Information Retrieval for Temporal Bounding

Leon Derczynski
University of Sheffield, UK
leon@dcs.shef.ac.uk

Robert Gaizauskas
University of Sheffield, UK
robertg@dcs.shef.ac.uk

ABSTRACT

The temporal bounding problem is that of finding the beginning and ending times of a temporal interval during which an assertion holds. Existing approaches to temporal bounding have assumed the provision of a reference document from which to extract temporal bounds. We argue that a real-world setting does not include a reference document and that an information retrieval step is often required in order to locate documents containing candidate beginning and end times. We call this task "Information Retrieval for Temporal Bounding". This paper defines the task and discusses suitable evaluation metrics, as well as demonstrating the task's difficulty using a reference dataset.

Categories and Subject Descriptors

H.3.3 [**Information Systems**]: Information Search and Retrieval

Keywords

Temporal bounding, Knowledge base population, Event mining

1. INTRODUCTION

The use of the temporal information in information retrieval is of increasing interest to researchers [1]. It promises enhanced search quality and is of great utility in language processing tasks that rely on selecting documents from large collections [3, 6, 12].

One may achieve a coarse ordering and clustering of documents based on document creation or publication time, but the temporal context of many events and pieces of information contained in a document can differ significantly from that time. Therefore, understanding the temporal structures inside a discourse is key to improving result quality and to achieving effective temporal search.

The temporal delimitation of any assertion is of great importance, for the assertion is true only inside these bounds. All empirical assertions are bound in time – even the observation that *"the sky is blue"* has a beginning and an end. We call the task of delimiting assertions in time **temporal bounding**.

While document timestamps provide explicit, machine-readable dates relevant to assertions, these document creation and publication dates are not applicable to much of the document's content.

Figure 1: Typical temporal anatomy of a news article

Figure 1 details the typical temporal anatomy of a newswire article. The first paragraph often contains a rough reference to the event; the middle of the text will contain precise times about sub-events discussed by the document; and the last paragraphs typically include mentions of historical, similar events. A publication date is typically precise and later than any of the other times mentioned in the document. While temporal "spikes" in a collection might help bound top-level events [11], these are an imprecise source of information and can mislead. For example, many articles related to the April 2013 earthquakes in Iran also referred to the 1978 Tabas earthquake, and searches for *"Iran Tabas earthquake"* using e.g. a vector space model yield a spike in documents written in 2013 – the wrong date for the event in the query. This illustrates the need to look within documents to determine temporal bounds.

The context for temporal bounding is often a corpus of thousands to millions of documents. Systems cannot look within each of these documents to determine temporal bounds. Therefore, information retrieval is used as a critical first step for reducing this search space. This task is information retrieval for temporal bounding (**IR4TB**).

2. TASK DEFINITION

Given an assertion, IR4TB of texts from a larger corpus that provide information concerning the beginning or end times of the interval during which the assertion holds. Many forms of expression contain information or partial information useful for constraining temporal bounds. For the purpose of this task, we call these expressions **temporal descriptions** (**TD**s). In many cases bounds are not explicitly stated but may need to be inferred from what is present in

Table 1: Minimum temporal description coverage and precision over the temporal query sets at document level.

Query type	At rank	1	2	5	10	15	20	30	40	50
Plain queries	Coverage	0.298	0.458	0.644	0.736	0.780	0.800	0.814	0.820	0.824
	Precision	0.298	0.229	0.148	0.106	0.089	0.080	0.070	0.065	0.062
Value near temporal expr.	Coverage	0.200	0.376	0.617	0.668	0.742	0.776	0.797	0.810	0.817
	Precision	0.200	0.188	0.137	0.090	0.077	0.069	0.060	0.055	0.052

the text. We do not try to exhaustively express a taxonomy of temporal descriptions here. Possible forms in text include mentions of a time or an event, as per TimeML [7].

For example, given the assertion `"USA" is-lead-by "Barack Obama"`, the text *"Barack Obama was inaugurated in January '09"* contains a TD. We also consider temporal inference, i.e., chains of events within and across documents. So, given the assertion `"Bill Gates" is-married-to "Melinda"`, if our corpus contains *"Bill Gates graduated in 1982"* and *"Bill married Melinda after he graduated"*, then both texts bear TDs.

IR4TB may be viewed as the inverse of conventional temporal IR. Where one might see typical temporal IR as the retrieval of assertions given a temporal constraint, in the IR4TB scenario, one retrieves temporal constraints given an assertion.

3. RELATED WORK

Specific information extraction tasks which require temporal information include question answering and knowledge-base population. In question answering, systems attempt to answer natural language questions. Information retrieval is used to select potential answer-bearing documents. Generating good IR results for question answering has been a research goal for some time and still contains important challenges [5]. In temporal QA, Saquete et al. [9] found their information retrieval system's temporal capabilities lacking, and introduced a post-processing layer specifically to handle temporal information through filtering retrieved results.

