RESEARCH ARTICLE

OPEN ACCESS

Relevance of Search Engines for Modern Generations

Trilok Gupta¹, Archana Sharma²

¹Pacific Academy of Higher Education and Research University, Udaipur, Rajasthan-India, ²Associate Professor, Department of Computer Science, Gurukul Institute of Engineering & Techchnology Institutional Area, Ranpur, Kota, Rajasthan-India

Abstract

Web search engines have major impact in people's everyday life. It is of great importance to test the retrieval effectiveness of search engines. However, it is labour-intensive to judge the relevance of search results for a large number of queries, and these relevance judgments may not be reusable since the Web data change all the time. Experiments on major search engines show that our approach mines many high-confidence rules that help understand search engines and detect suspicious search results. *Keywords* – Word Wide Web, Information Retrieval, Indexing

I. INTRODUCTION

The World Wide Web (Web) was invented in 1989 and the World Wide Web Consortium (W3C) was established in 1994 to lead the World Wide Web to its full potential. By the turn of the century the Web had entered most aspects of our lives from communication to e-Government, e-Commerce and e-Learning, making it much more than just an information repository.

"The World-Wide Web (W3) was developed to be a pool of human knowledge, and human culture, which would allow collaborators in remote sites to share their ideas and all aspects of a common project."

Sir Tim Berners-Lee invented the World Wide Web in 1989 while working as a software engineer at CERN, the large particle physics laboratory near Geneva , Switzerland. With many scientists participating in experiments at CERN and returning to their laboratories around the world, these scientists were eager to exchange data and results but had difficulties doing so. Tim understood this need, and understood the unrealized potential of millions of computers connected together through the Internet. Tim Berners-Lee, inventor of the Web and Director of the World Wide Web Consortium (W3C), is regularly cited for saying "The power of the Web is in its universality. Access by everyone regardless of disability is an essential aspect" and more recently "One Web for anyone, everywhere on anything" this is all part of the Web's 'full potential'. In 1999 the W3C Web Accessibility Initiative (WAI) published the first set of international guidelines for Web accessibility, the Web Content Accessibility Guidelines 1.0 (WCAG), documenting the essential requirements for Web content to be accessible to people with disabilities. Accessibility requirements for authoring tools (ATAG) and user agents

(UAAG), including browsers followed. At the time of writing (2008), the W3C had advanced drafts available of WCAG 2.0 and ATAG 2.0 along with a specification for Accessible Rich Internet Applications (WAI-ARIA) that will assist scripted Internet applications to become accessible. Traditional Web search engines mostly adopt a keyword based approach. When the keyword submitted by the user is ambiguous, search result usually consists of documents related to various meanings of the keyword, while the user is probably interested in only one of them. With the advent of the World Wide Web (Web), a new category of searching now presents itself. The Web has had a major impact on society (Lesk, 1997; Lynch, 1997) and comes the closest in terms of capabilities to realizing the goal of the Memex (Bush, 1945). In terms of quality, Zumalt and Pasicznyuk (1998) show that the utility of the Web may now match that of the skills a professional reference librarian. The Web possesses an ever- changing and extremely heterogeneous document collection of immense proportions. . Although developed in an apparently unstructured environment, Web document discovery is extremely structured in terms of its hyperlinks. The user population of the Web is enormous and extremely diverse, albeit with certain groups over represented (Hoffman, Kalsbeek, & Novak, 1996; NTIA, 1999).



Figure 1: Process of Information searching through search Search Engine

The Web's IR systems are also unique in terms of the interface, advertising constrains, bandwidth restrictions, and unique document indexing issues (e.g., spamming and URL hijacking). In sum, the Web appears to be a whole new searching environment (Sparck-Jones & Willett, 1997).

Studies can be viewed as a subset within the larger area of IR system evaluation, which typically focuses on measuring the recall and precision of the system (Sparck-Jones, 1981). The theoretical underpinnings for this type of IR evaluation are well defined (Salton & McGill, 1983), although the proper metrics are still a topic of debate (Saracevic, 1995). In this type of evaluation, one takes a known document collection with documents classified as relevant or non-relevant based on a set of queries. These queries are executed using a particular IR system against the document collection. Based on the number of relevant and non-relevant documents retrieved, one determines recall and precision. This is a systems view of relevance, with recall and precision directly related to the queries entered. The whole process is very systematic.

However, once a 'real' searcher is interjected into the system, the evaluation metrics are no longer so straightforward. Relevance to a searcher is not clearly defined (Mizzaro, 1997; Saracevic, 1975; Spink, Greisdorf, & Bateman, 1998). In fact, it is not even certain how a searcher conducts the search process, although there are several theories on the information seeking process (Belkin, Oddy, & Brooks 1982; Saracevic, 1996) that attempt to explain it. Most of these theories are based on empirical analyses of users and, in many cases, the studies do not agree with one another about usersearching processes.

