Time-Surfer: Time-Based Graphical Access to Document Content*

Hector Llorens[1], Estela Saquete[1], Borja Navarro[1], and Robert Gaizauskas[2]

[1] University of Alicante, Spain
{hllorens,stela,borja}@dlsi.ua.es
[2] University of Sheffield, United Kingdom
R.Gaizauskas@dcs.shef.ac.uk

Abstract. This demonstration presents a novel interactive graphical interface to document content focusing on the time dimension. The objective of Time-Surfer is to let users search and explore information related to a specific period, event, or event participant within a document. The system is based on the automatic detection not only of time expressions, but also of events and temporal relations. Through a *zoomable* timeline interface, it brings users an *dynamic* picture of the temporal distribution of events within a document. Time-Surfer has been successfully applied to history and biographical articles from Wikipedia.

1 Introduction

The amount of digitized information has grown to reach human tractable limits. Information retrieval (IR) studies methods to retrieve the most relevant documents in a collection given a user search query. However, these documents are often large and contain much more than the information the user is interested in. Consequently, methods to help the user locate the information he wants within the top-ranked retrieved documents are of considerable interest. Current search engines show the most relevant snippets of each search result, but the value of the temporal dimension in IR has been highlighted [2]. Temporal information is present not only in document timestamps or metadata, like creation or modification date, but also within the document content in time and event expressions.

Taking these observations as a starting point, the objective of our work is to demonstrate an information access approach within a document, using advanced temporal representational and navigational techniques. Specifically, our approach, Time-Surfer, relies on identifying times, events and temporal relations in documents and then utilizes this information within a graphical interface to provide users with dynamic time-based access to texts.

2 Related Work

Recent research work demonstrates the importance of the temporal dimension in IR [2,4]. The majority of this work uses the timeline as a basic time

* Paper supported by the Spanish Government, project TIN-2006-15265-C06-01.

P. Clough et al. (Eds.): ECIR 2011, LNCS 6611, pp. 767–771, 2011.

representation [3,1]. A clear sign of its usefulness is Google's timeline feature[1], and also Yahoo's Time Explorer[2] based on SIMILE (*www.simile-widgets.org*).

The cited approaches represent search results on a timeline using the time expressions contained in the documents or within metadata (e.g date of publication). Only [1] represents the main events as reported in the content of the retrieved documents, but they are extracted manually by humans.

Our approach differs from these proposals in the following respects. First, Time-Surfer focuses on accurately representing all the temporally grounded events *contained within large documents* and provides search, comparison and navigation facilities including *dynamic zooming in time*. Secondly, Time-Surfer uses computational tools to *extract references to events, times and temporal relations*, making *the whole process automatic*.

3 Time-Surfer

The Time-Surfer architecture consists of three main steps (see Fig. 1). The first step involves the extraction of the temporal expressions, events and temporal relations from an input document. Recent advances in the field of temporal information extraction (IE) have resulted in the proliferation of systems capable of extracting such elements following the TimeML [6] annotation scheme. In this demonstration, we use the TIPSem system [5] to annotate the input documents.

The second step gropus the events (TimeML) linked to the same time reference into *event groups*. The event class, location in the source text, participants and textual context are saved in a structured format for subsequent use.

Finally, this structured information is loaded into our interactive graphical interface. To develop this web-based interface, we adapted and extended the Flot jQuery library[3] with new representational, navigational and search features.

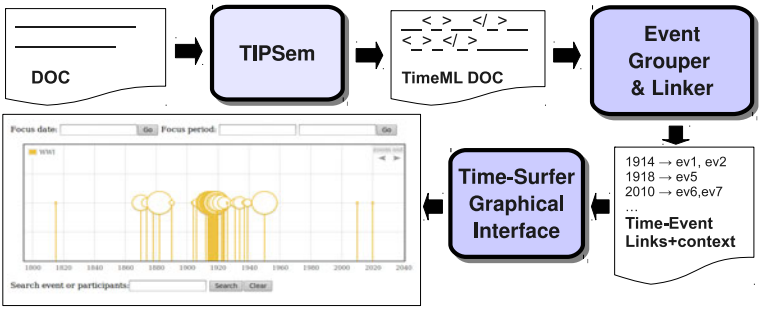

Fig. 1. Time-Surfer Architecture

[1] http://newstimeline.googlelabs.com/
[2] http://fbmya01.barcelonamedia.org:8080/future/
[3] http://flot.googlecode.com

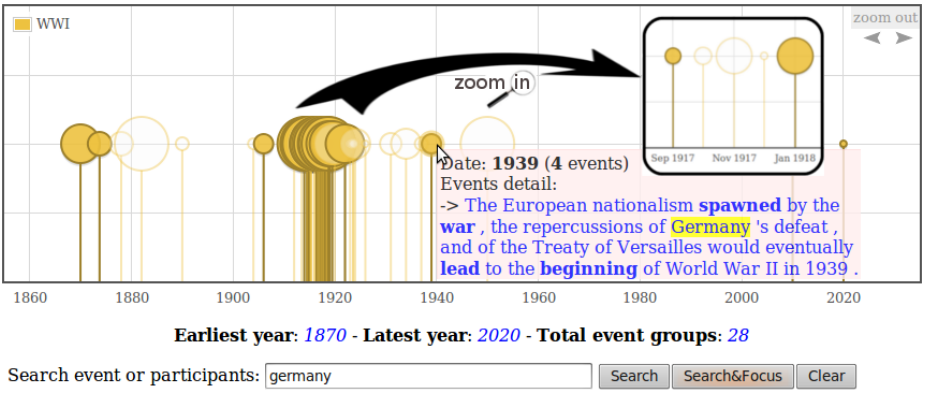

Fig. 2. Searching for "Germany" over WWI, hovering 1939, and zooming

The interface offers the following interactive capabilities:

- **Overview - *Timeline*:** Each event group is represented by a bubble whose position depends on the events' time, and whose radius on the number of events it contains. This gives a general view of the temporal distribution of events in the document, illustrating time bounds and hot periods. Fig. 1 shows this view for the First World War (WWI) Wikipedia article.
- **Surfing and searching:** The interface lets the user navigate the timeline.
  - **Hovering:** When hovering an event group, detailed information is shown (time reference, number of events, list of the events in context).
  - **Clicking:** By clicking an event group, the previous information is presented in a pop-up layer. From there, the user can click on a specific sentence and go to the text of the original document.
  - ***Zooming*:** Using the mouse-wheel the user can zoom-in and out in time. This dynamic zooming allows the exploring of overlapping event groups.
  - **Panning:** The user can scroll backwards and forwards along the timeline intuitively by drag and drop.

  Time-Surfer also includes forms to search a date (e.g., 1916) or period of focus (e.g., 1914-1918), and to search for events (e.g., "battle"), event participants (e.g., "Germany") or participants relations (e.g., "Hitler - Mussolini"). If a query is introduced, bubbles containing relevant events are colored, while others are made semi-transparent. In the text detail, matched instances are highlighted as well as the sentences containing them (see Fig. 2).
- **Comparing documents in time:** The interface allows multiple documents to be compared in time, maintaining the previous dynamic features. Here event groups from each document are displayed using a distinct colour, but are postioned on the same timeline (see Fig. 3).

Time-Surfer has been applied to a set of history and biographical articles from Wikipedia in order to demonstrate the described features – see the on-line demo at **http://gplsi.dlsi.ua.es/demos/TIMEE/Time-Surfer/**

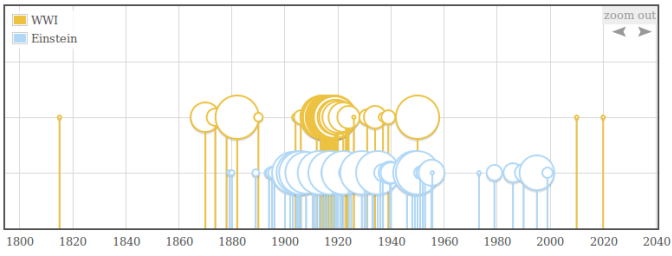

Fig. 3. WWI and Einstein searchable and navigable comparison

4 Analysis and Conclusions

This demonstration presents a novel dynamic user interface, Time-Surfer, that enables time-based access to document content. The system lets users search and navigate, through zooming and panning, the information via a combined graphical and textual representation of the temporal content of a text.

The strengths of the approach include: (1) the interactive interface helps users to *rapidly understand the temporal setting of a document* – the output clearly marks starting and ending time points, as well as the areas of concentration of event groups; (2) such concentration areas can be explored using the *zooming facility* which is missing in related works; (3) the *temporal IE, including events, is automatic*; (4) the search facilities make it easy to find information related to a specific date, period, event or event participant; (5) the *document comparison capability* supports finding relations in time between events from different documents. Limitations of the approach are: (1) substantial processing is needed to extract the temporal content from each document, requiring off-line preprocessing; (2) temporal IE performance is not perfect (85% approx).

In the future we expect both the speed and accuracy of temporal IE to improve, given that it has become the focus of an international research effort. Regarding Time-Surfer, a user-centered evaluation, ideally task-based, needs to be carried out, both to better understand the strengths and weaknesses of the system and to gather further ideas for refining the interface.

References

1. Alonso, O., Berberich, K., Bedathur, S., Weikum, G.: NEAT: News exploration along time. In: Gurrin, C., He, Y., Kazai, G., Kruschwitz, U., Little, S., Roelleke, T., Rüger, S., van Rijsbergen, K. (eds.) ECIR 2010. LNCS, vol. 5993, pp. 667–667. Springer, Heidelberg (2010)
2. Alonso, O., Gertz, M., Baeza-Yates, R.: On the Value of Temporal Information in Information Retrieval. SIGIR Forum 41(2), 35–41 (2007)
3. Alonso, O., Gertz, M., Baeza-Yates, R.: Search Results using Timeline Visualizations. In: SIGIR, pp. 908–908. ACM, New York (2007)

4. Berberich, K., Bedathur, S., Alonso, O., Weikum, G.: A language modeling approach for temporal information needs. In: Gurrin, C., He, Y., Kazai, G., Kruschwitz, U., Little, S., Roelleke, T., Rüger, S., van Rijsbergen, K. (eds.) ECIR 2010. LNCS, vol. 5993, pp. 13–25. Springer, Heidelberg (2010)
5. Llorens, H., Saquete, E., Navarro-Colorado, B.: TIPSem (English and Spanish): Evaluating CRFs and Semantic Roles in TempEval-2. In: Proceedings of the 5th International Workshop on Semantic Evaluation, pp. 284–291. ACL (2010)
6. Pustejovsky, J., Castaño, J.M., Ingria, R., Saurí, R., Gaizauskas, R., Setzer, A., Katz, G.: TimeML: Robust Specification of Event and Temporal Expressions in Text. In: IWCS-5 (2003)