

Hits Algorithm in SEO

Dr. Marathe Dagadu Mitharam

[Ph.D. MCM (Computer Management)]

ABSTRACT:-

In this paper we proposed the system for HITS in Search Engine Optimization. How captures the webpages from various web servers. HITS in Web crawling it is a program to downloading web pages, download only those web pages whose satisfies some criterion.

In this paper we see how to get webpages when we enter any query in search engine. Here we define one algorithm to specify how HITS work when crawler the web pages.

Goal: - HITS concepts

In this paper describes

- 1) *How get webpages in client machine in rank.*
- 2) *Hubs and Authorities.*
- 3) *To find answer for this question- "How does a search engine know that get all these pages contain the query term?"*

Ans: - Because all of those pages have been crawled and display page ranking using HITS and PAGERANK algorithms.

INTRODUCTION:-

Hyperlink Induced Topic Search (HITS) Algorithm is a Link Analysis Algorithm that rates webpages. This algorithm is used to the web link-structures to discover and rank the webpages relevant for a particular search. A good hub represented a page that pointed to many other pages. A good authority represented a page that was linked by many different hubs.

HITS first enlarges the list of relevant pages returned by a search engine and then produces two rankings of the expanded set of pages, authority ranking and hub ranking.

Web crawler is a relatively simple automated program or scripts that automatically scan or crawl through inter pages. Google is Google's web crawling bot (Sometimes also called a "Spider"). Crawling is the process by which Googlebot discovers new and updated pages to be added to the Google index. Web crawling has so many names just like Robot, Web Agent and Spider etc. Web crawler to looking for new web pages to index, and checking if pages already in its index have been updated or not.

Finally performing PageRank, HITS, Indexer to display the list of url in client machine based on web crawling webpages. The main strength of HITS is its ability to rank pages according to the query topic, which may be able to provide more relevant authority and hub pages. The ranking may also be combined with information retrieval based rankings.

HITS IN SEARCH ENGINE:-

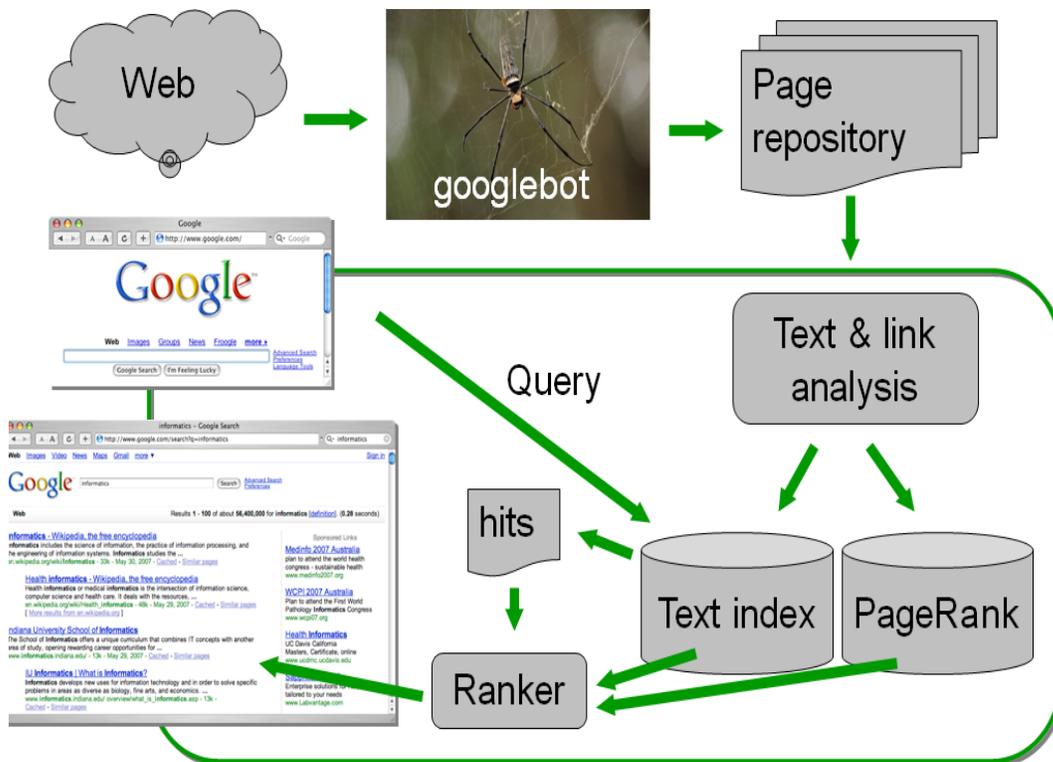
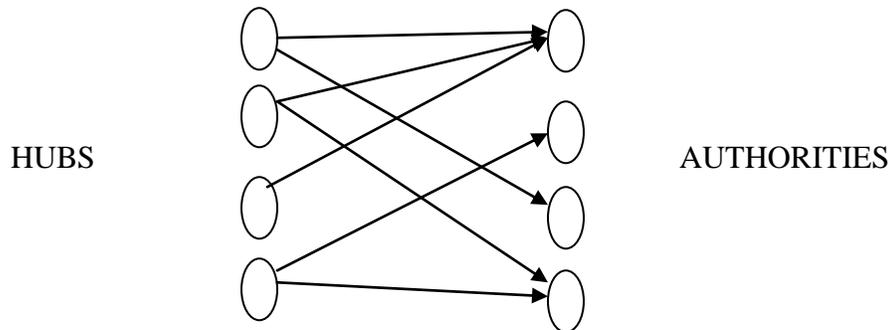


Fig.2 Process Webpage Retrieving.

