

AGH

**AKADEMIA GÓRNICZO-HUTNICZA
IM. STANISŁAWA STASZICA
W KRAKOWIE**

**Projekt na przedmiot Zaawansowane Techniki Integracji Systemów
Katedra Informatyki
Wydział Elektroniki Informatyki i Telekomunikacji AGH**

**Tytuł projektu “Opracowanie rozszerzalnego środowiska do
agregacji i analizy informacji pochodzącej z różnych portali
informacyjnych”**

Autorzy: Łukasz Pająk, Przemysław Furtak

DOKUMENTACJA

Zaawansowane techniki integracji systemów

Temat: Opracowanie rozszerzalnego środowiska do agregacji i analizy informacji pochodzącej z różnych portali informacyjnych

Spis treści:

[Cel](#)

[Plan pracy:](#)

[Architektura](#)

[Obraz 1. Schemat architektury.](#)

[Pobieranie danych](#)

[Obraz 2. Lista RSS feedów.](#)

[Baza danych](#)

[Obraz 3. Schemat bazy danych](#)

[Analiza danych](#)

[Komunikacja](#)

[Algorytmy](#)

[Testy](#)

[Wyniki](#)

[Analiza Statystyczna](#)

[Obraz 4. Wykres przedstawiający najczęściej występujące słowa w badanym przez nas zbiorze - z wyłączeniem 'stopwords' - słów przestankowych/nieznaczących np. 'the' 'as' 'on' 'an'.](#)

[Obraz 5. Statystyka przedstawiająca ilość feedów per źródło.](#)

[Obraz 6. Wykres przedstawiający rozkład ilościowy feedów z podziałem na źródło.](#)

[Obraz 7. Wykres kołowy przedstawiający rozkład ilościowy z podziałem na źródło.](#)

[Analiza z podziałem na kontynenty](#)

[Obraz 8. Wykres przedstawiający najczęściej używane słowa w USA, pominęte zostały wyrazy ze stop listy.](#)

[Obraz 9. Wykres przedstawiający najczęściej używane słowa w kontynentalnej europie, pominęte zostały wyrazy ze stop listy.](#)

[Obraz 10. Wykres przedstawiający najczęściej używane słowa w Afryce, pominęte zostały wyrazy ze stop listy.](#)

[Obraz 11. Wykres przedstawiający najczęściej używane słowa na bliskim wschodzie i w Azji \(Japonia, Chiny\), pominęte zostały wyrazy ze stop listy.](#)

[Ogólna analiza słów kluczowych](#)

[Obraz 12. Word cloud dla newsów z USA.](#)

[Obraz 7. Stopień słów dla newsów z Wielkiej Brytanii.](#)

[Wnioski](#)

[Bibliografia](#)

[* Lista feedów RSS - plik rsslist](#)

1. Cel

Celem naszego projektu jest wykonanie frameworka służącego do agregacji danych pochodzących z różnych portali informacyjnych. Pobierane będą w formie feedów RSS. Dane te zamierzamy analizować pod kątem częstości występowania słów kluczowych. Jednym z założeń naszej pracy jest łatwość i dostępność modyfikacji przez użytkowników - chcemy aby rozszerzanie, ograniczanie oraz zmienianie źródeł było zadaniem łatwym i przyjaznym użytkownikom. Dzięki naszemu komponentowi służącemu za interfejs pomiędzy danymi z feedów RSS a bazą danych, umożliwiamy każdemu zainteresowanemu tą tematyką wykonanie swoich własnych analiz, bez konieczności znajomości algorytmów czy zagłębiania się w szczegóły implementacyjne projektu.

Biorąc pod uwagę ilość oraz jakość pobieranych danych, naszym celem w temacie analizy tekstu jest wyszczególnienie najczęściej występujących słów kluczowych. Wytworzenie takiej bazy może służyć za fundament w dalszej analizie języka naturalnego. Ze względu na źródło pochodzenia nasze informacje to suche, krótkie notki dziennikarskie, przez co nie mogliśmy wykonać analiz charakteru czy sentymentu prezentowanej wiadomości.

2. Plan pracy

1. Zdefiniowanie portali informacyjnych.

Ze względu na słabą dostępność API umożliwiających pobieranie wiadomości na portalach informacyjnych zdecydowaliśmy się wykorzystać kanały RSS. Really Simple Syndication to rodzina języków znacznikowych bazująca na XMLu. Większość portali informacyjnych udostępnia kanały RSS co daje szeroki wybór źródeł informacji. Ponadto kanały RSS udostępniają wszystkie informacje wymagane do stworzenia planowanej przez nas aplikacji służącej agregacji i analizie informacji czyli tytuł, kategorie wiadomości, krótki opis oraz link prowadzący do pełnej informacji.

2. Wybranie kilku testowych portali informacyjnych.

Wybór portali informacyjnych z których będziemy korzystać pozostaje do uzgodnienia. Planujemy wybrać trzy lub cztery źródła - najlepiej takie, które reprezentują skrajnie odmienne poglądy polityczne (partie i przekonania). Przy wyborze planujemy kierować się stroną <https://rss.com/popular-rss-feeds/> gromadzącą najpopularniejsze kanały RSS oraz przekazanym nam projektem geomedia rozwijanym w poprzednich latach.

3. Przeanalizowanie struktury przechowywanych i udostępnianych informacji.

W związku z mnogością istniejących portali informacyjnych, chcemy uzyskać pełny obraz zbieranych danych. Zamierzamy zbierać jedynie podstawowe elementy opisujące artykuły na tychże portalach. Planujemy skupić się na przygotowaniu prostej, niezawodnej struktury, tak aby każdy korzystający z aplikacji wiedział co, i gdzie się znajduje bez konieczności pisania instrukcji dotyczących zawartości bazy.

4. Wybranie bazy danych oraz zdefiniowanie struktury agregowanych danych.

Zamierzamy wykorzystać jeszcze nie sprecyzowaną technologię bazodanową, obecnie zastanawiamy się nad wykorzystaniem sqllite. Będziemy agregować i przechowywać dane zbierane za pomocą feedparsera - biblioteki Pythona. Zamierzamy położyć nacisk głównie na komplementarność zbieranych danych. Wolimy uzyskać pełny obraz tego co zbieramy, z mniejszej ilości portali, niż uzyskiwać szczątkowe, niepełne informacje z wielu, niepowiązanych ze sobą portali. Dane które obecnie chcemy przechowywać to: tytuł artykułu, jego kategorię, datę zamieszczenia, autora, odnośnik do pełnej wersji. Pozostałe informacje zamierzamy określić w ramach implementacji.

5. Stworzenie rozszerzalnej aplikacji w języku Python umożliwiającej dodanie większej ilości connectorów do innych portali informacyjnych.

