



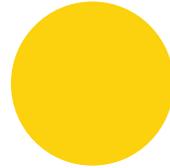
**A FAST EMPIRICAL
METHOD FOR DETECTING
FAKE NEWS ON
PROPAGANDISTIC NEWS
RESOURCES**

Conference proceeding

Behind the Digital Curtain. Civil Society vs State Sponsored Cyber Attacks

Brussels - 25/06/2019

DOI 10.34054/bdc000



Liubomyr Monastyrskyi,

Prof., Head of
Department, Ivan
Franko National
University of Lviv,
Department of
Radioelectronic and
Computer Systems

Oleh Sinkevych,

Assistance Prof.,
Ivan Franko National
University of Lviv,
Department of
Radioelectronic and
Computer Systems

Yaroslav Boyko,

Associate Prof., Ivan
Franko National
University of Lviv,
Department of
Radioelectronic and
Computer Systems

Bohdan Sokolovskyi,

Associate Prof., Ivan
Franko National
University of Lviv,
Department of
Radioelectronic and
Computer Systems

In this work, an effective method for detecting news reports about certain events with deliberately distorted content is proposed. The methodology consists in accumulating reliable information about a given event from reliable verified sources. Thereafter, there occurs a quantitative analysis of the reliability of information from the sources which are suspected in their integrity. To achieve this goal, we used the standard software tools for obtaining and processing information with the help of NLP techniques.

In our case, for collecting information the Python requests modules (data scraping) [1] are used and for working with HTML data the BeautifulSoup4 ones [2].

To date, there is a large number of effective tools for working with text information which use both direct rules based on rules and methods of artificial intelligence tools. Especially accurate results in many areas of NLP are provided by the machine learning, including deep machine

learning. At the same time, the use of the latter is connected with the problems of providing the necessary volumes of computing resources which are not always acceptable. For the preliminary evaluation analysis, in our opinion, the best methods are based on the reasonable mathematical approaches and, at the same time, are not related to large volumes of computations. To achieve our goal we propose to use the text-based properties based on the TF-IDF statistics [3] This is a statistical indicator used to evaluate the importance of words in the context of a document which is a part of collection of documents or corpus. The weight (significance) of words is proportional to the number of word uses in the document and is inversely proportional to the frequency of the use of words in other documents of collection. The TF-IDF indicator is used in the task of text analysis and information retrieval. It can be used as one of the criteria for the relevance of a document to a search query, as well as

To date, there is a large number of effective tools for working with text information which use both direct rules based on rules and methods of artificial intelligence tools.

when calculating the degree of affiliation of documents during clustering. The easiest ranking function can be defined as a total number of the TF-IDF of each term in the query. The most advanced ranking functions are based on this simple model [4].

In our work, the implementation of TF-IDF statistics computation in the Python sklearn module [5] was realized. On the

basis of the set of reliable news articles about a certain event, the main building was formed. In order to determine the degree of reliability of the article under study we added the latter to this set and a similarity matrix was calculated. It has been found that the degree of similarity of the articles that relate to the reliable ones exceeds an index of 0.5, while the articles with distorted data refer to the main corpus with similarity levels in the 0.1 - 0.4 range for all experiments conducted by us. The main disadvantage of our results is that we consider small volumes of sets of articles, so in future research, we plan to automate the process of obtaining data for the rapid formation of large corpus, as well as the improvement of computational procedures.

Sources

- [1] <https://2.python-requests.org/en/master/>
- [2] <https://www.crummy.com/software/BeautifulSoup>
- [3] Charu C. Aggarwal. Machine Learning for Text. – Springer. – 2018. – 493 p.
- [4] <https://uk.wikipedia.org/wiki/TF-IDF>
- [5] <https://scikit-learn.org/>

Web: www.promoteukraine.org
Contact: info@promoteukraine.org

Promote Ukraine is a non-profit start-up. It is a politically and governmentally independent organization situated in Belgium. It consists of a thriving team of professionals who on a pro bono basis seek to give voice to Ukrainian civil society in Europe and, in particular, throughout Belgium. We believe in European values such as civil rights, good governance and equal opportunities. Through connecting EU businesses and politicians with Ukrainian stakeholders, we facilitate the sharing of best practices between EU and Ukrainian partners with the goal to bring Ukraine closer to EU norms and values from a bottom-up perspective.

