

Estudo da qualidade dos links na Web através dos algoritmos PageRank e HITS

Thiago Locatelli dos Santos, Gudelia Morales

Introdução: A web cria muitos desafios para recuperação de informação. Sendo esta muito grande, heterogênea e com o aumento de usuários navegando, pesquisando e inserindo suas páginas online, a dificuldade para se recuperar a informação tende a ficar mais complicada. Estima-se que a quantidade de sites novos por ano deve dobrar. **Objetivos:** O objetivo deste artigo é analisar os algoritmos de classificação (*ranking*) utilizados pelos sites de pesquisa para classificar a importância de uma página na web. Resume-se o artigo aos algoritmos de PageRank e HITS, que são os mais utilizados pelos principais provedores de busca (Google, Yahoo, Bing, Ask, etc.). Além disso, tem-se como objetivo dar continuidade a produção do manual sobre a ferramenta Gephi, que vem sendo desenvolvido durante toda a pesquisa, incluindo as análises realizadas. **Metodologia:** Para a representação das redes web foi utilizado o software Gephi e, para o estudo e análise dos algoritmos, o software MS Excel 2013. Além disso, artigos relacionados ao tema. **Resultados:** Foi possível realizar o estudo dos algoritmos utilizando os conceitos algébricos de autovalores e autovetores de matrizes que representam os grafos, além dos exemplos extraídos da literatura pertinente e de redes já existentes que foram vistas durante a pesquisa. Após análise dos algoritmos, pôde-se constatar que ambos dimensionam a qualidade do link em questão através da qualidade dos links que estão ligados a ele. Ou seja, mesmo que um link A possua mais links ligados a ele do que um outro link B, se a qualidade dos links que possuem ligação com o link B forem de melhor qualidade, o link B possuirá melhor classificação que o link A. **Conclusão:** Pôde-se observar que os algoritmos de classificação de páginas web podem ser utilizados para outros meios de pesquisa, como, por exemplo, a avaliação de pontos centrais (hubs) e/ou mais importantes (autoridades) em uma rede social e em cadeias de suprimentos. Para o caso de redes web, os hubs são considerados os links que distribuem a informação, ou seja, fazem referência a outros sites que possuem uma boa qualidade, ou não, das informações que o usuário procura. Já as autoridades são os links que possuem boa qualidade, sendo os primeiros links a aparecerem quando o usuário digita o que está procurando na web.

Palavras-chave: Web, PageRank, HITS.

Instituição de fomento: CNPq