

# Ręczna anotacja korpusów językowych

Marcin Oleksy

[marcin.oleksy@pwr.edu.pl](mailto:marcin.oleksy@pwr.edu.pl)

Politechnika Wroclawska  
Katedra Sztucznej Inteligencji

# Sytuacja wyjściowa



Sformułowany problem  
badawczy

Materiał językowy

# Sytuacja wyjściowa



Zespół

# Sytuacja wyjściowa



## Narzędzia CLARIN



# Sytuacja wyjściowa



## Narzędzia CLARIN



# Do czego służy ?

1. tworzenie cyfrowych korpusów tekstowych i zarządzanie nimi
2. anotacja
  - automatyczna
  - ręczna
3. analiza i prezentacja danych
  - statystyki
  - konkordancje
  - rozkład (dystrybucja)

# Do czego służy INFOREX?

## 1. tworzenie cyfrowych korpusów tekstowych

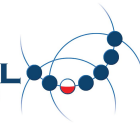
The image displays two screenshots from the INFOREX application. The top-left screenshot shows a button labeled "Add document(s)" with a plus icon. A callout box points to it, containing the text "Single document form" and "Upload zip file".

The bottom-left screenshot is titled "Upload a set of txt files." It features a "Zip file" section with a "Browse..." button and the text "No file selected." Below this, instructions state: "The Zip file must contain a set of **txt** files. File's metadata should be stored in a **ini** file with the same name as the **txt** file. The **ini** file must have the following format:"

```
[metadata]
url = "Path to a web with the document source"
publish_date = "Publish date in the format of YYYY-MM-DD"
author = "Author name"
title = "Document title"
```

Below the instructions is a "Subcorpus" dropdown menu set to "none" and a checkbox labeled "Split into subcorpora based on the file prefix: SUBCORPUS- DOCUMENT\_NAME..TXT." An "Upload" button is at the bottom.

The right screenshot is titled "Add document" and shows a form with two columns: "Metadata" and "Content". The "Metadata" column contains fields for Title, Author, Source, Filename, Subcorpus (set to "[unassigned]"), Status (set to "Przyjęty"), Publish date, Format (set to "plain"), Parent report ID, and Language. The "Content" column is a large text area with line numbers 1 through 32. A "Create document" button is at the bottom.



# Do czego służy ?

2. anotacja

# Role w procesie anotacji

- koordynator(ka)
- anotator(ka)
- superanotator(ka)
- badacz(ka)



# Co może i powinien robić koordynator?

## ludzie

- dodawanie użytkowników
- określanie ról

## dokumenty

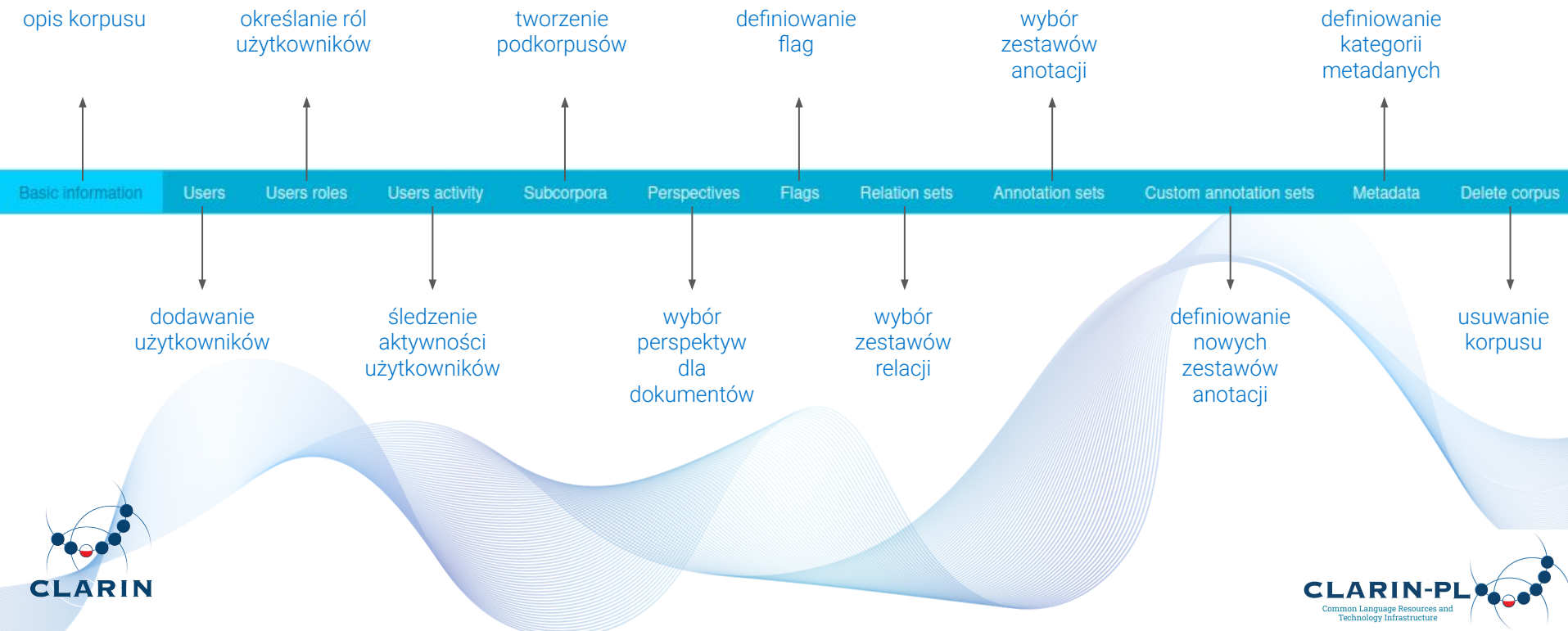
- dodawanie i usuwanie dokumentów
- przydzielanie dokumentów do anotacji
- opis korpusu (metadane)

