Contents | **Author index** | **Subject index** | **Search** | **Home**

The BBC, *Daily Telegraph* and Wikinews timelines of the terrorist attacks of 7th July 2006 in London: a comparison with contemporary discussions

Mike Thelwall

Statistical Cybermetrics Research Group, University of Wolverhampton, Wolverhampton WV1 1EQ, UK

Iina Hellsten

The Virtual Knowledge Studio for the Humanities and Social Sciences - VKS, Royal Netherlands Academy of Arts and Sciences, Cruquiusweg 31, 1019 AT Amsterdam, The Netherlands

Abstract

Introduction. *We assess the extent to which published media timelines reflect contemporary electronic discussions of major media events, using the London Attacks of July 2005 as a case study. The main objective is to judge whether timelines could, in principle, be automatically generated from contemporary debates. A secondary objective is to analyse the reasons for differences between contemporary debates and retrospective media timelines.*
Method. *Our method exploits the new opportunities for large-scale analysis afforded by electronic news feeds and blogs. We compared published media timelines with blog postings and news stories related to the London attacks of July, 2005. Rich Site Summary (RSS) technology was used to gather data from 19,587 blog and news sites. For the period of July 7 to July 31 we identified an average of 275 sources a day that posted new information containing the word 'London'. These postings, a combination of blogging and news stories, formed the raw data for our analysis. In particular, we extracted and counted the individual words used in the postings.*
Analysis. *We compared contemporary word usage statistics from our RSS data with the text of three published media timelines, particularly seeking words and types of words that were present in one but not the other.*
Results. *We found that the timelines tended to ignore the role of communication during the event itself, to use less emotionally charged language, and to gloss over to some extent the wider context of the event. Hence some aspects (frames) of the unfolding discussion were ignored by the timelines. Moreover, all sources included a range of relatively trivial details, the timelines apparently using them as concept markers (metonymy) for broader issues.*
Conclusions. *It seems that it would be difficult to automatically generate media timelines from contemporary discussions because of frame changes, and because of changes in terminology and the difficulty in separating highly discussed relatively insignificant details from the key facts. Nevertheless, the comparative analysis techniques reported in this paper may provide an interesting new window on contemporary discussions and media representations of major events.*

Introduction

The timeline is a useful device with which to summarise the key facts of an event, particularly when the sequence of occurrences is important. In principal, a timeline could cover any time period from the history of the universe to the lifecycle of a nuclear fusion reaction and could serve an illustrative device or

even an interface for an information retrieval system (Allen 2005). Given the increasing availability of digital information, it is logical to investigate the extent to which timelines may be automatically generated from coherent corpora of documents (Swan & Allan 2000). In particular, there is a now a wide variety of news-related online information ranging from newspaper Web sites to various kinds of diary-like blogs (Herring, Scheidt, Bonus, & Wright, 2004). Hence there is an opportunity to investigate both the extent to which news-related timelines can be automatically constructed from contemporary online sources and the extent to which existing media timelines reflect contemporary perceptions of major events.

In computational linguistics, a similar and established task is that of automatic summarisation: producing a short synthesis of one or more documents (Mani 2001). Various methods are available for this task and there are many different types of automatic summarisation. To give a simple example, in sentence extraction, weights are assigned to the sentences of a text using factors such as the existence of key phrases and the concentration of text keywords. The highest-weighted sentences could then be selected as the document summary. Timeline generation is different from automatic summarisation, however, because of the temporal component. In this paper we are specifically concerned with the automatic generation of timelines for events attracting considerable public discussion and extensive media coverage. Timelines are typically used in the coverage of sudden crises, such as terrorist attacks, but also in the representations of the spread of diseases, such as bird flu. In short, they are used as condensed stories about the *steps* in the development of the event. In this sense, timelines provide an anatomy of the event.

In this paper we use blogs and news feeds as a convenient source of ongoing time-stamped information and discussion, comparing them to event timelines published by authoritative sources for the London attacks of July, 2005. Although bloggers are not typical world citizens, being presumably US-centred, younger and richer than average (e.g., BBC 2005; Gill 2004; Lin & Halavais 2004), the ease of use of blog creation software allows blogs to be created by a wider section of the population than Web sites or any other current mass-publishing Internet technology. Hence it is reasonable to use blogs as a source of public opinion and attitudes, as many have already done (Fukuhara 2005; Glance *et al.* 2004; Gruhl *et al.* 2004; Pikas 2005). In particular, blogs offer, to a greater extent than previously available, unobtrusive access to the perspectives of the consumers of mass media information. Note that there are some free online resources that can be used to track blog discussions, such as blogpulse.com, which can generate graphs to illustrate the frequency in 'blogspace' of any given word. For example, a graph of the frequency of "London" in blogs could be expected to give insights into when the UK capital was a topic of blogger discussions. Nevertheless, there is a continuum between personal diary-like blogs with tiny friendship circle readerships and 'A-list' blogs with hundreds of thousands of readers (Herring *et al.* 2004), which is more than many national newspapers. These top blogs are almost mass media. Hence it is reasonable to analyse a collection of documents that includes both mass media and a wide range of types of blog.

