

Chapter 10

TempClass: Implicit Temporal Queries Classifier

Rahul Pradhan
GLA University, India

Dilip Kumar Sharma
GLA University, India

ABSTRACT

Information retrieval is a field that is emerging day by day as user needs are growing. Users nowadays are not satisfied with results that merely match the query textual words; they want the query to be understood well and then results to be retrieved. These changing requirements need the query to be processed and its hidden intent uncovered. The authors address this problem by creating a system that understands the hidden temporal intent of the query and classifies it into proposed classes. This chapter works on temporal expressions in the document and classifies the query with respect to the temporal expressions in the document. The work is not limited to just classifying the query but also explores how these classifications will help search engines to make modifications in their user interface, which helps users to reach their desired information faster. Temporal boundaries of queries can be found using this work, which will help to disambiguate certain queries.

INTRODUCTION

We are in the big pool of data and it is accumulating day by day. This data needs to be processed and requires to be understood through its attributes and features. These attributes are actually the signals that help us in process of judging whether data is relevant with respect to user needs and query (Sharma & Sharma, 2012). On the same path, most of the search engines look for features, signals and prioritize their results in the form of ranked list which goes through most relevant to least relevant. Many conventional search engines use only the Web structure information and pattern matching, rather than query intention; therefore, prioritizing or ranking results based on query intention is still an area of research which needs to be explored. In the aim of improving the ranking of search results, temporal information hidden in Web pages and documents can be exploited to give more meaningful ranking functions.

DOI: 10.4018/978-1-5225-5384-7.ch010

This chapter will discuss an approach that will consider ‘Query Understanding’ which is a widely used approach in most of the popular search engines; this technology plays an important role in judging the document relevance. While using the search engines or any other information retrieval system the user demands to get more improved results. Such as, the user does not just want the results that match the document which contains query terms, but requests search engine to first understand the query, its intention and then provide results. In the same way, we can observe that the user does not care or bother to write or spell query terms correctly; s/he depends upon spell correction facility provided by search engines. Query understanding technology will help search engines to understand the intention of the query, so that by the help of this understanding the ranking function can be modified or improved, aiming to meet the user’s changing demands. This technology does not just limit itself to ranking function but also helps search engines in presenting the information and results in more lucid and interesting way. For instance, when a user fires the query “India Gate”, then the user’s intention may be either information about India Gate or its history or its location, so search engine can also show map snippet in the result list to help user to find the route to India Gate. This task of improving visualization of results relies on different parameters such as language, context, and location (Sharma & Sharma, 2017; Singh & Sharma, 2013).

There are many queries that will be temporally ambiguous, such as “Milan Fashion Week”, that will not be a single event in time since it reoccurs. Another such queries is “Battle of Panipat” which is not reoccurring at same time interval. It is hard to predict which battle or which fashion week information the user is seeking. Such queries require analysis of returned documents based on temporal dimension (Pustejovsky, Knippen, Littman, & Sauri, 2005).

Conventional search engines cannot harness the temporal dimension in documents. If we are able to understand the query intent and integrate the knowledge gained through temporal analysis of documents we can return much better results, because now we can quantify the relevance on the basis of query intent and temporal aspect of the document. To collect the temporal information of the document and its content, we can use dates and time mentioned in the document, timestamp attached to blogs, Facebook posts, and microblog as tweets and emails (Salah Eldeen & Nelson, 2013). One can also perform carbon dating of the Web using the timestamp of server available in the metadata of Web page or document on the Web usually found in a form of creation date, the current timestamp of the server, or modification date. The major point of concern is about the response time of the system, as some methods taking time are actually futile in the field of information retrieval. For if ranking and results are not available in a timely manner then we will start losing users, which will impact on the economic feasibility of the system. Time itself poses a major issue in front of us concerning on how to manage and normalize the temporal information. Pin-pointing a timestamp to a timeline is tough if we get timestamps from different locations all having different time zones and therefore their meaning differs from one place to other (Alonso, Gertz, & Baeza-Yates, 2007; Alonso, Strötgen, Baeza-Yates, & Gertz, 2011).

Some of the issues that our research confronts are firstly seeking temporal information or features in the document, then understanding and processing it. The first issue our research faces regarding the identification of temporal information in a document is the form of “tomorrow”, “six months” or “today”. Alonso et al. (2007) categorized the temporal information present in the document content into three major classes as proposed by Schilder & Habel (2001). These categories will be discussed below.

- **Explicit Temporal Expression:** This is the category of temporal expressions that contains temporal information that is in the form of years, dates and timestamps. These expressions are easy to

23 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/tempclass/203424

Related Content

Block-Chain-Based Security and Privacy in Smart City IoT: Distributed Transactions

Thangaraj Muthuraman, Punitha Ponmalar Pichiah and Anuradha S. (2019). *Handbook of Research on Implementation and Deployment of IoT Projects in Smart Cities* (pp. 134-148).

www.irma-international.org/chapter/block-chain-based-security-and-privacy-in-smart-city-iot/233270

Adaptability and Adaptivity in The Generation of Web Applications

Raoudha Ben Djemaa, Ikram Amousand Abdelmajid Ben Hamadou (2009). *International Journal of Information Technology and Web Engineering* (pp. 20-44).

www.irma-international.org/article/adaptability-adaptivity-generation-web-applications/4033

Architecture, Specification, and Design of Service-Oriented Systems

Jaroslac Králand Michal Žemlicka (2008). *Software Engineering for Modern Web Applications: Methodologies and Technologies* (pp. 68-83).

www.irma-international.org/chapter/architecture-specification-design-service-oriented/29577

Smart Grid: A Study on Communication Technologies, Security, and Privacy

P. J. Beslin Pajila and E. Golden Julie (2019). *Handbook of Research on Implementation and Deployment of IoT Projects in Smart Cities* (pp. 117-133).

www.irma-international.org/chapter/smart-grid/233269

Multiple-User Simultaneous Testing: Experience with Two Methods

Borchuluun Yadamsuren, Anindita Paul, Sanda Erdelez and Joi L. Moore (2010). *Integrating Usability Engineering for Designing the Web Experience: Methodologies and Principles* (pp. 345-356).

www.irma-international.org/chapter/multiple-user-simultaneous-testing/40507