<https://www.buzzfeednews.com/article/craigsilverman/these-are-50-of-the-biggest-fake-news-hits-on-facebook-in> <https://spcollege.libguides.com/c.php?g=254319&p=1695321>

<https://www.theonion.com/> <https://mediabiasfactcheck.com/the-onion/>

A close up of a logo

Description automatically generated

Notice the fake news have longer headlines compared to real ones. A title presents the content to be analysed in the article in usually at most 10 words. That is because it is only stated the topic without any further details. Words such as a, an, and, they, etc are avoided as well as unnecessary adjectives and repetitions. For instance, “Great win today of Team A against Team B for the League” is unnecessarily long, while “Team A win” is good enough and all the rest are included in the text.

On the other hand, when one is writing headlines for fake news, he has to make sure that the headline can draw reader’s attention. The use of exclamation marks, 3 dots, capital letters, large numbers and adjectives that show exaggeration is very common in this case. Furthermore, the title usually includes some extra information just to make the reader want to know more about it. For instance, “Elderly woman accused of training her 65 cats to steal from neighbors” is a nice bait to “force” the reader to click on this article.

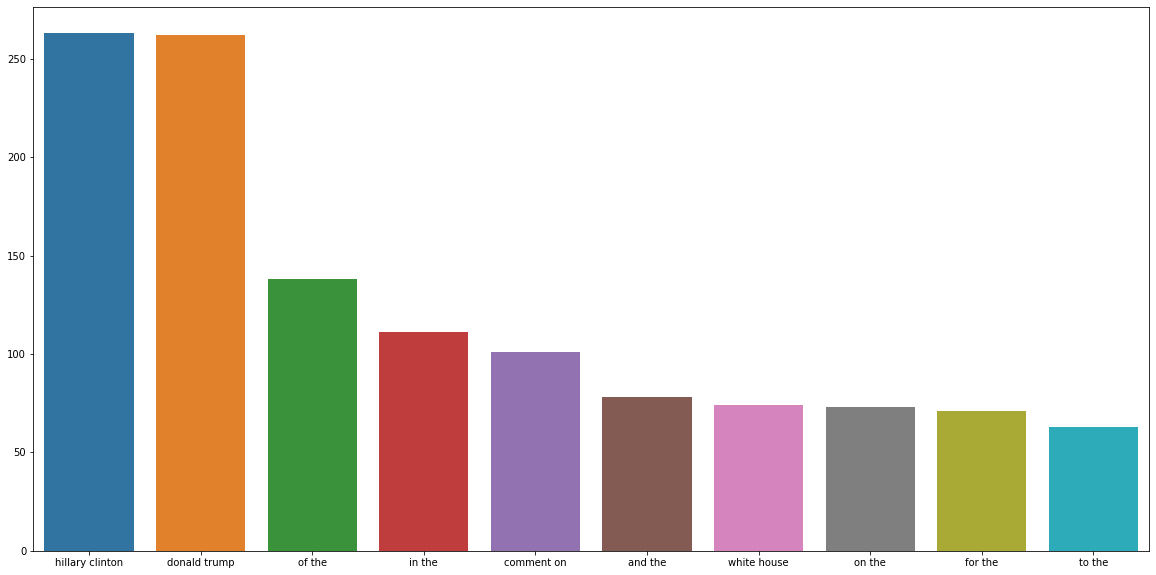
That explains quite accurately the above graph. Another point to mention is that real news headline’s length has greater consistency while for the fake ones there is bigger standard deviation. This is because writing a proper headline for an article follows some ground rules. On the contrary, fake news headline only requires drawing attention so it does not have a specific formal or structure.

