

Katarzyna Jarzyńska

Uniwersytet Warszawski

 <https://orcid.org/0000-0001-8181-963X>

BADANIE ANATOMII *FAKE NEWSA* JAKOŚCIOWE METODY ANALIZY DEZINFORMACJI ZAMIESZCZANEJ W INTERNECIE

***Fake newsy* – definicja i rodzaje**

Przekazywanie informacji jest nieodłącznym elementem egzystencji gatunku ludzkiego od początku jego istnienia. Komunikacja międzyludzka przeszła długą, wielowiekową drogę – od prostych znaków przez mowę aż do pisma. Wynalazek druku wprowadził nową jakość w dzieleniu się wiedzą i wiadomościami między ludźmi. Pojawienie się mediów tradycyjnych – prasy, radia i telewizji – było kolejnym przełomem na drodze do nowego typu komunikacji, o charakterze masowym. Największą rewolucję przyniosło pojawienie się internetu, który umożliwia przekazywanie informacji na niespotykaną wcześniej skalę z ogromną prędkością. Obecnie stanowi on podstawowy kanał porozumiewania się z innymi, zdobywania wiedzy czy poszukiwania rozrywki. Obok pozytywnych aspektów usieciowienia się społeczeństwa pojawiło się również wiele negatywnych, wśród których w ostatnich latach główną pozycję zajmuje zjawisko przekazywania fałszywych wiadomości, powszechnie zwanych *fake newsami*¹.

Fake newsy (FN) to nieprawdziwe informacje, przekazywane świadomie lub nieświadomie przez media różnego typu. FN bywają tworzone po to, by wyrządzić szkodę, np. żeby oszukać, zdyskredytować przeciwnika czy osiągnąć korzyści materialne bądź wizerunkowe².

¹ K. Rosińska, *Fake news. Geneza, istota, przeciwdziałanie*, Warszawa 2021, s. 7–14.

² K. Bąkowicz, *Fake news. Produkt medialny czasów postprawdy*, Warszawa 2020, s. 75–77.

Można wyodrębnić kilka typów *fake newsów*. Do najbardziej powszechnych należą typowe błędne informacje. Zamieszczane są nieintencjonalnie na skutek błędnego połączenia faktów. Zwykle nie stanowią zagrożenia i łatwo je zidentyfikować. Świadoma dezinformacja za pomocą FN dokonywana jest z intencją wyrządzenia szkody lub osiągnięcia zysku. Do tego celu wykorzystywane są zabiegi takie jak: manipulacja treścią lub jej fabrykowanie, zamieszczanie fałszywej treści oraz umieszczanie informacji w fałszywym kontekście. Nieprawdziwe wiadomości mogą zawierać szczegółowe opisy wydarzeń, cytaty, odniesienia do uczestników przytaczanych wydarzeń, a także przedstawiać bliższy i szerszy kontekst. Odpowiednie przetworzenie treści FN sprawia, iż upodabniają się one do treści opartych na faktach. Również materiały propagandowe, spreparowane i odbiegające od rzeczywistości, zaliczane są do grupy FN. Należy jednak zwrócić uwagę na wyjątki, które mogą mieć charakter FN, ale nimi nie są. Zastosowanie satyry, ironii i parodii upodabnia wiadomość do FN, ale jej nim nie czyni³.

W porównaniu do prawdziwej informacji tempo rozchodzenia się fałszywej wiadomości jest szybsze i dociera ona do większego grona odbiorców. Związane jest to z dużym nacechowaniem emocjonalnym FN, które dzięki temu są chętniej przekazywane. Oprócz mediów społecznościowych FN bywają także udostępniane za pośrednictwem mediów tradycyjnych. Przekazanie fałszywej informacji w prasie czy telewizji sprawia, że staje się ona bardziej wiarygodna⁴.

Ze względu na liczne zagrożenia wynikające z rozprzestrzeniania się szkodliwych, fałszywych wiadomości zasadne są próby opracowania narzędzi do wychwytywania tego typu informacji. Przeciwdziałanie rozpowszechnianiu się FN stanowi problem badawczy wielu polskich i zagranicznych instytucji naukowych. Tworzy się algorytmy rozpoznające i przewidujące fałszywe wiadomości. Ponadto dla użytkowników internetu przeprowadzane są akcje edukacyjne o zagrożeniach wynikających z braku weryfikacji FN. Jednym ze sposobów reagowania na *fake newsy* jest *fact-checking*, który polega na rzetelnym i bieżącym sprawdzaniu treści zamieszczanych w sieci i ich opiniowaniu przez ekspertów lub wolontariuszy⁵.

³ Ibidem, s. 77–78.

⁴ Ibidem.

⁵ K. Rosińska, op. cit., s. 173–204.

Narzędzia do analizy *fake newsów*

Pomocne przy identyfikowaniu FN jest poznanie ich cech głównych i opracowanie swego rodzaju opisu anatomii treści o charakterze fałszywym i dezinformacyjnym. Jedną z technik identyfikacji FN to analiza języka używanego w treści informacji spreparowanych i zmanipulowanych. Tego typu badania możliwe są do przeprowadzenia dzięki infrastrukturze polskiego konsorcjum naukowego CLARIN-PL. Stanowi ono część europejskiej infrastruktury badawczej CLARIN (Common Language Resources and Technology Infrastructure). W sześciu jednostkach naukowych i badawczych powstają elektroniczne zasoby językowe i środki techniczne do analizy dużych zbiorów tekstów w języku polskim. Narzędzia te służą głównie humanistom oraz przedstawicielom nauk społecznych do badań z wykorzystaniem najnowszych rozwiązań cyfrowych. Oprogramowanie i aparatura stworzone przez CLARIN są darmowe. Instytucja ta stara się upowszechniać wśród naukowców rozwiązania z zakresu inżynierii języka przez prowadzenie warsztatów, możliwość konsultacji, otwartość na współpracę z różnymi ośrodkami naukowo-badawczymi oraz redakcję materiałów szkoleniowych. Centrum Technologii Językowych CLARIN-PL funkcjonuje przy Politechnice Wrocławskiej, a nadzór nad nim sprawuje oraz zapewnia utrzymanie Ministerstwo Edukacji i Nauki Rzeczypospolitej Polskiej⁶.