## anotacje

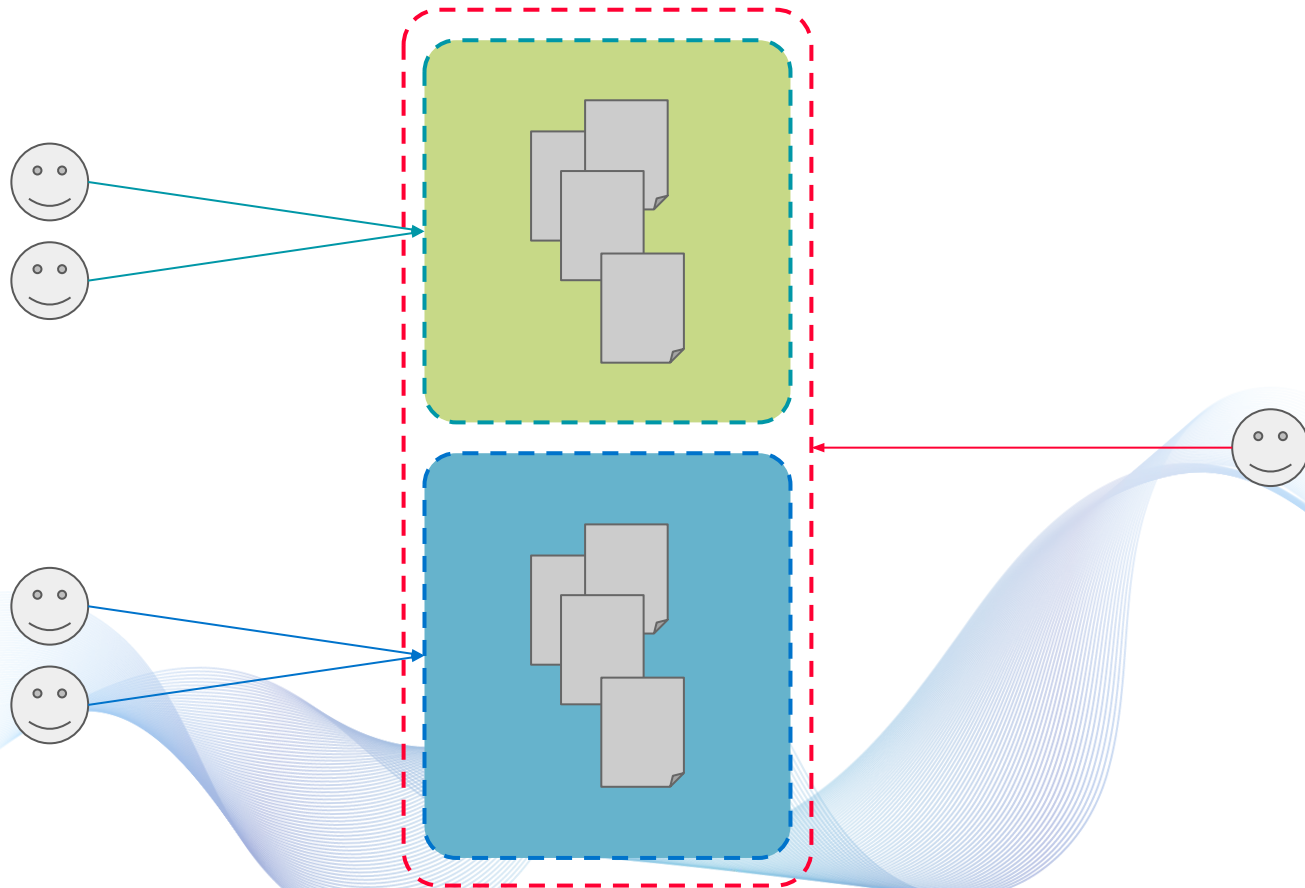
- dodawanie i usuwanie zestawów anotacji
- tworzenie zestawów anotacji

# Do czego służy ?

## 2. koordynowanie procesu anotacji



# Flagowanie dokumentów



# Do czego służy ?

## 2. anotacja korpusów językowych

- **oznaczenie** słów, innych fragmentów tekstu lub całych dokumentów **etykietami**, które przypisują dane jednostki do odpowiednich kategorii lub wnoszą dodatkowe informacje;
- na podstawie określonych **zasad**, z wykorzystaniem zdefiniowanych **zestawów znaczników**;
- obejmujące różne **warstwy** analizy, np.
  - znakowanie morfosyntaktyczne,
  - ujednocianie sensu słowa,
  - znakowanie składniowe,
  - znakowanie jednostek nazewniczych...
- Inne terminy: znakowanie, kodowanie, etykietowanie, tagowanie...

