for efficient Web search is to make search results helpful to the user even if the query is poorly formulated.

An argument in favor of the collaborative approach such as social information retrieval is that with the ever-increasing amount of electronically retrievable information, such as on the Internet, collaboration might be one of few ways to practically manage the complexity of information retrieval and management. To deal with the rate of growth in the information pool, which is in some cases already unmanageable by an individual, joining efforts in a group or community would enable us to manage the complexity of information retrieval such as indexing, organization, and storage, while maintaining reasonable coverage. Tools such as search engines and browsers enable users to engage in the exploration of a large information pool; there should also be effective tools for collaborative work that encourage such a participatory mode of information exploration and management.

One of the approaches to leverage groups' collaborative efforts in information retrieval is collaborative filtering, also referred to as social filtering. Essentially, collaborative filtering is an adaptation of relevance feedback by multiple members of a group. Relevance feedback can be seen as a form of query expansion through user feedback without the necessity to reformulate the query to improve the precision of the result. Being a query expansion implies that this applies effectively to single query. However, in the group context, the information retrieval task is often not carried out in isolation and the time scope is wider to allow for contributions from multiple members. When queries are made over time, the likelihood that the nature and context of query change also becomes higher. For such a situation, the model based on relevance feedback, such as collaborative filtering, may be too static, or not dynamic enough to capture such changes.

Instead, what we focus on here is the use of classifications based on conceptual structures. Such classifications can be shared and developed collaboratively by a group. They can be used to index documents or classify documents such as search results. If classifications can be dynamically restructured, the dynamic nature of group information needs may be dealt with more effectively.

A number of questions immediately arise from such an approach, especially for group scenarios.