

Research on Mechanism and Challenges in Meta Search Engines

Jyoti Mor1, Naresh Kumar, Dinesh Rai

Abstract— A Meta Search Engine (MSE) produces results gathered from other search engine (SE) on a given query. In brief MSEs have single interface corresponding to multiple searches. MSE employs their own algorithm to display search results. This paper reviews existing Meta Search Engines like Yippy, eTools.ch, Carrot2, qksearch and iBoogie commonly used for searching. This paper surveys and analysed the working of different result merging algorithms. Current research reviews MSE based on different approaches like clustering technique. Few MSEs are employing Neural networks for searching. Further it also discusses problem in existing MSEs.

Keywords. Search Engine, Meta Search Engine, Web page, Clustering

I. INTRODUCTION

Meta Search Engines are searching tool (as shown in Figure 1.1) which makes use of different Search Engine to produce resourceful and suitable results corresponding to a user query. The exposure of popular Search Engine e.g. Google, Alta Vista and Yahoo! is low and results are also low in terms of accuracy, thus Meta Search Engines are introduced. Meta Search Engines submits query from the user to other Search Engines such as , Yahoo!, Bing, Google, and Alta vista. Outcomes from the Search Engines are pooled to form clusters which are presented to the user as results [1]. Advantage of using Meta Search Engines is that there is an magnification of coverage area. Meta Search Engines employs indexing of other Search Engines to uphold search outcomes [1], form clusters and provide results to users in a modern way with appropriate labels [2]. Sometimes Search Engines provide more relevant than that of Meta Search Engines. Meta Search Engines cannot deduce query syntax completely and further do not interconnect with bulky search engines for their result development.

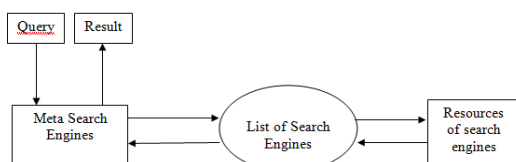


Figure 1.1 General Architecture of Meta Search Engine

Revised Manuscript Received on July 18, 2019.

Jyoti Mor1, Phd. Research Scholar, Computer Science, School of Engineering and Technology, Ansal University, Gurugram, India. (Email: jyotimor@ansaluniversity.edu.in)

Dr. Naresh Kumar, Associate Professor, Dept. of Computer Science & Engineering, MSIT, New De, lhi, India. (Email: 2narsumsaini@gmail.com)

Dr. Dinesh Rai, Associate Professor, School of Engineering and Technology, Ansal University, Gurugram, India (Email: dineshrai@ansaluniversity.edu.in)

II. LITERATURE SURVEY

Literature survey has been conducted for past 10 years for which details are given below:

1) The Global KE Algorithm analyzed [3] the results mined from combination of Weighted KE Algorithm, Geo KE Algorithm, and URL Aware KE Algorithm. Google, Bing, Yahoo and Ask Search Engines are used.

The key features of suggested system are:

- i. By passing on distributed user, different values of G are taken into account for ranking.
- ii. End Users can limit similar domain pages by transient over the Domain Awareness constant (D) to extracted results.
- iii. User be able to instruct to Search Engines so an enhanced process.
- iv. It makes use of area specific information for extracting results and also delivers language familiar web pages.

On the basis of user profile suggested algorithm leads to superior ranking.

2) Raghuvashi K.P. analyzed the influence of optimized keywords system on efficiency of Meta Search Engines. It confers detailed process of keywords optimization for Meta Search Engine (MSE). Such keywords centered Meta Search Engine helps users to retrieve results.

3) Vijaya P. , Raju G. , Ray S.K. , proposed Meta Search Engine using ontology & Semantic Similarity Measure (SSM) [5]. SSM is developed to take semantic in keyword matching. In this MSE using SSM Concepts Sets matched with Title Sets. With three different ranking measures : relevancy to contents, title sets and ranking value are combined to improve the effectiveness of MSE. Performance of the MSE is evaluated using TREC-style average precision (TSAP) by using different sets of queries. The proposed semantic Meta-Search Engine delivers 80% TSAP which shows better results as compared to existing search engine and Meta-Search Engine.

4) Usearch Meta Search Engine explains [6] 2 aspects of ranking (i) Similarity of queries taken & searched results description (ii) Weightage for extracted outcomes of each Search Engine depending on information requirements of user. Common and Non Common resultlist are displayed by its modules. Authors analyzed that top 20 results for merged list are always in uppermost 3 results for every Search Engine.

