

Revealing textual polarity patterns with a web browser extension

Mikko Tuomela, University of Illinois at Urbana-Champaign

Abstract

We describe a new method to combine sentiment analysis and web augmentation into a browser-based platform enabling visualization of a web document's opinionated expressions and patterns of polarity.

The Augmentator extension assists the reader recognizing keywords and paragraphs of polarity sentiment with the idea that by moving a part of the problem of text analysis from statistics and data mining into the realm of human vision and recognition, non-professionals can hopefully benefit more easily from powerful analysis and visualization tools.

Keywords: knowledge acquisition; web augmentation; cognition; annotation; sentiment analysis

Citation: Tuomela, M. (2015). Revealing textual polarity patterns with a browser extension. In *iConference 2015 Proceedings*.

Copyright: Copyright is held by the authors.

Contact: mstuomel@illinois.edu

1 Introduction

Sentiment analysis, also called opinion mining, is a set of technologies and methods primarily used by information scientists and other professionals. Typical applications include detecting emerging trends and evolving opinions from large data sets—such as from millions of tweets—or from product reviews (Hu & Liu, 2004).

Traditionally, the results could be presented as a summarization of text, focusing on extracted sentiments, and Hu and Liu describe their methods as a form of summarization. While being a powerful approach, its usual applications leave its use outside the grasp of non-professionals. At the other end of the process of knowledge acquisition is the user and the user interface. Today, the web browser is the de-facto interface to access online materials. The web interface offers several ways to improve, modify or *augment* the user experience (Díaz, Arellano, & Azanza, 2013).

Wikipedia is an excellent staging ground for both sentiment analysis and web augmentation applications. Because users are already familiar with the interface, the same familiarity could be leveraged to present them additional information (content, layout, navigation) associated with the part of web page they are viewing. One example of this is an automatically generated summary of a Wikipedia article, added as an ordinary-looking section at the end of the article (Díaz, Arellano, & Puente, 2012).

At the same time, there are specific expectations regarding the style of text used in an online encyclopedia, and topics of some of its articles are highly controversial, leading to "edit wars" that are detrimental to the quality of the article and make it more difficult for a reader to assess its reliability. These controversial topics vary from one language version to another, but certain topics and manifestations of controversy seem to be universal (Yasseri, Spoerri, Graham, & Kertész, 2014).

As annotation can be seen as a important method of how the value of a document grows. Keyword and/or context highlighting (among other methods, such as underlining) is a practical and widely used method of emphasizing relevant parts of a document, both in physical and digital texts (Zucker, 2010).

We want to find out if digital annotation by text highlighting is a reasonable approach for revealing positive and negative patterns in text in order to help understand the structure of arguments and take note of the specific tone used in sections of the text. We do not yet have empirical evaluation of benefits of this approach, however we believe the early findings are relevant enough to share.

2 Theoretical Background

2.1 Feasibility of Sentiment Analysis

Radev and McKeown used a knowledge-based summarization system to produce text summarizations of news articles (Radev & McKeown, 1998). A heuristic system like this seems usable in coping with the current challenges of the volume of data (*big data* or sometimes even *data deluge*), but the complex architecture in their work requires heuristic understanding of linguistic concepts and producing new content hides the structure of the original document(s) that the reader might want to retain. While Hu and Liu's work on feature-based opinion summarization system (Hu & Liu, 2004) does not produce or aggregate new content (template instantiation) but involves summarization by sentiment classification of individual texts.

Hu and Liu (Hu & Liu, 2004) showed that a straightforward rule-matching method can be an efficient and accurate classifier for product reviews. In their study, the texts were product reviews for personal media devices (cameras, media players, etc.) and the source texts were expected to express some opinion about the subject. With word lists of about 10,000 words in two categories ('positive' and 'negative'), they were able to correctly classify reviews' polarity with an accuracy of over 80 %, in some cases even over 90 %.

Instead of rule-matching, machine learning can be used to achieve better accuracy and to adapt to specific topics. Support Vector Machines, Naive Bayes, and Maximum Entropy are some more advanced methods of classification using machine learning. Pang, Lee and Vainyanathan (Pang, Lee, & Vainyanathan, 2002) were able to get consistent classification accuracies with SVM, NB, and ME using combinations of unigrams, bigrams, parts-of-speech tagging and other features—however, the combined accuracy for each result stayed between 80 and 90 percent.