Narzędzia do analizy języka polskiego udostępniane przez CLARIN-PL umożliwiły dokonanie analizy wybranych *fake newsów* w celu wyodrębnienia ich cech szczególnych i zbadania struktury fałszywych wiadomości. Pierwszym z zastosowanych narzędzi był Cat, służący do klasyfikacji tekstów na podstawie jednego ze wskazanych kryteriów: klasyfikacji tematycznej według modelu wyuczonego na pięciu kategoriach grupowania Wikipedii, klasyfikacji tematycznej według modelu wyuczonego na tematyce prasowej, klasyfikacji według podobieństwa stylu gramatycznego do stylu jednego ze znanych polskich pisarzy XIX i XX wieku, wykrywania udziału języka w całym korpusie⁷. Analizę można przeprowadzić w odniesieniu do dowolnego korpusu tekstowego w formacie .zip. Zastosowanie Cat pomaga w ustaleniu tematyki FN lub innego dowolnego tekstu bez konieczności czytania go.

W badaniu wykorzystano również narzędzie do spersonalizowanego rozpoznawania mowy nienawiści – Hatespeech. Określa ono udział procentowy słów o nacechowaniu negatywnym w badanym korpusie tekstów. Dzięki temu

⁶ O nas, CLARIN-PL, <https://clarin-pl.eu/index.php/o-nas/> (dostęp: 6.07.2022).

⁷ Cat, CLARIN-PL, <https://clarin-pl.eu/index.php/cat/> (dostęp: 6.07.2022).

możliwe jest prognozowanie poziomu emocji wywołanych przez daną informację po jej odczytaniu⁸.

Podobne rezultaty daje użycie narzędzia Sentemo, które służy do określania rozkładu wydźwięku emocjonalnego badanego dokumentu. Narzędzie to pomaga wskazać pozytywny lub negatywny aspekt wypowiedzi. Przetwarzanie i wyznaczanie sentymentu danych tekstowych pozwala określić rozkład polaryzacji emocjonalnego nacechowania publikacji. W sposób procentowy określony zostaje udział treści o pozytywnej i negatywnej polaryzacji, bez wydźwięku emocjonalnego oraz niejednoznaczne pod względem nacechowania emocjonalnego. Udział procentowy danych typów polaryzacji może być wspólny dla wielu FN, co stanowi pomoc we wczesnym wykrywaniu nieprawdziwych informacji⁹.

Możliwość dodatkowego wglądu w nacechowanie emocjonalne tekstu daje narzędzie Wydźwięk, za którego pomocą analizuje się wydźwięk emocjonalny FN. Przetwarza ono i analizuje dane tekstowe w celu oznaczenia ich nacechowania emocjonalnego, wykorzystując „podstawowe typy emocji oraz wartości uniwersalne przypisane do słów”¹⁰. Wynik w postaci wykresów określa liczbę wystąpień konkretnych emocji pozytywnych i negatywnych w analizowanym korpusie tekstów.

Określenie stopnia zrozumiałości tekstu jest wspomagane przez narzędzie Mgiistość tekstu, dostępne w ramach aplikacji LEM¹¹. Opiera się ono na indeksie czytelności FOG, który służy określeniu stopnia przystępności danej treści. Parametr ten wskazuje, ile lat kształcenia jest potrzebnych do prawidłowego odczytania tekstu. Wzór na wartość FOG sformułowano następująco¹²:

$$\text{FOG} = 0,4 \left(\frac{\text{liczba słów}}{\text{liczba zdań}} + 100 \left(\frac{\text{liczba słów długich}}{\text{liczba słów}} \right) \right)$$

W poniższej tabeli znajduje się szczegółowe wyjaśnienie odpowiednich wartości wskaźnika FOG (tab. 1).

⁸ *Alfabetyczny spis usług*, CLARIN-PL, <https://clarin-pl.eu/index.php/alfabetyczny-spis-uslug/> (dostęp: 6.07.2022).

⁹ *Sentemo*, CLARIN-PL, <https://clarin-pl.eu/index.php/sentemo/> (dostęp: 6.07.2022).

¹⁰ *Wydźwięk*, CLARIN-PL, <https://clarin-pl.eu/index.php/wydzwiek/> (dostęp: 6.07.2022).

¹¹ *LEM*, CLARIN-PL, <https://clarin-pl.eu/index.php/lem/> (dostęp: 6.07.2022).

¹² *Indeks czytelności FOG*, Wikipedia, https://pl.wikipedia.org/wiki/Indeks_czytelno%C5%9Bci_FOG (dostęp: 6.07.2022).

Tab. 1. Interpretacja wyniku testu FOG

Wartość FOG	Interpretacja
1–6	język bardzo prosty, rozumiały już dla uczniów szkoły podstawowej
7–9	język prosty, rozumiały już dla uczniów gimnazjum
10–12	język dość prosty, rozumiały już dla uczniów liceum
13–15	język dość trudny, rozumiały dla studentów studiów licencjackich
16–17	język trudny, rozumiały dla studentów studiów magisterskich
18 i więcej	język bardzo trudny, rozumiały dla magistrów i osób z wyższym wykształceniem

Źródło: J. Miodek et al., *Jak pisać o Funduszach Europejskich?*, Warszawa 2010, s. 21–24

TF-IDF (ang. TF – term frequency, IDF – inverse document frequency) to metoda służąca do określania wagi słów na podstawie częstotliwości ich występowania. Reprezentuje ona grupę algorytmów, które obliczają statystyczne wagi wyrazów. Każdy tekst reprezentowany jest przez wektor, na który składają się wagi słów pojawiających się w określonym dokumencie. TF-IDF informuje o częstotliwości pojawiania się terminów, wzięwszy pod uwagę odpowiednie zbalansowanie znaczenia lokalnego terminu i jego sensu w odniesieniu do całego zbioru dokumentów. Narzędzia TF-IDF używa się do oceny relewantności dokumentu lub podobieństw tekstów w systemach grupowania wyników oraz systemach antyplagiatowych¹³. Zastosowanie analizy TF-IDF do badania *fake newsów* umożliwia odtworzenie ich tematyki na podstawie wyznaczonych za pomocą tej metody najważniejszych słów w badanych tekstach.