5) An approach recommended in [7] emphasizes on outcomes in Meta Search Engine based on scoring of results

place in Search Engines. Descending order ranking is given to the extracted results from total results. After considering the ratio of total scores of every result by the total sum of result in to each Search Engine final scores gets calculated. Further the results gets calculated in Descending order of the concluding scores. Authors had taken 3 Search Engines and 5 examine results for experimental work, and conclusions demonstrated were decent.

6) Joojadeh A. R. and Hassanpour H., proposed a novel field for Meta Search Engine i.e A Medical Meta Search Engine (MMSE) [8]. It was aimed for the users having no expertise of medical field. To increase the relevancy of the searches it is enhanced with Unified Medical Language System (UMLS) domain knowledge. Its efficiency is established on the understanding of the Web Pages semantics and queries of users. In such system the keyword quest is transformed into conceptual searching. MMSE employs programmed usage of medical field information to its users. The outcomes of research shows, increasing user queries with specific field increases the precision of results and the quantity of extracted Web Pages which are significant for users search.

7) The author in [9] describes fuzzy logic for web pages ranking. Firstly, the system checks number of participants in Search Engine whose result needs demonstration and secondly it also plots the tally to the respective Fuzzy values. Ultimate results are shown just before the user in lessening fashion of fuzzy values.

8) A Meta Search Engine termed “iral” proposed in [10] to offer ranking of extracted results using SEO parameters. Advanced version of iral algorithm assign corresponding weights to SEO Parameters to calculate rank parameter value. Page rank computed taking product of weights & values of SEO parameters. Further it verifies results then determines accuracy of approx. 0.48% in view of the keywords “alcoholism” and “local computer shop” on data of 300 students.

III. EXISTING META SEARCH ENGINE

This section includes existing Meta Search Engines like Yippy, eTools.ch, Carrot2, qksearch and iBoogie which are very generally used for searching. It includes only those Meta Search Engines which are on the basis of clustering technique and challenges in these clusters are based on the understanding, experience, and problems faced till yet.

Yippy (also known as Clusty previously) is developed by Vivisimo, it is a research based Meta Search Engine [11]. Search Engines used are Gigablast, Live, Ask.com, ODP, Shopzilla, NY Times, Yahoo stock and Yahoo news.

Carrot2 Meta Search Engine : Carrot 2 is employed for mining of text and employs Clustering of Meta Search Engine Softwares. It is an open source software. It practices Lingo Clustering technique and K Means Clustering Technique [12]. Further the results are mined from eTools, SOLR, Bing, Lucene, Google Meta Search Engines, etc. Relevancy of retrieved results is good.

iBoogie: [13] This Meta Search Engine is based on document clustering. It uses complete web and MSN for its result generation. Also For searching it uses customizable tabs.

OpenText :[14] It increases searching magnitude and

improves user throughput . Also it accomplishes collection of outcomes dependent on the content or site. It works with Bing, Yahoo, Google, Wikipedia, Open Directory and Ask Search Engines.

qksearch Meta Search Engine achieves clustering search, blind search and split search [15].

eTools.ch Meta Search Engine highlight on refining the extracted results by making appropriate clusters and propose full confidentiality [16]. Google, Search, Ask, Exalead, Fastbot, Wikipedia, Tiger, Yahoo, BaseBing, Yandex, DuckDuckGo, FarooGoo, Webliste, DeuSu, and Moose Search Engines are used.

IV. CLUSTERING ANALYSIS IN META SEARCH ENGINES

Based on literature survey explained in section 2 following problems have been identified for Existing MSEs. Factors which mark the usefulness of a Meta Search Engine exists [17]:

i. Count of Searching Engines ii. Relevancy of the results iii. Clustering iv. Ranking v. Re-Ranking.vi. Dynamic clustering.

Identical group formed from similar objects from a set of objects which is called as Cluster Analysis [18]. Such indistinguishable grouping of resembling data is termed clusters. Cluster Labels have been employed as data mining technique to get results collected in associated clusters through which end users can classify their desired data [19]. For nominal clustering there should be least Intra Cluster Distance and extreme Inter Cluster Distance to accomplish.

Clustering Techniques are categorized as [20] - Partitioning Method, Density Based Method, Constraint Based Method Grid Based Method, Hierarchical Method and Model Based Method.

