Digital Innovations & Contemporary Research in SCIENCE & ENGINEERING
Vol. 4 No. 4, Dec. 2016

Digital Innovations & Contemporary Research in SCIENCE & ENGINEERING

The Digital Innovations & Contemporary Research in Science & Engineering Journal is an International Journal of
The Multidisciplinary Research Group
International Centre for IT & Development, USA

Article Citation Format
Yaya, J.O. & Robert, A.B.C (2016): Towards the Development of an Enhanced Recursive Model for Individualized Search.
Journal of Digital Innovations & Contempt Res. In Sc., & Eng Vol. 4, No. 4. Pp 109-121

Article Progress Time Stamps

Article Type: Research Article
Manuscript Received: 18rd September, 2016
Review Type: Blind
Word Count Post Review: 6107
Review/Acceptance Information Sent : 12th Nov, 2016
Final Acceptance:: 21st Dec, 2016
DOI Prefix: 10.22624

Towards the Development of an Enhanced Recursive Model for Individualized Search

Yaya, J.O & Robert, A.B.C.
Department of Computer Science
University of Ibadan
Ibadan, Nigeria

ABSTRACT

Various searching mechanisms and algorithms are used to query and rank information on the internet. QuickRank Algorithm (QRA) is a popular method for ranking entities based on various criteria selected by the user. However, this algorithm fails to handle evolving individualism and as a result ignored user biases and preferences. The aim of this work was to develop an enhanced recursive ranking model that could improve searching experience incorporating user biases.

Keywords: Development, Enhanced Recursive Model, Individualized Search & Ranking

1. BACKGROUND TO THE STUDY

Information retrieval plays significant role in all aspects of life, especially with the tremendous increase in inaccessible data. The problem of finding relevant information by searchers is made more difficult by the complex nature and structure of accessible data. A simple web query can return millions of results, through all of which practically nobody has the time or patience to go through. Users expect the most relevant results to be shown first, before the less relevant. Moreover, different users may have different search intent, and as such expect different results. Search efficiency becomes increasingly important as the number of internet users continues to increase.

With the popularity of online social networks, and the increasingly bigger role played by the Web in research, studies, and other areas of life, personalization of search results becomes a major issue. To solve this problem, there is a need to develop efficient algorithms and mechanisms that incorporate the specific needs of users in returning relevant information to them. Researches are being carried out in this area. This work is geared towards proposing solution to alleviate the difficulty that is associated with personalised searching experience. This chapter presents a brief summary on ranking algorithms, followed by the objectives, significance, and organization of this thesis.

The advent and introduction of internet is among the major inventions of the 21st century. Arguably, the origin can be traced back to when the Internet was employed as an instrument to inter-connect government research laboratories. From its inception in 1991, the internet has evolved as a global phenomenon interconnecting various networks from individuals' networks to corporate, private and public networks. It provides medium for exchanging information in various forms, including research papers, policy documents, learning documents, multimedia contents and private data. Internet as the largest source of data has revolutionised the availability of electronically accessible information. On the average, the Web is estimated to be growing by over a million pages per day, in addition to the hundreds of millions of pages that are already available online (Hobbes' Internet Timeline, 2017; Netcraft, 2017). The Internet presents an enormous amount of information, which consists of huge collection of webpages (currently estimated at over 4 billion) and countless hyperlinks.

Globally, demand for access to different types of information by diverse users is on the increase - content search in the digital libraries; media search like image, music, news, blog; social search, and video retrieval. As a result of this, in addition to the enormous surge in the quantity of text that is accessible online, web information retrieval grows beyond simple web page retrieval. It comprises of varying amount of processes including question answering, web object detection, and tracking, summarization, multimedia retrieval, searched results mining (Brin & Page, 1998). These features of the web, essentially bring about the creation of systems that are highly intelligent and capable of mining knowledge efficiently and effectively.

Due to exponential increase in the amount of information that is available online, it is now pertinent for users to use automated tools such as search engines while looking for desire sources and resources for information. These tools use web mining methods for data extraction in other to obtain vital information through the internet. Web structure mining is used in various aspect of web searching which include the following; information discovery, hyperlink inferences, authoritative web pages information, retrieval of relevant information and qualitative check of web page (Suri &Taneja, 2012). It has been in various research work that Hyperlink structure is the main focal point as regards the development of web mining techniques. A typical web graph is made up of webpages and hyperlinks. The webpages are referred to as nodes while the hyperlinks are the edges of the graph that connect one page to the other. A major application of web structure mining among others is web search results ranking. One of the important applications of web structure analysis is ranking the results of web search. In the search engine, ranking entails the ordering of search results or a query based on pre-defined measures or standards of relevance.

Qualitative measures in web searching process depends heavily on the quality of ranking algorithm that is implemented. In search engine mechanism, when query is issued to display relevant results, the quality of such results can be estimated by the quality of the ranking algorithm (Baeza-Yates et al., 2002). Ranking techniques greatly simplified the user interaction with large search results in the predefined relevance. Diverse kind of page ranking algorithm exist today, among the most popular ones are PageRank and HITS methods. In order to determine the relevance of the search results, most ranking mechanism use the underlying link structure of the Web.

