



# Prezentacja narzędzi CLARIN-PL

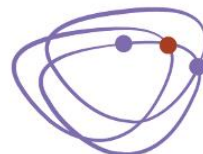
Michał Marcińczuk

[michal.marcinczuk@pwr.edu.pl](mailto:michal.marcinczuk@pwr.edu.pl)

Wrocław University of Science and Technology  
Department of Computational Intelligence  
G4.19 Research Group  
CLARIN-PL

Wykorzystanie technologii przetwarzania języka naturalnego w edukacji  
Warszawa, 12 grudnia 2019 r.

**CLARIN-PL**  
Common Language Resources and Technology Infrastructure



CENTRUM TECHNOLOGII  
JĘZYKOWYCH CLARIN-PL



Wrocław University  
of Science and Technology



Department of  
Computational  
Intelligence





- Archiwizacja i przechowywanie danych
  - DSpace, ClarinCloud
- Korpusy
  - korpusy dziedzinowe/tematyczne: Spokes, Paralela, Chronopress,
  - budowa własnych korpusów: Inforex, Korpusomat
- Zasoby leksykalne
  - Słownosieć, MWELexicon, Walenty, Słowniki kombinatoryczne HASK pol. i ang. kolokacji, Platforma leksykalna
- Podstawowe narzędzia
  - Any2txt, Txt clean, Morpho, WCRFT2, MorphoDita, Polem, Liner2, PolDeepNer, WSD, lobber, Parser, Spejd, Spatial
- Wydobywanie informacji z tekstu
  - Inkluz, Respa, Wydźwięk, Summarize, Tfidf, Cat, Sentemo
- Aplikacje badawcze
  - LEM, WebSty, Topic, TemoPL, MeWeX, CompCorp, Sentemo



- bez autoryzacji:
  - przeglądanie korpusów,
  - dostęp do zasobów leksykalnych,
  - dostęp do narzędzi badawczych,
- konto CLARIN-PL:
  - możliwość deponowania danych w archiwum długoterminowym DSpace,
  - dostęp do prywatnej chmury w ClarinCloud,
  - dostęp do narzędzi Inforex, MeWeX, Wosedon,
  - *kontext* w zakresie wgrania własnych korpusów,
- indywidualne zasady autoryzacji:
  - [korpusomat.pl](http://korpusomat.pl)



# Archiwizacja i przechowywanie danych



- [clarin-pl.eu/dspace](http://clarin-pl.eu/dspace)
- archiwum danych, metadanych, zasobów,
- deponowanie danych wymaga konta w CLARIN-PL,
- metadane są przeszukiwalne z poziomu VLO.

The screenshot shows the CLARIN-PL DSpace repository interface. At the top, there is a navigation bar with links for Home, Repository, Partners, and Contact. The CLARIN CENTRE B logo and the CENTRUM TECHNOLOGII JĘZYKOWYCH CLARIN-PL logo are also present. A search bar with a magnifying glass icon and a 'Search' button is located below the navigation bar. To the right of the search bar, there is a '2014 DSA 2017' badge. Below the search bar, there are three columns of filters: Author (Marcinczuk, Michał (33), Piasecki, Maciej (22), Kocot, Jan (20), Radziszewski, Adam (19), Kędzia, Paweł (18), ... View More), Subject (corpus (12), Polish (9), diachronic (8), plWordNet (7), syntax (7), ... View More), and Language (ISO) (Polish (223), English (11), Bulgarian (1), Lithuanian (1), Polish Sign Language (1), ... View More). Below the filters, there is a 'What's New' section with three items: 'CorpoGrabber' (ToolService), 'KPW Events' (Corpus), and 'Corpus IPI PAN' (Corpus). Each item has a brief description and a 'Publicly Available' or 'Distributed Under Creative Commons' label. On the right side of the interface, there is a 'What can you do?' section with links for DEPOSIT, CITE, Browse, My Account, Login, Register, General Information, RSS Feed, and Help Desk.



nextcloud.clarin-pl.eu

Prywatna chmura dla użytkowników CLARIN-PL do przechowywania danych

The screenshot displays the Nextcloud file manager interface. The left sidebar shows navigation options: Home, Recent, Favorites, Shared with you, Shared with others, Shared by link, and Tags. The main area shows the 'Documents' folder with a list of files. A context menu is open for the file 'result\_name.zip', showing options: Download, Download as zip, Export to Clarin services, and Delete. The 'Export to Clarin services' menu is expanded, showing options to export to dSpace, Inforex, Wosodon, and Mewex.

| File Name         | Size  | Modified    |
|-------------------|-------|-------------|
| About.odt         | 76 KB | 2 years ago |
| About.txt         | 1 KB  | 2 years ago |
| About_ccl.zip     | 3 KB  | a year ago  |
| result_name.zip   | 13 KB | 2 years ago |
| wikinews_test.txt | 2 KB  | 2 years ago |



---

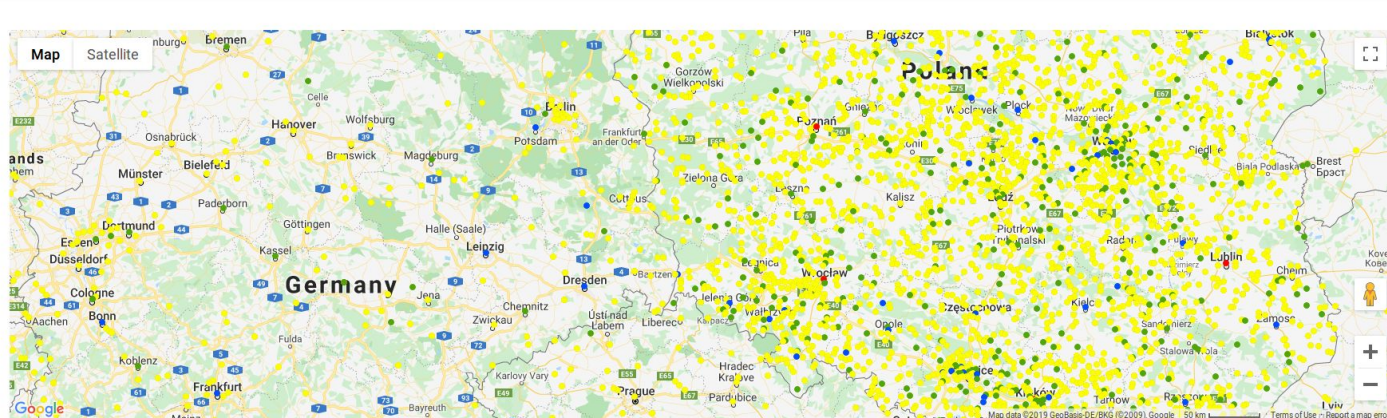
# Korpusy *dziedzinowe*



[chronopress.clarin.eu](http://chronopress.clarin.eu)

Korpus chronologiczny fragmentów tekstów prasowych z lat 1945–1954.

Podstawowe opcje: szeregi czasowe, profile wyrazów, konkordancje, mapa nazw, przeglądanie próbek.







## paralela.clarin-pl.eu

### Polsko-angielski korpus równoległy (320M słów)

Podstawowe opcje: wyszukiwanie fraz, przeglądanie kontekstów w obu językach

parallel corpus search

HOME • SEARCH • BROWSE • STATISTICS • HELP • LOGIN •

paralela

zawód

Slop: 0 7 Order: ✓

Options

Concordances Statistics Equivalence

Displaying segments 0 - 20 of 1632 in 2.479 s. Index contains 18,886,623 segments.

TMX Range

Show / hide columns Show 20 entries

First Previous 1 2 3 4 5 Next Last

Englism

| Lp | Polish                                                                                                                                                                                                                                                                                                                                                          | English                                                                                                                                                                                                                                                                                                                                                                                                                      |
|----|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 1  | I muszę to dzisiaj powiedzieć wprost Wysokiej Izbie i pani komisarz, moim zdaniem, rolników z Irlandii Północnej spotkał poważny <b>zawód</b> , nie tylko ze strony naszego Departamentu Rolnictwa, DARD, jak się go określa w Irlandii Północnej, ale najbardziej ze strony ministra, który wydaje się nieudolny, słaby i nie potrafi rozwiązać tego problemu. | They have done nothing wrong, committed no offence, and I have to state straightforwardly to this House tonight, and to the Commissioner, as far as I am concerned the farmers in Northern Ireland have been seriously let down, not only by our own Department of Agriculture, the DARD as it is known in Northern Ireland, and most especially by the Minister, who seems incapable, weak and unable to solve the problem. |
|    | Parlament Europejski                                                                                                                                                                                                                                                                                                                                            | European Parliament                                                                                                                                                                                                                                                                                                                                                                                                          |
| 2  | Chodzi tu o zezwolenia na prowadzenie taksówek, których otrzymanie wiąże się w wielu państwach członkowskich Unii Europejskiej z dużymi kosztami, oraz o <b>zawód</b> farmaceuty, w przypadku którego przy wydawaniu zezwoleń na prowadzenie działalności obowiązuje kryterium demograficzne i rozmieszczenia geograficznego.                                   | I refer to taxi licences, which in many Member States of the European Union are costly, and the <b>profession</b> of pharmacist, which is exercised under demographic and distribution measures.                                                                                                                                                                                                                             |
|    | Parlament Europejski                                                                                                                                                                                                                                                                                                                                            | European Parliament                                                                                                                                                                                                                                                                                                                                                                                                          |
| 3  | Należy przeprowadzić europejską ankietę dotyczącą koncepcji ostrzeżeń dla uzależnionych konsumentów, bez względu na ich wiek, <b>zawód</b> i płeć, ale również dla rynku wewnętrznego.                                                                                                                                                                          | It is necessary to carry out a European survey on the influence of this warning procedure for addicted consumers, regardless of their age, occupation and gender, but also for the internal market.                                                                                                                                                                                                                          |
|    | Parlament Europejski                                                                                                                                                                                                                                                                                                                                            | European Parliament                                                                                                                                                                                                                                                                                                                                                                                                          |
| 4  | Powinniśmy starannie rozważyć tę kwestię, nim założymy, że ktoś nie ma nieposzlakowanej opinii, która pozwoliłaby mu wykonywać ten <b>zawód</b> , jeśli nie zostało to dowiedzione.                                                                                                                                                                             | We should think carefully before assuming that someone is not of good repute to pursue an occupation if that is not proven.                                                                                                                                                                                                                                                                                                  |
|    | Parlament Europejski                                                                                                                                                                                                                                                                                                                                            | European Parliament                                                                                                                                                                                                                                                                                                                                                                                                          |
| 5  | Uważam, że praca rybaków, czy rybaczek, w zakresie w jakim kobiety wykonują ten <b>zawód</b> , jest jedną z najniebezpieczniejszych na świecie.                                                                                                                                                                                                                 | I believe that the job of a fisherman, or indeed a fisherwoman insofar as there are some, is one of the most dangerous jobs in the world.                                                                                                                                                                                                                                                                                    |
|    | Parlament Europejski                                                                                                                                                                                                                                                                                                                                            | European Parliament                                                                                                                                                                                                                                                                                                                                                                                                          |

Source

|                                     | Value              | Count |
|-------------------------------------|--------------------|-------|
| <input type="checkbox"/>            | 2ndjunglebook      | 0     |
| <input checked="" type="checkbox"/> | Academia           | 2     |
| <input type="checkbox"/>            | almayer            | 0     |
| <input type="checkbox"/>            | amyfoster          | 0     |
| <input checked="" type="checkbox"/> | argonautci         | 3     |
| <input type="checkbox"/>            | athtebay           | 0     |
| <input type="checkbox"/>            | bankholiday        | 0     |
| <input type="checkbox"/>            | Batory             | 0     |
| <input type="checkbox"/>            | billybudd          | 0     |
| <input type="checkbox"/>            | bisaraofpooree     | 0     |
| <input type="checkbox"/>            | bruhi              | 0     |
| <input type="checkbox"/>            | busan              | 0     |
| <input checked="" type="checkbox"/> | captainscourageous | 2     |
| <input checked="" type="checkbox"/> | chance             | 6     |
| <input type="checkbox"/>            | cienie             | 0     |
| <input checked="" type="checkbox"/> | ciezkieczasy       | 1     |
| <input checked="" type="checkbox"/> | CORDIS             | 19    |



## spokes.clarin-pl.eu

### Korpus nagrań i transkrypcji rozmów

Podstawowe opcje: pełnotekstowe wyszukiwanie, filtrowanie po metadanych, możliwość odsłuchania nagrania

[Home](#)
[Search](#)
[Browse](#)
[Stats](#)
[Help](#)

Slop: 
Order: ☒

[Concordances](#)
[Facets](#)
[Collocations](#)

Displaying utterances 0 - 20 of 24 in 0.497 s. Index contains 304,672 utterances.

