ISSN: 2278-4632
Vol-10 Issue-3 No. 1 March 2020

A Study on Knowledge based Propagation for Query Facets

P. Swapna[1], T. Swapna[2]
[1,2]*Assistant Professor, Department of IT*
[1,2]*Malla Reddy Engineering College, Hyderabad, Telangana, India*

Abstract

Important features of a query are generally presented and repeated in the top retrieved documents in the style of lists. Query facets can be extracted by collecting these significant lists. Query facets may provide direct information or instant answers that users are seeking i.e. user can select a particular facet item which he found relevant to his search need. So, the list format style is much more user friendly than displaying searches sentence wise. The scope of this survey is limited to get search results of a query in list format i.e. facets. Previously there has been lot of work done for retrieving more relevant data to users in order to meet their information needs thus improving performance of search engines. Search engine provide the platform for users to describe their information need more clearly by using query facets mining. Different approaches for extraction of query facets from web search results to assist information finding for queries are discussed along with similar techniques used earlier for information retrieval of queries. Query facets represents interesting facets of a query using groups of semantically related terms extracted from search results. This paper reviews techniques that represents interesting facets of a query using groups of semantically related terms extracted from search results along with approaches that are similar to query facets mining.

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].

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

ISSN: 2278-4632
Vol-10 Issue-3 No. 1 March 2020

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]. The user can search for a keyword by using the system.

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

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. Query Facets Mining

2.1. Query Reformulation

The reformulation process is an iterative attempt between users and search engines in getting a satisfactory set of results [4]. Query reformulation is the process of iteratively modifying a query to improve the quality of a search engine results. Reformulations are close to the previous query both syntactically, as sequences of characters or terms, and semantically, regularly involving clear naming conventions [5]. Herdagdelen. et. al. [5] proposed an approach to query reformulation that provide a principled framework for the combination of string similarity and corpus-based semantic association measures using generalized Levenshtein distance algorithms. An exploration of class of models viz. unsupervised, compact and efficient for query reformulation which combines the syntactic and semantic aspects are given. Huang et.al. [4] studied users reformulation strategies in the context of the AOL query logs and describes the human side of query reformulation. A taxonomy of query reformulation strategies is created and built a high precision rule-based classifier to detect each type of reformulation.

2.2. Entity Search

Entity retrieval is the task of finding objects related to an information need [10]. Balog et. al. [10] explore the potential of combining Information Retrieval (IR) with Semantic Web (SW) technologies to improve the end-to- end performance on a specific entity search task. To get the best of both worlds, Balog et. al. [10] proposed to combine text- based entity models with semantic information from the Linked Open Data (LOD) cloud. Approaches to the REF task are described using IR and SW techniques and aim to find a set of entities for each topic. Semantic class construction tries to discover the peer or sibling relationship among terms or phrases by organizing them into semantic classes. Zhang et.al. [11] presented an approach that employs topic modelling for semantic class construction. Given a query q, all raw semantic classes (RASCs) are

ISSN: 2278-4632
Vol-10 Issue-3 No. 1 March 2020

retrieved containing the item to form a collection CR (q) where CR is collection of RASCs. Latent Dirichlet Allocation (LDA) model and probabilistic Latent Semantic Indexing Model (pLSI) are used to generate semantic classes. Offline processing is performed for CR (q) that contains RASCs and store the results on disk, in order to reduce the online query processing time.

2.3. Search Results Diversification

It has been studied as a method that deals with ambiguous or multi-faceted queries while a ranked list of documents remains the primary output feature of web search engine today. Search result diversification tries to diversify the ranked list to account for different search target or query subtopics. A weakness of search result diversification is that the query subtopics are hidden from the user, leaving him or her to guess at how the results are organized. Query facet extraction addresses this problem by explicitly presenting different facets of a queries using groups of coordinate terms. T. Sakai et.al. [15] compare the properties of existing metrics related to the points viz. queries may have multiple intents, the likelihood of each intent given a query is available and graded relevance assessments are available for each intent. Also compared a wide range of traditional and diversified IR metrics for search results diversification after adding graded relevance assessments.

