

# Manipulative Propaganda Techniques

## Technical Report

Vít Baisa, Ondřej Herman, and Aleš Horák

Natural Language Processing Centre  
Masaryk University, Faculty of Informatics  
Botanická 68a, Brno

**Abstract.** Influencing the public attitude towards certain topics had become one of the strongest weapons in today's information warfare. The ability to recognize a presence of propaganda in newspaper texts is thus a treasured phenomenon, which is not directly transferable to algorithmic analysis.

In the current paper, we present the first steps of the project aiming at detection and recognition of selected propaganda elements in common texts. We introduce a developed tool used for annotating designated manipulative elements in newspaper texts and providing an overview of complex text characteristics of the underlying corpus data.

The presented Propaganda corpus consists of more than 5,000 newspaper articles by 4 publishing websites.

**Key words:** propaganda, manipulative techniques in text, propaganda corpus, propaganda detection

## 1 Motivation and Introduction

The freedom on the internet allows people with malicious intents to spread hatred, fake news, alternative truths, misinterpretations which might in turn instigate extreme nationalism, xenophobia, homophobia, class discrimination etc.

The work presented in this article is a part of an interdisciplinary project funded by Masaryk University. The idea of the project is a reaction to the current issue of politic propaganda of foreign entities via new media and social networks in Czech Republic. This phenomenon brings new opportunities for methodology, safety and law research and offers challenges for interdisciplinary research. On the example of pro-Russian information warfare the project develops methods of discerning, detecting and analysing of manipulative propaganda techniques in newspaper texts. Investigations also partly aim at users sharing manipulative content from the point of their motivation and evaluates security risks for the Czech Republic.

The collaborating parties include: political scientists from the Faculty of Social Studies, legal scientists from the Faculty of Law and computer scientists from

the Natural Language Processing Centre (NLP Centre), Faculty of Informatics (the authors of this paper).

The task of the NLP Centre group within the project is to develop a system capable of:

- 1) regular acquisition of web documents from a list of propaganda websites,
- 2) providing an annotation tools of the web documents for propaganda experts,
- 3) advanced search and data statistics acquisition based on the propaganda corpus data, and
- 4) developing automatic methods for discerning and classifying unseen web documents, stylometry for anonymous authorship recognition.

This paper describes the progress in the first two points.

## 2 Related Works

The information channels based on the World Wide Web environment approach the public via several basic access points such as the social networks (Facebook, Twitter, Instagram, ...), navigation from preselected news server website (on-line news sources) or the web search. The polarization of the first two channels is their inherent property and people expect them to be predisposed, but the web search is accepted as an *objective* tool in this respect. However, since the huge amounts of full text query results have surpassed human processable limits, the ordering and filtering of the web retrieval results can play a crucial role in polarizing them. Recent studies [1,2] show, how the topic of propaganda, fake news and manipulative text influence the current search engine techniques.

Social networks allow the speed-of-light dissemination of any viral kind of information, making it a perfect Swiss-army knife of possible propaganda and manipulation. [3] studies the tweeting behavior of Twitter propagandists and identify four features which can be used for distinguishing ideological tweeters from neutral users. The social network companies try to fight this situation using complex AI tools, user grouping rights, or fake news collaborative marking [4]. After the recent cases of massive propaganda during elections in several countries (USA, Germany, or France), the social networks employ community fact-checkers allowing thus a distributed way of manual fake news fight [5].

The actual propaganda devices and the propaganda model [6] have been theoretically studied for decades or centuries [7]. Formalizing these human techniques for computer processing is, however, a complex task possibly consisting on many subtasks such as the topic change identification [8], rumour identification and verification [9], or hoaxes and fake news detection [10].

In the following text, we concentrate on the ways of identification of possible manipulative techniques purely from the underlying text and style characteristics, without any factual verification.

### 3 The Annotation Scheme

Here we do not describe the annotation rules as it is out of the scope of this paper. We give a brief overview of attributes which are annotated and need to be stored in corpus data or in a database.