In link structure analysis, every webpage normally consist of both in-link to the page and out-link from the page. Analysis of both in-link and out-link of a particular webpage provides insight about the context of the page. PageRank mechanism which is use in Google search engine is a very good example of link analysis algorithm (Montenegro & Tetali, 2006). By allocating numerical weights to each of the element of a hyperlinked, PageRank is able to computing its importance.

2. RELATED WORKS

There have been serious research efforts to improve the performance of ranking algorithms (Salton & Buckley, 1988; Jones et al., 2000; Craswell et al., 2001; Brin & Page, 1998; Kleinberg, 1999). However, many issues still remain to be explored and addressed. One of such issues is the problem of personalisation of search results. Most of the ranking algorithms are user-neutral (Brin & Page, 1998; Kleinberg, 1999) and are very useful tools in the process of content-based document analysis. Despite the effectiveness of ranking algorithm in web searching process, a core of component of every searching domain is not fully taken into consideration. This major actor is the user that initiates who is the custodian of whatever that is generated from the searching output. Some attempts were made in the past researches to integrate users' parameters into ranking algorithms but the focus was mainly on using user information without full consideration for bias.

Various forms of users' information are used by ranking algorithm to ensure quality ranking and improve ranking results. Example of such users information include search context, location, history of previous searches, click-through logs, (Sun et al., 2005), topics of interest (Chirita et al., 2001) and personal bookmarks or frequently visited web pages (Jeh & Widom, 2003), to modify search result's weights. User friends' information needs are also taken into consideration in some algorithm. This is premised on believe that friends naturally share common interest. Consequently, the ranking algorithm were mainly concerned with learning the users previous online engagement while ignoring the effect of user bias in the current query. In reality, users may have certain preferences in their current information search, which differ from their previous activities.

A vivid example is a case of a mother looking for a good Nigerian university for her ward. She simply opened a search engine and typed "good universities in Nigeria", search engines simply return a generalised result as it cannot determine which criteria make a university good using the user's personalised bias criteria. Such searches normally return a result with generalised assumptions. In summary, the current personified search engine method is limited to an integrated information search about the visit and search histories of different individuals and their peers. With the integration of user bias, the searching process considers the specificity of each individual in the query criteria thereby delivering more relevant information.

Web content searching is known to be extremely difficult because such search has to traverse various information sources. These sources are very dynamic in term of content development and increasing volume of information on a continual basis. The Search and meta-search engine are tools that aid the user to easily identify information they are looking for. Search engines facilitate the retrieval of pages that contained necessary information based on the keyword supplied by the user. Web personalisation has recently gained interest from several stakeholders. Some websites are already involved with providing services including interface preferences, layout, and applications functions.

Interestingly, some of these services focus mainly on the presentation approach without considering issues such as retrieval model and ranking algorithm. Synthesizing web content in thematic categories is a vital function in searching techniques. With this, users are able to navigate individual search domain and retrieve desired information without much difficulties. Categorisation of information in thematic

hierarchies can be done using either "flatten the hierarchy" or text categorisation technique.

In flatten the hierarchy approach, categorisation is done in such a way that every section in the hierarchy has a correspondent grouping with personalised training data. In text categorisation technique, every new incoming web document is classified. This classification of document is developed in such a way that each classifying agent is able to determine when to categorise a document in a particular class among various classes that exist at the same level in a given hierarchy while traversing the hierarchy tree from the root node to the leave nodes

3. RESEARCH DIRECTION

This thesis addressed the problem of personalised search. As stated above, this problem is rooted in the fact that current ranking algorithms do not adequately provide room for users to search for items by specifying their preferences (explicitly or implicitly inferred). The popular ranking algorithms rank pages using only the search query keywords and the underlying link topology of the Web. Consequently, results of such searching exercise only reflects the matching criteria based on the keywords used in the query. Searching with criteria such as weighted age, gender and personality is not presently possible with the algorithm in the public domain. This is the reason why it is presently difficult to search for documents based of on characteristics. Underlying topology of a network is the techniques that is being used to rate documents (for instance web pages) within a network. In such case, ranking is done based on the search word(s) used by the user. Search engines basically rank documents based on word and not by users' traits. For instance, searching for individuals between the age of 30 and 40 who are interested in ice skinning with a certain degree of interest is almost impossible.

The objectives of the research work in this thesis are to understand the principles of operations of ranking algorithms, investigate the problem of personalised search, and evaluate ways to solve the problem. An efficient recursive ranking algorithm, which considered a searcher's preferences or bias when ranking the relevancy of documents, was proposed, developed, and tested on a standardised dataset. The results of the thesis revealed insights into methods of improving ranking algorithms to provide better personalised searching experience.

