



Instrukcja użytkownika

[Podstawowe informacje](#)

[Miejsce dostępu](#)

[Do czego służy Inforex?](#)

[Podstawowe funkcje](#)

[Dla kogo przeznaczone jest narzędzie](#)

[Przykłady zastosowań](#)

[Instrukcja obsługi](#)

[Inforex - podstawowe obszary](#)

[Panel sterowania](#)

[Ustawienia korpusu](#)

[Podstawowe informacje dotyczące korpusu](#)

[Użytkownicy](#)

[Uprawnienia użytkowników korpusu](#)

[Podkorpusy](#)

[Perspektywy](#)

[Flagi](#)

[Dodawanie nowej flagi](#)

[Zestawy kategorii](#)

[Własne zestawy kategorii](#)

[Usuwanie korpusu](#)

[Oznaczanie dokumentów korpusie](#)

[Widok listy dokumentów:](#)

[Widok zawartości dokumentu](#)

[Usuwanie dokumentu:](#)

[Perspektywy](#)

[Preview](#)

[Edycja metadanych dokumentów](#)

[Anotowanie korpusu](#)

[Annotator - znakowanie ręczne](#)

[Przygotowanie - wybór zestawu kategorii](#)

[Wybór trybu znakowania](#)

[Wprowadzanie anotacji](#)

[Flagowanie dokumentu](#)

[Edytowanie anotacji](#)

[Usuwanie anotacji](#)

[Bootstrapping - weryfikacja automatycznej anotacji](#)

[Wybór zestawu kategorii](#)

[Weryfikacja](#)

[Przeglądanie anotacji po weryfikacji](#)

[Przeglądanie anotacji](#)

[Filtry w sekcji Documents](#)

[Annotations:](#)

[Przeglądanie anotacji w kontekście w sekcji Annotation browser](#)

[Eksportowanie wybranych anotacji do CSV](#)

[Annotation frequency](#)

[Statystyki](#)

[Words frequency](#)

[Menadżer zadań](#)

[Dodawanie dokumentów do korpusu](#)

[Przesyłanie plików](#)

Podstawowe informacje

Miejsce dostępu

Inforex nie wymaga instalacji na urządzeniu użytkownika, nie potrzebuje dodatkowego oprogramowania, jest aplikacją webową, do której dostęp mamy bezpośrednio przez przeglądarkę internetową:

- bezpośredni adres: <https://inforex-work.clarin-pl.eu/>

Do czego służy Inforex?

Podstawowe funkcje

- Przy pomocy systemu Inforex możemy zarządzać, przeszukiwać i analizować zawartość wybranych korpusów. Pozwala on użytkownikom nie tylko na wyliczanie podstawowych statystyk i list frekwencyjnych, ale przede wszystkim umożliwia tworzenie jakościowych danych językowych, które posłużyć mogą jako podstawa do pogłębionej i usystematyzowanej analizy materiału badawczego.
- System umożliwia wprowadzanie różnorodnych poziomów opisu (anotacji), zarówno tych dodawanych ręcznie przez użytkowników, jak i automatycznych, dodawanych z wykorzystaniem wbudowanych w system narzędzi do rozpoznawania określonych kategorii językowych (m.in. nazw własnych, wyrażen temporalnych).
- Ze względu na to, że dane przechowywane są na bezpiecznym serwerze, różne osoby zaangażowane w prace badawcze mogą współpracować ze sobą zdalnie, mając wgląd we wzajemne postępy pracy. Inforex daje możliwość wygodnego przeglądania i porównywania anotacji przeprowadzonej przez różne osoby pracujące nad danym korpusem (badanie zgodności). Ułatwia i usprawnia kontrolę nad postępami pracy przez czytelny system flagowania (oznaczania) opracowywanej zawartości.

Dla kogo przeznaczone jest narzędzie

- Inforex ułatwia jakościowe opracowanie danych, które mogą posłużyć jako materiał do badań korpusowych na gruncie językoznawstwa, a także badań socjologicznych, historycznych, medioznawczych, analizy dyskursu i procesów komunikacyjnych. Analiza wybranych cech lingwistycznych pozwala uchwycić w badanym materiale prawidłowości, które mogą się stać podstawą dla stworzenia bardziej uniwersalnych modeli.