Ostatnie z narzędzi zastosowanych w badaniu to automatyczne ekstraktory słów kluczowych. W ramach opracowania skupiono się na YAKE! oraz RAKE, które są w stanie wyznaczać terminy kluczowe dla języka polskiego. YAKE! (Yet Another Keyword Extractor!)¹⁴ to prosta, nienadzorowana metoda automatycznego wyznaczania słów kluczowych, która opiera się na cechach statystycznych tekstu, wyodrębnionych z pojedynczych dokumentów. Algorytm nie musi być szkolony na konkretnym zestawie dokumentów, nie jest też zależny od słowników, korpusu zewnętrznego, rozmiaru tekstu ani języka¹⁵. RAKE (Rapid Automatic Keyword Extraction) to niezależny od domeny algorytm

¹³ TF-IDF, CLARIN-PL, <https://clarin-pl.eu/index.php/tf-idf/> (dostęp: 6.07.2022).

¹⁴ YAKE! Keyword Extractor dostępny jest na stronie: <http://yake.inesctec.pt/> (dostęp: 6.07.2022).

¹⁵ R. Campos et al., *YAKE! Keyword extraction from single documents using multiple local features*, „Information Sciences” 2020, Vol. 509, s. 257–289.

wyodrębniania słów kluczowych, który próbuje określić kluczowe frazy w tekście przez analizę częstotliwości pojawiania się słów i ich współwystępowania z innymi słowami w tekście¹⁶.

Anatomia fake newsów

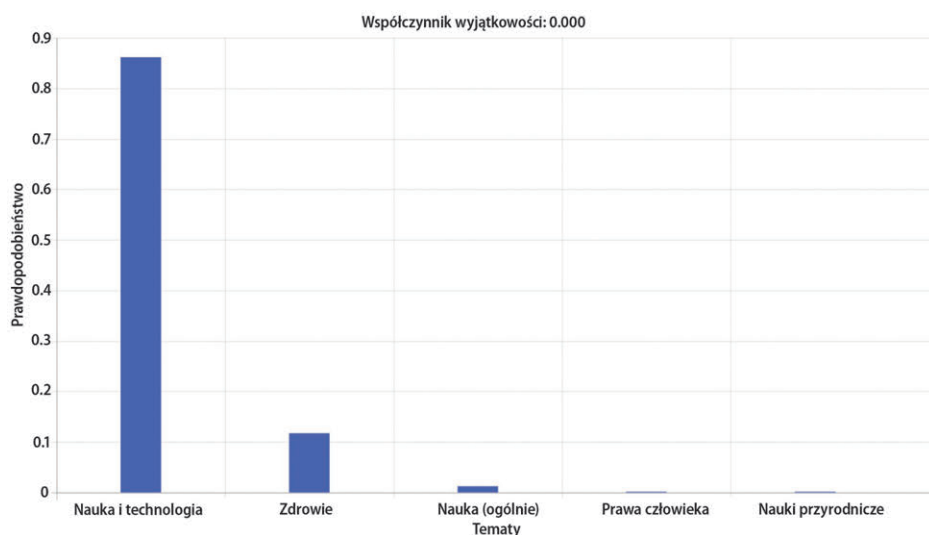
Analizie za pomocą omówionych wyżej narzędziami poddano trzy wyselekcjonowane *fake newsy*, pochodzące z różnych źródeł, licznie udostępniane m.in. w serwisach Facebook oraz Twitter. Każdy z FN dotyczył innej tematyki – pandemii COVID-19, obecnej sytuacji politycznej oraz ochrony środowiska.

Na blogu Lega Artis w kwietniu 2022 r. zamieszczono informację, iż grafen przenoszony jest od osób zaszczepionych do nieszczepionych, co ma stanowić zagrożenie dla zdrowia i życia tej drugiej grupy¹⁷. Postacią głównie odpowiedzialną za to odkrycie jest dr Philippe van Welbergen. Artykuł opatrzono licznymi cytatami z wypowiedzi doktora, fachowymi ilustracjami przedstawiającymi wnikanie grafenu do krwi ludzkiej oraz odniesieniami do innych stron uwiarygodniających treść zamieszczonego artykułu. Doniesienia doktora nie zostały jednak opublikowane w żadnym recenzowanym czasopiśmie naukowym, nie ma także dowodów, że badania zostały przez niego przeprowadzone w sterylnych warunkach, nie można więc uważać ich wyników za naukowe i rzetelne.

Analiza treści tego FN za pomocą narzędzia Cat wykazała, iż głównymi tematami poruszonymi w artykule Lega Artis na temat szkodliwości grafenu są te związane z nauką i technologią oraz zdrowiem (il. 1). Badanie TF-IDF wskazało na najczęściej występujące w tekście słowa: doktor, krew, Philippe, grafen, zaszczepić. Ekstrakcja słów kluczowych potwierdziła te wyniki w postaci słów: Lega Artis, doktor Philippe i krew.

¹⁶ K. Agarwal, *RAKE: Rapid Automatic Keyword Extraction Algorithm*, Medium, 8.04.2020, <https://medium.datadriveninvestor.com/rake-rapid-automatic-keyword-extraction-algorithm-f4ec17b2886c> (dostęp: 6.07.2022).

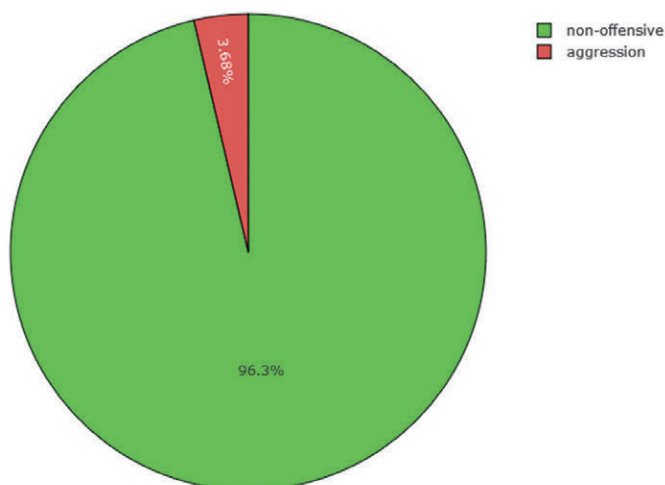
¹⁷ D. Głogowski, *Grafen jest przenoszony od zaszczepionych do nieszczepionych*, Kancelaria Lega Artis, 25.04.2022, <https://legaartis.pl/blog/2022/04/25/grafen-jest-przenoszony-od-zaszczepionych-do-nieszczepionych/> (dostęp: 6.07.2022).