# Profil anotatora

Balans pomiędzy

- analitycznym i krytycznym myśleniem a wykonywaniem poleceń
- samodzielnością a umiejętnością współpracy



# Anotacja korpusów w Inforexie

- anotacja ręczna
  - praca w różnych systemach (1, 1+1, 2+1)
  - bogaty repertuar środków
    - anotacje
    - atrybuty anotacji
    - relacje między anotacjami
- anotacja automatyczna
  - wgrywanie anotacji pochodzących z innych narzędzi
  - narzędzia zintegrowane z Inforexem (taggery)
  - weryfikacja

# Anotacja korpusów w Inforexie (środowisko pracy)

Preview Content Viewer Agreement Metadata Relation agreement Content History of changes Tokenization Annotator Anaphora Anaphora Viewer Bootstrapping WSD

Annotation lemmas Annotation attributes Import annotations Morphological Disambiguation

### Document content

Toronto Dominion Centre

Toronto Dominion Centre - kompleks handlowo - kulturalny w kanadyjskim mieście Toronto , w Financial District . Składa się z 3 czarnych budynków , zaprojektowanych przez architekta Ludwiga Mies van der Rohe .

Budynki tworzą odgradzony od ulic dziedziniec , na którym Joe Fafard ustawił 6 odpoczywających krów z brązu . Pomiedzy budynkami stoi także wielkie krzesło . W południe odbywają się koncerty jazzowe .

W kompleksie znajduje się jedna z najważniejszych galerii sztuki Inuitów – Toronto Dominion Gallery of Inuit Art .

### View configuration

Annotations

Stage: final

User: -

Annotation set, subset or type		Display
Names	0 (143)	<input type="checkbox"/>
WSD	0 (139)	<input type="checkbox"/>
Chunking	0 (13)	<input type="checkbox"/>
Anaphora	0 (17)	<input type="checkbox"/>

Apply configuration

### Annotations

### Relations

### Flags

- Tokens
- Clean
- Sent
- Key
- Date
- Theme
- export
- License
- Source
- Title
- Author
- WSD\_AO
- WSD\_AS
- WSD\_BN
- WSD\_BA
- WSD\_AD
- ch\_AK
- ch\_AW
- ch\_DD
- ch\_2.0
- ch\_1
- ch\_2
- ch\_3

# Anotacja korpusów w Inforexie (środowisko pracy)

Preview Content Viewer Agreement Metadata Relation agreement Content History of changes Tokenization **Annotator** Anaphora Anaphora Viewer Bootstrapping WSD

Annotation lemmas Annotation attributes Import annotations Morphological Disambiguation

### Document content

Toronto Dominion Centre

Toronto Dominion Centre - kompleks handlowo - kulturalny w kanadyjskim mieście Toronto , w Financial District . Składa się z 3 czarnych budynków , zaprojektowanych przez architekta Ludwiga Mies van der Rohe .

Budynki tworzą odgradzony od ulic dziedziniec , na którym Joe Fafard ustawił 6 odpoczywających krów z brązu . Pomiędzy budynkami stoi także wielkie krzesło . W południe odbywają się koncerty jazzowe .

W kompleksie znajduje się jedna z najważniejszych galerii sztuki Inuitów - Toronto Dominion Gallery of Inuit Art .

### View configuration

Annotations

Annotation set, subset or type		Display
Names	0 (143)	<input type="checkbox"/>
WSD	0 (139)	<input type="checkbox"/>
Chunking	0 (13)	<input type="checkbox"/>
Anaphora	0 (17)	<input type="checkbox"/>
TimeML	0 (25)	<input type="checkbox"/>
Capital	0 (3)	<input type="checkbox"/>

Apply configuration

### Annotation types

### Annotations

### Relations

### Flags

- Tokens
- Clean
- Sent
- Key
- Date
- Theme
- export
- License
- Source
- Title
- Author
- WSD\_AO
- WSD\_AS
- WSD\_BN
- WSD\_BA
- WSD\_AD
- ch\_AK
- ch\_AW
- ch\_DD
- ch\_2.0
- ch\_1
- ch\_2
- ch\_3

# Anotacja korpusów w Inforexie (środowisko pracy)

Preview Content Viewer Agreement Metadata Relation agreement Content History of changes Tokenization **Annotator** Anaphora Anaphora Viewer Bootstrapping WSD  
Annotation lemmas Annotation attributes Import annotations Morphological Disambiguation

Document content

Toronto Dominion Centre

Toronto Dominion Centre - kompleks handlowo - kulturalny w kanadyjskim mieście Toronto, w Financial District. Składa się z 3 czarnych budynków, zaprojektowanych przez architekta Ludwiga Mies van der Rohe.

Budynki tworzą odgradzony od ulic dziedziniec, na którym Joe Fafard ustawił 6 odpoczywających krów z brązu. Pomiędzy budynkami stoi także wielkie krzesło. W południe odbywają się koncerty jazzowe.