4. RESEARCH APPROACH

User bias was introduced as third parameter to local ranking and judgement in QRA. This parameter was applied at the tree, node and leaf levels to generate the resultant global ranking. The ranking dataset of top 100 Nigerian universities (University of Ibadan (UI), Covenant University (CU), University of Nigeria (UNN), Federal University of Technology Owerri (FUTO), University of Lagos (UNILAG), University of Port-Harcourt (UNIPORT), Federal University of Agriculture Abeokuta (FUNNAB), University of Ilorin (UNILORIN), Ahmadu Bello University (ABU) etc.) was obtained from 2016 webometric ranking data developed on the basis of presence, openness, impact, and excellence. The data were normalised to generate weighted criteria scores. Evolving individualism was taken into consideration at the criteria level using 870 sample candidates obtained through response driven survey. The model was validated using the data obtained from the 870 respondents. This was to make room for independent non-negative bias value for each of the criteria. The specified values for the four criteria were nomarlised to range from 0.00 – 1.00 with the cumulative value equal 1.00. These values were used to bias the criteria score for each University based on QRA. Cardinal and ordinal ranking were generated for each sample individual. Weighted mean was used to determine the weighted disparity between the actual webometric ranking and individually biased ranking.

5. CONCLUDING REMARKS

This study covers one of the major aspect of information retrieval (IR) technology, an area which has not been given adequate effort. There are no widely used IR systems and search engines that incorporates user preferences in ranking search results. This is quite an important application, since the expectation of users varies. Moreover, a personalised ranking algorithm will allow users of online social network to easily find and connect with people, whose characteristics closely match with the preferences of the users. All these have the potential to improve users' browsing experience. Moreover, the recursive nature of the algorithm provides valuable insight into the reduction of the computational complexity of ranking algorithms through the use of recursion.

The solution proposed in this research work can be applied to other scenarios. If it can be applied to a hierarchical social network, it is also mostly relevant in other similar situations. For instance, a person might search for an automobile, which best suits their needs from a database by indicating factors most important to them. One person might consider gas mileage the most important factor, followed by initial cost; while another might consider style as the most important, and performance as a close second.

In this case, the proposed solution could be modified to generate a ranking result on the basis of user's inherent uniqueness. For the two searchers, the algorithm would recommend quite different cars (Rosvall & Bergstrom, 2011).

BIBLIOGRAPHY

Ankur, G & Rajni, J (2008). An overview of ranking algorithms for search engines. *Proceedings of the 2nd National Conference; INDIACom-2008, New Delhi. Feb 08 – 09.*

Arasu, A; Novak, J; Tomkins, A; Tomlin, J (2002). PageRank computation and the structure of the web: Experiments and algorithms. *Proceedings of the Eleventh International World Wide Web Conference, Poster Track. Brisbane, Australia. 107-117.*

Baeza-Yates, R; Saint-Jean, F & Castillo, C (2002). *Web Dynamics, structure and page ranking*. Proceedings of 9th International Symposium on String Processing and Information Retrieval, SPIRE, Springer LNCS, Lisbon, Portugal, 117–130.

Bates, M, J; Wilde, D, N, & Siegfried, S (1993). An analysis of search terminology used by humanities scholars: *The Getty Online Searching Project report no. 1. Library Quarterly, 63(1), 1-39*

Bates, M. J (1979). Information search tactics. *Journal of the American Society for Information Science, 30(4), 205-214.*

Bates, M; Idea, J. (1979). Tactics. *Journal of the American Society for Information Science, 30 (5), 280-289.*

Beel, J; Gipp, B; Stiller, J (2009). Information retrieval on mind maps – what could it be good for? Proceedings of the 5th International Conference on Collaborative Computing: Networking, Applications and Worksharing (CollaborateCom'09). Washington, DC: IEEE.

Belkin, N, J & Croft, W, B (1987). "Retrieval Techniques," in Annual Review of Information Science and Technology, ed. M. Williams. New York: Elsevier Science Publishers, 109-145.

Belkin, N. J (1996). *Intelligent information retrieval: whose intelligence?* In Proceedings of the 5th International Symposium for Information Science (ISI '96): Humboldt-Universität zu Berlin, 17. -19. Oktober 1996; Krause, J., Herfurth, M., Marx, J., Eds.; Universitätsverlag Konstanz: Konstanz, Germany, 25-31.

Belkin, N. J. (1993). Interaction with texts: Information retrieval as information seeking behaviour. In Information Retrieval '93: Von der Modellierung zur Anwendung, Knorz, G., Krause, J., Womser-Hacker, C. Eds.; Universitaetsverlag Konstanz: Konstanz, Germany, 1993; 55-66.

Belkin, N. J.; Cool, C.; Stein, A.; Thiel, U (1995). Cases, scripts and information seeking strategies: On the design of interactive information retrieval systems. Expert Systems with Applications, 9 (3), 379-395.

Belkin, N. J; Marchetti, P, G; Cool, C (1993). Design of an interface to support user interaction in information retrieval. *Information Processing and Management, 29 (3), 325-344*.