Przykłady zastosowań

Korpus Politechniki Wrocławskiej (KPWr)

Otwarty (udostępniany na licencji Creative Commons) korpus języka polskiego o wielowarstwowej anotacji. Zawiera dokumenty tekstowe reprezentujące różne dyskursy, które pochodzą z różnorodnych źródeł m.in. z: blogów, artykułów i publikacji prasowych, prozy dawnej i współczesnej, tekstów religijnych, urzędowych, stenogramów sejmowych. KPWr zawiera około 450.000 słów, a prace nad jego powiększeniem wciąż trwają. Teksty wchodzące w skład korpusu zostały otagowane zarówno przy pomocy automatycznych narzędzi, jak i ręcznej, zespołowej anotacji, różnymi typami informacji, określającymi m.in.: jednostki identyfikacyjne, wyrażenia temporalne, wyrażenia przestrzenne czy słowa kluczowe.

Polski Korpus Listów Pożegnalnych (PKLP) [UWr, M. Zaśko-Zielińska]

Strona projektu: <http://www.pcsn.uni.wroc.pl/>

Publikacja: M. Zaśko-Zielińska, *Listy pożegnalne. W poszukiwaniu wyznaczników autentyczności tekstu*, Wrocław 2013.

PKLP jest specjalistycznym korpusem jednogatunkowym ze zbiorem metadanych autorów i szczegółową anotacją struktury tekstu, anotacją błędów, anotacją treści (w tym nazw własnych) oraz anotacją pragmatyczną. Korpus umożliwił opis wzorca maksymalnego gatunku. Stanowi rodzaj materiału porównawczego, który pozwolił oddzielić typowe cechy listu pożegnalnego od cech osobniczych autora. Badania wyłoniły różnice między autentycznymi listami pożegnalnymi a listami sfalszowanymi.

Korpus dokumentów dot. polityki terytorialnej UE [PWr, Ł. Damurski]

- listy frekwencyjne,
- znakowanie kategoriami związanymi z polityką terytorialną

Korpus tekstów hebrajskich [UWr, T. Bernaś]

- znakowanie pojęciami i ich tłumaczenie

Korpus konwersacji z Facebooka [UWr, E. Kaczmarz]

- statystyki błędów językowych i korelacje między nimi

Instrukcja obsługi

Inforex - podstawowe obszary

panel górny

korpora publiczne: dostępne dla wszystkich użytkowników

ID	Name	Description	Documents
5	CEN	Corpus of economic news from Polish Wikinews (from 17 September 2007 to 24 May 2010).	797
14	Demo (English)	Sample corpus with English news from Wikinews.	1
1061	Diachrono	próbki diachroniczne	0
1	GPW	Korpus raportów giełdowych.	99031
1156	Korpus_Testowy	Corpus imported from DSpace	10
4	Poligon	Korpus do eksperymentowania i testowania	5
15	Porównanie narzędzi PWi i IPI	Zbiór testów do porównania narzędzi PWi i IPI (chunkowanie, rozpoznawanie nazw własnych)	20

ID	Name	Description	Documents	Owner
1076	Diachrono	Corpus imported from DSpace	20	Aleksandra
37	Korpus	Corpus imported from DSpace	106	Marcin
1150	Korpus testowy	Corpus imported from DSpace	17	Aleksandra

korpora prywatne: dostęp do nich jest autoryzowany

This page was tested in FireFox Page generator in 0 seconds.