Show / hide columns Show  entries

| #  | Left                                                             | Match | Right                                                            | Source                  | Context | All | URL | Pitch |
|----|------------------------------------------------------------------|-------|------------------------------------------------------------------|-------------------------|---------|-----|-----|-------|
| 1  | jaki to byłby przyjemny                                          | zawód | gdymy nie te dzieci nie ?                                        | Babskie sprawy          | +       | +   | URL | +     |
| 2  | staż żeby zostać lekarzem żeby móc wykonywać po studiach         | zawód | lekarza podobno dwa lata                                         | Szkolne rozmowy na t    | +       | +   | URL | +     |
| 3  | nie no jak ktoś lubi pracę w szkole to jest to fajny             | zawód |                                                                  | O studiach i służbie zc | +       | +   | URL | +     |
| 4  | Bartek zmienia                                                   | zawód | na ogrodnika ?                                                   | Śniadanie Wielkanocn    | +       | +   | URL | +     |
| 5  | no fajny                                                         | zawód |                                                                  | Śniadanie Wielkanocn    | +       | +   | URL | +     |
| 6  | nie wybierając zz za                                             | zawód | jednych jednego nie                                              | Wywiad na zajęcia z s   | +       | +   | URL | +     |
| 7  | ee ale . po co ci to mówię bo . pytał eś się . czy wybrać em ten | zawód | jaki chciał em wybrać                                            | trasa                   | +       | +   | URL | +     |
| 8  | to zapytałem się jaki                                            | zawód | pan uprawia                                                      | Rozmowa z rodzicami     | +       | +   | URL | +     |
| 9  | no a niby jaki                                                   | zawód | by miał uprawiać ?                                               | Rozmowa z rodzicami     | +       | +   | URL | +     |
| 10 | jakiś wolny                                                      | zawód | bardzo wolny zawód                                               | Rozmowa z rodzicami     | +       | +   | URL | +     |
| 11 | jakiś wolny zawód bardzo wolny                                   | zawód |                                                                  | Rozmowa z rodzicami     | +       | +   | URL | +     |
| 12 | bo w Anglii jest tak że każdy                                    | zawód | ma przypisaną minimalną pensję roczną i dużo wyższa jest pens... | Spotkanie w pubie       | +       | +   | URL | +     |
| 13 | w tej szkole elektrycznej będziesz miał dobry                    | zawód | pojedziesz na Zachód                                             | Wywiad biograficzny 1   | +       | +   | URL | +     |
| 14 | niezależnie jaki                                                 | zawód | badac jaką uczelnię                                              | Dyskusja naukowa        | +       | +   | URL | +     |
| 15 | sztynnych mogą przecież w każdej chwili zmienić                  | zawód | yy                                                               | AW_LUZ_59               | +       | +   | URL | +     |

First Previous 1 2 Next Last

Age  
Sex  
Audio  
Education  
Places  
Regions  
Origin

# Korpus Dyskursu Parlamentarnego



kdp.nlp.ipipan.waw.p

Zbiór anotowanych lingwistycznie tekstów z posiedzeń plenarnych Sejmu i Senatu RP, interpelacji i zapytań poselskich oraz posiedzeń komisji od roku 1919 do chwili obecnej

KORPUS DYSKURSU PARLAMENTARNEGO

O KORPUSIEINSTRUKCJATEKSTYWYSZUKIWANIE

WYSZUKIWANIE SEGMENTÓW

Zapytanie  
podatki

KONSTRUKTOR ZAPYTANMETADANESTATYSTYKI

Liczba wyników na stronie  
10

Wyszukaj

Znaleziono 17689 wyników.

| Lp | Lewy kontekst                                                               | Rezultat                          | Prawy kontekst                                                                | Data       |
|----|-----------------------------------------------------------------------------|-----------------------------------|-------------------------------------------------------------------------------|------------|
| 1  | , a nie jako właściciel ziemi, pomimo że płaci                              | podatki [podatek:subst:pl:acc:m3] | i inne ciężary musi ponosić z morga. Pozatem gdyby                            | 1920-07-06 |
| 2  | przemysłowy wymaga ujednostajnienia w większej jeszcze mierze, aniżeli inne | podatki [podatek:subst:pl:nom:m3] | , gdyż przemysł i handel nie jest w tym stopniu                               | 1920-07-06 |
| 3  | tub 8% od części zysku, w Wielkopolsce zaś                                  | podatki [podatek:subst:pl:nom:m3] | przemysłowe stanowią tylko 10,4% od części zysków                             | 1920-07-06 |
| 4  | to przemysłu czy handlu, wychodząc z założenia, że                          | podatki [podatek:subst:pl:nom:m3] | , oparte na słusznym i sprawiedliwym zasadach powinny płacić wszystkie        | 1920-07-06 |
| 5  | , która bardzo wstydliwie przyznaje, aby przekonać, że                      | podatki [podatek:subst:pl:nom:m3] | są dobrze obliczone, iż budka z wodą sodową będzie                            | 1920-07-06 |
| 6  | o tem mówić. Pozatem sprawozdanie wspomina, że i                            | podatki [podatek:subst:pl:nom:m3] | gruntowe zostaną odpowiednio podniesione. Jest to obietnica bardzo miła       | 1920-07-06 |
| 7  | zostało już podkreślone, że Komisja proponując wydatne i silne              | podatki [podatek:subst:pl:acc:m3] | przemysłowo handlowe, chciała stworzyć przykład do naśladowania dla rolnictwa | 1920-07-06 |
| 8  | stanowisko tym, że ludność b. Galicji pamięta jeszcze                       | podatki [podatek:subst:pl:acc:m3] | indemnizacyjne, na poczet których wpłaciła w końcu 19-go wieku                | 1939-05-31 |
| 9  | jako to: surowy materiał, administracja, oprocentowanie i                   | podatki [podatek:subst:pl:nom:m3] | , Panowie poszli tylko po linii najmniejszego oporu, i                        | 1931-03-06 |
| 10 | , ażeby sprzedać za bezcen, zaspokoić daniny i ogromne                      | podatki [podatek:subst:pl:acc:m3] | niestosunkowe do jego dochodu. Wskutek tego, że niema                         | 1931-03-06 |

«12345...

17681769»

# kontekst



## kontekst.clarin-pl.eu

### Wyszukiwarka tekstów anotowanych lingwistycznie

Podstawowe opcje: dostęp do kilku wgranych korpusów, wyszukiwanie pełnotekstowe, statystyki częstości, kolokacje

- ▼ LINDAT speech corpora
  - [Otázky Václava Moravce - political debates - Czech - words only](#)
- ▼ Added corpora
  - [Susanne](#)
  - [COCA](#)
  - [KPW](#)
  - [KPW 1.2](#)
- ▼ Sherlock Holmes parallel corpora
  - [Sherlock Holmes- The Adventure of Speckled Band](#)
  - [Sherlock Holmes i Pstrokata Opaska](#)
- ▼ Parallel corpora
  - [Bossong\\_PL](#)
  - [Bossong\\_LT](#)

**kon**text

Query Subcorpora | Save Concordance Filter Frequency Collocations View options Help

Search in the corpus

Corpus:

Query Type:

Query:

Specify context

Specify query according to the meta-information



Query Subcorpora | Save Concordance Filter Frequency Collocations View options Help

info: [Europarlament](#)  
664,378 positions

Hits: 4 | i.p.m.: 6.02 (related to the whole Europarlament) | ARF: 3 | Result is shuffled

|          |                                                                                                                                                             |
|----------|-------------------------------------------------------------------------------------------------------------------------------------------------------------|
| doc#756  | m w ramach wystąpień jednogodzinnych jako wieloletni nauczyciel akademicki i <b>lekarz</b> przypomnieć, że poczęte dziecko nie jest częścią ciała kobiety   |
| doc#3526 | ograniczenia powoływania się na klauzulę sumienia, zgodnie z którą <b>lekarz</b> może powstrzymać się od wykonania świadczeń zdrowotnych niezgodnych z jego |
| doc#1243 | niezbędne na tym polu. W jednej sprawie mam jako <b>lekarz</b> inne zdanie: nie mogę zgodzić się ze szczególnymi propozycjami                               |
| doc#2408 | , zagrożenie dla zdrowia człowieka. Dlatego nie tylko jako <b>lekarz</b> , ale przede wszystkim jako konsument, w pełni popieram                            |

niezwykle smutno, najwyraźniej europejscy politycy przestali czytać Biblię. Szkoda. Panie Przewodniczący! W czasie debaty o prawach kobiet nie dopuszczono mnie do głosu, dlatego chciała bym w ramach wystąpienia jednogodzinnych jako wieloletni nauczyciel akademicki i **lekarz** przypomnieć, że poczęte dziecko nie jest częścią ciała kobiety (jak głoszą niektóre panie przedstawicielki narodów Europy, które nie powinny przekazywać na forum Parlamentu Europejskiego nieprawdziwych informacji), kiedy już naukowo udowodniono, że w chwili zapłodnienia



# Korpusy

*tworzenie i anotacja*



Webowy system przeznaczony do tworzenia jakościowych korpusów tekstowych.

Podstawowe opcje: współdzielenie danych, wsparcie dla równoległej anotacji i koordynacji prac, możliwość definiowania własnych schematów (struktura kategorii) anotacji.



This page was tested in  Firefox. Page generated in 0 sec(s).



Webowa aplikacja do samodzielnego tworzenia oznakowanych fleksyjnie korpusów tekstów, z których można korzystać za pomocą wyszukiwarki MTAS

Podstawowe opcje: konwersja, tagowanie, rozpoznawanie nazw, indeksowanie, statystyki, język zapytań

KORPUSOMAT

[NOWY KORPUS](#)
[MOJE KORPUSY](#)
[INFORMACJE](#)
[KONTAKT](#)
[PREFERENCJE](#)
[WYLOGUJ](#)

## ZAPYTANIA DO KORPUSU 'POLSKIE TŁUMACZENIA BAŚNI HANSA CHRISTIANA ANDERSENA I BRACI GRIMM'

Opis informacji wyświetlanych na tym ekranie

|   | Forma bazowa                    | C-value | Liczba wystąpień |
|---|---------------------------------|---------|------------------|
| 1 | wolna lektura                   | 120.00  | 120              |
| 2 | dotatkowy materiał udostępniony | 95.10   | 60               |
| 3 | majątkowe prawo autorskie       | 95.10   | 60               |
| 4 | reprodukcja cyfrowa wykonana    | 93.51   | 59               |
| 5 | Piotr Wik                       | 72.00   | 73               |
| 6 | nowoczesna Polska               | 70.00   | 70               |
| 7 | strona wolnelektury             | 60.00   | 60               |
| 8 | tekst opracowany                | 60.00   | 60               |

[illegible]

Zapytanie  
motyl

KONSTRUKTOR ZAPYTAŃ

METADANE ▾

STATYSTYKI ▾

Liczba wyników na stronie  
10

Wyszukaj

Znaleziono 10 wyników.

| Lp | Lewy kontekst                                               | Rezultat                                         | Prawy kontekst                                                |
|----|-------------------------------------------------------------|--------------------------------------------------|---------------------------------------------------------------|
| 1  | drzewami i zabrał chmurę liści; na gałęzie obnażonej usiadł | <a href="#">motyl [motyl;subst;sg;nom;m2]</a>    | , może ostatni motyl tego roku. Upłynęły dni i                |
| 2  | liści; na gałęzie obnażonej usiadł motyl, może ostatni      | <a href="#">motyl [motyl;subst;sg;nom;m2]</a>    | tego roku. Upłynęły dni i tygodnie. Mgła wilgotna             |
| 3  | ust, jeden mały płatek oderwał się leciuchny i jak          | <a href="#">motyl [motyl;subst;sg;nom;m2]</a>    | biały spadał ku ziemi. W zielonym lesie upadł na              |
| 4  | płynął dalej, aż do innego kraju. Prześliczny biały         | <a href="#">motyl [motyl;subst;sg;nom;m2]</a>    | usiadł na listku lilii, aby się lepiej przyjrzeć maleriżej    |
| 5  | starsza – odpowiedział kwiatek. – Aha! – zawołał            | <a href="#">motyl [motyl;subst;sg;nom;m2]</a>    | i uciekł co prędzej, gdyż nie chciał, aby                     |
| 6  | coś zarzucił. Upłynęła wiosna, kończyło się lato,           | <a href="#">motyl [motyl;subst;sg;nom;m2]</a>    | nie wybrał sobie dotąd żony. Jesień w nowe barwy              |
| 7  | wiatru i chłodu. Lecz nie miały zapachu. A                  | <a href="#">motyl [motyl;impt;sg;sec;imperf]</a> | teraz właśnie przekładał nad wszystko przyjemny zapach; on mu |
| 8  | się z nią może? – Tę wybiorę – postanowił                   | <a href="#">motyl [motyl;impt;sg;sec;imperf]</a> | nieodmiennie. Zbliżył się i oświadczył jej swoje zamiary.     |
| 9  | W naszym wieku przyjaźń wystarczyć powinna. I tak został    | <a href="#">motyl [motyl;subst;sg;nom;m2]</a>    | na koszu i nie znalazł żony. Sam sobie winien                 |
| 10 | ani sposób pokazać się w letniej sukni. Na szczęście        | <a href="#">motyl [motyl;subst;sg;nom;m2]</a>    | dostał się przypadkiem do ciepłego pokoju, gdzie było jak     |

Pobierz wyniki (CSV)