Il. 1. Analiza tematyczna artykułu Lega Artis o grafenie

Źródło: opracowanie własne w Cat CLARIN

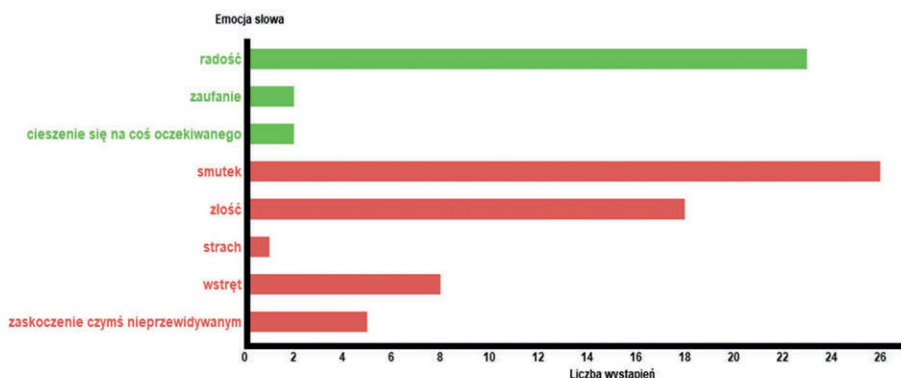
Badanie rozkładu obraźliwości (il. 2) wykazało niewielki stopień zawartości mowy nienawiści w artykule. Ma to wpływać na profesjonalny wygląd zawartych w nim informacji i zwiększać prawdopodobieństwo odebrania ich przez czytelników jako prawdziwe i zweryfikowane wyniki badań naukowca.



Il. 2. Rozkład obraźliwości artykułu Lega Artis o grafenie

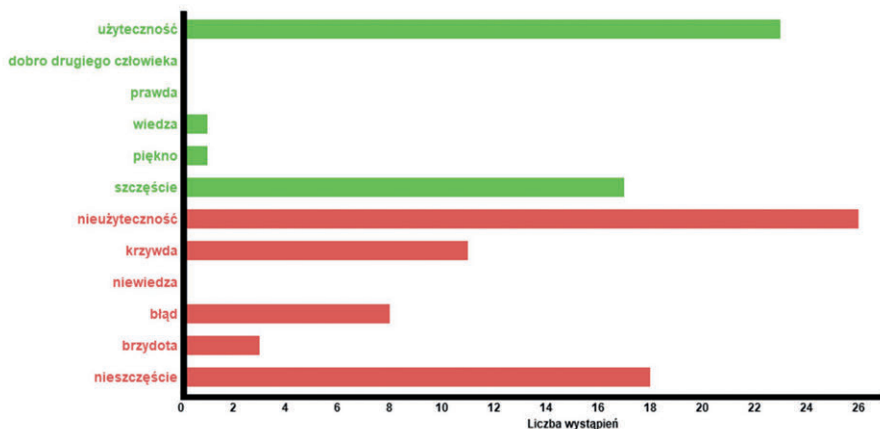
Źródło: opracowanie własne w Hatespeech CLARIN

Analiza wydźwięku emocjonalnego treści tego artykułu pomogła wykazać, iż jednoznacznie dominują w nim emocje negatywne, takie jak: smutek, złość czy wstręt (il. 3). Badanie dodatkowych cech wydźwięku ujawniło, że artykuł wykazuje emocje z zakresu nieużyteczności, krzywdy i nieszczęścia (il. 4). Wskazania narzędzia Wydźwięk zostały potwierdzone przez rozpoznanie tekstu za pomocą aplikacji Sentemo (il. 5). Łącznie ponad 16 procent treści tej informacji ma wydźwięk mniej lub bardziej negatywny, 66,7 procent nie ma nacechowania emocjonalnego, a jedynie 11,1 procent stanowią emocje pozytywne. Celem artykułu jest zatem wzbudzanie skrajnie złych emocji w jego odbiorcach.



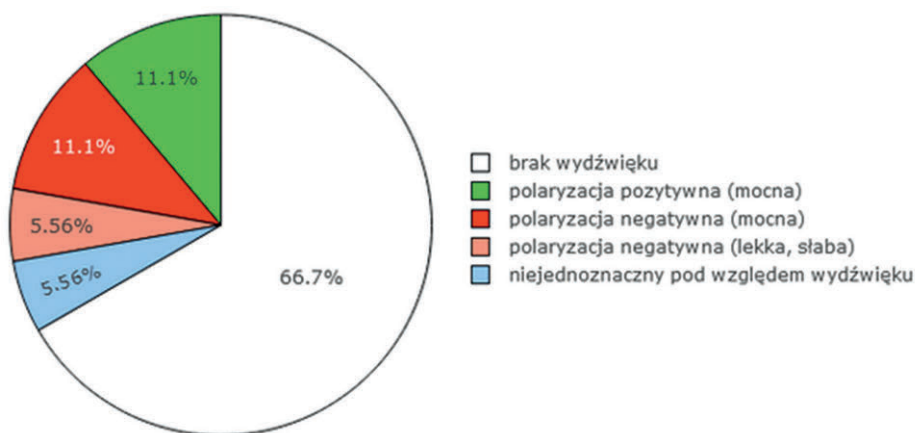
Il. 3. Analiza wydźwięku emocjonalnego artykułu Lega Artis o grafenie

Źródło: opracowanie własne w Wydźwięk CLARIN



Il. 4. Analiza wydźwięku emocjonalnego artykułu Lega Artis o grafenie

Źródło: opracowanie własne w Wydźwięk CLARIN



Il. 5. Polaryzacja wydźwięku emocjonalnego artykułu Lega Artis o grafenie

Źródło: opracowanie własne w Sentemo CLARIN

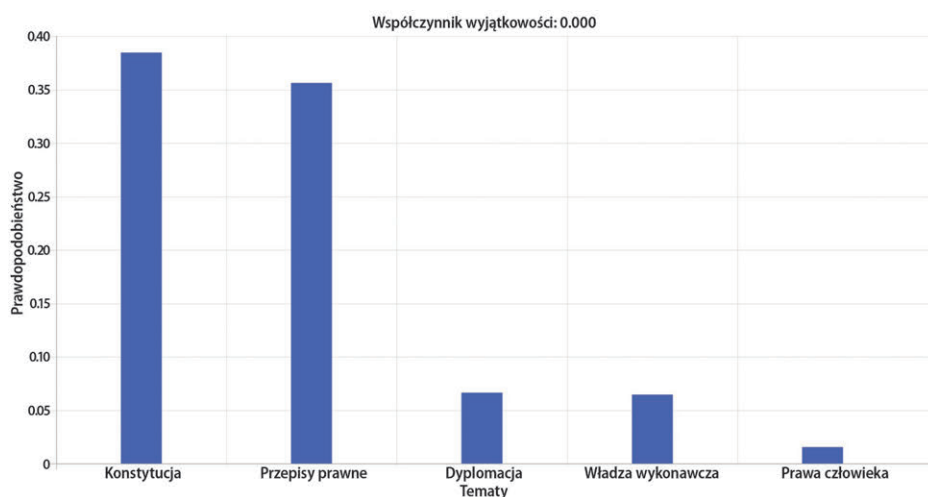
W wyniku analizy mglistości FN na temat szkodliwości grafenu dla zdrowia uzyskano współczynnik FOG na poziomie 10,77, co oznacza, iż tekst ten jest zrozumiały już dla uczniów liceum. Prosty język artykułu ma wspomagać jego szybsze przyswojenie przez odbiorców, zrozumienie, a na końcu usprawnić jego dalsze rozpowszechnianie wśród osób o podobnych poglądach.

