INTERNATIONAL JOURNAL OF ADVANCE RESEARCH, IDEAS AND INNOVATIONS IN TECHNOLOGY

ISSN: 2454-132X
Impact factor: 4.295
(Volume 4, Issue 3)
Available online at: www.ijariit.com

Automatically mining query facet from search results using text mining algorithm

Soniya Joy
soniyajoy15395@gmail.com
Mangalam College of Engineering, Ettumanoor, Kerala

Neena Joseph
neena.joseph@mangalam.in
Mangalam College of Engineering, Ettumanoor, Kerala

ABSTRACT

A query facet can be considered as a single word or multiple words which summarize and describe that query. Query facets may provide direct information that users are seeking. The existing algorithms for generating query facet can be used by extracting the frequent list in search results. The coverage of facet item must be limited because the only small number of search results can be used. In order to solve this kind of problem in the proposed system uses the format of the list is more user-friendly. Query facet is analyzing the text query the query facet provides useful knowledge about a query. In existing algorithms are used the coverage of facet item must be limited in order to solve this kind of problem propose an algorithm text mining and use the knowledge base to improve the quality and the coverage of facet item. Text Mining algorithms are used to extract the relevant information from available text.

Keywords— *Query facets, Text mining, Multi-faceted queries, Knowledgebase*

1. INTRODUCTION

Query facet is a collection of items which summarized the content of a query. In conventional method the user can browse a webpage user can view many documents for the information they are seeking, this takes a lot of time and confused the user [6]. Here use an automatic summarization of search result will produce it will help the user to know about the query they are searching without browsing many web pages. Mining query facets is an approach to solve the above-explained problem using text mining algorithms to mine the query facet. Table 1 shows an example of query facet the query is "Beijing subway," is a place in a European country. Its query facets cover aspects of related country lines temple, important city etc. These query facets help users learn about the topic "Beijing subway" without browsing so many web pages.

Query facets are good summaries of a query and are useful for users to understand the query and help them explore information [1]. Existing algorithm like QD Miner, QF-I, QF-J has used automatically mine query facets by aggregating frequent lists contained in the results. The facet item is extracted as a top search result from a search engine. One problem can arise by using this kind of methods the coverage of facet mined can be limited [6].

Table 1. Example of query facet

Query	Beijing subway
1	line 1, line 2, line 4, line 5, line 10, line 13, batong line
2	xizhimen, jianguomen, dongzhimen chongwenmen
3	forbidden city, the temple of heaven, Tiananmen square

To solve this problem use a knowledge base as a data source to improve the quality of query facet. Knowledgebase contains structured information such as entities and properties of the related query [6]. A text mining algorithm can be used to mine the query facet. Text mining is also known as text analytics, is the process of deriving high-quality information from text. Text mining is a process to extract interesting and significant patterns to explore knowledge from textual data sources [4]. Text mining is a multi-disciplinary field based on information retrieval, data mining, machine learning, statistics, and computational linguistics [4].

Fig. 1: Overview of proposed system

Figure 1 shows the overview of the proposed system. The user can search for a keyword by using the system. Then the URL of the search result is retrieved from the web and finally view the summarised search result and the user can download the search result. There are two methods are used to construct the final facet namely Facet Generation and Facet Expansion.

2. RELATED WORKS

Nowadays, search engines like Google have evolved to include in their results information from structured data sources along with text documents. These search engines provide a keyword

search capability to their user. But, users are mainly interested in exploring a structured collection of information than a query for a specific item. A commonly used interaction for structured information is faceted search. Faceted search provides a more user-friendly visual alternative to keywords for the user to explore the structured results. Faceted search is a technique for accessing information organized according to a faceted classification system [3].

The search engine is an important tool for the user to search for information. Query facet a set of items which summarise the important aspects of a query. Query facets may provide direct information that users are seeking. Direct access to digital information has completely changed the rules, users can directly browse the information, without consulting any complex systems. The spread of digital access to information has been a sudden increase in the volume of information available about any given query. Many websites now provide even more refined tools to help users find information. Filters are one such tool to find information [1].

Search queries are multi-faceted, which makes a simple ranked list of results. To finding information for such faceted queries, browse a technique that explicitly represents the facets of a query using groups of semantically related terms. Query facets can help users to find topics of the search results by applying multiple faceted. Construct a supervised method based on a graphical model for query facet extraction. The graphical model learns how a candidate term is to be a facet term as well as how likely two terms are to be grouped together in a query facet and captures the dependencies between the two factors [7].

