Sentiment Analysis of Controversial Topics on Wikipedia

ABSTRACT

- Wikipedia is the first point of contact for information for many people. Everyone from all around the world can create and post content.

- Widely differing opinions can incite arguments, thus, some topics are protected. We present a sentiment analysis between controversial and neutral topics

- We investigate the character of controversial topics and people, and compare them with neutral subjects to see how they are different and which attributes make them different.

- We present the sentiment attributes of eight different religions on Wikipedia
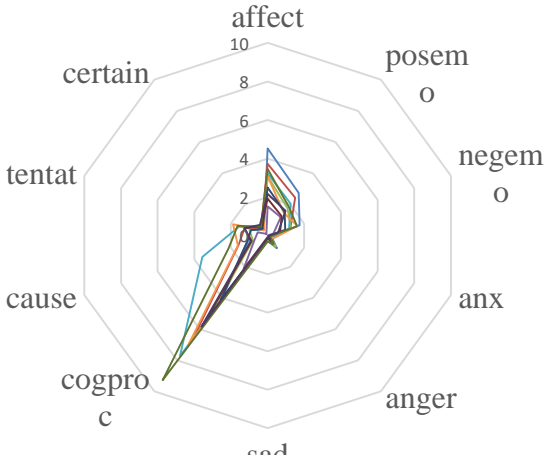
Motivation

- Wikipedia is accessible from worldwide and this can allow quality problems, vandalism, bias, and controversy. Many of these problems have been resolved by constraining access, limiting comments, using software and automatons for early detection of violation of rules.

- There are many studies on scrutinizing controversial and sensitive topics, addressing challenges of vandalism detection, finding bias, and detection of controversy.

- Following are the RQs we answer in this study:

- RQ1: What are the attributes of sentiment usage in controversial topics on Wikipedia?

- RQ2: What are the attributes of sentiment usage in different religions on Wikipedia?

- RQ3: What are the attributes of sentiment usage in neutral topics, and how are these different from controversial and/or religious topics sentiments?

Methodology

- Top ten controversial topics on Wikipedia

- George W. Bush, Anarchism, Muhammad, LWWEe (List of World Wrestling Entertainment, Inc. employees), Global warming, circumcision, United States, Jesus, Race and intelligence, and Christianity

- Top ten controversial people on Wikipedia

- George W. Bush, Michael Jackson, Jesus, Barack Obama, Adolf Hitler, Britney Spears, The Undertaker, Roger Federer, and Kane (wrestler)

- Ten random neutral (open access) topics on Wikipedia

- car, game, laptop, school, forest, factory, Convolutional Neural Networks, Spring (season), data mining, and passport

- Eight included religions

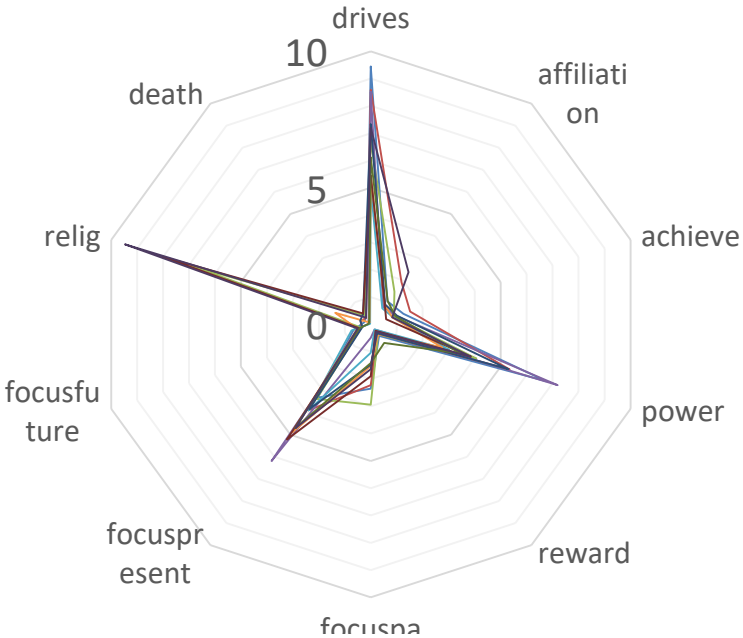- Islam, Christianity, Sikhism, Zoroastrianism, Judaism, Hinduism, Baha, Buddhism