W kompleksie znajduje się jedna z najważniejszych galerii sztuki Inuitów - Toronto Dominion Gallery of Inuit Art.

View configuration

Relations

Relation set	Display
Syntactic relations	<input type="checkbox"/>
Semantic relations	<input type="checkbox"/>
Coreference	<input type="checkbox"/>
Semantic relations by coreference	<input type="checkbox"/>
Spatial relations (final)	<input type="checkbox"/>

Apply configuration

Annotation types

Annotations

Relations

Flags

- Tokens
- Clean
- Sent
- Key
- Date
- Theme
- export
- License
- Source
- Title
- Author
- WSD\_AO
- WSD\_AS
- WSD\_BN
- WSD\_BA
- WSD\_AD
- ch\_AK
- ch\_AW
- ch\_DD
- ch\_2.0
- ch\_1
- ch\_2
- ch\_3

# Anotacja korpusów w Inforexie (praca)

Preview Content Viewer Agreement Metadata Relation agreement Content History of changes Tokenization **Annotator** Anaphora Anaphora Viewer Bootstrapping WSD

Annotation lemmas Annotation attributes Import annotations Morphological Disambiguation

Document content

Toronto Dominion Centre

Toronto Dominion Centre -1 -2 kompleks handlowo - kulturalny w kanadyjskim mieście 1 Toronto w 2 Financial District -1

Składa się z 3 czarnych budynków, zaprojektowanych przez architekta Ludwiga Mies van der Rohe

Budynki tworzą ogrodzony od ulic dziedziniec, na którym Joe Fafard ustawił 6 odpoczywających krów z brązu. Pomiedzy budynkami stoi także wielkie krzesło. W południe odbywają się koncerty jazzowe.

W kompleksie znajduje się jedna z najważniejszych galerii sztuki Inuitów - Toronto Dominion Gallery of Inuit Art

View configuration

Annotation types

- organization
  - nam\_org
  - nam\_org\_amb
  - nam\_org\_company
  - nam\_org\_group
  - nam\_org\_group\_band
  - nam\_org\_group\_team
  - nam\_org\_institution
  - nam\_org\_institution\_full\_deprecated
  - nam\_org\_nation
  - nam\_org\_organization
  - nam\_org\_organization\_sub
  - nam\_org\_organ

Annotations

Relations

Flags

- Tokens
- Clean
- Sent
- Key
- Date
- Theme
- export
- License
- Source
- Title
- Author
- WSD\_AO
- WSD\_AS
- WSD\_BN
- WSD\_BA
- WSD\_AD
- ch\_AK
- ch\_AW
- ch\_DD
- ch\_2.0
- ch\_1
- ch\_2
- ch\_3

# Anotacja korpusów w Inforexie (praca)

The screenshot displays the Inforexie web interface for corpus annotation. The top navigation bar includes tabs for Preview, Content Viewer, Agreement, Metadata, Relation agreement, Content, History of changes, Tokenization, Annotator (active), Anaphora, Anaphora Viewer, Bootstrapping, and WSD. Below this, there are sub-tabs for Annotation lemmas, Annotation attributes, Import annotations, and Morphological Disambiguation.

The main area shows document content with several annotations. A dialog box titled "Names" is open, displaying the "New state:" options:

- not ready
- ready to work
- under development
- ready
- ready and verified
- to correct

The right sidebar contains configuration panels: "View configuration", "Annotation types" (listing various organization-related types like `nam_org`, `nam_org_amb`, etc.), "Annotations", and "Relations". A "Flags" panel on the far right lists various flags such as `WSU_WK`, `WSD_DF`, `WSD_MO`, `WSD_a`, `QA_1_AD`, `Timex`, `TimexNorm`, `Zero verb`, `modality_ann`, `modality_rel`, `Name Rel`, `Names`, `Names (global)`, `Names_2`, `Nam_Rel_2`, `Chunks`, `Chunk Rel`, `Chunk Heads`, `Coref`, `SeRol_1`, `Coh`, `SeRol_2`, and `NE lemma`.

# Standardowy przebieg procesu anotacji

## (01) Projekt wytycznych

- Pierwsza wersja wytycznych
- Praca koncepcyjna

## (02) Doskonalenie wytycznych

- Kilka iteracji znakowania
- Wyjaśnienie kwestii dyskusyjnych i modyfikacja wytycznych
- Badanie zgodności między anotatorami
- Finalna wersja wytycznych

## (03) Faza produkcyjna

- Praca w wybranym systemie anotacji
- Stosowanie wytycznych
- Ewentualny powrót do dokumentów z fazy (02) i wprowadzenie koniecznych zmian

# Systemy anotacji

- 1
  - jedna osoba na dokument
- 1+1
  - dwie osoby na dokument
  - anotacja + weryfikacja
- 2+1 (fragmenty tekstu)
  - trzy osoby na dokument
  - niezależna anotacja prowadzona przez dwu anotatorów
  - uzgodnienie
- 2+1 (fragmenty tekstu + relacje)
  - 2+1 (fragmenty tekstu)
    - trzy osoby na dokument
    - niezależna anotacja prowadzona przez dwu anotatorów
    - uzgodnienie
  - 2+1 (relacje)
    - trzy osoby na dokument
    - niezależna anotacja prowadzona przez dwu anotatorów
    - uzgodnienie

