



International Journal of Software and Web Sciences (IJSWS)

www.iasir.net

A Novel Technique for Fetching User Relevant Pages Using Backlinks

Sanjeev Dhawan¹, Vinod²

1Assistant Professor, U.I.E.T., Kurukshetra University, Kurukshetra, Haryana, INDIA

2Research Scholar, U.I.E.T., Kurukshetra University Kurukshetra, Haryana, INDIA

Abstract: Backlinks, which are sometimes called inbound links, are incoming links to a web page or the entire website. Search engines have measure the number of backlinks a website or web page has, and ranks those web pages with more backlinks in a higher position as inbound links are important to search engine rankings. Back links play two important roles they direct traffic to your web site and factor in the search engines deciding the position of your web pages in the index of results. Web pages use different types of backlinks to make the linking a mutually beneficial objective, as well as to build the number of backlinks to the website. A web page that has more backlinks than another with similar content will rank higher than the other page, simply because it seems to be more popular with visitors and other websites.

I. Introduction

Most people use *Internet* and *World Wide Web* interchangeably, but fact in the two terms are not synonymous. The WWW is a huge situate of interlinked images, documents and other resources, linked by hyperlinks and URLs. Search engines are useful for finding information on the WWW, such as Google, Yahoo! and AltaVista. Search engines provide an interface to a group of items that enables users to specify criteria about an item of interest and have the engine find the matching items. Backlinks, which are sometimes called inbound links, are incoming links to a web page or the entire website. A web page that has more backlinks than another with similar content will rank higher than the other page, simply because it seems to be more popular with visitors and other websites. Many search engine use backlinks to determine page rank too. This means that many websites have engaged in paid linking, which boosts their backlink numbers. This has caused search engines to add in specifications to use to determine backlinks that now research whether the backlinks have been paid for, or are real. Only genuine backlinks help web pages rank well on most search engines.

II. Types of Backlinks

There are many types of backlinks which is given below [6]:

Links from scraper sites: Before submitting your website to a directory listing, be sure the site is a legitimate directory. Scraper directories steal traffic rather than driving more and typically use frames, no-follow tags, or simply omit the backlinks intended for your site. Gaining links from these websites will benefit their rankings rather than your own.

Backlinks from link farms, link exchanges, and similar groups: In Google's eyes, any types of link schemes are not approved and can seriously decrease your rankings. This includes buying, swapping, and giving away links simply to generate links. There are some reputable programs that have developed acceptable methods but traditional link farms and backlink exchanges are noticed by search spiders.

Links generated by scripts and software: There are a variety of programs available online that promise to increase backlinks to your website. Typically this is done with a software script that can be set to run automatically, leaving comments on forums and blogs that include your links. These scripts are annoying to other webmasters, are often ineffective because the comments are blocked by spam catchers, and can get your website blacklisted by search engines.

Backlinks from "bad" sites: Just as backlinks from respected, high ranking websites can provide great benefits to your own, backlinks from blacklisted websites or those deemed unsuitable by search engines can damage your standings. Avoid incoming links from gambling sites, adult entertainment, and other questionable niche communities.

Website directories designed to increase rankings: Directories are a great way to drive traffic to your website, but not all are effective link building tools. If you're hoping to gain a solid backlink from a directory listing, look for the following characteristics: The directory page is included in the website's navigational links and/or sitemap.

- The page is not using the no follow or no index attribute.
- The directory should be a static database rather than dynamic so it is easily read by search robots.

- Organized links arranged in categories or grouped by similar topics. A mish-mash of URLs thrown on a page generally signifies spam to the search engines.

This paper presents a novel technique for Fetching User Relevant Pages Using Backlinks. The rest of the paper is organized as follows. Section 2 discusses the related work. Section 3 presents our proposed technique and algorithm and finally Section 4 presents our conclusion.

