

Behavior of Search Engines in Popular Queries

Siddharth Ghansela* and Ashish Negi

Department of Computer Science and Application, GB Pant Engineering College, Garhwal - 246194, Uttarakhand, India; syahoocom@gmail.com, ashish.ne@gmail.com

Abstract

Objectives: Presently, various search engines are available in the web with huge database. Not only the available search engine but the query also plays important role for getting appropriate results from the search engines. Our objective is to show the importance of popular queries. **Methods/Statistical Analysis:** In this article, we have introduced two new categories of query the one is popular query and another one is non-popular query. We analyse the behaviour of search engines using popular queries in top three search engines and after that compared them with a traditional mathematical model for rank calculation along with user feedback method. **Findings:** By proposing new category of query we analyse how the behaviour of search engines changed. Here we are using three methods for calculating ranking in different types of search engines to give more strength to our results. Our findings are to show the importance of popular queries in different types of search engines. **Application/Improvements:** From this article, we conclude that the behaviour or search engine in popular query is different than a simple query; some of the search engine gives them more importance because of their popularities.

Keywords: Behaviour, Popular, Query, Rank, Search Engine

1. Introduction

The network has become crucial facet in the growth of many people, and search engines are the main gateway to the Web. Search engines are main apparatus for gaining the material, browsing sites, and services on the Web that many people use on a daily basis. Most common way used in search engine primarily focus on similarity of query and a page, as well as the overall page quality for ranking¹. From the past 15 years search engines plays significant act in knowledge retrieval. The first seek-

ing tool Archie was built by². Subsequently Gopher was popularized by³. A net toddler was introduced by⁴. One more search engine, Ali web further comes in 1993⁵. In 1998-2001, the Google search engine was developed by⁶. In 2004, Yahoo launched its own search engine. In 2005, MSN by Microsoft launch its search engine. In 2009 Bing was developed by Microsoft team⁷. When we want to search a document on the web not only search engine but query also plays important role for finding appropriate document from the existing database.

Table 1. Most searched query log

Most Searched								
S. No.	Trending	People	Movies	Sportsperson	Mobile Devices	Bollywood Actor (Male)	Bollywood Actor (Female)	Transaction Sites
1.	ICC Cricket World Cup 2015	Sunny Leone	Bahubali	Virat Kohli	YU Yureka	Salman Khan	Sunny Leone	Flipkart
2.	Bahubali	Salman Khan	Bajrangi Bhaijaan	Lionel Messi	Apple iPhone 6S	Shah Rukh Khan	Katrina Kaif	IRCTC

*Author for correspondence

3.	Bajrangi Bhaijaan	APJ Abdul Kalam	PremRatan DhanPayo	Sachin Tendulkar	Lenovo K3 Note	Akshay Kumar	Deepika Padukone	SBI - State Bank of India
4.	Premratan DhanPayo	Katrina Kaif	ABCD 2	M S Dhoni	Lenovo A7000	Shahid Kapoor	Alia Bhatt	Amazon
5.	Indian Premier League (IPL)	Deepika Padukone	I	Cristiano Ronaldo	Moto G	Hrithik Roshan	RadhikaApte	Snapdeal

2. Query

A search engine query is an appeal for data that is made using a search engine. The term query is to denote a word or collection or words or phrase. In this paper we are going to introduced two categories of query the first one is popular and another one is non-popular query. A popular query is one that is most visited in a particular duration (days or months). A non-popular query is a common type of query. The most searched queries are given⁸ in Table 1.

3. Traditional Mathematical Model for Ranking a Document

In this paper we used the vector space model⁹ as a traditional mathematical model for rank a document because it permits computing a regular degree of similarity between queries and documents, also it is easy to implement. Vector space prototype or term angle layout is an algebraic model for representing content documents. It is used in data retrieval, indexing and evaluation of documents.

Table 2. Weights of documents based on Google results

Terms	Term in Q	Count tf_i			df	$\frac{D}{df_i}$	$\log(\frac{D}{df_i})$	Weights, $W_i = tf_i \times IDF_i$			
		D1	D2	D3				Q	D1	D2	D3
ICC	1	1	1	1	3	1.0	0.0000	0.0000	0.0000	0.0000	0.0000
Cricket	1	1	1	1	3	1.0	0.0000	0.0000	0.0000	0.0000	0.0000
World	1	1	0	1	2	1.5	0.5849	0.5849	0.5849	0.0000	0.5849
Cup	1	1	0	1	2	1.5	0.5849	0.5849	0.5849	0.0000	0.5849
2015	1	1	0	1	2	1.5	0.0000	0.0000	0.0000	0.0000	0.0000
Live	0	0	1	0	1	3.0	1.5849	0.0000	0.0000	1.5849	0.0000
Scores	0	0	1	0	1	3.0	1.5849	0.0000	0.0000	1.5849	0.0000
News	0	0	1	0	1	3.0	1.5849	0.0000	0.0000	1.5849	0.0000
And	0	0	1	0	1	3.0	1.5849	0.0000	0.0000	1.5849	0.0000

4. User Feedback Session

User feedback session method is based on the query log. Many previous works has been investigated on problem of analysing user query logs¹⁰⁻¹⁴. Click division is a feature recommended by¹⁵. Generally, a period for web exploration is an array of subsequent queries to appease a single information need and some clicked search results. The proposed feedback session is based on clicked URLs. The single period includes all the three URLs. Each feedback session can tell what a user wants and what he/she does not requires. Therefore, for inferring user search goals, it is more efficient to analyse the feedback sessions than to analyse the search results or clicked URLs directly.