# Najczęstsze błędy

- Pamięć zbiorowa
- Nieuporządkowana dokumentacja
- Brak podziału na iteracje
- Brak wytycznych
- Zmiany wytycznych w trakcie iteracji
- Rozwarstwienie anotacji (niezależnie pracujące zespoły, np. zespoły 2+1)
- Brak reakcji w przypadku utykania anotatorów

# Dobre praktyki

- Konfrontacje anotatorów
  - spotkania
  - praca na dokumentach
- Dostrajanie metody pracy do:
  - problemu badawczego (tagowanie morfosyntaktyczne vs. akty dialogowe)
  - typu zadania anotacyjnego (nowy problem badawczy vs. szybka anotacja)
- Wykorzystywanie samodzielności anotatorów, także do tworzenia podzadań, podgrup etc.
- Wykorzystywanie otwartych zbiorów w przypadku tymczasowego braku danych docelowych

# Anotacja korpusów w Inforexie (wybór trybu pracy)

The screenshot displays the Inforexie web interface for document annotation. The top navigation bar includes tabs for Preview, Content Viewer, Agreement, Metadata, Relation agreement, Content, History of changes, Tokenization, Annotator (selected), Anaphora, Anaphora Viewer, Bootstrapping, and WSD. Below the navigation bar, there are sub-tabs for Annotation lemmas, Annotation attributes, Import annotations, and Morphological Disambiguation. The main content area is titled "Document content" and shows a text snippet: "Toronto Dominion Centre". Below this, there are three paragraphs of text, each with a small "i" icon indicating an annotation point. The right sidebar contains a "View configuration" panel with "Working mode" options: "final" (selected), "agreement", and "relation agreement". Below this is an "Annotations" section with an "Apply configuration" button. At the bottom of the sidebar are sections for "Annotation types", "Annotations", and "Relations". On the far right, a "Flags" panel lists various annotation types with checkboxes and icons, including Tokens, Clean, Sent, Key, Date, Theme, export, License, Source, Title, Author, WSD\_AO, WSD\_AS, WSD\_BN, WSD\_BA, WSD\_AD, ch\_AK, ch\_AW, ch\_DD, ch\_2.0, ch\_1, ch\_2, and ch\_3.

Document content

Toronto Dominion Centre

Toronto Dominion Centre - kompleks handlowo - kulturalny w kanadyjskim mieście Toronto, w Financial District. Składa się z 3 czarnych budynków, zaprojektowanych przez architekta Ludwiga Mies van der Rohe.

Budynki tworzą odgradzony od ulic dziedziniec, na którym Joe Fafard ustawił 6 odpoczywających krów z brązu. Pomiędzy budynkami stoi także wielkie krzesło. W południe odbywają się koncerty jazzowe.

W kompleksie znajduje się jedna z najważniejszych galerii sztuki Inuitów - Toronto Dominion Gallery of Inuit Art.

View configuration

Working mode

final  
 agreement  
 relation agreement

Annotations

Apply configuration

Annotation types

Annotations

Relations

Flags

- Tokens
- Clean
- Sent
- Key
- Date
- Theme
- export
- License
- Source
- Title
- Author
- WSD\_AO
- WSD\_AS
- WSD\_BN
- WSD\_BA
- WSD\_AD
- ch\_AK
- ch\_AW
- ch\_DD
- ch\_2.0
- ch\_1
- ch\_2
- ch\_3

# Anotacja korpusów w Inforexie (badanie zgodności)

Comparison

9511688	<b>narodów świata</b>	a	
9512424[2268,2274]	<b>Izraela</b>	n(nam_loc_gpe_country)	
9511689		a	
9512442[2298,2315]	<b>Orderem Orła Białego</b>	n(nam_pro_award)	
9511690		a	
9512443[2331,2343]	<b>Order Uśmiechu</b>	n(nam_pro_award)	
9511691		a	
9512449[2350,2358]	<b>prezydent</b>	n(nam_oth_position)	
9511692		a	
9512420[2359,2362]	<b>Lech</b>	n(nam_liv_person_first)	
9511694		a	
9512419[2359,2371]	<b>Lech Kaczyński</b>	n(nam_liv_person)	
9511693		a	
9512421[2363,2371]	<b>Kaczyński</b>	n(nam_liv_person_last)	
9511695		a	
9512444[2404,2424]	<b>Pokojowej Nagrody Nobla</b>	n(nam_pro_award)	
9511696		a	
9512425[2420,2424]	<b>Nobla</b>	n(nam_liv_person_last)	
		a	
9512427[2443,2451]	<b>Warszawie</b>	n(nam_loc_gpe_city)	
9511697		a	
<b>Report 754185</b>			
9512456[50,57]	<b>Ignacego</b>	n(nam_liv_person_first)	
9511798		a	
9512455[50,71]	<b>Ignacego</b>	n(nam_liv_person)	
9511797	<b>Maciejowskiego</b>	a	
9512457[58,71]	<b>Maciejowskiego</b>	n(nam_liv_person_last)	
9511799		a	
9512459[72,77]	<b>Ignacy</b>	n(nam_liv_person_first)	
9511801		a	
9512458[72,93]	<b>Ignacy Sewer</b>	n(nam_liv_person)	
9511800	<b>Maciejowski</b>	a	
9511802[78,82]	<b>Sewer</b>	n(nam_liv_person_add)	
		a	
9512460[78,82]	<b>Sewer</b>	n(nam_liv_person_first)	
		a	
9512461[83,93]	<b>Maciejowski</b>	n(nam_liv_person_last)	
9511803		a	
9512462[120,131]	<b>Kobiernikach</b>	n(nam_loc_gpe_city)	
9511804		a	
9512463[158,170]	<b>austriackiego</b>	n(nam_adj_country)	
		a	
9512509[252,257]	<b>Polskę</b>	n(nam_loc_gpe_country)	
9511827		a	
9512465[281,300]	<b>powstaniu styczniowym</b>	n(nam_eve_human)	
9511805		a	
9512466[319,353]	<b>komisarza</b>	n(nam_oth_position)	