Involved in development: Michał Marończuk, Adam Kaczmarek, Jan Kocot, Marcin Płak, Mikolaj Szwedczyk, Marcin Olekny
Grupa Technologii Językowych G4 15 Politechniki Wroclawskiej, 2009-2017

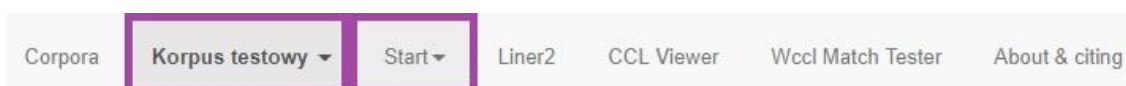
Panel górny:



Panel górny umożliwia dostęp do podstawowych zakładek systemu:

- Corpora - zestaw dostępnych korpusów
 - Public corpora** - korpusy tekstowe udostępnione publicznie, z których korzystać może każdy użytkownik Inforexa.
 - Private corpora** - lista korpusów prywatnych (udostępnianie korpusu innym użytkownikom: [Użytkownicy](#)).
- CCL Viewer - przeglądarka plików w formacie CCL.
- Wccl Match Tester - testowanie reguł dopasowania sekwencji tokenów i/lub anotacji w obrębie zdania.
- About & citing - informacje o systemie, materiały warsztatowe (prezentacja, film instruktażowy, plansze informacyjne), wskazówki do cytowania.

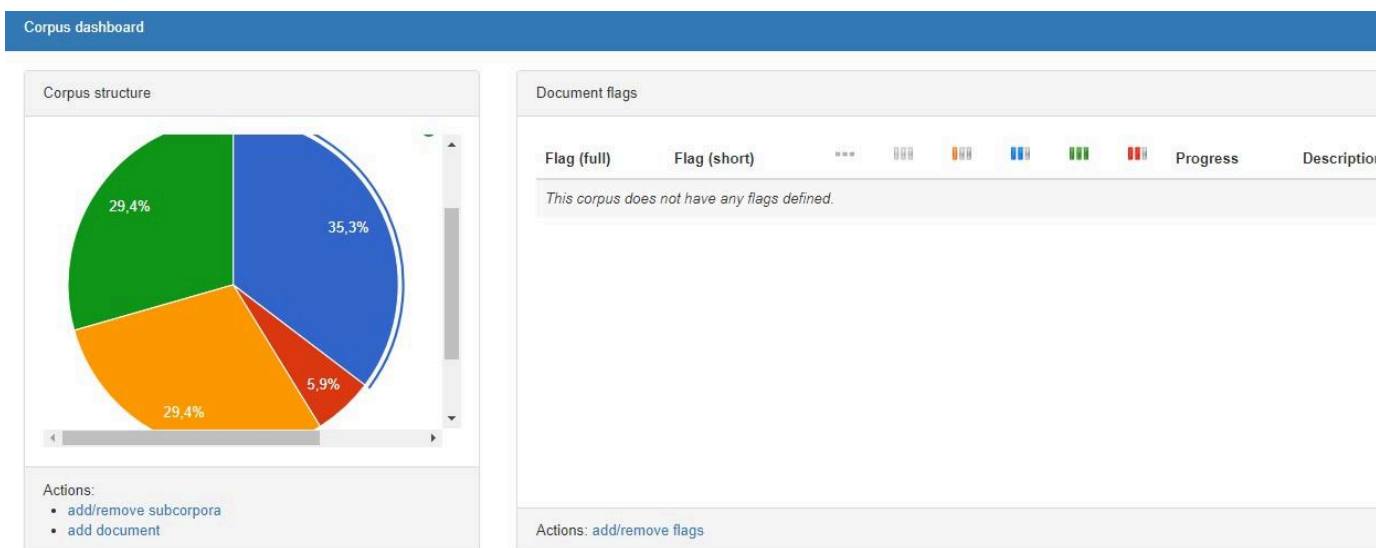
Po wybraniu dowolnego korpusu panel górny zostanie uzupełniony o dwie dodatkowe zakładki:



- Nazwa korpusu▼ - z rozwijanej listy możemy szybko przejść do innego korpusu.
- Start▼ - lista rozwijana opcji umożliwiających przeglądanie, anotowanie korpusu i zarządzanie nim.

