iMecho: a Context-aware Desktop Search System

Jidong Chen [*#], Hang Guo [#]
[#]EMC Research China, Beijing,China
{jidong.chen,hang.guo}@emc.com

Wentao Wu [*], Wei Wang [*]
[*]Fudan University, Shanghai,China
{wentaowu,weiwang1}@fudan.edu.cn

ABSTRACT

In this demo, we present iMecho, a context-aware desktop search system to help users get more relevant results. Different from other desktop search engines, iMecho ranks results not only by the content of the query, but also the context of the query. It employs an Hidden Markov Model (HMM)-based user model, which is learned from user's activity logs, to estimate the query context when he submits the query. The results from keyword search are re-ranked by their relevances to the context with acceptable overhead.

Categories and Subject Descriptors: H.3.3 [Information Search and Retrieval]: Search process

General Terms: Algorithms, Design.

Keywords: Context-aware search, User model, Desktop search.

1. SYSTEM OVERVIEW

Searching personal data with keywords is becoming increasingly challenging as users create more and more personal data, e.g., large amounts of local files, emails or cached web pages in their PCs. Current desktop search systems such as Google Desktop only support keyword searches that rank results by their relevance to the query. One problem with keyword search is that the result set usually contains many unrelated documents. To solve it, we need to find out what the user really needs besides using keywords.

In most cases of desktop search, the user is more interested in documents related to his context. By "*context*" we mean the user's task when he submits the query. A task can be simply modeled as a set of documents accessed by the user for a specific purpose. A user constantly goes back to his previous tasks when he is using his PC. A user study [2] shows that about 40% tasks are re-opened by the user within a week. Another 23% tasks are re-opened within a month.

Following this observation, we develop a context-aware desktop search system called iMecho. The idea is to build a user behavior model based on the user's previous activities.
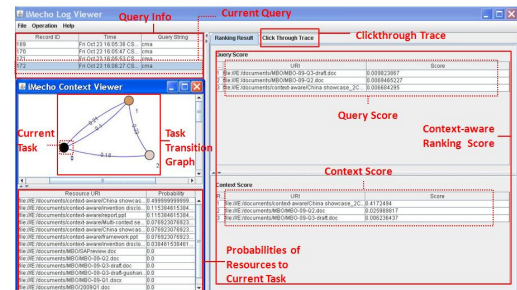
Figure 1: Interface of iMecho Query Viewer

When he submits a query, the model estimates his hidden task, i.e., the context of the query, and then the system refines the ranking results accordingly with reasonable overhead. iMecho employs the Hidden Markov Model (HMM) as the user model to estimate hidden tasks. In this model, a user keeps going from one task (a state of HMM) to another. Though a recent work [1] also applied HMM in context-aware search, it mainly focused on web searches and only provided context-independent results for a particular user. iMecho can record the complete trace of user activities on his desktop and the query context of the specific user can be comprehensively modeled by the sequence of user activities with accepted cost. iMecho works as follows:

- The offline part: the user's click sequence is collected to build a user model. The model includes 1) the probabilities of a task sequence; 2) the probability that a document is related to a specific task.

- The online part: when the user submits a query, the latest user model is used to estimate his current task given real time click sequence. And the ranking score of each retrieved document is adjusted accordingly.

Figure 1 shows the interface of iMecho query viewer, which gives details of the user model, the transition of query contexts and query-related information (e.g. ranking scores and the clickthrough data) in the query time. We will demonstrate how to build the user model from user activity logs and how to apply the model to rank the results.

2. REFERENCES

[1] H. Cao, D. Jiang, J. Pei, E. Chen, and H. Li. Towards context-aware search by learning a very large variable length hidden markov model from search logs. In WWW 2009.

[2] D. Elsweiler and I. Ruthven. Towards task-based personal information management evaluations. In SIGIR 2007.

3. SYSTEM REQUIREMENTS DESCRIPTION

iMecho is a desktop search system based on a general full text index (e.g. Lucene). It only needs a commodity personal computer/laptop with Window OS. Network connections are not necessary for iMecho.

iMecho collects user activities in the desktop by application plugins as background deamons. We have developed plugins for MS-Office documents, Emails (Outlook), Adobe *pdf* files and cached webpages (Firefox) in iMecho. Therefore the software requirements of iMecho contain JDK1.6.0 Runtime, Mozilla Firefox 3.0, Adobe Acrobat, and Microsoft Office 2003/2007. If JDK is not installed, the iMecho system will not work. If Firefox or Acrobat or MS Office is not installed, the corresponding add-on will not be installed and therefore the user activities in these applications will not be collected.