Tiwiki: Searching Wikipedia with Temporal Constraints

Prabal Agarwal
Max Planck Institute for Informatics
Saarland Informatics Campus
Saarbrücken, Germany
prabal.agarwal@mpi-inf.mpg.de

Jannik Strötgen
Max Planck Institute for Informatics
Saarland Informatics Campus
Saarbrücken, Germany
jannik.stroetgen@mpi-inf.mpg.de

ABSTRACT

Temporal information retrieval received a lot of attention during the last years and it is, in the meantime, widely accepted in the IR community that temporal information needs are important to tackle. A particular type of temporal queries are those with explicit temporal constraints, which make almost 15% of today's Web search queries. Although several approaches to allow textual search combined with temporal constraints regarding the content of the documents have been suggested, there are no publicly available search engines allowing for a time-centric search experience.

In this paper, we suggest TIWIKI, a time-aware search engine for Wikipedia. Relying on steadily updated Wikipedia dumps annotated with temporal expressions, queries with textual and temporal components can be formulated and are served by ranking the search results based on aggregated values of temporal and textual relevance. As the search results directly link to the original Wikipedia pages, the TIWIKI search engine can be considered as slightly delayed, yet timely access to Wikipedia.

Keywords

temporal information retrieval; Wikipedia; TIWIKI

1. INTRODUCTION

Temporal information needs are important to tackle. For instance, it was shown that almost 14% of queries in a Web search engine's query log contained explicit temporal information [25]. However, today's Web search engines still do not allow for an easy way to formulate temporal information needs in the form of explicitly specified time intervals of interest. Thus, it is difficult to retrieve relevant documents for information needs such as *events in Iraq between 2005 and 2010* or *famous speeches in Berlin between May 1961 and December 1961*.

As temporal expressions in the documents' texts are typically not detected and normalized, only explicitly mentioned expressions with a surface form matching the query terms are detected as relevant. In addition, year expressions might be matched with advanced features[1] to search for numbers between 2005 and 2010, but such numbers might occur as regular numbers in documents without temporal meaning. Moreover, underspecified (*July*) and relative expressions (*ten years later*) are not determined as relevant even if they refer to time points within the time interval of interest.

In the research community, several approaches have been suggested to formulate and serve queries with temporal constraints (e.g., [3, 14, 17, 20, 21]). However, while temporal information retrieval has become quite popular [2, 4, 8], only some aspects such as boosting recent documents in news-related queries [13] have found their way to the end user. In contrast, to the best of our knowledge, neither commercial systems nor research prototypes are publicly available to perform time-centric search on a dynamic set of documents.

In this paper, we suggest TIWIKI, a time-aware search engine for Wikipedia. Processing steadily updated Wikipedia dumps with a temporal tagger and indexing the revealed normalized temporal information, TIWIKI searches on up-to-date Wikipedia articles and allows the augmentation of temporal constraints with the textual queries. Thus, intervals of interest can be added to the topical information need description, and search results can be returned based on the topical and temporal relevance of the documents. Information needs, such as the examples above, can thus be easily formulated as ⟨events in Iraq, [2005, 2010]⟩ and ⟨famous speeches in Berlin, [1961-05,1961-12]⟩, respectively, and are well-served as all temporal expressions referring to any time points in the respective time intervals are considered independent of their surface forms.

Wikipedia is highly suitable as document collection as it is known to contain a lot of temporal information. It was thus frequently used for temporal knowledge harvesting, e.g., to extract temporal facts and events [11] and to build cooccurrence-based networks of time, locations, and entity information to summarize events [19]. Furthermore, Wikipedia is a constantly curated resource, which is often accessed to find sophisticated overviews of diverse topics – often expressed as informational queries, which can be nicely combined with temporal constraints.

The remainder of the paper is structured as follows: after surveying related work and explaining diverse types of temporal information available in Wikipedia, TIWIKI's query functionality is explained in Section 4. Section 5 and Section 6 cover its extraction and indexing pipeline and the realized ranking approach. In Section 7, a qualitative evaluation demonstrates TIWIKI's usefulness.