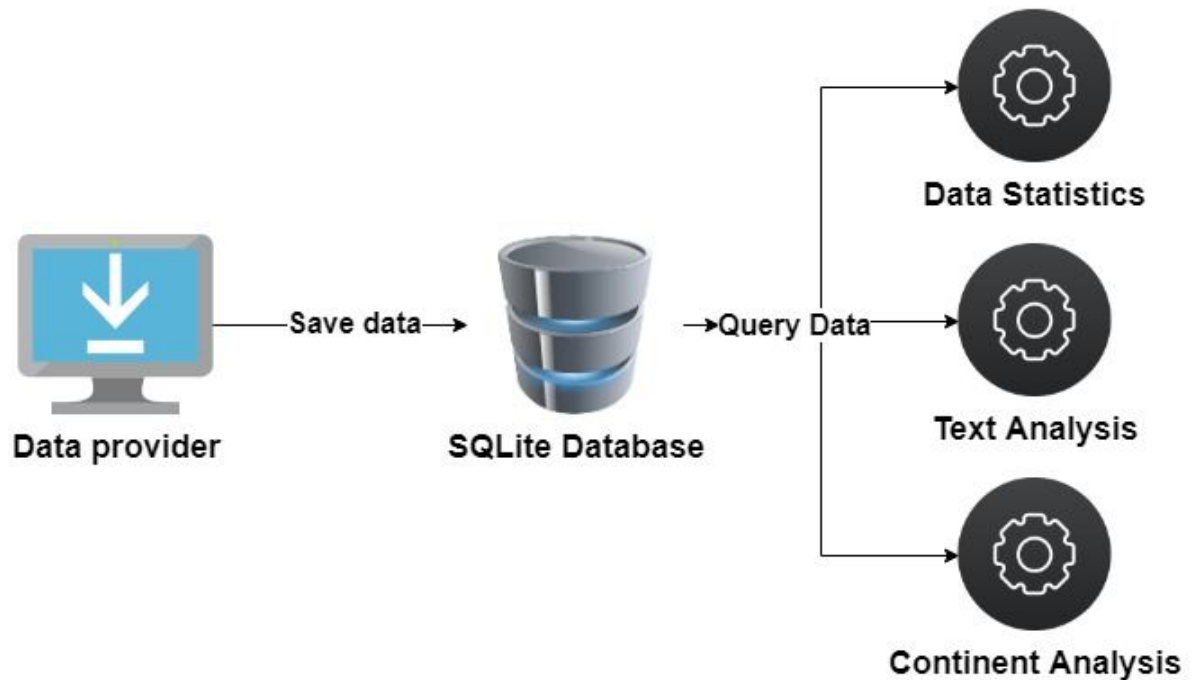
Aplikacja główna napisana zostanie w języku Python. Będzie implementacją wzorca strategii umożliwiającej rozwijanie projektu o nowe implementacje połączeń do innych portali informacyjnych. Jedna implementacja connectora będzie umożliwiała pobranie danych z jednego portalu informacyjnego i dopasowanie ich do zdefiniowanego przez nas schematu przechowywania wszystkich informacji. Do pobierania informacji z kanału RSS planujemy wykorzystać gotową bibliotekę pythonową feedparser - https://www.tutorialspoint.com/python/python_reading_rss_feed.htm, która w prosty sposób umożliwia pobieranie i przeglądanie informacji z kanałów RSS.

6. Przetestowanie działania aplikacji w połączeniu z bazą danych na wybranych wcześniej portalach informacyjnych.

Zamierzamy przeprowadzić testy naszej aplikacji na żywo, tj. uruchomić ją na okres 24 godzin, bądź dłużej, w zależności od potrzeb. Chcemy przetestować działanie naszej aplikacji, to czy poprawnie pobieramy i zapisujemy dane, jaką ilość wartościowych informacji udaje się nam przechwycić, oraz czy są one odpowiednio przygotowane do dalszej analizy.

3. Architektura

Cały projekt składa się z trzech głównych komponentów, które są odpowiedzialne za podstawowe części projektu. Została ona przedstawiona na Obrazie 1.



Obraz 1. Schemat architektury.

3.1. Pobieranie danych

Pierwszy komponent to prosta aplikacja napisana w języku Python służąca do pobierania danych z kanałów RSS i zapisywaniu ich do bazy danych. Służy do wybierania informacji pobieranych z feeda tak by nie zapisywać wszystkich niepotrzebnych oraz nadmiarowych danych. Dane te w kolejnych pętlach działania programu zapisują przefiltrowane informacje do bazy danych w postaci osobnych kolumn w tabeli. Działanie tego programu sterowane jest plikiem tekstowym *rsslist*, który zawiera wszystkie kanały RSS dostarczające informacje. Fragment zawartości tego pliku przedstawia Obraz 2.

```

$ cat rsslist
BBC
http://feeds.bbc.co.uk/news/politics/rss.xml
The Daily Telegraph
https://www.telegraph.co.uk/news/rss.xml
The Guardian
https://www.theguardian.com/politics/rss
Washington Post
http://feeds.washingtonpost.com/rss/politics
Japan Times
https://www.japantimes.co.jp/feed/topstories/
The Times of Israel
https://www.timesofisrael.com/feed/
Times of India
https://timesofindia.indiatimes.com/rssfeedstopstories.cms
New Zealand Herald
http://rss.nzherald.co.nz/rss/xml/nzhtsrsscid_000000698.xml
China Daily
http://www.chinadaily.com.cn/rss/world_rss.xml
CNN - Top Stories
http://rss.cnn.com/rss/edition.rss
CNN - Americas
http://rss.cnn.com/rss/edition_americas.rss
CNN - Asia
http://rss.cnn.com/rss/edition_asia.rss
CNN - Europe
http://rss.cnn.com/rss/edition_europe.rss
CNN - Middle East
http://rss.cnn.com/rss/edition_meast.rss
CNN - Africa
http://rss.cnn.com/rss/edition_africa.rss

```

Obraz 2. Lista RSS feedów.

3.2. Baza danych

Drugim komponentem jest baza danych służąca do trwałego przechowywania pobranych danych. Dane pobieramy przyrostowo uruchamiając każdorazowo program pierwszego komponentu. To również on jest odpowiedzialny za stworzenie bazy danych i tabeli na początku działania projektu oraz ewentualne usunięcie jej na końcu. Wykorzystana przez nas baza danych to relacyjna baza SQLite. Schemat bazy danych przedstawia Obraz 3.

Columns	Column Name	#	Data Type	Not Null	Auto Increment	Default
123 source	123source	1		<input type="checkbox"/>	<input type="checkbox"/>	
123 title	123title	2		<input type="checkbox"/>	<input type="checkbox"/>	
123 summary	123summary	3		<input type="checkbox"/>	<input type="checkbox"/>	
123 publication_date	123publication_date	4		<input type="checkbox"/>	<input type="checkbox"/>	
123 author	123author	5		<input type="checkbox"/>	<input type="checkbox"/>	

Obraz 3. Schemat bazy danych

3.3. Analiza danych

Trzeci i ostatni komponent wykorzystywany jest w naszym projekcie do analizy pobranych przez nas danych. Ze względu na łatwość uruchomienia i modyfikacji złożonych części kodu wykorzystaliśmy tutaj jupyter notebooks. Umożliwiają one niezależne uruchomienie segmentów z zaimplementowanymi algorytmami do analizy tekstu.