# Zasoby leksykalne



Relacyjny słownik semantyczny odzwierciedlający system leksykalny języka polskiego



# Wordnet viewer – Słownosiec



[wordnet-viewer.clarin-pl.eu](http://wordnet-viewer.clarin-pl.eu)

System do interaktywnego przeglądania grafu Słownosieci

>

Wordnet viewer- Słownosiec

Select relation

zawód

clear form

zawód 1 (czuj) plWN

zawód 2 (zwz) plWN

Tips:

Hold **ctrl** to mark many elements at once

Right click on a circle element to show hidden nodes

Synset

Properties

Lexical units:

- kwalifikacje 1 (zwz) plWN

Relations *kwalifikacje 1 (zwz) plWN*

From:

- kolokacyjność
- wyjatkowy 1 (jak)

Created by Wojciech Rauk and Tomasz Naskręt for Clarin-PL



[clarin-pl.eu/dspace/handle/11321/508](http://clarin-pl.eu/dspace/handle/11321/508)

Słownik 56 tys. wyrażen wielowyrazowych

Podstawowe opcje: schematy wyrażen z przykładami

```
-<mwegroup class="subst" id="SubstSubstFix;s:S1,s:S2" name="AllSubstSubst" type="fix">
  -<condition>
    //aniol stróż and( inter(base[0],$s:S1), inter(class[0],{subst,ger,depr}), inter(base[1],$s:S2), inter(class[1],{subst,ger,depr}), inter(cas[0], cas[1]), setvar($Pos1, 0), setvar($Pos2, 1) )
  </condition>
  -<instances>
    -<MWE base="co nieco">
      <head>in(class[0],{subst,ger,depr})</head>
      <var name="S1">co</var>
      <var name="S2">nieco</var>
    </MWE>
    -<MWE base="coś niecoś">
      <head>in(class[0],{subst,ger,depr})</head>
      <var name="S1">coś</var>
      <var name="S2">niecoś</var>
    </MWE>
    -<MWE base="gorączka Oropouche">
      <head>in(class[0],{subst,ger,depr})</head>
      <var name="S1">gorączka</var>
      <var name="S2">Oropouche</var>
    </MWE>
    -<MWE base="dzidzia piernik">
      <head>in(class[0],{subst,ger,depr})</head>
      <var name="S1">dzidzia</var>
      <var name="S2">piernik</var>
    </MWE>
    -<MWE base="cząstka alfa">
      <head>in(class[0],{subst,ger,depr})</head>
      <var name="S1">cząstka</var>
      <var name="S2">alfa</var>
    </MWE>
```

# Słownik walencyjny języka polskiego oparty na danych korpusowych

Hasła
Administracja
Rozwinięcia typów fraz
Pobierz słownik
Zaloguj się

---

**malować**

| Hasło          | Status         |
|----------------|----------------|
| maksymalizować | (S) gotowe     |
| małsterstwo    | sprawdzone     |
| małeć          | (S) gotowe     |
| <b>malować</b> | (S) w obsłudze |
| malretować     | (S) gotowe     |
| malwersacja    | (S) gotowe     |
| mało           | sprawdzone     |
| małodusznie    | sprawdzone     |
| malpować       | (S) gotowe     |
| mały           | (S) gotowe     |
| małżeństwo     | (S) gotowe     |
| mama           | sprawdzone     |
| mamić          | (S) gotowe     |
| mamlać         | (S) gotowe     |
| mamlęć         | gotowe         |
| mamrotić       | (S) gotowe     |
| mandat         | (S) gotowe     |
| manewr         | sprawdzone     |
| manewrować     | (S) gotowe     |
| mania          | gotowe         |
| mańić          | (S) gotowe     |
| manifest       | sprawdzone     |
| manifestacja   | sprawdzone     |
| manifestować   | (S) gotowe     |
| manipulacja    | (S) gotowe     |
| manipulować    | (S) sprawdzone |
| manto          | gotowe         |
| manuskrypt     | gotowe         |
| mapować        | (S) gotowe     |
| maraton        | gotowe         |
| marginalizacja | (S) gotowe     |
| marginalizować | (S) sprawdzone |
| mariaż         | (S) gotowe     |
| markotnie      | sprawdzone     |
| markotnić      | (S) gotowe     |
| markotno       | sprawdzone     |
| markować       | (S) gotowe     |

Schematy [17]
Semantyka [8+0]
Przykłady [1]

**Schemat:** pewny [52238] ✓

|            |         |          |                |           |       |
|------------|---------|----------|----------------|-----------|-------|
| Funkcja:   | subj    |          |                |           |       |
| Typy fraz: | np(str) | np(inst) | prepnp(na,acc) | xp(instr) | recip |

**Schemat:** pewny [52236] ✓

|            |         |          |                |           |      |
|------------|---------|----------|----------------|-----------|------|
| Funkcja:   | subj    |          |                |           |      |
| Typy fraz: | np(str) | np(inst) | prepnp(na,acc) | xp(instr) | refl |

**Schemat:** pewny [35530] ✓

|          |      |  |  |  |  |
|----------|------|--|--|--|--|
| Funkcja: | subi |  |  |  |  |
|----------|------|--|--|--|--|

**Przegląd przykładów**
Filtrowanie schematów ✕

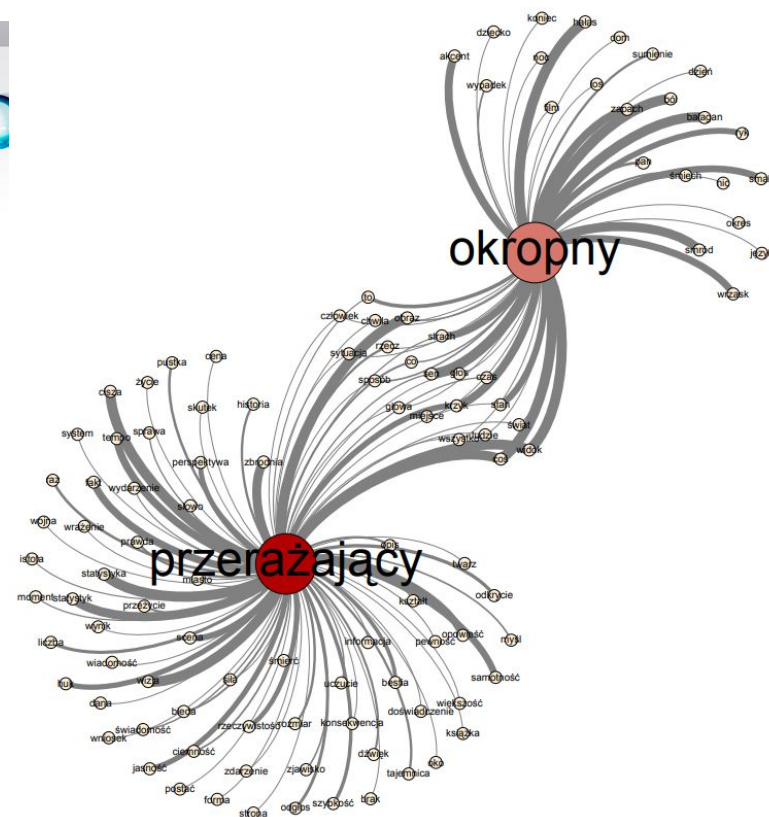
| Identyfikator: | Przykład:                                                                                                             | Źródło:                      | Ocena: |
|----------------|-----------------------------------------------------------------------------------------------------------------------|------------------------------|--------|
| 344724         | Jego uczestnicy odprawiają tańce Bharata Natyam, wystawiają tradycyjne sztuki indyjskie i malują się na kolory tęczy. | pełny NKJP (1800M segmentów) | dobry  |
| 344725         | Na zajęciach z makijażu, pod okiem doświadczonego wizażysty, panie malowały się kosmetykami firmy „Versace”.          | pełny NKJP (1800M segmentów) | dobry  |
| 344726         | - Malowała się reszkami starych szminek, uwielbiała się przebierać i marzyła o białej syplinie! [...]                 | pełny NKJP (1800M segmentów) | dobry  |

| Identyfikator: | Przykład:                                                                                                             | Źródło:                      | Ocen  |
|----------------|-----------------------------------------------------------------------------------------------------------------------|------------------------------|-------|
| 344724         | Jego uczestnicy odprawiają tańce Bharata Natyam, wystawiają tradycyjne sztuki indyjskie i malują się na kolory tęczy. | pełny NKJP (1800m segmentów) | dobry |
| 344725         | Na zajęciach z makijażu, pod okiem doświadczonego wizażysty, panie malowały się kosmetykami firmy „Versace”.          | pełny NKJP (1800m segmentów) | dobry |
| 344726         | - Malowała się resztkami szminek, uwielbiała się przebrać i marzyła o białej sylfide! [...] -                         | pełny NKJP (1800m segmentów) | dobry |



## The logo of the European Research Infrastructure Consortium (ERIC) is located in the top right corner. It features a stylized map of Europe in the background. Overlaid on the map are several blue circles connected by white lines, representing a network. In the center of the map, there is a circle of twelve yellow stars, similar to the European Union flag, with the text "A European Research Infrastructure" written below it.

Baza najczęstszych połączeń wyrazów wygenerowana z polskich i angielskich referencyjnych korpusów językowych.

[illegible]



[ws.clarin-pl.eu/w2vdemo.shtml](http://ws.clarin-pl.eu/w2vdemo.shtml)

Podobieństwo znaczeniowe z wykorzystaniem modelu word2vec zbudowanym na dużym korpusie tekstów w j. polskim

Podstawowe opcje: wyszukiwanie przez analogię, podobieństwo i eliminację niepasującego elementu

|                                                            |         |                                   |                         |                                       |         |                                      |                                  |
|------------------------------------------------------------|---------|-----------------------------------|-------------------------|---------------------------------------|---------|--------------------------------------|----------------------------------|
| <input type="text" value="mężczyzna"/>                     | jest do | <input type="text" value="król"/> | jak                     | <input type="text" value="kobieta"/>  | jest do | <input type="text" value="królowa"/> | <input type="button" value="↺"/> |
| <input type="text" value="podłoga"/>                       |         |                                   | jest podobne do         | <input type="text" value="posadzka"/> |         |                                      | <input type="button" value="↺"/> |
| <input type="text" value="śniadanie obiad droga kolacja"/> |         |                                   | nie pasującą frazą jest | <input type="text" value="droga"/>    |         |                                      | <input type="button" value="↺"/> |



# Platforma leksykalna



lexp.clarin-pl.eu

Integracja kilku zasobów leksykalnych

Podstawowe opcje: wyszukiwanie po formie tekstowej i lemacie

O APLIKACJI O NAS 

Opcje

RODZAJ ELEMENTU

JĘZYK

ELEMENT

**WYSZUKAJ**



▼ Słownica



▼ Słownik gramatyczny



▼ Walenty



▼ Inne języki



▼ Kolokacje



▼ Słownik wileński



▼ Podobieństwo



# Podstawowe narzędzia

# Podstawowe narzędzia – wstęp



- formularz webowy

- wgranie tekstu

- wyniki

- prosta wizualizacja
    - opcja pobrania w formacie csv, xml lub json (w zależności od narzędzia)

- usługa sieciowa (NlpRest2)

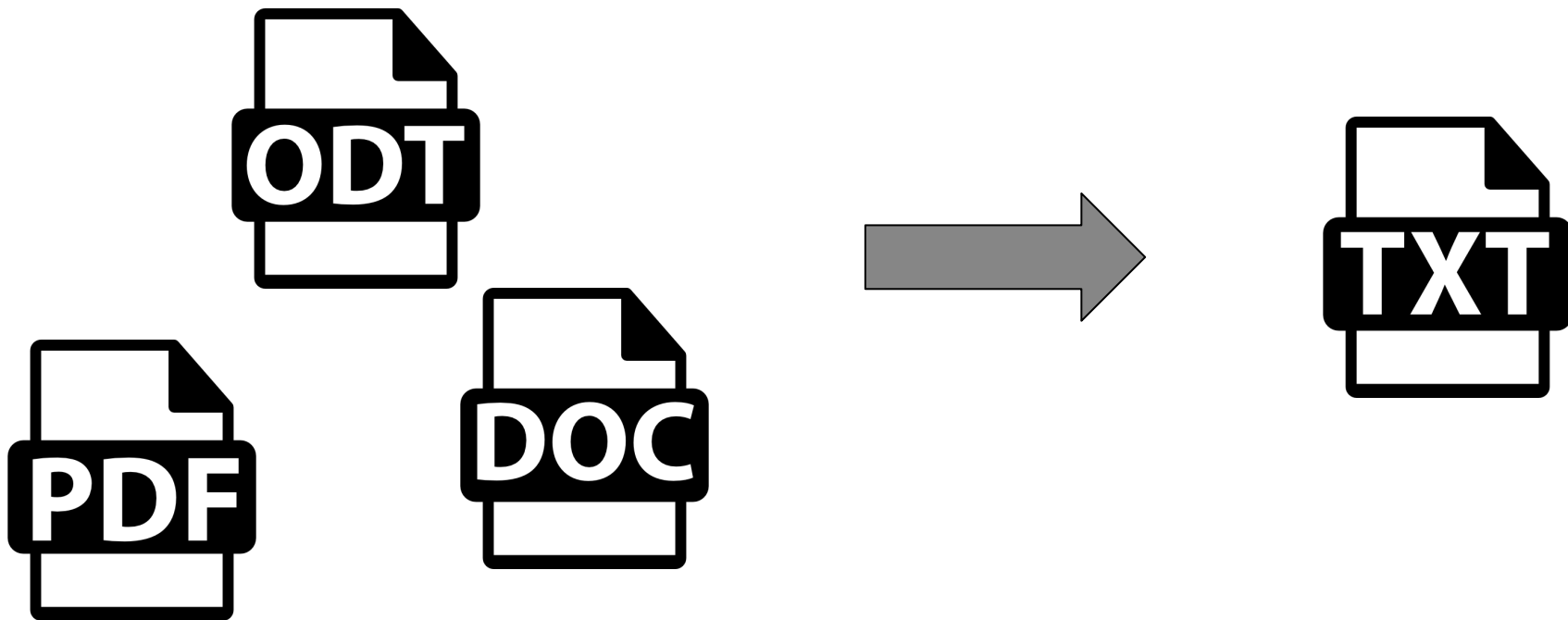
- dokumentacja — <http://nlp.pwr.wroc.pl/redmine/projects/nlprest2/wiki>
  - tutorial — <https://github.com/CLARIN-PL/NlpRest2-Tutorials>