We use a word frequency approach, focussing on words that enjoyed a significant increase in usage during the crisis but which were omitted from timelines, and for timeline events that did not have a high profile when they occurred. The purpose of the case study is to assess the extent to which words that increase in usage during a crisis flag events that should be included in a timeline. A secondary aim is to assess the extent to which media timelines reflect major events in terms of what could be considered the essential facts at the time and retrospectively. Another secondary aim is to assess whether existing timelines could be used as benchmarks for fine-tuning timeline generation systems, in the way that corpuses of correctly classified documents are used as benchmarks for information retrieval systems (see the Text Retrieval Conference (TREC) site). In other words, would it be a good idea to design a timeline generation system specifically to replicate existing media timelines? Finally, we are also interested in the wider potential of the techniques introduced here to cast light on social science issues such as the role of blog communication in politics and its relationship to the media. This is an exploratory paper: since these issues do not seem to have been researched before, we do not have natural choices for specific research hypotheses.

Fact selection in media reporting

In the mass media, news events are not reported from a purely factual perspective but are *framed* in various ways. As Gamson and Modigliani (1989: 2) claim, 'Public discourse is carried on in many different forums. Rather than a single public discourse, it is more useful to think of a set of discourses that interact in complex ways.' In the complex nexus of various competing discourses in the public media, there is a need for tools that provide some coherence to the issues. Frames provide a central organising idea that helps to put the particular news item in a wider context.

> *Framing essentially involves selection and salience. To frame is to select some aspects of a perceived reality and make them more salient in a communicating text, in such a way as to promote a particular problem definition, causal interpretation, moral evaluation, and/or treatment recommendation for the item described. (Entman, 1993: 52 (italics original)).*

Frames, such as *catastrophe*; or *progress*, are used by the media to provide the core idea what an event or issue is about, with the event's reported facts selected to fit the frame. Both timelines and frames, however, provide a view on the dynamics of a public event; they are communicative tools that aim to provide coherence and wider context for making sense of an issue.

Frames are not unique to the media but are a natural method organise reality (Goffman 1974). Nevertheless, the use of frames by the media implicitly suggests limiting of the range of facts judged salient to an event. Given the influence of the mass media, it seems reasonable to suppose that even in blogspace the most discussed aspects of an event would be those aligned with the main current media frames. Despite this, blogspace is also known for its use by individuals and groups that campaign against mainstream perspectives (Gorgura 2004; Kim 2005), and so it is possible, in theory, that the blogspace facts of an event would be different from the media facts of an event, simply because of the use of different frames. An example of this is the issue of whether the London attacks were related to the UK's invasion of Iraq: the prime minister stated categorically that they were not and hence put implicit pressure on the media not to incorporate Iraq into the 'facts' of the London attacks.

Frames are not static but develop over time in the coverage of a specific topic. The life-cycle of a topic in the mass media has been described by Downs (1972) as an issue-attention-cycle where issues develop through specific phases. We expect that the frames or phases manifest themselves in the differences in the word usage. In the London attacks, for example, the first news items reported about the happenings based on the information given by the police and fire brigade as well as the eye-witnesses. In this phase, the main aim is to provide the public with information of *what* has happened and *where*. Second, the separate events were drawn together and framed as a catastrophe and terrorist attacks. In this phase, the media reporting focused on the question *why* has this happened and *who* has done it. Afterwards other frames appeared, such as the consequences of the bombings for the individuals affected as well as for the state, in emotional, political and economic terms. In each of these frames the facts discussed could be expected to be noticeably different.

Method

We built an RSS feed corpus containing both blogs and news feeds as the raw data for this paper. Rich Site Summary (RSS) is an Internet-based format used to disseminate concise summaries of blog postings (Gill 2005; Hammersley 2005) and this makes it a more practical data source then the blogs themselves, which are difficult to automatically analyse in large numbers because of their complex, repetitive structure; although some companies such as Intelliseek Inc. (Glance *et al.* 2004) and some researchers have achieved this (Kumar *et al.* 2003). Each unit *item*, as collected by us from an RSS feed, is a single news story (e.g., a BBC news headline) or a single blog posting. Some of the items are complete stories or postings, whereas others are brief summaries. Whether to summarise or not is the choice of the RSS feed owner. Each feed may report zero, one or more new items a day. While active bloggers and major news sites may have tens or hundreds of new *items* a day, occasional bloggers may just post once a week or less. The contents of the items vary as much as blogs, from very personal, diary-like, comments to political debates on contemporary issues.

We analyse news sources together with blogs: both extensively use RSS and a combination of the two provides a broad spectrum of contemporary discussion. It is quite difficult to separate media from blog RSS feeds on a large scale without using considerable human labour to check each one. This is partly a result of the similarity of popular blogs to media sources: both may post frequently on current news topics. Since we need a large number of RSS feeds for our method to work, we are constrained to a broad-based collection of both blogs and news sources. A disadvantage of the broad coverage is that the data are difficult to analyse from a *social science* perspective in terms of cause-and-effect relationships because of the wide variety of types of source, presumably having varying communication strategies and motivations. For example, had we restricted the sample to a much narrower set, such as male, US, student bloggers, then the data could have been analysed with a deeper understanding of its creators. Our aim in this paper is not to analyse one sector of society, however, nor to compare blogs to news (e.g., Thelwall 2006), but to compare the words used in timelines to those used in the general combined blogs and news corpus.

The method for generating RSS data is based upon previous research (Prabowo & Thelwall 2006; Thelwall *et al.* forthcoming). A collection of 19,587 RSS feeds was built from browsing major news sites, searching RSS and blog databases, and Google searches. The feeds were monitored hourly (daily if rarely updated), starting on January 31, 2005. We report results from the date of the first London bombings, July 7, 2005 to July 31, 2005 to include both the London attacks and their immediate aftermath.