Bellardo, T (1985). What do we really know about online searchers? *Online Review, 9 (3), 223-239.*

Bhavnani, S. K (2002). *Important cognitive components of domain-specific search knowledge*. In The Tenth Text REtrieval Conference, TREC-2001; Voorhees, E.M.; Harman, D.K. Eds.; Information Today: Medford, NJ, 571-578.

Bilal, D (2002). Perspectives on children's navigation of the World Wide Web: Does the type of search task make a difference. *Online Information Review, 26 (2), 108-177.*

Bjørklund, T. A., Götz, M., Gehrke, J., & Grimsmo, N. (2011, October). Workload-aware indexing for keyword search in social networks. In Proceedings of the 20th ACM international conference on Information and knowledge management (pp. 535-544). ACM

Bonacich, P (1972). Factoring and weighting approaches to status scores and clique detection. *Journal of Mathematical Sociology, pages 113–120*

Brereton, P., Kitchenham, B. A., Budgen, D., Turner, M., & Khalil, M. (2007). Lessons from applying the systematic literature review process within the software engineering domain. *Journal of Systems and Software, 80(4), 571-583*.

Broder A. Z., Kumar R., Maghoul F., Raghavan P., Rajagopalan S., Stata R., Tomkins A. & Wiener J.L. (2000). "Graph structure in the Web", Proceedings of the 9th International World Wide Web Conference, Amsterdam, The Netherlands, pp. 309-320.

Bruza, P. D; Dennis, S (1997). *Query-reformulation on the Internet: Empirical data and the hyperindex search engine.* In RIAO 97: Conference proceedings with prototype and operational systems demonstrations: Computer-assisted information searching on Internet, McGill University, Montreal, Quebec, Canada, 25th-27th June 1997; RIAO 97, Ed.; CID: Paris, 1997; Vol. 1, 488-499.

Bryman, A. (2012). *Social research methods*. OUP Oxford.

Byström, K (2002). *Information and information sources in tasks of varying complexity*. *Journal of the American Society for Information Science and Technology, 53 (7), 581-591*.

Byström, K; Järvelin, K (1995). Task complexity affects information-seeking and use. *Information Processing and Management, 31(2), 191-213.*

Callahan, E. (2005). Interface design and culture. *Annual Review of Information Science and*

Technology, 39, 257-310.

Chang, S (1995). Toward a multidimensional framework for understanding browsing. Unpublished doctoral dissertation, Rutgers University: New Brunswick, N, J.

Chen, H & Dhar, V (1991). Cognitive processes as a basis for intelligent retrieval system design. Information Processing and Management, 27 (5), 405-432.

Cheng, A & Friedman, E (2006). Manipulability of PageRank under sybil strategies. In First Workshop on the Economics of Networked Systems (NetEcon06), URL http://www.cs.duke.edu/nicl/netecon06/papers/ne06-sybil.pdf.

Cheng, Y., Park, J., & Sandhu, R. (2012). A user-to-user relationship-based access control model for online social networks. *Data and Applications Security and Privacy XXVI, 8-24.*

Chu, H (2003). *Information Representation and Retrieval in the Digital Age*; Information Today: Medford, N. J

Cole, C (2001). Intelligent information retrieval: Part IV. Testing the timing of two information retrieval devices in a naturalistic setting. *Information Processing and Management, 37 (1), 163-182.*

Craswell, N., Hawking, D & Robertson, S. (2001). *Effective site finding employing link anchor information.* In Proceedings of the ACM Conference on Information Retrieval (SIGIR '01), 250-257.

Dalal, M. (2007). *Personalized social & real-time collaborative search*. In Proceedings of the International World Wide Web Conference (WWW '07), 1285-1286.

Dieste, O., Grimán, A., Juristo, N., & Saxena, H. (2011, September). Quantitative determination of the relationship between internal validity and bias in software engineering experiments: consequences for systematic literature reviews. In Empirical Software Engineering and Measurement (ESEM), 2011 International Symposium on (pp. 285-294). IEEE.

Drabenstott, K. M (2003). Do nondomain experts enlist the strategies of domain experts? *Journal of the American Society for Information Science and Technology, 54 (9), 836-854.*

Duhan, N., Sharma, A. K., & Bhatia, K. K. (2009, March). *Page ranking algorithms: a survey*. In Advance Computing Conference, 2009. IACC. IEEE International (pp. 1530-1537). IEEE.

Dumais, S, T; Belkin, N, J (2005). The TREC interactive tracks: Putting the user into search. In TREC: Experiment and Evaluation in Information Retrieval; Voorhees, E.M.; Harman, D.K., Eds.; The MIT Press: Cambridge, MA, 123-152.

Earhart, S. (1986). The UNIX Programming Language, vol. 1. New York: Holt, Rinehart, and Winston.

Ellis, D & Haugan, M (1997). Modeling the information seeking patterns of engineers and research scientists in an industrial environment. *Journal of Documentation, 53 (4), 384-403*.

