Extracting Event-Related Information from Article Updates in Wikipedia

Mihai Georgescu, Nattiya Kanhabua, Daniel Krause, Wolfgang Nejdl, and Stefan Siersdorfer

L3S Research Center, Appelstr. 9a, Hannover 30167, Germany

Abstract. Wikipedia is widely considered the largest and most up-to-date online encyclopedia, with its content being continuously maintained by a supporting community. In many cases, real-life events like new scientific findings, resignations, deaths, or catastrophes serve as triggers for collaborative editing of articles about affected entities such as persons or countries. In this paper, we conduct an in-depth analysis of event-related updates in Wikipedia by examining different indicators for events including language, meta annotations, and update bursts. We then study how these indicators can be employed for automatically detecting event-related updates. Our experiments on event extraction, clustering, and summarization show promising results towards generating entity-specific news tickers and timelines.

1 Introduction

Wikipedia is a free multilingual online encyclopedia covering a wide range of general and specific knowledge in about 23 million articles (∼4 million in the English version). It is continuously kept up-to-date and extended by a community of over 100,000 contributors, with an average of 3.5 million edits *per month* observed in 2011.[1] One of the reasons that drives editing and updating in Wikipedia is the occurrence of new events in the real world such as elections, accidents, political conflicts, or sport events. In the context of a political argument between the US president Obama and the Republican Wilson, which immediately lead to a burst of edits and discussions in Wikipedia, the New York Times wrote: "If journalism is the first draft of history, what is a Wikipedia entry when it is updated within minutes of an event to reflect changes in a person's biography?"[2]. As another example, Figure 1 shows typical updates as well as a plot depicting the burst of edits triggered by Rumsfeld's resignation in November 8, 2006.

Wikipedia articles and associated edits constitute a potentially interesting data source to mine for obtaining knowledge about real-world events. In this paper, we conduct a study on this information with several complementary goals. On the one hand, we study the viability of using the edit history of Wikipedia for extracting event-related updates. This has direct applications to building annotated timelines and news tickers for specific entities featured in Wikipedia

[1] http://en.wikipedia.org/wiki/Special:Statistics

[2] http://bits.blogs.nytimes.com/2009/09/10/the-wikipedia-battle-over-joe-wilsons-obama-heckling/

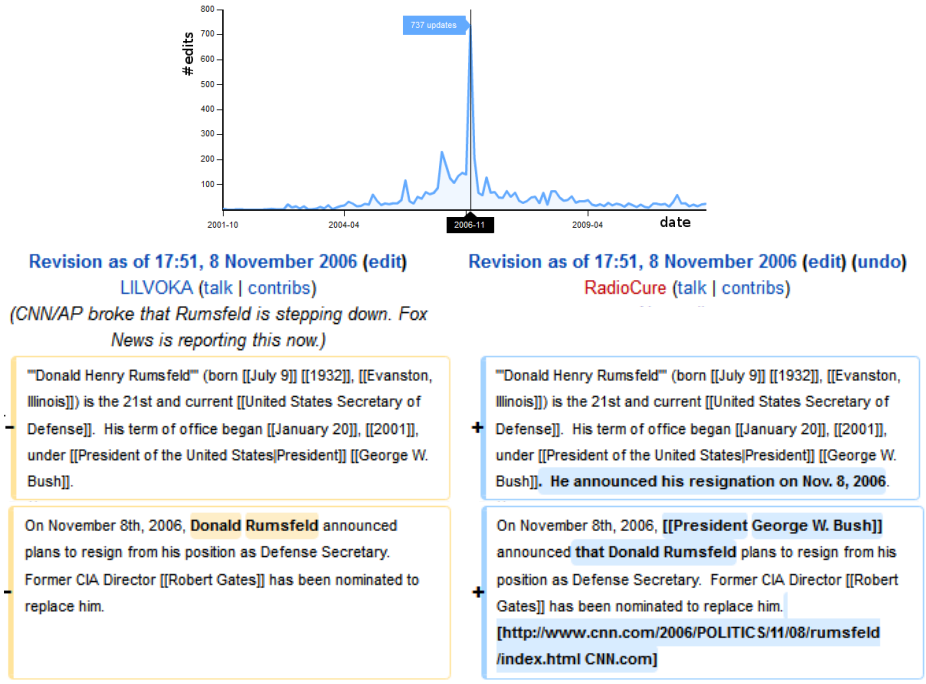


Fig. 1: On November 8, 2006, the resignation from the U.S. Secretary of Defense of Donald Rumsfeld, caused a burst of updates. Two event-related updates are shown, and contributors, timestamps, comments, and the differences of two revisions highlighted.