Since our feed corpus is a general collection, most of the information was unrelated to the London attacks. RSS information is posted in item units, with each item normally covering a single topic or piece of information, for example a blog posting. Hence the first task was to select all the items relating to the London attacks. This was achieved by the simple expedient of automatically removing all items not containing the word *London*, creating a London sub-corpus. Of course, many items relating to the event would not contain this word and many unrelated items were included but this seemed reasonably effective at generating a large number of relevant items. A total of 34,880 feed-days (i.e., feeds containing at least one relevant item on any given day), was extracted and, based upon the difference in average numbers during the crisis (275) and the average number for the month before (115) there were probably about 160 relevant feed-days a day from July 7 to July 31: a total of 4,000, accounting for a majority of the London-subcorpus during the period examined.

The algorithm for extracting significant events is based in part upon previous research (Gruhl *et al.* 2004). A daily time series was generated for each word in our London sub-corpus, giving the percentage of feeds containing the word on each day, out of the total number of feeds posting in the sub-corpus on that day. For each word, the day on which this frequency experienced the biggest jump (spike) was recorded as well as the size of the spike. Spikes were measured against the average frequency of words in the London sub-corpus over all previously recorded days (from January 31, 2005).

After excluding temporal words (days, months and their abbreviations) we selected the top 1,600 spiking words as candidates for association with or describing the major events related to the London attacks. The number 1,600 was arbitrarily chosen due to system limitations. We searched the Web for timelines of the London attacks published by authoritative sources and giving a reasonable amount of detail over several days. We selected three: from Wikinews (accessed December 6, 2005) the BBC (accessed December 21, 2005) and the *DailyTelegraph* newspaper (accessed December 6, 2005). The words in each timeline were compared with the 1,600 selected from the London sub-corpus and matches recorded in two ways. First, a simple count of occurrences of each word from the 1,600 in each timeline as compared to the sub-corpus was recorded. Second, in each timeline all words used were checked and all those occurring in the 1,600 from the London subcorpus were highlighted (in yellow). The purpose of this was to be able to manually identify and analyse in context words that occurred in each timeline but did not significantly increase in usage within the London subcorpus (i.e. the unhighlighted terms, see Figure 2).

The first author performed an inductive classification exercise on the top 200 spike words that were missing from all timelines. The purpose of this was to identify common classes of reasons for omission. We then conducted a qualitative analysis of words missing from each timeline but appearing in the sub-corpus, seeking an initiative explanation both in terms of the specific timeline sources and in general terms for retrospective reporting.

The research method is information-centred in the following sense. A media studies approach might use methodologies such as content analyses to describe an event or test specific hypotheses about its reporting (Weare & Lin 2000) or frame analysis to study the development of a public debate (Gamson & Modigliani 1989; Scheufele 1999). In contrast, a computer science approach might develop an algorithm to extract particular facts, such as the dates of key events: a text mining (Kleinberg 2003) or data mining (Han & Kamber, 2000) approach. An information-centred approach (Thelwall & Wouters, 2005) is exploratory: to investigate the kind of information that a data source could reveal, but without a pre-selected research hypothesis.

Results

Figure 1 summarises the number of timeline mentions of each word extracted from the corpus. The tendency is for words with a higher jump in usage frequency during the crisis to be found more often in the timelines. This shows that the timelines tend to reflect the words selected by our method and so it seems reasonable to attempt to generate timelines from the selected words. Moreover, the graph is evidence that words found in one source but not the other are to some extent unusual and worthy of investigation.
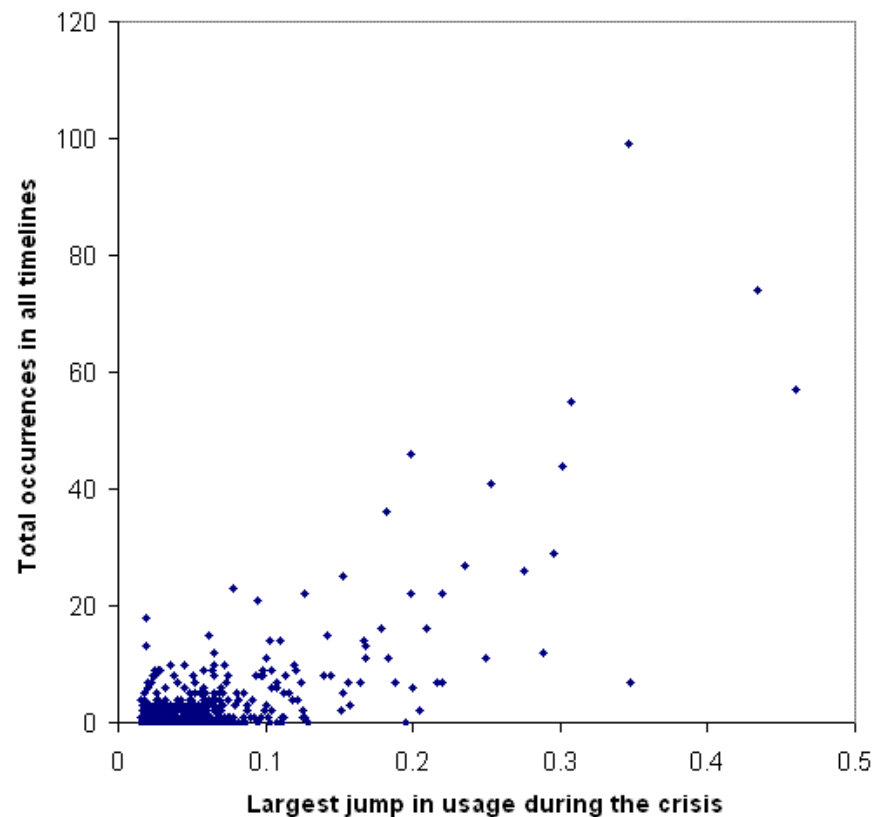
Fig. 1. Timeline occurrences for the top 1,600 spike terms (with the largest usage jump) in the crisis.

