

A Comparative Study of Link Analysis Algorithms

Chamath Palihawadana
 Department of Software Engineering
 Informatics Institute of Technology
 Colombo, Sri Lanka
 chamath.2014245@iit.ac.lk

Guhanathan Poravi
 Department of Software Engineering
 Informatics Institute of Technology
 Colombo, Sri Lanka
 guhanathan.p@iit.ac.lk

Abstract— With the advancement of technology, the internet has become a growing place with huge amounts of information. Retrieving information is not a hard task, but ranking the retrieved information according to the relevancy is a very complex task. Search engines provides users to query and assess results instantly. Behind these search engines there runs vast amounts of algorithms to give out the best results. One such algorithm type is the link analysis algorithms which is used to rank pages and content. This paper reviews and compares the most commonly used link analysis algorithms.

Keywords—component; Link Analysis, HITS, PageRank, Page Ranking, Web ranking, Information Retrieval.

I. INTRODUCTION

With the rapid growth of the world wide web the information that is available also grows at a very high rate. It is estimated that there are 25 billion pages on the internet as of October 2008 [1]. Recently with this growth, it has been very difficult to retrieve relevant information at high speed and accuracy. Search Engines are playing a major role at offering services to billions of users to find their required content [2]. Also Search engines are given the task to rank resulted web pages in a very effective manner so that the most relevant is always being at the fingertip of a user. Also, users tend to pick the most readily available information thus neglect on searching the most relevant information. This impatience and reluctance of users causes the second page of almost any search engine to be untouched [3]. It has become a very crucial area for search engines to optimize their core function of ranking web pages for a given query [4]. Users depend on search engines not only to get pages related to a query but also to make sure the good sites are separated from the bad sites among the deep links within the internet [5]. Web page ranking has different criteria used by different ranking algorithms, however the linked structure (hypertext) of the web pages has made the researchers work more efficiently on developing these algorithms from a long time [4]. Through this linked (connectivity) based nature, derived the link analysis algorithms to rank pages [2].

Link analysis algorithms are used to measure relationships among nodes. It also plays a major role in information retrieval and delivery of information in a desired order [2]. In this paper, a comparative study is done on the widely used link analysis algorithms which are used to rank webpages.

II. PAGERANK

PageRank is widely discussed as the best approach for ranking web pages and it also became popular as early Google versions used PageRank for their search engine [5]. PageRank Algorithm was developed by Larry Page and Sergey Brin at the Stanford University in 1996. The authors [4] have taken the inspiration from previous techniques which were used to analyze academic citations done by others in the world wide web. This has given them the motivation to go beyond just academic citations and rank over 150 million web pages at the start of the billion-dollar company Google Inc. Since simple citation counting had many drawbacks when counting web hyperlinks PageRank algorithm tried solving it by looking at the relative “importance” of a link.

The PageRank Algorithm only looks at the link structure towards a page [4]. So this algorithm doesn't have any effect on the page content. Most importantly this algorithm only considers the backlinks which are called inbound links and outbound links which are pointing towards other pages. These links are the ones which point towards a specific site.

Figure 1 below provides a high level view on the link structure of web pages including its in-pages and out-pages.

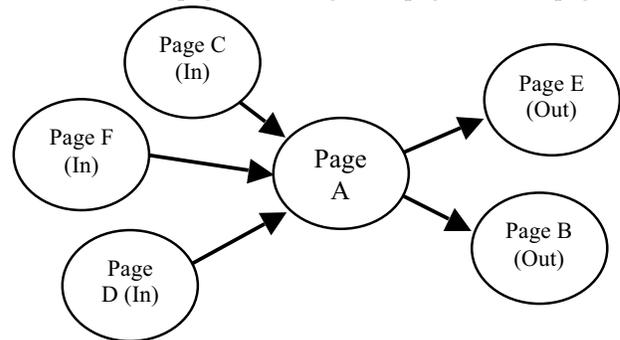


Figure 1. Inbound and Outbound links

The authors have clearly defined that a page will have high rank if the inbound pages (backlinks) are high. And also, when a page has few high ranked backlinks then the page will get a high rank.