A Brief Study on Query Searching Techniques in Biomedical Databases and Clinical Information Retrieval Systems

Dr. Maria Lourdes Santos[1]*
[1]University of the Philippines Manila, Department of Medical Informatics and Health Information Systems, Manila, Philippines

ABSTRACT

Query mining and query processing were the twomain research domains that needed customization. Every business analyst needs to identify the growth potential of business products. So, they depended on querying techniques and tended to get the connected data including query-based summarization, query recommendation, facet mining and so on. These techniques had some merits and demerits. Facet mining was a recent technique in query searching. However, it has a lot of challenges in the web community. According to this paper, a comparative study of query mining techniques and facet mining had made to enable the researcher to work on it.

KEYWORDS: Facet mining, Query reformulation, Query recommendation, Query summarization.

1. INTRODUCTION

Query mining has a vital role in business individuals. They depend on queries and query related process for exploring the information. So, the analysts come across the mining problem during query mining. Mining problem generates weak and invalid results. Hence, the business community explores the useful information. To minimize the time and to extract efficient and accurate results, techniques such as query reformulation, query recommendation, query summarisation, facet mining are available in the literature. Query-based text summarisation is aimed to extract essential information that answers the question from the original text. The answer is conferred in a very marginal, often predefined, range of words. The computer program users are not comfortable with the result of an input query; query recommendation suggests connected queries for them. Query reformulation is a method of reformulating a question to improve retrieval performance in information retrieval.

Figure:

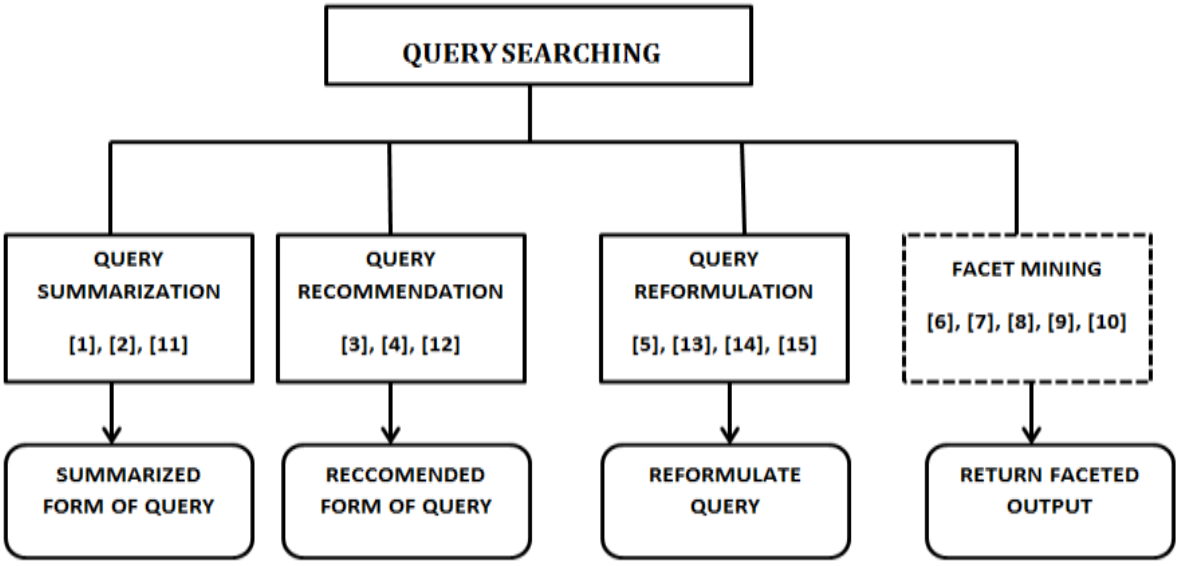

Fig: Classification of Query Searching Techniques

These techniques have some merits and demerits. Facet mining is an emerging technique to solve existing query mining problems. A query facet is a set of things that describe and summarise complete information about a query. Here a facet item is often termed as a word or a phrase. A facet mining technique summarizes the knowledge and provides a list of information according to the queries. This literature bestows with query mining techniques, facet mining and its importance, the emerging trend in the society to optimize query mining and discussion of its application. The figure below shows the entire topic discussed in this literature survey.

2. LITERATURE REVIEW

Query-Based summarisation

A lengthier searching process will result in wastage of time. An efficient way to get the output for a query searching is text summarisation. Text summarisation [1] helps the analyst to read the entire document in a summarized form. A comprehensive survey on text summarisation system has done, and its categorization is also specified here [1]. Query-based summarization techniques and its applications on different fields such as e-commerce, education etc. are specified in the paper [2]. It also portraits the idea of making e-books more intelligent by using text-based summarisation. Fastest retrieval of query related information is an emerging area in this field. Here the researchers explore approaches for a single document and multi-document summarisation [11], [12]. Moreover knowledge-based and machine learning methods are used for choosing the most relevant sentences from documents for a particular query and tailored summarisation techniques for medical texts [2], [13], [14], [15].