Panel sterowania

Sekcja Start - panel sterowania korpusem. Z tego miejsca użytkownik zyska dostęp do podstawowych informacji o korpusie, zwizualizowanych za pomocą graficznych wskaźników. Wszystkie najważniejsze dane, dotyczące struktury korpusu i postępów nad opisem jego zawartości, zawarte są w obrębie jednego ekranu, którego ważną częścią jest także ułatwiony dostęp do podstawowych opcji systemu.

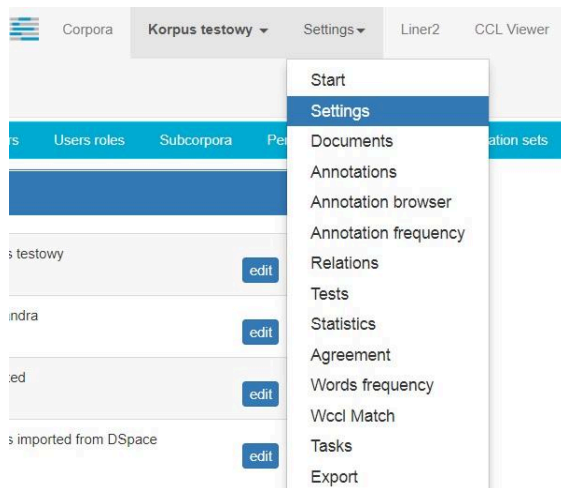


Obszary:

- Corpus structure - wizualizacja struktury korpusu, pokazuje ilościowy udział zawartości poszczególnych podkorpusów (o ile zostały wyodrębnione, więcej: [Podkorpusy](#)).
- Actions - szybki dostęp z poziomu pulpitu do podstawowych opcji:
 - [add/remove subcorpora](#) - po wybraniu tej opcji użytkownik zostanie przeniesiony do panelu zarządzania strukturą korpusu, gdzie będzie mógł dodać lub usunąć podkorpus (więcej: [Podkorpusy](#)).
 - [add document](#) - przenosi użytkownika do panelu tworzenia dokumentów (więcej: [Dodawanie dokumentów](#)).

- Document flags - widok postępów anotacji dokumentów oznaczonych flagami. Flagi umożliwiają szybkie oznaczanie danych partii materiału tekstowego (materiał gotowy do pracy/skończony/nieskończony), tak, by były one czytelne dla innych współpracowników (więcej: [Flagi](#)).
- Actions: [add/remove flags](#) - szybki dostęp do opcji dodawania/usuwania flagi.

Ustawienia korpusu



Użytkownik może dostosowywać zawartość korpusu, udostępnić go innym użytkownikom, a także określić zakres działań, jaki mogą na nim przeprowadzić. Te i szereg innych opcji dostępne są w zakładce Settings:



Widok na pasek zakładek w sekcji ustawień

Basic information

Podstawowe informacje dotyczące korpusu

Użytkownik ma wgląd i możliwość edycji ogólnych informacji dotyczących korpusu. Może zmienić jego nazwę, rodzaj dostępu (z otwartego na ograniczony), przypisać do niego innego właściciela lub zmienić jego opis. Każdą z tych operacji zatwierdza przycisk [Confirm](#).

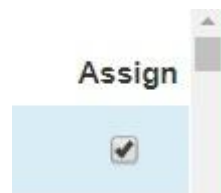
Users

Użytkownicy

W tym miejscu możemy wybrać spośród zarejestrowanych użytkowników tych, którym korpus zostanie udostępniony i którzy będą mogli wykonywać określone działania (szczegółowe informacje dotyczące dostępu znajdziemy w zakładce [Users roles](#)). Znajdujemy właściwych użytkowników z listy (nazwa i adres e-mail), a następnie włączamy ich do pracy nad korpusem poprzez zatwierdzenie okna wyboru. Aby cofnąć udostępnienie korpusu wybranemu użytkownikowi, wystarczy ponownie zatwierdzić pole wyboru.

