Real-time monitoring of sentiment in business related Wikipedia articles

Finn Årup Nielsen
DTU Compute, Technical University of Denmark
Kongens Lyngby, Denmark
faan@dtu.dk

Michael Etter
Copenhagen Business School
Frederiksberg, Denmark
me.ikl@cbs.dk

Lars Kai Hansen
DTU Compute, Technical University of Denmark
Kongens Lyngby, Denmark
lkai@dtu.dk

ABSTRACT

We present an online service with real-time monitoring of Wikipedia pages for companies and detects sentiment with respect to the edits, the companies and editors. It monitors the IRC stream, detects company-related articles using a small hand-built list and performs sentiment analysis using a sentiment-annotated word list. The system generates a report that can be emailed to users.

Categories and Subject Descriptors

[**Information Systems Applications**]: Collaborative and social computing systems and tools—*Reputation systems, Wikis*

General Terms

Measurement, Human Factors

Keywords

Wikipedia, Sentiment analysis

1. INTRODUCTION

Wikipedia is widely used and appears frequently in the top results in search engines. Business-related Wikipedia articles, thus, are quite visible and are considered important for the public image of companies [3, 2]. Companies, most of their stakeholders and Wikipedia editors have an interest in high information quality on Wikipedia pages on companies, e.g., in terms of accuracy and unbiasedness. Companies may want to improve articles about themselves, but Wikipedia policies[1] discourage conflict of interest (COI) editing on Wikipedia as COI edits may be biased in favor of the company. On the other hand activists, disgruntled employees and other critical editors may also have the potential for biased edits, since it is often their aim to portray the respective company in a bad way. Given the limited technical abilities in the MediaWiki software running Wikipedia to ear mark the insertion of biased content, we here consider real-time monitoring of Wikipedia targeted at detecting potential biased content.

[1]English Wikipedia:Conflict of interest (547005788)

Wikipedia anti-vandalism bots already perform real-time monitoring. Prolific *ClueBot NG* are reported to make tens of thousands vandalism reversions a months [7], and the semi-automated STiki system may catch vandalism that automated systems (that require high precision) fails to detect [14]. However, we are not interested in mere vandalism, but rather well-done and competent edits that nevertheless show a slant to one side or another and may not be detected by standard bots and, hence persists.

There are at least three groups of researchers that have analyzed (potentially) biased (or "slanted") edits on Wikipedia [1, 4, 3, 5]. The WikiScanner by Virgil Griffith allowed web-users to more quickly identify COI edits than would have been possible by browsing through the Wikipedia history pages. It used "whois" information from IP addresses and can thus only work on "anonymous" Wikipedia edits (although their "Poor man's check user" may link IP addresses with non-anonymous user names for some users). The WikiScanner enabled news media to report a whole series of different COI edits from a wide range of companies and organizations. Chandy, a student of Griffith, applied a word-list-based sentiment analysis on Wikipedia in a system called Wikiganda [1].

DiStaso and Messner examined 10 selected Fortune 500 companies and downloaded 2006, 2008 and 2010 versions of Wikipedia articles for each of them [3]. They split the articles to analyze more than 3,800 sentences and labeled for topic (e.g., historical, financial, corporate social responsibility, etc.) and "tonality" (i.e., sentiment analysis). Their analysis was manual. The findings suggest that a changes in contested and debated areas, such as scandals and legal issues, are prevalent.

Greenstein and Zhu examined English Wikipedia pages on US Politics with a automated technique using phrases obtained from Congressional Record [5].

Other Wikipedia research has looked at emotion (sentiment) expressed on the discussion pages of Wikipedia [6, 10]. In one study the SentiStrength method was used for sentiment analysis, — in another study the ANEW word list.

Beyond sentiment analysis there are several other Wikipedia monitoring tools, see [12, section "Trend spotting and prediction"]. At the most basic level registered Wikipedia users each have the ability to construct an individual watch list, monitoring changes on listed articles, that are displayed in a temporally sorted list. By default is listed only the most recent changes to a page, potentially hiding relevant change. An option in the preferences can be switched to show all changes. Other online Wikipedia monitoring systems, e.g., WikiTrust (http://www.wikitrust.net/) color-codes text according to reputation, the WikiWatchdog (http://wikiwatchdog.com) shares functionality with the WikiScanner, and the recently introduced Wikipedia Live Monitor (http://wikipedia-irc.herokuapp.com/) detects breaking news.

