

# Generating Query Facet Based on Knowledge Base

Sandeep M S<sup>1</sup>, Hashik H<sup>2</sup>, Sandra k raj<sup>3</sup> Nisha Mohanan<sup>4</sup>

<sup>1,2,3</sup>B.Tech Student, Mount Zion College of Engineering, Kadammanitta, Kerala, India

<sup>4</sup>Assistant Professor, Department of Computer Science and Engineering, Mount Zion College of Engineering, Kadammanitta, Kerala, India

**Abstract-** A query facet is a significant list of information nuggets that explains an underlying aspect of a query. Existing algorithms mine facets of a query by extracting frequent lists contained in top search results. The coverage of facets and facet items mined by this kind of methods might be limited, because only a small number of search results are used. In order to solve this problem, we propose mining query facets by using knowledge bases which contain high-quality structured data. Specifically, we first generate facets based on the properties of the entities which are contained in Freebase and correspond to the query. Second, we mine initial query facets from search results, then expanding them by finding similar entities from Freebase. Experimental results shows that our proposed method can significantly improve the coverage of facet item over the state-of-the-art algorithms.

**Index Terms-** Facets, query, freebase

## 1. INTRODUCTION

Conventional search engine results usually takes a lot of time and troubles the users .An automatic summarization of search results may help users understand the query without browsing many pages .Automatic summarization may save lot of time for user .Mining query facets is an emerging approach to solve the conventional problem. In existing solutions, facet items are extracted from the top 10 search results of a search engine .In existing system .Facet is an another form of the content(query)we used to search in the data base, all the information (terms/results) fetched while searching the query can also be termed as facets.

We use QD-Miner, SimHash & Shingling technique for generating results .In the case of proposed system, we use both existing techniques with Boyer-more algorithm. Proposed system generates results according to the value set by the user. Sim-hash

technique is a technique for quickly estimating how similar two sets are.

The algorithm is used by the Google Crawler to find near duplicate pages. It was created by Moses Charikar. Shingling in natural language processing a w-shingling is a set of unique shingles each of which is composed of contiguous subsequence of tokens within a document.

It can then be used to ascertain the similarity between documents. The symbol w denotes the quantity of tokens in each shingle selected, or solved for QDMiner which discovers query facets by aggregating frequent lists within the top results. QDMiner /QDM for query facet extraction, which appears to be the first work that addressed the problem of query facet extraction. To solve the problem of finding query facets from the noisy candidate lists extracted, they used an unsupervised clustering approach.

## 2. AIM

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. Our aim is to help the users by preserving their previous time by combining displaying the search results with the significance of each query.

## 3. SCOPE

An automatic summarization of search results may help users understand the query without browsing many pages. An automatic summarization may save lot of time for user. Mining query facets is an emerging approach to solve the conventional problem. A QDMiner system that can automatically mine query facets by aggregating frequent lists

contained in the results. In existing solutions, facet items are extracted from the top 10 search results of a search engine. Users can understand the features of query without browsing large number of pages. Query facets can also use reformulations and providing suggestions to the query.

#### 4. RELATED WORKS

Automatically mining facets for queries from their search results by Z. Dou, Z. Jiang, S. Hu, J. Wen, and R. Song (2016)

QD Miner is used to automatically mine query facets by aggregating frequent lists from free text, HTML tags, and repeat regions within top search results. Create two human annotated data sets and apply existing metrics and two new combined metrics to evaluate the quality of query facets. The problem of duplicated lists, and find that facets can be improved by modelling fine-grained similarities between lists within a facet by comparing their similarities.

Beyond basic faceted search by O. Ben-Yitzhak, N. Golbandi, N. Har'El, R. Lempel, and A. Neumann

Extended traditional faceted search to support richer information discovery tasks over more complex data models. Describes a reference implementation of a middleware for faceted search and two extensions to the basic faceted search paradigm. First extension adds flexible, dynamic business intelligence aggregations to the faceted application, enabling users to gain insight into their data that is far richer than just knowing the quantities of documents belonging to each facet.

#### 5. SYSTEM CONFIGURATION

##### A. System Overview

The basic system requirements would be enough for the smooth functionality of the program since there is no use of hardware components. The basic operating system that the device should be having is Windows 7. The operating system helps in ensuring the smooth running of the program with the essential Integrated Development Environment Tool (IDE).