[1] E.g., Google: `https://www.google.com/advanced_search`.

WWW'17 Companion, April 3–7, 2017, Perth, Australia.
ACM 978-1-4503-4914-7/17/04.
http://dx.doi.org/10.1145/3041021.3051112

2. RELATED WORK

Temporal information can be exploited in diverse ways to improve information retrieval (IR) approaches. While opportunities and challenges have been discussed already several years ago [2], good overviews of the state-of-the-art and more current trends in temporal IR research are covered in [4] and [8].

In IR, time can be considered as dimension of relevance, e.g., news-related queries can be better served when considering the freshness of documents [13]. Time can also be exploited as context information as identical queries may represent different information needs depending on when they were formulated. Thus, temporal information can be used to improve search results or query auto-completion [18]. Both, time as dimension of relevance and time as context information can be tackled without considering the content of the documents. However, there are also many applications that exploit temporal expressions occurring in the documents, e.g., to cluster search results along timelines [1] or time intervals of interest [5] or to improve results for implicitly temporal queries [9, 16].

A prerequisite to exploit temporal information occurring in the documents' content is temporal tagging, i.e., to extract and normalize temporal expressions [23]. While temporal tagging was tailored towards processing news-style documents for quite a long time, it is important to take care of domain-sensitive characteristics when processing other types of documents, e.g., narrative documents such as Wikipedia articles [15, 23]. Thus, first publicly available tools support domain-sensitive temporal tagging, namely HeidelTime [22] and UWTime [12], with HeidelTime being much faster, as UWTime requires a deeper linguistic analysis.

Most similar to our work are approaches that address explicit temporal queries, i.e., queries in which temporal expressions occur. Early approaches were presented to allow temporal Web search for Chinese [7] and Spanish [24]. Both respective systems are not available anymore. A general approach for spatio-temporal search in the form of a Lucene extension was suggested in [14]. Furthermore, to address temporal information needs, temporal language models have been suggested [3] as well as learning to rank techniques to serve time-sensitive queries [10]. Two further approaches to spatio-temporal search have been proposed in [17] and [20], however, without providing any techniques to re-rank search results for the different query dimensions. Such a re-ranking was suggested in [21], where the proximity between terms matching the different query dimensions was considered in addition to the textual, temporal, and geographic relevance.

A further topic related to our work, is the extraction of temporal knowledge from Wikipedia. Temporal facts and events have been extracted in [11], and cooccurrences of temporal information as well as location and entity information have been extracted in a network structure to summarize events [19].

Finally, similar as we consider temporal information, a smart search solution is suggested in [6] to allow querying with entity and category names in addition to standard text. However, while some standard semantic search facilities are supported by nowadays popular web search engines, there are – to the best of our knowledge – no systems that allow for a time-sensitive search experience by allowing temporal constraints on the documents' topic – neither in the form of web-scale search engines nor as research prototypes.

3. TIME INFORMATION IN WIKIPEDIA

As of January 13, 2017, the English Wikipedia[2] contains more than 5.3 million articles about concepts, topics, or entities, and careful monitoring of the edits done on these articles ensures the authenticity and quality of information. A temporal context can be associated with each article on Wikipedia. In this section, we describe some Wikipedia aspects that can be leveraged to create an article's timeline.

3.1 Infoboxes

Wikipedia infoboxes are fixed format tables summarizing the page content on the basis of predefined aspects. Although multiple templates might exist for particular types of pages, the aspects that can be associated with a class or type of article are rather fixed. For example, infoboxes in articles related to persons can contain fields such as `birth_date`, `death_date`, and `years_active`. Similarly, infoboxes about organizations can contain fields like `formation`, `extinction`, `membership_year`, `budget_year`, and `revenue_year`.