Outside Wikipedia monitoring of Internet content is legio, e.g., for monitoring of news, blogs and microblogs and analyzing the collected text with sentiment analysis.

2. DATA

It is possible to acquire Wikipedia data from the XML dumps or via the API, but to get real-time information about changes we monitor the Internet Relay Chat (IRC) streams, that the Wikimedia Foundation provides, — an approach also followed by semi-automated anti-vandalism tool STiki [14] and ClueBot NG. The stream pushes real-time summaries of the edits on Wikipedias, such as timestamp, the integer identifier for the revision, the page name, the editor, but not the page text or text difference. We monitor the English Wikipedia and to get the text itself we query the Wikipedia API.

We hand-build a small database of companies presently stored in a JSON file. In the data structure we record "main" country (usually the location of the headquarter) and where possible: company reputation indexes, stock symbol, the main Wikipedia page for the company as well as Wikipedia subpages (if any): company history article, subsidiaries, parent companies, products, controversy pages, etc. For some companies we record the brand page as the main article rather than the company page. For example, for the company "Lego" we recorded 13 subpages, e.g., the articles "History of Lego", "The Lego Group", "Lego Duplo", "Legoland" and "Kirkbi AG v. Ritvik Holdings Inc.". We do not associate more general articles to the companies. In the case of Lego we do not associate "List of companies of Denmark", "Toy" and "Acrylonitrile butadiene styrene" with Lego. We diverge a bit from that rule in the case of pharmaceutical companies, as pharmaceutical products are usually not listed on Wikipedia by trade name but rather by compound name, e.g., "Fluoxetine" vs. "Prozac".

It is not always simple to make a clear definition of what is related to a company/brand or not. For example, the Hilton brand was at one point split between two companies, and the Toyota brand is split between the companies Toyota Industries and Toyota Motor.

3. METHODS

With a continuously running Python program we monitor IRC recent changes stream of the English Wikipedia, match entities (e.g., page, user) with regular expressions and store

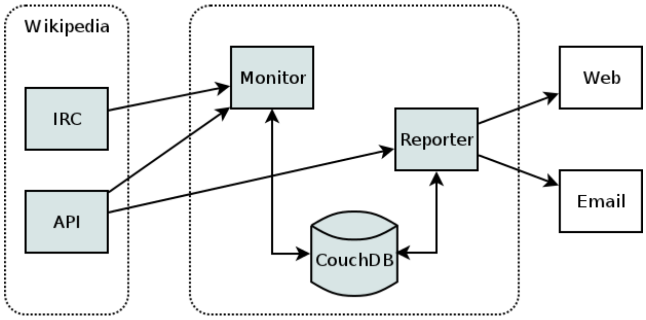
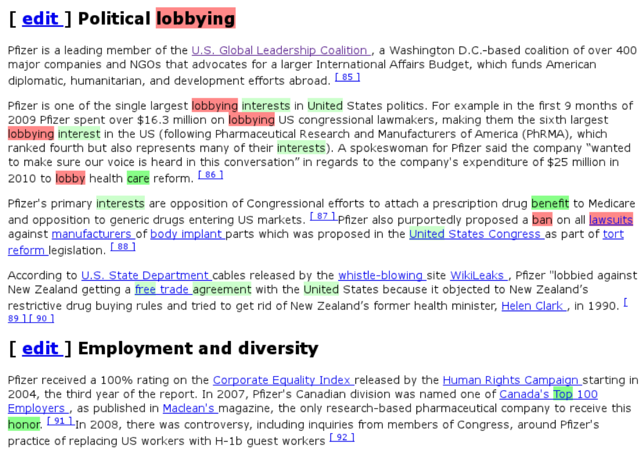


Figure 1: Workflow diagram.

the parsed data in a CouchDB document database. When the text of the revision is needed we look in the database and if the revision text is not available we query the Wikipedia API, store the revision in the database and return the text along with its metadata, see Figure 1 for an overview of the system. If we lack revisions, e.g., due to system administration downtime, we can fetch these by iterating back in the revisions.

For automated sentiment analysis we use our AFINN word list [11]. It contains 2477 terms and on an annotated Twitter data set we found it to performed reasonable well compared to other word lists. A preliminary evaluation of an extended version of the word list indicated that AFINN could perform on par with SentiStrength. The words are scored in a range between −5 (most negative) and +5 (most positive). The operation of the sentiment analyzer, its advantages and faults, can be gauged by looking at the individual words color-coded for sentiment in Figure 2. We store the computed sentiments of the revision and the difference between two revisions in the CouchDB as a part of the document, and use it as a lazy property: only computing it when requested.