articles such as persons and countries. On the other hand, we perform an in-depth analysis of event-related updates in Wikipedia, including qualitative and quantitative studies for sets of samples gathered using different filtering mechanisms. How many updates in Wikipedia are related to events? Is there a connection between bursts of edits and real-life events? Are there indicators for event-related updates in the textual content and meta annotations of the Wikipedia edits? Can we automatically detect event-related updates? These are some of the questions we investigate in this paper by analyzing Wikipedia's publicly available edit history.

For extracting event-related information from Wikipedia edits, we first identify event-related updates; then we cluster these updates in order to map the updates to their corresponding events and to generate summaries (cf. Figure 2). In order to identify event-related updates we employ different filters and extraction methods. First, we apply burst detection because events of interest tend to trigger peaks of attention for affected entities. Date detection helps to identify event-related updates that contain dates in the proximity of the update creation time. Finally, we build classification models based on the textual content of the updates as well as meta annotations. To summarize event-related information, we perform clustering of edits by exploiting different types of information such as update time, textual similarity, and the position of edits within an article.

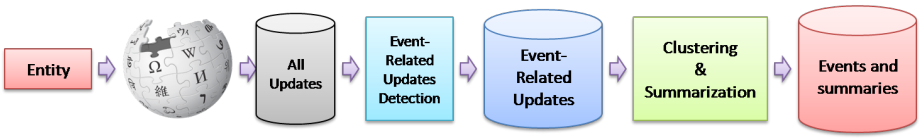


Fig. 2: Pipeline for identifying and presenting the events related to an entity.

2 Event Extraction Methods

An *update* in Wikipedia represents the modifications present in one revision when compared to the previous revision of an article. It is accompanied by its creation time (timestamp), its author, and, possibly, comments provided by the updater. For a given update, we further consider the blocks of text added and removed, the title of the section where the modification occurred, and the relative and absolute positions of the blocks in their sections and in the article.

In order to extract event-related information from Wikipedia edits for a given entity and its corresponding article, we first identify event-related updates; in a second step we cluster these updates in order to map the updates to their corresponding events and to generate summaries. The pipeline for this process is depicted in Figure 2. In the following subsections we describe the methods we employ for event-related update detection and summarization.

2.1 Detection of Event-Related Updates

For detecting event-related updates we make use of a combination of filters and classifiers based on burst detection, temporal information, and textual content.

Burst Detection Filter: Bursts of updates (peaks in the update activity) in a Wikipedia article are indicators for periods with an increased level of attention from the community of contributors. As we will discover later in our analysis in Section 4, bursts often co-occur with real-life events, making burst detection a promising filter for gathering event-related updates. In order to detect bursts, we apply a simplified version of the burst detection algorithm presented in [21] on the temporal development of the update frequency of an article. The algorithm employs a sliding time window for which the number of updates is counted. The corresponding time intervals for which the update rate exceeds a certain threshold are considered *bursty*; our burst detection filter extracts the updates within those bursty periods. The parameters of the algorithm are ω - the size of the sliding window (e.g., day, week, or month), and θ - a threshold for the number of standard deviations above the average update number over the whole lifetime of the article for a time interval to be considered as bursty.

Date Extraction Filter: This filter makes use of the following heuristic: If the textual content of the update contains a date which is in close temporal proximity to the timestamp of the update, then this is an indicator that the update might be connected to an event. More specifically, our filter identifies temporal expressions in updates matching the format recommended by Wikipedia[3], and checks if these expressions fall into the interval within one month before or after the update was done.