View configuration

Annotation types

Annotation set, subset or type	Count	Display
Names	143 (143)	<input checked="" type="checkbox"/>
TimeML	0 (25)	<input type="checkbox"/>
keywords	0 (4)	<input type="checkbox"/>
Spatial_3	0 (13)	<input type="checkbox"/>

Documents

By flag

Apply configuration

Agreement

Annotation category	Only A	A and B	Only B	PCS
nam_liv_person	72	808	23	94%
nam_liv_person_first	66	934	17	96%
nam_liv_person_last	65	990	33	95%
nam_loc_gpe_city	67	651	55	91%
nam_adj_country	58	385	40	89%
nam_eve_human	2	23	36	55%
nam_fac_goe	44	106	27	75%
nam_org_institution	46	176	72	75%
nam_pro_title_book	0	8	11	50%

# Anotacja korpusów w Inforexie (uzgadnianie)

Preview Content Viewer Agreement Metadata Relation agreement Content History of changes Tokenization Annotator Anaphora Anaphora Viewer Bootstrapping WSD

Annotation lemmas Annotation attributes Import annotations Morphological Disambiguation

Resolve annotations agreement

					<b>fno</b>
89	95	Toronto	fno	fno	Create annotation of matching types (more) <b>fno</b>
97	114	w Financial District	fpm	fpm	Create annotation of matching types (more) <b>fpm</b>
116	124	Sklada się	fwe	fwe	Create annotation of matching types (more) <b>fwe</b>
125	134	z 3 czarnych	-	fpmp	Do not create an annotation (more)
125	142	z 3 czarnych budynków	fpm	-	Do not create an annotation (more)
135	142	budynków	-	fno	Do not create an annotation (more)
144	159	zaprojektowanych	fpt	fpt	Create annotation of matching types (more) <b>fpt</b>
160	174	przez architekta	fpm	fpm	Create annotation of matching types (more) <b>fpm</b>

Apply actions

Filter annotations: All: 50 Final: 0 Agreed: 38 Choose: 12

Document content

Toronto Dominion Centre  
Toronto Dominion Centre - kompleks handlowo-kulturalny w kanadyjskim mieście Toronto, w Financial District. Składa się z 3 czarnych budynków, zaprojektowanych przez architekta Ludwiga Mies van der Rohe. Budynek tworzą odgradzony od ulic dziedziniec, na którym Joe Fafard ustawił 6 odpoczywających krów z brązu. Pomiedzy budynkami stoi także wielkie krzesło. W południe odbywają się koncerty jazzowe. W kompleksie znajduje się jedna z najważniejszych galerii sztuki Inuitów – Toronto Dominion Gallery of Inuit Art.

# Role w procesie anotacji

- koordynator(ka)
- anotator(ka)
- superanotator(ka)
- badacz(ka)



# Narzędzia analityczne w Inforexie

01

## Statystyki

- Anotacje i relacje
- Filtrowanie po metadanych, statusie anotacji i użytkownikach

02

## Dystrybucja

- Filtrowanie po statusie, kategorii, formie bazowej i formie z tekstu
- Eksport do CSV

03

## Konkordancje

- Filtrowanie po statusie, kategorii i podkorpusach
- Wizualizacja i eksport do PNG

# Narzędzia analityczne w Inforexie

Annotation statistics (click the row to expand/roll-back)

Annotation			Count		
Group	Subgroup	Category/Value	docs	unique	all
Names					55118
	adjective				2208
	event				1106
	facility				2078
	living				22283
	location				9507
	numex				1483
	organization				8593
		nam_org	8	9	16
		nam_org_amb	2	2	3
		nam_org_company	358	620	1146
		nam_org_group	78	91	133
		nam_org_group_band	51	123	167
		nam_org_group_team	68	404	630
		nam_org_institution	542	1167	3277
		nam_org_institution_full_deprecated	69	106	225
		nam_org_nation	200	207	544
		nam_org_organization	520	839	2101
		nam_org_organization_sub	9	11	15
		nam_org_politicalParty	108	119	336
	other				2277
	product				5583