In the infoboxes, temporal information usually occurs as explicit, structured expressions, and the temporal values can be easily scrapped from the html content. A drawback of using only this type of temporal information is the incompleteness of the data, as infoboxes do not represent a complete temporal picture associated with the topic or entity.

3.2 Revision History

Wikipedia articles are constantly updated.[3] Edits are often made to incorporate new information to existing content. For instance, if a person with a Wikipedia page resumes a public office, respective information is typically appended soon after. Thus, revision dates can represent an active timeline of an entity or topic.

The revision history is maintained by the Wikimedia Foundation.[4] A drawback of using revision history is that pages related to dormant entities, e.g., a dead person or a closed organization, may also be revised by the community just for the sake of adding or correcting content. Such revision dates do not reflect any temporal context related to the entities. Another drawback is the range of temporal information. The edit history is of limited range from the beginning of Wikipedia till today, resulting in pages related to an ancient concept receiving a temporal value of the 21st century.

3.3 Page Views

The number of views received by a Wikipedia page on a particular date is another temporal aspect. Articles receive relatively higher number of views on dates that can be associated with them. For instance, *Graham Taylor* was the most popular page on January 13, 2017,[5] due to his death on the previous day.

The page view data is maintained by Wikimedia.[6] Using just page views as temporal context has the drawback of limited temporal range. Moreover, there could be a considerable fraction of pages not receiving any uptick in the view counts throughout the history of Wikipedia, leading to the association of a null temporal context with the entity or topic.

[2] `https://en.wikipedia.org/wiki/Main_Page`
[3] `https://stats.wikimedia.org/EN/TablesWikipediaEN.htm`
[4] `https://dumps.wikimedia.org/enwiki/latest/`
[5] `http://www.wikipediatrends.com/`
[6] `https://dumps.wikimedia.org/other/pagecounts-ez/`

3.4 Page Content

The main page content is an important and abundant source of temporal information that can be mapped to a Wikipedia page. For example, the abstract of the Wikipedia page about *Roger Federer* contains the following temporal information (marked in bold): *Roger Federer (born **8 August 1981**) is a Swiss professional tennis... ranked in the top 10 from **October 2002** to **November 2016**).*

Such in-content temporal information can be identified and normalized by temporal taggers. Temporal context derived using such an approach does not suffer from temporal range limitation, as the range can be as long as the difference in the date values mentioned in the page content. However, a challenge when using temporal information extracted from the page content is the normalization of underspecified and relative expressions (e.g., *two year later*) as reference dates for them have to be identified in the documents' texts. Therefore, a temporal tagger with a normalization strategy tailored towards processing narrative texts should be applied.

Currently, we solely rely on this type of temporal information as it serves all types of temporal queries equally well. However, we plan to evaluate how the other types of temporal information can be exploited to improve our approach.

4. QUERIES WITH TIME CONSTRAINTS

The primary purpose of a temporal search engine is to rank relevant documents by factoring in the temporal constraints. The retrieved results need to be bounded by a temporal range, with or without a textual component. In this section, we elaborate the possible representations of such queries.

4.1 Temporal Range Queries

A query can consist of a textual and a temporal part, i.e., $q = \langle q_{text}, [t_{begin}, t_{end}] \rangle$. Note that a query can also be specified to a particular point in time by setting the same begin and end date value.

Text query with temporal constraints. Date bounded queries can be issued by a user to get documents with matching text aspect of a query and pertaining to a certain date range. For example, `<brazilian economy, [1990, 1995]>`, where documents related to the *brazilian economy* in the date range of *1990* to *1995* are required. To rank documents, the textual relevance and the temporal relevance have to be combined in a meaningful way.

Query with only temporal constraints. Another use case is that the user is interested in a particular time interval without specifying any textual constraints. Such a query requires a list of most relevant articles from a given date range. For example, the query `< , [1910-11, 1910-12]>` specifies the time interval of interest as *November 1910* to *December 1910*, and documents containing time references to this time interval have to be ranked.