Ellis, D. A (*1989).* Behavioural approach to information retrieval system design. *Journal of Documentation, 45 (3), 171-212.*

Faloutsos, C. (1985). "Access Methods for Text," *Computing Surveys, 17(1), 49-74*.

Fenichel, C. H (1981). Online searching: Measures that discriminate among users with different types of experience. *Journal of the American Society for Information*

Science, 32 (1), 23-32.

Fidel, R & Pejtersen, A, M (2004). From information behavior research to the design of information systems: The cognitive work analysis framework. *Information Research, 10 (1). http://informationr.net/ir/10-1/paper210.html*.

Fidel, R & Soergel, D (1983). Factors affecting online bibliographic retrieval: A conceptual framework for research. *Journal of the American Society for Information Science, 34 (3), 163-180.*

Fidel, R (1985). Moves in online searching. *Online Review, 9 (1), 61-74*

Foote, Jonathan (1999). "An overview of audio information retrieval". Multimedia Systems. Springer.

Ford, N, Wilson, T, Foster, D, Ellis, A, & Spink, A, D (2002). Information seeking and mediated searching. Part 4.Cognitive styles in information seeking. *Journal of the American Society for Information Science and Technology, 53 (9), 728-735.*

Ford, N; Miller, D, & Moss, N (2002). Web search strategies and retrieval effectiveness: An empirical study. *Journal of Documentation, 58 (1), 30-48.*

Frakes, W, B. (1992). Information retrieval data structures & algorithms. Prentice-Hall, Inc

Franceschet, M (2002). "PageRank: Standing on the shoulders of giants".

Fujimura K., Inoue R., and Sugisaki M. (2005). "EigenRumor Algorithm for Ranking Blogs". May 10-14, 2005, Chiba, Japan.

G. Jeh, G & Widom, J (2003). Scaling personalized web search. In Proceedings of the International World Wide Web Conference ('03), 271-279, 2003.

García-Crespo, Á, Colomo-Palacios, R., Gómez-Berbís, J. M., & García-Sánchez, F. (2010). SOLAR: social link advanced recommendation system. *Future Generation Computer Systems, 26(3), 374-380*

Ghafari, M., Saleh, M., & Ebrahimi, T. (2012). A federated search approach to facilitate systematic literature review in software engineering. *International Journal of Engineering, 23(5), 12-36*

Goodrum, Abby A. (2000). "Image Information Retrieval: An Overview of Current Research". Informing Science. 3 (2).

Google technology overview (2004) "http://www.google.com/intl/en/corporate/tech.html.

Google. Personalized search for everyone. http://googleblog.blogspot.com/2009/12/personalized-search-for-everyone.html. Last recoup on April, 16, 2010.

Greenwald, A., & Wicks, J. (2006). QuickRank: A recursive ranking algorithm. In Proc. of the 1st International Workshop on Computational Social Choice.

Greenwald, A., & Wicks, J. (2006). QuickRank: A recursive ranking algorithm. In Proc. of the 1st International Workshop on Computational Social Choice.

Gupte, M., Shankar, P., Li, J., Muthukrishnan, S., & Iftode, L. (2011, March). Finding hierarchy in directed online social networks. In Proceedings of the 20th international conference on World wide web (pp. 557-566). ACM.

Gyongyi Z. & Garcia-Molina H. Web spam taxonomy (2004). Technical report, Stanford University Technical Report.

Haveliwala, T. (2002). *"Topic-Sensitive PageRank*", In Proceedings of the 11th World wide Web Conference.

Hawk, W. B.; Wang, P (1999). *Users' interaction with the World Wide Web; Problems and problem solving.* Proceedings of the 62nd ASIS Annual Meeting, 36, 256-270.

Hema, D, B. & Roy, N. (2011). An improved Page Rank Algorithm based on Optimized Normalization Technique. International Journal of Computer Science and Information Technologies, 2 (5), 2183-2188

Hotho, A; Jäschke, R; Schmitz, C, & Stumme, G. (2006). *Information retrieval in folksonomies: Search and ranking.* In Lecture Notes in Computer Science. Vol. 4011/2006(the Semantic Web: Research and Applications), 411-426

Howard, H (1982). Measures that discriminate among online users with different training and experience. *Online Review, 6 (4), 315-326*.

Hsieh-Yee, I (1993). Effects of search experience and subject knowledge on the search tactics of novice and experienced searchers. *Journal of the American Society for Information Science, 44 (3), 161-174.*

Hyldegard, J (2006). Collaborative information behaviour: Exploring Kuhlthau's information search process model in a group-based educational setting. *Information Processing and Management, 42 (1), 276-298.*

Ingwersen, P & Järvelin, K (2005). The Turn: Integration of Information Seeking and Retrieval in Context; Springer. Germany: Heidelberg

Ingwersen, P (1992). *Information retrieval interaction*; Taylor Graham: London.

Ingwersen, P (1996). Cognitive perspectives of information retrieval interaction: Elements of a cognitive IR theory. *Journal of Documentation, 52 (1), 3-50*.