Common filters Reset filters Copy URL

Status: Nieznany, Przyjęty, Załącznik, Odrzucony, Odłożony, Do usunięcia, **all**

Flags:

style: press, colloquial, rhetorical, scientific, legal, literary, **all**

Annotation filters

Stage:

User:

# Narzędzia analityczne w Inforexie

Annotation frequency

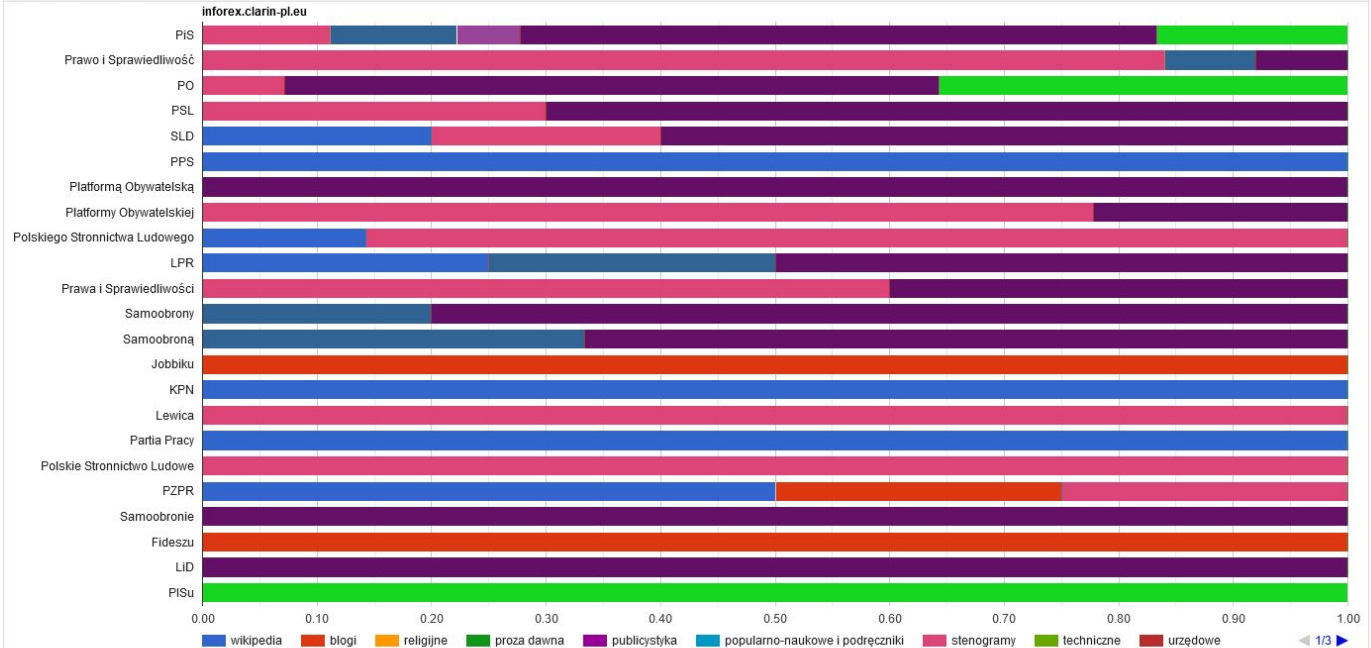
Filters: Annotation stage:  Annotation set:  Annotation type:  Subcorpus:  Phrase:

Annotation frequency

No.Base	Type	Count	Docs
1	PIS	41	18
2	Prawo i Sprawiedliwość	29	25
3	PO	19	14
4	PSL	17	10
5	SLD	13	11
6	PPS	11	1
7	Platformą Obywatelską	10	8
8	Platformy Obywatelskiej	10	9
9	Polskiego Stronnictwa Ludowego	9	7
10	LPR	7	4
11	Prawa i Sprawiedliwości	6	5
12	Samobrony	6	5
13	Samobroną	5	3
14	Jobbiku	4	1
15	KPN	4	1
16	Lewica	4	3
17	Partia Pracy	4	1
18	Polskie Stronnictwo Ludowe	4	3
19	PZPR	4	4
20	Samobronie	4	2
21	Fideszu	3	2
22	LID	3	2
23	PISu	3	2

Page 1 of 6 Displaying 1 to 23 of 119 items

Annotation distribution across subcorpora



Open chart as a PNG file

# Narzędzia analityczne w Inforexie

Annotation browser

Annotation stage

405623	final
917	discarded
516842	agreement

Annotation types

6292	nam_liv_person
950	nam_liv_person_add
5265	nam_liv_person_first
4572	nam_liv_person_last
12	nam_liv_plant
61	nam_loc
1	nam_loc_amb
115	nam_loc_astronomical
97	nam_loc_countryRegion
1	nam_loc_gpe
7	nam_loc_gpe_admin
304	nam_loc_gpe_admin1
142	nam_loc_gpe_admin2
263	nam_loc_gpe_admin3