4.2 Query Granularity

The dates specified in the temporal query can be of different granularities. For the sake of simplicity we limit them to day, month, and year. As years consist of months, and months consist of days, the temporal relevance should also factor in the inclusion hierarchy. For instance, given the query `< ,[1990, 1995]>`, the mentions of the dates *1990-01, 1992-01, 1992-02-01*, should also be considered as relevant. Similarly, the query `<Cricket World Cup, [2011-03, 2011-04]>` should also consider all date references of day granularity falling into March or April 2011.

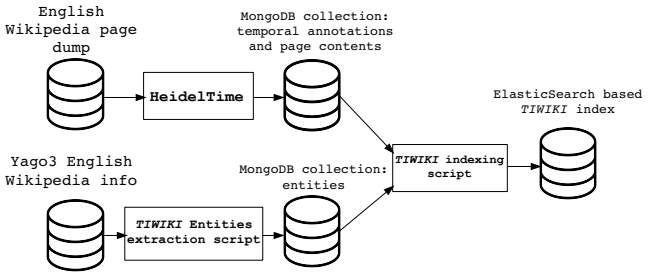

Figure 1: TIWIKI extraction and indexing pipeline.

5. TIWIKI: EXTRACTION AND INDEXING

In this section, we elaborate the TIWIKI extraction and indexing pipeline, which is depicted in Figure 1. First, we describe the extraction and normalization of temporal information from the Wikipedia documents (Section 5.1) and the extraction of entities (Section 5.2).

In Section 5.3 and Section 5.4, we explain how the temporal information is indexed and stored to finally create time-centric snippets. In general, the indexing of the title, entities, temporal context, and page content is done using ElasticSearch, while the MongoDB collections are used to store all information required for the final indexing.

5.1 Extraction of Temporal Expressions

The extraction and normalization of temporal expressions from Wikipedia articles is done with HeidelTime [22]. In contrast to most other temporal taggers, HeidelTime is not tailored towards processing news documents, but contains domain-specific normalization strategies so that it performs similarly well on Wikipedia articles as on news articles. A further advantage of HeidelTime is that it is multilingual [22]. Thus, our approach can easily be extended to further languages, which we aim at as future work.

All normalized information – as well as offset information about the expressions in the documents' content to allow the creation of snippets – is stored in a MongoDB collection before it will be finally indexed for TIWIKI.

5.2 Extraction of Entities

Often, standard textual query terms contain named entity references. Though not focusing on entities but on temporal information, TIWIKI stores an additional field of entities extracted from the documents. This allows to include a further relevance score, i.e., the entity similarity score, in order to improve the retrieval performance. The entities related to a document are extracted using YAGO3 Wikipedia info for English,[7] which contains the out-link information for each YAGO entity having a Wikipedia page. These out-links are links to other Wikipedia pages related to the current page. As each Wikipedia page describes a real-world entity, topic, or concept, adding the page titles of all the referenced pages also adds semantic information to the index.

[7] `http://www.mpi-inf.mpg.de/departments/databases-and-information-systems/research/yago-naga/yago/downloads/`

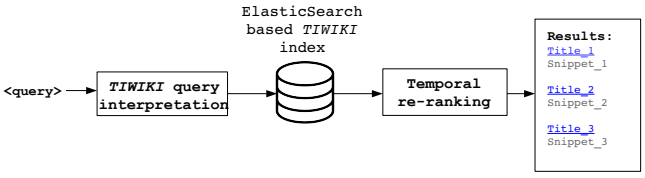

Figure 2: TIWIKI querying and ranking pipeline.

