

In-depth Accounts and Passing Mentions in the News: Connecting Readers to the Context of a News Event

Earl J. Wagner, Jimmy Lin

The iSchool

University of Maryland, College Park

College Park MD USA

ewagner@umiacs.umd.edu, jimmylin@umd.edu

ABSTRACT

Software that models how types of news events unfold can extract information about specific events and explain them to a news reader. This support can be useful when the background provided by an article is insufficient, if other news coverage exists from which an event's history can be extracted. For extended sequences of related events, it is reasonable to expect that articles published after the sequence concludes include less background coverage of the sequence. Focusing on two stereotypical types of event sequences – kidnappings and corporate acquisitions – we distinguish between articles providing *in-depth coverage*, those having multiple sentences mentioning the same event sequence, from articles making a *passing mention* in just one sentence. We find that, after an event sequence concludes, passing mentions become more common and there are significantly fewer mean mentions per article.

Categories and Subject Descriptors

I.2.4 [Knowledge Representation Formalisms and Methods]: Frames and scripts.

General Terms

Measurement.

Keywords

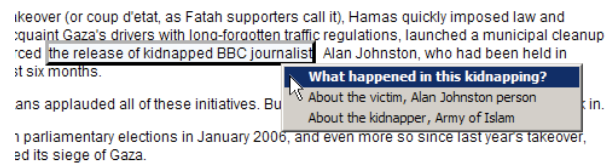
News, Scripts, Computational Journalism.

1. INTRODUCTION

Reading a news article that references an event may provoke interest in earlier events that preceded it. While reading an article mentioning a corporate acquisition, an individual may be curious about the associated events such as an increased or decreased bid, or an offer by a “white knight,” and so on. In many cases she may find this information later in the article, in the background it provides. Because news articles focus on details that are new or have changed, however, they often leave out contextual

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

iConference 2011, February 8-11, 2011, Seattle, WA USA.
Copyright 2010 ACM 978-1-4503-0121-3/11/02...\$10.00.

A screenshot of a news article snippet. The text reads: "ikeover (or coup d'etat, as Fatah supporters call it), Hamas quickly imposed law and equalit Gaza's drivers with lono-forgotten traffic regulations, launched a municipal cleanup 'ced' the release of kidnapped BBC journalist Alan Johnston, who had been held in it six months." A tooltip box is overlaid on the text, titled "What happened in this kidnapping?" and contains two links: "About the victim, Alan Johnston person" and "About the kidnapper, Army of Islam".

ans applauded all of these initiatives. Bu
1 parliamentary elections in January 2006, and even more so since last year's takeover, ad its siege of Gaza.

Figure 1. Inspecting a passing mention of a news situation.

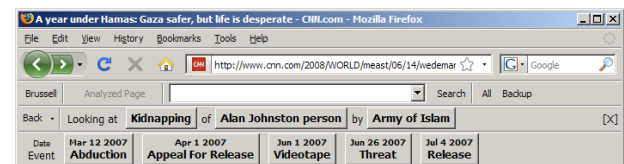


Figure 2. Viewing the history of a news situation.

information of this sort. In an ethnographic study of young news readers conducted by the Associated Press, one individual remarks “if you want background, it’s up to you.” [1] Further, the study also found “subjects were overloaded with facts and updates and were having trouble moving more deeply into the background and resolution of news stories.” It goes on to note that in response to this sense of being “overloaded”, more motivated readers would often investigate this background themselves to determine how stories unfold.

Users’ need to search for the background of events suggests a role for content-aware support. Part of the background of a news event resides in the history of the events that caused it. This history can be gathered by software with human-like expectations of how a sequence of news events is likely to have proceeded [2]. Indeed, the script formalism was originally developed to facilitate computer understanding of stories by employing the expectations a human reader uses to make inferences about a sequence of related events, such as those that make up a restaurant visit [3].

Recent research has used scripts to not only read news but also provide background to news readers [4]. Textual mentions of a news event in a web page (Figure 1) are linked to a timeline-like “storyline” for the sequence previously extracted from news (Figure 2). But the benefit of these storylines in explaining the background of an event depends upon the extent of coverage already provided by the web page. If the page already provides thorough coverage of the course of events, then no further explanation may be necessary. On the other hand, if the page

includes just a passing mention of an event, then a link to an extended explanation of its history may be useful. Further, the extent of a storyline is limited by the progress of the sequence and is by definition incomplete until the sequence concludes – the kidnapping victim is freed, or the acquisition is completed. Thus it would be useful to know how extensively articles cover a sequence before it concludes versus after, when full information about the sequence can be extracted.