The following attributes are manually annotated for each document. We give also examples of values. The first part consists of attributes which can be assigned to a sequence of words from a document (we call them *range attributes*):

- location (a town, a country: *EU, Česká republika, USA, other country, Rusko, NATO, Rusko + USA*),
- blame (is anyone blamed for something?, boolean),
- labelling (is somebody or something labelled?, boolean),
- argumentation (does the article include arguments in favour or against something?, boolean),
- emotions (*fear, outrage, sympathy, hatred, other, missing*),
- demonizing (is something demonized?, boolean),
- relativizing (is something relativized?, boolean),
- fear (boolean),
- fabrication (boolean),
- Russia (how is the Russian Federation depicted?, *positive, neutral, negative, victim, hero, missing*),
- genre (*news, commentary, interview*),
- expert (is an expert mentioned or cited?, boolean),
- politician I–III (the name of a mentioned politician),
- sentiment I–III (*negative, neutral, positive, acclaim, hateful, missing*),

These attributes are bound to a document as a whole, but the current aim is to bind them also with the particular words and phrases from the documents to be able to capture significant correspondences which will be used in the subsequent machine learning techniques for automatic annotation (see Section 6).

The second group of attributes consists of document-level attributes, which generally cannot be reflected in concrete phrases of the document:

- topic (*migration crisis, domestic policy, international relations/diplomacy, society, other, energetics, Ukraine conflict, culture, Syria conflict, warfare policy, economics, conspiracy*),
- focus (*foreign, domestic, both, indeterminate*),
- opinion (does the article states an opinion?, boolean),
- overall sentiment (*positive, neutral, negative*),
- picture (does the article contain a picture/photo?, boolean),
- video (does the article contain a video, boolean), and
- source (does the article mention or cite a source?, boolean).

**Fig. 1.** Advanced full-text search entry form with two selected attributes of “*Argumentace*” (*argumentation*) and “*Obsažené emoce*” (*present emotions*) with the values of “*ano*” (*yes*) and “*rozhořčení*” (*outrage*).

## 4 The Propaganda Corpus

In the first phase, we were provided with a list of processed web document annotations prepared by experts. We have downloaded all the referenced documents and extracted, using templates specific for each of the four servers, the texts of the articles and additional metadata – the document title, text of the leading article, date of publication, and, where possible, the author of the document.

Contrary to our expectations, the format of the pages is not stable in time and the templates needed to be updated for every new batch of articles. The changes consisted of different nesting of the elements and their identifiers, some of the text could not be discovered using the old template and conversely, some unwanted boilerplate would be included.

The extracted text was then tokenized using *unitok* [11] and morphologically annotated using *majka* [12] and *desamb* [13].

## 5 The Annotation Editor

The editor is designed as a server-client application ready-to-use without installation via standard web browser. The server part is written using a Python web library `bottle.py`<sup>1</sup>, the client part uses the `jQuery`<sup>2</sup> and `Bootstrap`<sup>3</sup> libraries.

<sup>1</sup> `bottlepy.org`

<sup>2</sup> `jquery.org`

<sup>3</sup> `bootstrap.org`

The screenshot shows a search engine interface with a dark header. The header contains the word 'Propaganda' in white and a blue button labeled 'test'. On the right side of the header is a hamburger menu icon. Below the header, the search results are displayed. The title is 'Výsledek vyhledávání'. The search query is 'Dotaz Putin'. Below the query, it says 'Zobrazeno prvních 100 výsledků z 528 v 5322 dokumentech.' The search results are listed in a table with seven rows, each containing a snippet of text from a document.

Search Result Snippet
'Jak může garantovat bezpečnost režim, kde mizí novináři?', 'Pane ministře, nechte toho!' Stropnický a Filip se v ČT tak hádali, že je musel roztrhávat Moravec
Naštvaná Ukrajina a zvěsti o novém Majdanu. Moldavsko žije volbami
'Sedí prase na Hradě, libuje si v úřadě.' Co chcete od společnosti, která tomu tleská! Čechokanaďan Jírovec ukazuje v zásadní věci na Schwarzenberga
'Systém EET zkolabuje a přivodí Babišovi porážku!' uvedl v rádiu senátor Valenta. Promluvil také o ParlamentníchListech.cz
Válka na Krymu, pravda o Savčenkové, tajnosti kolem malajsijského letadla. Pozorovatel přináší žhavé novinky z Ukrajiny
Víme, co marketingoví odborníci říkají o ruském dokumentu Prezident v hlavní roli s Vladimírem Putinem
Volby v USA: Co jste nevěděli o pánovi od Hillary Clintonové. Tu prý platí velké korporace, Trumpa drobní střádalové. No a Putin...