Classification

Fifty-nine of the top 200 words from the London sub-corpus (i.e., words automatically selected for large jumps in usage during the crisis) were not used in any of the three timelines. These were classified and the results are presented in Table 1. Some of the words could have been classified differently, for example the word *subway* is US-English and was used in the context of US events related to London. The purpose of the classification is to provide evidence that the described phenomena occur, rather than to measure the extent to which they occur; hence a formal classifications procedure, such as content analysis (Krippendorff 1980), would be inappropriate.

Classification	Terms	Examples or information
Emotional expressions or reactions	innocent, evil, tragedy, dramatic, ideology prayer, safe	'the war between good and evil'; 'i kinda felt nievely [sic] safe from them here'
Dead end or incorrect information	biochemist, egypt, zambia, questioned, camp, militant, lahore, magdy, nail	'Egyptian biochemist is arrested in Cairo'; 'man detained in Zambia'
Other events related to London	olympic, random, bag, patriot	'NY police begin random bag searches on subways'
Temporal shift in grammar	shoot, evacuation, silent	'Shot', 'evacuated' and 'silence' were used instead.
Discussion of the event and its wider context	madrid, afghanistan, radical, perpetrator, cop, spreading, surveillance, religion, islamist, shoot-to-kill	'today the frontline crosses Madrid and London'; 'uk police defend shoot-to-kill after fatal mistake'

Very specific information about the event	admit, detonator, smoke, widen, buses, regret, troop, immigration, chased, issued, manhunt, normal	'suspect held in Italy said to admit carrying bomb in train'; 'hunt widens for evidence in London blasts'; 'police arrest under immigration law'; 'London transport network is returning to normal'
Discussion of information sources	coverage, flickr, sky	'please do send in examples of more coverage'; 'Livingstone told Sky News'
Change in terminology/language	civilian, wounded, botched, remaining, unconnected, briton	Military terminology tended to be dropped, e.g., *wounded* replaced by *injured*. 'Botched attacks' replaced by 'failed attacks'. The information in the last three words the was given in a different way or implicit in the timeline
Non-UK English or alternative spelling	subway, transit, qaeda	'NY police to search backpacks on transit system'
Unrelated to the London attacks	harry, potter	'new Harry Potter adventure flies off shelves', the release of a Harry Potter novel

Table 1: A classification of automatically selected spiking terms found in the London sub-corpus but not used in any timeline.

It is noticeable that most of these categories of the words that were missing from the timelines reflect the various phases in the development of the issue, some of them framing the event as war against terrorism (use of military words and the search for the suspects), some as the consequences of the attacks. These different categories also reflect the development of the reporting of the issue over time from describing what has happened to searching for the reasons of the attacks.

Analysis of words used in the three media timelines

Wikinews Figure 2 shows an extract from the annotated Wikinews timeline to illustrate a section of the output of the annotation software. Wikinews is an open access data repository where anyone can write text on current issues (http://en.wikinews.org/). The Wikinews timeline included many specific details that were not alluded to in the blog/news corpus, for example that passengers tried to 'break windows with umbrellas in order to escape'. In Figure 2 it can be seen that some of the names of the blast locations featured in the top 1,600 London sub-corpus words, but not all, showing that the spike words do not seem to represent a *systematic* set of facts of the event.

17:30: Having flown back from Scotland, from the G8 Summit, Prime Minister Tony Blair emerges from a meeting in Downing Street and urges the public not to 'be terrorised'.
18:13: Deputy Assistant Police Commissioner Brian Paddick confirms 37 fatalities: two in the bus explosion at Upper Woburn Place/Tavistock Square, seven at Liverpool Street/Aldgate, seven at Edgware Road, and twenty-one in the King's Cross/Russell Square blast, as well as around seven hundred injuries, with roughly three hundred of those being transported by ambulance to London hospitals.

Figure 2: A section of the Wikinews timeline: highlighted words are in the top 1600 from the London sub-corpus (July 7, 2005).

Notable missing terms include 'Scotland' and 'Blair': both were relatively frequently alluded to in the news and blogs before the blasts because of the important meeting that was taking place in Scotland during the first terrorist attacks of 7 July, masking their importance in July because the relative frequencies did not increase significantly. Wikinews gives both highly specific information about the crisis and information that was later superseded, for example giving casualty estimates at different points in the crisis. Hence it reports contemporary perspectives in addition to the facts known in retrospect. One very specific fact given, for July 13th, was 'In response to public pressure, the United States Air Force bases at RAF Lakenheath and RAF Mildenhall in Suffolk lift travel bans imposed on service personnel in the wake of the bombings'. This was mentioned just twice in the London sub-corpus.

