# Visualizing fakenews as reported by euvsdisinfo.eu

Jesper Henrichsen
Supervised by: Pedro Ferreira

December 8, 2017

## Contents

# 1 Introduction

Since the introduction of the internet, information has been able to reach further and faster than ever before. Recent developments of omnipresent social media has created the perfect platform for the spread of this information within the internet. For the sake of advertisement social media platforms has, since their introduction, only become better at targeting their audience with information that will capture their attention. As put by the team behind the newsfeed at Facebook: "*The goal of News Feed is to deliver the right content to the right people at the right time (...)*". However, a relevant question would be to whom it is implied to be *right* for.
That it is not necessarily the user, has recently emerged as a likely answer.

Because of the nature of social media, it has proved an effective way to spread misinformation. This report will describe the process of scraping and visualizing information from euvsdisinfo.eu.

# 2 Background

Since 2015 the campaign `euvsdisinfo.eu`, has been run by the European External Action Service East Stratcom Task Force. The primary focus of the campaign is to identify and debunk pro-Kremlin disinformation. According to the The campaign enlists cases of debunked information from sources spanning from official news sites to twitter accounts, to non-online content such as interviews. The campaign included more than 3400 such cases that became part of the dataset this report is based on. The cases are all reported either by the campaign staff themselves, or one of the 400 collaborating organizations and individuals. In **??** is seen how cases are listed on the website, the order of which is chronological with respect to the date it was reported.

Figure 1: Example listing of disinformation cases on euvsdisinfo.eu

| | | | |
|---|---|---|---|
| 02.11.2017 | Brussels are closing the door they opened with visa freedom for Georgians | Rezonansi | Europe, Georgia |
| 02.11.2017 | EU wants to ban information about the of country of origin on the food labels | Vlastenecké Noviny, eOdborar.cz | Italy, EU |
| 02.11.2017 | The West supported the terrorists during hostage crisis in Dubrovka in Moscow ("Nord-Ost attack") in 2002 and in the school of Beslan, North Ossetia, in 2004 | Vremya pokazhet @Pervyi kanal, 19:15 | Russia, The West |
| 02.11.2017 | The US destroyed the European values and culture, so now Russia is the only flagship of the European civilization | Vremya pokazhet @Pervyi kanal' TV-channel, 41:08 | Europe, US |
| 01.11.2017 | The EU teaches journalists how to properly inform about Islam and migrants | ac24.cz | Ireland, Italy, Austria, Slovenia, Hungary, Greece, EU, Germany, Spain |
| 31.10.2017 | Finland wants Russia to join the European Centre of Excellence for Countering Hybrid Threats | Sergei Lavrov, Russian Foreign Ministry's website | Russia, Finland |
| 31.10.2017 | The West, primarily the United States, is collecting biological material in Russia to create a biological weapon that destroys the Russians. | Mesto vstrechi @NTV TV-channel, 1:12:48 | Russia, The West, US |
| 31.10.2017 | Czech MP prefers there were 5 million Muslim migrants in Czech Republic rather than 5 million voters of Czech president Miloš Zeman | BezPolitickeKorektnosti | Czech Republic |
| 31.10.2017 | Estonia has opened a new military base in the town of Tapa under the pretext of fear of a Russian attack. | cz.Sputniknews | Russia, Baltic states, Estonia |
| 31.10.2017 | US is using sanctions in trying to push Russia out of the European energy and arms market | cz.sputniknews | Europe, Russia, US |

Figure 2: An example of the information for each case

# EU wants to ban information about the of country of origin on the food labels

## Summary of Disinformation

EU wants to ban information about the of country of origin on food labels. This measure is applied in order to fight against the nationalistic tendencies in member states.

VIEW ORIGINAL PUBLICATION / MEDIA

## Disproof

The original text from Italian website Afffaritalini is about a quarrel between the European Commision and Italy, since Italy started to demand from producers to write on food labels where the product was produced and packed.

According to the European Commission, the current approach is that the country of origin or place of provenance labelling on food is voluntary, unless its absence could mislead consumers.

The Regulation introduces mandatory origin labelling for fresh meat from sheep, goat, poultry and pigs. As of 1 April 2015, with some exemptions, the Member State or third country where the animal was reared and slaughtered will appear on the label of such meats.

For foods bearing origin indications, the country of origin or place of provenance of the main ingredients must also be listed if those ingredients originate from a different place than the declared origin of the finished product. For example, butter churned in Belgium from Danish milk could be labelled as "produced in Belgium from Danish milk." The application of these rules is subject to the adoption of implementing acts which have not yet been adopted by the Commission.