[3] http://en.wikipedia.org/wiki/Wikipedia:Manual_of_Style/Dates_and_numbers

Text Classification: Language and terms used in the update text can serve as an indicator whether an update is related to an event. For instance, we observed that terms like *death*, *announce*, and *outburst* are typical for event-related updates. In addition, Wikipedia updates are often accompanied with meta annotations such as "{current}" (explicitly marking current events) or "rvv" in comments (indicating vandalism rather than events) which can provide additional clues on the event-relatedness of updates. In order to exploit that type of information we trained Support Vectors Machine classifiers [6] on manually labeled samples to distinguish between "event-related" and "not event-related" updates. We tested different bag-of-words based feature vector representations of updates, which will be described in more detail in Section 5.

2.2 Clustering and Summarization of Event-related Updates

The stream of event-related updates determined in the previous step serves as a starting point for identifying the events themselves and creating a meaningful summarization. In order to present event-related information in a understandable way, instead of using the detected event-related updates for summarization, we use the sentences that were modified by them. To this end, we start by identifying the sentences where the event-related updates were done, and assign to them a weight, corresponding to the number of times they were updated, and a list of positions at which the sentences appeared within the Wikipedia articles.

Temporal Clustering: As already observed in Section 2.1 events are signaled in Wikipedia by a burst of updates. Therefore, in order to identify the distinct events, we first resort to a temporal clustering by identifying the bursts among the event-related updates. Each burst of event-related updates corresponds to a distinct event.

Text-Based Clustering: Within a burst of updates, in order to eliminate the duplicate sentences and group together the sentences that treat the same topic we employ an incremental clustering based on the Jaccard similarity as a distance measure. Each *sentence cluster* is characterized by the *aggregated weight* of member sentences, and represented by the longest member sentence, that serves as a candidate for summarization.

Position-Based Clustering: Assuming that sentences that treat the same topic are located in spatial proximity of each other on the article page, by investigating the positions of all sentences modified in a burst we can identify *position clusters*. Each sentence cluster belongs to the position cluster that has the maximum overlap of positions with member sentences.

Summarizing Detected Events: Each identified event, corresponding to a burst of updates is summarized using a ranked list of sentences. We rank the position clusters by how many sentence clusters are assigned to them, ignoring the position clusters that are not well represented and we rank the sentence clusters by the *aggregated weight* of their member sentences. The proposed summarization for an individual event consists of displaying for each of the top-N identified position clusters, the representative sentences for the top-M clusters of sentences.

3 Datasets

We downloaded the dump of the whole Wikipedia history (version from 30 January 2010). The history dump contains more than 300 million updates with the size of approximately 5.8 TB covering the time period between 21 January 2001 and 30 January 2010. We discarded updates made by *anonymous* users, resulting in a dataset containing 237 million updates belonging to 19 million articles. In this work, we studied our proposed method for extracting event-related information using different datasets created by randomly selecting Wikipedia updates for: 1) articles from all categories, and 2) only those belonging to *people* category. Note that, we discarded all the articles that had less than 1,000 updates.

By considering articles from all categories, we can investigate the domains on which our proposed methods can be applied without any limitation on some particular types of articles. In this case, we sampled updates in three ways:

- **ALL-Random** was collected by randomly sampling from all available updates in our history dump collection.
- **ALL-Burst** was collected taking into account the time dimension by sampling updates coming from bursts, where bursts were identified by using the detection algorithm described in Section 2.1 with the empirically chosen parameters $\omega = 2$ days and $\theta = 4$.
- **ALL-Date** was gathered using a constraint in which article updates contain at least one *date mention* in the proximity of their timestamps. More precisely, we checked whether the month and year of timestamps occurred inside the text added, removed or inside the comments. This dataset was also selected from burst periods determined using $\omega = 2$ days and a higher $\theta = 32$ in order to filter just the updates done in highly salient bursts and to increase the chances of finding event-related updates.