[edit] Political lobbying

Pfizer is a leading member of the U.S. Global Leadership Coalition , a Washington D.C.-based coalition of over 400 major companies and NGOs that advocates for a larger International Affairs Budget, which funds American diplomatic, humanitarian, and development efforts abroad. [85]

Pfizer is one of the single largest lobbying interests in United States politics. For example in the first 9 months of 2009 Pfizer spent over $16.3 million on lobbying US congressional lawmakers, making them the sixth largest lobbying interest in the US (following Pharmaceutical Research and Manufacturers of America (PhRMA), which ranked fourth but also represents many of their interests). A spokeswoman for Pfizer said the company "wanted to make sure our voice is heard in this conversation" in regards to the company's expenditure of $25 million in 2010 to lobby health care reform. [86]

Pfizer's primary interests are opposition of Congressional efforts to attach a prescription drug benefit to Medicare and opposition to generic drugs entering US markets. [87]Pfizer also purportedly proposed a ban on all lawsuits against manufacturers of body implant parts which was proposed in the United States Congress as part of tort reform legislation. [88]

According to U.S. State Department cables released by the whistle-blowing site WikiLeaks , Pfizer "lobbied against New Zealand getting a free trade agreement with the United States because it objected to New Zealand's restrictive drug buying rules and tried to get rid of New Zealand's former health minister, Helen Clark , in 1990. [89][90]

[edit] Employment and diversity

Pfizer received a 100% rating on the Corporate Equality Index released by the Human Rights Campaign starting in 2004, the third year of the report. In 2007, Pfizer's Canadian division was named one of Canada's Top 100 Employers , as published in Maclean's magazine, the only research-based pharmaceutical company to receive this honor. [91]In 2008, there was controversy, including inquiries from members of Congress, around Pfizer's practice of replacing US workers with H-1b guest workers [92]

Figure 2: Example of sentiment coloring of sections of the Pfizer article. Words colored according to sentiment with a online service working on the live version of the English Wikipedia. The stronger the red, the more negative; the stronger the green, the more positive.

To test accuracy of the detected sentiment changes, manual coding of all detected and undetected changes was conducted for selected time range and companies. We monitored 13 companies over a 5 week period starting 2012-12-03. The researcher evaluated and labeled the data according to if a change was correctly detected as positive, negative or neutral. We only recorded errors, not which type of errors, thus we are not able to give precision-recall performance.

For presentation of relevant text paragraphs with interesting changes in a compact way, we strip wikitext formatting (tags, templates, tables, etc.) and perform sentence tokenization with the *Natural Language Toolkit* (NLTK) followed by a wikitext list detector (as the list items in Wikipedia are not segmented into items by the NLTK tokenizer). For comparing two revisions we send the two list of sentences through Python's standard difflib. For each changed sentence we detect whether it has a sentiment different from zero, and then only show that sentence as an interesting change.

Using Python's Matplotlib with output to the SVG format we are able to generate hyperlinks in point in the plot corresponding to individual edits. We show the relative change through the week rather than the absolute change. We also generate bar plots of the sentiment of users for each company as well as *sequential collaboration networks* (SCN) [9] In SCN the nodes indicate users, the links subsequent edits and the nodes/users are colored according to the user sentiment. The user sentiment is here simply the sum of the change in sentiment made by a user to the articles associated with a company.

4. RESULTS

We monitored 649 English Wikipedia pages associated with 278 companies. At http://rb.compute.dtu.dk we generate a page with plots and listing of relevant changes color-coded according to sentiment.

Figure 3 shows an example output of a weekly report with a plot showing the relative change in sentiment through time for a company together with summary statistics for user sentiment and the individual changes.

Figure 4 is an example of a sequential collaboration network for a company based on edits on its main Wikipedia page and subpages across two months. The largest red node indicates a user making several edits in the period and changing the sentiment in a negative direction. That user is directly connected to green nodes changing the sentiment in a positive direction. An examination of the individual edits reveals a small edit war between the "red" and "green" users, where the red users insert negative text on multiple pages related to the company, while other users subsequently remove it. Some of the corresponding to edits are shown in Figure 3.

