

## Techniques for Improving Search Efficiency and Scalability

Ismail Sengör Altingövde  
Bilkent University, Turkey

### Abstract:

In the last decade, academia and industry sought for and found out several optimization techniques to allow an inverted index to cope with the massive amount of text data on the Web. In this talk, I will broadly focus on two such directions for optimization, namely, index/query pruning and result caching, and present some approaches developed in our research group. In the first part of this talk, I will describe a new data structure, namely, cluster-skipping inverted index that is specifically tailored for cluster-based data collections, and present results for employing it in a category-restricted search scenario for Web directories. Next, I will argue using the same data structure for dynamic pruning purposes, i.e., to prune the least promising Web sites as a whole during the query processing. In the second part of the talk, I will briefly discuss our on-going research on static index pruning, and present our findings on a detailed comparison of two well-known techniques from the literature. In the last part of my talk, I will present a cost-based caching approach, which takes the query processing costs into account, for static and dynamic caching. Finally, I will also outline a result presentation strategy for navigational queries that is proposed to utilize the bandwidth and cache space usage.

Short biography: Ismail Sengör Altingövde is a PhD candidate in the Computer Engineering Department of Bilkent University in Ankara, Turkey. His research interests include efficiency issues for large scale information retrieval, machine learning, Web mining and query processing in relational database systems. He received his M.S. in computer engineering from Bilkent University in 2001. Contact him at the Department of Computer Engineering, Bilkent University, Ankara 06800, Turkey; [ismaila@cs.bilkent.edu.tr](mailto:ismaila@cs.bilkent.edu.tr).