International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395 -0056
Volume: 04 Issue: 03 | Mar -2017 www.irjet.net p-ISSN: 2395-0072

A Survey on Automatically Mining Facets for Queries from Their Search Results

Anusree Radhakrishnan[1] Minu Lalitha Madhav[2]

PG Scholar[1], Asst. Professor[2]

1 Sree Buddha College of Engineering, Alappuzha, India

2 Sree Buddha College of Engineering, Alappuzha, India

Dept. of Computer Science & Engineering , Sree Buddha College of Engineering, Pattoor, Alappuzha

---

Abstract - *Now a days we address the time consuming problem of web searching. Continuously navigating through a number of pages is a difficult task. So query facet is an optimal solution for this. Query facet can be considered as a single word / multiple words which summarize and describe that query. A query facet can be obtained by aggregating the significant lists. The query facet engine will automatically fetch the facets associated with a query. Searching will be easier with the help of facets .It also add the concept of frequent item mining. The facets are assigned a weightage value. In order to display the facets in priority wise manner utility mining concept is also integrated with it. It improves the searching*

Key Words*: *Facet, weightage, utility mining

1.INTRODUCTION

Query facet is derived by analyzing the text query .It allows the users to explore collection of information by applying multiple filters. Faceted search / Faceted navigation is a technique for accessing information organized according to a faceted classification system. Query facets provide interesting and useful knowledge about a query. It improve search experiences. Query facet generate significant aspects. from a large list of queries based on a particular product/ services. Facets access a recommendation for searched users .Automatically mine query facets that exhibits the characteristics of product/ service . A query may have multiple facets that summarize information from a query from different perspectives

2 Literature Survey

In [1] S. Gholamrezazadeh describes about Query-Based Summarization Query facets are a specific type of summaries that describe the main topic of given text. Existing summarization algorithms are classified into different categories in terms of their summary construction methods (abstractive or extractive),the number of sources for the summary (single document or multiple documents), types of information in the summary (indicative or informative), and the relationship between summary and query (generic or query-based). Brief introductions to them can be found. QDMiner aims to offer the possibility of finding the main points of multiple documents and thus save users' time on reading whole documents. The difference is that most existing summarization systems dedicate themselves to generating summaries using sentences extracted from documents, while we generate summaries based on frequent lists. In addition, we return multiple groups of semantically related items, while they return a flat list of sentences..

[2] A. Herdagdelen proposes Query reformulation and query recommendation (or query suggestion) are two popular ways to help users better describe their information need. Query reformulation is the process of modifying a query that can better match a user's information need , and query recommendation techniques generate alternative queries semantically similar to the original query. The main goal of mining facets is different from query recommendation. The former is to summarize the knowledge and information contained in the query, whereas the latter is to find a list of related or expanded queries.

However, query facets include semantically related phrases or terms that can be used as query reformulations or query suggestions sometimes. Different from transitional query suggestions, we can utilize query facets to generate structured query suggestions, i.e., multiple groups of semantically related query suggestions. This potentially provides richer information than traditional query suggestions and might help users find a better query more easily. We will investigate the problem of generating query suggestions based on query facets in future work.

[3] K. Shinzato and T. Kentaro describes about entity search . Some existing entity search approaches also exploited knowledge from structure of webpages . Finding query facets differs from entity search in the following aspects. First, finding query facets is applicable for all queries, rather than just entity related queries. Second, they tend to return different types of results. The result of an entity search is entities, their attributes, and associated homepages, whereas query facets are comprised of multiple lists of items, which are not necessarily entities.

In [4] O. Ben-Yitzhak introduces a technique called faceted search. Faceted search is a technique for allowing users to digest, analyze, and navigate through multidimensional data. It is widely applied in e-commerce and digital libraries. A robust review of faceted search is beyond the scope of this paper. Most existing faceted search and facets generation systems An unsupervised technique for automatic extraction of facets that are useful for browsing text databases. Facet hierarchies are generated for a whole collection, instead of for a given query. Facetedpedia, , a faceted retrieval system for information discovery and exploration in Wikipedia. Facetedpedia extracts and aggregates the rich semantic information from the specific knowledge database Wikipedia.