In addition to the selection methods described above, we investigated updates of Wikipedia articles from the category *people* in particular because the updating of personal information is highly relevant to some events, e.g., professional achievement, changing of civil status, or health issues. We randomly selected 185 Wikipedia articles, whose categories start with "peopl" and contain at least a burst of updates. In detail, we sampled updates in three ways:

- **PPL-Burst** was created by randomly selecting 10 updates for each article coming from the identified bursts using $\omega = 2$ days and $\theta = 12$. The parameters of the burst detection algorithm were chosen in order to offer a reasonable number of candidates to sample from.
- **PPL-Date** was collected by randomly choosing 10 updates for each article with dates in the vicinity of their timestamps, i.e., in the window of one month before/after timestamps. Date mentions were identified by looking for date mentions in the standard formats provided by Wikipedia. Note that, we filtered out date mentions found in an administrative context because they might not be related to events.
- **PPL-Random** was created by randomly selecting 10 updates for each article without considering bursts or containing date mentions close to their creation timestamps.

Our last dataset, denoted **DETAIL**, was created by selecting four particular entities: Jerry Fallwell, Donald Rumsfeld, Alexandr Solzhenitsyn and Kosovo. Each of those entities is associated to one or more important events, and we aimed at performing a detailed analysis of bursts. For each article, we used all updates from bursts identified using the narrower parameter choice, $\omega = 2$ days and $\theta = 32$, in order to perform further investigation of update dynamics.

4 Data Analysis

In this section, we perform an in-depth analysis of event-related updates in Wikipedia gathered using the different filtering mechanisms as explained in the previous section.

4.1 Data Labeling

There exists no ground truth dataset for evaluating the task of event extraction from Wikipedia updates. In order to identify which of the updates are related to events we therefore manually labeled the updates in the datasets described in the previous section. More precisely, for each article update we provided a human assessor with the *differences* (i.e., text added or removed) between the revision before and after the update using Wikipedia's *diff* tool [4]. In addition, we provided the *comment* made by the editor of an update as additional context. The human assessor was asked to assign one of the following labels to each update: 'event-related" or "not event-related". The updates on which the assessor was unsure about, were discarded in the experiments and analysis. Vandalizing updates were regarded as *not event-related*. For the event-related updates, we also determined whether they were *controversial* or not. An update was considered as controversial if it: 1) contained a point of view, 2) was repeatedly added and removed, and 3) exhibited a dispute between the contributors. These annotations help to understand the effect of controversy in the process of updating an article in the case of an event, and show how many of the event related updates are likely to be disputed. In order to gain further insight into the types of edits that occur during bursty periods, we performed a detailed investigation by categorizing them into the following classes: *fact* (modifying facts presented in the article), *link* (adding/removing links within or outside Wikipedia), *markup* (changing cosmetic appearance or Wikipedia markup), *vandalism* (vandalizing of an update), *spelling* (editing punctuation, spelling or formulation of facts without modification), and *category* (changing the category of a Wikipedia article). Finally, there were approximately 10,000 article updates labeled and the dataset is publicly available for download[5].

4.2 Data Statistics

Table 1 shows statistics of our datasets including the total number of labeled updates, the number of *event-related* updates (number of *controversial* updates in

[4] `http://en.wikipedia.org/w/index.php?diff=prev&oldid=[REVISION_NO]`

[5] `http://www.l3s.de/wiki-events/wiki-dataset.zip`.

parentheses), and the number of *non event-related* updates (number of *vandalizing* updates in parentheses). We observe that filtering by bursts increases the number of event-related updates found. The percentage of event-related updates for **ALL-Burst** increases up to 10% compared to just 1% for **ALL-Random**. The burst detection increases the number of event-related updates from 3% in **PPL-Random** to 11% in **PPL-Burst**, amplified to 41% in **PPL-Date**. We further observe a substantial increase in the number of *event-related* updates when filtering by date mentions. For the **ALL-Date** dataset, 66% of the updates are related to events, and 30% of those are controversial. More event-related updates took place during bursty periods showing that burst detection helps increasing the percentage of event-related updates while reducing the overall number of updates to choose from. This effect can be further amplified by using date filtering. The number of vandalism updates is steady across our samples with a slight increase in the case of the **ALL-Date** and **PPL-Date** samples.