In the above fig.2 show that whenever we enter query in search engine then it perform various process. Googlebot download fresh content WebPages from web. It transfers to pagerank or indexer. Finally it displays the output in client machine.

HITS stands for **Hypertext Induced Topic Search**. Unlike PageRank which is a static ranking algorithm, HITS is search query dependent. When the user issues a search query, HITS first expands the list of relevant pages returned by a search engine and then produces two rankings of the expanded set of pages, **authority ranking** and **hub ranking**. An **authority** is a page with many in-links.



The idea is that the page may have good or authoritative content on some topic and thus many people trust it and link to it. A **hub** is a page with many out-links. The page serves as an organizer of the information on a particular topic and points to many good authority pages on the topic.

PageRank is a static ranking of Web pages in the sense that a PageRank value is computed for each page off-line and it does not depend on search queries. Since PageRank is based on the measure of prestige in social networks, the PageRank value of each page can be regarded as its prestige. We now derive the PageRank formula. Let us first state some main concepts again in the Web context.

OBJECTIVES:-

- 1) The main objective is to developed HITS algorithm for page ranking.
- 2) To study process of web page retrieving.
- 3) To study the use of HITS and Web Crawler algorithms.

EXPERIMENT:-

In this paper to say how the web crawling is done. Here we developed one algorithm for Page ranking.

Algorithm

Input

A=Query box value/ Search Engine Box

B=Number of webservers.

I=1

Ht=0

Algo.

I to B

1. L= **POP URL** BASED ON QUERY TERM (A)
2. If already visited L page (Content) then Continue loop
Download page in P for L
3. If P Not HTML then
Continue loop
4. If cannot download P then
(Error/ Page cannot open)
5. P is Assign to new links in N
6. N assign to R1 (Page repository or data set)

LOOP Ht to R1

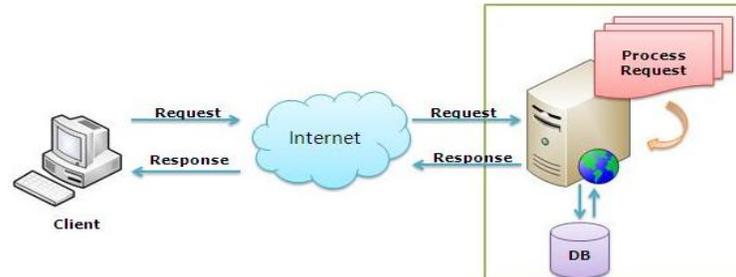
7. Assign every page in R1 an Authority Score and Hub Score
To Find Authority Ranking (H1) from N
8. To Find Hubs Ranking (H1) from N

End LOOP

9. Send to Ranker
10. Append N to the end of B

END

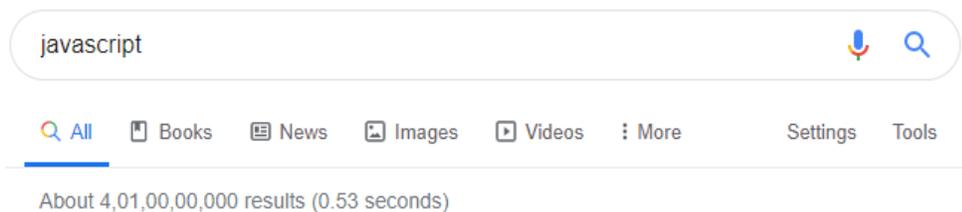
Executing above algorithm obtain ranking webpage from seed URL or new URL using Hits algorithm. Our institute near about 500 computers for educational study. We store data in 3 different data servers. Execute this algorithm on server. Client request to server then the server executes that code and finally display on the client machine. Server transfer this output in HTML Language. In this process we use as Web crawler, PageRank and Indexer etc as other algorithms. Algorithm use php code.



RESULT:-

In this paper we mention the experiment is done in our institute. In this lab 50 computers as client machine and 3 are servers. Whenever enter any query then it gets the output. In output display number of links in Hubs ranking and Authority Ranking In this paper one think is main Hits algorithm. HITS Algorithm is nothing but programs to calculate hubs ranking and authority ranking. This process we done in my college lab.

e.g.



[JavaScript Language Learning | Introductory Course - Free](#)

CONCLUSIONS:-

This paper has attempted to for the purpose of HITS algorithm in SEO. The proposed methods were successfully tested webservers. If we want to rank downloaded WebPages using HITS algorithm then refer this paper. The results which were obtained after the analysis were satisfactory and contained valuable information Hits algorithm.

REFERENCES:-

1. The Impact of Office Automation on the Organization Margrethe H. Olson 25 Issue 11, Nov 1982.
2. Impact of computers in organizations Thomas L. Whisler 1970 Year.
3. Fundamentals of computers, V. Rajaraman, PHI publication.
4. <https://archive.org/details/1.EngImpactOfComputerMARATHE>.
5. https://www.researchgate.net/journal/0166-3615_Computers_in_Industry.
6. Balaguruswamy, E. (2009). Fundamentals of Computers. Tata McGraw Hill.
7. Kothari, C. (2004). Research Methodology- methods and techniques. new age international publisher new Delhi.
8. Liu, B. (n.d.). Web Mining. Springer.
9. Summers, D. W. (n.d.). Impact of ICT Innovations on Business .
10. Web data mining – Bing Liu
11. PPT for Web usage mining- Bing Liu
12. Srivastava, J., Cooley, R., Deshpande, M., Tan, P.N.(2000). Web Usage Mining: Discovery and Applications of Usage Patterns from Web Data. ACMSIGKDD, Jan 2000.
13. Jaideep Srivastava Paper .