Fig. 2. Example search result

## 5.1 The Server Side

The documents are stored in so-called vertical format<sup>4</sup> where the main structure “<doc>” contains the structure attributes corresponding to the annotation scheme. Only a few extra attributes are processed automatically and added to the metadata, namely the second level web domain, top level domain, document word count, original URL and the date of publication (if available on the webpage).

The data is indexed by the modular corpus manager *manatee* [14] which allows fast full-text search within the documents.

The annotation data which comes from the client during manual annotation is stored in a separate SQLite database. The document and range attribute annotations are stored separately.

The HTTP API is maintained by a Python HTTP server provided by the *bottle* library. It can serve either JSON outputs or HTML responses generated by a template engine which is a part of the library.

<sup>4</sup> <https://www.sketchengine.co.uk/documentation/preparing-corpus-text/>

ak ať drží hubu. A další výroky z drsné

u. A další výroky z drsné politické besedy

granty a příležitost pro čečenskou vládu, ukrajinské setkání příznivců a členů Úsvitu – národní koalice a Bloku / desítky lidí, aby si mimo jiné poslechly i bezpečnostního

Idát na hejtmana Plzeňského kraje **Roman Bakalá**, v tebný, protože jsou tam dvě firmy, které dělají halal čné názvy ulic. Pokud menšina přesáhne deset procent v mě týká Slezska, kde je velká polská komunita, ale když si dočkáme toho, že některá náměstí a ulice budou možná oval místní přítomné, za což si zasloužil hlasitý šum v sále.

: „Zažil jsem islamizaci v Iráku a Kuvajtu. V Kuvajtu chodily ních. Dnes už to není možné. Prakticky je situace taková, trstva, jejich salaafistické Ideologie a výsledek je takový, že echny příkazy Koránu, i když některé příkazy nejsou ani ladu v Kuvajtu došlo k tomu, že jeden profesor islámského

Atributy s rozsahem	
Místo	Ceska republika ▼
Vina	ano ▼
Nálepkování	ano ▼
Argumentace	ano ▼
Obsažené emoce	strach ▼
Démonizace	ano ▼
Relativizace	ne ▼
Strach	ano ▼
Fabulace	ano ▼
Rusko	pozitivni priklad ▼
Žánr	zpravodajstvi ▼
Odborník	ano ▼
Politik 1	<input type="text"/>
Vyznění 1	pozitivni ▼
Politik 2	<input type="text"/>
Vyznění 2	negativni ▼
Politik 3	<input type="text"/>
Vyznění 3	negativni ▼
Atributy dokumentu	
Téma	migracni krize ▼
Zaměření	domaci ▼
Názor	ne ▼

Fig. 3. Annotation of range attributes, the highlighting

## 5.2 The Client Side

The client serves as the central point of all operations related to the processing of the corpus data:

- 1) basic and advanced full-text search,
- 2) viewing of documents with their metadata, and
- 3) annotation of document and range attributes.

The simple search operation accessible on every page (in the top bar) can be used for searching a word or a phrase. What is important is that not just the exact word or phrase is searched for but also their other word forms. E.g. for query “*Putin*” the system will retrieve documents containing the word in other cases (useful especially for morphologically-rich languages as Czech).

The advanced search allows to combine this full-text search with querying the metadata annotated in the documents. Each attribute from the annotation schema can be used together with a particular value which narrows the results of such a search. Figure 1 shows an example advanced query specification with two attributes “*Argumentace*”—*argumentation* and “*Obsažené emoce*”—*present emotions* with two particular values “*ano*”—*yes* and “*rozhořčení*”—*outrage*, respectively. An example of search results is shown in Figure 2.