5.3 Indexing Temporal Information

The temporal information extracted by HeidelTime can be stored as a list of objects of the following datatypes in ElasticSearch: String, Date or Integer (by removing the delimiters). In TIWIKI, the temporal ranking is done on top of the scores computed by ElasticSearch. Hence the selection of the correct datatype is important in order to achieve good query run-time performance. In TIWIKI, the temporal context of a Wikipedia page is stored as a list of tuples `<expression, count>`, with `count` being the number of occurrences of the date in the document and `expression` being a distinct normalized date value extracted from the content and cast into an integer type, as string and date comparison would be expensive in terms of computational cost.

5.4 Creating Time-centric Snippets

ElasticSearch has a functionality to construct snippets of matched text from the retrieved documents, which uses the stored offset positions. In TIWIKI, the goal is to provide snippets matching the textual and temporal components of the query. A challenge in the snippet generation process is that the temporal queries do not match with the surface form of the temporal expressions. For example, for a temporal query `<2016-07-10>`, the unnormalized forms of the query mentions need to be highlighted. Therefore, the original temporal expression in the document is replaced by another expression, specifically designed for effective snippet generation. Assume the following sample text with the temporal expression *July 10* (with 2016 as the reference year).

> The final match was played on July 10 between Portugal and France.

Using the temporal tagger's output, TIWIKI indexes:

> The final match was played on 2016_d 2016-07_d 2016-07-10_d t_x July 10 _t_x between Portugal and France.

In a post-processing step of the snippet generation process, irrelevant meta information is removed and only the normalized temporal form specific to the granularity of the temporal query and the temporal expression are shown.

> The final match was played on **July 10** [2016-07-10] between Portugal and France.

6. TIWIKI: DOCUMENT RANKING

TIWIKI takes into account the temporal constraints and ranks the relevant documents accordingly. In Figure 2, the TIWIKI querying and ranking pipeline is shown. After identifying the granularity of the temporal range (if any), a request body of the query is formed. The ElasticSearch based TIWIKI index is used to retrieve relevant documents, which are finally re-ranked on the basis of the temporal bounds.

6.1 Text Queries with Temporal Constraints

For queries with a text part and temporal bounds, relevant documents with respect to the text part of the queries are retrieved. The textual score of a document d for a query q is computed as weighted sum of scores for the textual similarities sim between the textual part of the query and the title of the page, named entities extracted from the page, and the page content.

$$\begin{aligned} s_{text}(q,d) =& \alpha_{title} \cdot sim(q_{text}, d_{title}) + \\ & \alpha_{entities} \cdot sim(q_{text}, d_{entities}) + \\ & \alpha_{content} \cdot sim(q_{text}, d_{content}) \end{aligned}$$

The temporal score of each retrieved document is computed by aggregating the number of date values in the given range it covers (weighted by β_{cover}) and the frequency of those date values (weighted by β_{count}).

$$\begin{aligned} s_{temp}([t_1,t_2], d_{temp}) =& \beta_{cover} \cdot \sum_{t \in d_{temp}} I(t, [t1, t2]) + \\ & \beta_{count} \cdot \sum_{t \in d_{temp}} count(t, [t1, t2]) \end{aligned}$$

with $I(x,y)$ being 1, if x is contained in y and 0 otherwise, and $count(x,y)$ being the number of occurrences of x in y. The final ranking score is the weighted aggregation of the textual and temporal scores, normalized by the maximum textual and temporal scores of all relevant documents D_r for the given query, respectively.

$$\begin{aligned} s(q_{text}, [t_1,t_2], d) =& \gamma_{text} \frac{s_{text}(q, d_{text})}{\max_{d \in D_r} s_{text}(q,d)} + \\ & \gamma_{temp} \frac{s_{temp}([t1,t2], d_{temp})}{\max_{d \in D_r} s_{temp}([t_1,t_2], d)} \end{aligned}$$

6.2 Queries with only Temporal Constraints

To rank documents given only temporal constraints, all documents with mentions of date values within the queried range are filtered out. The resulting documents are ranked using the same scoring function as for queries with textual and temporal parts, with the textual score being set to 0.

