

CHANGING VALUE OF STORED INFORMATION FOR SEARCH ENGINE

Ajeet Mishra

e-mail: sidhiopen@gmail.com

Assistant Professor [as Guest Faculty]

Awadhesh Pratap Singh University, Rewa (M.P.)

1. Abstraction:

There are so many search engines that are used by internet user to retrieve and seeking websites with special keyword. There are so many record in world wide web that related to one keyword that entered by user then how google.com, yahoo.com, bing.com decide to show which site got top in list of result during key-word searching. In this document we will work to explain a way to change value of information for search engine, after that we can see good result of searching because there are not good searching in present time by any search engine because it targeted certain criteria particularly targeted traffic, hit and view. This criterion is achieved by any website by using simple tricks and by using third party that give insurance to website for top searching taking fees for SEO.

2. Introduction

Today internet is most popular place which is used for business, entertainment, education and e-governing. Daily more than 100 website are registered. About 350000 website has been registered between January 2013 to 2014. There is a list of registered website with years which is given below.

Year(June)	Websites	Change
2014	968,882,453	44%
2013	672,985,183	-3%
2012	697,089,489	101%
2011	346,004,403	67%
2010	206,956,723	-13%
2009	238,027,855	38%
2008	172,338,726	41%
2007	121,892,559	43%

2006	85,507,314	32%
2005	64,780,617	26%
2004	51,611,646	26%
2003	40,912,332	6%
2002	38,760,373	32%
2001	29,254,370	71%
2000	17,087,182	438%
1999	3,177,453	32%
1998	2,410,067	116%
1997	1,117,255	334%
1996	257,601	996%
1995	23,500	758%
1994	2,738	2006%
1993	130	1200%
1992	10	900%
Aug. 1991	1	

As you can see from above table number of website are registered every day, each website is for individual use but people cannot remember so many website's address in their mind and there are so many website that's name and use, user don't know. At this problem, user use search engine and search website with keyword. For example I am preparing competition exams for banks job and suddenly I hear from any one IBPS [Institute of Banking Personnel Selection is an autonomous agency in India] conduct an exam and form portal is open. I use Google and type IBPS form and hit enter and result displayed in my monitor. Then I see that my required website is listed in 7th or 8th position while from all over India candidates are trying to access this website and position of this website should be top in result page. And many search engines give preference to that website which are old and get more traffic. Although new website that come in light for internet should get importance because they are new and they will collect good and updated information.

Therefore by this article/paper we describe here a method to help improving search techniques and give value to stored data that come from web crawling and saved in database.

3. Collecting Data in tables:

First we should have our database that will contain tables to hold data. These data is important because this data is help to searching any website and there are some table that hold value of

information means how any data is searching by user and which website is clicked by user after showing result.

3.1 Table first: Data_of_web

Here “Data_of_web” is our table name that store information about web pages like url, information, keyword etc.

- 1- **Url** : this field is used to hold address of web page .
- 2- **Content**: this field is used to hold content of web page that are check by keyword that entered in search engine text box by user.
- 3- **Content_type** : This field hold value for recognized that which type data contained by webpage.
- 4- **Hit_rate** : this field of table contain number that tell, how many time this page accessed by searching process.
- 5- **Purpose_of_page** : this hold information of webpage like purpose of page.
- 6- **Entry_number** : this is special unique number to target page that address is stored in url Field. And this is primary key for this table.
- 7- **Created_date** : this hold date when it crawled and stored in this table.

Url	Content	Content_type	Hit_rate	Purpose_of_page	Entry_number	Created_date

3.2 Second Table : Statiscs_of_webpage

This table is used to collect value of information this table hold a value that tell to importance of data. By this table our program decides that which web site gets top preference to show result of any search according any keyword by user.

- 1- **Entry_number** : This number is foreign key for table first which name is “data_of_web” and this field hold value that target specifics web address according to “data_of_web”.
- 2- **Keyword** : this field hold data that was searched by user. And this field is maintain only specific key word.
- 3- **Number_of_hit** : this column is used hold value of number, how many time open web page during searching of keyword that are stored in keyword field.

- 4- **Country_entry_num** : this is primary key of this table that is unique value of this table and used as foreign key in third table which name is “country_index”.

3.3 Third Table : Country_table

This table will be used to analysis that which keyword is most searched by which country. Search engine check the nation status of searcher and give result according to country location of user.