5. Experimental Results The popular query selected is as follows:

Q: ICC cricket world cup 2015 and the results of top three search engines are given as follows:

Google

D1: Cricket World Cup 2015-ICC Cricket.

D2: Live Cricket Scores and News-ICC.

D3: Results Cricket World Cup 2015 - ICC Cricket.

And the weights of documents based on Google results are given in Table 2.

Bing

D1: 2015 Cricket World Cup.

D2: News about ICC Cricket World Cup 2015.

D3: Videos of ICC cricket world cup 2015.

And the weights of documents based on Bing results are given in Table 3.

Yahoo

D1: ICC Cricket World Cup 2015 News.

D2: 2015 Cricket World Cup.

D3: ICC Cricket World Cup 2015 - ICC.

And the weights of documents based on Yahoo results are given in Table 4.

Table 3. Weights of documents based on Bing results

Terms	Term in Q	Count tf_i			df_i	$\frac{D}{df_i}$	$\log\left(\frac{D}{df_i}\right)$	IDF_i	Weights, $W_i = tf_i \times IDF_i$			
		D1	D2	D3					Q	D1	D2	D3
ICC	1	0	1	1	2	1.5	0.5849	0.5849	0.0000	0.5849	0.5849	
Cricket	1	1	1	1	3	1.0	0.0000	0.0000	0.0000	0.0000	0.0000	
World	1	1	1	1	3	1.0	0.0000	0.0000	0.0000	0.0000	0.0000	
Cup	1	1	1	1	3	1.0	0.0000	0.0000	0.0000	0.0000	0.0000	
2015	1	1	1	1	3	1.0	0.0000	0.0000	0.0000	0.0000	0.0000	
News	0	0	1	0	1	3.0	1.5849	0.0000	0.0000	1.5849	0.0000	
About	0	0	1	0	1	3.0	1.5849	0.0000	0.0000	1.5849	0.0000	
Videos	0	0	0	1	1	3.0	1.5849	0.0000	0.0000	0.0000	1.5849	

Table 4. Weights of documents based on Yahoo results

Terms	Term in Q	Count tf_i			df_i	$\frac{D}{df_i}$	$\log\left(\frac{D}{df_i}\right)$	IDF_i	Weights, $W_i = tf_i \times IDF_i$			
		D1	D2	D3					Q	D1	D2	D3
ICC	1	1	0	1	2	1.5	0.5849	0.5849	0.5849	0.0000	0.5849	
Cricket	1	1	1	1	3	1.0	0.0000	0.0000	0.0000	0.0000	0.0000	
World	1	0	1	1	2	1.5	0.5849	0.5849	0.0000	0.5849	0.5849	
Cup	1	0	1	1	2	1.5	0.5849	0.5849	0.0000	0.5849	0.5849	
2015	1	1	1	1	3	1.0	0.0000	0.0000	0.0000	0.0000	0.0000	
News	0	1	0	0	1	3.0	1.5849	0.0000	0.0000	0.0000	0.0000	

6. Similarity Analysis of Search Engines

The similarity function is

$$\text{Cos } Q.D_1 = \frac{Q.D_1}{|Q| \times |D_1|}$$

So from this formula the rank of all the three documents are given as follows:

Google:

$$\text{Cos } Q.D_1 = 8.5525$$

$$\text{Cos } Q.D_2 = 0.0000$$

$$\text{Cos } Q.D_3 = 8.5525$$

Bing:

$$\begin{aligned} \text{Cos}Q.D1 &= 0.0000 \\ \text{Cos}Q.D2 &= 1.9737 \\ \text{Cos}Q.D3 &= 2.4180 \end{aligned}$$

Yahoo:

$$\begin{aligned} \text{Cos}Q.D1 &= 1.0676 \\ \text{Cos}Q.D2 &= 1.5098 \\ \text{Cos}Q.D3 &= 0.9367 \end{aligned}$$

7. Comparison of Methods

After comparing the three methods based on the above results rank comparison of the three documents are shown in Table 5.