User	Login	Assign
	stasiek_bond@wp.pl	<input type="checkbox"/>
	wspoloch.gawron@wp.pl	<input type="checkbox"/>
	tomasz.walczak@wp.pl	<input type="checkbox"/>
	marcin.zawadzki@wp.pl	<input type="checkbox"/>
		<input type="checkbox"/>
A	sczytko.alex@gmail.com	<input type="checkbox"/>
Adam	adam.zacharycz@gmail.com	<input type="checkbox"/>
Adam Pawlaczek	adam.pawlaczek	<input type="checkbox"/>
Adam Radziszewski	aradzi	<input checked="" type="checkbox"/>
Adam Warotyński	awarot	<input type="checkbox"/>
Agata	agata@wp.pl	<input type="checkbox"/>
agata	agata@wp.pl	<input type="checkbox"/>
Agata	agata@technet.com	<input type="checkbox"/>
Agnieszka	agnieszka.mielga@wp.pl	<input type="checkbox"/>
Agnieszka	a_gonka_wp@wp.pl	<input type="checkbox"/>
Agnieszka	agnieszka.pawlaczek@gmail.com	<input type="checkbox"/>
Agnieszka	agnieszka.nemczak@gmail.com	<input type="checkbox"/>
Agnieszka	agnieszka714@gmail.com	<input type="checkbox"/>
Agnieszka	nikolagnieszka@gmail.com	<input type="checkbox"/>
Agnieszka	agnieszka_kaczmarek@wp.pl	<input type="checkbox"/>

Dodanie użytkownika należy potwierdzić w kolumnie **Assign**.



Widok na zarejestrowanych użytkowników Inforexa w kolejności alfabetycznej - wybieramy spośród nich tych, którym chcemy udostępnić korpus.

Users roles

Uprawnienia użytkowników korpusu

W tej zakładce możemy edytować uprawnienia poszczególnych użytkowników zaangażowanych w prace nad danym korpusem (m.in. ustalić kto może dodawać i usuwać dokumenty, jaki ma dostęp do edycji anotacji, przypisać uprawnienia administrowania korpusem, itp.). Aby wprowadzić zmiany, wystarczy zatwierdzić pole wyboru pod odpowiednią opcją.

	Dodawanie dokumentów	Sprawdzanie zgodności anotatorów	Edycja anotacji publicznych	Edycja anotacji prywatnych	Dostęp do listy anotacji	Dostęp do listy relacji	Prawo do usuwania anotacji	Usuwanie dokumentów
	[add_documents]	[agreement_check]	[annotate]	[annotate_agreement]	[browse_annotations]	[browse_relations]	[delete_annotations]	[delete_documents]
Adam Radziwiłowicz	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Agnieszka Dziub	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Agnieszka Turak	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Andrzej Bruchowski	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Michał Rywik	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
olam.pl	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Diogenes Kalkus	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Dariusz Kucielny	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
idm	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
dominka	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Dominka Film	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Filip Marzec	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Gregorz Zaręczyński	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
IT-PL	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Jan Kowal	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Jan Włoczek	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Joanna Jaszczyńska	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Justyna Łaszczyk	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

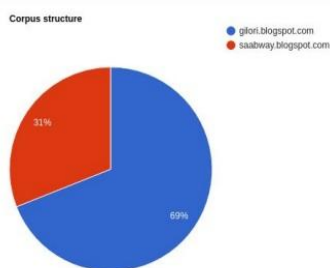
Subcorpora Podkorporusy

Strukturę korpusu możemy w każdej chwili zmienić, dodając, zmieniając lub usuwając poszczególne podkorporusy. W tym celu wybieramy odpowiednią opcję spośród trzech możliwych:

Create Edit Delete

Uwaga! Usunięcie podkorporusu spowoduje **usunięcie wszystkich zawartych w nim dokumentów!**

Aby tego uniknąć, należy **przypisać wybrane dokumenty do innego podkorporusu**, zmieniając ich metadane: Documents → Metadata → Subcorpus, więcej: [przypisywanie tekstów do podkorporusów](#).



Nowy podkorpus będzie uwzględniony w wykresie kołowym w naszym panelu sterowania (zakładka **Start**), jednak dopiero wtedy, gdy dodamy do niego dokumenty; więcej → [przypisywanie dokumentów](#) a także → [dodawanie dokumentów](#).