V. CHALLENGES ASSOCIATED WITH META SEARCH ENGINES & RESULTS

The following challenges are experienced by the different authors while literature studies of existing Meta Search Engines:

i. Precise and Genuine identification of user perspective and user manners is required for formulating of framework as proposed in is done on the basis of available information. Questionnaire survey is adopted to get the accurate information. Thus to organize information to match a particular context is a crucial problem of [21].

ii. In fuzzy clustering algorithm defining of number of clusters by the user is a bad practice because user is not having any approximation of retrieved results and the count of clusters formed from them [2].

iii. In [22], DDC technique is not suitable for datasets of single dimension and leads to slow processing of results.

iv. The major challenges of [20] are as:

- Clustering can lead to the formation of inappropriate clusters if the understanding of semantic relationship among the web documents is not used properly.

- To define the number of clusters System is dependent on the user which creates problem as user is not aware of clusters to be formed to include all the search results.
- The proposed method digs out terms from Title Tags, URL, and Meta Tags to put together the results into clusters. It may not cover the entire keywords essential to explain a web page entirely.
- The proposed algorithm used stemming in its steps which usually leads to the creation of insignificant words.
 - v. Group clustering stage in [23] leads overlapping of groups in a cluster which affects the uniqueness of a cluster.
 - vi. Topic generation module in [24] consider that starting and finishing keywords are more important than the information available in core of any document but it is not always true. Moreover, it does not support Ontology and web 2.0.
 - vii. Though ITSE in [25] there is improvement in the exactness of the clusters formed, but in comparison to other search engines it is taking more time means it is not efficient for high dimensional data sets.
 - viii. Extraction method in [25] finds appropriate documents depending on the retrieval of URLs. It may not include all the appropriate documents as per user requirements. Also it does not satisfy the uniqueness and accord of clusters.
 - ix. Overlapping of clusters generated by [11] means web page availability in more than one cluster. It affects the uniqueness of a cluster. Further ranking of clusters has not been done efficiently which leads to low relevancy of results and contains too several sponsored Web Pages.
 - x. A search result in [11] present in multiple cluster leads to increased redundancy.
 - xi. Literature [12] describes overlapping clusters which results in low relevancy of results.
 - xii. A cluster created by [13] containing multiple documents makes it less memory efficient. Also such clusters lack in relevancy of results. Therefore, [27] fails to create high class and clean clusters.
 - xiii. The major issue of [14] is that it displays the same message "Our server could not serve your request. Please try again later." for every query keyword. It is wastage of search time.
 - xiv. A cluster containing the similar web page more than once causes the creation of superfluous and unrelated results by [15]. Also labels of some clusters are not self-descriptive.

VI. CONCLUSION

Meta Search Engines employed to deal enormous data or domain specific data. It is very complex to build a Meta Search Engine demonstrating entirely effective outcomes. The key objective stays to reduce clustering concerns & identify problem areas which provide possible results to improvise clustering. Efficient Meta Search Engine based on clustering can be developed to show results better than existing ones. The above challenges representing the research gaps that can be classified into Presentation of results, Relevancy of results and Linking issues which can form the basis of future research.

REFERENCES

1. Kumar N., Nath R., "A Meta Search Engine Approach for Organizing Web Search Results using Ranking and Clustering", International Journal of Computer (IJC), ISSN 2307-4531.
2. Portmann E., "Weblog Extraction with Fuzzy Classification Methods," in Second International Conference on the Applications of Digital Information and Web Technologies, pp. 411-416, ISSN: 978-1-4244-4457-1, 2009. DOI:10.1109/ICADIWT.2009.5273930.
3. Akritidis L., Katsaros D. and Bozani P., "Effective Ranking Fusion Methods for Personalized Meta Search Engines", DOI: 10.1109/PCI.2008.31.
4. Raghuvashi K.P., "Empirical Study on the Meta- Search Engine Optimization Technique Based on Keyword: A Review" in "IBMRD's Journal of Management and Research, Print ISSN: 2277-7830, Online ISSN: 2348-5922, Volume-3, Issue-2, September 2014 pp 51-56
5. Vijaya P., Raju G., Ray S.K., "S-MSE: Asemantic Meta search engine using semantic similarity and reputation measure" in "Journal of Theoretical and Applied Information Technology" 2014/03/01 222- 230
6. Joojadeh A. R. and Hassanpour H., " MMSE: Design & Implementation of a Medical Meta Search Engine" ,in International Journal of Machine Learning and Computing, Vol. 2, No. 4, August 2012, pp 453-457
7. Alloui T., Boussebough I., Chaoui A., Nouar A. Z. and Chettah M. C., "Usearch: A Meta Search Engine based on a New Result Merging Strategy", In 7th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management (IC3K), 2015 URL: <http://ieeexplore.ieee.org/document/7526966/>.
8. Patel B., Shah D. and Patel J., "Fuzzy logic implementation for ranking of search results in meta-search engine", at Patel et al., International Journal of Advanced Engineering Research and Studies, E-ISSN2249-8974, Int. J. Adv. Engg. Res. Studies/III/IV/July- Sept., 2014.
9. Manral J., Hossain M. A., "An Innovative Approach for online Meta Search Engine Optimization", In 6th Conference on Software, Knowledge, Information Management and Applications, Chengdu, China, URL:<https://arxiv.org/abs/1509.08396>, September 9-11 2012.
10. The yippy website [Online]. Available: <http://www.yippy.com/> , Last Accessed 23 January 2018 at 21:29(IST).
11. The carrot2 website [Online]. Available: <http://search.carrot2.org/stable/search> , Last Accessed 21 January 2018 at 22:17(IST).
12. The iboogie website [Online]. Available: <http://www.iboogie.com/> , Last Accessed 22 January 2018 at 19:12(IST).
13. The opentext website [Online]. Available: <http://www.opentext.com/> , Last Accessed 23 January 2018 at 21:18(IST).
14. The qksearch website [Online]. Available: <http://qksearch.com/> , Last Accessed 23 January 2018 at 21:16(IST).
15. The etools website [Online]. Available: <https://www.etools.ch/> , Last Accessed 23 January 2018 at 21:21(IST).
16. Ball S. and Nath R., "To Evaluate the Performance of Meta Search Engines: A Comparative Study," Journal of Technology and Engineering Sciences, vol.1, issue1, pp.70-78, 2009.

