구글과 바이두 검색엔진의 개인정보에 대한 감성분석과 마이닝

이가베[*] · 이효맹[*] · 유효문[*] · 강선경[*] · 이현창[*] · 신성윤[**]

[*]원광대학교

[**]군산대학교

Current State of the Art and Review of Google and Baidu Search Engines' Privacy Policies Using Sentiment Analysis and Opinion Mining

Jiapei Li[1*] · Xiaomeng Li[1] Xiam· Sun-kyung Kang[1] · Hyun Chang Lee[1],· Seong-yoon Shin[**]

[*]Wonkwang University

Kunsan National University

E-mail : [*]jiapei@gmail.com [*]hclglory@gmail.com

요 약

감성분석은 의사 소통에서 감정이나 느낌의 측정을 결정하기 위해 서면 또는 구두로 의사 소통을 검토하는 것이다. 검색 엔진은 인터넷에서 방문한 가장 인기있는 사이트 중 하나이며 전세계에서 수 십억 건의 조회가 발생한다. 이러한 검색 사이트와 관련된 개인 정보 취급 방침은 분명히 중요하다. 본 연구는 두 개의 가장 큰 검색 엔진 인 Google과 Baidu의 개인 정보 보호 정책을 검토하여 개인 정보 보호 정책의 전반적인 정서를 결정한다. 몇 가지 정서 및 의견 분석 방법을 사용하여 중요한 개인적인 발견과 중요한 차이가 발견하여 향후 이를 이용한 정책반영에 활용할 수 있다.

ABSTRACT

Sentiment analysis is the review of written or verbal communications to determine some measure of emotion or feeling in the communication. Search engines are one of the most popular sites visited on the Internet generating hundreds of billions of hits per month worldwide. Obviously privacy policies related to these search sites are extremely important. Our study reviews the privacy policies of the two largest search engines, Google and Baidu to determine the overall sentiment of their privacy policies. Significant individual findings and significant differences were found using several sentiment and opinion analysis methods.

키워드

Ⅰ. 서 론

The field of sentiment analysis and opinion mining is exploding. There is a virtual flood of qualitative data available from a wide variety of sources on the web that can be used to analyze the attitudes behind textual material. Millions of Twitter posts or tweets, millions of Facebook posts and billions of web pages and other documents can be reviewed to determine the opinions behind the words. This analysis can be extremely useful for both researchers and practitioners. Marketing professionals can

monitor text communications to determine current attitudes towards their products. Politicians can analyze text communications to determine popularity and feelings toward their candidacy and their stands on issues. Researchers can likewise study text data to find differences, patterns, or trends in a wide variety of text, from policies to presentations, from documents to websites. Sentiment analysis including its definition, history, and a review of the literature. Following this is a review of current tool terms and dictionaries that are used in contemporary sentiment and opinion analysis tools. Finally, a detailed example of the use of these tools is presented comparing the sentiment of the privacy policies of two major search engines, Google and Baidu. A statistical comparison is made of the sentiment results of these two documents and statistical conclusions are made with regard to their sentiment differences.

II. Methodology

There are a variety of tools that can be used for sentiment analysis. We will use chinese BosonNLP to compare results where, and perform a review of two the most popular search engines privacy policies, Google and Baidu and analyze their sentiment with these tools.

For our analysis, we utilize several online and commercial products as well as several new analyses we have developed using extensive new affective word dictionaries. Specifically we used BosonNLP (Boson Chinese Semantic Open Platform) and NLPIR(ICTCLAS2016 Huaping Zhang 2016) With these two analyses we are able to calculate significant differences between the two policies. This was because our self-developed algorithm had values for every word in the document and was thus able to allow calculation of means, standard deviations and perform t-tests on the data.
The Google and Baidu search engine privacy policies (henceforth known as policies or privacy policies) were download from their respective sites and used to perform all analysis.

III. Discussion and Conclusions

Overall this study has contributed to the literature in three ways, first it defines, presents and demonstrates different methods of sentiment analysis. Researchers and practitioners can use this manuscript as a source, primer and guide for developing their own sentiment analysis of any communication. Second, the study illustrates the inexact but relatively consistent results that are generated by several sentiment analysis tools and dictionaries. Researchers and practitioners can reliably use any of the tools and obtain similar results regardless of the tools used. Note that there is some small variation that will be experienced. Finally, the study analyzes the privacy policies and sentiment and tone of the two largest search engines. The results show little differences in any of the sentiment measures between Google and Baidu. Both are somewhat positive in sentiment, Researchers can use these findings to compare to other search engines policies or other privacy policies for other type sites to compare and contrast their sentiment characteristics. Search engines companies can use these findings to improve their overall sentiment if they choose. Potential changes in privacy policies for companies could be to make privacy policies, happier, less controlled, provide a change in arousal, and be more concrete.

REFERENCES

[1] NLPIR(2017) text analysis API http://ictclas.nlpir.org/nlpir/
[2] BosonNLP text analysis API http://bosonnlp.com/
[3] Liu, B. (2012). Sentiment analysis and opinion mining. Synthesis lectures on human language technologies, 5(1), 1-167.