博士論文要旨

Web Information Search and Sharing: A Human-Centric Integrated Approach

Roman Y. Shtykh Director : Qun Jin

In the situation of information overload we are experiencing today, conventional Web search systems taking one-size-fits-all approach are often not capable to effectively satisfy individual information needs. To improve the quality of Web information retrieval, information systems need to know every individual user's information needs. However, knowing and correctly applying information needs is extremely difficult, often impossible. Yet knowing multiple contexts of user information behaviour can give us some conception of conceivable information a user tries to obtain in a particular context.

Recognising the importance of human factor for overcoming problems related to information overload and the significance of knowing a user and his/her context for effective personalisation as a means to overcome these problems, we are seeking for solutions to capture contexts from user-system interaction data and employ the power of collaboration in order to achieve better search and sharing in information seeking tasks.

In our research project we set a user in the centre of information seeking task. Although fragmentarily, we endeavoured to capture users' contexts through relevance feedback obtained from user-system interactions and organise them as concepts of user interests. Concepts are stored in user profiles, which are used for personalisation of users' search experiences. We recognise that contexts are changeable and they change along with users' activities - therefore dynamism is an essential condition for organisation of contexts. In addition, in order to enforce the user-centrism of the approach we are seeking to employ 'social' elements of collaborative information search.

In the thesis we discuss a collaborative personalised search approach that makes an attempt to "understand" and better satisfy information needs for each and every searching user. The approach is realised with a Web information retrieval framework called BESS (BEtter Search and Sharing). In order to facilitate a user's information seeking activities, the framework captures his/her information contexts into dynamically changing user profiles, which are further used as a source of the user's interests and expertise and applied for search personalisation.

The thesis is organised in seven chapters.

In Chapter 1 we pose the problem of information overload and discuss its issues for information acquisition and processing. We emphasise the subjective nature of the problem and thus assume that it has to be solved with close consideration of human factor, including its social aspects.

In Chapter 2 we discuss human-centric solutions for information seeking and exploration with the main focus on personalisation, its advances in academy and business, and its problems, and deliberate on some personalisation-related issues, privacy in particular. The survey and discussion on user profiles as the main modelling approach in personalisation in general follow. Further, we give an extensive overview of user interest inference, its methods, design and use, as an important part of personalisation process, and outline the peculiarities of the proposed approach.

In Chapter 3 we propose the BESS collaborative information search and sharing framework in attempt to incorporate the discussed user-centeredness into information seeking tasks by harnessing relevance feedback from users in order to estimate their interests, construct user profiles reflecting those interests and, finally, apply them for information acquisition in online collaborative information seeking context. We discuss the conceptual basis of the framework and show its model and architecture.

To know user contexts, which are, as we have noted, very important to achieve the goals of this research and make search and sharing user-centred, we extract concepts that reflect a user's interests and organise them in multi-layered user profiles. User profile is the central part of the proposed framework. It contains one static and one dynamic part, where the latter reflect the current, short-term and long-term user interests. The layers are dynamically updated with changes of the user's interests. Their construction and update are based on *interest-change-driven profile construction* mechanism that adopts *recency, frequency* and *persistency* as the three important criteria reflecting the volatility of user interests and emphasising the steadiness of persistent preferences. User modelling is discussed in Chapter 4.

Concepts for profile construction can be obtained, for instance, by grouping/classifying user relevance feedback items, such as the objects selected by a user during his/her interaction with a search system or explicitly stated user preferences. In Chapter 5 we propose *High Similarity Sequence Data-Driven clustering* algorithm for concept extraction based on similarity characteristics of uniform user relevance feedback. The algorithm is online and incremental, and does not require any prior knowledge of data to be processed. Having been evaluated against different data sets, the algorithm proves to be fast, easy to implement and produces reasonable clustering results.

In Chapter 6 we show how the dynamics of interest-change-driven user profile, its contextual information and its construction process are employed for contribution co-evaluation and search. In BESS, interacting with *objective index* data of conventional search services a user creates his/her individual repository (*subjective index*) and thus contributes to the whole community of the system. Such contributions are different from objective index data - they have an additional value inferred from users' expertise derived from their individual user profiles. Therefore, when searching subjective index repositories, a document evaluated by users with deep expertise in the topic the document belongs to get higher ranking. We propose the following two factors for users' expertise assessment - *contribution activeness* and *contribution popularity*. Furthermore, in order to narrow down the number of documents to be searched, a searcher can use *focused search*, a search on the documents contributed by users with similar contexts.

Finally, Chapter 7 concludes the thesis with the summary of the presented research and outlines future research issues.

In this study we have proposed a human-centric integrated approach for Web information search and sharing. The approach has incorporated the important user-centric elements, namely a user's individual context and 'social' factor realised with collaborative contributions and co-evaluations, into Web information search. The major contributions of this study are as follows.

1) A framework for better search and sharing has been proposed;
2) The dynamic interest-change-driven model for accurate user contextual information organisation has been devised;
3) The Similarity Sequence Data-Driven clustering algorithm for concept construction has been developed;
4) Measures to evaluate our proposed model and algorithms have been defined and introduced.

The major differences of our approach from many modern conventional personalisation solutions on the conceptual level are

1) separation of subjective value-added resources from impersonal objective data of conventional search services - creation of *'social' sub-space* for subjective information accumulation and sharing through search;
2) tight coupling of information sharing and search in the following personalisation cycle (Figure 1)

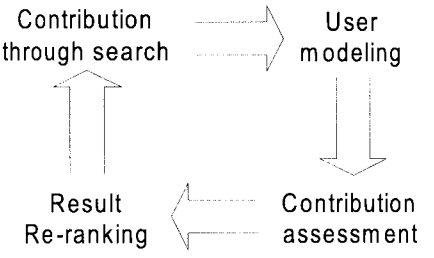

Figure 1. Personalization Cycle

The validity and expected quality of the personalisation cycle are ensured by several important dynamic mechanisms, such as H2S2D similarity-based online incremental algorithm and *interest-change-driven profile construction*. In spite of its relative simplicity, the proposed H2S2D method has demonstrated reasonably good clustering results in terms of accuracy and precision and has proved to be suitable for fast real-time relevance feedback processing to guarantee always-updated concepts. Further, the proposed *interest-change-driven profile construction* mechanism ensures proper organisation of volatile and persistent user interests, thus, together with H2S2D clustering, always providing accurate information about the user's context change. The experiments have demonstrated the validity and efficacy of the proposed human-centric integrated approach for Web information search and sharing. BESS is a general framework following our ideas of collaborative search and sharing, and its components implementing most methods can be separated from it to be applied to other solutions.

The achievements of our study are highly expected to contribute to the field of Web information search in general, and contextual personalisation in particular.