III. Related Work

Krishna and George [1] presented the response to a query a search engine returns a ranked list of documents. If the query is on a popular topic (i.e., it matches many documents) then the returned list is usually too Long to view fully. Studies show that users usually look at only the top 10 to 20 results. However, the best targets for popular topics are usually linked to by enthusiasts in the same domain which can be exploited. Clara Yu *et al.* [2] described the need for smarter search engines is a research suggesting several methods of improving search engine relevancy including latent semantic indexing and multi-dimensional scaling. The network structure of a hyperlinked environment can be a rich source of information about the content of the environment. A set of algorithmic tool was developed for extracting information from the link structure of an environment that effectiveness in a variety of contexts on the World Wide Web. Sepandar Kamvar [3] observed that the convergence patterns of pages in the Page Rank algorithm have a non uniform distribution. Specifically, many pages converge to the true Page Rank quickly, while relatively few pages take a much longer time to converge. Furthermore, it has been observed that these slow-converging pages are generally those pages with high Page Rank. This observation was used to devise a simple algorithm to speed up the computation of Page Rank, in which the Page Rank of pages that have converged are not recomputed at each iteration after convergence. This algorithm, which we call Adaptive Page Rank, speeds up the computation of Page Rank by nearly 30%. Brandman *et al.* [4] introduced simple and easy to incorporate modifications to web servers so that there are significant bandwidth savings. When crawlers periodically revisit pages at a website to gain updates, the web server uses its network bandwidth to respond to these crawlers. Media-specific crawlers that index audio, video and images are gaining in popularity. After studying this paper crawler identify recently changed material. The only way a crawler can find the location of pages is by downloading HTML pages and following links. Much of the downloaded HTML is just a means to discover some smaller set of desired pages. Meta-data describe the entire website.

Hector and Molina [5] elaborated their research work to generate significant loads on Web servers. A crawler (or spider) is a program that visits Web pages, using links on seen pages to identify more pages. The crawled pages are typically used to build indexes for search engines such as AltaVista and Google. However, crawlers are also used together pages for data mining, or to produce Web caches than can be more conveniently accessed by users. Wu and Davison [7] presented the ideas of generating a seed set of spam pages and then expanding it to identify link farms. First, we used a simple but effective method based on the common link sets within the incoming and outgoing links of Web pages for selecting the seed set. Trust Ranks is that good pages usually point to good pages and seldom have links to spam pages. They first select a bunch of known good seed pages and assign high trust scores. The spam page set by adding pages is expanded that have at least 3 outgoing links to the pages that have already been marked as scam pages. They found that pages with in link farms are densely connected with each other and many common pages exist both in the incoming and the outgoing link sets for a page in a link farm.

IV. Proposed Technique

All search engines determine what will be returned in the SERPs (Search Engine Results Pages) by using two measures authority and relevance. Relevance means the page contains the keywords and authority means the page has back links to it from other web pages. The search engines determine the order in which web pages are indexed on the results pages by the number of back links to the page and their respective authority. Back links play two important roles they direct traffic to your web site and factor in the search engines deciding the position of your web pages in the index of results. Browsers who discover and click on back links containing keywords associated with their interests will be directed to your web pages. This text in the back link is known as ‘anchor text’ and this too is taken into consideration by the search engines. Whilst all back links are of value some are more valuable than others. The search engines give authority to web pages which can be passed on to your web pages through the back links. The higher the authority of the web page ‘sending’ the back link the more authority is likely to receive by your web pages. According to relevance is Structure of web page shown in the figure 1.

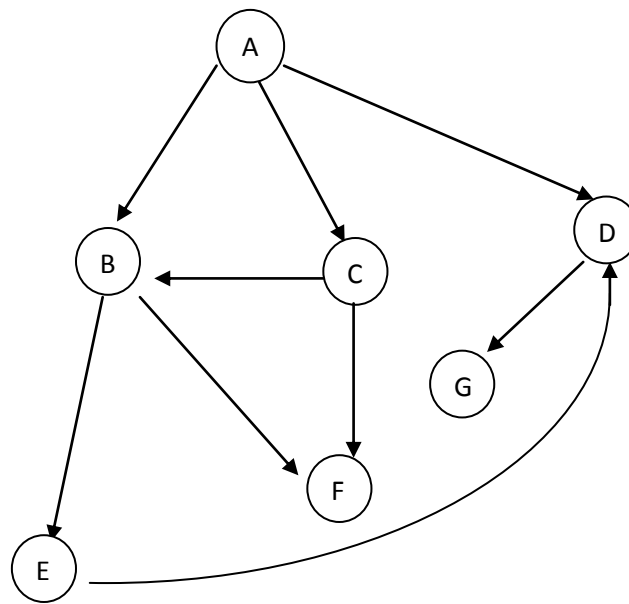


Figure 1. Structure of web page

Figure 1 consists of 7 classes for backlinks web taxonomy (i.e. A, B, C, D, E, F, G)

