

The Influence of URL Shortener on Pagerank

M. Zainal Arifin

State University of Malang
Jl. Semarang 5, Malang
Indonesia

Author Email : zainal@um.ac.id

Abstract. Ranking in website of a search engine or what we call as the Pagerank is very important in the search results to display relevant and appropriate to the level of importance. Search results in the search engines always associate with the relevant results on the first page, other than that on page 5 and onwards. Using URL Shortener has now become a trend, especially for those who use a shortener of links that are very long. For example, when an upload document in google doc, url given long enough and need a short address which can be used to share on the web or other site and URL shortener provide answers to these questions. But there are some obstacles in the search engine to determine the content of the page, because the address given by the URL Shortener, unlike Permalink URL used by blogs, that are easy on the search engine crawlers by contemporaries, such as Google, Yahoo or Bing. Analysis on the URL shortener on the ranking examined in this study, and showed that shorten a web page that are using URL Shortener adversely doesn't affect the ranking of search results and outcomes.

Key-Words: *url shortener, pagerank, link analysis, search engine ranking*

1. INTRODUCTION

TinyURL is one of the oldest, most visible and popular on Social-Media websites (Orkut, Facebook, Twitter, Identica, etc.) url shortener around. They really useful from the SEO and web traffic point of view and what are the adverse effects of using a URL shortener to publish links on social media and other such websites on internet.

All the url shortener services are not alike and some of them are less worthy from the search engine optimization point of view. They are all useful without any doubt for micro blogging websites like Twitter and Identica where limited to 140 characters and this shorteners can help extensively. Links are connectivity of the internet and from a link can have an idea about the content.

Shortening urls thing become more popular after the rise of micro blogging services like Twitter and thousands of new sites providing the url shortening services chipped in. There are few who left this bandwagon now and very few survived.

2. LITERATURE REVIEW

in this chapter will be discussed the basis for the conduct of research, starting from the notion of url shortener to discuss the results of the study.

2.1. URL Shortener

URL shortening is a technique on the World Wide Web in which a Uniform Resource Locator (URL) may be made substantially shorter in length and still direct to the required page. This is achieved by using an HTTP Redirect on a domain name that is short, which links to the web page that has a long URL. For example, the URL http://en.wikipedia.org/wiki/URL_shortening can be shortened to <http://bit.ly/urlwiki>, <http://tinyurl.com/urlwiki>, <http://is.gd/urlwiki> or <http://goo.gl/Gmzqv>. This is especially convenient for messaging technologies such as Twitter and Identica which severely limit the number of characters that may be used in a message. Short URLs allow otherwise long web addresses to be referred to in a tweet. In November 2009, the shortened links of the URL shortening service Bitly were accessed 2.1 billion times [1].

Another use of URL shortening is to disguise the underlying address. Although this may be desired for legitimate business or personal reasons, it is open to abuse and for this reason, some URL shortening service providers have found themselves on spam blacklists, because of the use of their redirect services by sites trying to bypass those very same blacklists. Some websites prevent short, redirected URLs from being posted [1].

2.2 Pagerank

The PageRank algorithm for determining the “importance” of Web pages has become a central technique in Web search. The core of the PageRank algorithm involves computing the principal eigenvector of the Markov matrix representing the hyperlink structure of the Web. As the Web graph is very large, containing over a billion nodes, the PageRank vector is generally computed offline, during the preprocessing of the Web crawl, before any queries have been issued.

The development of techniques for computing PageRank efficiently for Web scale graphs is important for a number of reasons.

For Web graphs containing a billion nodes, computing a PageRank vector can take several days. Computing PageRank quickly is necessary to reduce the lag time from when a new crawl is completed to when that crawl can be made available for searching. Furthermore, recent approaches to personalized and topic-sensitive Page-Rank schemes require computing many PageRank vectors, each biased towards certain types of pages. These approaches intensify the need for faster methods for computing PageRank [2].