Results

Top ten psychological features for top ten controversial topics

Other psychological features for top ten controversial topics

Results

Top ten psychological features for top ten controversial people
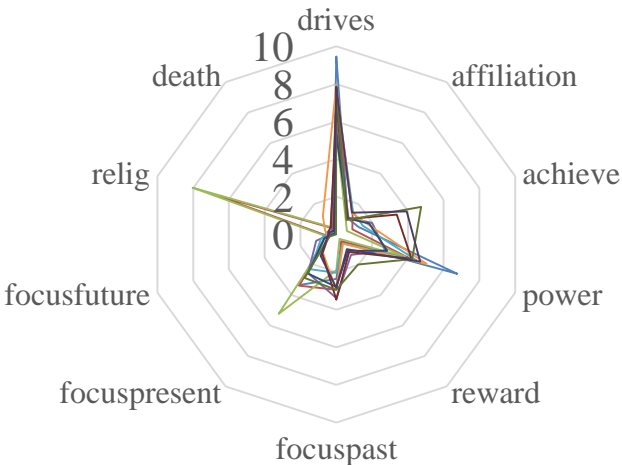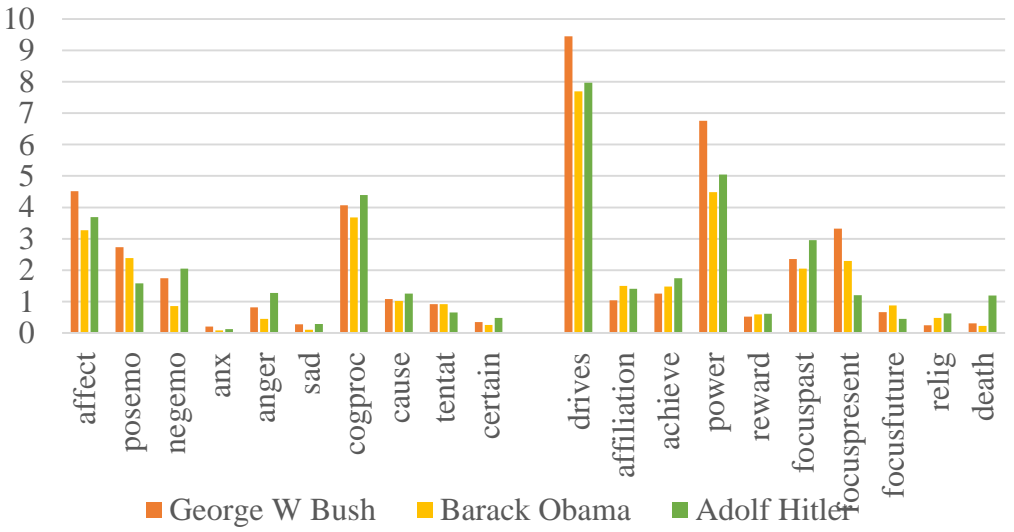
George W Bush   Muhammad

Jesus   Michael Jackson

Barack Obama   Adolf Hitler

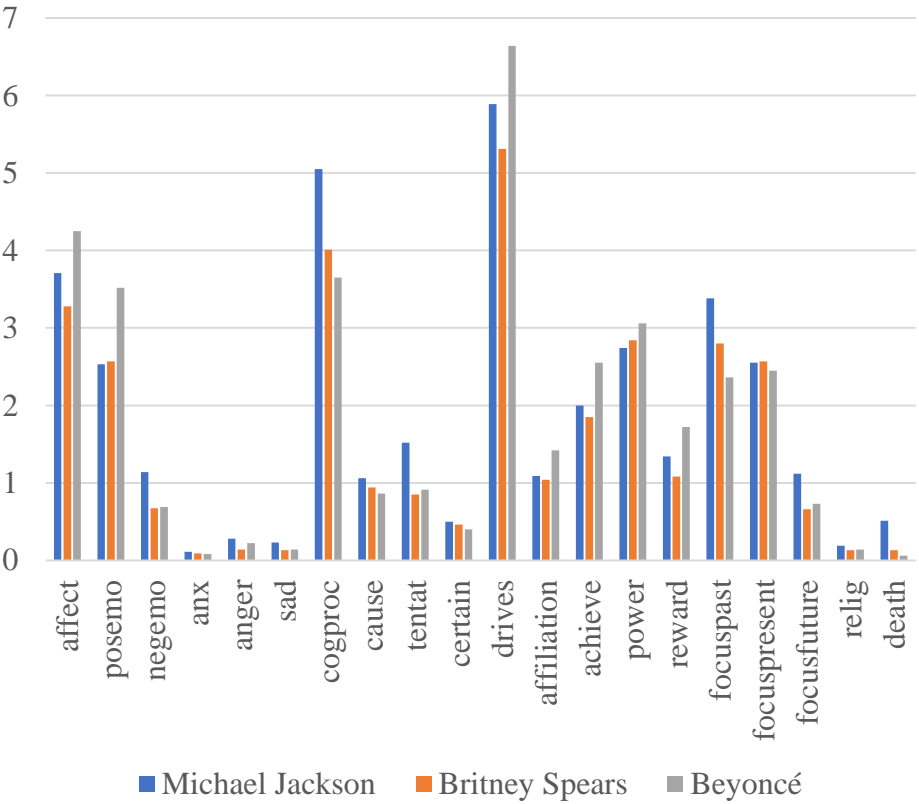Britney Spears   The Undertaker

Roger Federer   Kane(wrestler)