Perspectives Perspektywy

W tym miejscu użytkownik może określić “perspektywy” czyli zakres dostępnych opcji, które będą widoczne dla niego i poszczególnych użytkowników w sekcji Documents. Szczegółowo poszczególne opcje zostały opisane w dalszej części instrukcji: [Perspektywy](#).

active	title	description	access
<input type="checkbox"/>	Agreement		loggedin ▼
<input type="checkbox"/>	Anaphora	Show and edit anaphora relations.	loggedin ▼
<input type="checkbox"/>	Anaphora Viewer	Show anaphora relations.	loggedin ▼
<input type="checkbox"/>	Annotation lemmas	Allows to edit lemmas for existing annotations.	loggedin ▼
<input checked="" type="checkbox"/>	Annotator	Show and edit document annotations, relations and events.	loggedin ▼
<input checked="" type="checkbox"/>	Bootstrapping	Verify automatically added annotations.	loggedin ▼
<input checked="" type="checkbox"/>	Content	Show and edit document content.	loggedin ▼
<input type="checkbox"/>	Content Viewer	Display document content with hidden sensitive information.	loggedin ▼
<input type="checkbox"/>	History of changes	Show history of document content changes.	loggedin ▼
<input type="checkbox"/>	Images	Manage images attached to the document.	loggedin ▼
<input type="checkbox"/>	Metadata	Show and edit document metadata.	loggedin ▼

[Confirm](#)

Perspektywy przeglądania i edycji dokumentów, spośród których możemy wybierać:

- **Preview** - podgląd treści;

- **Metadata** - metadane dokumentów;
- **Annotator** - anotacja dokumentów;
- **Bootstrapping** - weryfikacja automatycznej anotacji;
- **Content** - zawartość dokumentu z możliwością edycji.

Agreement

Anaphora

Anaphora Viewer

Annotation lemmas

Content Viewer

History of changes

Images

Tokenization

Topic

Transcription

WSD

Flags

Flagi

Flagowanie korpusu ułatwia współpracę między użytkownikami, pozwala nie tylko oznaczać stopień zaawansowania pracy nad daną partią materiału, ale jest też pomocne przy badaniu zgodności anotatorów (więcej: agreement).

Dodawanie nowej flagi

- Sposób 1: Z sekcji ustawień **Settings** wybieramy zakładkę **Flags**. Następnie wybieramy przycisk **New** i tworzymy nową flagę. Podajemy nazwę, nazwę skróconą i opis nowego oznaczenia. Zatwierdzamy **Confirm**.
- Sposób 2: W panelu startowym sekcji **Start** znajdujemy opcję **add/remove flags**, która

The screenshot shows a 'Create flag' dialog box with the following fields and values:

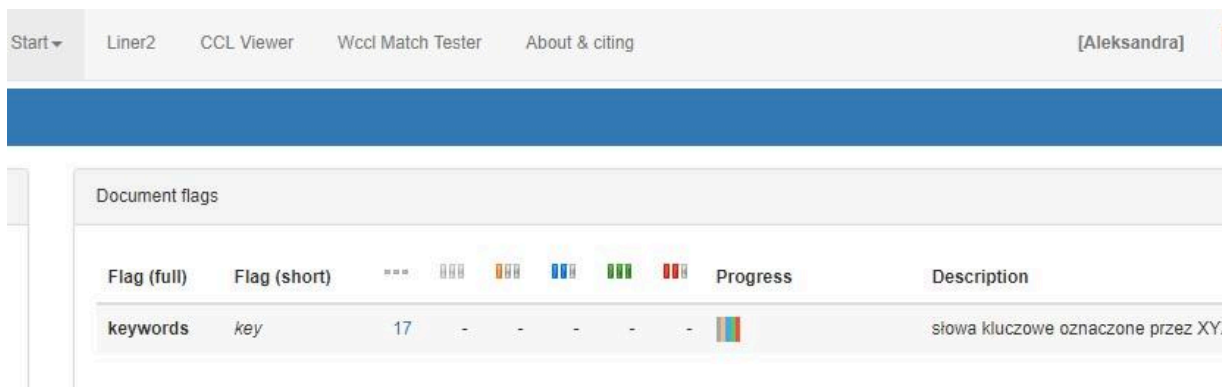
- Name: keywords
- Short: key
- Description: słowa kluczowe oznaczone przez XYZ
- Sort: 1

Buttons: 'New' (bottom left), 'Confirm' (bottom right).

przenosi nas do ustawień dotyczących flagowania korpusu. Dalej postępujemy tak, jak w sposobie pierwszym.

