Minning Queries From Search Results : A Survey

Anju G R¹ & Karthik M²

¹Student , Mohandas College of Engineering and Technology
²Asst Professor , Mohandas College of Engineering and Technology

Abstract: Address the problem of finding facets from query which explain the content covered by query. Most queries in web search are mainly ambiguous in nature . Understanding the search intent of users is essential for satisfying a user’s search needs. Most of the queries are ambiguous. There are many methods for mining queries from their search results .This paper presents a survey about mining queries from their search result.

Keywords: query facet, facet search, summarization

Introduction

Automatically mining facets from the search result is a difficult task. Many approaches for this. Most queries are ambiguous or multifaceted For example, ‘harry shum’ is an ambiguous query, which may refer to an American actor, a vice president of Microsoft, or another person named Harry Shum. ‘Xbox’ is a multifaceted query. When people search for ‘xbox’, they may be looking for information on different facets of an Xbox, such as ‘online game’, ‘homepage’, and ‘marketplace’. To address the issue of discovering inquiry aspects which are numerous gatherings of words and expression A question may have various features that compress the data about the question from alternate points of view indicates test features for a few inquiries. Aspects for the question "watches" spread the information about watches in five extraordinary perspectives, including brands, sexual orientation classes, supporting components, styles, and hues. The question "visit Beijing" has an inquiry aspect about prominent resorts in Beijing.

We watch that vital bits of data around a question are normally introduced in rundown styles and rehashed commonly among top recovered reports. Therefore we expert posture totaling continuous records inside the top indexed lists to mine question features and execute a framework. All the more particularly, extricates records from free content, HTML labels, and rehash areas contained in the top indexed lists, bunches them into groups taking into account the things they contain, then positions the groups and things in light of how the rundown and things show up in the top results. We expert stance two models, the Unique Website Model and the Context Similarity Model, to rank inquiry features. In the Unique Website Model, we expect that rundowns from the same site may contain copied data, while distinctive sites are free and each can contribute an isolated vote in favor of weighting aspects. Be that as it may, we find that occasionally two records can be copied, regardless of the fact that they are from various sites. For instance, mirror sites are utilizing diverse area names yet they are distributed copied content and contain the same records. Some substance initially made by a site may be re-distributed by different sites, henceforth the same records contained in the substance may show up multiple times in various sites. Moreover, distinctive sites may distribute content utilizing the same programming and the product may create copied records in various sites.

2. Mining Approaches

There are basically three types of media retrieval is available

1) Query based reformulation
2) Query based summarization
3) Entity Search

2.1 Query Based Reformulation

Most standard information retrieval models use a single source of information for query formulation task such as query weighting etc. The process of query formulation modifies the keyword submitted by the user to the search engine. The formulated query is used as the input to search engine ranking algorithm. Therefore the primary goal of the query reformulation is to improve the overall quality of the ranking process. For example, for the query “what is the fastest animals in the world” in Table 1, we generate a dimension “cheetah, pronghorn antelope, lion, Thomson’s gazelle, wildebeest, ...” which includes animal names that are direct answers rather than query reformulations to the query.
It is based on two steps

- The first is the processing stage also called query refinement
- The second the processing stage alters the query on structural level

2.1.1 Advantages

- It is very fast retrieval method.
- More accurate result is obtained

2.2 Query-Based Summarization

Producing a summary a document satisfying a request for information expressed by a query. The summary is the sequences of sentences which can be extracted from documents. The number of sources for the summary that are single document summary and multiple document summaries.

Summary construction methods are of two types

- Abstractive
- Extractive

Steps in query based summarization

- Identification of the correct sections
- Summary generation

Advantages:

- Easy to implement
- Fast retrieval

2.2.1 Disadvantages:

- User must read all the documents
- Time consuming

2.3 Entity Search

The Entity search engines aim at providing the user with entities and relationships between these entities, instead of providing the user with links to web pages. Information is everywhere. The traditional search approach is to scan every document in any order. Its is time consuming for users

An entity refers to any object or thing that can be uniquely identified in the world. Each entity is match with other entity

Benefits of entity search

- It is categorized into taxonomy
- The first task is to make a decision
- More structured than compared to others
More understandable by human
Increase precision
Less time consuming

Fig 4: Idea about entity search

4. Conclusion

As the primary methodology of discovering question features, can be enhanced in numerous angles. For instance, some semi administered bootstrapping list extraction calculations can be utilized to iteratively extricate more records from the top results. Particular site wrappers can likewise be utilized to concentrate top notch records from legitimate sites. Including these rundowns may enhance both precision and review of inquiry features. Grammatical feature data can be utilized to further check the homogeneity of records and enhance the nature of inquiry aspects. We will investigate these points to refine aspects later on. We will likewise research some other related themes to discovering inquiry aspects. Great portrayals of question aspects might be useful for clients to better comprehend the features. Automatically create significant depictions is an intriguing examination subject.

5. References