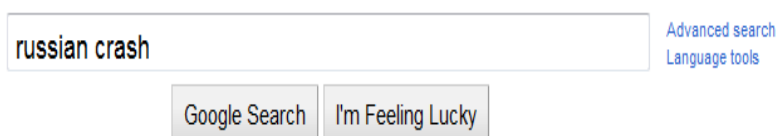
- 1- **Country_name**: this column store country name that will help to identify which keyword mostly searched from which country.
- 2- **Hit_rank**: This column is used to store value that point that how many time this page was visited according both keyword and country.
- 3- **Country_entry_num**: It will hold special value in form of number.

4. Problem by other Search engine and Solution hints:

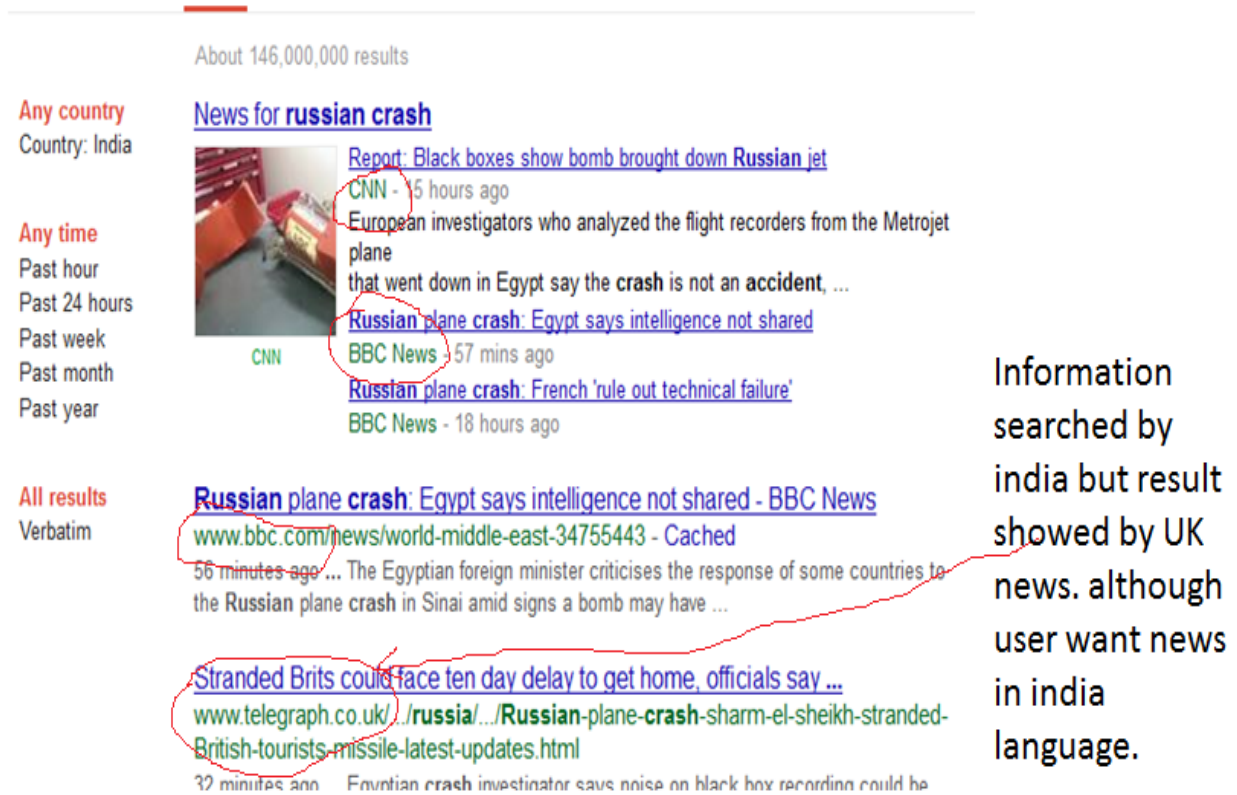
In search engine there are a big problem that if there is new website that recently launched they not get preference according keyword only those website got preference that are old and hit rate is good.

Google give preference youtube, blogspot more as compare to other website that is not fine if any new article launched with new website and keyword match with that website search engine should preference new websites.

Here I want show an example, I am Indian and recently an accident occurred Russian aircraft crashed, but I don't know detail so I did used google and search “Russian crash” like below picture.



Now user hit enter to get result of entered keyword “Russian crash”. But result that will display in below page it has got preference by other language and geographic while user are from India and he want know information about latest news and accident of Russian airplane crash. This result show that search engine should have knowledge that type of information source should provide to user, in this case search engine provide result according number of hit for specific website behalf of news.

