DEPARTMENT OF MATHEMATICS AND COMPUTER SCIENCE
UNIVERSITY OF SOUTHERN DENMARK, ODENSE

COMPUTER SCIENCE COLLOQUIUM

Inside Google's Web Page Scoring System

Marco Gori
Dipartimento di Ingegneria dell'Informazione
Universita' di Siena

Tuesday, September 18, 2001, at 2:15 PM
The Seminar Room

Abstract: A crucial problem for search engines is that of sorting answers to a given query according to the relevance of the retrieved documents. By and large this is regarded as a very difficult problem which involves the content of single pages as well as their hyperlinks. In this talk, I'll focus attention on recent developments on web page scoring systems, which only depend on the topology of the graph expressing the links among web pages. In particular, I'll look inside Google's approach to page scoring which relies on the principle that the authority of a page depends on the selective citations of highly authoritative pages. I'll present some properties of the corresponding scoring scheme and discuss the page score for special graph topologies.

I'll argue that the horizontal search scheme of nowadays search engines is inherently limited, in the sense that the coverage of the web and the freshness of the information cannot be simultaneously reached (coverage-freshness dilemma). I'll advocate the importance of focussed search engines to overcome the previous dilemma and, beginning from Google, I'll present novel web page scoring systems properly conceived to retrieve information on specific topics.

Finally, I'll advocate the importance of exporting learning methodologies from the machine learning literature to the Web. This is likely to open the doors to the study of novel page scoring systems oriented to specific topics, as well as to the development of personalized interactions.

Host: Klaus Meer