Actually, in some cases, relying entirely on unigrams (single words, as in Hu and Liu (Hu & Liu, 2004)) yielded better accuracy than using bigrams or combinations of other methods, which asserts us that unigram-based rule-matching method is a reasonably accurate approach for detecting and highlighting polarity patterns in a document (Pang et al., 2002).

Even though these accuracy levels exhibit some challenges in determining sentiment, they also show that a simpler, unigram-based approach can still be viable and even justified, even if potentially more advanced methods exist.

Using various different methods and lexicons, sentiment analysis has been successfully used to extract sentiments along the axis of subjectivity-objectivity (Pang & Lee, 2004), emotional emphasis on various levels (Boldrini, Balahur, Martínez-Barco, & Montoyo, 2010) or even gender of the author of a text (Mukherjee & Liu, 2010).

2.2 Web Augmentation and Annotation

Web augmentation's applications include improving, "fixing" and modifying web pages' navigation and layout, and adding new content to the hypertext document, thus *augmenting* it. This is typically achieved by means of browser scripting, using JavaScript and some software framework made for this purpose, e.g. Greasemonkey (Díaz et al., 2013; Zucker, 2010).

Even though benefits of text highlighting and other types of annotation in documents are not entirely clear—and annotations can be distracting as well as beneficial (Marshall, 1997)—it has been shown that they can assist in communication by highlighting keywords in a machine-translated text (Gao, Wang, Cosley, & Fussell, 2013) or help a reader understand the discourse (structure of opinionated text) in a document (Mihăilă et al., 2013). This type of highlighting would be feasible to be used in a web interface in a digital annotation fashion that Marshall considers a "fundamental aspect" of hypertext (Marshall, 1998).

Díaz et al. argue that augmentation as an annotation-type mechanism can make it easier for people to participate in editing Wikipedia (Díaz et al., 2012). In their approach, a custom Wikipedia plugin supplements existing content with a summary of aggregated information retrieved from other Wikipedia articles and websites, each user potentially having customized content and experience fit to his or her needs. They argue that this approach is perceived as appropriate, helps monitor quality of articles and inclines readers to become editors. It is also important for the user to have the content customized to his or her own purposes.

Medical effects [\[edit\]](#)

Cocaine is a **powerful nervous** system stimulant.^[11] Its effects can last from fifteen to thirty minutes to an hour. That is all depending on the amount of the intake dosage and the route of administration.^[12] Cocaine can be in the form of **fine** white powder **bitter** to the taste. When inhaled or injected it causes a numbing effect. "Crack" cocaine is a smokeable form of cocaine made into small "rocks" by processing cocaine with sodium bicarbonate (baking soda) and water.

Cocaine increases alertness feelings of **well-being** and **euphoria** energy and motor activity feelings of competence and sexuality. Anxiety **paranoia** and **restlessness** can also occur especially during the comedown. With **excessive** dosage tremors convulsions and increased body temperature are observed.^[11] Severe cardiac **adverse** events particularly **sudden cardiac death** become a serious **risk** at high doses due to cocaine's blocking effect on cardiac sodium channels.^[8] Occasional use of cocaine rarely causes permanent or **severe trouble** to the body and mind. ^{[13][14]}

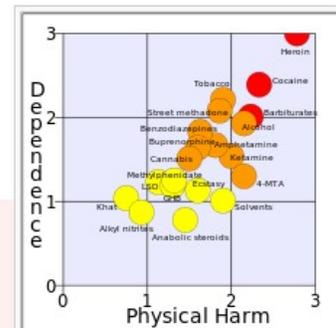
Acute [\[edit\]](#)

Main article: Cocaine intoxication

With **excessive** or prolonged use the drug can cause **itching tachycardia hallucinations** and **paranoid delusions**.^[15] Overdoses cause **hyperthermia** and a marked elevation of blood pressure which can be **life-threatening** ^[15] **arrhythmias** ^[8] and death.^[8]

Chronic [\[edit\]](#)

Chronic cocaine intake causes **strong** imbalances of transmitter levels in order to compensate extremes. Thus receptors disappear from the cell surface or reappear on it resulting more or less in an "off" or "working mode" respectively or they change their susceptibility for binding partners (ligands) – mechanisms called **down-/upregulation**. However studies suggest cocaine abusers do not show normal age-related **loss of striatal dopamine transporter** (DAT) sites suggesting cocaine has neuroprotective properties for dopamine neurons.^[16] Possible side effects include **insatiable** hunger **aches**



