

Chapter 49

Temporal Classifier: Classification of Implicit Query on Temporal Profiles

Rahul Pradhan
GLA University, India

Dilip Kumar Sharma
GLA University, India

ABSTRACT

Users issuing query on search engine, expect results to more relevant to query topic rather than just the textual match with text in query. Studies conducted by few researchers shows that user want the search engine to understand the implicit intent of query rather than looking the textual match in hypertext structure of document or web page. In this paper the authors will be addressing queries that have any temporal intent and help the web search engines to classify them in certain categories. These classes or categories will help search engine to understand and cater the need of query. The authors will consider temporal expression (e.g. 1943) in document and categories them on the basis of temporal boundary of that query. Their experiment classifies the query and tries to suggest further course of action for search engines. Results shows that classifying the query to these classes will help user to reach his/her seeking information faster.

1. INTRODUCTION

We have so much information and most of the information or data provide us signals that will help us in observing its usage and importance. (Sharma & Sharma, 2012). In similar way many search engines also look for these signals and rank the result on the basis of these signals, but so far most of the search engines only rely on hypertext structure and string matching, hence the understanding of query intent and using this understanding in, ranking the result, still remains an open question. For this inclusion, temporal dimension will prove to be an asset in improving the quality of retrieved results.

DOI: 10.4018/978-1-5225-5191-1.ch049

Our paper will discuss Query Detection (QD) Technology which is an emerging and important part or component of nowadays leading commercial search engines, as our contemporary users have become very demanding, they are not satisfied by just getting the results which literally matches the query. For instance, we observe that, nowadays we don't bother to correctly spell the words in our query, we rely on '*did you mean?*' suggestions. It is desirous that search engine to understand the query implied intent or the meaning that query seeks and will search the web or document collection and produce the results accordingly. This technology helps us to modify ranking function so that user will get his desired result on top of result list. Most of the time this technology also helps in transforming the user interface as per query's demand, it is something like if you fire query "Taj Mahal" then in right hand corner it will show directions and spatial position of that place. Another example is "Sangam" most of search engines will show map as well as meanings, songs and public parks; this is because Sangam means rendezvous in Hindi as well as there is a popular Hindu pilgrim name Sangam at Allahabad City where three river meets. This is possible by analyzing the information needs of user with respect to different parameters such as their geographical location, language, time, user profile as personalized search and many more.

In case of queries like "Cricket world Cup" many search engines will retrieve the results from the last Cricket World Cup, while user may be interested in information of world cup took place in 1983 rather than the one going to take place in 2015. For this we would like the retrieval system should detect and form clusters of retrieved document, on the basis of temporal dimension of these document (Pustejovsky, Knippen, Littman & Saurí, 2005).

If our systems of retrieval can understand the temporal aspect of queries, we can build a system that take decision about the relevance of result with query intent and modify the current ranking function of retrieval systems accordingly. For this purpose, we access and exploit the temporal information as dates present in document, timestamp of blogs, tweet, and emails as well as for carbon dating the web (Sala-hEldeen & Nelson, 2013) by using server timestamp of web page which found in form of Creation date, Modification date or current server datetime. The major issue is the retrieval system should be fast and capable of providing results to user within short duration of time otherwise we will start losing our users. There will be an issue of normalizing the timestamp as sometime these timestamps are local to server and every geographical located server has its own time zone according to which it record the timestamp, so to make them pin point to single time is a challenge but due to absence of global clock make this problem more difficult, therefore we need a mechanism to normalize time and to make system effective understand them (Alonso, Gertz & Baeza-Yates, 2007; Alonso, Strötgen, Baeza-Yates & Gertz, 2011).

There are few issues our research faces, as we are looking for temporal signals in document, then try to understand and process them. The first and foremost issue is of finding the temporal expression, this is similar to finding a date or an expression that represent time such as "today", "tomorrow", "two months" and many more. Alonso et al. (Alonso, Gertz & Baeza-Yates, 2007) has segregate temporal expression into three major categories which are similar to Schilder et al. (Schilder & Habel, 2001), we discuss these categories below:

- **Explicit Temporal Expression:** These are the temporal expression that holding the temporal information as dates, years or timestamps. These expressions pinpoint to specific point in time. For example, "February 2013" is an explicit temporal expression. These expressions require less processing and easy to normalize, although there are some issues such as 12/02/2014 can be 12th February, 2014 or 2nd February, 2014; these require certain knowledge about document which are usually extracted from document's metadata;

21 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage:

www.igi-global.com/chapter/temporalclassifier/198592

Related Content

Compressing and Vague Querying (XCVQ) Design

Badya Al-Hamadani and Joan Lu (2013). *Design, Performance, and Analysis of Innovative Information Retrieval* (pp. 117-139).

www.irma-international.org/chapter/compressing-vague-querying-xcvq-design/69133

Traffic-Signs Recognition System Based on FCM and Content-Based Image Retrieval

Yue Lian and Wei Wang (2013). *Modern Library Technologies for Data Storage, Retrieval, and Use* (pp. 60-72).

www.irma-international.org/chapter/traffic-signs-recognition-system-based/73766

Analysis of SMOTE: Modified for Diverse Imbalanced Datasets Under the IoT Environment

Ankita Bansal, Makul Saini, Rakshit Singhand Jai Kumar Yadav (2021). *International Journal of Information Retrieval Research* (pp. 15-37).

www.irma-international.org/article/analysis-of-smote/274038

A Knowledge Based Approach for Accident Analysis: Application to the Safety of Automated Rail Transport Systems

Lassad Mejri (2018). *International Journal of Information Retrieval Research* (pp. 51-66).

www.irma-international.org/article/a-knowledge-based-approach-for-accident-analysis/210062

Towards Semantic Mashups: Tools, Methodologies, and State of the Art

Aikaterini K. Kalou and Dimitrios A. Koutsomitropoulos (2015). *International Journal of Information Retrieval Research* (pp. 1-25).

www.irma-international.org/article/towards-semantic-mashups/130005