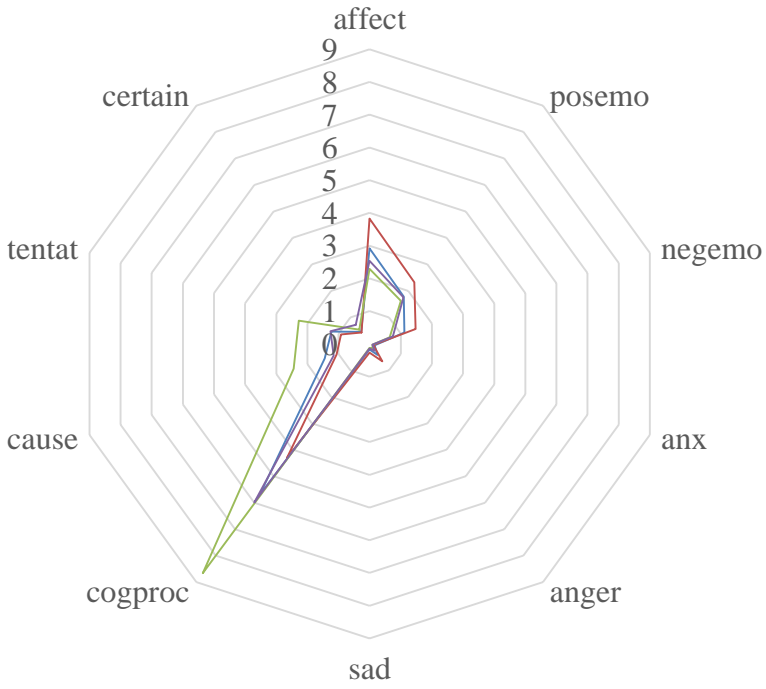Other psychological features for top ten controversial people

Results

Comparison of sentiment dimensions between controversial politicians

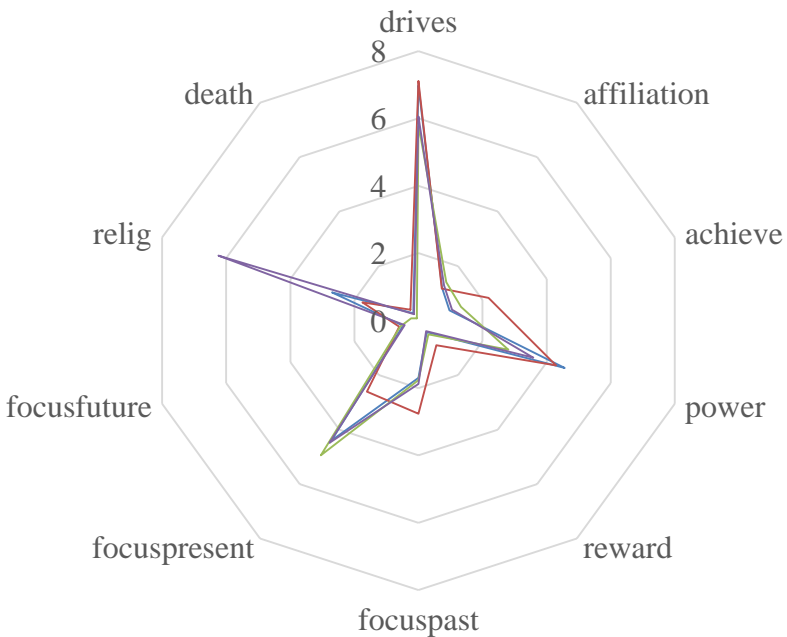Comparison of sentiment dimensions between controversial singers

Results

Comparison of average of ten sentiment dimensions between controversial topics, controversial people, neutral topics, and religions

Comparison of average of sentiment dimensions between controversial topics, controversial people, neutral topics, and religions

Conclusion and Future Directions

- In this study, we investigated the character of the Wikipedias' most controversial topics and compared them with neutral topics.

- Our analysis of Wikipedias' controversial and neutral topics sentiment demonstrates differences between the degree of sentiment dimensions among controversial and neutral topics, implying that the represented information about the topics on Wikipedia is indicative of the personality (nature or character) of the subject

- It shows close correspondence to the real nature of the subjects designating that the content of the topics represents the offline landscape.

- Via content and opinion mining we find that controversial people and subjects contain more intense feelings and display attitudes not contained within neutral topics

- This work sheds light on the latent sentiment that world-wide users are inadvertently consuming. Readers have a reasonable expectation that the content of Wikipedia is neutral. This work indicates that this is not the case.

- There are unknown impacts to be discovered discussed in future works about the subtle nuances contained in Wikipedia's content.