II. LITERATURE REVIEW

The literature review is divided into two sections. **Section one**, Primary Web-searching Studies, is a review of all Web searching studies that deal with studies of searching using Web search engines. These studies contained a substantial amount of data and addressed a broad range of Web searching characteristics. **Section two**, Secondary Web-searching Studies, is a review of Web searching studies that are more limited in scope in that they do not present enough data to give a full picture of Web searching. Most of these studies analyzed Web searching on a singular Web site that was not a search engine.

2.1 PRIMARY WEB-SEARCHING

Once read that the average person living in a modern industrialized society is exposed to as many different pieces of information in a single day as a person living 100 years ago would have seen in a year. That includes advertisements, newspaper headlines, websites, text messages, traffic signs, Tshirt slogans, and on and on and on. It's hardly surprising that attention spans are getting shorter and that the majority of people believe themselves to be busier than ever.

With this information overload, it is next to impossible to remember everything we need to, to call up names, dates, figures, phone numbers, email addresses and all the corporate and client information we need to do business effectively. That's why we use tools to do the remembering and information retrieval for us.

My company uses Salesforce.com to handle the bulk of our customer relationship management information. I use Microsoft Outlook to manage my email. When I want to find a product, service or piece of information online, I use a Search Engine.

I'm not alone in using Search Engines. Far from it. In the month of March 2006 alone, there were 6.4 billion searches. Assuming each user looks at an average of two search results pages, each of which displays 10 search results, that gives an average of 128 billion search results shown to Internet users in a single month. Search Engines are ubiquitous, and so accepted in contemporary culture that the word "Google" now appears in the dictionary as verb (as in "to Google something").

Search Engines essentially act as filters for the wealth of information available on the Internet. They allow users to quickly and easily find information that is of genuine interest or value to them, without the need to wade through numerous irrelevant web pages. There is a lot of filtering to donine years ago in 2004 the number of pages in Google's index exceeded the number of people of the planet, reaching the staggering figure of over 8 billion. With that much content out there, the Internet would be essentially unworkable without the Search Engines, with Internet users drowning in sea of irrelevant information and shrill marketing messages.

The goal of the Search Engines is to provide users with search results that lead to relevant information on high-quality websites. The operative word here is "relevant". To attain and retain market share in online searches, Search Engines need to make sure they deliver results that are relevant to what their users search for. They do this by maintaining databases of web pages, which they develop by using automated programs known as "spiders" or "robots" to collect information. The Search Engines use complex algorithms to assess websites and web pages and assign them a ranking for relevant search phrases. These algorithms are jealously guarded and frequently updated. Google looks at over 200 different metrics when assessing websites, including copy, in-bound links, and website usability and information architecture.

What this means is that the Search Engines provide users with the information they are looking for, and not necessarily the information that marketers would like them to see. Type the name of a major brand into Google, and you will most probably be served a wide range of search results that include not only the official website of the brand you searched for, but also other websites, consumer review sites, Blogs, online articles on Web 2.0 sites and press releases on news syndication channels. Of course, not all searches are for brand names. The majority of searches are for non-brand key phrases for example, "Hong Kong luxury hotel" rather than "The Peninsula Hong Kong". With key phrases that are service or product-specific rather than brandspecific, results pages will also include many competitors, which makes acquiring a prominent position at the top of the page even more crucial.

There are two major ways to make sure a website appears in a prominent location on the major Search Engines for relevant key phrases: Paid Search (also known as Pay-Per-Click) and Organic Search Engine Optimization. Of the two, Organic Search Engine Optimization tends to yield the best long-term results and the optimum return on investment, for the simple reason that Internet users are four times as likely to click an Organic search result as they are a Pay-Per-Click ad on the same results page. In a September 2006 poll by Marketing Sherpa, 68.7% of marketers in the US identified Search Engine Optimization as yielding the best Return on Investment for product marketing. I will discuss Paid and Organic search in much more depth in a separate article. It is enough here to state that companies doing business or marketing online should look at striking a healthy balance of both techniques to make the most of the potential of marketing through the major Search Engines.

Search Engines matter because they increasingly determine the information about brands, products and services that customer's access online. Being easy to find on Google, Yahoo and MSN is now as much of a marketing necessity as having a strong presence in print and broadcast media, or an effective traditional direct marketing program. And as consumers and organizations come to rely more heavily on them to find the goods, services and suppliers they need, the importance of the Search Engines to modern businesses can only increase.

2.2 SECONDARY WEB-SEARCHING

Secondary Web-searching Studies, is a review of Web searching studies that are more limited in scope in that they do not present enough data to give a full picture of Web searching. Most of these studies analyzed Web searching on a singular Web site that was not a search engine.