Figure 3 illustrates the percentage of updates labeled into different classes for **ALL-Random** and **ALL-Burst**. We can observe differences between updates made in general and updates made during bursty periods. The samples taken from the detected bursts contain substantially more updates related to facts rather than changing the cosmetic appearance and style of the articles.

Dataset	Updates	Event-related (Contro.)	Unrelated (Vandalism)
ALL-Random	961	13(0)	948(63)
ALL-Burst	1331	133(21)	1198(141)
ALL-Date	1626	1037(256)	589(51)
total	*3918*	*1183(277)*	*2735(255)*
PPL-Random	1850	62	1788(329)
PPL-Burst	1850	199	1651(159)
PPL-Date	1448	604	844(310)
total	*5148*	*865*	*4283(798)*
DETAIL	1614	568(280)	1046(108)

Table 1: Statistics of datasets.

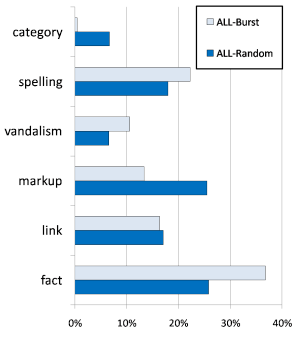
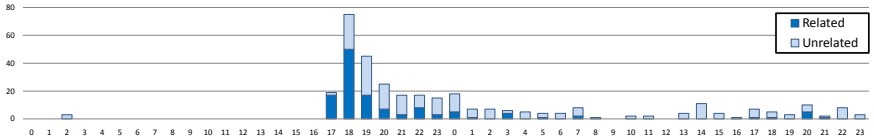


Fig. 3: Classes of updates.

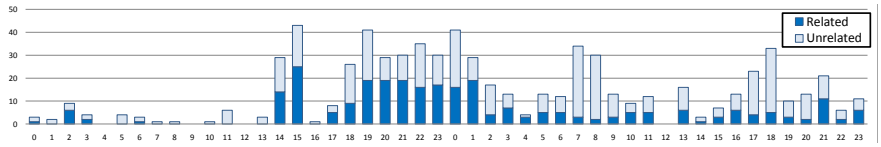
4.3 Investigating the Burst of Updates

We investigated the updates made on the four articles in the **DETAIL** dataset in order to better understand the process of event-triggered updating. Figure 4 shows the distribution over time of the number of updates for the Wikipedia articles on Donald Rumsfeld and Kosovo. For every hour of the day since the beginning of the burst, we plot the number of updates composed of the event-related and not event-related updates. We observe that *not* all of the updates done during a burst period are related to an event. After a burst, the updates are no longer related to the events; instead, the attention is rather directed towards making the article more accurate, giving raise to correction of unrelated facts, punctuation and cosmetic changes. For the resignation of Donald Rumsfeld, we notice that the burst of updates contains a small number of peaks, which are

bigger at the beginning of the event and then become smaller as the overall number of updates and the number of event-related updates decrease towards the end of the burst. This might be a characteristic of the type of event or entity. If the event is not controversial, or no other information becomes available, the interest in editing the article drops. In contrast, if the entity or the event is controversial or the event develops over a longer period of time, as in the case of Kosovo's independence declaration, the interest decreases much slower.



(a) **Donald Rumsfeld** and the corresponding event of *resignation*.



(b) **Kosovo** and the corresponding event of *independence declaration*.

Fig. 4: Distribution over time (in hours) of updates for two Wikipedia articles: Donald Rumsfeld and Kosovo.

4.4 Discriminative Term Analysis

In order to assess the feasibility of building a term-based classifier, we studied the differences between the terms used in event-related updates and non event-related updates by conducting a discriminative term analysis . For computing ranked lists of stemmed terms from the set of event-related updates, and the updates unrelated to events, we used the information-theoretic Mutual Information (MI) measure [15]. Table 2 shows the top-20 stemmed terms computed from the datasets containing a sufficient number of event-related updates. For all of the updates we considered words added and removed, as well as words from comments and meta annotations denoting the type of the update. We observe that time-related terms (*date, time, current*), sports-event related terms (*championship, sport, schedul*), news-related terms (*news, announc, publish, releas, stori, report*) or status change terms (*die, death, outburst*) characterize the event-related updates as opposed to Wikipedia administrative terms (*sysop, delet, wikifi, page*) or general terms (*common, street, king, power*) that characterize updates that are unrelated to events.