We introduce two categories of coverage of event sequences. We define articles with *in-depth accounts* as those with multiple sentences identifying an event sequence by explicitly naming its principals. This is in contrast to articles, often about another topic, that make a *passing mention* and name the principals of an event sequence in only one sentence. If in-depth accounts are prevalent, then a reader seeing a reference to an event is likely to already have an explanation of its history within the article. On the other hand, if passing mentions are common, then readers could benefit by having them linked to further information. Further, we sought to test our expectation that coverage becomes increasingly sparse as time passes after a sequence concludes, when the storyline can provide the most information.

2. EXPERIMENT METHODOLOGY

We gathered two ground-truth test corpora of news situations from Wikipedia¹. One is a set of 164 kidnappings of foreigners in Iraq between the beginning of the US invasion in 2003 and August 2008. For each kidnapping, the full name of the victim including given, middle and surname was collected and converted to a regular expression extended to also match with just a middle initial or without the middle name. The last dated event specified in each entry was used as the final event in the kidnapping.

The other corpus is a set of 892 corporate acquisitions, gathered from list pages within Wikipedia’s category of “Lists of corporate acquisitions” as of January 2010. This category includes lists of acquisitions by 16 technology companies including Google, Microsoft, Apple and Sony, as well as the publisher Condé Nast. For each acquisition in each of these lists, the acquirer, acquiree and the completion date were automatically extracted and converted to regular expressions. Entities’ regular expressions were also extended to match any term redirecting to the associated entity’s page such as “HP” for “Hewlett-Packard.”

A news article was selected if it was expected to be about the news event sequence, indicated by a word matching “acquir*” or “acquis*” for acquisitions, or “kidnap*”, “abduct*” or “hostage*” for a kidnappings. Each selected article was split into sentences and each sentence was scanned looking for explicit identification of the principal(s) of the event. For acquisitions, a matching sentence included both the acquirer and the acquiree, while for kidnappings a matching sentence need only include the victim’s last name if the victim’s name had appeared in full elsewhere in the article.

As input, we used the AQUAINT-2 corpus of 907K English-language news articles published by six sources between October 2004 and March 2006².

¹ <http://www.wikipedia.org>

² Available as LDC2008T25 from <http://www ldc.upenn.edu/>

3. EXPERIMENT RESULTS

According to the criteria above, the news corpus included 1725 acquisition mentions in 562 articles and 16916 kidnapping mentions in 4043 articles. Of the 7149 occurrences in which an event sequence was mentioned at least once in an article, 3888 or 54.4% were passing mentions. For articles published before or on the day after the completion of the sequence, the mean number of mentions was 2.89 (SD = 3.13), with 49.3% being passing mentions. For articles more than one day after the end of the sequence, the mean number of mentions was 2.04 (SD = 2.20) with 64.5% being passing mentions. This difference in mean number of mentions is significant (t-test, $p < 0.005$). The distribution of mentions for these two categories can be seen in Figure 3.

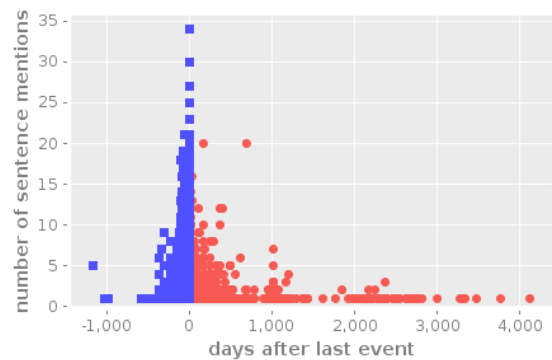


Figure 3. Distribution of mentions in articles before and after news situations end.

4. CONCLUSION

We see that articles with passing mentions are common, and indeed become more common than in-depth accounts after event sequences complete. This supports the observation of the Associated Press that news readers often have to work on their own to learn more the background of mentioned events. With the thoroughness of coverage diminishing after a sequence ends, it is reasonable to expect that automated tools could help in supporting this task.

5. REFERENCES

- [1] Associated Press. 2008. A new model for news: Studying the deep structure of young-adult news consumption. <http://www.ap.org/newmodel.pdf>
- [2] Cullingford, R. 1981. SAM. In Schank, R. C. and Riesbeck, C.K. (Eds.), *Inside Computer Understanding*, (pp. 75--119). Lawrence Erlbaum Associates, Hillsdale, New Jersey.
- [3] Schank, R. C. and Abelson, R. P. 1977. *Scripts, Plans, Goals and Understanding*. Lawrence Erlbaum Associates, Hillsdale, New Jersey.
- [4] Wagner, E. J., Liu, J., Birnbaum, L., Forbus, K. D. 2009. Rich interfaces for reading news on the web. In *Proceedings of the 2009 International Conference on Intelligent User Interfaces (IUI 2009)*.