III. MOST POPULAR SEARCH ENGINES

Knowing which search engines are getting the largest percentage of search traffic plays a big role in deciding the focus of optimization efforts. Google search statistics, for example, show us that Google is still the king of search traffic, accounting for 66.52% of all search traffic in July 2013. Bing and Yahoo! follow further behind with 11.40% and 8.40% respectively, while Ask is at 4.14% and AOL Search is at 1.84%.



Figure 2 : Search Engine usage for July 2013

Most people don't want 290 search engines, especially people who are internet beginners. Most users want a single search engine that delivers three key features:

- 1. Relevant results (results you are actually interested in)
- 2. Uncluttered, easy to read interface
- 3. Helpful options to broaden or tighten a search

With these criteria, 10 Reader Favourite Search Engines come to mind. These 10 search sites should meet 99% of the searching needs of a regular everyday user.

Below is a changing list of user favourites, compiled from reader email suggestions. The sites below are in random order, and are updated regularly to reflect changes and user suggestions.

3.1. Ask (aka 'Ask Jeeves')



The Ask/AJ/Ask Jeeves search engine is a long-time name in the World Wide Web. The superclean interface rivals the other major search engines, and the search options are as good as Google or Bing or DuckDuckGo. The results groupings are what really make Ask.com stand out. The presentation is arguably cleaner and easier to read than Google or Yahoo! or Bing, and the results groups seem to be more relevant. Decide for yourself if you agree... give Ask.com a whirl, and compare it to the other search engines you like.

3.2. Bing



Bing is Microsoft's attempt at unseating Google. Bing used to be MSN search until it was updated in summer of 2009. Touted as a 'decision engine', Bing tries to support your researching by offering suggestions in the leftmost column, while also giving you various search options across the top of the screen. Things like 'wiki' suggestions, 'visual search', and 'related searches' might be very useful to you. Bing is not dethroning Google in the near future, no. But Bing is definitely worth trying.

3.3. Duck Duck Go



At first, DuckDuckGo.com looks like Google. But there are many subtleties that make this spartan search engine different. DuckDuckGo has some slick features, like 'zero-click' information (all your answers are found on the first results page). DuckDuckgo offers disambiguation prompts (helps to clarify what question you are really asking). And the ad spam is much less than Google. Give DuckDuckGo.com a try... you might really like this clean and simple search engine.

3.4. Dogpile

dögpile

Years ago, Dogpile was the fast and efficient choice before Google. Things changed, Dogpile faded into obscurity, and Google became king. But today, Dogpile is coming back, with a growing index and a clean and quick presentation that is testimony to its halcyon days. If you want to try a search tool with pleasant presentation and helpful crosslink results, definitely try Dogpile.

3.5. <u>Webopedia</u>

(Wěbopēdia)

Webopedia is one of the most useful websites on the World Wide Web. Webopedia is an encyclopedic resource dedicated to searching techno terminology and computer definitions. Teach yourself what 'domain name system' is, or teach yourself what 'DDRAM' means on your computer. Webopedia is absolutely a perfect resource for nontechnical people to make more sense of the computers around them.

3.6. <u>Yippy (formerly 'Clusty')</u>



Yippy is a Deep Web engine that searches other search engines for you. Unlike the regular Web, which is indexed by robot spider programs, Deep Web pages are usually harder to locate by conventional search. That's where Yippy becomes very useful. If you are searching for obscure hobby interest blogs, obscure government information, tough-to-find obscure news, academic research and otherwise-obscure content, then Yippy is your tool.

3.7. The Internet Archive



The Internet Archive is a favorite destination for longtime Web lovers. The Archive has been taking snapshots of the entire World Wide Web for years now, allowing you and me to travel back in time to see what a web page looked like in 1999, or what the news was like around Hurricane Katrina in 2005. You won't visit the Archive daily, like you would Google or Yahoo or Bing, but when you do have need to travel back in time, use this search site.

3.8. <u>Mahalo</u>



Mahalo is the one 'human-powered' search site in this list, employing a committee of editors to manually sift and vet thousands of pieces of content. This means that you'll get fewer Mahalo hit results than you will get at Bing or Google. But it also means that most Mahalo results have a higher quality of content and relevance (as best as human editors can judge). Mahalo also offers regular web searching in addition to asking questions. Depending on which of the two search boxes you use at Mahalo, you will either get direct content topic hits or suggested answers to your question.

Try Mahalo. You might like it enough to even become a editor there.

3.9. <u>Yahoo!</u>



Yahoo! is several things: it is a search engine, a news aggregator, a shopping center, an emailbox, a travel directory, a horoscope and games center, and more. This 'web portal' breadth of choice makes this a very helpful site for Internet beginners. Searching the Web should also be about discovery and exploration, and Yahoo! delivers that in wholesale quantities.