5 Evaluation of Event-Related Information Extraction

In this section, we investigate more closely the components of the pipeline described in Section 2, by evaluating methods for event-based classification as the final step in the detection of event-related updates and presenting some examples of extracted and summarized events.

Table 2: Top (stemmed) terms ranked by MI values for two types of updates.

Dataset	Event-related Terms	Not Event-related Terms
PPL-Date	2006 second state schedul date add championship announc time releas presid report current year publish contract news titl sport web	2007 2004 sysop delet excess 18 use 15 juli protect expir march level wp:vandal decemb expiri autoconfirmed:mov 22 edit utc
ALL-Date	reaction stori 2009 2006 2007 state bhutto 12 report die presidenti wilson decemb obama www.cnn.com 08 news death outburst septemb	squar common tavistock street use wikifi pancra king bma network destroy life page fix name woburn power edgwar terrorist russel april

5.1 Event Classification

For text-based classification of updates into categories "event-related" and "not event-related" we used the LIBSVM [4] implementation of linear support vector machines (SVMs) with the default parameters.

We conducted our evaluation on **ALL-Burst**, **ALL-Date**, **PPL-Burst**, and **PPL-Date** as these datasets contain a sufficient number of event-related updates for experiments (cf. Section 4). We experiment with different feature representations of the updates. If some of these feature representations generate empty documents, they are excluded from the experiments. To avoid an imbalance towards one category or the other, for our experiments we randomly chose a number of instances from the bigger category equal to the number of instances contained in the smaller category. For testing the classification performance on the thus generated balanced datasets we used 5-fold cross-validation. We repeated this procedure 100 times and averaged over the results.

Our quality measures are precision, recall as well as the break-even points (BEPs) for precision-recall curves (i.e. precision/recall at the point where precision equals recall, which is also equal to the F1 measure, the harmonic mean of precision and recall in that case). We also computed the area under the ROC curve values (AUC) [7]. ROC (Receiver Operating Characteristics) curves depict the true positive with respect to the false positive rate of classifiers.

We compared the following update representations for constructing bag-of-word based tf*idf feature vectors (using stemming and stop word elimination for each of the options):

- *wordsAdd* - terms added in an update
- *wordsRmv* - terms removed in an update
- *All* - terms added in an update, terms removed, and terms from comments
- *P.text* - terms from text added and removed treating added and removed terms as different dimensions in the feature vector
- *P.all* - terms added in an update, terms removed, and terms from comments treating added, removed, and comment terms as different dimensions in the feature vector
- *P.T.all* - *P.all* with the titles of the updated sections as additional context

Table 3 shows the results of our experiments. We achieve the best performance for the feature representation using a combination of terms added in an

Table 3: Classification performance using different textual representations.

Features	ALL-Burst				ALL-Date				PPL-Burst				PPL-Date			
	AUC	BEP	P	R	AUC	BEP	P	R	AUC	BEP	P	R	AUC	BEP	P	R
wordsAdd	.75	.69	.77	.53	.76	.70	.72	.58	.75	.69	.77	.53	.77	.71	.77	.57
wordsRmv	.78	.72	.82	.53	.70	.66	.69	.56	.78	.72	.82	.53	.73	.67	.70	.62
All	.80	.73	.80	.54	.80	.74	.78	.61	.80	.73	.80	.54	.87	.79	.81	.78
P.text	.75	.68	.78	.51	.75	.69	.72	.58	.75	.68	.78	.51	.77	.71	.76	.60
P.all	.76	.69	.77	.51	.77	.71	.74	.58	.76	.69	.77	.51	.86	.79	.78	.82
P.T.all	.73	.67	.74	.47	.72	.68	.71	.52	.73	.67	.74	.47	.81	.74	.70	.89

update, terms removed, and terms from comments (*All*), with an AUC value of 0.87 and a BEP value of 0.79.