Table 5. Rank of documents based on three engines

Document	Rank Based On VSM		
	Google	Bing	Yahoo
D1	1	3	2
D2	3	2	1
D3	2	1	3

Table 6. User feedback sessions

User ID	Query	Clicked URL	Marks Given By User out of 10	Search Engine	Time Spend
00A	ICC cricket world cup 2015	Cricket World Cup 2015-ICC Cricket	8	Yahoo	1.39 min
00B	ICC cricket world cup 2015	Live Cricket Scores & News-ICC	6	Google	0.40 min
00C	ICC cricket world cup 2015	Results Cricket World Cup 2015 - ICC Cricket	9	Bing	1.53 min
00D	ICC cricket world cup 2015	Cricket World Cup 2015-ICC Cricket	10	Google	2.06 min
00E	ICC cricket world cup 2015	Cricket World Cup 2015-ICC Cricket	7	Yahoo	1.42 min
00F	ICC cricket world cup 2015	Live Cricket Scores & News-ICC	9	Google	2.19 min
00G	ICC cricket world cup 2015	Cricket World Cup 2015-ICC Cricket	7	Bing	0.59 min
00H	ICC cricket world cup 2015	Live Cricket Scores & News-ICC	10	Bing	3.29 min
00I	ICC cricket world cup 2015	Live Cricket Scores & News-ICC	7	Yahoo	2.29 min

8. User Feedback Sessions

Here we used user feedback session which was based on browsing the web content of the given query. In our experimental result we have taken ten academic students give them query with pen and paper. After some time the user give us a feedback about the web content shown in Table 6 and the result of user feedback sessions are given in Table 7 also the comparative chart of user feedback sessions is given in Figure 1.

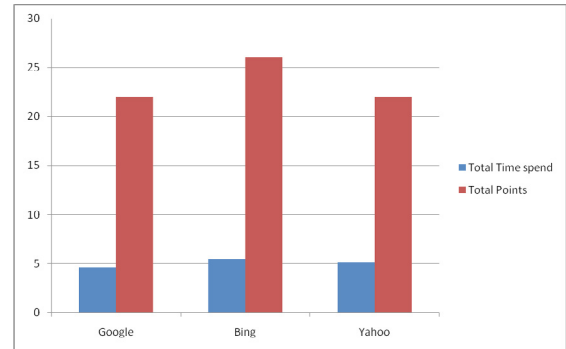


Figure 1. Comparison of user feedback sessions.

9. Conclusion and Future Work

In this paper we have taken one popular query based on the query log and the top most three search engines. After

Table 7. User feedback sessions results

Search Engines	Total Time spend	Total Points
Google	4.6	22
Bing	5.4	26
Yahoo	5.1	22

entering the query in all three search engines we have taken top most three results. From the shown table we are able to understand that the frequency of all the three top most search engines does not match. But at the same time when we started a user feedback session with nine users and three search engines with same numbers of users divided among three search engines we found that Bing got more points comparisons of Google and yahoo also people spent more time in Bing.

10. References

1. Rezaei HR, Dehkordi MN, Moghadam RA. Improving performance of search engines based on fuzzy classification. *Indian Journal of Science and Technology*. 2012 Nov; 5(11):1–5.
2. Leiden U. Archie, Internet History - Search Engines (from Search Engine Watch). Universite it Leiden, Netherlands, web: Leiden U-Archive; 2001.
3. Randall J, Neil N. *The World Wide Web unleashed*. Sams Publishing; 1994. p. 20.
4. Gray M. Internet growth and statistics: Credits and background. Available from: <http://www.mit.edu/people/mkgray/net/background.html>
5. Gray M. Internet growth and statistics: Credits and background. Available from: <http://www.mit.edu/people/mkgray/net/background.html>
6. Senthil Kumar NK, Kumar KK, Rajkumar N, Amsavalli K. Search engine optimization by fuzzy classification and prediction. *Indian Journal of Science and Technology*. 2016 Jan; 9(2):1–5.
7. Microsoft's new search at bing.com helps people make better decisions: Decision engine goes beyond search to help customers deal with information overload. Microsoft. Available from: www.techmeme.com/090528/p55
8. Hacking the Google Trends API. Available from: <http://www.google.co.in/trends/topcharts>
9. Shalton G, Wong A, Yang CS. A vector space Model for automatic indexing. *Communications of the ACM*. 1975; 18(11): 613–20.
10. Joachims T. Optimizing search engines using clickthrough data. *Proc 8th ACM SIGKDD Int'l Conf Knowledge Discovery and Data Mining (SIGKDD)*; 2002. p. 133–42.
11. Beeferman D, Berger A. Agglomerative clustering of a search engine query log. *Proc 6th ACM SIGKDD Int'l Conf Knowledge Discovery and Data Mining (SIGKDD '00)*; 2000. p. 407–16.
12. Wen JR, Nie JY, Zhang HJ. Clustering user queries of search engine. *Proc 10th Int'l Conf World Wide Web (WWW '01)*; 2001. p. 162–8.
13. Shen D, Sun J, Yang Q, Chen Z. Building bridges for web query classification. *Proc 29th Ann Int'l ACM SIGIR Conf Research and Development in Information Retrieval (SIGIR '06)*; 2006. p. 131–8.
14. Baeza-Yates B, Hurtado C, Mendoza M. Query recommendation using query logs in search engines. *Proc Int'l Conf Current Trends in Database Technology (EDBT '04)*; 2004. p. 588–96.
15. Lee U, Liu Z, Cho J. Automatic identification of user goals in web search. *14th WWW Conference*; 2005. p. 1–10.