3.10. Google



Google is the undisputed king of 'spartan searching'. While it doesn't offer all the shopping center features of Yahoo!, Google is fast, relevant, and the largest single catalogue of Web pages available today. Make sure you try the Google 'images', 'maps' and 'news' features... they are outstanding services for locating photos, geographic directions, and news headlines.

week ending August 24, 2015					
Web sites	Total Visits	Visits Share	Rank 08/17	Rank 08/10	Rank 08/03
Google	2,468,944,85 6	67.26%	1	1	1
Bing	401,735,522	10.94%	2	2	2
Yahoo! Search	301,801,156	8.22%	3	3	3
Ask	151,567,463	4.13%	4	4	4
AOL Search	62,010,716	1.69%	5	5	5

Top 5 Search Engines by Total Visits Week ending August 24, 2013

Source: Hitwise US

© 2013 Hitwise Pty Ltd. All rights reserved.

IV. CONCLUSION

Successful searching of information on the internet depends on techniques such as going straight to the information source, guess work and developing strategies for when to use subject directories and [13] 2002.

search engines. By indexing a target Web page more accurately, and allowing each user to perform more fine-grained search that satisfy his/her information need.

REFERENCES

- [1] R. Baeza-Yates, F. Saint-Jean, and C. Castillo. Web Structure, Dynamics and Page Quality. In Proc. of the 9th International Symposium on String Processing and Information Retrieval (SPIRE 2002), pages 117–130, 2002.
- [2] Henzinger et. al. "Challenges in web search engine."ACM SIGIR Forum, 2002 dl.acm.org.
- [3] Abhishek Das, Ankit Jain "Indexing the World Wide Web : The Journey So Far."Publisher: IGI Global, Pages: 1-24.
- [4] Monika R.Henzinger"Algorithmic in Web Search Engines" Internet Mathematics Vol.1, No.1:115-126.
- [5] K. Sugiyama. A Method of Reranking Search Results Using their Hidden Hyperlink Structure. In Doctoral Posters Program of 28th International Conference on Very Large Data Bases (VLDB2002), http://www.cs.ust.hk/vldb2002/VLDB2002proceedings/papers/S34P14.pdf, 2002.
- [6] http://www.webopedia.com.
- [7] http://www.google.com.
- [8] http://docs.google.com.
- [9] Rainer Olbrich, Carsten D. Schultz "Search Engine Marketing and Click Fraud" Business Information Review (2003) Volume: 20, Issue: 4, Pages: 195-202 ISSN: 02663821 DOI: 10.1177/0266382103204005
- [10] TamarSadeh "User centric solutions for scholarly research in the library." LIBER QUARTERLY, ISSN 1435-5205 © LIBER 2007. All rights reservedIgitur. Utrecht Publishing & Archiving Services.
- [11] Jansex et. al. "Determining the user intent of web search engine queries." WWW '07 Proceedings of the 16th international conference on World Wide Web ACM New York, NY, USA ©2007 table of contents ISBN: 978-1-59593-654-7.
- [12] J. Wang, Z. Chen, L. Tao, W.-Y. Ma and L. Wenyin. Ranking User's Relevance to a Topic through Link Analysis on Web Logs. In Proc. of the 4th ACM CIKM International Workshop on Web Information and Data Management (WIDM'02), pages 49–54,
- [14] Monika Hezinger, Jeffrey Beall "The

solution to weaknesses of web search engine" Science Publishing Online august 2007.

- [15] www.cloaking.com.
- [16] Prof. Dr. Andreas Meier "Meta Search Engine Analysis"University of Fribourg Faculty of Economics Social Sciences Information Systems Research Group Meta-Search Engine Analysis Seminar Thesis September 2010.

Author's Profile



Trilok Gupta was born in Kota (Rajasthan). He received his Master Degree in Computer Science from Janardan Rai Nagar Rajasthan Vidhyapeeth University, Udaipur, Rajasthan -India. He is pursuing Ph.D. Computer Science from Faculty of Computer Application, Pacific University Udaipur-PAHER, Rajasthan -India. His area of interest includes Web Applications, Search Engines and Information Exploring.



Archana Sharma was born in Ajmer (Rajasthan). She is Ph.D in Computer Science and Engineering with specialization in Simulation and Modeling. She completed her M.Tech in Computer Science from Banasthali Vidyapith, India. Her field of study is Simulation and modeling, data mining, database, Artificial Intelligence.

She is working in the field of education for last 15 years. She has taught many subjects at undergraduate and postgraduate level. She has published several research papers in national and International Journals. She is currently working in the field of Cloud Computing, Artificial Intelligence and Educational Data Mining. She is Senior Member of International Association of Computer Science and Information Technology (IACSIT). She is also the Board member in Seventh Sense Research Group Journals.

Professor Sharma is member of Indian Society of Theoretical and Applied Mechanics. She worked as Editor in Journal of Management and IT 'OORJA'.