Aby pokryć jak największą część publikowanych feedów, wykonujemy skrypt do pobierania danych co godzinę - dzięki temu, oraz zaimplementowanym mechanizmie usuwania duplikatów - uzyskujemy bardzo dokładny obraz wszystkich pobieranych przez nas serwisów.

W naszym projekcie skorzystaliśmy z trzech osobnych notebooków, są to odpowiednio:

- A. Data statistics - workbook odpowiadający za przeliczenie statystyk głównych dla naszego zbioru danych - ile feedów przypada na dany kontynent, na źródło, jak często feedy są publikowane.
- B. Text analysis - workbook odpowiadający za analizę tekstu - najczęściej występujących wyrażań, zbitek słów, najważniejszych słów
- C. Continent Analysis - workbook odpowiadający za analizę względem geograficznym - kontynentalny podział feedów, najczęściej wykorzystywane słowa na badanych kontynentach, różnice pomiędzy najczęstszymi słowami na różnych szerokościach geograficznych

4. Komunikacja

Cała komunikacja i przepływ danych w naszym projekcie odbywa się za pośrednictwem relacyjnej bazy danych SQLite. Komponent służący do pobierania danych dba o przyrostowe dopisywanie ich do tabeli natomiast komponent służący do analizy każdorazowo odczytuje dane nie modyfikując ich.

Na godzinę uzyskujemy przyrost średnio około 5 feedów. Fluktuacje godzinne wynikają z różnych stref czasowych w których znajdują się centrale portali z których wiadomości korzystamy. Pojedynczy portal informacyjny wysyła średnio 84 feedy tygodniowo czyli około 12 dziennie.

5. Algorytmy

W projekcie planujemy wykorzystać bibliotekę pythonową NLTK (Natural Language Toolkit) umożliwiającą analizę tekstu oraz bazującą na niej RAKE-NLTK. Ta druga pozwala na automatyczną ekstrakcję słów kluczowych. Do analizy tekstu będziemy korzystać z kolumny summary w naszej bazie. Zawiera ona najwięcej tekstu opisującego informację pobraną z feeda.

Metody które obecnie wykorzystujemy w naszej analizie:

- `get_ranked_phrases` - metoda służąca do wydobywania słów kluczowych
- `get_ranked_phrases_with_scores` - metoda służąca do wydobywania słów kluczowych wraz z punktowaniem
- `get_ranked_phrases_with_scores with custom made stopwords and omitment of certain punctuations` - metoda służąca do wydobywania słów kluczowych przy niestandardowych zmiennych
- `get_word_frequency_distribution` - metoda służąca do wydobywania dystrybucji częstości występowania słów
- `get_word_degrees` - metoda służąca do wydobywania stopnia

6. Testy

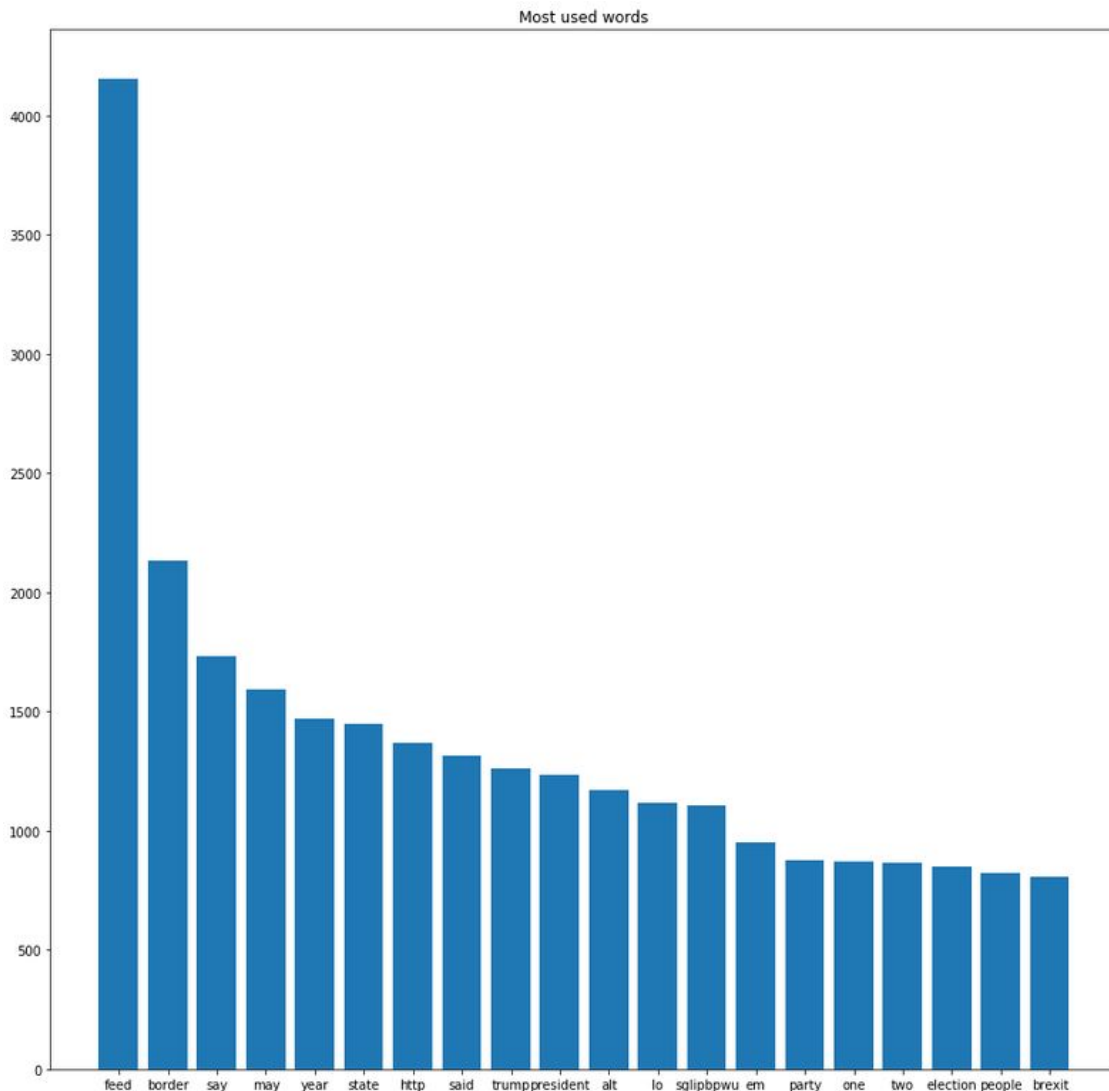
Testy naszej aplikacji zostały przeprowadzone przy każdym rozszerzeniu liczby połączeń z feedami RSS. Jako pierwsze testowaliśmy połączenie, czy uzyskujemy dane ze źródła. Wykonaliśmy również testy kontekstowe - sprawdziliśmy jakość przychodzących danych, czy odpowiadają one zdefiniowanym przez nas kategoriom, czy przychodzące dane mieszczą się w predefiniowanych przez nas tematykach. Sprawdziliśmy również dane pod kątem języka w którym zostały one opublikowane - skupiliśmy się tylko i wyłącznie na feedach, które postują w języku angielskim. Testy wydajnościowe naszej aplikacji zostały przez nas traktowane z należytą dokładnością, jednak ze względu na wolumen zagregowanych danych, nie mieliśmy w tej materii dużego pola do popisu.