BBC For compatibility with the duration of the blog data collection, the BBC timeline is only considered for the period up to the end of July, although it continued to November in the version accessed. Whilst the Wikinews timeline covered the few weeks of the attacks, the BBC's devoted much less space to the attacks themselves and covers a longer time period. One consequence of this is that it did not give a flavour of the uncertainty at the time of the attacks,

for example not reporting incorrect casualty estimates. This may be because the Wikinews is open access media while BBC is an established news agency that relies on other ethical norms (such as those of the Society of Professional Journalists (1996)) in reporting than Wikinews, such as that the facts have to be confirmed or tested before they can be published. Like Wikinews, the BBC timeline reported a range of specific details that did not feature in the corpus. These included a joint statement between the UK and Spanish prime ministers, and the fact that the Police at one stage cordoned off 'a white VW Golf in East Finchley.' Moreover, the timeline reported some significant dead ends, as shown in Figure 3.

A Leeds chemistry student, arrested in Cairo as part of the investigation, has nothing to do with the bombing, the Egyptian government says.

Figure 3: A snippet from the BBC timeline (July 19, 2005).

Daily Telegraph The *Daily Telegraph* presented a themed 'London terror factfile' rather than a direct timeline, although it serves a similar purpose, and was more provocative and discursive than the others, see for example Figure 4. This difference is again based on the differences in the expected audiences of the BBC and the *Daily Telegraph*. It did not seek to give just the facts but attempted to keep the drama of the event and to present a debate around what were perceived to be the key issues (or frames), particularly in the 'opinion' section at the bottom of the page.

France saw attack risk
A serious threat of al-Qa'eda terrorist attacks on Britain was identified in a confidential French intelligence report completed days before the London bombings of July 7.

Figure 4: A snippet from the *Daily Telegraph* timeline (undated in 'the aftermath' timeline section).

Discussion

The differences between the timelines and the spike words in the London sub-corpus sheds some light on the timelines as well as on the development of discursive frames in the corpus. It is interesting that the timelines are all different in the extent of their coverage and their style even though they all aim to provide a summary of the most important phases of the event. The timeline device is quite flexible and, even for a single major event, can be used very differently. Of the categories in the classification exercise, some seem to be exclusive to the contemporary discussions whereas others are not. Apart the 'unrelated' category, the exclusively contemporary categories seem to be: non-UK English and alternative spellings; discussion of information sources; and change in terminology and language. The reason for the first is the US influence on the corpus but the second highlights the importance of finding information about a major event at the time (e.g., Bucher 2002), something that does not seem to be recognised afterwards when the initial uncertainty about the basic facts has passed. Changes in terminology and language perhaps do not seem worthy of pointing out in a timeline, so their omission is not surprising. Yet, changes in terminology and language can be used, perhaps, for automatically detecting the development of the event as reflected in the changes in the frames.

Most of the classes represent categories that were present in the timelines, but through different instances than the corpus. Emotional language was present in the timelines in the evocative quotes and comments of the *Daily Telegraph* (e.g., 'family distraught'), but much less in the other two. Figure 5 illustrates the relative frequencies of emotional words, showing that most tend to decrease in usage during the crisis. The careful management of emotion is crucial to media coverage of violent events partly because it is a device to capture audience attention in a busy media environment (Seaton 2005: 230-260), and the decrease in usage of emotional words may reflect a decreasing ability for the topic to engage the public's attention. Note that the word *innocent* was also used in a factual context, when reporting de Menezes, the man shot by the police, as not being associated with the attacks. The peak for *evil* reflects this word's usage in a widely reported speech by the UK prime minister on July 16, 2005 (Blair 2005). These two examples suggest that individual events and influential people's comments could have a significant overall effect on the tone of the discussion.
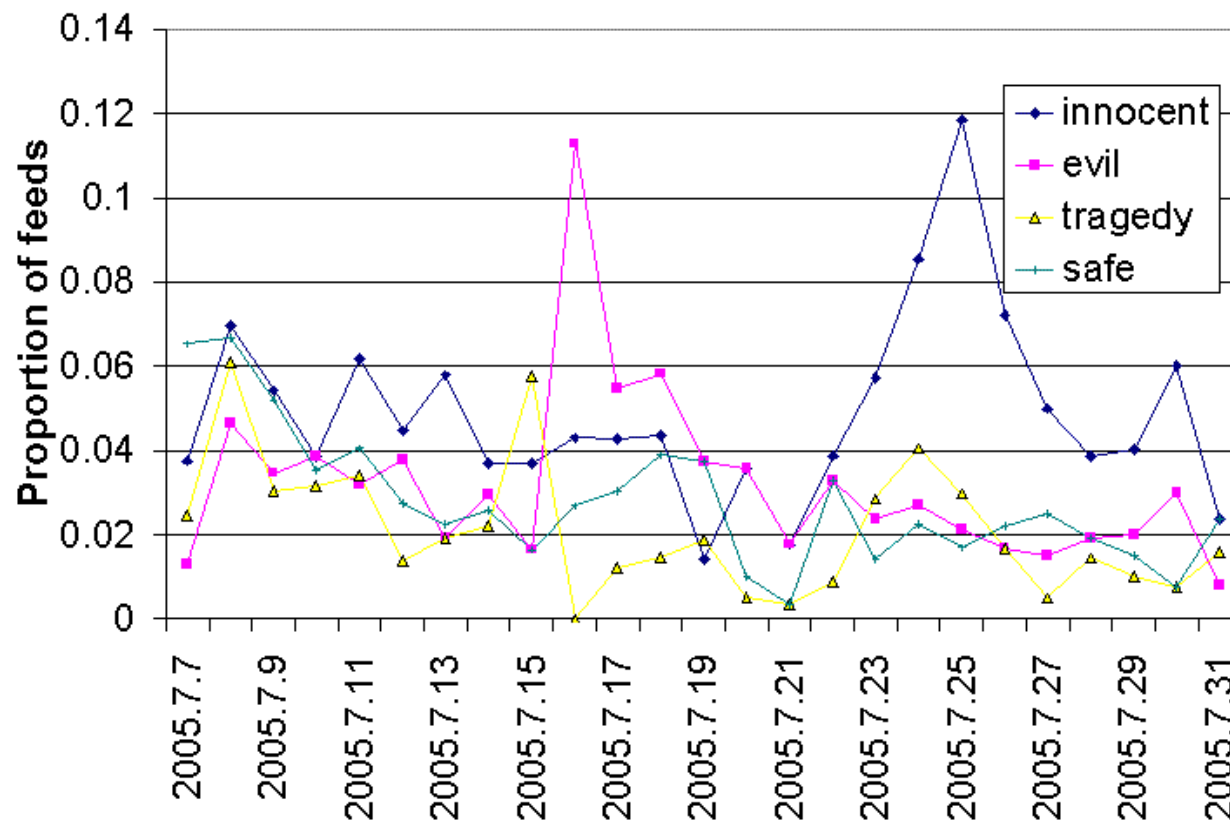
Figure 5: Time series for selected emotion-related terminology.