3. PROPOSED SYSTEM

Search engines currently have become the vital tools for web users to locate information [1]. A knowledge base use as a data source to improve the quality of query facets [6]. Knowledge bases hold numerous prominent organized majority of the data for, such as their properties. In the proposed system, the user can search a keyword the search result are retrieved from the web. Then check the URL of search keywords if the URL is valid then extract the URL and then extract the search result. Apply the facet generation and facet expansion method to construct the final facet.The facet candidates are constructed by facet generation and expansion are further merged, because there might be duplicate items within these candidates. Then apply the facet grouping and facet weighting and finally produce the final facet. Our focus will be the system can made more effective by improving the recall of facet items by using entities and their properties of the query and at increase the accuracy of facet item. There are two methods are used, facet generation and facet expansion to produce the final facet. In facet generation, straightly utilize the properties about entities relating with a query. In facet expansion, expand starting facets mined by using existing algorithm. Those facets constructed using this two techniques would further consolidated and positioned should produce last query facets. Facet grouping can done in the system is all the facet candidates constructed by facet generation and expansion might have duplicate entities cluster them into the final facets by grouping similar candidates together. Facet Weighting can be done to weight each final facet.

The text mining algorithm is used, Text mining is the method of extracting meaningful information or knowledge or patterns from the available text documents from various sources [5]. Text mining is also referred as text data mining deriving high-quality information from text. Text mining usually involves the process of structuring input text deriving patterns from structured data finally evaluate the output. Text mining has a higher economic value than data mining. Text mining tasks consist of three steps: text preprocessing, text mining operations, text post-processing. Text preprocessing includes data selection text categorization and feature extraction. Text mining operations are the core part of text mining that includes association rule discovery, text clustering, and pattern discovery. Post-processing tasks modify the data after text mining operations are completed such as selecting, evaluating and visualization of knowledge [2].

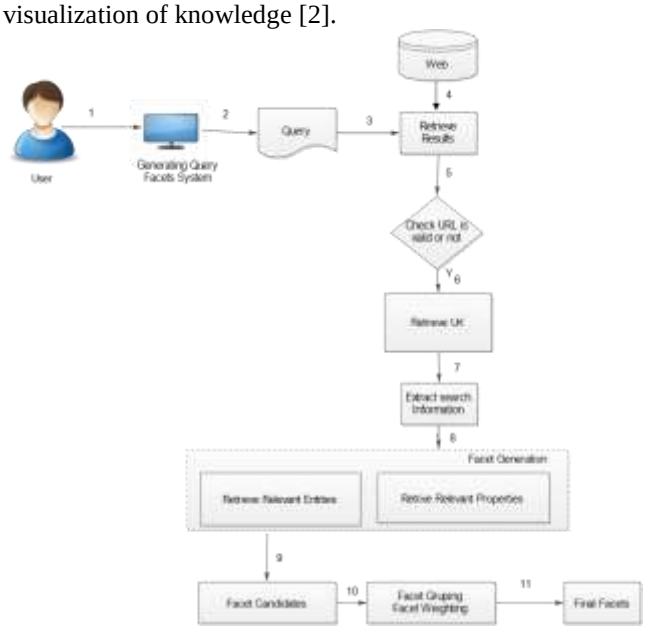

Fig. 2: System Architecture

Text Mining is finding unknown hidden information. The information extracted from different written resources is done automatically. Text mining is a process that employs a set of algorithms for converting unstructured text into structured data items. Generic process of text mining performs the following steps [4]. (Figure 3)

1. Collecting unstructured data from different sources available in different file formats such as plain text, Web pages, pdf files etc.
2. Pre-processing and cleansing operations are performed to detect and remove anomalies. The cleansing process makes sure to capture the real essence of text available and is performed to remove stop words stemming (the process of identifying the root of the certain word) and indexing the data.
3. Processing and controlling operations are applied to audit and further clean the data set by automatic processing.
4. Pattern analysis is implemented by Management Information System (MIS).
5. Information processed in the above steps are used to extract valuable and relevant information for effective and timely decision making and trend analysis.

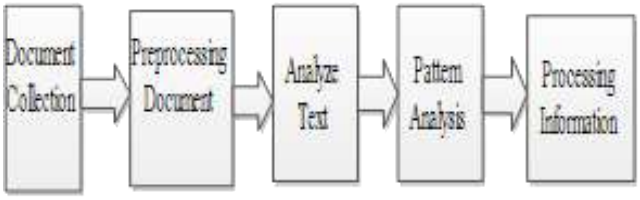

Fig. 3: Text mining process

Text mining deals with natural language text which is stored in semi-structured and unstructured format. The selection of an appropriate technique for mining text reduces the time and effort to find the relevant patterns for analysis and decision making [4].There are some basic text mining technologies they are Information Retrieval, Information Extraction, Categorization, Clustering, Summarization.