J.-T. Sun, H.-J. Zeng, H. Liu, Y. Lu & Z. Chen. (2005). *Cubesvd: A novel approach to personalized web search.* In Proceedings of the International World Wide Web Conference (WWW'05), pp. 382-390.

Jamshidi, P., Ghafari, M., Aakash, A., & Pahl, C. (2012). *A protocol for systematic literature review on Architecture-Centric Software Evolution Research*. Technical Report, Lero-The Irish Software Engineering Research Centre, Dublin City University.

Jansen, B. J. & Rieh, S. (2010). The Seventeen Theoretical Constructs of Information Searching and Information Retrieval. Journal of the American Society for Information Sciences and Technology. 61(8), 1517-1534.

Jansen, B. J.; Spink, A.; Saracevic, T (2000). Real life, real users, and real needs: A study and analysis of user queries on the Web. *Information Processing and Management 2000, 36 (2), 207-227.*

Joachims, T. (2002). *Optimizing search engines employing clickthrough data.* In Proceedings of the ACM Conference on Knowledge Discovery and Data Mining (SIGKDD'02), 133-142

Jones, K. S., Walker, S & Robertson, S. E (2000). A probabilistic model of information retrieval: Development and comparative experiments. Information Processing and Management. 36 (6), 779-808, 2000.

Jones, S.; Cunningham, S.J.; McNab, R.; Boddie, S (2000). Human-computer interaction for digital libraries: A transaction log analysis of a digital library. *International Journal on Digital Libraries , 3 (2), 152-169.*

K. Bharat and G.A. Mihaila (2002). When experts agree: Employing Non-Affiliated Experts to Rank Popular Topics. *ACM Transactions on Information Systems, 20(1), 47-58,.*

Kitchenham, B., Pretorius, R., Budgen, D., Pearl Brereton, O., Turner, M., Niazi, M., & Linkman, S. (2010). Systematic literature reviews in software engineering–a tertiary study. *Information and Software Technology, 52(8), 792-805*.

Kleinberg J. M. (1999). Authoritative sources in a hyperlinked environment. *Journal of the*

ACM, 46(5), 604 –632.

Kracker, J (2002). Research anxiety and students' perceptions of research: An experiment: Part 1. Effect of teaching Kuhlthau's ISP model. Journal of the American Society for Information Science and Technology, 53 (4), 282-294.

Kuhlthau, C. C. (1991). Inside the search process: Information seeking from the user's perspective. Journal of the American Society for Information Science 1991, 42 (5), 361-371.

Kumar, R., Novak, J., & Tomkins, A. (2010). Structure and evolution of online social networks. Link Mining: Models, Algorithms, and Applications, 337-357.

L. Srour, A Kayssi and A. Chebab (2007). Personalized Web Page Ranking Employing Trust and Similarity", IEEE 2007.

Lau, T.; Horvitz, E (1999). Patterns of search: Analyzing and modeling Web query refinement. In Proceedings of the 7th International Conference on User Modeling Banff, Canada, June 1999; Kay, J., Ed.; Springer-Wien: New York, 119-128.

Lazonder, A. W.; Biemans, H. J. A.; Wopereis, I. G. J. H (2000). Differences between novice and experienced users in searching information on the World Wide Web. *Journal of the American Society for Information Science, 51 (6), 576-581.*

Lempel R. & Moran S. (2000). *The stochastic approach for link-structure analysis (SALSA) and the TKC effect.* Proceedings of the 9th. *International World Wide Web Conference, Amsterdam, The Netherlands, pp. 387 – 401*

Liu N. C. and Cheng Y. (2005). The academic ranking of world universities. Higher Education in Europe, 30(2), 1-14

M. Montaner, B. López, & J. L. de La Rosa (2003). A taxonomy of recommender agents on the internet. *Artificial Intelligence Review. 19(4), 285-330.*

M. Richardson and P. Domingos (2002). The Intelligent Surfer: Probabilistic Combination of Link and Content Information in PageRank", Advances in Neural Information Processing Systems, 14, 1441-1448, MIT Press.

M.S. Aktas, M.A. Nacar, and F. Menczer (2004). Personalizing PageRank Based on Domain Profiles, WebKDD.

Manning, C., Raghavan, P & Schutze, H (2008). *Introduction to information retrieval*. Cambridge: University Press.

Marchionini, G (1995). *Information-Seeking in electronic environments*; Cambridge University Press: Cambridge,.

Marchionini, G.; Dwiggins, S.; Katz, A.; Lin, X (1993). Information seeking in full-text end-user-oriented search-systems - The roles of domain and search expertise. *Library and Information Science Research, 15 (1), 35-69.*

Markey, K., Atherton, P (1978). *Online training and practice manual for ERIC Database Searchers*; ERIC Clearinghouse on Information Resources: Syracuse, NY

Marques, A. B., Rodrigues, R., & Conte, T. (2012, August). Systematic Literature Reviews in Distributed Software Development: A Tertiary Study. In Global Software Engineering (ICGSE), 2012 IEEE *Seventh International Conference on (pp. 134-143). IEEE.*

Micarelli, A. Gasparetti, F; Sciarrone, F & Gauch, S (2007). *Personalized search on the World Wide Web*. In Lecture Notes in Computer Science. Volume 4321 (the Adaptive Web), pp. 195-230, 2007.