##### B. NetBeans 7.2.1

NetBeans is an Integrated Development Environment for Java. It allows the development of applications using a modular approach. NetBeans platform is a

framework for simplifying the development of Java based desktop applications. NetBeans is the tool that is essentially used for the development of the program by bifurcating it into modules such as admin module and user module.

##### C. MySQL

MySQL is a relational database management system which is open source. MySQL is used by many database driven web applications like Word press. It can be built and installed manually from the source code. However it is more commonly installed under a binary package until special customizations seem necessary. MySQL is used here for the storage of data with respect to the user module which can be monitored by the admin module.

#### 6. PROPOSED SYSTEM

Here we propose the development and use of generating facets with the help of freebase and Boyer Moore technique. We propose 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. Our target is to improve the recall of facet and facet items by utilizing entities and their properties contained in knowledge bases. But at the same time make sure that the accuracy of facet items are not harmed too much.

Our approach consists of two methods which are facet generation and facet expansion. In facet generation, we directly use properties of entities corresponding to a query as its facet candidates.

In facet expansion, we expand initial facets mined by traditional algorithms. The facets constructed by the two methods are further merged and ranked generated final query facets.

It improves the quality of the facets. The most important is that the different aspects of the query can be displayed together with search results. Users can understand the important features of query without browsing large number of pages and query facets can be also used in reformulations and providing suggestions to the query.

In FACET GENERATION we first retrieve relevant entities from Freebase and obtain all the properties of these entities. Sometimes, such properties are not sufficient. To get more properties of an entity, we

further obtain “properties of properties”. Uses Direct and Second-hop properties. If query matches no entities, no facet candidates will be generated in this step.

FACET EXPANSION is the process the QD Miner mines initial query facets, then use Freebase to expand these facets. “Property based” & “Type based” expansion First expanding by assigning each facet to a suitable property of the entities corresponding to the query Add the target entities of the property to enrich the facet. Query can be expanded by the type in which it includes. A web search query is a query based on a specific term that a user enters into a web search engine to satisfy her needs. Enter the query into the text box for obtaining result. From each document  $d$  in the search result set  $R$ . List and context extraction is a method in which, we extract a set of lists  $L_d = \{ l \}$  from the HTML content of  $d$  based on three different types of patterns, namely free text patterns, HTML tag patterns, and repeat region patterns. For each extract list, we extract its container node together with the previous and next sibling of the container node as its context.

A good list contains items that are informative to the query. list weighting is the process in which we propose to aggregate all lists of a query, and evaluate the importance of each unique list  $l$  by the following component. list clustering is the process in which the two list can be list together if they share enough items, we define the distance  $dl(l_1, l_2)$  of two list  $l_1$  and  $l_2$ . In Facet ranking the candidate query facets are generated, we evaluate the importance of facets and items, and rank them based on their importance. Item ranking depends on how many lists contain the item and its ranks in the lists. As a better item is usually ranked higher by its creator than a worse item in the original list.

## 7. CONCLUSION

To satisfy the needs of any user, is an important challenge task of search engine. Improving search experience and providing diversified result set is an ultimate aim of any search engine. This paper describes about the idea of mining facets which can be used to improve the search experience. Facet generation directly uses properties in Freebase as candidates, while facet expansion intends to expand

initial facets mined by QD Miner in property based and type-based manners.

## ACKNOWLEDGEMENT

The authors would like to thank everyone for having given us this opportunity to conduct this study. We would also like to thank Mount Zion College of Engineering and APJ Abdul Kalam Technological University for giving us this platform.

## REFERENCES

- [1] H. Shum, “Bing dialog model: Intent, knowledge, and user interaction,” [http://research.microsoft.com/enus/redmond/events/fs2010/presentations/Shum\\_Bing\\_Di\\_Model\\_RFS\\_71210.pdf](http://research.microsoft.com/enus/redmond/events/fs2010/presentations/Shum_Bing_Di_Model_RFS_71210.pdf).
- [2] J. Nicolai, “Yahoo vows death to the ’10 blue links’,” [http://www.pcworld.com/businesscenter/article/165214/yahoo\\_vows\\_death\\_to\\_the\\_10\\_blue\\_links.html](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://blogscoped.com/archive/2007-06-28-n28.html>.
- [4] Z. Dou, S. Hu, Y. Luo, R. Song, and J.-R. Wen, “Finding dimension for queries,” in Proceedings of CIKM’11, 2011, pp. 1311–1320.
- [5] W. Kong and J. Allan, “Extending faceted search to the general web,” in Proceedings of CIKM’14, 2014, pp. 839–848.