Comparison of the perceived harm for 20 popular recreational drugs from a poll among medical psychiatrists specialized in addiction treatment. Cocaine is ranked the 2nd most addictive and the 2nd most harmful of 20 popular recreational drugs.^[17]

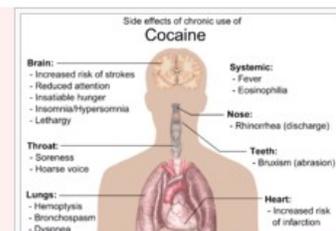


Figure 1: The Wikipedia article on cocaine with keyword and paragraph highlighting description of negative medical effects.

Article title	Controversiality	Quality	General sentiment	Observations
<i>Westboro Baptist Church</i>	medium	medium	negative	Almost the whole article is highlighted with red colors, with many references to 'protest'. This corresponds to the public's sentiment towards the organization.
<i>cocaine</i>	high	high	negative	Corresponding to the negative health effects and associated crime, parts of the article are clearly deemed 'negative', but highlighting makes it clear where alternative viewpoints and history of the substance are presented.
<i>Optimus Prime</i>	low	low	neutral/mixed	Even though the topic is not controversial per se, the highlighting shows a large number of polarized expressions. The article has been tagged as needing attention, and the number of highlighted expressions might reflect this.
<i>machine gun</i>	low	medium	positive	Machine gun is a lethal, powerful weapon. Its advantages over ordinary rifles are clear, which makes the article ripe with positive expressions—even though the effects of the weapon itself are terrible.
<i>Toyota Tacoma</i>	low	medium	neutral	As a somewhat technical description, the article is straightforward in style with a low number of polarized expressions.

Table 1: Observations from several Wikipedia articles

Wikipedia is a convenient environment for studies of web augmentation because it has a consistent structure and interface, with a lot of freely accessible content available. At the same time, due to its open nature, edit wars happen around controversial topics. In an edit war, competing editors defend their edits and submit new revisions without reaching a consensus. Analyzing these conflicts can lead to understanding of collaborative content creation as well—and conflicts in general (Yasseri et al., 2014).

According to Gao et al., highlighting keywords in otherwise irrelevant or unreliable text can help a reader to focus on relevant parts of a text (Gao et al., 2013), while Zucker has quantitatively shown (Zucker, 2010) that assisted keyword and context highlighting can improve reader's comprehension and knowledge acquisition. Use of digital annotation as means of assisting discourse analysis and need for aggregation of different means of annotation has been noted by Mihăilă et al. (Mihăilă et al., 2013).

3 Augmentator Browser Extension

3.1 Aims and Approaches

In order to make sentiment contained in discourse visible, we propose a system for automatically augmenting a web document by highlighting expressions according to their opinion polarity classification. The prototype system for sentiment indication is a standard extension for the Google Chrome web browser. When the user navigates to Wikipedia, it detects the Wikipedia web layout and analyzes the main article content, classifying expressions according to Hu & Liu's opinion lexicon of roughly 7,000 keywords¹ (Hu & Liu, 2004).

¹Hu & Liu's opinion lexicon is periodically updated and is available at Bing Liu's website at <http://www.cs.uic.edu/~liub/FBS/sentiment-analysis.html#lexicon>

The Augmentator prototype (which includes the lexicon) described herein is available for testing and experimentation at <http://people.lis.illinois.edu/~mstuomel/augmentator/>

The extension visualizes the classification in two ways: (1) comparing each word with the opinion lexicon and highlighting every matched keyword with either green ("positive") or red ("negative") background color; and (2) highlighting the paragraph where matched keywords are detected with either light green or light red background color, depending on general sentiment. Paragraphs that contain an equal number of positive and negative keywords are not highlighted this way.

There are several reasons for this approach:

- We want the reader to be able to distinguish between augmented content and original article by not modifying any other elements than words' and paragraphs' background color.
- The main content of a Wikipedia article is textual, hence graphical elements stand out from the text.
- Text on a Wikipedia article is black with white background, hence changing a word's background is a convenient way to emphasize a word's importance and it does not obscure the text in any way.
- A text paragraph is a conveniently-sized block of text providing context for highlighted expressions.
- Thanks to Wikipedia's consistent web layout, it facilitates systematic analysis and augmentation of its textual content (Díaz et al., 2012).