Testowanie części analitycznej naszego projektu przebiegało inkrementacyjnie. Na początku testowaliśmy działanie różnych algorytmów używanych do analizy tekstu. Pierwsze próby przeprowadzaliśmy na kilku wybranych feedach celem ustalenia zasadności wykorzystania danego algorytmu. Jeżeli algorytm okazywał się przydatny i spełniał ustalone przez nas założenia analizy informacji zawartej w feedach, rozszerzyliśmy jego działanie na wszystkie feedy zagregowane w bazie danych.

7. Wyniki

Podczas analizy danych przechowywanych w bazie stworzyliśmy trzy niezależne playbooki analizujące dane na różne sposoby.

1. Analiza Statystyczna



Obraz 4. Wykres przedstawiający najczęściej występujące słowa w badanym przez nas zbiorze - z wyłączeniem 'stopwords' - słów przestankowych/nieznaczących np. 'the' 'as' 'on' 'an'.

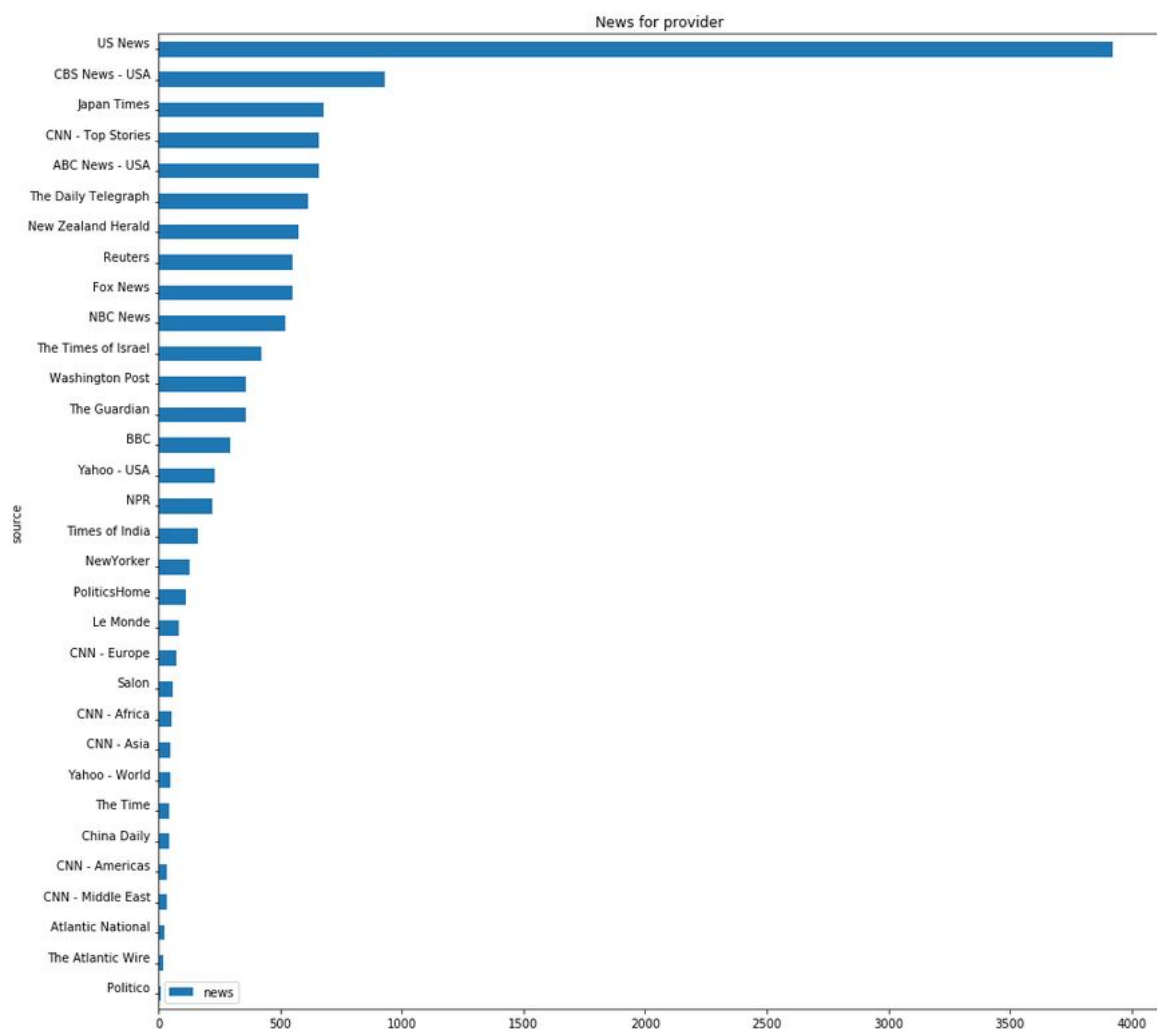
Pierwszy z wykresów które przedstawiamy na Obrazie 4 dotyczy najmniej zaawansowanej analizy tekstu - częstość występowania poszczególnych pojedynczych słów dla całego zbioru danych. W tym jak i każdym kolejnym przypadku usuwaliśmy z analizy słowa

przestankowe z języka angielskiego, tak aby nie zaburzać naszych analiz wyrazami jedno literowymi np. 'a', 'the', 'an'.

	source	news
0	US News\n	3924
1	CBS News - USA\n	928
2	Japan Times\n	676
3	CNN - Top Stories\n	661
4	ABC News - USA\n	657
5	The Daily Telegraph\n	616
6	New Zealand Herald\n	573
7	Reuters\n	552
8	Fox News\n	551
9	NBC News\n	519
10	The Times of Israel\n	420
11	Washington Post\n	358
12	The Guardian\n	356
13	BBC\n	292
14	Yahoo - USA\n	229
15	NPR\n	222
16	Times of India\n	161
17	New Yorker\n	128
18	PoliticsHome\n	114
19	Le Monde\n	82
20	CNN - Europe\n	75
21	Salon\n	58
22	CNN - Africa\n	52
23	CNN - Asia\n	50
24	Yahoo - World\n	48
25	The Time\n	45
26	China Daily\n	42
27	CNN - Americas\n	34
28	CNN - Middle East\n	33
29	Atlantic National\n	25
30	The Atlantic Wire\n	20
31	Politico\n	10

Obraz 5. Statystyka przedstawiająca ilość feedów per źródło.