2.4. Facets Extraction according to Data Types

Existing automatic facets extraction methods can be divided into three categories corresponding to the different data types: unstructured, semi- structured and structured. They are explained below.

2.4.1. Facet Extraction of Unstructured Data

Unstructured data refers to information that either does not have a pre-defined data model or is not organized in a predefined manner. Unstructured information is typically text heavy, but may contain data such as dates, numbers, and facts as well. In facet term extraction, the most common form of unstructured data is the natural language text, which is always ambiguous and ill-formed.

2.4.2. Facet Extraction of Semi-structured Data

Semi-structured data is a form of structured data that does not match with the formal structure of data models associated with relational databases or other forms of data tables i.e. does not conform to an explicit data schema but on the other hand contains tags or other markers to separate semantically related elements. Semi-structured data lies somewhere between the structured and unstructured data. Examples of the semi-structured data include HTML pages, XML pages, JSON or JavaScript Object Notation. E.g. A Word document is generally considered to be unstructured data. It is possible to can add metadata tags in the form of keywords and other metadata that represent the document content and make it easier for that document to be found when people search for those terms , the data is now semi-structured. Semi-structured data has an implicit formal structure, which can be exploited to improve the quality of facet term extraction. For example, the hyperlinks of web pages can be used to evaluate the importance of facet terms.

ISSN: 2278-4632
Vol-10 Issue-3 No. 1 March 2020

3. System Analysis

3.1. Existing System

Query facets summarize a query in different aspects. They may help users quickly understand important aspects of the query and help them explore information. Dou et al. first introduced this problem and addressed QDMiner algorithm, which first extracts frequent lists in top search results using predefined patterns, and then weights each list and group them into final facets. Like QDMiner, Kong and Allan developed supervised approaches, namely QF-I and QF-J, to mine query facets. Facet item candidates are extracted from frequent lists which are obtained in a similar way as QDMiner. Then two Bayesian models are learned to estimate how likely a candidate is a facet item and how likely two candidates belong to the same facet. All the existing works are based on top search results; hence the quality of final facets might be limited. If some words or phrases don't appear in a list within top search results, they have no opportunity to be facet items.

3.1.1. Disadvantages

- There are no efficient techniques to generate Query facets like Query Based Summarization.
- There is no option to Provide Protection on User Query facets data.

3.2. Proposed System

In the proposed system, the system proposes leveraging a knowledge base as a complementary data source to improve the quality of query facets. Knowledge bases contain high quality structured information such as entities and their properties and are especially useful when the query is related to an entity. The system also proposes using both knowledge bases and search results to mine query facets in this paper. The reason why we don't abandon search results is that search results reflect user intent and provide abundant context for facet generation and expansion. The system's target is to improve the recall of facet and 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 is not harmed too much.

3.2.1. Advantages

- Automatic Query Recommendation.
- The system is more efficient due to Query.
- Based Summarization.

3.3. Implementation

3.3.1. Admin

In this module, admin server has to login with valid username and password. After login successful he can do some operations such as View all users and authorize, View all Researchers and authorize and Add category and subcategory, View all all products details based on clusters with rank, rating, reviews and View all products details with rank, rating, reviews, View all users search transaction based on Search type and View

all user purchased products, View all Keyword Facet by rank and give link to show in chart, View product's rank in Chart

3.3.2. User

In this module, User should register before searching the Website contents. After registration successful the user can login by using valid username and password. After Login successful the user will do some operations Manage Accounts and Query Facet Mining and Faceted Search.

1. Single Facet Query Search

Enter your keyword or keyword+cat, or keyword+cat+Sub Cat, each based on pdesc_Cat+Sub cat and display products Facet details and view details(inc rank), review, Rate, view similar products facets, purchase

2. Multiple Facet Query Search

Select Category and Subcategory and display products Facet details and view facet details(inc rank), review, Rate, view similar products facets, purchase

3.Question Type Query Search

Type some question and match any one word with cat or sub cat or pdesc and diplay products. Facet details and view details (inc rank), review, Rate, view similar products facets, purchase View all search transaction based on Search type and Enter top k value and view searched Keyword facets.