17. Kumar N., Nath R. and Kherwa P., "An Automated Framework based on Three Level Scoring Method to Choose Best Search Engine in a Particular," International Journal Of Computer Applications, vol. 83, issue 14, pp. 42-48, ISSN 0975 – 8887, 2013.
18. Poomagal S. and Hamsapriya T., "K-means for Search Results clustering using URL and Tag contents," International Conference on Process Automation, Control and Computing, pp. 1-7, ISSN: 978-1-61284-764-1, 2011. DOI:10.1109/PACC.2011.5978906.
19. Bijuraj L.V., "Clustering and its Applications," In Proceedings of National Conference on New Horizons in IT NCNHIT, pp. 169-172, ISBN: 978-93-82338-79-6, 2013.
20. Kadir N. A., Lokman A. M. and A. Ahmad, "Dominant User Context (DUC) Filtering Framework for Web Personalized Search," In IEEE Symposium on Wireless Technology and Applications, pp. 280-285, ISSN: 978-1-4673-2210-2, 2012. DOI: 10.1109/ISWTA.2012.6373862.
21. Juasiripukdee P., Wiyartanti L. and Kim L., "Clustering Search Results of Non-text User Generated Content," In: Digital Information Management, Fifth International Conference, pp. 95-100, ISSN: 978-1-4244-7571-1, 2010. DOI: 10.1109/ICDIM.2010.5663293.
22. Tsai C.W., Huang K., Chiang M. C., and Yang C. S., "A Fast Tree-Based Search Algorithm for Cluster Search Engine," In: Proceedings of the 2009 IEEE International Conference on Systems, Man, and Cybernetics San Antonio, TX, USA, pp. 1603-1608, ISSN: 978-1-4244-2794-9, 2009. DOI: 10.1109/ICSMC.2009.5346100.
23. Li H. J. and Wang J. K., "Precise Image Retrieval on the Web with a Clustering and Results Optimization," In International Conference on Wavelet Analysis and Pattern Recognition, Beijing, Vol. 1, pp.188-193, ISSN: 1-4244-1066-5, 2007. DOI:10.1109/ICWAPR.2007.4420661.
24. Campos R., Dias G. and Nunes C., "WISE: Hierarchical Soft Clustering of Web Page Search Results based on Web Content Mining Techniques," In Proceedings of the 2006 IEEE/WIC/ACM International Conference on Web Intelligence, pp. 301-304, ISSN: 0-7695-2747-7, 2006. DOI: 10.1109/WI.2006.201.
25. Vijaya P., Raju G., Ray Santosh Kumar, "S-MSE: Asemantic Meta Search Engine Using Semantic Similarity and Reputation Measure", Journal of Theoretical and Applied Information Technology, 20th February 2014. Vol. 60 No.2, ISSN: 1992-8645.
26. Haveliwala T. H., Gionis A., Indyk P. , "Scalable Techniques for Clustering the Web", In: Proceeding of Web DB Workshop 2000, <http://www.stanford.edu/~taherh/papers/scalable>.