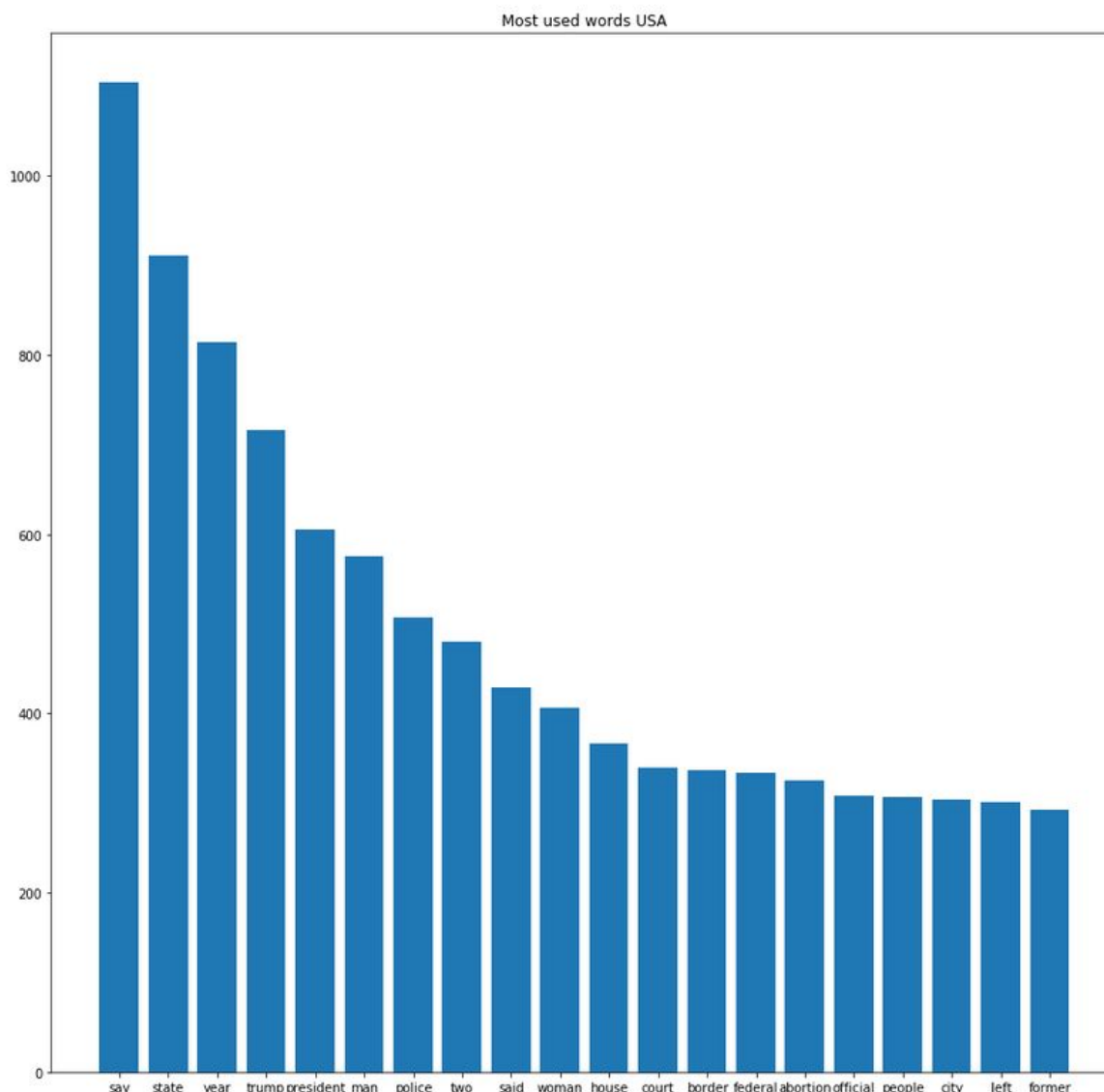
Na Obrazie 5 przedstawiamy ilość newsów z podziałem na nadawcę. Jak łatwo zauważyć, najbardziej często publikującym dziennikiem/źródłem jest US News. Jest on w przypadku naszego zbioru danych prawie 40 razy bardziej znaczący od serwisu Politico.



Obraz 6. Wykres przedstawiający rozkład ilościowy feedów z podziałem na źródło.

Na Obrazie 6 przedstawiamy w sposób bliższy dla oka jak znaczącą różnicę prezentuje serwis US News, w porównaniu do pozostałych, mniej częstych portali.

