

Position Paper for Design to Read workshop

Pictorial Temporal Structure of Documents to Help People who have Trouble Reading or Understanding

Naushad UzZaman, Jeffrey P. Bigham and James F. Allen

University of Rochester,

Rochester, NY

{naushad, jbigham, james}@cs.rochester.edu

Introduction

Our goal is to explore how automatically generated pictorial representation of the temporal information contained within text may help people read. We hypothesize that such structure may help a diversity of people who have difficulty reading, including people with disabilities, people whose first language is not English, people with low literacy, children, older people, and even those who just want to quickly skim a document. Although we could manually create contents targeting people who have difficulty reading, this approach does not address the large amount of text available on the web already. Instead, we plan to design representations of text that are easier to read and also *automatically* generated from existing documents. This representation, Pictorial Temporal Structure, will be presented to users in a way that is likely to improve reading and unlikely to make reading more difficult. In this paper, we will describe our intended approach to automatically generate pictorial temporal structure and explain why we believe they may make reading easier.

Pictorial Temporal (PicTemp) Structure

A Temporal structure of document orders the events¹ of a document in terms of temporal relations like before, after, overlap and other temporal relations. It gives reader a sense of which event occurred when in relation with other events and also in terms of specific times when possible. A Pictorial Temporal (PicTemp) Structure is a temporal structure of documents using pictures and a screen reader.

Figure 1 (in next page) is an example of PicTemp representation for the text extracted from a Reuters' article². The PicTemp representation in Figure 1 shows the ordering of events in terms of temporal relations, instead of document ordering. Using this representation, we can have a quick overview of event sequence and using images we can get an idea of the events. Combining both, it will give a better understanding of the content to anyone and by clicking any event readers can listen to text. For the example in Figure 1, we have matched the color of event text with associated sentences in the text.

Generating PicTemp structure requires both generating the temporal structure of the document and creating a pictorial representation. Temporal structure help ease the understanding of temporal relations in text because they follow the true sequence of events, based on events' temporal relations, instead of following the document event sequence. Natural writing often interleaves events in order to achieve particular effects in writing. PicTemp shows events that occur first at the top and other events follow temporally below. This can help reader to see the event flow at a glance.

The next component is generating a pictorial representation. We will generate the image of actors (subject) of each event and associated information by searching the web. This way each event node will have associated pictures that will give a sense of who did this event, where, when and in what order.

The PicTemp representation will explain each event with associated images and a textual temporal summary in a timeline. In addition, we will associate the sentence that describes an event with the appropriate event node, so readers can click on the event image and start reading from that sentence. For this purpose, we plan to leverage existing screen reader technology that is already available for blind peoples to read the text (Bigham et al 2008).

¹ Our *event* is according to definition of Pustejovsky et al 2003a

² <http://www.reuters.com/article/idUSTRE60211720100103>

Text from Reuters' article:

Burj Dubai, whose opening has been delayed twice since construction began in 2004, will mark another milestone for the deeply indebted emirate with a penchant for seeking new records.

Dubai, one of seven members of the United Arab Emirates, gained a reputation for excess with the creation of man-made islands shaped like palms and an indoor ski slope in the desert.

With investor confidence in Dubai badly bruised by the emirate's announcement in November that it would seek a debt standstill for one of its largest conglomerates, the Burj Dubai is seen as a positive start to the year after a bleak 2009.

PicTemp representation of the text:

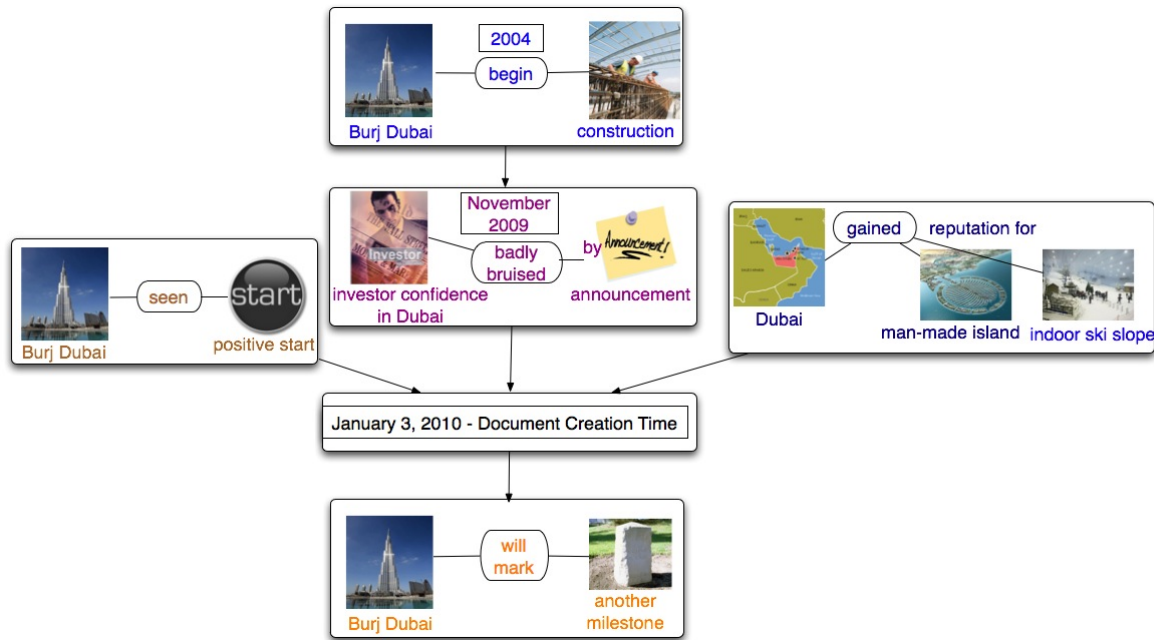


Figure 1: PicTemp representation for a news text. Temporal events are ordered in the figure top down, and their origins are color-coded to help readers connect the pictorial information back to the original text.

Design Motivation

We are designing for people who have trouble reading. Frith (1986) has pointed out, when one component of the developing reading process is dysfunctional, other skills will likely become highly developed in compensation. Our motivation is based on this idea that these people will be able to take benefit of other modalities. This motivated us to propose a multi-modal system with picture, screen reading and temporal ordering.

Several sources (details in Farmer and Klein 1995) support the claim that some dyslexics are more reliant on visual or orthographic coding in reading-related tasks. In addition, Design to Read framework³ also suggest the inclusion of pictures for people with learning disabilities. Inclusion of images in our representation may help many people understand it more easily.

On the other hand, understanding the temporal ordering of events is significant aspect of text comprehension. It gives the ability to summarize the document, do question answering, and language understanding in general. There is a good amount of literature on building temporally aware system in different areas like Artificial Intelligence, Computational Linguistics, Philosophy, etc (Mani et al 2005). In Design to Read framework it was also suggested to build content with simple ideas for people with learning

³ Design to Read framework: <http://www.designtoread.com/Framework>

disabilities. Presenting the events in terms of temporal ordering will assist people to comprehend content more easily, and will help to get a quick summary of the text.

Prior work on screen reading has shown that reading out loud can help people who have difficulty reading text (Bigham et al 2008). Adding a screen reader provides assistance to people, who have problem reading but comfortable in listening, e.g. people with low literacy or aging, and many others.

Combining pictorial temporal ordering, and screen reading, we get a representation that would help many people with reading problem, including some people with learning disabilities, low literacy, aging, stressful context, unfamiliar with the concepts, complex materials. In addition to that, people who do not have trouble reading could use our representation to get quick summary of topic as well.

Past Work and Interest in the Topic

For last several years, the first author has been working on natural language processing and for last few months he was working particularly in understanding temporal aspects of language, by extracting events and temporal expressions, finding temporal relations between event and time locations and between events in the news documents. His work is based on temporally annotated corpus TimeBank, by Pustejovsky et al (2003a). TimeBank is based on TimeML (Pustejovsky et al 2003b) specification, which is a language that describes how to capture time related information. At present, he is extending the work from pair of events to document level, i.e. extracting temporal relation between all events and generating a temporal summary of a document.

Previously, the first author also worked to support his language, Bengali, spoken by more than 200 million people, most of whom are illiterate. His main motivation is to build a tool that will help people of low literacy to access knowledge, i.e. helping people from countries like Bangladesh. This work will be beginning of that and we hope it will also help children, older people, and people with disabilities using English. We also hope to extend it to other languages to help people from countries like Bangladesh.

The second author is the creator of the WebAnywhere screen reader and has become interested in reading as people have come to them describing how they are already using WebAnywhere to help them read, and the third author is a renowned researcher in Natural Language Understanding, Artificial Intelligence and the author of the book Natural Language Understanding.

The Reason why I want to come to this Workshop

I am still in the initial phase of this work and I would like to get feedback in the early stage of my design. This workshop is perfect match for my work, and I hope to learn from the experiences of the participants of the workshop. That's why I wanted to take the opportunity to participate in the workshop and learn more about designing for people who have trouble reading. With our experience in web technology, artificial intelligence, and natural language understanding, we hope that we can also be useful resources for other workshop attendees.