## Usługa sieciowa

Konwersja dowolnego formatu tekstowego (doc, docx, odt, rtf, pdf) na czysty tekst

Podstawowy element wielu potoków przetwarzania



# Txt clean



[ws.clarin-pl.eu/txtclean.shtml](https://ws.clarin-pl.eu/txtclean.shtml)

Oczyszczanie tekstów na potrzeby dalszego przetwarzania

Podstawowe opcje: zamiana symboli, oczyszczanie z przeniesień słów, oczyszczanie z adresów URL

WYBIERZ FUNKCJĘ ⓘ

ZAMIANA Z

ZAMIANA NA

☒ Z UWZGLĘDNIENIEM WIELKOŚCI LITER

ZRESETUJ WSZYSTKIE OPCJE

PRZETWORZENIA

Reset

1. Funkcja: oczyszczanie z przeniesień  
2. Funkcja: oczyszczanie z URL

Cofnij się o krok

Tekst

Korpus

ALICJA ORZESZKOWA <https://adres1.com>

Pani budkowa  
WSPOMNIENIE [www.adres2.com](http://www.adres2.com)

Opowiem zdarzenie prawdziwe. Nic nie dodam, nic nie ujmę; opowiem tak jak było.  
Historja nieduża i jak prawie wszystkie historyjki tego świata, z początkiem wesołym,  
a z zakończeniem smutnym. Jeszcze teraz, gdy ją czasem wspominam, czyni mi się smutno  
i jakoś wstyd. Nie za siebie. Za kogoś? Może za ród ludzki. Przed kim? Przed ptakiem...

Zmiany

Zamiana przed/po

ALICJA ORZESZKOWA

Pani budkowa  
WSPOMNIENIE

Opowiem zdarzenie prawdziwe. Nic nie dodam, nic nie ujmę; opowiem tak jak było.  
Historja nieduża i jak prawie wszystkie historyjki tego świata, z początkiem wesołym,  
a z zakończeniem smutnym. Jeszcze teraz, gdy ją czasem wspominam, czyni mi się smutno  
i jakoś wstyd. Nie za siebie. Za kogoś? Może za ród ludzki. Przed kim? Przed ptakiem...  
Dawno... dawno... dawno temu, – jeszcze słońce życia na południowym punkcie  
niebios stało, – przyjechaliśmy z miasta na wieś, na kilkomiesięczny pobyt na wsi pięknej,

★

POBIERZ

# WCRFT2 i MorphoDita



[ws.clarin-pl.eu/tager.shtml](http://ws.clarin-pl.eu/tager.shtml)

Analiza morfo-syntaktyczna wraz z podziałem tekstu na zdania i tokeny

Podstawowe opcje: wybór tagera, słownika morfologicznego, opcja zgadywania nieznanych słów

|            |   |
|------------|---|
| WCRFT2     | ▼ |
| WCRFT2     |   |
| MorphoDiTa |   |

|                          |            |                                     |            |                                     |            |
|--------------------------|------------|-------------------------------------|------------|-------------------------------------|------------|
| <input type="checkbox"/> | ODGADYWACZ | <input checked="" type="checkbox"/> | MORFEUSZ 1 | <input checked="" type="checkbox"/> | MORFEUSZ 2 |
|--------------------------|------------|-------------------------------------|------------|-------------------------------------|------------|

XML

Tekst

Woda jest jedną z **najpospolitszych** substancji we Wszechświecie.  
Cząsteczka wody jest trzecią **naibardziej rozpowszechnioną** molekułą w ośrodku międzygwiazdowym, po cząsteczkowym wodorze i tlenku węgla.  
Jest również szeroko rozpowszechniona w kosmosie: stanowi istotny element budowy Ceres i księżyców lodowych krążących wokół planet-olbrzymów, jako domieszka występuje w ich atmosferach, a przypuszcza się, że duże jej ilości znajdują się we wnętrzach tych planet.  
Jako lód występuje także na części planetoid, a zapewne również na obiektach transneptunowych.  
Woda jest bardzo rozpowszechniona także na powierzchni Ziemi.  
Występuje głównie w oceanach, które pokrywają 70,8% powierzchni globu, ale także w rzekach, jeziorach i w postaci stałej w lodowcach.  
Część wody znajduje się w atmosferze (chmury, para wodna).  
Niektóre związki chemiczne zawierają cząsteczki wody w swojej budowie (hydraty – określa się ją wówczas mianem wody krystalizacyjnej).  
Zawartość wody włączonej w strukturę minerałów w płaszczu Ziemi może przekraczać łączną zawartość wody w oceanach i innych zbiornikach powierzchniowych nawet dziesięciokrotnie.  
Woda występująca w przyrodzie jest roztworem soli i gazów.  
Najwięcej soli mineralnych zawiera woda morska i wody mineralne; najmniej woda z opadów atmosferycznych.  
Wodę o małej zawartości składników mineralnych nazywamy wodą miękką, natomiast zawierającą znaczne ilości soli wapnia i magnezu – wodą twardą.  
Oprócz tego wody naturalne zawierają rozpuszczone substancje pochodzenia organicznego, np. mocznik, kwasy humusowe itp.



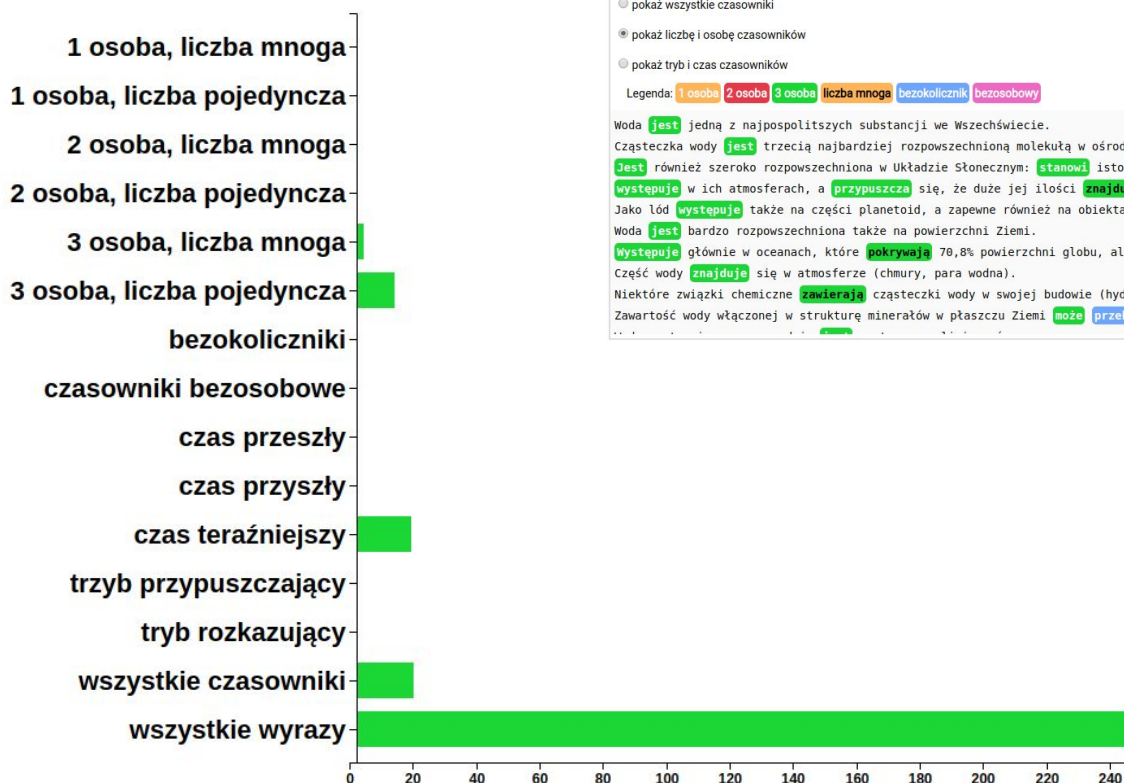
# Verbs



[ws.clarin-pl.eu/verbs.shtml](http://ws.clarin-pl.eu/verbs.shtml)

## Anotacja czasowników

Podstawowe opcje: kategoryzacja czasowników ze względu na liczbę, osobę i czas



XML Wykres Tekst Pobierz

☐ pokaż wszystkie czasowniki  
☒ pokaż liczbę i osobę czasowników  
☐ pokaż tryb i czas czasowników

Legenda: 1 osoba 2 osoba 3 osoba liczba mnoga bezokolicznik bezosobowy

Woda **jest** jedną z najpospolitszych substancji we Wszechświecie.  
Cząsteczka wody **jest** trzecią najbardziej rozpowszechnioną molekułą w ośrodku międzygwiazdowym, po cząsteczkowym wodorze i tlenku węgla.  
**Jest** również szeroko rozpowszechniona w Układzie Słonecznym: **stanowi** istotny element budowy Ceres i księżyców lodowych krążących wokół planet-olbrzymów, jako domieszka **występuje** w ich atmosferach, a **przypuszcza** się, że duże jej ilości **znajdują** się we wnętrzach tych planet.  
Jako lód **występuje** także na części planetoid, a zapewne również na obiektach transneptunowych.  
Woda **jest** bardzo rozpowszechniona także na powierzchni Ziemi.  
**Występuje** głównie w oceanach, które **pokrywają** 70,8% powierzchni globu, ale także w rzekach, jeziorach i w postaci stałej w lodowcach.  
Część wody **znajduje** się w atmosferze (chmury, para wodna).  
Niektóre związki chemiczne **zawierają** cząsteczki wody w swojej budowie (hydraty – **określa** się ją wówczas mianem wody krystalizacyjnej).  
Zawartość wody włączona w strukturę minerałów w płaszczu Ziemi **może** **przekraczać** łączną zawartość wody w oceanach i innych zbiornikach powierzchniowych nawet dziesięciokrotnie.



[ws.clarin-pl.eu/ner.shtml](http://ws.clarin-pl.eu/ner.shtml)

Rozpoznaje wystąpienia nazw własnych, wyrażeń temporalnych i sytuacji

Podstawowe opcje: wybór modelu znaczników

|                            |   |
|----------------------------|---|
| 9 kategorii                | ▼ |
| 82 kategorie               |   |
| 9 kategorii                |   |
| 5 kategorii                |   |
| timex1                     |   |
| timex4                     |   |
| granice nazw               |   |
| atrybuty i relacje zdarzeń |   |
| grupa modeli               |   |

XML

Tekst

Przebieg dyskusji obu liderów nie należał do przyjemnych.

Wołodomyr Zełeński był sfrustrowany.

Nie ukrywał, że Władimir Putin jest "trudnym" przeciwnikiem politycznym.

Sam przywódca Rosji zdawał się zupełnie przeciwnego zdania.

Zełeński przyznał w rozmowie z dziennikarzami, że liczył na większe osiągnięcia.

Tymczasem Władimir Putin oznajmił, że "jest szczęśliwy" z wyniku rozmów.

Co ciekawe, w podobnym tonie wypowiadali się m.in. Angela Merkel i Emmanuel Macron.

Negocjacje w tzw. formacie normandzkim odbyły się w Pałacu Elizejskim, a gospodarzem spotkania był Emmanuel Macron.

Rozmowy przywódców Francji, Niemiec, Ukrainy i Rosji trwały 5,5 godziny i dotyczyły uregulowania konfliktu w Donbasie.

Władimir Putin i Wołodomyr Zełeński przeprowadzili także pierwszą w historii rozmowę w cztery oczy.

W pisemnym oświadczeniu przywódcy Rosji i Ukrainy zgodzili się na wprowadzenie „pełnego i kompleksowego” zawieszenia broni we wschodniej Ukrainie do końca roku.