Kolejnym FN poddanym analizie był wpis z bloga Wielka Biała Polska, którego autor donosi, że zgodnie z oświadczeniem ONZ państwa Polska nie ma i nigdy nie było ze względu na brak oficjalnie zatwierdzonej granicy¹⁸. W rzeczywistości jednak żadne z państw nie ma obowiązku tego robić, a Polska jest krajem oficjalnie uznanym w istniejących granicach *de iure* (według prawa)¹⁹.

Analiza tematyczna tego artykułu za pomocą narzędzia Cat wykazała, że dotyczy on spraw związanych z konstytucją, przepisami prawnymi, dyplomacją, władzą wykonawczą oraz prawami człowieka (il. 6). Badanie TF-IDF ujawniło, iż główne tematy poruszane we wpisie dotyczą granic Polski i Ukrainy, a także ONZ, umów i prawa. Słowa kluczowe wyekstrahowane z tego tekstu wskazały na poruszanie w nim spraw związanych z granicą Polski oraz bezpieczeństwem ONZ, a także prawem międzynarodowym.

¹⁸ Upadliniec, *Oświadczenie ONZ – państwa Polska nie ma i nigdy nie było!*, Wielka Biała Polska, 5.01.2016, <https://archive.ph/8U8vc#selection-228.0-228.1> (dostęp: 6.07.2022).

¹⁹ *Polska ma niezarejestrowane granice w ONZ? Powtarzalna dezinformacja!*, Demagog, 3.03.2022, https://demagog.org.pl/fake_news/polska-ma-niezarejestrowane-granice-w-onz-powtarzalna-dezinformacja/ (dostęp: 6.07.2022).



Il. 6. Analiza tematyczna wpisu z bloga Wielka Biała Polska

Źródło: opracowanie własne w Cat CLARIN

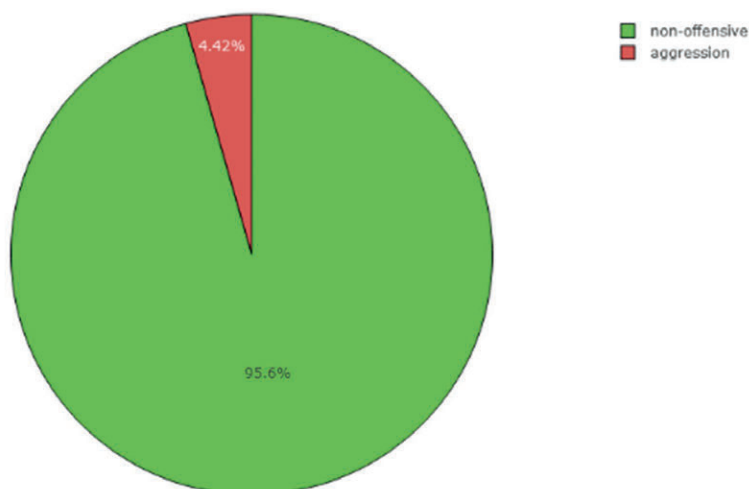
Badany tekst wykazuje nieco większy stopień zawartości mowy nienawiści w swojej treści od FN na temat grafenu, gdyż jest to wartość na poziomie blisko 4,5 procent (il. 7). Znalazło to potwierdzenie w analizie wydźwięku emocjonalnego, w której dominują emocje takie jak smutek, złość, strach (il. 8) oraz nieużyteczność, krzywda czy nieszczęście (il. 9). Narzędzie Sentemo wskazało jednoznacznie, iż 14,3 procent tekstu wykazuje nacechowanie negatywne, a pozostała jego treść nie ma żadnego wydźwięku (il. 10). Jest to wpis, który w odbiorcach ma wywoływać silne negatywne emocje, dzięki czemu zostanie lepiej zapamiętany. Dzięki zaangażowaniu emocjonalnemu czytelnicy szybciej zaczną się nim dzielić z innymi, by im uświadomić, jaka jest prawda.

Tekst napisany został językiem prostym, zrozumiałym już dla uczniów liceum, na co wskazuje wartość współczynnika FOG na poziomie 11,9. W połączeniu z wysokim stopniem zaangażowania emocjonalnego odbiorców ma to wpływać na szybsze przyswojenie przez nich treści tego FN, a także dalsze rozpowszechnianie go wśród osób o podobnych poglądach.

Trzecim FN poddanym analizie jest artykuł ze strony nczas.com, którego autor donosi, iż zmianę klimatu powoduje jedynie orbita słoneczna Ziemi, a nie działalność człowieka²⁰. Mimo udziału zmian orbity Ziemi w przeobrażeniach