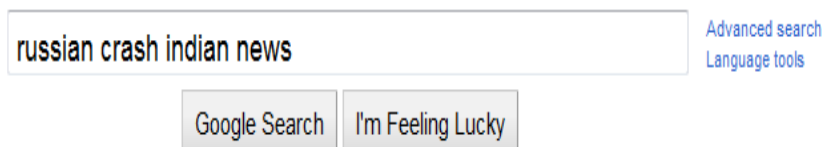


The screenshot shows a Google search interface with the following elements:

- Search results count: "About 146,000,000 results"
- Filters on the left:
 - Any country**: Country: India
 - Any time**: Past hour, Past 24 hours, Past week, Past month, Past year
 - All results**: Verbatim
- Search results:
 - Section: **News for russian crash**
 - Image: A small image of a cockpit interior with the text "CNN" below it.
 - Result 1: [Report: Black boxes show bomb brought down Russian jet](#) - CNN - 5 hours ago. Snippet: "European investigators who analyzed the flight recorders from the Metrojet plane that went down in Egypt say the crash is not an accident, ..." (circled in red).
 - Result 2: [Russian plane crash: Egypt says intelligence not shared](#) - BBC News - 57 mins ago. Snippet: "Russian plane crash: French 'rule out technical failure'" (circled in red).
 - Result 3: [Russian plane crash: French 'rule out technical failure'](#) - BBC News - 18 hours ago.
 - Result 4: [Russian plane crash: Egypt says intelligence not shared - BBC News](#) - www.bbc.com/news/world-middle-east-34755443 - Cached. Snippet: "56 minutes ago ... The Egyptian foreign minister criticises the response of some countries to the Russian plane crash in Sinai amid signs a bomb may have ..." (circled in red).
 - Result 5: [Stranded Brits could face ten day delay to get home, officials say ...](#) - www.telegraph.co.uk/.../russia/.../Russian-plane-crash-sharm-el-sheikh-stranded-British-tourists-missile-latest-updates.html. Snippet: "37 minutes ago Egyptian crash investigator says noise on black box recording could be ..." (circled in red).

Information searched by india but result showed by UK news. although user want news in india language.

In above picture you can see clearly that result showing, a result of some website that are not Indian website specially Hindi news website while I am seeking Hindi website. Then I try once again search engine text box to find Indian news website and changed search keyword like given picture.



The screenshot shows a search box containing the text "russian crash indian news". To the right of the search box are links for "Advanced search" and "Language tools". Below the search box are two buttons: "Google Search" and "I'm Feeling Lucky".

In above figure we can see that there is textbox that are in search page now we are searching “Russia crash Indian news” then we see below figure that give the result of given keyword, and user get Indian news website while search engine should be such as that if user search an key word without pointing country wise news website then search engine get user location by network ip and give information according nation wise.

Any country

Country: India

Any time

- Past hour
- Past 24 hours
- Past week
- Past month
- Past year

All results

Verbatim

News for russian crash indian news



Times of India

[Black boxes of Russian plane that crashed in Egypt point to 'bomb attack'](#)

Times of India - 23 hours ago

READ ALSO: **Russian plane crash**: What will Putin do if ISIS bomb brought ... And

The Times newspaper reported on Friday that electronic ...

[US adds air security 'enhancements' amid Russian plane crash probe](#)

The Indian Express - 15 hours ago

[Russia awaits concrete evidence on plane crash](#)

Zee News - 10 hours ago

Crashed Russian plane may have been brought ... - Times of India

timesofindia.indiatimes.com/.../Crashed-Russian.../49664675.cms - Cached

2 days ago ... Britain said on Wednesday that the Russian plane that crashed in Egypt after taking ... The Times of India ... News Home » World » Middle East ...

Russia, Egypt dismiss suggestions a bomb caused crash - Times of ...

timesofindia.indiatimes.com/world/.../Russia...crash/.../49681667.cms - Cached

1 day ago ... The Times of India Russia and Egypt insist the investigation into the crash must run its course before any conclusion is reached. ... to suspend flights to Sharm el-Sheikh, the Egyptian state-run news agency MENA reported.

Russian plane crashes in Egypt, about 150 bodies ... - India Today

indiatoday.intoday.in/story/russian-jet...crashes-in.../511454.html - Cached

31 Oct 2015 ... Russian plane crash in Egypt: About 150 bodies, including those of five ... For more news from India Today, follow us on Twitter @indiatoday ...

5. Set value of Information to get Result from search engine according user:

First we should focus that on keyword which keyword tried to search by user and compare it form our data base.

For example: any user search “narendra modi”

*Select * from Data_of_web where content like “%narendra modi%”*

After implementing this query we get result of entity that contain “narendra modi” now from here this is very simple but trick and planning needed from here to ahead. Now we have number of rows as a output. And which website get preference how it will decided.

