Max−Planck−Institut für Informatik

Databases and Information Systems Group (AG 5)

Prof. Dr.-Ing. G. Weikum   Dipl.-Inform. Hanglin Pan
Dipl.-Inform. Sergej Sizov
Dipl.-Inform. Stefan Siersdorfer

Selected Topics in Web Information Retrieval and Mining (WS 03/04)

Assignment 2

Handout on: Thursday, November 6, 2003

Due on: solutions will be discussed on Friday, Nov 14, 2003

Exercise 2.1:

Write pseudocode (in Java, C++, C, or Pascal-like notation) for the Adaptive OPIC method that can handle an evolving link graph and uses time windows.

Exercise 2.2:

To what extent and how can you combine OPIC with other extended features of link-analysis-based authority ranking, most importantly, topic-specific authority or text-relevance-based edge weighting (as used in the HITS extension by Bharat and Henzinger)?

Exercise 2.3:

Suppose you have temporal information about Web pages and hyperlinks (as assumed in Exercise 1.4).
How could you incorporate this information into an OPIC-style authority computation?

Exercise 2.4:

Prove the Decomposition Theorem by Jeh and Widom.

Hint: You can exploit the fact that a solution to the fixpoint equation for a Page-Rank vector is unique; so syntactically different solutions that all satisfy the fixpoint equation must be equal.

Exercise 2.5:

Analyze the amount of storage space that the Jeh/Widom algorithm for personalized Page-Rank computation needs
a) in the precomputation phase and
b) at the time when a user provides her specific personalized preference vector.
Compute the required space in terms of the cardinalities of V (the total node set), L (the link graph), H (the hub set), and u (the non-zero entries of the actual preference vector of the user), as well as relevant tuning parameters such as Q (the node set in the Selected Expansion Algorithm).

Exercise 2.6:

How can you exploit the notion of a Web Skeleton, as proposed by Jeh and Widom,
for fast computation of an approximate global Page-Rank vector?