The media timelines reported some dead ends, perhaps to give a flavour of the uncertainty of the times. Nevertheless it is interesting that some widely-reported apparent dead ends appear to have disappeared from the record. Two of the timelines used selected events to set the London attacks in a wider context and to illustrate its repercussions, particularly the more discussion-oriented Telegraph version, but Wikinews kept to the facts of the event itself. The selected events could thus be described as 'class representatives' or 'concept markers' (Case and Higgins 2000). In semiotics, this is 'metonymy', the selection of a part to represent the whole. Metonymy has long been known as a standard device used in realist literature (e.g., Fiske 1991: 95). Hence it should not be surprising to find it in media reports of all kinds, and in blog discussions of current events.

The temporal shift in grammar was not universal because the Wikinews and BBC timelines both used the present tense, although the *Daily Telegraph* used the past tense. Many relatively specific facts about the London attacks were omitted from the timelines, but the timelines all also reported specific facts that were not significantly discussed at the time, such as place names associated with events. In this sense, the timelines give a hindsight perspective on the event while frames develop historically during the reporting. In the case of the timelines the extra details appear to be used for artistic purposes to make the timelines more readable, rather than for a reassessment of the importance of the information.

There did not seem to be any important types of information reported in the timelines but not discussed at the time, although the sub-corpus missed some key terms because of its method (e.g., Blair, Scotland). These were words that were previously significantly discussed in relation to London and so their increase in use during the crisis was not enough to trigger inclusion in the top 1,600 spike words.

From the perspective of the automatic generation of timelines the differences identified above present several problems. If based primarily upon word

frequency increases then an automatic timeline would tend to include some unrelated events, relatively minor details, obsolete terminology, emotional language and discussions of information sources. All of these might not be desirable. It seems unlikely that automatic process would be sophisticated enough to produce a readable media timeline in the sense of effectively using concept markers (metonymy) and evocative fine details. Nevertheless, it might be able to identify the main facts in conjunction with a selection of less important or irrelevant information, and perhaps for detecting the evolution of frames of the event, although automatic frame detection is probably difficult (Maher 2001). In this sense, the method may be more useful for the analysis of the anatomy of news events than for generating complete timelines.

In this paper the details of potential algorithms to create timelines have not been discussed, and these might affect the types of information that could be extracted and the overall interpretation of results. For example computational techniques such as noun and noun-phrase extraction (Mitkov 2003) could have bypassed some difference in grammar use, and as could more extensive word stemming (Porter 1980).

Finally, this study has a number of limitations for generalisation that future work may address. The raw data was an undifferentiated collection of blogs and news sites; it would be interesting to break down this data source in terms of geography, genre, and author characteristics. Our results are also based upon a single case study and only three media timelines; different events may have different characteristics. Moreover, given the flexibility of the timeline device, there may be others that deal with issues such as media coverage that were missing from our three. Nevertheless, the three timelines discussed here do all have some claim to authority and, consequently, are important examples.

Conclusion

The results suggest that there are some key differences between contemporary discussions of events and retrospective timelines. Timelines seem not to report obsolete contemporary terminology and discussions of information seeking. They probably under-report the wider context, related events elsewhere, dead-end or incorrect information, and the emotional context, but use specific facts as concept markers for these.

For computer scientists the results are disappointing. It seems that the automatic generation of timelines from blog and media sources for major events would not be able to produce something that could match the readability of a human-generated media timeline and may not be able to separate the major facts from relatively minor details that seemed important at the time. Moreover, using existing timelines as a goal with which to fine-tune an automatic system would not be ideal because of the relatively arbitrary nature of the illustrative fine details and concept marker or metonymy examples chosen for any given timeline.