3.3.3. Researcher

In this module, Researcher should register before searching the Website contents. After registration successful the user can login by using valid username and password. After Login successful the researcher will do some operations Select category and sub category and Add products according to category and sub cat with pname, Pdesc(enc), Manufacturer name, Price, Nodel no ,Color, Add image, View all your products details with rank, rating, reviews, View all purchased your products with total bill and View all Keyword Facet by rank.

4. Conclusions

Existing query facet mining algorithms, including QDMiner, QF-I, and QF-J mainly rely on the top search results from the search engines. The coverage of facets mined using this kind of methods might be limited, because usually only a small number of results are used. We propose leveraging knowledge bases as complementary data sources. We use two methods, namely facet generation and facet expansion, to mine query facets. Facet generation directly uses properties in Freebase as candidates, while facet expansion intends to expand initial facets mined by QDMiner in property based and type-based manners. Experimental results show that our approach is effective, especially for improving the recall of facet items.

References

ISSN: 2278-4632
Vol-10 Issue-3 No. 1 March 2020

[1] H. Shum, ―Bing dialog model: Intent, knowledge, and user interaction,‖ http://research.microsoft.com/en-us/um/ redmond/events/fs2010/presentations/Shum_Bing_Dialog_ Model_RFS_71210.pdf.

[2] J. Niccolai, ―Yahoo vows death to the ’10 blue links’,‖ http://www.pcworld.com/businesscenter/article/ 165214/ yahoo_vows_death_to_the_10_blue_links.html.

[3] R. Browne, ―No longer just ’ten blue links’,‖ http://blogoscoped. com/archive/2007-06-28n28.html.

[4] J. Huang, E. N. Efthimiadis, "Analyzing and evaluating query reformulation strategies in Web search logs", Proc. 18th ACM Conf. Inf. Knowle. Manage., pp. 77-86, 2009.

[5] A. Herdagdelen et al., "Generalized syntactic and semantic models of query reformulation", Proc. 33rd Int. ACM SIGIR Conf. Res. Develop. Inf. Retrieval, pp. 283-290, 2010.

[6] W. Kong, J. Allan, “Extracting query facets from search results,‖ in Proceedings of SIGIR ’13, 2013, pp. 93–102.

[7] Z. Dou, Z. Jiang, S. Hu, J. Wen, and R. Song, “Automatically mining facets for queries from their search results,” IEEE Trans. Knowl. Data Eng., vol. 28, no. 2, pp. 385–397, 2016.

[8] O. Ben-Yitzhak, N. Golbandi, N. Har’El, R. Lempel, A. Neumann, S. Ofek-Koifman, D. Sheinwald, E. Shekita, B. Sznajder, and S. Yogev, “Beyond basic faceted search,” in Proceedings of WSDM ’08, 2008, pp. 33–44.

[9] 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, pp. 1029–1038.

[10] D. Dash, J. Rao, N. Megiddo, A. Ailamaki, and G. Lohman, “Dynamic faceted search for discovery-driven analysis,” in CIKM ’08, 2008, pp. 3–12.

[11] H. Zhang, M. Zhu, S. Shi, J.-R. Wen, "Employing topic models for pattern-based semantic class discovery", Proc. Joint Conf. 47th Annu. Meeting ACL 4th Int. Joint Conf. Natural Language Process. AFNLP, pp. 459-467, 2009

[12] W. Dakka and P. G. Ipeirotis, “Automatic extraction of useful facet hierarchies from text databases,” in Proceedings of ICDE ’08, 2008, pp. 466–475.

[13] W. Dakka, R. Dayal, and P. G. Ipeirotis, “Automatic discovery of useful facet terms,” in SIGIR Faceted Search Workshop, 2006, pp. 18– 22.

[14] W. Dakka, P. G. Ipeirotis, and K. R.Wood, “Automatic construction of multifaceted browsing interfaces,” in Proceedings of CIKM ’05, 2005, pp. 768–775.

[15] R. C. Wang, W. W. Cohen, "Language-independent set expansion of named entities using the Web", Proc. 7th IEEE Int. Conf. Data Mining, pp. 342-350, 2007.