Mislove, A; Gummadi, K. P & Druschel, P (2006). Exploiting social networks for internet

search. In Proceedings of the 5th Workshop on Hot Topics in Networks (HotNets'06).

Montenegro R. & Tetali P. (2006). Mathematical aspects of mixing times in Markov chains. *Foundations and Trends in Theoretical Computer Science, 1(3), 237 – 354.*

Moukdad, H.; Bulk, A (2001). Users' perceptions of the Web as revealed by transaction log analysis. *Online Information Review, 25 (6), 349-359.*

Novak, J. D. (1998). Learning, creating, and employing knowledge: Concept maps as facilitative tools in schools and corporations. Mahwah, NJ: Lawrence Erlbaum Associates.

Novak, J. D., & Gowin, D. B. (1984). Learning how to learn. New York, NY: Cambridge University Press.

Novak, J., Fleishmann, M., Strauss, W., Schneider, M., Wurst, M., Morik, K., & Kunz, C. (2002). Augmenting the knowledge bandwidth and connecting heterogeneous expert.

Okoli, C. and Schabram, K. (2010). A Guide to Conducting a Systematic Literature Review of Information Systems Research. Sprouts. Working Papers on Information Systems, 10(26).

P. A. Chirita, W. Nejdl, R. Paiu & C. Kohlschütter (2001). Employing odp metadata to personalize search. In Proceedings of the ACM Conference on Information Retrieval (SIGIR '01), pp. 250-257.

Palmquist, R. A., & Kim, K. S (2000). Cognitive style and online search experience on Web search performance. *Journal of the American Society for Information Science and Technology, 51 (6), 558-567*.

Pennanen, M., Vakkari, M, & Students', P. (2003). Conceptual structure, search process and outcome while preparing a research proposal. *Journal of the American Society for Information Science, 54 (8), 759-770.*

Prasad Chebolu & Pall Melsted (2008). "*PageRank and the random surfer model",* Proceedings of 19th annual ACM-SIAM Symposium on Discrete Algorithms, pp. 1010-1018

Ready, R. C & Hu, D. (1995). Statistical Approaches to the Fat Tail Problem for Dichotomous Choice".

Rieh, S. Y.; Xie, H (2006). Analysis of multiple query reformulations on the web: The interactive information retrieval context. *Information Processing & Management, 42 (3), 751-768.*

Rosvall, M., & Bergstrom, C. T. (2010). Mapping change in bulk networks. *PloS one, 5(1), e8694*.

Rosvall, M., & Bergstrom, C. T. (2011). Multilevel compression of random walks on networks reveals hierarchical organization in bulk integrated systems. *PloS one, 6(4), e18209*.

S. Brin & L. Page (1998). The anatomy of a bulk-scale hypertextual Web search engine, In Proceedings of the International World Wide Web Conference (WWW '98), pp. 107-117, 1998. *International Journal on Web Service Computing (IJWSC), 3(1),*

Salton, G & Buckley, C (1988). Term-weighting approaches in automatic text retrieval. Information Processing and Management. 24(5), 513-523.

Saracevic, T (1996). Modeling interaction in information retrieval (IR): *A review and proposal. Proceedings of the 59th ASIS Annual Meeting 1996, 33, 3-9.*

Saracevic, T (1997). *The stratified model of information retrieval interaction: Extension and applications.* Proceedings of the 60th ASIS Annual Meeting 1997, 34, 313-327.

Schacter, J.; Chung, G. K. W. K.; Dorr, A (1998). Children's Internet searching on complex problems: Performance and process analyses. *Journal of the American Society for Information Science, 49 (9), 840-849.*

Schenkel, T. Crecelius, M. Kacimi, S. Michel, T. Neumann, J. X. Parreira, X. & Weikum, G. (2008). *Efficient top-k querying over social-tagging networks.* In Proceedings of the ACM Conference on Information Retrieval (SIGIR '08), pp. 523-530, 2008.

Shiri, A. A.; Revie, C (2003). The effects of topic complexity and familiarity on cognitive and physical moves in a thesaurus-enhanced search environment. *Journal of Information Science, 29 (6), 517-526.*

Shute, S. J.; Smith, P. J (1993). Knowledge-based search tactics. *Information Processing and Management, 29 (1), 29-45.*

Siegfried, S.; Bates, M.J.; Wilde, D.M (1993). A profile of end-user searching behavior by humanities scholars: The Getty online searching project (Rep. No. 2). *Journal of the American Society for Information Science, 44 (5), 273-291.*

Silverstein, C.; Henzinger, M.; Marais, H.; Morica, M (1999). Analysis of a very bulk Web search engine query log. *SIGIR Forum, 33 (1), 6-12.*

Soloman, P (1993). Children's information retrieval behavior: A case analysis of an OPAC. *Journal of the American Society for Information Science, 44 (5), 245-264.*

Spink, A.; Jansen, B. J (2004). Web Search: Public Searching of the Web; Kluwer Academic Publishers: Boston.