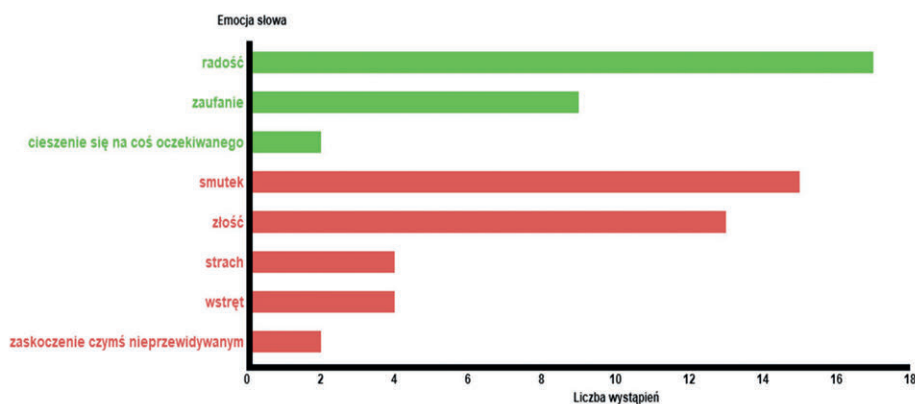
²⁰ KM, *Sensacyjny raport NASA! Zmianę klimatu powoduje orbita słoneczna Ziemi, a nie działalność człowieka*, nczas.com, 27.09.2019, [https://nczas.com/2019/09/27/sensacyjny-raport-nasa-zmiane-klimatu-powoduje-orbita-sloneczna-ziemi-a-nie-](https://nczas.com/2019/09/27/sensacyjny-raport-nasa-zmiane-klimatu-powoduje-orbita-sloneczna-ziemi-a-nie)

klimatycznych w głównej mierze na owe zmiany mają wpływ efekty działalności człowieka, a przede wszystkim zwiększona emisja gazów cieplarnianych²¹.



Il. 7. Rozkład obraźliwości wpisu z bloga Wielka Biała Polska

Źródło: opracowanie własne w Hatespeech CLARIN

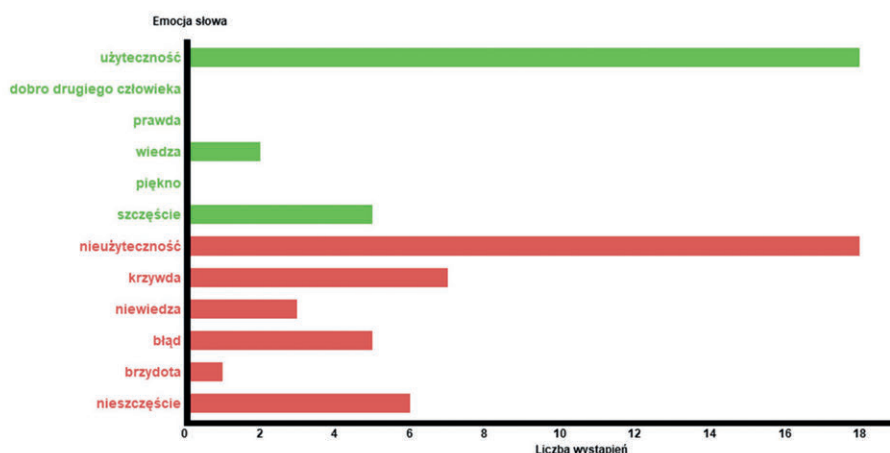


Il. 8. Analiza wydźwięku emocjonalnego wpisu z bloga Wielka Biała Polska

Źródło: opracowanie własne w Wydźwięk CLARIN

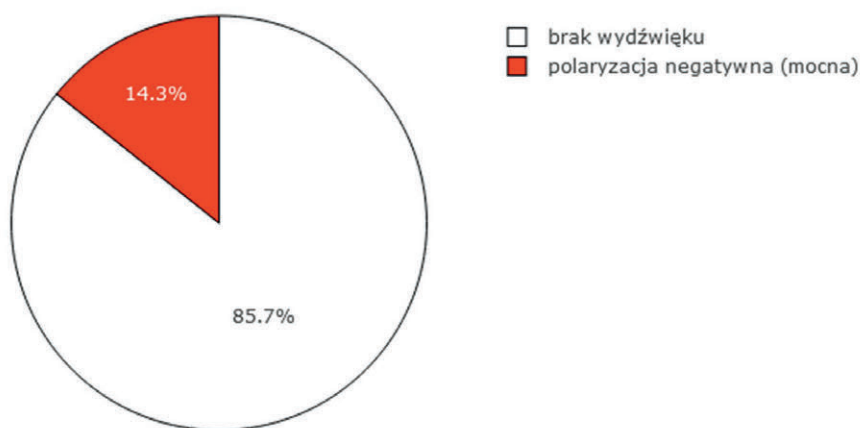
dzialalnosc-czlowieka/?fbclid=IwAR1nfGtoklU17tqs50ptzAK-fMZflao-nETXYJd9264xhteIvAQrDirsoY (dostęp: 6.07.2022).

²¹ Nie, zmiany orbity Ziemi nie są przyczyną obecnych zmian klimatu, Demagog, 18.10.2019, https://demagog.org.pl/fake_news/nie-zmiany-orbity-ziemi-nie-sa-przyczyna-obecnich-zmian-klimatu/ (dostęp: 6.07.2022).



Il. 9. Analiza wydźwięku emocjonalnego wpisu z bloga Wielka Biała Polska

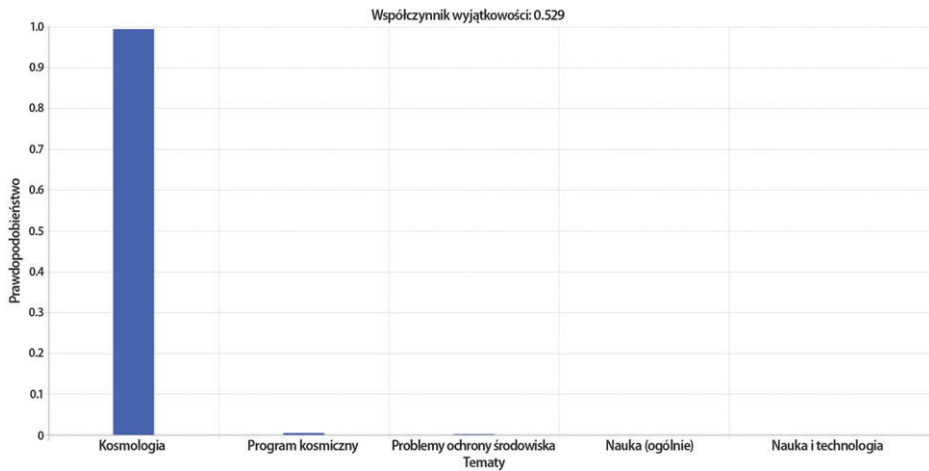
Źródło: opracowanie własne w Wydźwięk CLARIN