The Critical Issues in the Topic in my Point of View

Our PicTemp representation will be based on sophisticated natural language processing tools that will generate the temporal structure of documents and will also generate pictures for events' associated information. These tasks are even hard for humans, and so we believe main challenges will be (i) creating an accurate PicTemp representation, (ii) creating the interface to PicTemp structure, and (iii) evaluating the PicTemp interface.

The automatic generation of PicTemps will be quite difficult, and this difficulty will make creating useful interfaces and conducting evaluations difficult. The first step in generating a PicTemp is to create the temporal structure of the document. Recent emergence in natural language processing applications like searching, question answering, document summarization that process wide range of text has drawn attention to the need for system that are temporally aware. To accommodate these intelligent application there have been recent interest and investment on temporal information processing. One such initiative is TempEval contest (Verhagen et al 2007), where six different teams participated to identify temporal relations of events in the text. The contest had the following three different subtasks: (i) temporal relations

between event and time expression in the same sentence, (ii) temporal relations between event and document creation time, and (iii) temporal relation between main events of adjacent sentences.

To create the temporal structure of a document, or to build a temporally aware system, we will need all three subtasks' relations. However, for building something like temporal structure, temporal relation between main events of adjacent sentences will be the most useful. To give an idea how hard this task is, this particular task has an inter-annotator agreement of 65% (average of precision and recall), according to Verhagen et al (2007). So far the best performing team has the performance of around 55%, which explains the hardness of this small subtask. This explains, when we get an overall temporal structure, it will be even less accurate, which means it will be a hard task to automatically generate the temporal structure of documents.

The next task is to generate associated images of each event. We haven't fairly explored this area yet; so far we are only considering identifying the agents (subject) of the event and few related information like where it happened, what happened, etc. However, after finding out what to search for, finding the right image is challenge of its own! As an illustration, the following is the first five search results for word "milestone".



Figure: Image Search Result for "milestone" using images.google.com

It will be hard to pick an image from this to represent "milestone" and its equally hard for many other words like this. Expecting worst case, i.e. the image won't represent the word correctly, we also write "milestone" or the words in the representation along with the image. Our challenge in this task would be to extract the right search term.

Finally, generating all these in real-time is very expensive in computation time, and it will be challenging to present the assistive PicTemp representation to the users in short time (something equivalent to loading the webpage). We hope to explore methods (both computation and related to the user interface) that can help us decrease these costs while maintaining the benefits for users.

Conclusion

In this paper, we presented a representation, PicTemp – Pictorial and Temporal representation to assist people, who have trouble reading, e.g. people with low literacy, disabilities, children, older people, and others, to read and understand the contents of text better. This is a very challenging task, as it is even difficult for humans to generate these representations. We are still in our initial stage of work, and would appreciate the feedback of other people in the workshop to refine our design.

References

- (Pustejovsky et al 2003a) J. Pustejovsky, P. Hanks, R. Saur, A. See, R. Gaizauskas, A. Setzer, D. Radev, B. Sundheim, D. Day, L. Ferro, and M. Lazo. 2003. The TIMEBANK corpus. *Corpus Linguistics*, 2003, Lancaster, March 2003.
- (Pustejovsky et al 2003b) James Pustejovsky, Jos M. Castao, Robert Ingria, Roser Sauri, Robert J. Gaizauskas, Andrea Setzer, Graham Katz, and Dragomir R. Radev. 2003. TimeML: Robust Speciation of Event and Temporal Expressions in Text. In Mark T. Maybury, editor, *New Directions in Question Answering*, pages 2834. AAAI Press, 2003.
- (Verhagen et al 2007) Marc Verhagen, Robert Gaizauskas, Frank Schilder, Mark Hepple, Graham Katz and James Pustejovsky, *SemEval-2007 Task 15: TempEval Temporal Relation Identification*, Proceedings of 4th International Workshop on Semantic Evaluations (SemEval 2007).
- (Bigham et al 2008) Jeffrey P. Bigham, Richard E. Ladner, and Craig M. Prince, *WebAnywhere: A Self-Voicing, Web-Browsing Web Application*, In Proceedings of the International Conference on the World Wide Web (WWW 2008), Beijing, China.
- (Frith 1986) U. Frith, A developmental framework for developmental dyslexia, *Annals of Dyslexia*, 36, 69-81.
- (Farmer and Klein 1995) ME Farmer, RM Klein, The evidence for a temporal processing deficit linked to dyslexia: A review, *Psychonomic Bulletin and Review*, 1995
- (Mani et al 2005) Inderjeet Mani, James Pustejovsky, and Rob Gaizauskas. *The Language of Time: A Reader*. Oxford Linguistics, 2005.