Knowledge base population (**KBP**) involves extracting attribute values from a document collection, given a set of entities. IR is used to find documents that may contain information about values of entity-attribute pairs. In 2011, a temporal bounding subtask was added to the TAC shared KBP challenge, determining start and end times of the values for these attribute-entity pairs [6, 4].

4. EVALUATION METRICS

A text useful for temporally bounding an assertion is one that contains at least one relevant TD. Coverage and redundancy, metrics used for evaluating information retrieval for question answering [8], are also useful for the temporal bounding task. Given a set of assertions, coverage is the proportion of assertions that have at least one relevant TD-bearing document in the results set. Redundancy is the mean number of relevant TD-bearing documents conveying the same temporal bound information, per assertion.

Spurious results should incur a penalty. We use precision for this purpose. In the context of temporal bounding, we define precision as the proportion of texts returned by a query which contain a relevant TD. E.g., if a system returns 50 results for an assertion and 12 contain relevant TDs, the precision is 0.24.

Regarding extrinsic evaluation, one may see Amigó et al.'s metric [2] for temporal bounding systems, which evaluates systems in terms of vagueness and over-constraint.

5. TRIAL APPROACH

The TAC KBP data [6] includes a 1.8 million document mixed-genre collection. This is accompanied by gold-standard temporal bound information, listing a set of assertions as entity-relation-value triples each with dates specifying the bounds of the assertion, and a reference document for each assertion. Though each of these dates may be from different documents, in this dataset the same document is given for all bounds per assertion.

The merged test and training data contains 295 assertions. Only one TD-bearing document is given for each assertion, so we can only measure the minimum temporal bounding coverage of an information retrieval system, and cannot measure redundancy.

We indexed the data with Lucene. Queries comprised an entity name, and one of its attributes' values, as phrases. Table 1 shows that precision falls rapidly before coverage becomes acceptable.

To explore the effect of textual proximity to dates, we annotated temporal expressions in the source data using HeidelTime [10] and then refined our queries to prefer documents where the target entity's name occurs within ten tokens of a temporal expression. The latter section of Table 1 shows the results of this exercise, indicating the resilience of the problem to naïve approaches.

6. CONCLUSION

This paper defines a new task: information retrieval for temporal bounding. We have described the types of information that a system should retrieve and provided metrics for evaluating an IR system in the context of temporal bounding. The problem we outline is complex and difficult and involves challenging research opportunities in both information retrieval and natural language processing.

7. REFERENCES

[1] O. Alonso, J. Strötgen, R. Baeza-Yates, and M. Gertz. Temporal Information Retrieval: Challenges and Opportunities. In *Proc. 1st Temporal Web Analytics Workshop*, pages 1–8, 2011.

[2] E. Amigó, J. Artiles, Q. Li, and H. Ji. An Evaluation Framework for Aggregated Temporal Information Extraction. In *Proc. 1st Workshop on Entity-Oriented Search*, pages 59–64, 2011.

[3] K. Berberich, S. Bedathur, et al. A Language Modeling Approach for Temporal Information Needs. *Advances in IR*, 5993:13–25, 2010.

[4] A. Burman, A. Jayapal, et al. USFD at KBP 2011: Entity linking, slot filling and temporal bounding. In *Proc. TAC*, 2011.

[5] R. Gaizauskas, M. Hepple, and M. Greenwood. Information Retrieval for Question Answering. *SIGIR Forum*, 38(2):41–44, 2004.

[6] H. Ji, R. Grishman, et al. Overview of the TAC2011 Knowledge Base Population Track. In *Proc. TAC*, 2011.

[7] J. Pustejovsky, J. Castano, et al. TimeML: Robust specification of event and temporal expressions in text. In *The language of time: a reader*. Oxford University Press, 2005.

[8] I. Roberts and R. Gaizauskas. Evaluating passage retrieval approaches for question answering. *Advances in IR*, 2997, 2004.

[9] E. Saquete, J. Vicedo, P. Martínez-Barco, R. Munoz, and H. Llorens. Enhancing QA systems with complex temporal question processing capabilities. *JAIR*, 35(1):755–811, 2009.

[10] J. Strötgen and M. Gertz. HeidelTime: High quality rule-based extraction and normalization of temporal expressions. In *Proc. SemEval*, pages 321–324, 2010.

[11] P. P. Talukdar, D. Wijaya, and T. Mitchell. Coupled temporal scoping of relational facts. In *Proc. WSDM*, pages 73–82, 2012.

[12] Y. Wang, B. Yang, L. Qu, M. Spaniol, and G. Weikum. Harvesting facts from textual web sources by constrained label propagation. In *Proc. CIKM*, pages 837–846, 2011.