5.2 Clustering and Summarization of Event-Related Updates

Table 4 shows some example outputs of the clustering and summarization step described in Section 2.2. For each event we show its date and the top-2 sentence cluster representatives along with the cluster weight. For Paul Newman the event detected is his death. Most of the edits occurred in the introduction of his Wikipedia entry, where contributors added his death date. The high number of edits is due to the sentence having been added and removed several times until a trusted source confirmed the information. The second sentence provides more details about his death. For Donald Rumsfeld the most frequently edited sentence is the announcement of his planned resignation, and the second most frequently edited one is related to the nomination of a successor and includes a link to the mainstream media. For Charlie Sheen the summarized event that drew the attention of the Wikipedia community is his provocative comment on the 9/11 attacks.

Table 4: Examples of extracted and summarized events.

Entity	Event date	Weight	Representative Sentence
Charlie Sheen	12 September 2009	26	Days before the eight anniversary of the 9/11 attacks, Sheen publicly requested a meeting with President Obama to discuss a list of 20 questions he had about the September 11th attacks which he says remain unanswered and is demanding an investigation into the attacks be reopened
Charlie Sheen	12 September 2009	19	On September 8, 2009, Sheen released an open letter to President Barack Obama outlining his concerns and questions relating to a possible new investigation into the WTC attack.
Paul Newman	27 September 2008	9	"'Paul Leonard Newman"' (January 26, 1925 - September 26, 2008)
Paul Newman	27 September 2008	5	On September 26, 2008, Newman died at his longtime home in Westport, Connecticut, of complications arising from cancer
Donald Rumsfeld	8 November 2008	13	On November 8th, 2006, the GOP announced that Rumsfeld plan to resign from his position as Defense Secretary.
Donald Rumsfeld	8 November 2008	11	President Bush has nominated Robert Gates, former head of the CIA, to replace Rumsfeld http://www.cnn.com/2006/POLITICS/11/08/ rumsfeld.ap/index.html

6 Related Work

Event detection has been applied in many contexts including topic detection and tracking [2, 10, 13], tracking of natural disasters [20], and event-based epidemic intelligence [3, 9]. Previous work has focused on detecting events from unstructured text like news, using features such as key words or named entities. In this work, we employ Wikipedia article updates for event detection instead of using traditional news streams. We show that crowd behavior of editing provides strong indicators for events, and enables focused detection of events connected to a *particular entity* by analyzing the corresponding Wikipedia article.

There is a variety of applications leveraging information from Wikipedia - see Medelyan et al. [16] for a survey. Adler et al. [1] make use of the edit history to estimate the reputation of contributors. Nunes et al. [17] generate term clouds over edits made in a particular time period in order to visualize the evolving popularity of different topics. In [19] machine learning techniques are applied for detecting vandalism in Wikipedia. In the context of retrieval in document archives, Kanhabua and Nørvåg [11] extract time-based synonyms (i.e., terms semantically related to a named entity in a particular time period) from the Wikipedia history, and employ these synonyms for query reformulation. In contrast, in this work we focus on extracting and summarizing *events*.

There is preliminary work on detecting events using Wikipedia. In the earliest work studying the link between Wikipedia and news events [14], the author noticed that exposure through press citation results in an increasing amount of traffic for articles. Ciglan and Nørvåg [5] proposed to detect events by analyzing trends in page view statistics. Osborne et al. [18] propose to use Wikipedia page views for improving the quality of *first story detection* in Twitter data streams. In their recent work, Keegan et al. [12] studies the temporal dynamics of editorial patterns of news events using structural analysis, while Ferron and Massa [8] proposed different representations of events related to disasters by analyzing language usage. However, none of the aforementioned works makes use of *Wikipedia updates*, and, to the best of our knowledge, we are the first to study and analyze the edit history of Wikipedia in the context of event detection.