6.3 Queries with only Text Aspect

This query functionality has been added to ensure that TIWIKI can be utilized as a standard text matching search engine for Wikipedia documents as well. In this case, the documents are ranked using just the textual score, i.e., the temporal score is set to 0. This functionality is also used in our evaluation to compare the top ranked documents for queries with and without temporal constraints.

7. QUALITATIVE EXPERIMENTS

Due to no publicly available benchmarks for temporal information needs, a large quantitative evaluation is difficult. Thus, our evaluation aims at qualitatively demonstrating TIWIKI's usefulness based on diverse information needs.[8] For this, we compare the top-ranked documents given textual queries and combined textual-temporal queries.

[8]We use ElasticSearch's BM25 for title similarity; its classical tf-idf for content and entities. Weights are heuristically determined and set to α_{title}=$\alpha_{content}$=0.45, $\alpha_{entities}$=0.1; β_{cover}=0.6, β_{count}=0.4; γ_{text}=0.25, γ_{temp}=0.75.

Table 1: Queries with/without temporal constraints.

Query	Ranking
`<george bush, []>`	George W. Bush George H. W. Bush
`<george bush, [1990, 1993]>`	George H. W. Bush Presidency of George H. W. Bush
`<george bush, [2001, 2009]>`	George W. Bush Public image of George W. Bush
`<clinton, []>`	Hillary Clinton pres. prim. camp., 2008 Bill Clinton
`<clinton, [1995, 2000]>`	Bill Clinton Hillary Clinton
`<clinton, [2014, 2016]>`	Political positions of Hillary Clinton Hillary Clinton pres. campaign, 2016

Table 1 shows that temporal constraints can be effectively used to disambiguate queries. We also use the queries collected by Berberich et al. using Amazon Mechanical Turk [3]. These have temporal constraints of different granularities along with textual aspects covering different topics (*Sports*, *Technology*, *Culture*, and *World Affairs*). The queries aim at particular information needs that can be associated with the respective time ranges. As Berberich et al. used the queries to test their approach on a news archive, we do not aim at comparing the performance of the approaches. In Table 2, we instead show, compare, and explain Tiwiki's results for a subset of the queries with the temporal constraints being formulated as text and separately. Note, however, that these queries contain temporal constraints that can be easily expressed with words, and that Tiwiki's full power can be exploited when setting the temporal constraints to time intervals that cannot be expressed as single textual expressions. Nevertheless, the results clearly demonstrate Tiwiki's usefulness to serve temporal information needs.

8. CONCLUSIONS & ONGOING WORK

Queries with a temporal dimension require an IR system that can index and query not only the terms of documents, but also the temporal context of documents. The idea of exploiting the temporal content of the documents is not new, and there has been extensive research in the field of temporal IR. However, Tiwiki is an effort to make a time-aware search engine for Wikipedia available to the general public.[9]

The work on Tiwiki is to be continued. We plan to develop an evaluation dataset to optimize the weights of the ranking functions. As around 800 documents are added to Wikipedia each day, our updating pipeline for indexing new content in Tiwiki needs to be constantly maintained. Finally, Tiwiki could be extended to further languages.

9. REFERENCES

[1] O. Alonso, M. Gertz, and R. Baeza-Yates. Clustering and Exploring Search Results using Timeline Constructions. In *CIKM*, 2009.

[2] O. Alonso, J. Strötgen, R. Baeza-Yates, and M. Gertz. Temporal Information Retrieval: Challenges and Opportunities. In *TempWeb*, 2011.

[3] K. Berberich, S. J. Bedathur, O. Alonso, and G. Weikum. A Language Modeling Approach for Temporal Information Needs. In *ECIR*, 2010.

[9] `http://www.mpi-inf.mpg.de/departments/databases-and-information-systems/research/yago-naga/TimeSEA`

[4] R. Campos, G. Dias, A. M. Jorge, and A. Jatowt. Survey of Temporal Information Retrieval and Related Applications. *ACM Computing Surveys*, 47(2):15:1–15:41, 2014.