From the point of view of media theory, the comparison between the spike words in the timelines and the blogs and news items is a first effort to automate the analysis of the anatomy of media debates of crises. The timelines are used to give a condensed package of information on the development of the event. Complementing this, frames are used to provide non-temporal structures that emphasise certain aspects or facts in an event. Comparisons between these two strategies for providing a larger context for events opens up a new way to analyse the discussion and media reporting of contemporary debates. Whilst the large scale analysis of heterogeneous sources (blogs and mass media) presented here can identify broad patterns, future researchers may wish to apply similar methods to more homogeneous corpora, e.g., a specific type of bloggers, or media feeds alone, and may be able to produce interesting quantitative evidence with which to cast new light on theories of communication and the impact of mass media. In particular, if blogs and news feeds can be separated in large enough numbers (tens of thousands) to apply our methods then the potential exists to explore in detail the political relationship between the blogging public and the media. For example, this might reveal the extent to which 'ordinary' non-A-list blogs support a public sphere (Habermas 1991) for political discussion that is independant of the mass media (e.g., Thompson 2003).

Acknowledgement

References

- Allen, R. B. (2005). A focus-context browser for multiple timelines. In Mary Marlino (Ed.), *5th ACM/IEEE-CS joint conference on Digital libraries, Denver, CO, USA. Proceedings* (pp. 260-261). New York, NY: ACM Press.
- British Broadcasting Corporation. (2005). *Blog reading explodes in America*. Retrieved 20 August, 2006 from http://news.bbc.co.uk/1/hi/technology/4145191.stm
- Blair, A. (2005). *Full text: Blair speech on terror*. London: BBC. Retrieved 20 August, 2006 from http://news.bbc.co.uk/1/hi/uk/4689363.stm
- Bucher, H.J. (2002). The Internet in crisis: communication in the case of September 11th. *First Monday*, **7**(4). Retrieved 20 August, 2006 from http://www.firstmonday.org/issues/issue7_4/bucher/
- Case, D. O., & Higgins, G.M. (2000). How can we investigate citation behavior? A study of reasons for citing literature in communication. *Journal of the American Society for Information Science & Technology*, **51**(7), 635-645.
- Downs, A. (1972). Up and down with ecology - the issue attention cycle. *The Public Interest*, **28**(2), 38-50.
- Entman, R. M. (1993). Framing: toward clarification of a fractured paradigm. *Journal of Communication*, **43**(4), 51-58.
- Fiske, J. (1991). *Introduction to communication studies* (2nd. ed.). London: Routledge.
- Fukuhara, T. (2005). *Analyzing concerns of people using Weblog articles and real world temporal data*, Paper presented at the WWW 2005 2nd Annual Workshop on the Weblogging Ecosystem: Aggregation, Analysis and Dynamics, Chiba, Japan. May 10th 2005. Retrieved 20 August, 2006 from http://www.blogpulse.com/papers/2005/fukuhara.pdf
- Gamson, W.A., & Modigliani, A. (1989). Media discourse and public opinion on nuclear power: A constructionist approach. *American Journal of Sociology*, **95**(1), 1-37.
- Gill, K.E. (2004). *How can we measure the influence of the blogosphere?* Paper presented at the WWW 2004 Workshop on the Weblogging Ecosystem: Aggregation, Analysis and Dynamics. New York, May 18th 2004. Retrieved 20 August, 2006 from http://faculty.washington.edu/kegill/pub/www2004_blogosphere_gill.pdf
- Gill, K. E. (2005). *Blogging, RSS and the information landscape: a look at online news*. Paper presented at the WWW 2005 2nd Annual Workshop on the Weblogging Ecosystem: Aggregation, Analysis and Dynamics, Chiba, Japan. May 10th 2005. Retrieved 20 August, 2006 from http://www.blogpulse.com/papers/2005/gill.pdf
- Glance, N. S., Hurst, M., & Tomokiyo, T. (2004). *BlogPulse: Automated trend discovery for Weblogs* Paper presented at the WWW 2004 Workshop on the Weblogging Ecosystem: Aggregation, Analysis and Dynamics. New York, May 18th 2004. Retrieved 20 August, 2006 from http://www.blogpulse.com/papers/www2004glance.pdf
- Goffman, E. (1974). *Frame analysis: an essay on the organization of experience*. New York, NY: Harper & Row.
- Gorgura, H. (2004). *The war on the terror consensus: anti-war blogs as an online sphere of dissensus*. Paper presented at AoIR 5.0, Brighton, UK, 19-22nd September 2004. [Abstract available at http://gsb.haifa.ac.il/~sheizaf/AOIR5/207.html]
- Gruhl, D., Guha, R., Liben-Nowell, D. & Tomkins, A. (2004). *Information diffusion through Blogspace.* Paper presented at the WWW 2004 Workshop on the Weblogging Ecosystem: Aggregation, Analysis and Dynamics. New York, May 18th 2004. Retrieved 20 August, 2006 from http://www.www2004.org/proceedings/docs/1p491.pdf
- Hammersley, B. (2005). *Developing feeds with RSS and Atom*. Sebastopol, CA: O'Reilly.
- Habermas, J. (1991). The public sphere. In C. Mukerji & M. Schudson (Eds.), *Rethinking popular culture: contemporary perspectives in cultural studies* (pp. 398-404). Berkeley, CA: University of California Press.
- Han, J., & Kamber, K. (2000). *Data mining: concepts and techniques*. New York, NY: Morgan Kaufmann Publishers.
- Herring, S.C., Scheidt, L.A., Bonus, S. & Wright, E. (2004). Bridging the gap: a genre analysis of Weblogs. In *R. Sprague and J. Nunamaker (Eds.) Proceedings of the Thirty-seventh Hawaii International Conference on System Sciences (HICSS-37), January 5-8, 2004, Big Island, Hawaii. Los Alamitos: IEEE Press.* Retrieved 20 August, 2006 from http://csdl.computer.org/comp/proceedings/hicss/2004/2056/04/205640101b.pdf
- Kim, J. H. (2005). *Blog as an oppositional medium? A semantic network analysis on the Iraq war blogs.* Paper presented at Internet Research 6.0: Internet Generations, October 5-9, 2005, Chicago.