The highlighting resides in a background layer that would otherwise be plain white; Augmentator does not change the textual content, but provides a set of visual elements that make an alternative approach to the article possible.

By using this extension, we expect to see the reader from the following:

- The reader is able to use pattern recognition abilities to gain additional knowledge about the article—instead of systematically reading the textual content.
- By observing the highlighted patterns, the reader is able to get an overview of the structure of opinions and polarized expressions within a paragraph and the whole article.
- By noticing exceptions in expected patterns, the reader is able to recognize points of interest, such as alternative viewpoints or to take note of controversies.
- The reader's own judgement is assisted by the ability of being aware of the number of opinionated expressions contained in the text.
- The reader is able to assess the overall quality, style and trustworthiness of the article, based on the amount and patterns of highlighted expressions and paragraphs.
- Non-native speakers will be able to quickly get an overview of the general sentiment of the text.

3.2 Evaluation of Wikipedia Articles

It is expected that different types of articles yield different polarity patterns and the accuracy of the polarity analysis may vary as well. To test this, we identified a set of Wikipedia articles with varying quality and structure and observed how the extension highlights their contents (summary in Table 1). This preliminary one-person evaluation is to be used in designing a further empirical study, and while Yasseri et al. note there are several efforts to statistically detect controversies in Wikipedia articles (Yasseri et al., 2014), the controversiality as discussed herein is a subjective assessment.

For each article, three ratings are listed: (1) *Controversiality* is a measure of how contrasting opinions and strong sentiments the article is expected to attract; (2) *Quality* is the subjective perception of how well written and informative the article is; and (3) *General sentiment* is the overall polarity sentiment as judged by the author.

The article for *Westboro Baptist Church* is consistently highlighted with the red color, reflecting the controversy and strong opinions regarding the group. Numerous negative keywords were detected, rendering

determine. [reliable sources. \(March 2014\)'>citation needed](#)

In 1879 cocaine began to be used to treat [morphine](#) addiction. Cocaine was introduced into clinical use as a [local anesthetic](#) in Germany in 1884 about the same time as [Sigmund Freud](#) published his [work](#) *Über Coca* in which he wrote that cocaine causes [reliable sources. \(January 2014\)'>citation needed](#)

Exhilaration and lasting [euphoria](#) which in no way differs from the normal [euphoria](#) of the [healthy](#) person. You perceive an increase of self-control and possess more vitality and capacity for work. In other words you are simply normal and it is soon [hard](#) to believe you are under the influence of any drug. Long intensive physical [work](#) is performed without any fatigue. This result is [enjoyed](#) without any of the [unpleasant](#) after-effects that follow [exhilaration](#) brought about by alcohol. Absolutely no craving for the further use of cocaine appears after the first or even after repeated taking of the drug.

In 1885 the U.S. manufacturer [Parke-Davis](#) sold cocaine in various forms including cigarettes powder and even a cocaine mixture that could be injected directly into the user's veins with the included needle. The company [promised](#) that its cocaine products would "supply the place of food make the [coward brave](#) the [silent eloquent](#) and render the [sufferer insensitive](#) to pain."

By the late [Victorian era](#) cocaine use had appeared as a [vice](#) in [literature](#). For example it was injected by [Arthur Conan Doyle's](#) [fictional Sherlock Holmes](#) generally to offset the [boredom](#) he felt when he was not working on a case.

In early 20th-century [Memphis, Tennessee](#), cocaine was sold in neighborhood drugstores on [Beale Street](#), costing five or ten cents for a small boxful. Stevedores along the Mississippi River used the drug as a stimulant, and white employers encouraged its use by black laborers. ^[88]

In 1909 [Ernest Shackleton](#) took "Forced March" brand cocaine tablets to [Antarctica](#) as did [Captain Scott](#) a year later on his [ill-fated](#) journey to the [South Pole](#). ^[89]

During the mid-1940s, amidst WWII, cocaine was considered for inclusion as an ingredient of a future generation of 'pep pills' for the German military code named [D-JX](#). ^[90]

Prohibition [\[edit\]](#)

See also: [Legal status of cocaine](#)

 The examples and perspective in this section **deal primarily with the United States**



Figure 2: Highlighting by the Augmentator extension reveals a section that details positive effects of cocaine use. Note the high number of detections in the quote from a manufacturer in the third paragraph.

most of the article red—which was the naïve assumption, given the general sentiment and media coverage of them. However, even if the group itself is controversial, the content of the article is not, and there are no visible irregularities in the text.

