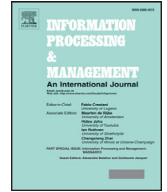




Contents lists available at ScienceDirect

# Information Processing and Management

journal homepage: [www.elsevier.com/locate/infoproman](http://www.elsevier.com/locate/infoproman)

## Computing controversy: Formal model and algorithms for detecting controversy on Wikipedia and in search queries



Kazimierz Zielinski\*, Radoslaw Nielek, Adam Wierzbicki, Adam Jatowt

Polish-Japanese Academy of Information Technology, Warsaw, Poland, Kyoto University, Kyoto, Japan

### ARTICLE INFO

#### Article history:

Received 22 January 2017  
 Revised 11 August 2017  
 Accepted 23 August 2017

#### Keywords:

Controversy  
 Wikipedia  
 Web search

### ABSTRACT

Controversy is a complex concept that has been attracting attention of scholars from diverse fields. In the era of Internet and social media, detecting controversy and controversial concepts by the means of automatic methods is especially important. Web searchers could be alerted when the contents they consume are controversial or when they attempt to acquire information on disputed topics. Presenting users with the indications and explanations of the controversy should offer them chance to see the “wider picture” rather than letting them obtain one-sided views. In this work we first introduce a formal model of controversy as the basis of computational approaches to detecting controversial concepts. Then we propose a classification based method for automatic detection of controversial articles and categories in Wikipedia. Next, we demonstrate how to use the obtained results for the estimation of the controversy level of search queries. The proposed method can be incorporated into search engines as a component responsible for detection of queries related to controversial topics. The method is independent of the search engine’s retrieval and search results recommendation algorithms, and is therefore unaffected by a possible filter bubble.

Our approach can be also applied in Wikipedia or other knowledge bases for supporting the detection of controversy and content maintenance. Finally, we believe that our results could be useful for social science researchers for understanding the complex nature of controversy and in fostering their studies.

© 2017 The Authors. Published by Elsevier Ltd.

This is an open access article under the CC BY-NC-ND license.

<http://creativecommons.org/licenses/by-nc-nd/4.0/>

## 1. Introduction

Controversy abounds in the world. At times it seems that almost any subject can be a source of controversy since it is virtually impossible for everyone to agree on any subject of interest. However, such assumption is a fallacy: by focusing on the presence of minor disagreements, we ignore the existence of an organized, encyclopedic collection of agreed-upon facts. Still, it is quite clear that controversial topics and controversial information exist, and compared with non-controversial topics or information, they are usually more prominent and easily noticeable in media and on the Web. Even topics that are well supported by facts and evidence may not constitute a guaranteed consensus. People tend to argue about opinions, interpretations, and points of view. These disagreements are not always counterproductive; they may allow participants to

\* Corresponding author.

E-mail address: [kazimierz.zielinski@pjwstk.edu.pl](mailto:kazimierz.zielinski@pjwstk.edu.pl) (K. Zielinski).

adjust their points of view or to inspire a search for truth or common ground. Yet, controversy can also impede or prevent cooperation. This problem has been recognized and addressed by communities that generate peer-produced content, such as Wikipedia, which attempts to limit the impact and spread of controversy and improve the article editing process. Therefore, detection of controversy may be crucial for ensuring high-quality peer-produced content and more balanced viewpoints.

While browsing the Web in search for information about a topic of interest, individuals may find content that supports only one point of view. The threat of receiving biased information or misinformation is even more serious with the pervasive personalization of search engines and the associated filter bubble effect (Nguyen, Hui, Harper, Terveen, & Konstan, 2014; Pariser, 2011). Unless users persevere in their search for acquiring more information to corroborate and to find information from diverse viewpoints, they are likely to learn about only one-side of any argument regarding the topic of their interest. Consequently, a person may make wrong decisions as a result of being exposed to misleading (or one-sided) information about a controversial topic, without even realizing that controversy exists. If the person was made aware of the controversy, she/he would be able to make more informed and better decisions based on more comprehensive knowledge about the topic. The truism 'forewarned is forearmed' also applies to the context of Web searches.

With the above in mind, a timely warning about the existence of controversy could be beneficial for users who may be unaware of it before they trust incomplete or biased information on important topics. In our study, we focus on the two most widely used "starting points" for users who are looking for information on the Internet: Wikipedia and search engines. With close to 500 million unique visitors per month Wikipedia is immensely popular knowledge base that has powerful educational impact on our society. Often users use it as a springboard when researching variety of topics including complex ones in scientific domains or more trivial ones like celebrity information. Wikipedia is also commonly used for supporting various data processing tasks.

Search engines are gateways to the Internet. They assist users in locating any desired information or websites. As such, it would be beneficial to searchers if they got alerted on topic controversy when interacting with search engines – before even seeing the returned results. In this work we consider only search queries and not the results returned by search engines to avoid the need to account for the search engine type and limitations associated with personalizing the search progress.

### 1.1. Research objectives

Given the described importance and complexity of the concept of controversy, we identified the following research objectives.

While prior work has already demonstrated methods of identifying controversy, there is no unified framework that would define the problem and represent it formally. We believe that a conceptual model of controversy is necessary. It would provide basis for computational approaches towards building applications aiming at detecting, estimating and understanding controversy in texts. Such a model would be useful for establishing detection methods, but it could also allow us to compare existing approaches.

Given the importance of Wikipedia and search engines in acquiring information and learning, one would expect a strong support for online users to prevent acquisition of one-sided information. However, no such support has been explicitly offered so far, besides solutions that rely on manual detection of controversy or focus on related tasks (e.g., content relevance estimation). To implement such functionalities, the development of effective algorithmic approaches for treatment of controversy is required.

### 1.2. Contributions

In this work, we make the following contributions:

**Definition and formal representation of controversy:** We propose a definition of controversy to serve as the basis for our research. This definition is based on the social sciences and epistemology. It may also be useful for researchers in computer science, as it draws a line between controversial and non-controversial content and topics. The main advantage of this definition is that the level of controversy can be quantified. Moreover, our formal model classifies errors as having three main sources; such classification is indispensable when attempting to predict controversy.

**Estimating controversy of Wikipedia articles and categories:** We develop new methods for identifying controversial articles on Wikipedia by analyzing the sentiments of discussions on the article talk pages. Using these detection methods, we can pinpoint the sections of articles that generate the greatest controversy. Based on the proposed classifiers of Wikipedia article controversy, we can estimate the levels of controversy associated with topical categories on the English-language Wikipedia. This approach can be then used to predict the level of controversy in new articles within a category.

**Method for estimating controversy levels of search queries for offering early warning about high controversy:** Our third contribution is a method for estimating controversy and generating early warnings about high controversy associated with the topics of Web search queries. This method is based on a ranking of the level of controversy characterizing Wikipedia content categories. We then consider the similarity of a query to Wikipedia content categories associated with known controversy levels. This prediction tool can be used by searching Web users to receive early warnings prior to the submission of a search query to a Web search engine. Most importantly, the proposed method only depends on the user's query and information from Wikipedia, not on the search engine or its search results. Because our method is independent of the search engine's algorithm, it also works independently of the filter bubble. Note that two different users may receive

Download English Version:

<https://daneshyari.com/en/article/4966378>

Download Persian Version:

<https://daneshyari.com/article/4966378>

[Daneshyari.com](https://daneshyari.com)