

Analyzing the popularity of Wikipedia Articles

Briceag Marian

Master thesis in Computer Science

Nowadays, people use search engines to get informed about topics of interest or any important piece of news. Among all the results retrieved by any search engine, usually at least one points to a Wikipedia entry as the latter serves as a platform where people try to find reliable, detailed and consolidated answers to all the questions related to the event of interest. The main motivation behind this work was to find if based on the information provided by the pageviews or edit information of Wikipedia, we can detect anomalies that might correspond to real life events. Our focus is limited to Wikipedia as search engines do not provide detailed access to the search logs for research purposes. For example, Google Trends offers only normalized popularity of top search queries in Google Search. Our work focuses on two questions. Firstly, we verify the hypothesis that if a keyword is in the top results of a search query, the corresponding article on Wikipedia is also highly viewed. This is done by studying correlation between Google Trends output for few relevant keywords and the corresponding Wikipedia pageview statistic. The second question was to find if we can detect anomalies, that might correspond to real life events, based on the information provided by the pageviews of Wikipedia. In the following work, we have analyzed these two questions based on datasets for the following topics: Mass shootings in the US, Capital punishment in the US, Linux distributive releases, Stock Exchange (NASDAQ, NYSE), Elections worldwide.

Prof. Philippe Cudré-Mauroux