Those rules intend to protect consumers from misleading origin indications and will ensure a level playing field between food business operators.

**Reported in:**
Issue 86

**Date:**
02.11.2017

**Language:**
Czech

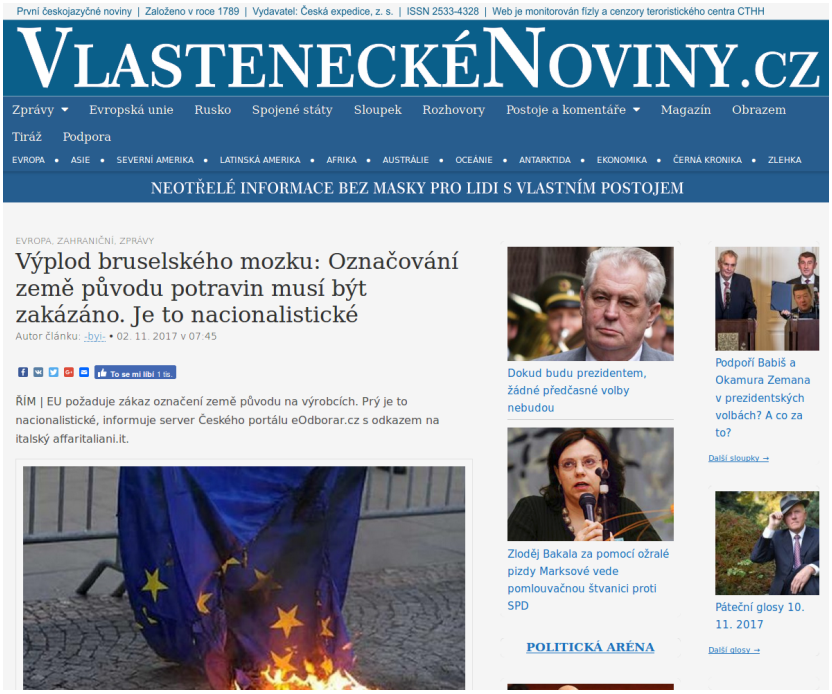
**Country of Origin:**
Italy, EU

**Reported by:**
Prague Security Studies Institute (PSSI)

**Keywords:**
European Commission

**Disinforming outlet:**
Vlastenecké Noviny, eOdborar.cz

Figure 3: An example of a debunked news article from a Czech website




# 3 Acquiring information

In this section, the approach to acquiring and extracting information for the visualization will be described.

## 3.1 Article scraping and content extraction

A general approach to extracting content is difficult because source of information can be any sort of media, whether digital or physical, in writing or a video. And because each website is different, then even building a scraper to extract the content of the subset that are online articles, will be difficult. Furthermore, in order to consider the names of locations that are being mentioned in the articles, the entire content is not necessary. For this reason I chose to use the meta tags to extract information about the content, summary, titles and descriptions. I would also filter on html tags such as the header tags, h1, h2, h3, h4, h5, h6 as well as the title tag used for setting the window title. I found through experiment that only considering the first found header tag worked best, the reason is that other header tags than the first one would often be titles of other news articles that the news site wants the user to click on.

The meta tag proved to be a very reliable way, since most news sites are interested in their content being shared, so its almost a must, in order to get

content shared on social media.

In order to avoid duplicating content as much as possible I compared the content of a meta tag or html tag to the content that was already found in previous html tags. This approach resulted most often in a short paragraph of information about the article, including keywords, title and summary. Before applying the approach to the articles scraped from euvsdisinfo.eu, it was tested on sampled articles, such as danish news outlets. It was also tried on a sample of articles taken from the subreddits: r/politics, r/news, and r/worldnews, because these were good collections of all different news articles and news outlets. The results was used to evaluate qualitatively on the smaller sample. None of which turned out to have no content, and only one resulted in an extract only consisting of keywords. However, interestingly among the keywords were also the name of the country that the news story revolved around. One thing that was clear from this however, was that locations were not always mentioned if the news revolved around well known state leaders such as Putin, Merkel or Trump. In such cases, only the names of the state leaders were present as indication of what countries were mentioned in the articles.

### 3.2 Named entity recognition

Named entity recognition is a research field within natural language processing concerned with recognizing names of things such as people, company names, or, as relevant for this project, locations. The study of recognizing location names within texts is one of the most studied ares of named entity recognition, however, still an ongoing research. A more in depth explanation of named entity recognition or even natural language processing, is however, beyond the scope of this project. Instead for the purposes of this project, I will rely on existing tools.

## 4 Results

## 5 Discussion

## 6 Conclusion