In this paper, we explore to automatically find query dependent facets for open-domain queries based on a general Web search engine. Facets of a query are automatically mined from the top web search results of the query without any additional domain knowledge required. As query facets are good summaries of a query and are potentially useful for users to understand the query and help them explore information, they are possible data sources that enable a general open-domain faceted exploratory search. Another supervised approach based on a graphical model to mine query facets. The graphical model learns how likely a candidate term is to be a facet item and how likely two terms are to be grouped together in a facet. Different from our approach, they used the

[5] Azilawati Azizan describes query formulation using crop characteristics in specific domain search. Retrieving relevant information from web search is an important task. This is because the web content is of large size and rapid growth happens ..Users are not aware of translating their search content into query. So this paper tries to present seven different query reformulation techniques . A lot of efforts have been made to help users to build their own query. Dome of the techniques involved here are query refinement, query expansion, query disambiguation, query reformulation. For each of the approaches the researchers use different techniques. Query formulation is forming query that represents the users search intent to format that can be used by the search system to process. Query reformulation is modifying initial query to improve the search results.

In [6] Zhicheng Dou describes Generating Query Facets using Knowledge Bases . 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. Here include the following steps which are query facet generation, Facet expansion. The facet candidates constructed by facet generation and expansion are further merged, because there might be duplicate items within these candidates. We then re-weight the final facets by checking the occurrence of the facet items within top search results. Knowledge bases act not only as supplemental data sources, but also bring structured information to query facets. Different items among facets mined by traditional methods are isolated and lean, while during the process of our algorithm, we actually link some facet items to knowledge bases, which could yield many benefits such as (a) finding more information related to each facet item through the link structure of knowledge bases; (b) using the types or properties in knowledge bases as a potential explanation of the meaning of each facet.

In [7] Wisam Dakka explains Faceted Browsing over Large Databases of Text-Annotated Objects. Here, we demonstrate our techniques [1] that discover automatically the facets that can be used to browse an underlying database . It also demonstrate how to enhance the ability of users to identify items of interest in the underlying database, by using ranking algorithms that take into consideration the available screen real estate and with the use of RVSP, an advanced visualization technique that exposes the contents of the underlying database, with minimal use of the screen real estate (Section 3). Finally our system demonstrates how to enhance the browsing experience by using predictive prefetching techniques. It includes the following steps Automatic Facet Discovery, Browsing through Multiple Hierarchies, Adaptive Category Ranking, Rapid Serial Visual Presentation, Prefetching for Interactive Browsing We prioritize the SQL precomputation, by giving higher priority to actions that can be generated by mouse clicks that are closest to the current mouse position. The time that the user spends browsing through the results is typically enough for our system to precompute all the SQL statements that can be generated from the next two clicks of the user.

[8] Damir Vandic introduces Faceted browsing is widely used in Web shops and product comparison sites. In these cases, a fixed ordered list of facets is often employed. This approach suffers from two main issues. First, one needs to invest a significant amount of time to devise an effective list. Second, with a fixed list of facets it can happen that a facet becomes useless if all products that match the query are associated to that particular facet. In this work, we present a framework for dynamic facet ordering in e-commerce. Based on measures for specificity and dispersion of facet values, the fully automated algorithm ranks those properties and facets on top that lead to a quick drill-down for any possible target product. In contrast to existing solutions, the framework addresses e-commerce specific aspects, such as the possibility of multiple clicks, the grouping of facets by their corresponding properties, and the abundance of numeric facets. In a large-scale simulation and user study, our approach was, in general, favorably compared to a facet list created by domain experts, a greedy approach as baseline, and a state-of-the-art entropy-based solution.