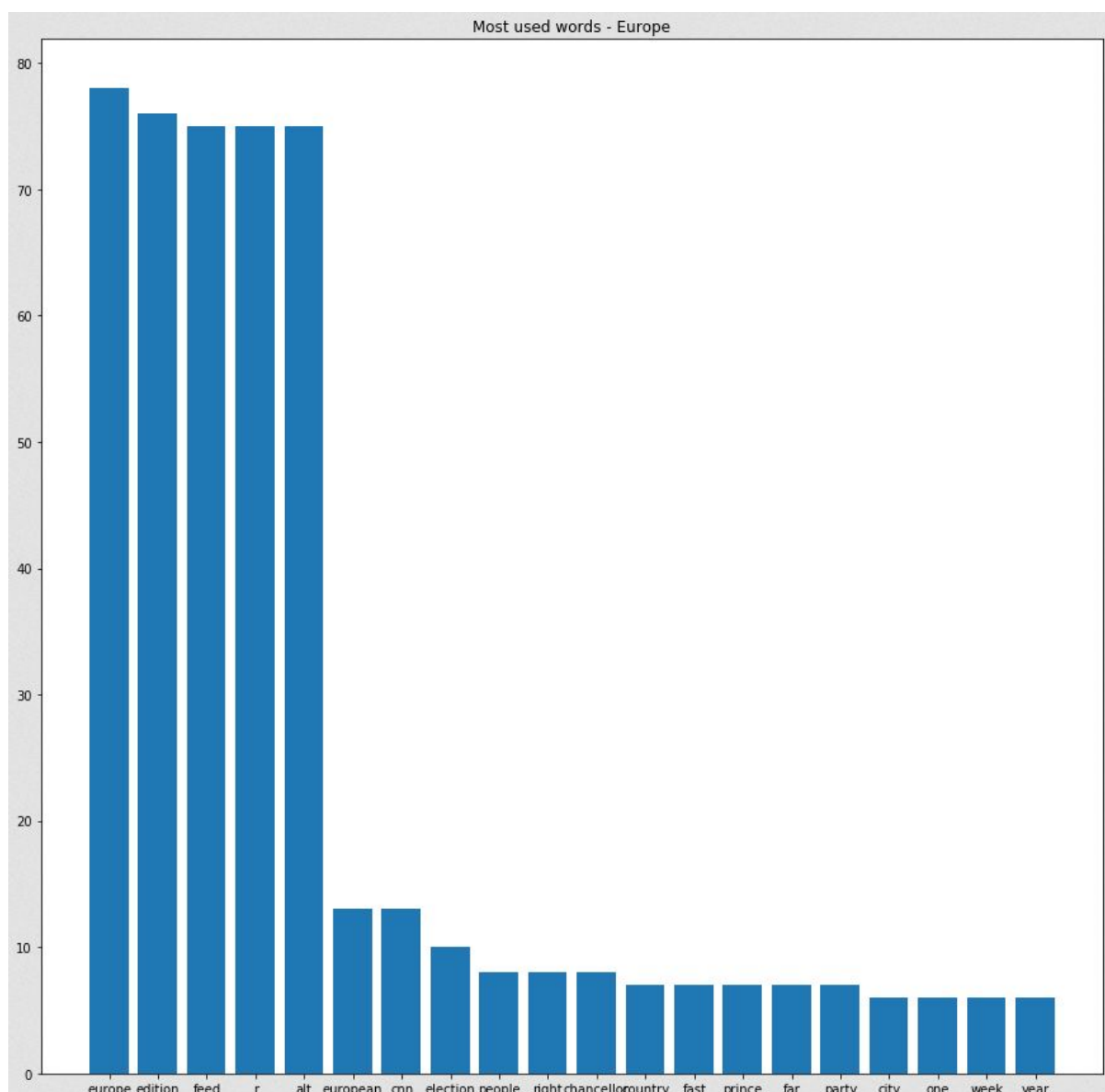
2. Analiza z podziałem na kontynenty



Obraz 8. Wykres przedstawiający najczęściej używane słowa w USA, pominięte zostały wyrazy ze stop listy.

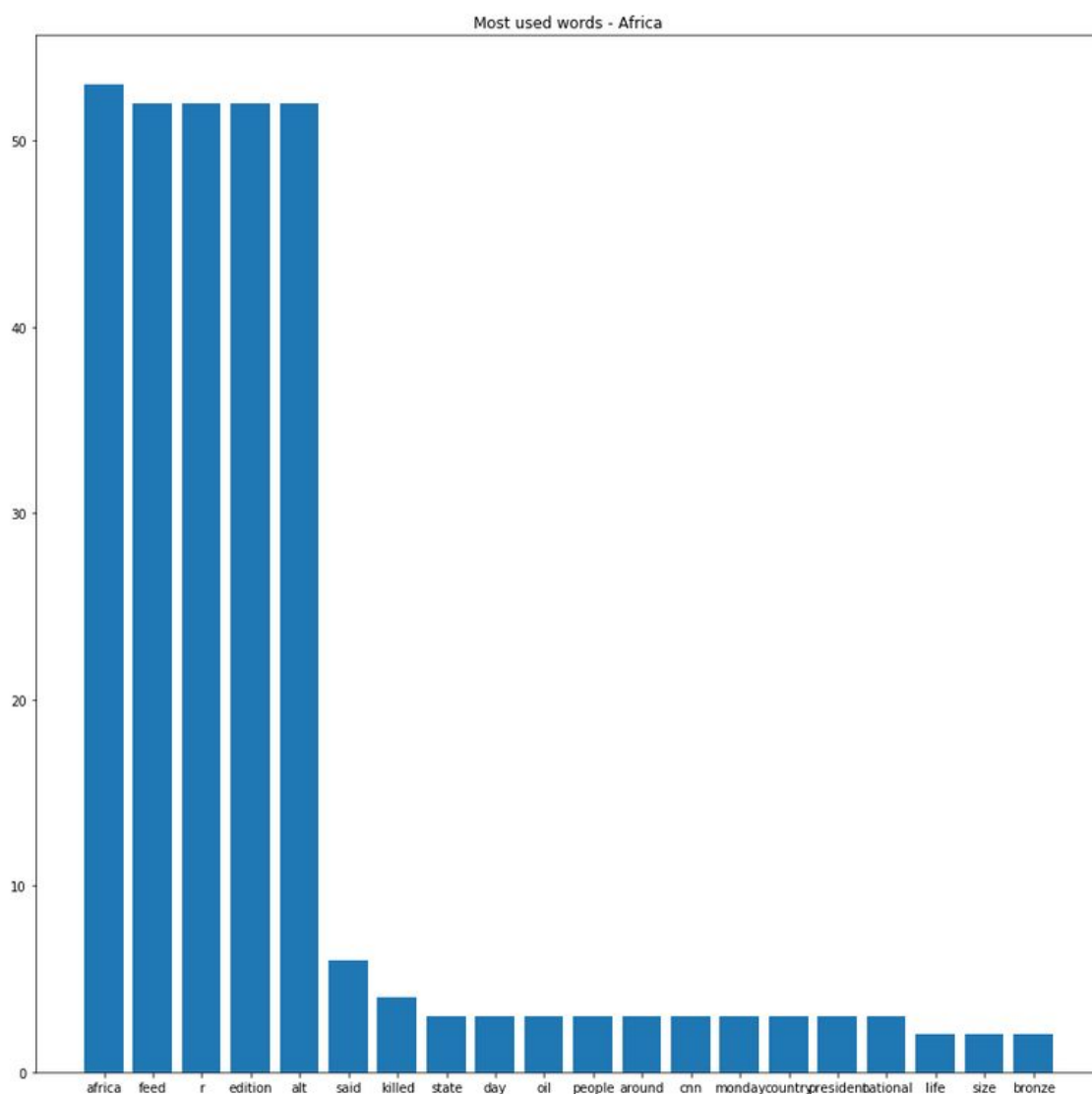
Obraz 8 rozpoczyna serię analiz kontynentalnych z naszego trzeciego workbooka. Zostały w nim zawarte częstości występowania słów kluczowych z podziałem na kontynent amerykański północnej. Przydziału odpowiednich źródeł dokonaliśmy na bazie

- siedziby głównej gazety/dziennika
- przynależności danego kanału np. w przypadku CNN dysponowaliśmy kilkoma różnymi źródłami, które publikowały newsy dotyczące danego kontynentu, np. CNN Europe. W takim przypadku, pomimo tego że centrala CNN znajduje się w USA, dany feed klasyfikowaliśmy do Europy