Zobowiązali się także do wymiany wszystkich więźniów związanych z konfliktem i wycofania sił wojskowych w 3 kolejnych regionach Ukrainy do końca marca 2020 r.

Nie określono jednak, o które obszary chodzi dokładnie.

Za 4 miesiące mają odbyć się kolejne rozmowy, podczas których przywódcy omówią postępy zawieszenia broni.



[ws.clarin-pl.eu/chunker.shtml](http://ws.clarin-pl.eu/chunker.shtml)

## Znakowanie fraz składniowych

Rodzaje fraz: uzgodnione, rzeczownikowe i czasownikowe

XML

Tekst

Woda jest jedną z najpospolitszych substancji we Wszechświecie.  
 Częsteczką wody jest trzecią najbardziej rozpowszechnioną molekułą w ośrodku międzygwiazdowym, po cząsteczkowym wodorze i tlenku węgla.  
 Jest również szeroko rozpowszechniona w Układzie Słonecznym: stanowi istotny element budowy Ceres i księżyców lodowych krążących wokół planet-olbrzymów, jako domieszka występuje w ich atmosferach, a przypuszcza się, że duże jej ilości znajdują się we wnętrzach tych planet.  
 Jako lód występuje także na części planetoid, a zapewne również na obiektach transneptunowych.  
 Woda jest bardzo rozpowszechniona także na powierzchni Ziemi.  
 Występuje głównie w oceanach, które pokrywają 70,8% powierzchni globu, ale także w rzekach, jeziorach i w postaci stałej w lodowcach.  
 Część wody znajduje się w atmosferze (chmury, para wodna).  
 Niektóre związki chemiczne zawierają cząsteczki wody w swojej budowie (hydraty – określa się ją wówczas mianem wody krystalizacyjnej).  
 Zawartość wody włączonej w strukturę minerałów w płaszczu Ziemi może przekraczać łączną zawartość wody w oceanach i innych zbiornikach powierzchniowych nawet dziesięciokrotnie.  
 Woda występująca w przyrodzie jest roztworem soli i gazów.  
 Najwięcej soli mineralnych zawiera woda morska i wody mineralne; najmniej woda z opadów atmosferycznych.  
 Wodę o małych zawartościach składników mineralnych nazywamy wodą miękką, natomiast zawierającą znaczne ilości soli wapnia i magnezu – wodą twardą.  
 Oprócz tego wody naturalne zawierają rozpuszczone substancje pochodzenia organicznego, np. mocznik, kwasy humusowe itp.



[ws.clarin-pl.eu/spejd.shtml](http://ws.clarin-pl.eu/spejd.shtml)

Płytki parser składniowy

Wynik w postaci pliku XML.

1 Woda jest jedną z najpospolitszych substancji we Wszechświecie. Częsteczka wody jest trzecią najbardziej rozpowszechnioną molekułą w ośrodku międzygwiazdowym, po cząsteczkowym wodorze i tlenku węgla. Jest również szeroko rozpowszechniona w Układzie Słonecznym: stanowi istotny element budowy Ceres i księżyców lodowych krążących wokół planet - olbrzymów, jako domieszka występuje w ich atmosferach, a przypuszcza się, że duże jej ilości znajdują się we wnętrzach tych planet. Jako lód występuje także na części planetoid, a zapewne również na obiektach transneptunowych. Woda jest bardzo rozpowszechniona także na powierzchni Ziemi. Występuje głównie w oceanach, które pokrywają 70,8 % powierzchni globu, ale także w rzekach, jeziorach i w postaci stałej w lodowcach. Część wody znajduje się w atmosferze (chmury, para wodna). Niektóre związki chemiczne zawierają cząsteczki wody w swojej budowie (hydraty - określa się ją wówczas mianem wody krystalizacyjnej). Zawartość wody włączonej w strukturę minerałów w płaszczu Ziemi może przekraczać





[github.com/CLARIN-PL/NlpRest2-Tutorials](https://github.com/CLARIN-PL/NlpRest2-Tutorials)

Lematyzacja fraz wielowyrazowych i nazw własnych

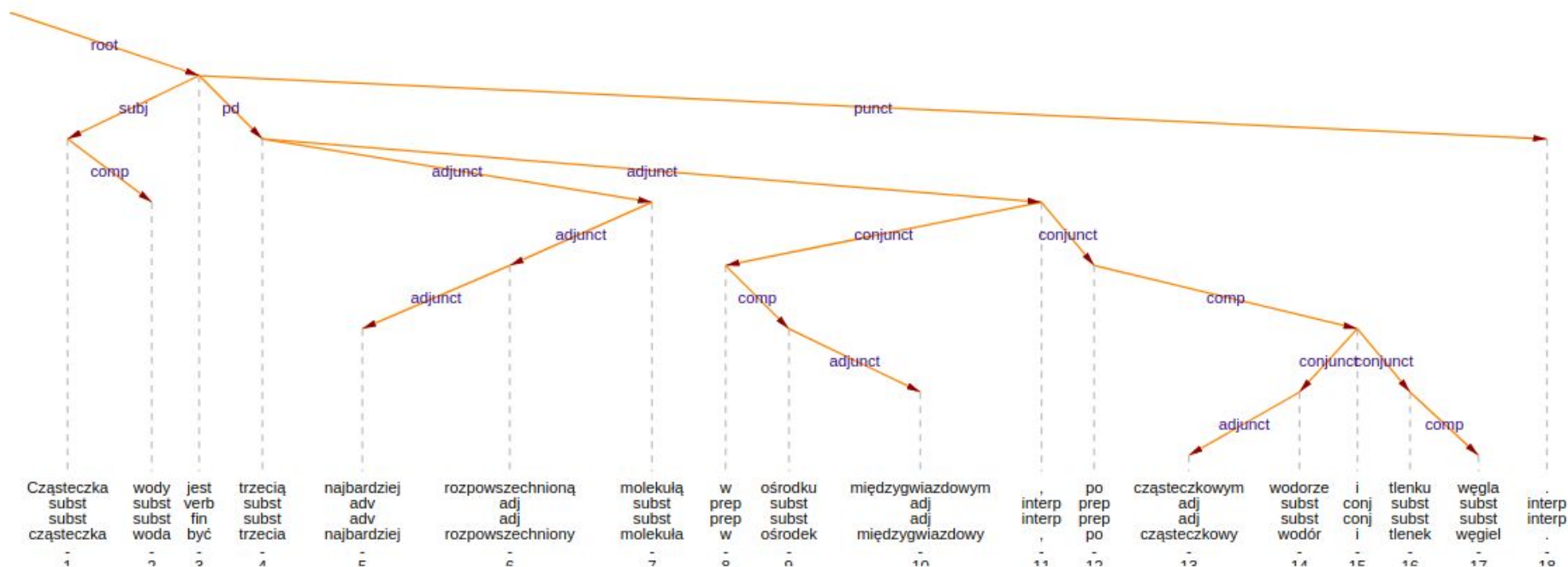
Dostępne tylko jako usługa sieciowa. Przetwarza listę fraz z podaną analia morfologiczna.

| Forma tekstowa             | Formy bazowe            | Lemat                   |
|----------------------------|-------------------------|-------------------------|
| Ministerstwie Finansów     | ministerstwo finanse    | Ministerstwo Finansów   |
| Irlandii Północnej         | Irlandia północny       | Irlandia Północna       |
| Brytyjskiej Izby Handlowej | brytyjski izba handlowy | Brytyjska Izba Handlowa |

## Ustalanie znaczeń słów na bazie Słownosieci



## Parser zależnościowy uczony na korpusie *Składnica* zależnościowa





[ws.clarin-pl.eu/tfidf.shtml](http://ws.clarin-pl.eu/tfidf.shtml)

Generowanie listy lematów z wartościami częstości, td, idf oraz tf-idf

| Lemat            | Częstość | TF       | IDF       | TFIDF    |
|------------------|----------|----------|-----------|----------|
| woda             | 16       | 0.114286 | 3.104936  | 0.354850 |
| sol              | 3        | 0.021429 | 7.450061  | 0.159644 |
| rozpowszechniony | 3        | 0.021429 | 6.887348  | 0.147586 |
| mineralny        | 3        | 0.021429 | 5.754460  | 0.123310 |
| zawartość        | 3        | 0.021429 | 4.748247  | 0.101748 |
| zawierać         | 4        | 0.028571 | 3.374734  | 0.096421 |
| cząsteczka       | 2        | 0.014286 | 6.705490  | 0.095793 |
| występować       | 4        | 0.028571 | 3.273876  | 0.093539 |
| planet           | 2        | 0.014286 | 6.450824  | 0.092155 |
| transneptunowych | 1        | 0.007143 | 12.268493 | 0.087632 |
| krystalizacyjny  | 1        | 0.007143 | 11.720853 | 0.083720 |
| hydrat           | 1        | 0.007143 | 10.957910 | 0.078271 |



[ws.clarin-pl.eu/spatial.shtml](http://ws.clarin-pl.eu/spatial.shtml)

Identyfikacja przyimkowych wyrażeń przestrzennych. Wyrażenie przestrzenne składa się z trójki: obiekt lokalizowany, przyimek lokatywny i lokalizator.

Wynik w postaci pliku json.

```
{
  "frames": [
    {
      "slots": {
        "debug": {
          "attributes": {
            "schema": "w,we#5a; w,we#w1",
            "pattern": "MALT"
          }
        },
        "trajector": {
          "attributes": {
            "sumo": "Entity, EngineeringComponent, Object"
          },
          "id": "a11"
        },
        "landmark": {
          "attributes": {
            "sumo": "Attribute, Atmosphere, Gas, EarthsAtmosphere, Region, Air"
          },
          "id": "a12"
        },
        "region": {},
        "spatial_indicator": {
          "id": "a10"
        }
      },
      "id": "x",
      "type": "spatial"
    },
    [...]
  ]
}
```

- figury w świątyniach

|                  |             |
|------------------|-------------|
| <b>Trajector</b> | figury      |
| <b>Indicator</b> | w           |
| <b>Landmark</b>  | świątyniach |

Pattern: <NG|Pact|PrepNG>

Schema: w,we#1; w,we#w1;

- koń w Radogoszczu

|                  |             |
|------------------|-------------|
| <b>Trajector</b> | koń         |
| <b>Indicator</b> | w           |
| <b>Landmark</b>  | Radogoszczu |

Pattern: <NG|Pact|PrepNG>

Schema: w,we#w1;



# Wydobywanie informacji z tekstu

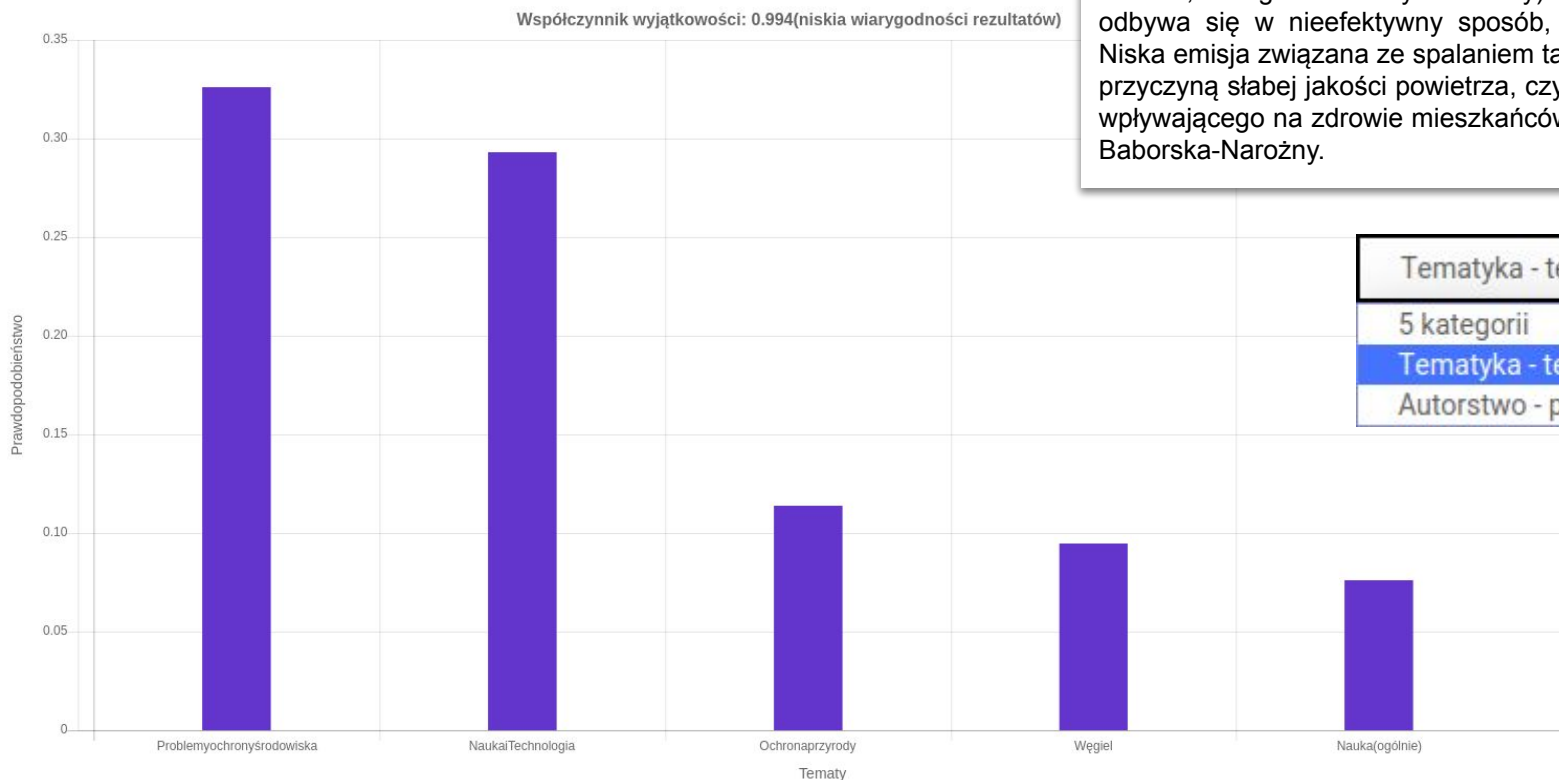


[ws.clarin-pl.eu/cat.shtml](http://ws.clarin-pl.eu/cat.shtml)