Query Recommendation

Query recommendation is a process by which search engine listout some queries based on the past search experience of the users. As a result, the user can navigate to the valid resources and explore thereby improve the search accuracy. The most powerful technique used for query recommendation is the clustering process. In the clustering process [3] semantically similar queries are identified, and preferences are given to these queries based on search engine log history. Then it is finally ranked based on the relevance. The sequential search behavior of search engine users is studied and analyzed and then use this analysis for a query introducedby a client in [4]. This method is also combined with traditional content-based similarity model to get more efficient results.

Query Reformulation

Query reformulation is a technology for modifying a query. Every user modifies queries during the time of information exploration to get an accurate result. If a user is not satisfied with the currently entered query, the query reformulation technique modifies the query and explores information on behalf of users. A rule-based classifier to detect such type of reformulation methods are portrayed in [5], [16].

Facet Mining

Facet mining is a technique in which the facet can provide output for a query in different aspects. It is fast and time efficient as compared to the existing techniques [22], [23] such as query reformulation, query recommendation, text summarisation and so on. The current facet mining techniques are confined to specific domains such as Wikipedia, e-commerce applications and so on. So to improve the searching experience more efficient a proper facet mining technique is needed [17], [18], [19].

Audios and videos are the two methods for easier communication. Spoken web [8] is similar to the World Wide Web. But the only difference is that, in information retrieval, exchange of ideas and explores information through voice. Searching queries on spoken web results in many problems such as delay in information retrieval inconvenience during query processing etc. [20], [21]. To overcome these problems, indexing the metadata associated with the audio content with the help of facet. Finally, rank the facets and develop an interactive query interface for easy browsing of search results.

Facetopedia is a facet retrieval system in which the information exploration and related searches are confined to a database called Wikipedia. The facetopedia generates faceted result with the help of facets. So the user can easily explore information about the query. Here they propose metrics for ranking individual faceted interfaces by the navigational cost of users. Here [6] facet generation and hierarchy construction are done automatically. It also recommends metrics for ranking individual faceted interface based on navigational cost, average pairwise similarities. For achieving this, a faceted interface discovery algorithm is used to optimize the ranking matrices. A dynamic faceted search for analyzing the data with textual content and structured attributes are proposed in [7]. Navigational expectations and new interestingness measures are the new two contributions by [7]. A user-defined expectation value can be set based on the user navigation result is described as navigational expectation. Surprising aggregated information to users based on their expectation is called interestingness. Single facet instance and whole facet instance are the two interesting measures proposed here. The widely used query interface is facets. Information contains structured and unstructured data. Most of the enterprises want to work on data analytics to extract valuable structured information.

Most of the facets related searches are confined to a specific domain. A new technique called QD Miner is introduced in [10]. It applies to all domains and is query dependent. As compared to text summarisation, query reformulation and query recommendation, facets are the best way to explore useful information. Facet related researches have some disadvantages such as they are query independent and applicable only for some specific

domain [24],[25], [26], [27]. So QDMiner is an efficient way to retrieve and explore information for query searches in search engines in the fastest way. The table below shows the merits and demerits of some query searching techniques as compared to facet mining.

Table:

Table 1: Merits and Demerits of Query Searching Techniques

SL. NO	METHOD	MERIT	DEMERIT
1	Text summarisation and query based summarisation	An efficient way to retrieve the output of a whole document in summarised form	Unable to gather the required information
2	QueryRecommended-tion	Helps to suggest new query terms	Cannot offer complete information regarding an input query, Provides rec ommendation only
3	Query reformulation	Helps in modifying the query	Cannot provide the valuable information to the user
4	Facet mining	Helps gather data to in an exceedingly re-stricted time, To re-trieve data as a faceted output	Confined to specific domains such as Wikipedia, E-commerce, spoken web etc.
5	Advanced facet mining	Efficient way to retrieve and explore information for query searches in search engines	Applicable to all domain

3. SCOPE AND FUTURE WORK

Facet mining plays a vital role in e-commerce applications. Its existing works were not much efficient because mining was confined to restricted domains like Wikipedia, spoken internet etc. As a future work facetmining [28], [29] will be extended to a wide range of fields such as medical, banking, weather forecast etc. It will be applied among these domains to improve the searching more efficient. Facet mining will have been extended to findthe list duplication during query processing.

4. CONCLUSION

Every query retrieval process depended on time. So the business analysts relied upon time efficient works. If a query processing method had taken a massive amount of time for the completion of the work, it would be an inefficient one. There were many techniques available for query searching which encompass text summarisation, query recommendation, query reformation and so on. These techniques had some merits and demerits. So the best way to improve the query searching process will be facet mining. Facet could provide the output in a listed form covering all the information related to the input query. It was time efficient as compared to other query retrieval methods. So it will be used for faster query searching among broader domain.

5. ACKNOWLEDGEMENT

The authors gratefully thank to Computer Science department, CEMP, Punnapra, Kerala for providing the facilities in conducting this research.