- 1- First we check field “hit_rate” of first table where that tell that how many time this webpage hit by this content.
- 2- Now second thing to necessary to know that “narendra modi” is politician of India now what type of content should be education, news, or entertainment? Politician actually related to news therefore give preference to those rows that select by above query that content type is “politics” and “news”.
- 3- Check from which country, user search keyword according ip address and now check the result that got by above query, that which content belong to user’s location and represent it in search engine.
- 4- There are content in table that upload in website a year ago and a content uploaded recently therefore search engine should provide information in two type first popular clicked searched information and second should according latest uploaded in new website.

6. Conclusion:

Search engine behavior that used by other search engine website give many time unexpected result that not wanted by user, and many time user search e keyword that was based on recently incident but search engine provide result that based hit rate of keyword while a solution that given in this article is thinkable that give variety of conditions to search engine to check keyword from available data from database according to country location of user, data type means relation of keyword for example “sachin tendulkar” is cricketer and its relation will be news, education and entertainment then sql query should be arrange in this form that they provide information that satisfy by user needs. Idea share by this article solve this problem, after all there should be two type of value of information that stored in search engine database one is high and second is low where high means information entered recently and watched by many times while low value of information set that this value is old. Now what benefit we gain after managing two type values, first people need current data therefore by this value search engine provide data front of user by preference of value of information and high value should be first and then other value got preference.

7. References:

[1] Abhishek Das,*Google Inc., USA* And Ankit Jain,*Google Inc., USA* Indexing The World Wide Web: The Journey So Far

- [2] Alpert, J., and Hajaj, N. (2008). *We knew the web was big...*, Google Blog, Retrieved on 10/13/2010 from <http://googleblog.blogspot.com/2008/07/we-knew-web-was-big.html>
- [3] Asanovic, K., Bodik, R., Demmel, J., Keaveny, T., Keutzer, K., Kubiawicz, J., Lee, E., Morgan, N., Neula, G., Patterson, D., Sen, K., Wawrzynek, J., Wessel, D. & Yelick, K. (2008). *The Parallel Computing Laboratory at U.C. Berkeley: A Research Agenda Based on the Berkeley View*. Technical Report, UC Berkeley
- [4] Brin, S., Page, L. (1998). *The Anatomy of a Large-Scale Hypertextual Web Search Engine*. Proceedings of the 7th international conference on World Wide Web (WWW). Brisbane, Australia. pp. 107–117. 1998
- [5] Kesmodel, D. (2005-09-22). *Sites Get Dropped by Search Engines After Trying to 'Optimize' Rankings*. Wall Street Journal, Retrieved on 10/13/2010 from <http://online.wsj.com/article/SB112714166978744925.html>
- [6] Manning, C., Raghavan, P., and Schütze, H. (2008) *Introduction to Information Retrieval*. Cambridge University Press.
- [7] Patterson, A.L. (2004), *Phrase-based searching in an information retrieval system*. US Patent Application 20060031195: Published February 9, 2006. Filed July 26, 2004
- [8] http://en.wikipedia.org/wiki/Web_search_engine
- [9] Christopher D. Manning, Prabhakar Raghavan, and Hinrich Schütze (2007), *Introduction to Information Retrieval*.
- [10] S. E. Robertson and K. S. Jones. Relevance Weighting of Search Terms. *Journal of the American Society for Information Sciences*, 27(3): pages 129–146, 1976.
- [11] G. Salton. *The Smart Retrieval System: Experiments in Automatic Document Processing*. Prentice - Hall, Englewood Cliffs, NJ, 1971.
- [12] <http://www.pewinternet.org/Reports/2012/Search-Engine-Use-2012/Summary-of-findings.aspx>
- [13] T. Hofmann. Collaborative Filtering via Gaussian Probabilistic Latent Semantic Analysis. In Proc. of the 26th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '03), pages 259– 266, 2003.
- [14] G. Jeh and J. Widom. Scaling Personalized Web Search. In Proc. of the 12th International World Wide Web Conference (WWW 2003), pages 271–279, 2003.
- [15] Salton, G., Wong, A., And Wang, C. S (1975). *A vector space model for automatic indexing*. *Comm. ACM* 18, 11 (Nov.) 613–620.
- [16] <http://infolab.stanford.edu/~backrub/google.html>

[17] <http://blog.woorank.com/2014/07/mp-explain-me-how-a-crawler-works-like-im-five/>

[18] Web Crawling and Data Mining with Apache Nutch Paperback – 1 Jul 2014 Author by Dr. Zakir Laliwala

[19] <https://www.safaribooksonline.com/library/view/the-art-of/9780596809133/ch02.html>