2.3 Link Analysis

The analysis of hyperlinks and the graph structure of the Web has been instrumental in the development of web search. In this chapter, we focus on the use of hyperlinks for ranking web search results. Such link analysis is one of many factors considered by web search engines in computing a composite score for a web page on any given query. Link analysis of web search has intellectual antecedents in the field of citation analysis, aspects of which overlap with an area known as bibliometrics. These disciplines seek to quantify the influence of scholarly articles by analyzing the pattern of citations amongst them. Much as citations represent the conferral of authority from a scholarly article to others, link analysis on the Web treats hyperlinks from a web page to another as a conferral of authority. Clearly, not every citation or hyperlink implies such authority conferral, for this reason, simply measuring the quality of a web page by the number of *in-links* (citations from other pages) is not robust enough. For instance, one may contrive to set up multiple web pages pointing to a target web page, with the intent of artificially boosting the latter's tally of in-links. This phenomenon is referred to as *link spam*. Nevertheless, the phenomenon of citation is prevalent and dependable enough that it is feasible for web search engines to derive useful signals for ranking of the more sophisticated link analysis. Link analysis also proves to be a useful indicator of what page(s) to crawl next while crawling the web [3].

2.4 URL Redirection

URL redirection, also called URL forwarding, is a World Wide Web technique for making a web page available under more than one URL address. When a web browser attempts to open a URL that has been redirected, a page with a different URL is opened. For example, www.example.com is redirected to www.iana.org/domains/example/. Similarly, Domain redirection or domain forwarding is when all pages in a URL domain are redirected to a different domain, as when wikipedia.com and wikipedia.net are automatically redirected to wikipedia.org. URL redirection can be used for URL shortening, to prevent broken links when web pages are moved, to allow multiple domain names belonging to the same owner to refer to a single web site, to guide navigation into and out of a website, for privacy protection, and for less innocuous purposes such as phishing attacks using URLs that are similar to a targeted web site [4].

2.4.1 HTTP 301

The HTTP response status code 301 Moved Permanently is used for permanent redirection, meaning current links or records using the URL that the 301 Moved Permanently response is received for should be updated to the new URL provided in the Location field of the response. This status code should be used by the location header. RFC 2616 states that:

1. If a client has a link-editing capabilities, it should update all references to the Request URI.
2. The response is cachable.
3. Unless the request method was HEAD, the entity should contain a small hypertext note with a hyperlink to the new URI(s).
4. If the 301 status code is received in response to a request of any type other than GET or HEAD, the client must ask the user before redirecting [5].

3. RESEARCH METHODOLOGY

In this study, the researchers write methodologies activity into multiple steps in the flowchart below:

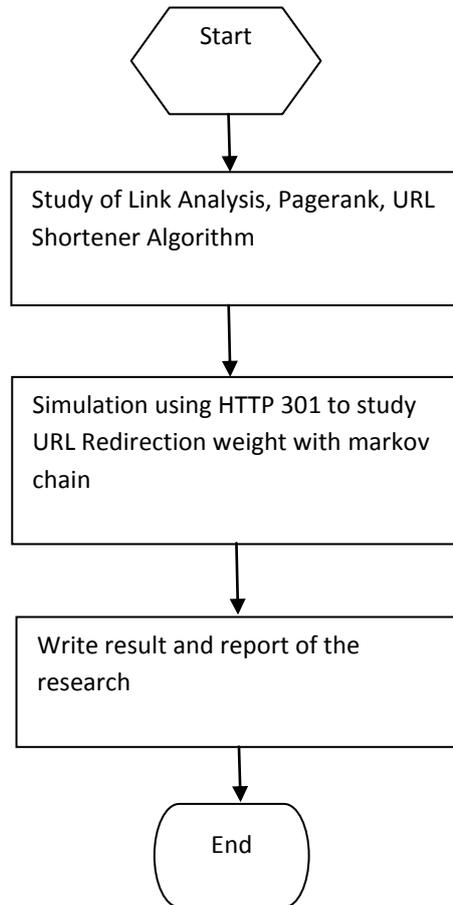


Figure 1. Flowchart of research

The following will explain the steps in this research:

1. The initial stage is the preparation by studying the study of existing literature, then explore topics related to this study. In this study the url shortener on, researchers use base-36 algorithm in the long url address change to a shorter form, using the primary key in the table that stores a long list of urls.
2. Next is to apply HTTP 301 crawler that gives instructions to the address that will be referenced permanently so it will be proven whether there is a significant influence in the calculation of the url shortener ranking in search engines like Google.
3. Write reports and research results.

4. Result and Discussion

The results of this study have provided the desired results, early beginning of the study, created the design for the first URL Shortener to the ranking calculation.

Table 1. Table map of decimal conversion

Decimal	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
Base 36	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F	G	H
Decimal	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35
Base 36	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z

Table 1 describes the basics of the changes to the way the manufacture and design of URL Shortener. While the images shown in the table below, also explained about the characters long url generated by the number recorded in the database, the number of urls shown by the first column, where the integer is the unique id of each record in the database.

Table2. Length of the decimal number in base-36

Decimal	Base 36	Base 36	Decimal	Fraction	Decimal	Base 36
1	1	1	1	1/2	0.5	0J
10	A	10	36	1/3	0.333	0C
100	2S	100	1,296	1/4	0.25	09
1,000	RS	1,000	46,656	1/5	0.2	07
10,000	7PS	10,000	1,679,616	1/6	0.166	06
100,000	255S	100,000	60,466,176	1/7	0.142857	05
1,000,000	LFLS	1,000,000	2,176,782,336	1/8	0.125	04I
1,000,000,000	GJGGS	1,000,000,000	79,364,164,096	1/9	0.111	04
1,000,000,000,000	CRE60RS	1,000,000,000,000	2,821,109,907,456	1/10	0.1	03C

Table 3 is a database design to create a url shortener, some hidden columns that the view we remain focused on what is observed.

Table3. Number of URL count visited

ID	URL	VisitCount
1	www.um.ac.id	23
2	Google.com	231
3	Yahoo.com	52
4	Detik.com	12
.....		
20102	Sctv.co.id	2
20103	Tvri.co.id	1

To do the parsing of the url in order to produce efficient results, then do the following:

```
$expectedURL = trim($_SERVER['URL']);
$split = preg_split(":{:80\\}", $expectedURL);
$shortURL = $split[1];
// security: strip all but alphanumerics & dashes
$shortURL = preg_replace("/[^a-z0-9-]+/i", "", $shortURL);
```

Then further to show the original address of the url shorter, are presented below:

```
$isShortURL = false;
$result = getLongURL($shortURL);
if ($result) { $isShortURL = true; }
$longURL = $result['longURL'];
```

The result of this research show that from the previous code that with HTTP 301, crawler still can find the destination page and information directed to that page.

5. Conclusion

At the end of the study, we will give you exposure to the use of url shortener did not have a major impact on the ranking in the search engines, this is because the site is using a url shortener that transfers HTTP 301 permanent link was visited bot crawler.

References

- [1] Anonymous, 2011. *URL Shortening*. http://en.wikipedia.org/wiki/URL_shortening.
- [2] Anonymous, 2013. *URL Redirection*. http://en.wikipedia.org/wiki/URL_redirection.
- [3] Anonymous, *HTTP 301*, 2013. http://en.wikipedia.org/wiki/HTTP_301, 7 february 2013
- [4] Haveliwala, Kamvar, et all, 2003. *Computing Pagerank Using Power Extrapolation*, Stanford University.
- [5] Manning, Ragavar, 2008. Schutze, *Introduction to Information Retrieval*, Cambridge University Pers.