7 Conclusions

We conducted an in-depth analysis of Wikipedia to shed some light on how real-world events such as political conflicts, natural catastrophes, and new scientific findings are mirrored by article updates in Wikipedia. To this end, we gathered and annotated random samples from Wikipedia updates as well as samples obtained using various filters, in order to investigate different characteristics of the Wikipedia edit history. We found that events are correlated with bursts of edits, identified connections between events and language as well as meta annotations of updates, and showed that temporal information in edit content and from timestamps can provide clues on the event-relatedness of updates. The results of our experiments on automatic extraction and summarization of events from Wikipedia updates are promising, with possible applications including the construction of entity-specific, annotated timelines and news tickers.

Acknowledgments

This work was partially funded by the European Commission FP7 under grant agreements No. 287704 and No. 600826 for the CUBRIK and ForgetIT projects respectively.

References

1. B. T. Adler and L. de Alfaro. A content-driven reputation system for the Wikipedia. In *Proceedings of WWW '07*, 2007.
2. J. Allan, R. Papka, and V. Lavrenko. On-line new event detection and tracking. In *Proceedings of SIGIR '98*, 1998.
3. E. Aramaki, S. Maskawa, and M. Morita. Twitter catches the flu: Detecting influenza epidemics using Twitter. In *Proceedings of EMNLP '11*, 2011.
4. C.-C. Chang and C.-J. Lin. LIBSVM: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology*, 2:27:1–27:27, 2011. Software available at `http://www.csie.ntu.edu.tw/~cjlin/libsvm`.
5. M. Ciglan and K. Nørvåg. WikiPop: personalized event detection system based on Wikipedia page view statistics. In *Proceedings of CIKM '10*, 2010.
6. C. Cortes and V. Vapnik. Support-vector networks. *Mach. Learn.*, 20(3):273–297, 1995.
7. T. Fawcett. An introduction to ROC analysis. *Pattern Recogn. Lett.*, 27:861–874, June 2006.
8. M. Ferron and P. Massa. Psychological processes underlying wikipedia representations of natural and manmade disasters. In *Proceedings of WikiSym '12*, 2012.
9. M. Fisichella, A. Stewart, K. Denecke, and W. Nejdl. Unsupervised public health event detection for epidemic intelligence. In *Proceedings of CIKM '10*, 2010.
10. Q. He, K. Chang, and E.-P. Lim. Analyzing feature trajectories for event detection. In *Proceedings of SIGIR '07*, 2007.
11. N. Kanhabua and K. Nørvåg. Exploiting time-based synonyms in searching document archives. In *Proceedings of JCDL '10*, 2010.
12. B. Keegan, D. Gergle, and N. Contractor. Staying in the loop: Structure and dynamics of wikipedia's breaking news collaborations. In *Proceedings of WikiSym '12*, 2012.
13. Z. Li, B. Wang, M. Li, and W.-Y. Ma. A probabilistic model for retrospective news event detection. In *Proceedings of SIGIR '05*, 2005.
14. A. Lih. Wikipedia as participatory journalism: Reliable sources? metrics for evaluating collaborative media as a news resource. In *the 5th International Symposium on Online Journalism*, 2004.
15. C. Manning and H. Schütze. *Foundations of statistical natural language processing*. MIT Press, 1999.
16. O. Medelyan, D. Milne, C. Legg, and I. H. Witten. Mining meaning from Wikipedia. *Int. J. Hum.-Comput. Stud.*, 67:716–754, September 2009.
17. S. Nunes, C. Ribeiro, and G. David. WikiChanges: exposing Wikipedia revision activity. In *Proceedings of WikiSym '08*, 2008.
18. M. Osborne, S. Petrovic, R. McCreadie, C. Macdonald, and I. Ounis. Bieber no more: First story detection using Twitter and Wikipedia. In *SIGIR 2012 Workshop on Time-aware Information Access (TAIA'12)*, 2012.
19. M. Potthast, B. Stein, and R. Gerling. Automatic vandalism detection in wikipedia. In *Proceedings of ECIR '08*, 2008.
20. T. Sakaki, M. Okazaki, and Y. Matsuo. Earthquake shakes Twitter users: real-time event detection by social sensors. In *Proceedings of WWW '10*, 2010.
21. Y. Zhu and D. Shasha. Efficient elastic burst detection in data streams. In *Proceedings of KDD '03*, 2003.