The article for *cocaine* is long, informative and otherwise of apparently high quality. Even though drugs in general may be controversial, there is a consensus that cocaine is a dangerous substance if abused, and it is involved in much of the drug-related crime in the United States. Its detrimental effects on health and other negative issues (death, suffering, crime, abuse) are highlighted in red, as could be expected (Figure 1). The word pair "strong imbalances" demonstrates shortcomings of the system: the word "strong" is highlighted as being positive whereas "imbalances" is not highlighted at all. Other negative words in the same paragraph still turn the context negative. One more positive passage that stands out is the part about historical recreational and medicinal use, and is reflected in the use of such words as *euphoria*, *healthy*, and *enjoyed* (Figure 2).

Machine gun is an advanced, powerful weapon and the article describes it in a neutral tone. The positive expressions describing its features do not reflect its use as a means of killing people. Even though the function of the weapon involves death and destruction, the article discusses its technical merits, effectiveness and power, and the general sentiment is thus, perhaps surprisingly, positive.

There is no controversiality in the article for *Toyota Tacoma*, which is an article about a popular automobile and its technical features. There have been no recalls or court cases regarding safety and the car is described in general, neutral terms. An article for an engine (instead of a branded automobile) could be expected to be even more neutral in style.

Regardless of individual misclassifications of keywords, paragraph classifications generally corresponded to subjective evaluation.

3.3 Accuracy Considerations

The accuracy of classification by sentiment analysis is somewhat dependent on the domain in which it is used. Automobile and bank reviews seem to be easier to classify than reviews for movies and travel destinations. Comments about the film's narrative and themes can be misleading if taken out of context and thus a lexicon gathered from a single domain might be significantly less accurate in other domains (Turney, 2002).

This is shown in some cases with the augmentator extension as well. In the article for *Toyota Tacoma*, the expression "downhill" was classified as negative (assuming it was a figure of speech), even though it was an neutral (explicit) reference to the vehicle's performance. In the article for *cocaine*, the word "fictional" in the passage "Arthur Conan Doyle's fictional Sherlock Holmes" was classified as negative using Liu & Hu's lexicon (Figure 2).

In the case of Wikipedia augmentator, the classification process is straightforward and visible detections of keywords is shown to the reader, helping to resolve any potential misclassifications. Nevertheless, observations from the articles discussed above imply that context (paragraph) highlighting still does adequate job in visualizing the prevalence of opinionated expressions in text as it is now.

Prominent types of disagreements between Augmentator and a human evaluator regarding an individual word's classification include (1) homonyms ("fine" as being of high quality versus a punishment); (2) domain confusion ("downhill" as something becoming worse versus moving down a hill); and (3) objectivity disagreement (a neutral, objective expression perceived as opinionated in a certain context). In addition, there are a number of obviously opinionated words, words with suffixes and spelling variations not included in the lexicon. Surrounding HTML tags and also pose a challenge to the current version, leading to fewer matches.

Difficulty of improving sentiment analysis results—or generally, to develop an effective classifier without building a complex natural language processing system—is notable. The rule-based matching can still be improved in a number of ways, such as detection of features and subfeatures, in some cases using bigrams (and n-grams) instead of unigrams (Hu & Liu, 2004; Pang et al., 2002) or introducing intermediate elements that negate or neutralize the sentiment of matched keywords (Pang et al., 2002; Reckman et al., 2013). In the Augmentator extension's case, we fear that more advanced techniques might be less comprehensible to the reader, while not necessarily offering better precision—and we postulate that for this application, a high precision rate is not essential.

4 Conclusions

Even though this is an early work, we believe that the approach has great promise. The Augmentator prototype shows that using sentiment analysis approach for visualizing the polarity structure of hypertext is feasible and practical. Highlighting articles using sentiment analysis has a variety of potential uses including aiding comprehension of the content, clarifying rhetorical structure, and helping assessment of an article to decide whether it needs editing for certain kinds of bias.