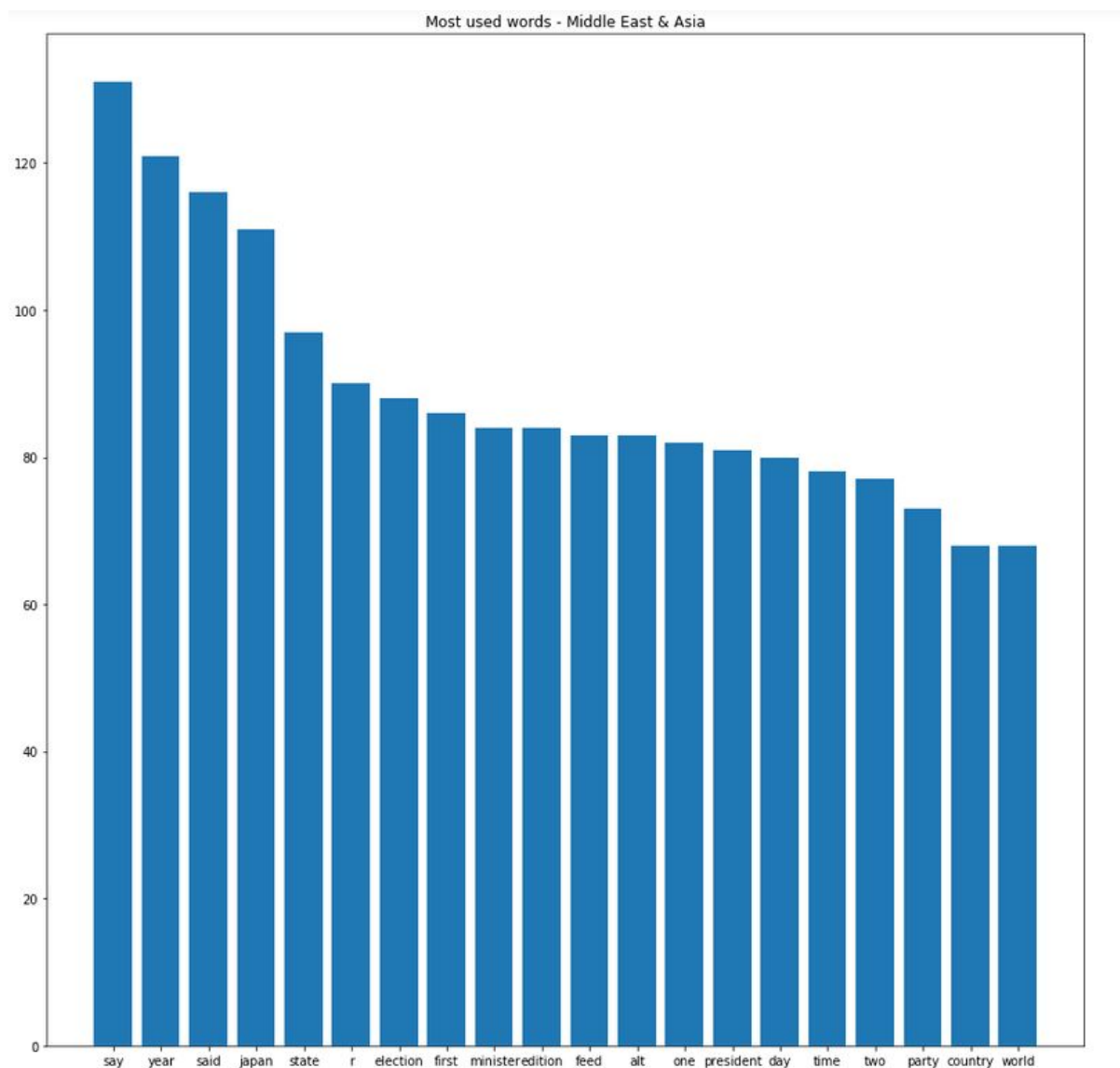
Obraz 9. Wykres przedstawiający najczęściej używane słowa w kontynentalnej europie, pominięte zostały wyrazy ze stop listy.

Obraz 9. przedstawia analogiczną sytuację jak wykres dotyczący USA, w przypadku Europy skupiliśmy się głównie na dziennikach brytyjskich, ponieważ znalezienie żywego i sensownego feeda z innego kraju, który obejmowałby tematykę europejską, oraz był pisany po angielsku, okazało się nietrywialne.



Obraz 10. Wykres przedstawiający najczęściej używane słowa w Afryce, pominięte zostały wyrazy ze stop listy.

Obraz 10 przedstawia analizę częstości występowania słów w Afryce. W przypadku Afryki mieliśmy do dyspozycji jedynie oddziały ogólnowiadomościowych gazet, np. CNN - Africa.

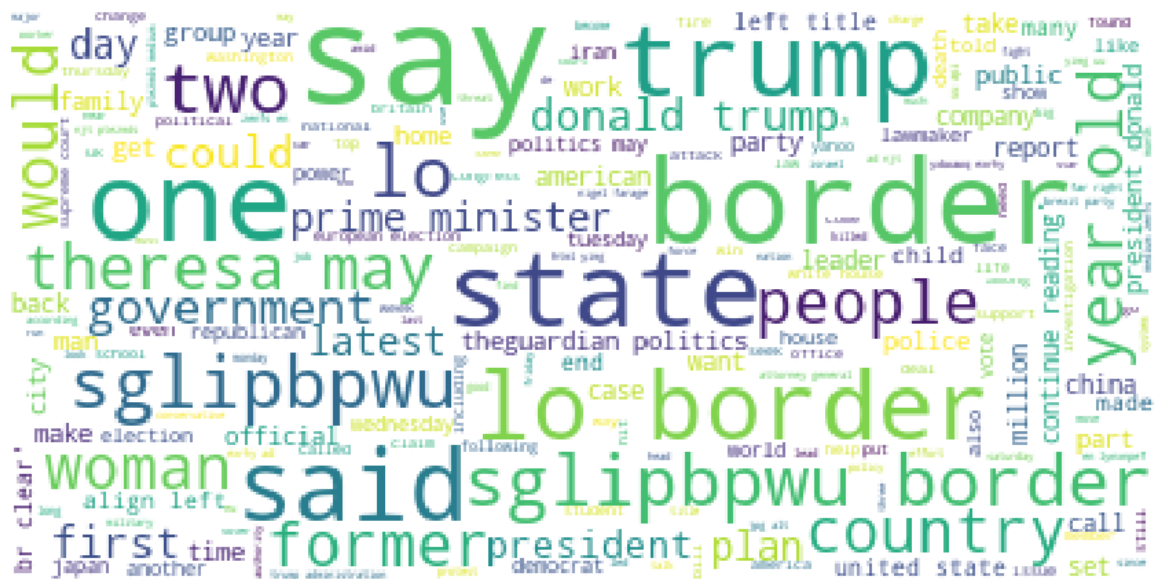


Obraz 11. Wykres przedstawiający najczęściej używane słowa na bliskim wschodzie i w Azji (Japonia, Chiny), pominięte zostały wyrazy ze stop listy.

Obraz 11. przedstawia analizę częstości występowania słów na kontynencie Azjatyckim oraz w rejonie bliskiego wschodu. W przypadku tych dwóch rejonów, opieraliśmy się zarówno na oddziałach gazet ogólnowiadowych, jak i na natywnych dziennikach np. Japan News.

3. Ogólna analiza słów kluczowych

Word Cloud



Obraz 12. Word cloud dla newsów z USA.