Il. 10. Polaryzacja wydźwięku emocjonalnego wpisu z bloga Wielka Biała Polska

Źródło: opracowanie własne w Sentemo CLARIN

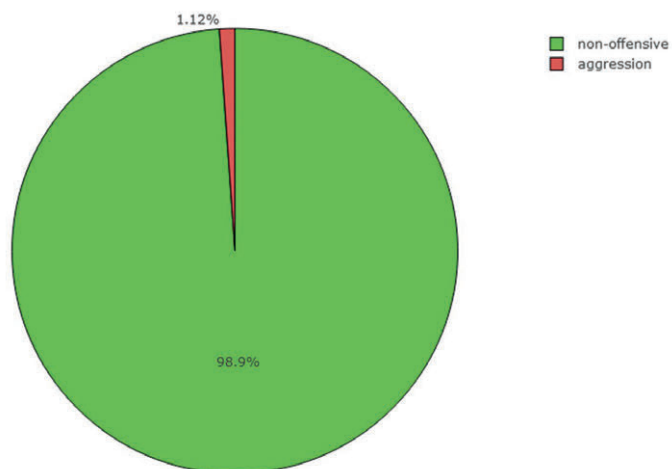
Analiza tematyczna treści artykułu wykazała, iż dotyczy on głównie kosmologii, programu kosmicznego oraz problemów ochrony środowiska (il. 11). Badanie metodą TF-IDF ujawniło, że najważniejszymi poruszonymi w nim zagadnieniami są te powiązane z orbitą, NASA, zmianą, klimatem, ziemią i człowiekiem. Ekstrakcja słów kluczowych wytypowała jako główne tematy tego FN działalność człowieka, agencję aeronautyczną, upał oraz przestrzeń kosmiczną.



Il. 11. Analiza tematyczna artykułu ze strony nczas.com

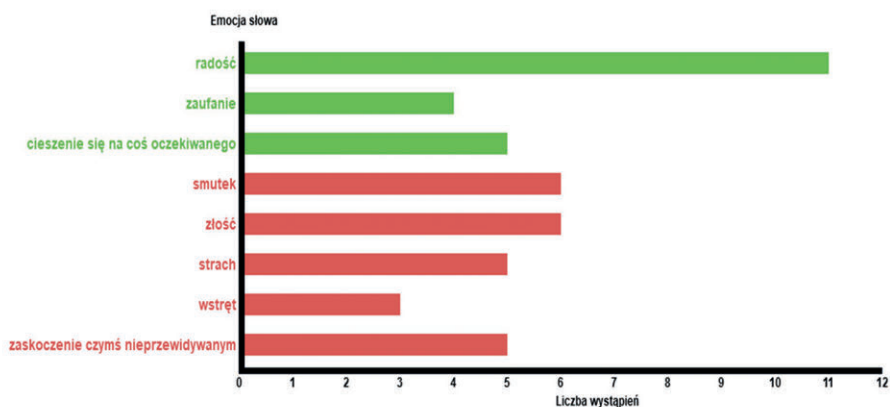
Źródło: opracowanie własne w Cat CLARIN

Badany tekst cechuje bardzo niski stopień zawartości mowy nienawiści, nieco ponad 1 procent (il. 12). Analiza wydźwięku emocjonalnego wykazała równy udział emocji negatywnych i pozytywnych w treści artykułu. Wśród emocji pozytywnych przeważają radość, zaufanie, użyteczność oraz szczęście (il. 13 i 14). Emocje negatywne to głównie strach, smutek, złość, krzywda, nieszczęście oraz nieużyteczność (il. 13 i 14). Dodatkowo narzędzie Sentemo wskazało, iż tekst nie jest spolaryzowany ani negatywnie, ani pozytywnie (il. 15). Taka konstrukcja artykułu może wskazywać, że głównym celem jego twórcy było uwiarygodnienie przekazywanych treści oraz nadanie im stylu pseudonaukowego. Potwierdzeniem tej tezy może być podwyższony współczynnik mglistości tekstu – na poziomie 14,43 – co oznacza tekst zrozumiały dla studentów studiów licencjackich. Zabiegi redaktora artykułu mają wpłynąć na docieranie zawartych w nim informacji do lepiej wykształconych osób, interesujących się aspektami ochrony środowiska i badaniami prowadzonymi przez NASA.



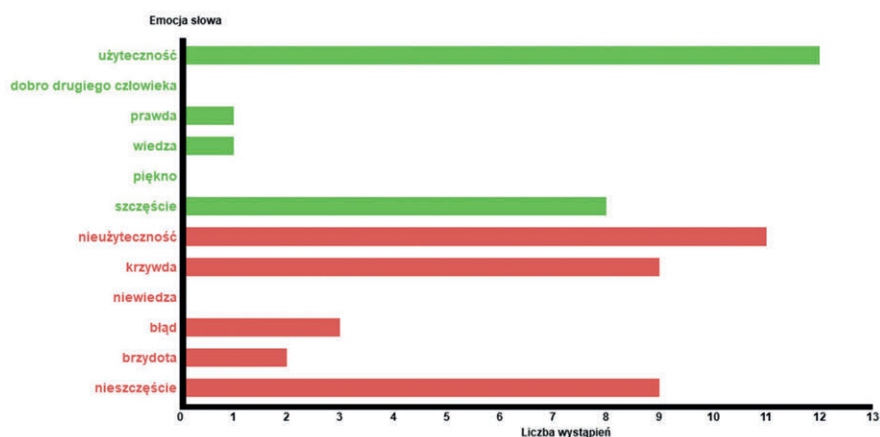
Il. 12. Rozkład obraźliwości artykułu ze strony nczas.com

Źródło: opracowanie własne w Hatespeech CLARIN



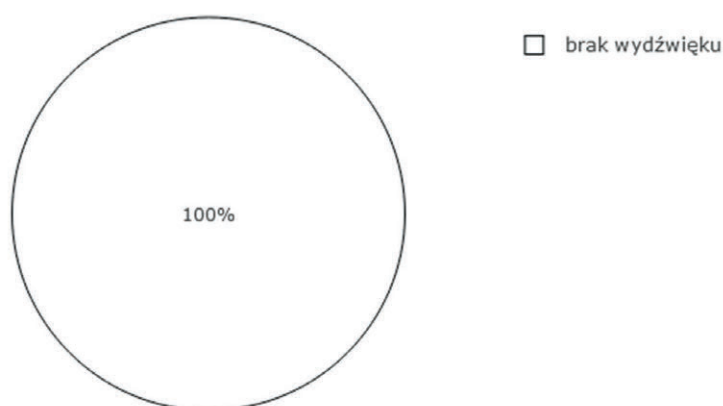
Il. 13. Analiza wydziwisku emocjonalnego artykułu ze strony nczas.com

Źródło: opracowanie własne w Wydziwisk CLARIN



Il. 14. Analiza dodatkowych cech wydźwięku emocjonalnego artykułu ze strony nczas.com

Źródło: opracowanie własne w Wydźwięk CLARIN



Il. 15. Polaryzacja wydźwięku emocjonalnego artykułu ze strony nczas.com

Źródło: opracowanie własne w Sentemo CLARIN

Podsumowanie

Badanie anatomii *fake newsów* za pomocą narzędzi dostępnych w ramach usług platformy CLARIN umożliwia wyróżnienie cech charakterystycznych FN. Najczęściej są to teksty pisane prostym językiem, zrozumiałym dla osób, które ukończyły edukację na poziomie szkoły średniej. Zdarzają się teksty

pisane bardziej skomplikowanym językiem, jednak może mieć to na celu nadanie im pseudonaukowego charakteru i uwiarygodnienie ich treści.