Of course empirical evaluation is the next step, but already we have some thought-provoking insights from the activity of building the tool, presenting it and simply trying it out on a variety of different articles. Seeing it operating gives ideas of different use-scenarios that will inform the particular kinds of evaluation to undertake.

The visualized polarity classification seems to help identify where different aspects of an issue are discussed and how opinionated a text is. This would hopefully assist a critical reader in comprehending the structure of the content and assessing its quality and potential problematic areas. Nevertheless it is clear that the sentiment analysis algorithm used has a considerable error rate in individual classifications. The examples in the paper clearly illustrate the kinds of errors that can occur. Certain words are highlighted that should not be, given their use in context, and others that should be are not highlighted.

Fortunately there is the intriguing possibility that the substantial error rate in individual word classifications might not be as catastrophic for actual use as might be expected. Again this must be verified, but it seems that certain users may be able to tolerate quite a lot of individual word errors and still find the tool useful overall. Classification of paragraphs may be good enough. Acceptability may also be helped by the simplicity of the approach—that when the algorithm gets it wrong by an incorrect highlighting or failure to do so, it is easy to see why and to perhaps ‘forgive’ or tolerate a good enough algorithm. We invite readers to try the tool out for themselves on different Wikipedia articles and see if they agree with this interim hypothesis, as we work towards obtaining some empirical validation.

The accuracy of classification varies by the domain—in some cases, erroneous classification can be confusing. As the use of ‘weasel words’ (ambiguous, opinionated expressions) are not desirable in an encyclopedia text, the ones overlapping with the used lexicon are easily detected. Hence, the Augmentator could be developed into a quality-assessment tool to help editors vet out improper language and patterns of text. One direction for future research would be to find out what the correlation between the detected expressions and other metrics of text quality is—many high-quality articles are designated as ‘featured’ or ‘starred’ so they would be a logical choice for evaluations.

Even though polarity classification based on the unigram lexicon is somewhat crude, it is efficient and easily approachable. The accuracy could be improved using different methods, but dimensions other than polarity could be detected as well, and visualized either separately or fused into one unified augmentation layer. It seems that despite of the considerable error rate in individual classifications, classification of paragraphs is practical and accurate enough to be used in assessment of polarity in a Wikipedia article.

The Augmentator extension makes it easier, faster and cognitively cheaper to skim through an article while observing its general characteristics, assisting the reader to find relevant passages therein. We expect future empirical research to yield quantitative data on the cognitive benefits of Augmentator.

References

- Boldrini, E., Balahur, A., Martínez-Barco, P., & Montoyo, A. (2010). Emotiblog: A finer-grained and more precise learning of subjectivity expression models. In *Proceedings of the fourth linguistic annotation workshop* (pp. 1–10). Stroudsburg, PA, USA: Association for Computational Linguistics. Retrieved from <http://dl.acm.org/citation.cfm?id=1868720.1868721>
- Díaz, O., Arellano, C., & Azanza, M. (2013, May). A language for end-user web augmentation: Caring for producers and consumers alike. *ACM Trans. Web*, 7(2), 9:1–9:51. Retrieved from <http://doi.acm.org/10.1145/2460383.2460388> doi: 10.1145/2460383.2460388
- Díaz, O., Arellano, C., & Puente, G. (2012). Wikipedia customization through web augmentation techniques. In *Proceedings of the eighth annual international symposium on wikis and open collaboration* (pp. 11:1–11:10). New York, NY, USA: ACM. Retrieved from <http://doi.acm.org/10.1145/2462932.2462947> doi: 10.1145/2462932.2462947