[5] D. Gupta and K. Berberich. Identifying Time Intervals of Interest to Queries. In *CIKM*, 2014.

[6] J. Hoffart, D. Milchevski, and G. Weikum. STICS: Searching with Strings, Things, and Cats. In *SIGIR*, 2014.

[7] P. Jin, J. Lian, X. Zhao, and S. Wan. TISE: A Temporal Search Engine for Web Contents. In *IITA*, 2008.

[8] N. Kanhabua, R. Blanco, and K. Nørvåg. Temporal Information Retrieval. *Foundations and Trends in Information Retrieval*, 9(2):91–208, 2015.

[9] N. Kanhabua and K. Nørvåg. Determining Time of Queries for Re-ranking Search Results. In *ECDL*, 2010.

[10] N. Kanhabua and K. Nørvåg. Learning to Rank Search Results for Time-sensitive Queries. In *CIKM*, 2012.

[11] E. Kuzey and G. Weikum. Extraction of Temporal Facts and Events from Wikipedia. In *TempWeb*, 2012.

[12] K. Lee, Y. Artzi, J. Dodge, and L. Zettlemoyer. Context-dependent Semantic Parsing for Time Expressions. In *ACL*, 2014.

[13] X. Li and W. B. Croft. Time-based Language Models. In *CIKM*, 2003.

[14] J. Machado, B. Martins, and J. Borbinha. LGTE: Lucene Extensions for Geo-Temporal Information Retrieval. In *GIIW*, 2009.

[15] P. Mazur and R. Dale. WikiWars: A New Corpus for Research on Temporal Expressions. In *EMNLP*, 2010.

[16] D. Metzler, R. Jones, F. Peng, and R. Zhang. Improving Search Relevance for Implicitly Temporal Queries. In *SIGIR*, 2009.

[17] D. Pfoser, A. Efentakis, T. Hadzilacos, S. Karagiorgou, and G. Vasiliou. Providing Universal Access to History Textbooks: A Modified GIS Case. In *W2GIS*, 2009.

[18] M. Shokouhi and K. Radinsky. Time-sensitive Query Auto-completion. In *SIGIR*, 2012.

[19] A. Spitz and M. Gertz. Terms over LOAD: Leveraging Named Entities for Cross-Document Extraction and Summarization of Events. In *SIGIR*, 2016.

[20] J. Strötgen and M. Gertz. TimeTrails: A System for Exploring Spatio-Temporal Information in Documents. In *VLDB*, 2010.

[21] J. Strötgen and M. Gertz. Proximity^2-aware Ranking for Textual, Temporal, and Geographic Queries. In *CIKM*, 2013.

[22] J. Strötgen and M. Gertz. A Baseline Temporal Tagger for All Languages. In *EMNLP*, 2015.

[23] J. Strötgen and M. Gertz. *Domain-sensitive Temporal Tagging*. Morgan & Claypool Publishers, San Rafael, CA, 2016.

[24] M. T. Vicente-Diez and P. Martinez. Temporal Semantics Extraction for Improving Web Search. In *DEXA*, 2009.

[25] R. Zhang, Y. Konda, A. Dong, P. Kolari, Y. Chang, and Z. Zheng. Learning Recurrent Event Queries for Web Search. In *EMNLP*, 2010.

Table 2: Comparing rankings for queries with text-based and separate temporal constraints; queries from [3].