REFERENCES

[1] S. Gholmrezazadeh, M. A. Salehi and B. Gholamzadheh,"A comprehensive survey on text summarization systems", in proceedings of CSA'09,2009,pp.1-6.
[2] M. Damova and I. koychev,"Query based summarization: A survey", in s3T'10,2010.
[3] R. Baeza - Yates, C. Hartado and M. Mendoza, "Query recommendation using query logs in search engines", in proceedings of EDBT'04,2004,pp.588-596.
[4] Z. Zhang and O.Nasraoui,"Mining search engine query logs for query recommendation", in proceedings of www'06,2006.
[5] J. Huang and E.N. Ethimiadis, "Analyzing and evaluating query reformulation strategies in web search logs", in proceedings of CIKM.Newyork,NY,USA:ACM,2009,pp.77-86.
[6] C. Li, N. Yan, S.B. Roy, L. Lisham and G. Das,"Facetedpedia: dynamic generation of query dependent faceted interfaces for Wikipedia", in proceedings of www'10.ACM,2010.
[7] D. Dash, J .Rao, N. Megiddo, A. Ailamaki, and G .Lohman, "Dynamic faceted search for discovery driven analysis", in CIKM'08,2008.
[8] M. Diao, S. Mukherjea, N. Rajput, and K. Srivastava, "Faceted search and browsing of audio content on spoken web", in proceedings of CIKM'10,2010.
[9] O. Ben-Yitzhak, N. Golbandi, N. Har'El, R. Lempel, A. Neumann, S. Ofek-koifman, "Beyond basic faceted search", in proceedings of WSDM'08,2008.
[10] Zhicheng Dou, Zhenghaoj. Lang, Shattu, "Automatically mining facets for queries from their search results", 2015.
[11] Hovy Eduard, Lin Chin Yew, "Automated text summarization in SUMMARIST" in , MIT Press, pp. 81-94, 1999.
[12] I Mani, "Automatic Summarization" in , John Benjamin's Publishing Co., pp. 1-22, 2001.
[13] Karel Jezek, Josef Steinberger, "Automatic Text Summarization (the state of the art 2007 and new challenges)", Znalosti 2008, pp. 1-12.
[14] Al-sanie Waleed, "Towards an infrastructure for Arabic text summarization using rhetorical structure theory", 2005.
[15] S. Horacio, L. Guy, "Generating indicative-informative summaries with SumUM: Summarization", Computational linguistics - Association for Computational Linguistics, vol. 28, pp. 497-526, 2002.
[16] R. Baeza-Yates. Query usage mining in search engines. Web Mining: Applications andTechniques, Anthony Scime, editor. Idea Group, 2004.
[17] R. Baeza-Yates and B. Ribeiro-Neto. Modern Information Retrieval, chapter 3, pages 75–79.Addison-Wesley, 1999.
[18] D. Beeferman and A. Berger. Agglomerative clustering of a search engine query log. InKDD, pages 407–416, Boston, MA USA, 2000.
[19] B. M. Fonseca, P. B Golgher, E. S. De Moura, and N. Ziviani. Using association rulesto discovery search engines related queries. In First Latin American Web Congress (LA-WEB'03), November, 2003. Santiago, Chile.
[20] M. Jansen, A. Spink, J. Bateman, and T. Saracevic. Real life information retrieval: a studyof user queries on the web. ACM SIGIR Forum, 32(1):5-17, 1998.
[21] Agichtein, E., Brill, E., and Dumais, S. (2006). Improving web search ranking by incorporating user behavior informa-tion. In SIGIR '06, 19-26.
[22] Anick, P. (2003). Using terminological feedback for web search refinement: a log-based study. In SIGIR '03, 88-95.
[23] Baeza-Yates, R., Calderón-Benavides, L., and González-Caro, C. (2006). The intention behind Web queries. In SPIRE '06, 98-109.
[24] O. Ben-Yitzhak, N. Golbandi, N. Har'El, R. Lempel, A. Neumann, S. Ofek-Koifman, D. Sheinwald, E. Shekita, B. Sznajder, S. Yogev, "Beyond basic faceted search", Proc. Int. Conf. Web Search Data Mining, pp. 33-44, 2008.
[25] M. Diao, S. Mukherjea, N. Rajput, "Faceted search and browsing of audio content on spoken web", Proc. 19th ACM Int. Conf. Inf. Knowl. Manage., pp. 1029-1038, 2010.
[26] D. Dash, J. Rao, N. Megiddo, A. Ailamaki, G. Lohman, "Dynamic faceted search for discovery-driven analysis", ACM Int. Conf. Inf. Knowl. Manage., 2008.
[27] W. Kong, J. Allan, "Extending faceted search to the general web", Proc. ACM Int. Conf. Inf. Knowl. Manage., pp. 839-848, 2014.
[28] Duhita Pawar, Vina M. Lomte ,"A Survey on Automatically Mining Facets for Web Queries ", 2017.
[29] Shraddha Dinanath Londhe,Pallavi Dhande,Dr.Rajendra Gode , "Review Paper on Implementation of Mining Facets Automation for the Searched Queries" 2018.