Text forms

39	Jezus
24	Adam
21	Stefan Niesiolowski
17	Jerzy Szmajdziński
17	Kasia
16	Karol
15	Ewa Kierzkowska
15	Jarek
15	Magda

Lemmas

15	Jarosław Kaczyński
15	Piotrek
14	Krzysztof Putra
14	Maciolus
14	Marek
14	Tomek
13	Chopin
13	Ela
13	Keraban

Contexts

Document	Left	Annotation	Right	Source	Stage
101563	Trzeba wrócić do dobrych projektów opracowanych przez rząd pana	Jarosława Kaczyńskiego	, przez panią minister Gęsicką, które znalazły się na liście	user	final
101591	ataki na schorowanego generała Januzelskiego. To przeciwie	Jarosław Kaczyński	porównał generała do Adolfa Eichmanna. Tymczasem dla wielu	user	final
107255	2006-01-21: Nowa stabilizacyjna propozycja PiS-u	Jarosław Kaczyński	kuje do wszystkich partii z wyjątkiem SLD propozycję stworzenia	user	final
107255	stabilizującego zaplecze rządu w Sejmie". Jak podkreślił	Jarosław Kaczyński	, formacja ta powstałaby po wypełnieniu założeń wspomnianego	user	final
107256	Platformą Obywatelską. Prawdopodobnie jednak bez udziału	Jarosława Kaczyńskiego	i Jana Rokity. Rozmowy zapowiedział wczoraj publicznie Kaczyński	user	final
107258	2006-01-28:	Jarosław Kaczyński	: Koalicja rządowa z Samoobroną i LPR na razie niemożliwa Jarosław	user	final
107258	Kaczyński: Koalicja rządowa z Samoobroną i LPR na razie niemożliwa	Jarosław Kaczyński	powiedział w Sygnale Dnia, że koalicja rządowa z Samoobroną	user	final
107258	trzeba będzie odwołać się do demokracji - zaznacza premier.	Jarosław Kaczyński	stwierdził także, że jeszcze wczoraj była szansa na porozumienie	user	final
107268	2006-11-24: Kaczyński: Rosja nie może nas tak traktować Premier	Jarosław Kaczyński	powiedział, że Rosja nie może prowadzić rozmów z Polską tak,	user	final
107272	potwierdzają inne pracownice Samoobrony. Po rozmowie z premierem	Jarosławem Kaczyńskim	prokurator krajowy Janusz Kaczmarek postanowił wszcząć postępowanie	user	final
107296	Lublinie Podczas dzisiejszej konwencji wyborczej PiS w Lublinie	Jarosław Kaczyński	przekonywał słuchaczy, że według politycznych przeciwników	user	final
107296	innego jak tylko zrzucenie tego worka" - powiedział premier.	Jarosław Kaczyński	stwierdził, że aktualny spór w kraju dotyczy jednego zagadnienia:	user	final
107303	prowadził bardziej zdecydowanie, niż jego poprzednicy w rządzie	Jarosława Kaczyńskiego	i że nie zrezygnuje z pewnych warunków, które uwzględniają	user	final
107319	posłaniu o rok młodszych dzieci do podstawówek od 2009 r. partia	Jarosława Kaczyńskiego	może zyskać przychylność wyborców. W głównej mierze rodziców,	user	final
107346	dopuszczalny limit emisji CO2 na 208,5 mln ton. Ówczesny rząd	Jarosława Kaczyńskiego	obliczył, że potrzeby polskiej gospodarki sięgają rocznego	user	final

Page 1 of 1 | Displaying 1 to 15 of 15 items

Export selected annotations to CSV

# Cytowania

- Marcińczuk, M., Oleksy, M., & Kocoń, J. (2017). Inforex—a collaborative system for text corpora annotation and analysis. In Proceedings of the International Conference Recent Advances in Natural Language Processing, RANLP. (pp. 473-482);
- Marcińczuk, M., & Oleksy, M. (2019). Inforex—a collaborative system for text corpora annotation and analysis goes open. In Proceedings of the International Conference on Recent Advances in Natural Language Processing (RANLP 2019) (pp. 711-719).

## Finansowanie

2025-2027

w ramach Funduszy Europejskich dla Nowoczesnej Gospodarki 2021-2027, priorytet 2. „Środowisko sprzyjające innowacjom”, nabór „FENG.02.04-IP.04-001/24 Badawcza Infrastruktura Nowoczesnej Gospodarki – ścieżka dla projektów z PMIB

2024-2026

w ramach projektu pt.: „Wniesienie wkładu własnego do europejskiej infrastruktury badawczej “CLARIN ERIC – European Research Infrastructure Consortium: Common Language Resources and Technology Infrastructure - Wspólne Zasoby Językowe i Infrastruktura Technologiczna” na lata 2024-2026” dofinansowanego przez Ministra Nauki na podstawie umowy nr 2024/WK/01 w ramach programu „Wsparcie udziału polskich zespołów naukowych w międzynarodowych projektach infrastruktury badawczej.

2020-2023

w ramach Programu Operacyjnego Inteligentny Rozwój 2014-2020, Priorytet IV: Zwiększenie potencjału naukowo-badawczego, Działanie 4.2: Rozwój nowoczesnej infrastruktury badawczej sektora nauki, nr POIR.04.02.00-00C002/19, CLARIN - Wspólne zasoby językowe i infrastruktura technologiczna”.