## Klasyfikacja tekstów

Podstawowe opcje: predefiniowane zestawy kategorii

Rząd Miejski we Wrocławiu sfinansuje badania dotyczące wpływu ogrzewania budynków na jakość powietrza w naszym mieście. Projekt realizowany jest przez naukowców z Wydziału Architektury i Wydziału Inżynierii Środowiska pod kierunkiem dr inż. arch. Magdaleny Baborskiej-Narożny i dr inż. Małgorzaty Szulgowskiej-Zgrzywy. Dużym problemem, z którym borykają się miasta w Polsce, a w tym też Wrocław, jest tzw. niska emisja związana ze spalaniem paliw stałych do ogrzewania budynków (czyli m.in. węgla kamiennego, węgla brunatnego, drewna, ekogroszku czy biomasy). – Ogrzewanie często odbywa się w nieefektywny sposób, paliwem o złej jakości. Niska emisja związana ze spalaniem takiego paliwa jest główną przyczyną słabej jakości powietrza, czyli smogu tak negatywnie wpływającego na zdrowie mieszkańców – uważa dr Magdalena Baborska-Narożny.



Tematyka - teksty prasowe

5 kategorii

Tematyka - teksty prasowe

Autorstwo - proza XIX i pocz. XX w



[ws.clarin-pl.eu/respa.shtml](http://ws.clarin-pl.eu/respa.shtml)

Generowanie tekstowych słów kluczowych w oparciu o PageRank i TfIdf.

|                                   |                 |
|-----------------------------------|-----------------|
| część woda                        | 0.0666276662297 |
| woda morski                       | 0.0666276662297 |
| łączny zawartość woda             | 0.0666276662297 |
| woda naturalny                    | 0.0666276662297 |
| cząsteczka woda                   | 0.0666276662297 |
| woda miękki                       | 0.0666276662297 |
| woda twardy                       | 0.0666276662297 |
| woda mineralny                    | 0.0666276662297 |
| cząsteczka woda                   | 0.0666276662297 |
| rozpowszechniony molekuła         | 0.0261846660711 |
| stan istotny element budowa ceres | 0.0235746260476 |
| swój budowa                       | 0.0235746260476 |
| znaczny ilość sol wapń            | 0.0224270149646 |
| mała zawartość składnik mineralny | 0.021638149797  |
| atmosfera                         | 0.0193451395253 |
| atmosfera                         | 0.0193451395253 |
| sol mineralny                     | 0.0191332236514 |
| roztwór sol                       | 0.0191332236514 |
| ocean                             | 0.0184437461077 |
| ocean                             | 0.0184437461077 |
| księżyc lodowy                    | 0.0143120091609 |
| olbrzym                           | 0.0143120091609 |

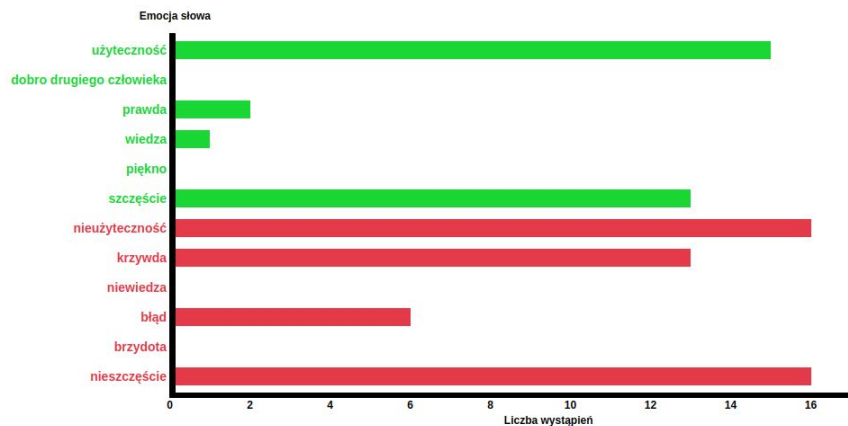
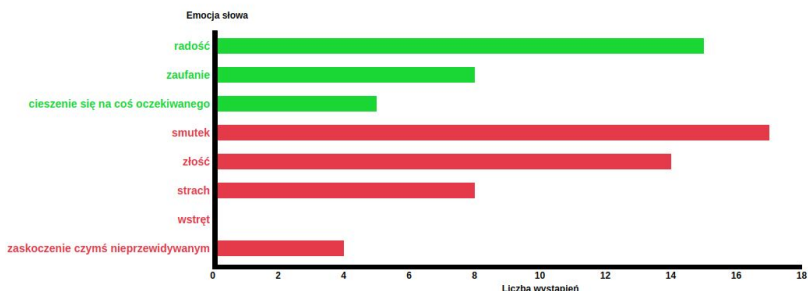
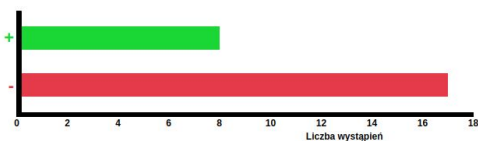


# Wydźwięk



[ws.clarin-pl.eu/sentyment.shtml](http://ws.clarin-pl.eu/sentyment.shtml)

Analiza wydźwięku ze względu na podstawowe typy emocji przypisane do słów



+ (8) - (17)

Radość(15) Zaufanie(8) Cieszenie się na coś oczekiwanego(5)  
Smutek(17) Złość(14) Strach(8) Zaskoczenie czymś  
nieprzewidywanym(4)

Użyteczność(15) Prawda(2) Wiedza(1) Szczęście(13)  
Nieużyteczność(16) Krzywda(13) Błąd(6) Nieszczęście(16)

Piją, jadą i okazują się niewinni. Ministerstwo Zbigniewa Ziobry chce to ukrócić. Nietrzeźwy kierowca bierze udział w wypadku, a potem twierdzi, że pił dopiero po kraksie – policja coraz częściej spotyka się z takimi sytuacjami. Resort sprawiedliwości chce je **całkowicie** wyeliminować. Jeśli sprawca wypadku sięgnie po butelkę, zanim zostanie poddany badaniu alkomatem, będzie traktowany tak, jakby spowodował go w stanie nietrzeźwości lub zbiegł z miejsca zdarzenia – zapowiada ministerstwo Zbigniewa Ziobry. W ten sposób resort zamierza walczyć z kierowcami, którzy przed sądem zarzekają się, że pili dopiero po zdarzeniu. Żeby ukoić nerwy. Co jakiś czas zdarzają się sytuacje, w których oskarżony próbuje w ten sposób, mówiąc kolokwialnie, **zrobić** sąd w konia. Powołujemy wtedy biegłych, ale nie zawsze są oni w stanie **jednoznacznie** określić, jakie stężenie miał kierujący w momencie wypadku – przyznaje sędzia Rafał Lisak z Sądu Okręgowego w Krakowie. Adam Reza z Polskiego Stowarzyszenia Biegłych Sądowych do Spraw Wypadków Drogowych tłumaczy, że wyniki badań retrospektywnych są obarczone dużą niepewnością. Zwłaszcza gdy kierowca **faktycznie** pije po wypadku, by maksymalnie utrudnić ustalenie stanu trzeźwości w określonym czasie. Dodatkowo zdarza się, że wychodzi z samochodu i sięga po butelkę tak, by widziało to jak najwięcej osób. W ten sposób uprawdopodobnia **swoją** linię obrony. Nie

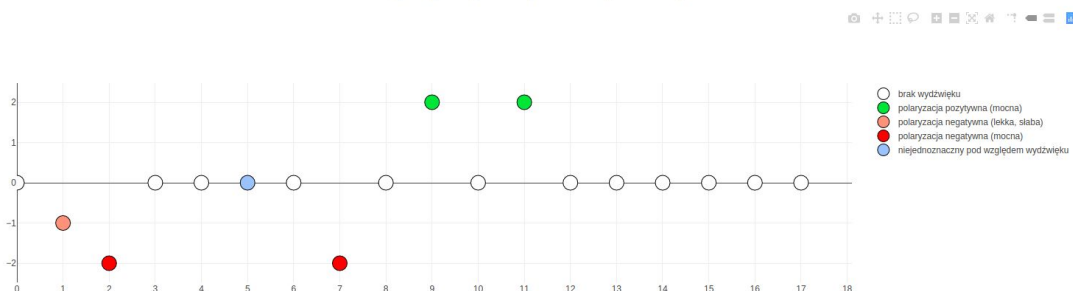


[ws.clarin-pl.eu/sentemo.shtml](http://ws.clarin-pl.eu/sentemo.shtml)

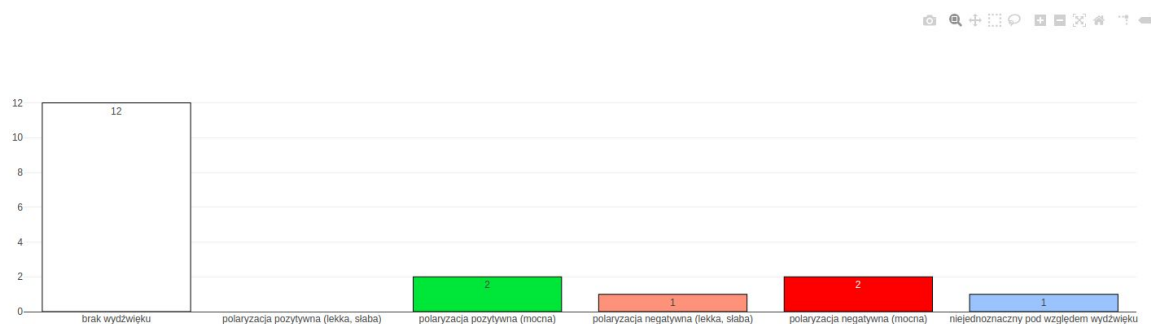
## Analiza wydźwięku

Wykrycie polaryzacji pozytywnej, negatywnej (mocnej, lekkiej), niejednoznaczności wydźwięku

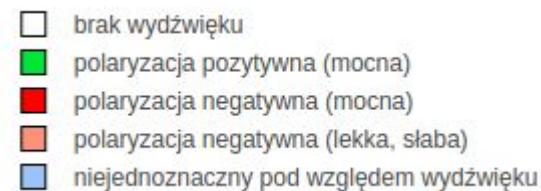
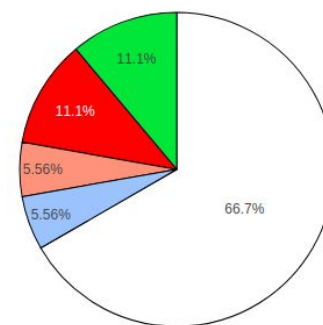
Polaryzacja wydźwięku emocjonalnego



Histogram wydźwięku emocjonalnego



Rozkład wydźwięku emocjonalnego





[ws.clarin-pl.eu/summarize.shtml](http://ws.clarin-pl.eu/summarize.shtml)

Streszczanie tekstów w oparciu o wybór najistotniejszych zdań

Woda jest jedną z najpospolitszych substancji we Wszechświecie.

Cząsteczka wody jest trzecią najbardziej rozpowszechnioną molekułą w ośrodku międzygwiazdowym, po cząsteczkowym wodorze i tlenku węgla. Jest również szeroko rozpowszechniona w Układzie Słonecznym: stanowi istotny element budowy Ceres i księżyców lodowych krążących wokół planet-olbrzymów, jako domieszka występuje w ich atmosferach, a przypuszcza się, że duże jej ilości znajdują się we wnętrzach tych planet. Jako lód występuje także na części planetoid, a zapewne również na obiektach transneptunowych. Woda jest bardzo rozpowszechniona także na powierzchni Ziemi. Występuje głównie w oceanach, które pokrywają 70,8% powierzchni globu, ale także w rzekach, jeziorach i w postaci stałej w lodowcach. Część wody znajduje się w atmosferze (chmury, para wodna). Niektóre związki chemiczne zawierają cząsteczki wody w swojej budowie (hydraty – określa się ją wówczas mianem wody krystalizacyjnej). Zawartość wody włączonej w strukturę minerałów w płaszczu Ziemi może przekraczać łączną zawartość wody w oceanach i innych zbiornikach powierzchniowych nawet dziesięciokrotnie.