Spink, A.; Wolfram, D.; Jansen, B. J.; Saracevic, T (*2001)*. Searching the Web: The public and their queries. *Journal of the American Society for Information Science, 52 (3), 226-234*.

Suri, P. & Taneja, H. (March 2012). An Integrated Ranking Algorithm for Efficient Information Computing in Social Networks. International Journal on Web Service Computing (IJWSC), 3(1), 31 – 44.

Sutcliffe, A. G.; Ennis, M.; Watkinson, S. J (2000). Empirical studies of end-user information searching. *Journal of the American Society for Information Science, 51 (13), 1211-1231.*

Vakkari, P (2000). *ECognition and changes of search terms and tactics during task performance: A longitudinal study*. In RIAO' 2000 Conference Proceedings, Content-Based Multimedia Information, Collège de France, Paris, France, April 12-14, 2000; RIAO, Eds.; C.I.D.: Paris, 2000; 1, 894-907. http://www.info.uta.fi/vakkari/Vakkari_Tactics_RIAO2000.html (accessed July 15, 2018).

Vakkari, P (2000). *Relevance and contributory information types of searched documents in task performance*. In Proceedings of the 23rdAnnual International ACM SIGIR Conference on Research and Development in Information Retrieval; Belkin, N.J., Ingwersen, P., Leong, M-K, Eds; SIGIR forum; ACM Press: New York, Vol. 34, 2-9.

Vakkari, P. A (2001). Theory of the task-based information retrieval process. *Journal of Documentation, 57 (1), 44-60.*

Vakkari, P.; Hakala, N (2000). Changes in relevance criteria and problem stages in task performance. *Journal of Documentation, 56, 540-562*.

Vakkari, P.; Pennanen, M.; Serola, S (2003). Changes of search terms and tactics while writing a research proposal. *Information Processing and Management, 39 (3), 445-463*.

W. B. Frakes & R. Baeza-Yates. Information Retrieval: Data Structures & Algorithms, (2009).

Walker, G.; Janes, J (1999). Online Retrieval: A Dialogue of Theory and Practice, 2nd Ed.; Libraries Unlimited: Englewood, Colorado.

Wang, P.; Berry, M.; Yang, Y (2003). Mining longitudinal Web queries: Trends and patterns. *Journal of the American Society for Information Science and Technology, 54 (8), 743-758*.

Wang, P.; Hawk, W. B.; Tenopir, C (2000). Users' interaction with World Wide Web resources: An exploratory study employing a holistic approach. *Information Processing & Management, 36 (2), 229-251.*

Webometrics Website (2016).

Wildemuth, B. M (2004). The effect of domain knowledge on search tactic formulation. *Journal of the American Society for Information Science and Technology, 55 (3), 246-258*.

Wilson, T. D (2000). Human information behaviour. *Informing Science, 3(2), 49-56*.

Wohlin, C., Runeson, P., Höst, M., Ohlsson, M. C., Regnell, B., & Wesslén, A. (2012). Systematic Literature Reviews. *Experimentation in Software Engineering, 45-54*.

Wolfram, D.; Xie, H (2002). Traditional IR for Web users: A context for general audience digital libraries. *Information Processing & Management, 38 (5), 627-648*.

X. Shen., B.Tan, & C. Zhai, C (2002). Ucair: Capturing and exploiting context for personalized search. In Proceedings of the Information Retrieval in Context Workshop, SIGIR IRiX'05, 2005.

Xie, H (2000). Shifts of interactive intentions and information-seeking strategies in interactive information retrieval. *Journal of the American Society for Information Science, 51 (9), 841-857.*

Xie, H (2002). Patterns between interactive intentions and information-seeking strategies. *Information Processing & Management, 38 (1), 55-77.*

Xie, H. Understanding human-work domain interaction: Implications for the design of a corporate digital library. *Journal of the American Society for Information Science and Technology, 57 (1), 128-143.*

Xie, I. Interactive information retrieval in digital environments; IGI Global Inc.: Hershey, Pennsylvania, 2008.

Xing, W & Ghorbani, A. (2004). *"Weighted PageRank algorithm".* Proceedings of the Second Annual Conference on Communication Networks and Services Research 21-24 May 2004. Fredericton, Canada.

Y. Hu, G. Xin, R. Song, G. Hu, S. Shi, Y. Cao, & H. Li, H. (2005). *Title extraction from bodies of html documents and its application to web page retrieval.* In Proceedings of the ACM Conference on Information Retrieval (SIGIR '05), pp. 250-257.

Zareh Bidoki A. and Yazdani N. (2007). "DistanceRank: An intelligent ranking algorithm for web pages". *Information Processing and Management* (2007), doi:10.1016/j.ipm.2007.06.004