Category	Query	Ranking	Explanations
Sports	`<boston red sox october 27 2004, []>`	Yankees-Red Sox Rivalry Boston Red Sox 2004 World Series	Query with temporal constraint retrieves the relevant season's page and the full series page as the top documents.
	`<boston red sox, [2004-10-27, 2004-10-27]>`	2004 World Series 2004 Boston Red Sox History of Massachusetts	
Sports	`<ac milan may 23 2007, []>`	Milan War of Currents Lockheed AC-130	That day, AC Milan won the *2006-07 UEFA Champions League* – though originally ejected due to the *2006 Italian football scandal*; *Jankulovski* was one of Milan's players. Text-only query returned *Milan* (the city), a page about electricity and one about a gunship.
	`<ac milan, [2007-05-23, 2007-05-23]>`	2006 Italian football scandal Marek Jankulovski 2006-07 UEFA Champions League	
Sports	`<italian national soccer team july 2006, []>`	Soccer in the United States United States Soccer Federation U.S. men's national soccer team	Results of the time-aware search are more relevant: *Jürgen Klinsmann* coach of German team, which lost against Italy in July 2006, *Francesco Totti* part of Italy's national soccer team, *Bruce Arena* coach of US team, which also played against Italy at the world cup.
	`<italian national soccer team, [2006-07, 2006-07]>`	Jürgen Klinsmann Francesco Totti Bruce Arena	
Sports	`<new york yankees 1910s, []>`	New York Yankees Yankees-Red Sox rivalry Boston Red Sox	The results for the query with the temporal constraint are much more temporally relevant, with *Babe Ruth* being one of the most famous players at that time – who played for the Red Sox (until 1919) and the Yankees (from 1920).
	`<new york yankees, [1910, 1919]>`	History of the New York Yankees Babe Ruth New York Yankees	
Technology	`<mac os x march 24 2001, []>`	Microsoft Office OS X Mac OS 8	On *March 24, 2001*, *Mac OS X 10.0* was released. By setting the temporal constraint, the page about this specific version is ranked higher and the temporally rather irrelevant page on *Mac OS 8* disappears.
	`<mac os x, [2001-03-24, 2001-03-24]>`	OS X Internet Explorer for Mac Mac OS X 10.0	
Technology	`<internet 1990s, []>`	Internet access Internet Explorer Net neutrality	The pages *History of the Internet* and *Internet* give good overviews of the internet in the queried time period.
	`<internet, [1990, 1999]>`	History of the Internet Internet Explorer Internet	
Technology	`<siemens 19th century, []>`	Siemens Siemens Healthineers Unify Software and Solutions	*Carl Wilhelm Siemens* and *Werner von Siemens* are clearly more 19th century relevant than the companies *Siemens Healthineers* and *Unify Software and Solutions*.
	`<siemens, [1800,1899]>`	Carl Wilhelm Siemens Werner von Siemens Siemens family	
Culture	`<woodstock august 1994, []>`	Woodstock Limp Bizkit Woodstock, Ontario	The *Woodstock '94* was a music festival organized to commemorate the 25th anniversary of the original *Woodstock* festival of 1969. The *Woodstock Jimi Hendrix album* – though recorded in 1969 – was released in August 1994.
	`<woodstock, [1994-08, 1994-08]>`	Woodstock '94 Woodstock Jimi Hendrix album Michael Wadleigh	
Culture	`<pink floyd march 1973, []>`	Roger Waters Pink Floyd Pink Floyd live performances	*The Dark Side of the Moon* was released in March 1973.
	`<pink floyd, [1973-03, 1973-03]>`	The Dark Side of the Moon Pink Floyd Pink Floyd live performances	
Culture	`<michael jackson 1982, []>`	Michael Jackson Janet Jackson Scream Childhood	In 1982, *Michael Jackson's* album *Thriller* was released. In contrast *Scream/Childhood* was released in 1995.
	`<michael jackson, [1982, 1982]>`	Michael Jackson Thriller Michael Jackson album Bo Jackson	
World affairs	`<berlin october 27 1961, []>`	West Berlin Berlin Tempelhof Airport History of Berlin	In the *Berlin Crisis of 1961*, soon after the Berlin wall was built, Soviet and American tanks stood 100 yards apart on either side of the *Checkpoint Charlie* on October 27, 1961.
	`<berlin, [1961-10-27, 1961-10-27]>`	Berlin Crisis of 1961 History of Berlin Checkpoint Charlie	