Woda występująca w przyrodzie jest roztworem soli i gazów. Najwięcej soli mineralnych zawiera woda morska i wody mineralne; najmniej woda z opadów atmosferycznych. Wodę o małej zawartości składników mineralnych nazywamy wodą miękką, natomiast zawierającą znaczne ilości soli wapnia i magnezu – wodą twardą. Oprócz tego wody naturalne zawierają rozpuszczone substancje pochodzenia organicznego, np. mocznik, kwasy humusowe itp.

Jest również szeroko rozpowszechniona w Układzie Słonecznym: stanowi istotny element budowy Ceres i księżyców lodowych krążących wokół planet-olbrzymów, jako domieszka występuje w ich atmosferach, a przypuszcza się, że duże jej ilości znajdują się we wnętrzach tych planet. Występuje głównie w oceanach, które pokrywają 70,8% powierzchni globu, ale także w rzekach, jeziorach i w postaci stałej w lodowcach. Zawartość wody włączonej w strukturę minerałów w płaszczu Ziemi może przekraczać łączną zawartość wody w oceanach i innych zbiornikach powierzchniowych nawet dziesięciokrotnie.



[ws.clarin-pl.eu/inkluz.shtml](http://ws.clarin-pl.eu/inkluz.shtml)

Wykrywanie w polskim tekście obcojęzycznych wtrąceń

Tekst

Muszę lecieć bo mam zaraz **meeting**. Dzisiaj mamy **calla** z nowym klientem z **corpo**.



# Aplikacje badawcze



[ws.clarin-pl.eu/termopl.shtml](http://ws.clarin-pl.eu/termopl.shtml)

Narzędzie do automatycznego wyciągania polskiej terminologii w oparciu o C-value

Podstawowe opcje: tylko terminy wielowyrazowe, predefiniowane stopwords, eliminacja zaimków złożonych

Tabela

Pobierz

| #  | Ranking | Pojęcie                             | Pojęcie zlematyzowane                    | C-value | Długość | Freq_s | Freq_in | Kontekst |
|----|---------|-------------------------------------|------------------------------------------|---------|---------|--------|---------|----------|
| 1  | 3       | akcjonariusz większościowy          | [akcjonariusz większościowy]             | 7.0     | 2       | 8      | 2       | 2        |
| 2  | 3       | sąd wysoki                          | [sąd najwyższy]                          | 7.0     | 2       | 9      | 4       | 2        |
| 3  | 3       | użytkownik wieczysty                | [użytkownik wieczysty]                   | 7.0     | 2       | 12     | 5       | 1        |
| 4  | 3       | podział nieruchomości               | [podział nieruchomości]                  | 7.0     | 2       | 8      | 1       | 1        |
| 5  | 3       | wartość nieruchomości               | [wartość nieruchomości]                  | 7.0     | 2       | 8      | 3       | 3        |
| 6  | 4       | opłata adiacencki                   | [opłata adiacencka]                      | 6.8     | 2       | 8      | 5       | 4        |
| 7  | 5       | użytkownik wieczysty nieruchomości  | [użytkownik wieczysty nieruchomości]     | 6.3     | 3       | 4      | 0       | 0        |
| 8  | 5       | umowa kredyt indeksować             | [umowa kredytu indeksowanego]            | 6.3     | 3       | 4      | 0       | 0        |
| 9  | 5       | obniżyć wartość nieruchomości       | [obniżenie wartości nieruchomości]       | 6.3     | 3       | 4      | 0       | 0        |
| 10 | 7       | średnica obiektyw                   | [średnica obiektywu]                     | 6.0     | 2       | 7      | 2       | 2        |
| 11 | 7       | nowy płyta                          | [nowa płyta]                             | 6.0     | 2       | 6      | 0       | 0        |
| 12 | 7       | tytuł obniżyć wartość nieruchomości | [tytuł obniżenia wartości nieruchomości] | 6.0     | 4       | 3      | 0       | 0        |
| 13 | 7       | postępowanie administracyjny        | [postępowanie administracyjne]           | 6.0     | 2       | 8      | 4       | 2        |
| 14 | 7       | honor 9                             | [honor 9]                                | 6.0     | 2       | 6      | 0       | 0        |
| 15 | 9       | cena wykup                          | [cena wykupu]                            | 5.3     | 2       | 7      | 5       | 3        |



# Interfejs webowy do wydobywania z korpusów kolokacji w języku polskim

Podstawowe opcje: wybór funkcji dyspersji, miary ekstrakcji, typów strukturalnych

Funkcja dyspersji:

BRAK LUB KORPUS 1-PLIKOWY

ODCHYLENIE STANDARDOWE

JUILLAND D

LYNES D3

TF-IDF

WSPÓŁCZYNNIK ZMIENNOŚCI

DYSTRYBUCYJNA ZGODNOŚĆ

Miara ekstrakcji:

☒ FREQUENCY ⓘ

☒ W SPECIFIC CORRELATION ⓘ

☒ FREQUENCY BIASED MUTUAL DEPENDENCY ⓘ

☒ W SPECIFIC EXPONENTIAL CORRELATION ⓘ

☒ LOGLIKELIHOOD ⓘ

☒ W PEARSON'S CH2 ⓘ

☒ Z-SCORE ⓘ

☒ DICE ⓘ

☒ UNIGRAM SUBTUPLES ⓘ

☒ CONSONNI T1 ⓘ

☒ EXPECTED FREQUENCY ⓘ

☒ POINTWISE MUTUAL INFORMATION ⓘ

☒ MUTUAL DEPENDENCY ⓘ

☒ SPECIFIC EXPONENTIAL CORRELATION ⓘ

☒ MUTUAL EXPECTATION ⓘ

☒ PEARSON'S CH2 ⓘ

☒ T-SCORE ⓘ

☒ JACCARD ⓘ

☒ ODDS RATIO ⓘ

☒ SORGFENFREI ⓘ

☒ CONSONNI T2 ⓘ

☒ INVERSED EXPECTED FREQUENCY ⓘ

| Typy strukturalne:                                                                                                                                              |                                                                                                                                                                    |
|-----------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <input checked="" type="checkbox"/> REZIDENCJA / PRZYMOTNIK (ZOBOWIĄZENIE)   | <input checked="" type="checkbox"/> BURAKOWSKA / REZIDENCJA                     |
| <input checked="" type="checkbox"/> BERKOWSKI / KUBEL                        | <input type="checkbox"/> MIEKLEW PRZYMOTNIKOWY / PRZYGŁOSIAK                    |
| <input checked="" type="checkbox"/> COŚ / PRZYMOTNIK W UŚM                   | <input checked="" type="checkbox"/> REZIDENCJA / PRZYMOTNIK P                   |
| <input checked="" type="checkbox"/> REZIDENCJA / REZIDENCJA                  | <input checked="" type="checkbox"/> REZIDENCJA / BERE                           |
| <input checked="" type="checkbox"/> ZAMKÓW S. CH. W UŚM. / REZIDENCJA        | <input checked="" type="checkbox"/> LICZBNIK / REZIDENCJA                       |
| <input type="checkbox"/> DWA WTRYBY                                          | <input type="checkbox"/> CZADOWNIA / REZIDENCJA                                 |
| <input type="checkbox"/> WYRZĄCENIE PRZYMOTNIKA                              | <input checked="" type="checkbox"/> PRZYMOTNIK / REZIDENCJA / PRZYMOTNIK        |
| <input type="checkbox"/> CZADOWNIA / PRZYMOK / REZIDENCJA                    | <input type="checkbox"/> PRZYMOTNIK / REZIDENCJA / REZIDENCJA                   |
| <input checked="" type="checkbox"/> REZIDENCJA / PRZEMOK / REZIDENCJA        | <input type="checkbox"/> REZIDENCJA / PRZYMOTNIK (ZOBOWIĄZENIE + NEGOCJALNOŚĆ)  |
| <input type="checkbox"/> BURAKOWSKA / REZIDENCJA (NEGOCJALNOŚĆ)              | <input type="checkbox"/> BERKOWSKI / KUBEL (NEGOCJALNOŚĆ)                       |
| <input type="checkbox"/> MIEKLEW PRZYMOTNIKOWY / PRZYGŁOSIAK (NEGOCJALNOŚĆ)  | <input type="checkbox"/> COŚ / PRZYMOTNIK W UŚM (NEGOCJALNOŚĆ)                  |
| <input type="checkbox"/> REZIDENCJA / PRZYMOTNIK P (NEGOCJALNOŚĆ)            | <input type="checkbox"/> REZIDENCJA / REZIDENCJA (NEGOCJALNOŚĆ)                 |
| <input type="checkbox"/> REZIDENCJA / BERE (NEGOCJALNOŚĆ)                    | <input type="checkbox"/> ZAMKÓW S. CH. W UŚM. / REZIDENCJA (NEGOCJALNOŚĆ)       |
| <input type="checkbox"/> LICZBNIK / REZIDENCJA (NEGOCJALNOŚĆ)                | <input type="checkbox"/> DWA WTRYBY (NEGOCJALNOŚĆ)                              |
| <input type="checkbox"/> CZADOWNIA / REZIDENCJA (NEGOCJALNOŚĆ)               | <input type="checkbox"/> PRZYMOTNIK / REZIDENCJA / PRZYMOTNIK (NEGOCJALNOŚĆ)    |
| <input type="checkbox"/> CZADOWNIA / PRZYMOK / REZIDENCJA (NEGOCJALNOŚĆ)     | <input type="checkbox"/> PRZYMOTNIK / REZIDENCJA / REZIDENCJA (NEGOCJALNOŚĆ)    |
| <input type="checkbox"/> REZIDENCJA / PRZEMOK / REZIDENCJA (NEGOCJALNOŚĆ)    |                                                                                                                                                                    |

| Tabela  |    |                                        |   |                                   |                       | Pobierz                                                                                                                                                                                                                                                  |                 |
|---------|----|----------------------------------------|---|-----------------------------------|-----------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------|
| Ranking | φ  | Ilość                                  | φ | Relacja                           | Kolokacja             | Forma zlematyzowana                                                                                                                                                                                                                                      | Wszystkie formy |
| 1.77    | 13 | RZECZOWNIK + PRZYMIOTNIK (uzgodnienie) |   | subst:użytkownik adj:wieczysty    | użytkownik wieczysty  | 6:użytkownik wieczysty (subst.sg.nom.m1 adj.sg.nom.m1.pos), 3:użytkownika wieczystego (subst.sg.gen.m1 adj.sg.gen.m1.pos), 3:użytkownikowi wieczystemu (subst.sg.dat.m1 adj.sg.dat.m1.pos), 1:użytkownicy wieczysti (subst.pl.nom.m1 adj.pl.nom.m1.pos). |                 |
| 1.59    | 11 | LICZEBNIK + RZECZOWNIK                 |   | num:36 subst:usta                 | 36 ust                | 11:36 ust (num.pl.nom.m1 rec. subst.pl.gen.n).                                                                                                                                                                                                           |                 |
| 1.57    | 10 | RZECZOWNIK + RZECZOWNIK                |   | subst:sąd subst:najwyższy         | sąd najwyższy         | 6:Sąd Najwyższy (subst.sg.nom.m3 subst.sg.nom.m1), 4:Sądu Najwyższego (subst.sg.gen.m3 subst.sg.gen.m1).                                                                                                                                                 |                 |
| 1.42    | 8  | RZECZOWNIK + PRZYMIOTNIK (uzgodnienie) |   | subst:opłata adj:adiacenski       | opłata adiacenska     | 5:opłaty adiacenskiej (subst.sg.gen.f adj.sg.gen.f.pos), 2:Opłata adiacenska (subst.sg.nom.f adj.sg.nom.f.pos), 1:opłata adiacenska (subst.sg.nom.f adj.sg.nom.f.pos).                                                                                   |                 |
| 1.36    | 15 | RZECZOWNIK + RZECZOWNIK                |   | subst:wartość subst:nieruchomości | wartość nieruchomości | 10:wartości nieruchomości (subst.sg.gen.f subst.pl.nom.n), 4:wartości nieruchomości (subst.sg.nom.f subst.pl.nom.n), 1:wartością nieruchomości (subst.sg.instr.f subst.pl.nom.n).                                                                        |                 |
| 1.36    | 9  | RZECZOWNIK + RZECZOWNIK                |   | subst:wypuk subst:akcja           | wypuk akcji           | 6:wypuku akcji (subst.sg.gen.m3 subst.sg.gen.f), 2:wypukie akcji (subst.sg.loc.m3 subst.sg.gen.f), 1:wypuk akcji (subst.sg.nom.m3 subst.sg.gen.f).                                                                                                       |                 |
| 1.35    | 7  | RZECZOWNIK + RZECZOWNIK                |   | ign:Beautiful subst:trauma        | Beautiful Trauma      | 7:Beautiful Trauma (ign subst.sg.nom.f).                                                                                                                                                                                                                 |                 |
| 1.31    | 7  | RZECZOWNIK + RZECZOWNIK                |   | subst:średnica subst:obiektyw     | średnica obiektywu    | 3:średnicy obiektywu (subst.sg.gen.f subst.sg.gen.m3), 1:Średnica obiektywów (subst.sg.nom.f subst.pl.gen.m3), 1:średnicą obiektywu (subst.sg.instr.f subst.sg.gen.m3), 1:średnica obiektywów (subst.sg.nom.f subst.pl.gen.m3).                          |                 |
| 1.29    | 7  | RZECZOWNIK + RZECZOWNIK                |   | subst:obniżenie subst:wartość     | obniżenie wartości    | 3:obniżenia wartości (subst.sg.gen.m3 subst.sg.gen.f), 2:obniżeniu wartości (subst.sg.dat.n subst.sg.gen.f), 1:Obniżenie wartości (subst.sg.nom.n subst.sg.gen.m1), 1:obniżenie wartości (subst.sg.gen.m3 subst.sg.gen.f).                               |                 |



