

Review of “What is this Page Known for? Computing Web Page Reputations”
(by Vinay Goel)

This paper presents a web search approach that returns a ranked set of topics on which a particular page has a reputation. The authors provide several algorithmic formulations of the notion of reputation using simple random walk models.

They present (convincing) potential applications of such an approach: self evaluations by organizations, page classification, reputation of home pages to determine what topics a person is known for etc.

The authors propose two methods for computing the reputations of a page. The first method is based on one-level weight propagation, generalizing the PageRank model. The second method is based on two-level weight propagation, generalizing the Hubs and Authorities model.

The authors do a good job of presenting their methods in the form of formal models and then going on to prove that the notion of reputation rank for the particular models are well defined. The random walk formulation of the hubs-and-authorities model presents an interesting way to look at such a model; it makes intuitive sense.

They present approximate algorithms of their approach in the case of a lack of access to a large Web crawl or the lack of time to run full computation. They also highlight the duality of topics/terms and pages. But the idea of ‘topic’ used is highly generalized. The approach in Section 3.1 bases its decision-making on the absence/presence of a term in a page, which is probably too simplistic. Using ‘synonymy’ of terms to capture similar terms may help provide a better sense of a ‘topic’.

The evaluation section is not too convincing. It would be desirable to have results presented for a larger test set --- to possibly provide some kind of topic distribution of the web.

In conclusion, this is an interesting paper but could use a little work in formalizing the notion of a ‘topic’ and presenting more evaluation results.