- Gao, G., Wang, H.-C., Cosley, D., & Fussell, S. R. (2013). Same translation but different experience: The effects of highlighting on machine-translated conversations. In *Proceedings of the sigchi conference on human factors in computing systems* (pp. 449–458). New York, NY, USA: ACM. Retrieved from <http://doi.acm.org/10.1145/2470654.2470719> doi: 10.1145/2470654.2470719
- Hu, M., & Liu, B. (2004). Mining opinion features in customer reviews. In *Proceedings of the 19th national conference on artificial intelligence* (pp. 755–760). AAAI Press. Retrieved from <http://dl.acm.org/citation.cfm?id=1597148.1597269>
- Marshall, C. C. (1997). Annotation: From paper books to the digital library. In *Proceedings of the second acm international conference on digital libraries* (pp. 131–140). New York, NY, USA: ACM. Retrieved from <http://doi.acm.org/10.1145/263690.263806> doi: 10.1145/263690.263806
- Marshall, C. C. (1998). Toward an ecology of hypertext annotation. In *Proceedings of the ninth acm conference on hypertext and hypermedia : Links, objects, time and space—structure in hypermedia systems: Links, objects, time and space—structure in hypermedia systems* (pp. 40–49). New York, NY, USA: ACM. Retrieved from <http://doi.acm.org/10.1145/276627.276632> doi: 10.1145/276627.276632
- Mihăilă, C., Kontonatsios, G., Batista-Navarro, R. T. B., Thompson, P., Korkontzelos, I., & Ananiadou, S. (2013, August). Towards a better understanding of discourse: Integrating multiple discourse annotation perspectives using uima. In *Proceedings of the 7th linguistic annotation workshop and interoperability with discourse* (p. 79-88). Retrieved from <http://www.aclweb.org/anthology/W13-2310> (LAW Challenge Award)
- Mukherjee, A., & Liu, B. (2010). Improving gender classification of blog authors. In *Proceedings of the 2010 conference on empirical methods in natural language processing* (pp. 207–217). Stroudsburg, PA, USA: Association for Computational Linguistics. Retrieved from <http://dl.acm.org/citation.cfm?id=1870658.1870679>
- Pang, B., & Lee, L. (2004). A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts. In *Proceedings of the 42nd annual meeting on association for computational linguistics*. Stroudsburg, PA, USA: Association for Computational Linguistics. Retrieved from <http://dx.doi.org/10.3115/1218955.1218990> doi: 10.3115/1218955.1218990
- Pang, B., Lee, L., & Vaithyanathan, S. (2002). Thumbs up?: Sentiment classification using machine learning techniques. In *Proceedings of the acl-02 conference on empirical methods in natural language processing - volume 10* (pp. 79–86). Stroudsburg, PA, USA: Association for Computational Linguistics. Retrieved from <http://dx.doi.org/10.3115/1118693.1118704> doi: 10.3115/1118693.1118704
- Radev, D. R., & McKeown, K. R. (1998, SEP). Generating natural language summaries from multiple on-line sources [Article]. *Computational Linguistics*, 24(3), 469-500.
- Reckman, H., Baird, C., Crawford, J., Crowell, R., Micciulla, L., Sethi, S., & Veress, F. (2013, June). teragram: Rule-based detection of sentiment phrases using sas sentiment analysis. In *Second joint conference on lexical and computational semantics (*sem), volume 2: Proceedings of the seventh international workshop on semantic evaluation (semeval 2013)* (pp. 513–519). Atlanta, Georgia, USA: Association for Computational Linguistics. Retrieved from <http://www.aclweb.org/anthology/S13-2085>
- Turney, P. D. (2002). Thumbs up or thumbs down?: Semantic orientation applied to unsupervised classification of reviews. In *Proceedings of the 40th annual meeting on association for computational linguistics* (pp. 417–424). Stroudsburg, PA, USA: Association for Computational Linguistics. Retrieved from <http://dx.doi.org/10.3115/1073083.1073153> doi: 10.3115/1073083.1073153
- Yasseri, T., Spoerri, A., Graham, M., & Kertész, J. (2014). Global wikipedia: International and cross-cultural issues in online collaboration. In P. Fichman & N. Hara (Eds.), (chap. The most controversial topics in Wikipedia: A multilingual and geographical analysis). Scarecrow Press.
- Zucker, R. (2010). The effects of annotated web documents, using context highlighting, on quiz performance and preparation time. In *Proceedings of the 48th annual southeast regional conference* (pp. 36:1–36:5). New York, NY, USA: ACM. Retrieved from <http://doi.acm.org/10.1145/1900008.1900059> doi: 10.1145/1900008.1900059

Table of Figures

Figure 1	The Wikipedia article on cocaine with keyword and paragraph highlighting description of negative medical effects.	3
Figure 2	Highlighting by the Augmentator extension reveals a section that details positive effects of cocaine use. Note the high number of detections in the quote from a manufacturer in the third paragraph.	6

Table of Tables

Table 1	Observations from several Wikipedia articles.....	4
---------	---	---