[ws.clarin-pl.eu/lem.shtml](http://ws.clarin-pl.eu/lem.shtml)

System eksploracji tekstów literackich

Podstawowe opcje: wybór predefiniowanego zadania do wykonania na wybranym zbiorze tekstów

Opcje

RODZAJ TAGERA ⓘ  ▼ Lematyzacja, cz  
- leksemu. Form

☐ UŻYCIE STOPLISTY ? ⓘ

WYBÓR ZADANIA ⓘ  ▼

- Lematyzacja - plik tekstowy
- Wyznaczanie części mowy
- Charakterystyki czasownikowe wypowiedzi
- Statystyki słów i części mowy
- Charakterystyki częstotliwościowych zdań i tokenów
- Wyznaczanie nazw własnych - posortowana lista nazw własnych
- Statystyki nazw własnych
- Ujednaczanie znaczeń leksykalnych
- Hiperonimy i hiponimy
- Wydźwięk
- Analiza stylometryczna (WebSty)
- Modelowanie tematyczne
- Konwersja do txt



POBIERZ XLSX



POBIERZ CSV

[ws.clarin-pl.eu/websty.shtml](http://ws.clarin-pl.eu/websty.shtml)

System do analizy stylometrycznej i podobieństwa tekstów

Podstawowe opcje: predefiniowane schematy, wybór cech, redukcja wymiarów, miary podobieństwa, wizualizacja wyników.

Wybór cech ^

GRAMATYCZNE I LEKSYKALNE SŁOWNIKOWE MODELOWANIE TEMATYCZNE WEKTORY DYSTYBUCYJNE

Elementy

☒ LEMATY ☐ Z listy ☐ FORMY WYRAZOWE ☐ Z listy

Interpunkcja

☐ DOWOLNE ZNAKI ☒ POZOSTAŁE ZNAKI Z LISTY

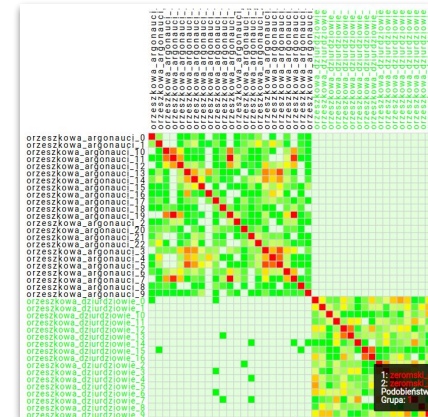
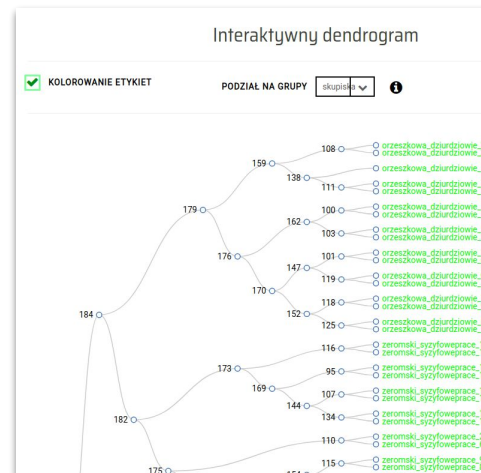
Części mowy

☐ CZASOWNIKI ☐ PRZYIMOTNIKI ☐ PRZYIMKI

☐ RZECZOWNIKI ☐ PRZYSŁÓWKI

Pozostałe klasy gramatyczne

☐ RZECZOWNIKI POSPOLITE (SUBST W NIKU) ☐ FORMY WINIEN ☐ PSEUDOSŁOWY



Opcje wyznaczania podobieństwa ^

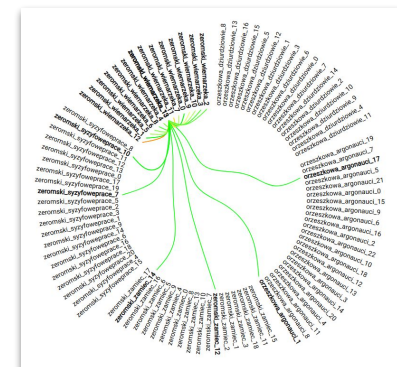
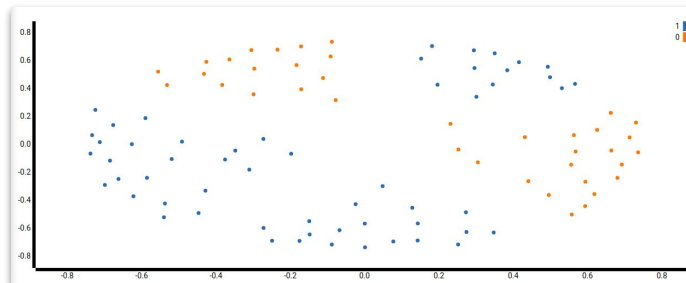
Wyznaczanie podobieństwa

METODA REDUKCJI WYMIARU

MIARA PODOBIEŃSTWA

Dan

Adres pliku (zip)



# Topic



[ws.clarin-pl.eu/topic.shtml](http://ws.clarin-pl.eu/topic.shtml)

## Modelowanie tematyczne

Opcje podstawowe

☒ PODZIAŁ PLIKÓW WEJŚCIOWYCH  METODA

USUWANIE RZADKICH LEMATÓW  LICZBA TEMATÓW

USUWANIE CZĘSTYCH LEMATÓW  LICZBA ITERACJI

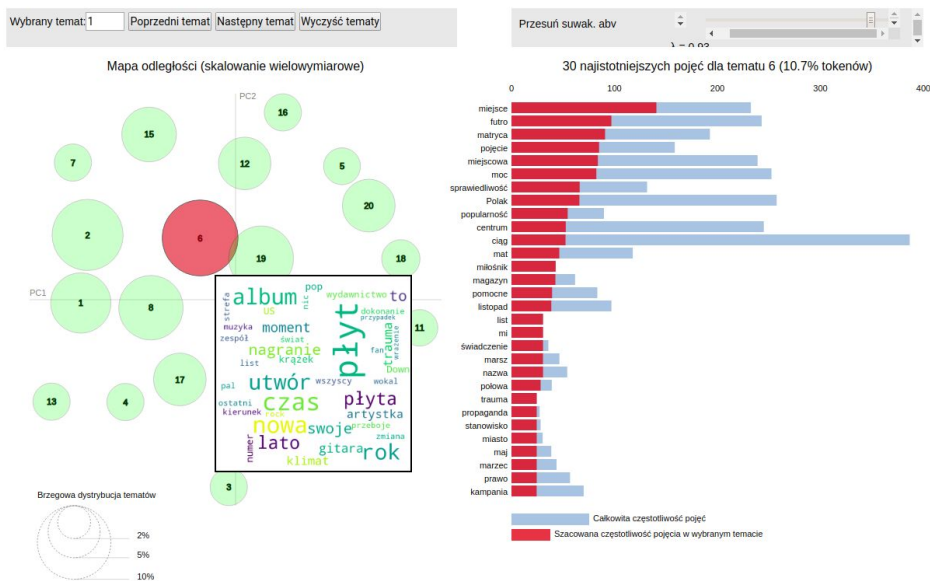
Stop lista

TYP LISTY

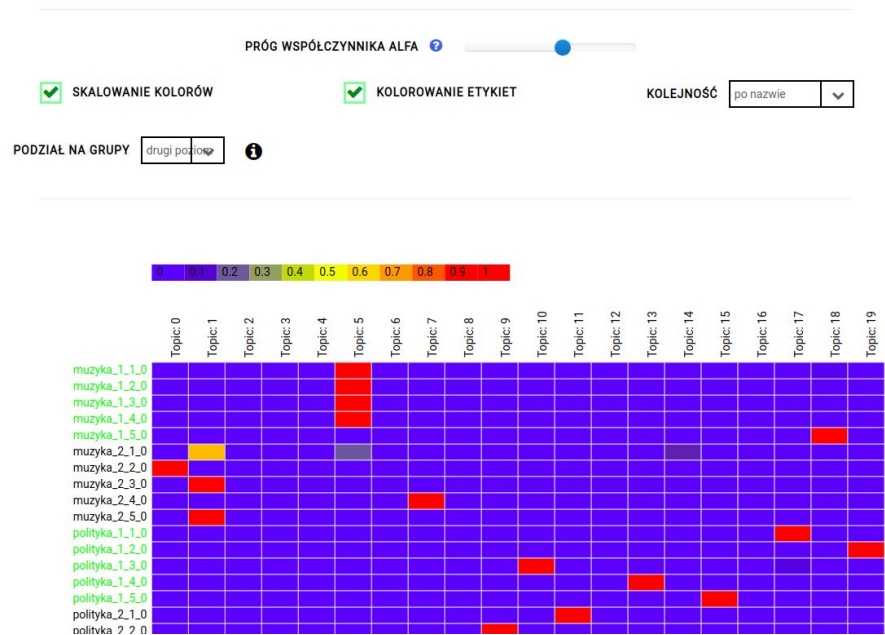
FILTR DOPUSZCZA DO ZLICZANIA WYŁĄCZAJĄCE CECHY

CZĘŚCI MOWY: ☐ CZASOWNIKI ☒ RZECZOWNIKI ☐ PRZYMIOTNIKI ☐ PRZYŚŁÓWKI ☐ PRZYIMIKI

### Dwuwymiarowa mapa odległości tematów

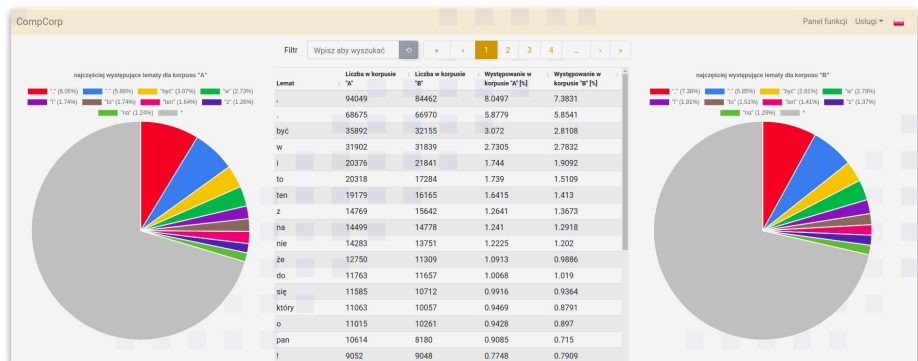
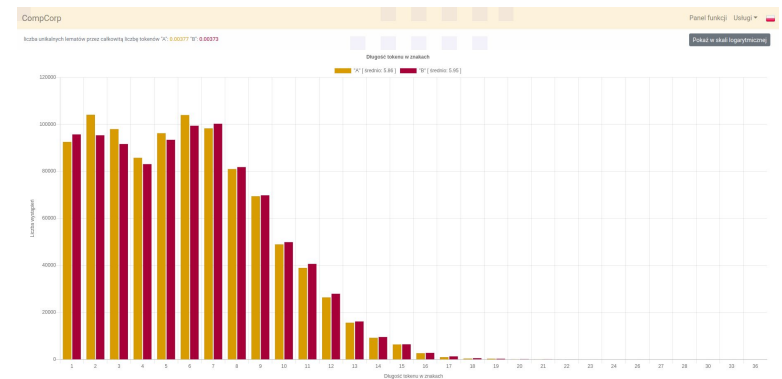


### Mapa ciepła ilustrująca udział tematów (alfa) w poszczególnych dokumentach





Podstawowe opcje: porównanie lematów, tagów, nazw własnych, długości zdań, terminów wielowyrazowych





---

Dziękuję za uwagę





W każdej sprawie:  
[clarin-pl@pwr.edu.pl](mailto:clarin-pl@pwr.edu.pl)

W sprawie działania usług sieciowych i narzędzi  
[webserwisy@clarin-pl.eu](mailto:webserwisy@clarin-pl.eu)

Opiekun użytkowników  
[jan.wieczorek@pwr.edu.pl](mailto:jan.wieczorek@pwr.edu.pl)