Table 1. The distribution of links to classes

A	B	C	D	E	F	G
1 URL FOR B	1 URL FOR E	1 URL FOR B	1 URL FOR G	1 URL FOR D		
1 URL FOR C	2 URL FOR D	3 URL FOR D		2 URL FOR G		
3 URL FOR D	1 URL FOR F	2 URL FOR E				
2 URL FOR E	3 URL FOR G	2 URL FOR F				
3 URL FOR F		4 URL FOR G				
2 URL FOR G						

EXAMPLE (for table 1)

BackLinks for F URL

F has 3 URL

F is available in the B & C Classes.

B is available in the A & C Classes.

C is available in the B Class.

Step 1

F->1 B, F->2C

Step2

F->1 B->2C->Root

F->2C->Root

F->1 B->Root

*C Path Merge
(As common Path)*

Similarly Backlinks for G

G has 4 Paths

G is available in the B, C, D & E Classes.

B is available in the C Classes.
C is available in the B Classes.
D is available in the E, C, & B Classes.
E is available in the C&B Classes.

Step 1

$G \rightarrow 3B, G \rightarrow 4C, G \rightarrow 1D, G \rightarrow 2E$

Step 2

$G \rightarrow D \rightarrow \text{Root}, G \rightarrow D \rightarrow E \rightarrow B \rightarrow \text{Root}, G \rightarrow D \rightarrow E \rightarrow C \rightarrow \text{Root}$

Algorithm backlinks

Step 1: Total no of path for back links

Step 2: If match the path in the table then

Go to that in table (path will be available then carries on continued)

Step 3: URL available more than two in table then merge

Step 4: Shortest path will be considered

Step 5: Link is stored

Step 6: Finishing the links

V. Conclusion

The backlink search engine results are the web pages that point to the compact URL. The search engines determine the order in which web pages are indexed on the results pages by the number of back links to the page and their respective authority. Back links play two important roles they direct traffic to your web site and factor in the search engines deciding the position of your web pages in the index of results. This paper discusses the structure of a web page and various backlinks used in it. The proposed technique finds relevant web pages for this structure using backlinks. The proposed technique is less time consuming, do not contains duplicate backlinks and contains relevant web pages.

VI. References

- [1] Krishna Bharat, "Search Pad: Explicit Capture of Search Content to Support Web Search" The International Journal of computer Telecommunication Networking Volume Issue 1-6, June 2000 pages 493-501.
- [2] Clara Yu, John Luadardo, Maciej Ceglowski, J. ScottPayne "Patterns in unstructured data" National Institute for Technology & Liberal Education (NTILE).
- [3] Kamvar, Sepander & Haveliwala, Taher & Golub, "Adaptive methods for the computation of page rank" Gene (2003) Technical Report Standford.
- [4] Onn Brandman, Junghoo Cho, Hector Garcia-Molina, Narayanan Shiva Kumar "Crawler-Friendly Web Servers" Standford, CA 94305.-2000.
- [5] Wang Lam and Hector Garcia-Molina, "Multicasting a Web Repository" International Workshop on the Web and Database.
- [6] Gyongyi, Zoltan and Garcia-Molina, Hector, "Link Farm alliance to increase the ranking of certain target Web pages" Technical Report, Standford.
- [7] Baoning Wu, Brain D.Davison, "Identifying Link Farm Spam Pages" publication in proceeding of the 14th International WWW Conferences 2005 International World Wide Wed Conference Committee.
- [8] Ezz Hattab, "Updating Search engine using meta updates" the international arab journal of information technology vol4, no.4, October 2007
- [9] Lawrence & Brin, Sergery & Motwani, Rajiv and Winograd, Terry "The Page Rank Citation Ranking: Bringing order to" Technical Report Stanford Infolab.
- [10] Jeffery Deans & Sanjay Ghemawat, "Map Reduce: data Processing on Large Clusters" Communication of the ACM Simplified 50th anniversary issue: 1958-2008 CACM Homepage Archive Volume 51 issue 1, Jan. 2008, page 107-113.