- Kleinberg, J. M. (2003). Bursty and hierarchical structure in streams. *Data Mining and Knowledge Discovery*, **7**(4), 373-397.
- Krippendorff, K. (1980). *Content analysis: an introduction to its methodology*. Beverly Hills, CA: Sage.
- Kumar, R., Novak, J., Raghavan, P. & Tomkins, A. (2003). *On the bursty evolution of blogspace.* Paper presented at WWW2003, Budapest, Hungary, 20-24 May 2003. Retrieved 20 August, 2006 from http://www2003.org/cdrom/papers/refereed/p477/p477-kumar/p477-kumar.htm.
- Lin, J. & Halavais, A. (2004). *Mapping the blogosphere in America.* Paper presented at WWW 2004 Workshop on the Weblogging Ecosystem: Aggregation, Analysis and Dynamics. New York, May 18th 2004. Retrieved 20 August, 2006 from http://www.blogpulse.com/papers/www2004linhalavais.pdf
- Maher, T.M. (2001). Framing: an emerging paradigm or a phase of agenda setting. In S. D. Reese, O. H. Gandy & A. E. Grant (Eds.), *Framing public life: perspectives on media and our understanding of the social world*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Mani, I. (2001). *Automatic summarization*.Amsterdam: John Benjamins.
- Mitkov, R. (2003). *The Oxford handbook of computational linguistics*.Oxford: Oxford University Press.
- Pikas, C. K. (2005). Blog searching for competitive intelligence, brand image, and reputation management. *Online*, **29**(4), 16-21.
- Porter, M. (1980). An algorithm for suffix stripping. *Program*, **14**(3), 130-137.
- Prabowo, R. & Thelwall, M. (2006). A comparison of feature selection methods for an evolving RSS feed corpus. *Information Processing & Management*, **42**(6), 1491-1512
- Scheufele, D. A. (1999). Framing as a theory of media effects. *Journal of Communication*, **49**(4), 103-122.
- Seaton, J. (2005). *Carnage and the media.* London: Allen Lane.
- Society of Professional Journalists. (1996). *Code of ethics.* Indianapolis, IN: Society of Professional Journalists. Retrieved 20 August, 2006 from http://www.spj.org/ethicscode.asp?
- Swan, R. & Allan, J. (2000). Automatic generation of overview timelines. In E. Yannakoudakis, N. J. Belkin, M.-K. Leong & P. Ingwersen (Eds.), *Proceedings of the 23rd annual international ACM SIGIR conference on research and development in information retrieval, July 24-28, 2000, Athens, Greece.* (pp. 49-56). New York, NY: ACM Press.
- Thelwall, M. (2006). *Bloggers during the London attacks: top information sources and topics*. Paper presented at the WWW 2006 3rd Annual Workshop on the Weblogging Ecosystem: Aggregation, Analysis and Dynamics, Edinburgh, Scotland. May 23rd 2006. Retrieved 20 August, 2006 from http://www.blogpulse.com/www2006-workshop/papers/blogs-during-london-attacks.pdf
- Thelwall, M., Prabowo, R. & Fairclough, R. (forthcoming). Are raw RSS feeds suitable for broad issue scanning? A science concern case study. *Journal of the American Society for Information Science and Technology*.
- Thelwall, M. & Wouters, P. (2005). What's the deal with the Web/Blogs/the next big technology: a key role for information science in e-social science research? In Fabio Crestani and Ian Ruthven, (Eds.). *Information Context: Nature, Impact, and Role 5th International Conference on Conceptions of Library and Information Sciences, CoLIS 2005, Glasgow, UK, June 4-8, 2005 Proceedings.* (pp. 187-199). Berlin: Springer. (Lecture Notes in Computer Science, 3507). Retrieved 20 August, 2006 from http://www.springerlink.com/content/kgkcc88h0230503q/fulltext.pdf
- Thompson, G. (2003). Weblogs, warblogs, the public sphere, and bubbles. *Transactions* 7(2). Retrieved 20 August, 2006, from http://transformations.cqu.edu.au/journal/issue_07/article_02.shtml.
- Weare, C. & Lin, W. Y. (2000). Content analysis of the World Wide Web—opportunities and challenges. *Social Science Computer Review*, **18**(3), 272-292.

How to cite this paper

Thelwall, M. & Hellsten, I. (2006). "The BBC, *Daily Telegraph* and Wikinews timelines of the terrorist attacks of 7th July 2006 in London: a comparison with contemporary discussions" *Information Research*, 12(1) paper 84 [Available at http://InformationR.net/ir/12-1/paper284.html]

Find other papers on this subject

Contents | **Author index** | **Subject index** | **Search** |
Home