

# iMiner: From Passive Searching to Active Pushing

Deyi Li

Institute of Electronic System Engineering, Beijing 100039, China

The success of a search engine in cloud computing environment relies on the numbers of users and their click-through. If we take the previous search key words as tags of users to study and differentiate the user interaction behaviors, the search engine is able to actively push related and useful information to users based on their previous actions instead of passively waiting for users' queries. However the user searching behavior is affected by lots of factors, and it is quite complex and uncertain. The log files provided by a search engine have recorded all the information of the user interaction process on their servers or browsers, such as key words, click-through rate, time stamp, time on page, IP address, browser type and system stats, even the user location etc, which are all important information to understand and categorize users' searching behavior. Is there any statistical property almost independent to search key words? How to push recommendation based on the queried key words? And how to extract user behavior models of searching actions in order to recommend the information to meet users' real needs more timely and precisely?

In order to answer these questions, we don't think there is only one correct or optimal solution. We take the do-the-best strategy with uncertainty. Statistics play a great rule here. From the statistical point of view there are two very important distributions: the Gauss distribution and the power-law distribution. We suggested and have already studied a cloud model to bridge the gap between the two by using 3 mathematical characteristics: Expectation, Entropy and Hyper entropy. The collective intelligence may appear under certain conditions. The cloud model can well describe the preferential attachment in users' habits, which means that a continuous or periodic habit may reappear again and again, and even the user interest is changed from time to time.

We are developing an iMiner system at present, a mining platform that can actively push useful information to users, rather than a traditional search engine which only passively waits for users' key words. To support this kind of user-oriented pushing service, iMiner models users' behaviors according to their identities, status, interests, and previous querying records, preference of results, temporal search patterns, and interest shifts. It is made possibly with the modules of huge log data pre-processing, users' actions modeling, query modeling and query matching etc. We try to implement all of the mining services on a virtual cloud mining service center. The supported services in cloud computing are user-oriented to meet various groups which may be small but very active and frequently click-through the pages. The supported services may even be sensitive to the users' environment such as the present location and time.