Ładunek emocjonalny FN jest zwykle negatywny, emocje pozytywne i neutralne występują w nich w niewielkim stopniu. Przeważają emocje takie jak strach, smutek, złość, krzywda, błąd i nieszczęście. Wśród najczęściej pojawiających się emocji o nacechowaniu pozytywnym występują użyteczność oraz radość. Ma to na celu wzbudzenie w odbiorcach zaufania i zachęcenie ich do dalszego rozpowszechniania treści.

Analiza metodą TF-IDF za pomocą narzędzia Cat oraz automatyczna ekstrakcja słów kluczowych umożliwiają odtworzenie zawartości treściowej FN bez konieczności czytania ich treści. Badanie dużych zbiorów FN może doprowadzić do wyodrębnienia najczęściej powtarzających się w nich słów kluczowych oraz kategorii tematycznych. Na tej podstawie artykuły zawierające słowa z takiego wyselekcjonowanego prototypowego zbioru będzie można wstępnie kwalifikować jako wiadomości nieprawdziwe.

Badania nad metodami detekcji i zwalczania *fake newsów* są jednym z wyzwań stojących przed współczesnymi naukami o mediach i komunikacji społecznej. Rozwój narzędzi do analiz jakościowych treści wiadomości zamieszczanych w internecie przyczyni się do wzrostu wykrywalności informacji fałszywych w sieci. Dalsze studia w tym zakresie powinny dążyć do połączenia metod ilościowych i statystycznych badania *fake newsów* z jakościową analizą ich treści. Odpowiednia kombinacja dostępnych narzędzi może wspomóc zbudowanie algorytmu wykrywającego *fake newsy* odpowiednio wcześniej i zapobiegającego ich rozpowszechnianiu.

Bibliografia

- Agarwal K., *RAKE: Rapid Automatic Keyword Extraction Algorithm*, Medium, 8.04.2020, <https://medium.datadriveninvestor.com/rake-rapid-automatic-keyword-extraction-algorithm-f4ec17b2886c> (dostęp: 06.07.2022).
- Alfabetyczny spis usług, CLARIN-PL, <https://clarin-pl.eu/index.php/alfabetyczny-spis-uslug/> (dostęp: 6.07.2022).
- Bąkowicz K., *Fake news. Produkt medialny czasów postprawdy*, Warszawa 2020.
- Campos R. et al., *YAKE! Keyword extraction from single documents using multiple local features*, „Information Sciences” 2020, Vol. 509.
- Cat, CLARIN-PL, <https://clarin-pl.eu/index.php/cat/> (dostęp: 6.07.2022).
- Głogowski D., *Grafen jest przenoszony od zaszczepionych do nieszczepionych*, Kancelaria Lega Artis, 25.04.2022, <https://legaartis.pl/blog/2022/04/25/grafen-jest-przenoszony-od-zaszczepionych-do-nieszczepionych/> (dostęp: 6.07.2022).

- Indeks czytelności FOG*, Wikipedia, https://pl.wikipedia.org/wiki/Indeks_czytelno%C5%9Bci_FOG (dostęp: 6.07.2022).
- KM, *Sensacyjny raport NASA! Zmianę klimatu powoduje orbita słoneczna Ziemi, a nie działalność człowieka*, nczas.com, 27.09.2019, <https://nczas.com/2019/09/27/sensacyjny-raport-nasa-zmiane-klimatu-powoduje-orbita-sloneczna-ziemi-a-nie-dzialalnosc-czlowieka/?fbclid=IwAR1nfGtoklU17tqs50ptzAK-fMZflao-nETXYJd9264xhteIvAQrDirsdY> (dostęp: 6.07.2022).
- LEM, CLARIN-PL, <https://clarin-pl.eu/index.php/lem/> (dostęp: 6.07.2022).
- Miodek J. et al., *Jak pisać o Funduszach Europejskich?*, Warszawa 2010.
- Nie, zmiany orbity Ziemi nie są przyczyną obecnych zmian klimatu*, Demagog, 18.10.2019, https://demagog.org.pl/fake_news/nie-zmiany-orbity-ziemi-nie-sa-przyczyna-obecnich-zmian-klimatu/ (dostęp: 6.07.2022).
- O nas*, CLARIN-PL, <https://clarin-pl.eu/index.php/o-nas/> (dostęp: 6.07.2022).
- Polska ma niezarejestrowane granice w ONZ? Powtarzalna dezinformacja!*, Demagog, 3.03.2022, https://demagog.org.pl/fake_news/polska-ma-niezarejestrowane-granice-w-onz-powtarzalna-dezinformacja/ (dostęp: 6.07.2022).
- Rosińska K., *Fake news. Geneza, istota, przeciwdziałanie*, Warszawa 2021.
- Sentemo*, CLARIN-PL, <https://clarin-pl.eu/index.php/sentemo/> (dostęp: 6.07.2022).
- TF-IDF*, CLARIN-PL, <https://clarin-pl.eu/index.php/tf-idf/> (dostęp: 6.07.2022).
- Upadliniec, *Oświadczenie ONZ – państwa Polska nie ma i nigdy nie było!*, Wielka Biała Polska, 5.01.2016, <https://archive.ph/8U8vc#selection-228.0-228.1> (dostęp: 6.07.2022).
- Wydźwięk*, CLARIN-PL, <https://clarin-pl.eu/index.php/wydzwiek/> (dostęp: 6.07.2022).
- YAKE! Keyword Extractor*, <http://yake.inesctec.pt/> (dostęp: 6.07.2022).