The accuracy of the automated sentiment analysis compared to the ground truth of manually labeled data was from 72%-90% across the 5 weeks and the overall accuracy $81\% - 82\%$, — depending on whether "questionable"-labeled errors are counted. The data set is unbalanced as the number of neutral changes exceeds the number of changes with positive or negative sentiment: Counting on all 12490 tweets collected up to 2013-03-23, our system identified 9016 neutral changes, 1895 positive and 1579 changes. Thus by declaring all changes to be neutral we would gain a baseline accuracy on 72%.

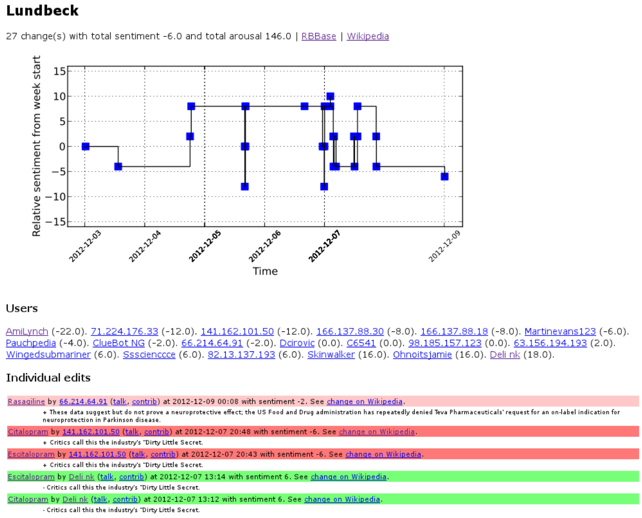


Figure 3: Weekly report from December 2012 with sentiment graph and edit differences for the Lundbeck company. The blue squares are each hyperlinked to Wikipedia's difference pages.

5. DISCUSSION

Simple word-list-based sentiment analysis may fail because a word does not appear in the list, in case of homonymy or negation or because the grammatical and semantic context is complex. Common errors we observed relate to the words "United" and "Limited", the former slightly positive, the latter slightly negative in our word list. "United" appear in country names and "Limited" in company names, see Figure 2. With exclusion of these words from our word list, handling upper and lower case or identifying named entities accuracy would possible increase slightly. Wikipedia text presents another issue as one may handle different wiki text components, such as categories, interwiki (language) links and templates, in various ways for the sentiment anal-

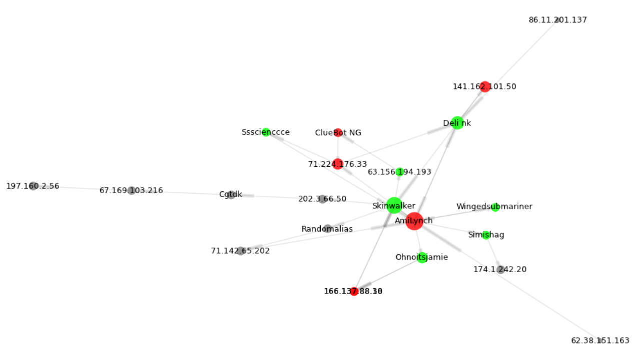


Figure 4: Sequential collaboration network for a company in the months November and December 2012. The colors depend on user sentiment.

ysis, e.g., ignoring interwiki links and expanding templates. We used the raw text without ignoring these component, enabling us to, e.g., use the word “controversies” in the “GlaxoSmithKline” article which has been categorized under the “Medical controversies” category. Using raw text one error occurs with “no” in the interwiki language link. “no” is slightly negative in our word list. With the Addbot “user” recent removal of interwiki language links due to wikidata conversion we see slight increases in sentiment each time Addbot removes the interwiki links if the Norwegian Wikipedia is linked. Bot addition of automated titles for hyperlinked references also produces change in sentiment if the title has a word matching the sentiment word list. This change is usually relevant to detect: A title word with sentiment from a referenced work should contribute to the sentiment of the Wikipedia article.

Better accuracy in the sentiment analysis could presumable be obtained with a machine learning approach, but that would require a large labeled data set. Machine learning-based ClueBot NG uses crowd-sourcing at the review interface (http://review.cluebot.cluenet.org), and similar crowd-sourcing could possibly also be used to generate a training set with labeled edits for bias.

Wiki anti-vandalism tools may use many different features [13]. Sentiment analysis have been used as one of the features [13], e.g., Harpalani et al. used LingPipe’s sentiment analyzer trained on a movie review data set [8]. It is unclear if other features from vandalism detection are relevant for bias detection, e.g., a capitalization feature probably only detects vandalism and not competent biased edits, while edit-time-of-day could possibly be useful as a feature if paid editing mostly occurs during normal working hours. A small examination of a set of 3 likely COI edits showed that these edits were performed Monday early afternoon, Tuesday morning and Thursday early afternoon. However, good bias detection should probably focus on more complex computational linguistic features together with user modeling.