Obraz 12 prezentuje najczęściej używane słownictwo dla całego naszego zbioru danych przedstawiony w formie tzw. word-clouda. Uwagę przykuwają największe, najczęściej używane zlepki słów takie jak: Trump, Theresa MAY, state, people, border czy country.

Stopień słów:

```
{'labour': 7994,  
  'pledge': 557,  
  'minimum': 1933,  
  'wage': 1376,  
  'jeremy': 3571,  
  'corbyn': 4128,  
  'unveils': 1376,  
  'plan': 3571,  
  'paid': 2195,  
  'hour': 2195,  
  'rather': 1376,  
  'current': 1376,  
  'brexit': 17593,  
  'gavin': 3833,  
  'williamson': 3833,  
  'attack': 557,  
  'theresa': 4980,  
  'may': 9108,  
  'talk': 1671,  
  'recently': 557,  
  'sacked': 2195,  
  'defence': 4390,  
  'secretary': 6618,  
  'say': 15759,  
  'negotiation': 557,  
  'deal': 3309,  
  'fail': 1671,  
  'could': 557,  
  'set': 1114,  
  'exit': 557,  
  'date': 1114,  
  'week': 1933,  
  'sir': 2752,  
  'graham': 1114,  
  'brady': 1114,  
  'senior': 1933,  
  'tory': 3866,  
  'expects': 557,  
  'learn': 557,  
  'departure': 1114,
```

Obraz 13. Stopień słów dla newsów z Wielkiej Brytanii.

W przypadku Obrazu 13 prezentujemy stopień słów dla newsów pochodzących jedynie z wielkiej brytanii. Widać tutaj kilka najbardziej wyróżniających się słów, np 'brexit', 'corbyn', 'labour', ma to związek z okresem w jakim zbieraliśmy dane, który był mocno skorelowany z zawirowaniami politycznymi dotyczącymi brexitu w Wielkiej Brytanii

8. Wnioski

Po wykonaniu analiz nasuwają się nam następujące wnioski:

- większy nakład pracy jest wymagany przy tworzeniu środowiska do zbierania danych, tak aby były one poprawne, gotowe pod analizę, czyste, niż do samej analizy
- Feedy dziennikarskie mogą być z łatwością przypisane do jednej z trzech grup - aktywne (ponad 50 feedów dziennie), średnioaktywne - od 10 do 50 feedów dziennie, oraz mało aktywne - poniżej 10 feedów dziennie, lub nieregularnie
- analiza tekstów pochodzących z RSS feedów ma swoje bardzo ograniczone zastosowanie, nie można wykonać na tym zbiorze analizy sentymentu - ponieważ dane te są z natury pisane bezosobowo, bez emocji
- analiza feedów względem kontynentu pokazuje, że nie ma znacznej różnicy w częstości najbardziej używanych słów - jest to bardziej podyktowane językiem oraz formą sformułowanych wiadomości
- wybrany przez nas stack technologiczny jest odpowiedni do problemu, umożliwia on szybkie, proste, efektywne analizy tekstu, czyszczenie danych oraz tworzenie środowiska do integracji wielu serwisów RSS

9. Bibliografia

- <https://towardsdatascience.com/extracting-keywords-from-short-text-fce39157166b?fbclid=IwAR1X6HeBUhTT34Kx7b2sygW4mdgPgMYaAVxwB08bge2FzBj0lME5ieubkRQ>
- <https://stackoverflow.com/questions/29903025/count-most-frequent-100-words-from-sentences-in-dataframe-pandas>
- https://medium.com/analytics-vidhya/automated-keyword-extraction-from-articles-using-nlp-bfd864f41b34?fbclid=IwAR3TaWJn0Giv9_eBQFZacmhxDUQOH9susqsswTI8P4w5DyrMju2dVESR78s
- <https://www.nltk.org/>
- <https://pypi.org/project/wordcloud/>
- <https://docs.python.org/2/library/collections.html>
- <https://pypi.org/project/rake-nltk/>
- https://matplotlib.org/users/pyplot_tutorial.html

10. * Lista feedów RSS - plik rsslist

BBC

<http://feeds.bbc.co.uk/news/politics/rss.xml>

The Daily Telegraph

<https://www.telegraph.co.uk/news/rss.xml>

The Guardian

<https://www.theguardian.com/politics/rss>

Washington Post

<http://feeds.washingtonpost.com/rss/politics>

Japan Times

<https://www.japantimes.co.jp/feed/topstories/>

The Times of Israel

<https://www.timesofisrael.com/feed/>

Times of India

<https://timesofindia.indiatimes.com/rssfeedstopstories.cms>

New Zealand Herald

http://rss.nzherald.co.nz/rss/xml/nzhtsrsscid_000000698.xml

China Daily

http://www.chinadaily.com.cn/rss/world_rss.xml

CNN - Top Stories

<http://rss.cnn.com/rss/edition.rss>

CNN - Americas

http://rss.cnn.com/rss/edition_americas.rss

CNN - Asia

http://rss.cnn.com/rss/edition_asia.rss

CNN - Europe

http://rss.cnn.com/rss/edition_europe.rss

CNN - Middle East

http://rss.cnn.com/rss/edition_meast.rss

CNN - Africa

http://rss.cnn.com/rss/edition_africa.rss

Politico

<http://www.politico.com/rss/politics08.xml>

Real Clear Politics

<https://www.realclearpolitics.com/rss/>

PoliticsHome

<https://www.politicshome.com/news/rss>

ABC News - USA

<http://feeds.abcnews.com/abcnews/usheadlines>

CBS News - USA

<http://www.cbsnews.com/latest/rss/main>

NY Times

<http://www.nytimes.com/services/xml/rss/nyt/National.xml>
Wall Street Journal
http://online.wsj.com/xml/rss/3_7085.xml
USA Today
<http://content.usatoday.com/marketing/rss/rsstrans.aspx?feedId=news2>
NBC News
<http://feeds.nbcnews.com/feeds/topstories>
Reuters
<http://feeds.reuters.com/Reuters/worldNews>
Yahoo - USA
<http://news.yahoo.com/rss/us>
Yahoo - World
<http://rss.news.yahoo.com/rss/world>
Newsweek
<http://www.newsweek.com/rss>
NewYorker
<http://www.newyorker.com/feed/news>
US News
<http://www.usnews.com/rss/news>
NPR
<http://www.npr.org/rss/rss.php?id=1003>
Atlantic National
<http://feeds.feedburner.com/AtlanticNational>
The Atlantic Wire
<http://feeds.feedburner.com/TheAtlanticWire>
LA Times - USA
<http://www.latimes.com/nation/rss2.0.xml>
La Times - World
<http://www.latimes.com/world/rss2.0.xml>
Vice
<https://news.vice.com/>
Salon
<http://www.salon.com/category/news/feed/rss/>
The Time
<http://time.com/newsfeed/feed/>
Fox News
<http://feeds.foxnews.com/foxnews/latest?format=xml>