Any document from the corpus (accessible through the search results or from a preassigned list denoted by the user name) can be viewed together with its annotations. When viewing the document, it is possible to directly annotate both range and document attributes. Once logged in (the editor does not allow anonymous editors to save any annotations), the document attributes can be

Neméně šokující bylo další Kerryho prohlášení <sup>✖ politik2: John Kerry</sup> Kerry v němž uvedl, že do voleb v Sýrii by měl být zapojen současný <sup>✖ politik1: Bašár Asad</sup> prezident Bašár Asad, a to navzdory skutečnosti, že oficiální <sup>✖ politik3: Barack Obama</sup> postoj Obamy s něčím takovým nepočítá – Bílý dům trvá na odstoupení syrského vůdce.

Fig. 4. Visualization of annotated ranges in documents.

annotated by selecting them in the right hand table containing all attributes and their values (together with the immutable attributes like URL, date etc.).

The range attributes can be annotated once a phrase selection (range) is highlighted: by clicking on first word of a sequence and then on the last word of the sequence, the range is highlighted and a value can be set to a particular attribute in the highlighted part of the metadata table. An example of this operation can be seen in Figure 3.

When viewing documents, badges are shown next to the annotated attributes and a user can open the appropriate ranges with values by clicking the badges. The annotated ranges can be easily removed (see Figure 4).

## 6 Future Development and Conclusions

With the first version of the editor developed, the new (and possibly also the previous annotations) can express the manipulative features with precise references on word level which will provide valuable data for a) exact specification of the particular subtasks, such as the identification of the propagandistic phenomena in texts, and b) training automatic methods: features from attributes, their values and words within annotated ranges will be extracted and used within selected machine learning techniques. This will allow the next version of the system to be able to pre-annotate the document-level and word-level attributes so annotators have them at the disposal to verify and/or amend them.

**Acknowledgements.** This project was partially supported by the Grant Agency of Masaryk University within the project MUNI/G/0872/2016.

## References

1. Castillo, C., Davison, B.D., et al.: Adversarial web search. *Foundations and Trends in Information Retrieval* 4(5) (2011) 377–486
2. Metaxas, P.T.: Web spam, social propaganda and the evolution of search engine rankings. In: *International Conference on Web Information Systems and Technologies*, Springer (2009) 170–182

3. Lumezanu, C., Feamster, N., Klein, H.: # bias: Measuring the tweeting behavior of propagandists. In: Sixth International AAAI Conference on Weblogs and Social Media. (2012)
4. Weedon, J., Nuland, W., Stamos, A.: Information operations and Facebook. Facebook on-line report (April 27, 2017)
5. Wakabayashi, D., Isaac, M.: In Race Against Fake News, Google and Facebook Stroll to the Starting Line. *The New York Times* 4 (2017)
6. Herman, E.S., Chomsky, N.: A propaganda model. *Manufacturing Consent: The Political Economy of the Mass Media* (1988)
7. Lee, A., Lee, E.B.: *The fine art of propaganda*. (1939)
8. Holz, F., Teresniak, S.: Towards automatic detection and tracking of topic change. *Computational linguistics and intelligent text processing* (2010) 327–339
9. Vosoughi, S.: Automatic detection and verification of rumors on Twitter. PhD thesis, Massachusetts Institute of Technology (2015)
10. Janze, C., Risius, M.: Automatic detection of fake news on social media platforms. (2017)
11. Michelfeit, J., Pomikálek, J., Suchomel, V.: Text tokenisation using unitok. In Horák, A., Rychlý, P., eds.: *RASLAN 2014*, Brno, Czech Republic, Tribun EU (2014) 71–75
12. Šmerk, P.: Fast Morphological Analysis of Czech. In Sojka, P., Horák, A., eds.: *Third Workshop on Recent Advances in Slavonic Natural Language Processing*, Masarykova univerzita, Masaryk University (2009) 13–16
13. Šmerk, P.: K počítačové morfologické analýze češtiny (in Czech, Towards Computational Morphological Analysis of Czech). PhD thesis, Faculty of Informatics, Masaryk University (2010)
14. Rychlý, P.: Manatee/Bonito – a modular corpus manager. In: *1st Workshop on Recent Advances in Slavonic Natural Language Processing*. (2007) 65–70