A screenshot of a cell phone

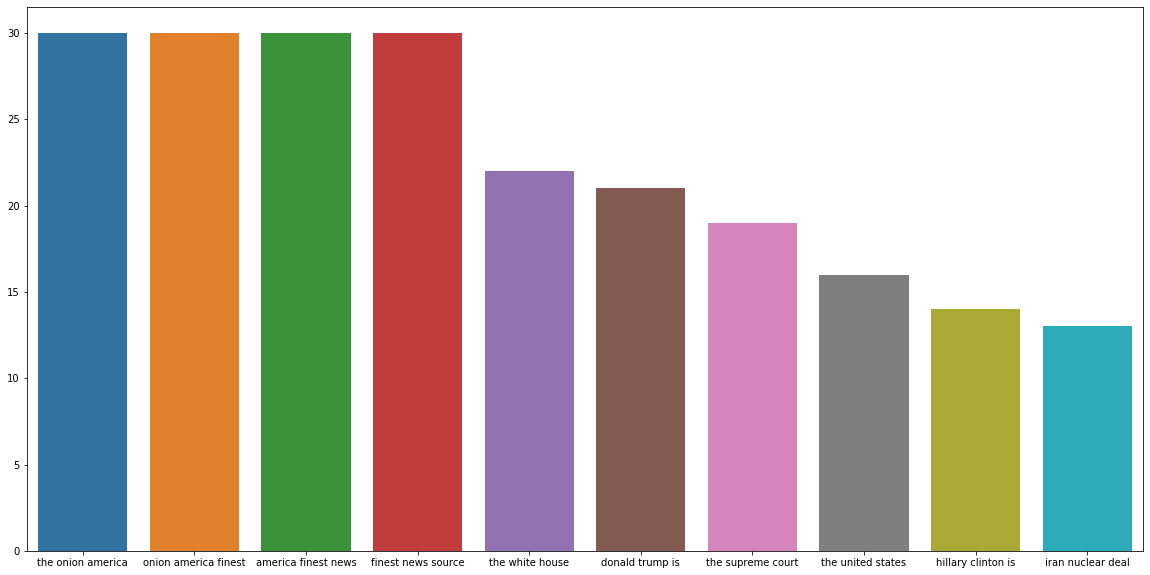
Description automatically generated

This graph can be misleading. It seems that the word “trump” appears too many times and one can guess that it refers to Donald Trump. By making a quick search on the data we can observe that this is not always that case. We find many words such as “trumpet”, “trumped”, “trumper”, “trumps”, etc. Same thing applies for the word “clinton”. One can guess that it refers to Hillary Clinton, which is true for most cases but not for all of them. Again, by doing some search we can see that it can also be Bill Clinton or sometimes it is just Clinton without specifying for which one of them it refers to. Furthermore, notice the word “us” in the fourth place which can either refer to the country USA or it could just mean ourselves. Overall though, in most cases it refers to Donald Trump, Hillary Clinton and USA.

Another very important point we need to make here is that we observe other key words such as “Obama”, “war”, “Iraq”, “police”, “election”, etc. We can easily conclude that the majority of articles have serious political content related to crime, violence, war, propaganda, foreign policy, etc. These are the main topics where fake news are spotted. All the above-mentioned subjects are controversial due to people with different political beliefs may have different opinions around them. In addition, it is usually quite hard for one to know all the facts about an incident and so even harder to distinguish between real and fake facts. Moreover, these are all sensitive issues and the most naïve ones can easily get fanatic about them just because they believe some fake news they read. With these arguments we can justify the high proportion of fake news in our data.



This graph does not contain anything surprising; it has what we expected. The top bigrams are of course Hillary Clinton and Donald Trump. The rest are just stop words apart from the “White House”. We cannot draw any new conclusions as bigrams such as “on the” or “and the” would appear pretty much everywhere. However, the next graph contains more meaningful information.



The top trigrams have to do with “The Onion”. This is an American journalistic site which calls itself “America’s Finest News Source”. Just by reading a few headlines in its home page and conducting a quick research one can realise that all its news is fake. Apparently, its only purpose is to make satire, making fun of both Democrats and Republicans. It is one of the most popular websites for fake news on the internet and so their income comes from online advertising. Anyone who does not have critical thinking and does not know what “The Onion” is all about can mistake it for a reliable source of information.

This reinforces the previous argument that it is easier to create fake news for political topics compared to any other. When we started searching about The Onion we did not know what it is about but we started realising something is not right when we read a sport’s headline. Whatever happens to sports is a clear fact, so it is much harder to produce fake news for this kind of topic.