4. EXPERIMENTAL RESULTS

4.1 Comparison of QDMiner and Text Mining Algorithm

QDMiner extracts lists from free text, HTML tags, and repeat regions contained in the top search results, groups them into clusters based on the items they contain, then ranks the clusters and items based on how the lists and items appear in the top results [8]. To summarize the information contained in the query to find a list of related queries. QDMiner, to automatically mine query aspects by way of extracting and grouping common lists from loose textual content, HTML tags, and repeat areas inside top search effects [9]. The facets in QDMiner are generated using four essential phases such as List extraction, list weighting, list clustering and list ranking [9].

Text mining is extracting meaningful information from the available text document. Text mining usually involves the process of structuring the input text deriving patterns from structured data and finally evaluation and interpretation of the output. There are different tasks performed in text mining algorithm, Text categorization, Text clustering, Concept mining, Information retrieval, Information Extraction. .Text mining generally consists of the analysis of text documents by extracting key phrases, concepts, etc. and the preparation of the text processed in that manner for further analyses with numeric data mining techniques[10].

From the figure 4 shows the comparison of QDMiner algorithm i.e., the traditional method used and web data i.e., proposed algorithm it is text mining algorithm. Here we can clearly conclude that by using a traditional method different search keys are used the number of the search result can decrease. By using the proposed method the number of the search result can be increased rapidly. So by using the proposed method we can improve the system efficiency and at the same time improve the accuracy of facet item.

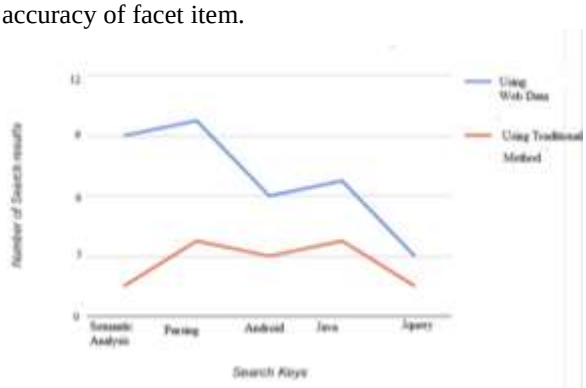

Fig. 4: Comparision Results of QDMiner and Text Mining Algorithm

5. CONCLUSION

In this paper, we can automatically generate query facet by using text mining algorithm. Query Facet is a set of items which describe and summarize one important aspect of a query. Text mining is extracting meaningful information from the available text document. The user can search for a keyword by using the system. Then the URL of the search result is retrieved from the web and finally view the summarised search result and the user can download the search result. There are two methods are used to construct the final facet namely Facet Generation and Facet Expansion.

In future, the system can be made more effective by improving the recall of facet items by utilizing entities and their properties contained in knowledge bases, and at the same time, make sure that the accuracy of facet items.

6. REFERENCES

[1] Survey on Query Facets Mining Approaches, Sheetal Sonwane, Nilam Patil.

[2] Feature Extraction and Duplicate Detection for Text Mining: A Survey, Ramya R S, Venugopal K R, Iyengar S S, Patnaik L M.

[3] Facet Discovery for Structured Web Search: A Query-log Mining Approach, Jeffrey Pound University of Waterloo Waterloo, Canada, Stelios Paparizos Microsoft Research Mountain View, CA, USA, Panayiotis Tsaparas Microsoft Research Mountain View, CA, USA.

[4] Text Mining: Techniques, Applications and Issues, Ramzan Talib, Muhammad Kashif Hanify, Shaeela Ayeshaz, and Fakeeha Fatimax, Department of Computer Science, Government College University, Faisalabad, Pakistan.

[5] A Review on Various Text Mining Techniques and Algorithms, R. Balamurugan, Dr. S. Pushpa.

[6] Generating Query Facets using Knowledge Bases, Zhengbao Jiang, Zhicheng Dou, Member, IEEE, and Ji-Rong Wen, Senior Member, IEEE.

[7] Extracting Query Facets from Search Results, Weize Kong and James Allan Center for Intelligent Information Retrieval School of Computer Science University of Massachusetts Amherst Amherst, MA 01003.

[8] Automatically Mining Facets for Queries from Their Search Results, Zhicheng Dou, Member, IEEE, Zhengbao Jiang, Sha Hu, Ji-Rong Wen, and Ruihua Song.

[9] Useful Query Facets Extracting Automatically from Top Retrieved Documents by Using QDMiner System, Bhagya Varsha S, Y. Sucharitha, Dr. D. Baswaraj, Dr.M.Janga Reddy.

[10] A tutorial review on Text Mining Algorithms, Mrs. Sayantani Ghosh, Mr. Sudipta Roy, and Prof. Samir K. Bandyopadhyay.