Initially, we considered a more automated approach for identifying relevant subpages for a company. Using DBpedia’s “subsidiary” field would, e.g., in the case of the company *Pfizer* give 6 relevant Wikipedia articles, but miss “Kelo v. City of New London”, “Peter Rost (doctor)” and “Donepezil”. The approach by Erenrich of using USPTO trademark databases [4] could potential detect edits related to Donepezil.

Short-lived vandalism edits could be left out of the reporting. Pure vandalism tends to generate large chunks of text changes which are less relevant to show if it is quickly reverted. However, the intensity of such edits may be of interest to businesses as an indicator of the number of adversaries. Handling vandalism will require the detection of vandalism, e.g., by examining whether the edit is reverted.

6. ACKNOWLEDGMENTS

We would like to thank Annemette Leonhardt Kjærgaard for useful discussions, Christoffer Toyberg-Frandzen for manually coding Wikipedia sentiment and the Danish Council for Strategic Research for funding the *Responsible Business in the Blogosphere* research project.

7. REFERENCES

[1] R. Chandy. Wikiganda: Identifying propaganda through text analysis. *Caltech Undergraduate Research Journal*, 9(1):6–11, Winter 2008–2009.

[2] M. W. DiStaso. Measuring public relations Wikipedia engagement: how bright is the rule? *Public Relations Journal*, 6(2), 2012.

[3] M. W. DiStaso and M. Messner. Forced transparency: corporate image on Wikipedia and what it means for public relations. *Public Relations Journal*, 4(2), Spring 2010.

[4] D. Erenrich. Wikiscanner: Automated conflict of interest detection of anonymous Wikipedia edits. In *Student-Faculty Programs: 2008 Abstract Book*, page 22. California Institute of Technology, 2008.

[5] S. Greenstein and F. Zhu. Is Wikipedia biased? *American Economic Review*, 102(3):343–348, May 2012.

[6] M. Grigore and C. Rosenkranz. Increasing the willingness to collaborate online: an analysis of sentiment-driven interactions in peer content production. In *Proceedings of the 32nd International Conference on Information Systems*, page 20. Association for Information Systems, December 2011.

[7] A. Halfaker, R. S. Geiger, J. T. Morgan, and J. Riedl. The rise and decline of an open collaboration system: how Wikipedia’s reaction to popularity is causing its decline. *American Behavioral Scientist*, page 0002764212469365, December 2012.

[8] M. Harpalani, T. Phumprao, M. Bassi, M. Hart, and R. Johnson. Wiki vandalysis — Wikipedia vandalism analysis. In *Lab Report for PAN at CLEF 2010*, 2010.

[9] T. Iba, K. Nemoto, B. Peters, and P. A. Gloor. Analyzing the creative editing behavior of Wikipedia editors: through dynamic social network analysis. *Procedia - Social and Behavioral Sciences*, 2(4):6441–6456, 2010.

[10] D. Laniado, C. Castillo, A. Kaltenbrunner, and M. F. Morell. Emotions and dialogue in a peer-production community: the case of Wikipedia. In *WikiSym ’12 - 8th International Symposium on Wikis and Open Collaboration, Linz, Austria*, 2012.

[11] F. Å. Nielsen. A new ANEW: evaluation of a word list for sentiment analysis in microblogs. In M. Rowe, M. Stankovic, A.-S. Dadzie, and M. Hardey, editors, *Proceedings of the ESWC2011 Workshop on ’Making Sense of Microposts’: Big things come in small packages*, volume 718 of *CEUR Workshop Proceedings*, pages 93–98, May 2011.

[12] F. Å. Nielsen. Wikipedia research and tools: Review and comments. IMM Publication Database, February 2013.

[13] M. Potthast, B. Stein, and T. Holfeld. Overview of the 1st international competition on Wikipedia vandalism detection. In *PAN 2010*, 2010.

[14] A. G. West, S. Kannan, and I. Lee. Spatio-temporal analysis of Wikipedia metadata and the STiki anti-vandalism tool. In *Proceedings of the 6th International Symposium on Wikis and Open Collaboration*, page 18, New York, NY, USA, 2010. ACM.