[9] K Latha proposes An Automatic Facet Generation Framework for Document Retrieval method . This paper presents an automatic Facet Generation Framework (AFGF) for an efficient document retrieval. Facet generation is the task of automatically discovering facets of documents from text descriptions. In this paper, we propose a new approach which is both unsupervised and domain independent to extract the facets. We also discover an efficiency improving semantically related feature sets with the help of Wordnet, Which carves out a structure that

reflects the contents of the target information collection. Empirical experiments on different text of data show that our approach can effectively generate multi-faceted arbitrary topics; and are comparable with those generated by traditional approaches like Baseline, Greedy and Feedback Language models.

3. CONCLUSIONS

This survey has been performed for collecting the details of different facet mining mechanisms. Different methods were analysed and each has its advantages and disadvantages. Query facet is a single word or multi word that summarises the characteristics of the query. So it is necessary to represent the facet properly. We address the problem of finding query facets which are multiple groups of words or phrases that explain and summarize the content covered by a query. Query facets can be mined out by aggregating significant lists. Query facet is a systematic solution to automatically mine query facets by extracting and grouping frequent lists from free text. Facet based mining will help to find the attributes of a product which are prominant Facet may eliminate multi linking and multi page search method on e- commerce application

ACKNOWLEDGEMENT

We are grateful to our project guide and PG Coordinator Prof. Minu Lalitha Madhav for her remarks, suggestions and for providing all the vital facilities like providing the Internet access and important books, which were essential. We are also thankful to all the staff members of the Department

REFERENCES

[1] S. Gholamrezazadeh, M. A. Salehi, and B. Gholamzadeh, "A comprehensivesurvey on text summarization systems," in Proc. 2nd Int. Conf. Comput. Sci. Appli., 2015, pp. 1–6.

[2] A. Herdagdelen, M. Ciaramita, D. Mahler, M. Holmqvist, K. Hall, S. Riezler, and E. Alfonseca, "Generalized syntactic and semanticmodels of query reformulation," in Proc. 33rd Int. ACM SIGIR Conf. Res. Develop. Inf. retrieval, 2014, pp. 283–290.

[3] T. Cheng, X. Yan, and K. C.-C. Chang, "Supporting entity search: A large-scale prototype search engine," in Proc. ACM SIGMOD Int. Conf. Manage. Data, 2007, pp. 1144–1146.

[4] 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 Proc. Int. Conf. Web Search Data Mining, 2008, pp. 33–44.

[5] Azilawati Azizan, Zainab Abu Bakar (2014)" Query Reformulation Using Crop Characteristic in Specific Domain Search", COMSWARE IEEE European Modelling Symposium, pp. 791--798.

[6] Zhengbao Jiang, Zhicheng Dou (2015) " Generating Query Facets using Knowledge Bases" , IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL.

[7] Wisam Dakka, Panagiotis G. Ipeirotis , Kenneth R. Wood (2013)" Faceted Browsing over Large Databases of Text-Annotated Objects", Journal of Computer and Communications, 2015, 3, 9-20

[8] Damir Vandic, Steven Aanen, Flavius Frasincar (2015) , "Dynamic Facet Ordering for Faceted Product Search Engines", IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING Volume 3 Issue 1 1000140

[9] K.LATHA,K.RATHNA VENI (2014)," AFGF: An Automatic Facet Generation Framework for Document Retrieval" , IEEE International Conference on Electro/Information Technology (EIT) 2010 International Conference on Advances in Computer Engineering , vol., no., pp.602,607,5-7

BIOGRAPHIES

Anusree Radhakrishnan received B.Tech. degree in Computer Science and Engineering from Mahathma Gandhi University, India. Pursuing M.Tech. degree in Computer Science and Engineering from Kerala Technical University, India. Technical University, India.

Minu Lalitha Madhavu received B.Tech. degree in Computer Science and Engineering from Rajiv Gandhi Institute of Technology , MG University, India, received M.Tech. degree in Technology Management from Kerala University, India. Currently, She is Assistant Professor at Sree Buddha College Engineering, Kerala University