Utworzoną flagę możemy edytować, wybierając opcję **Edit** lub usunąć **Delete**.

Po dodaniu przynajmniej jednej flagi zmieni się wygląd panelu startowego:



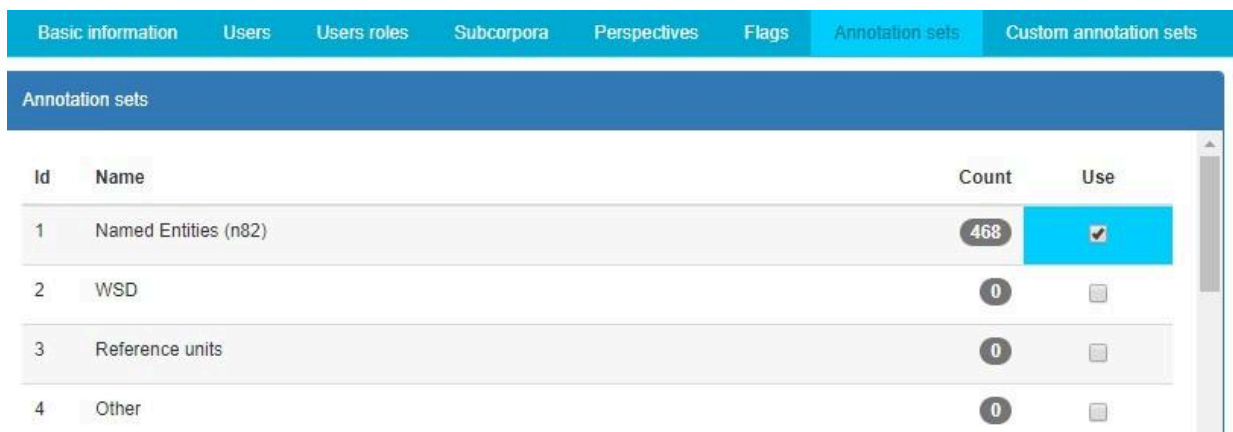
Dodawanie oznaczeń w obrębie konkretnego dokumentu w korpusie zostało opisane w części:

[Flagowanie dokumentu](#)

Annotation sets

Zestawy kategorii

W tym miejscu możemy zdecydować, które spośród gotowych (uprzednio zdefiniowanych) zestawów kategorii chcemy wykorzystać do oznaczania dokumentów w korpusie. Wybór oznaczamy w kolumnie **Use** przy wybranej pozycji z listy . Aby sprawdzić, jakie dokładnie kategorie wchodzą w skład wybranego zestawu, należy przejść do sekcji **Annotations** (więcej: [Annotations](#)).



Custom annotation sets

Własne zestawy kategorii

Użytkownik może tworzyć własne zestawy kategorii, które wykorzysta przy oznaczaniu dokumentów. W tym celu należy wybrać przycisk **Create** w zakładce **Custom annotation sets**. Operację zatwierdzamy przyciskiem **Confirm**.

Gdy stworzymy nowy zestaw kategorii, możemy dodać do niego podzestawy i kategorie szczegółowe (np. Zdrobnienia → Nazwy Własne → Imiona Męskie), a także przypisać im odpowiednie oznaczenia kolorystyczne i nazwę skróconą, które będą widoczne w sekcji wprowadzania anotacji (Documents/Annotator).

Uwaga: przy podawaniu nazwy symbolicznej dla kategorii (symbolic name) nie możemy używać polskich znaków ani spacji (zastępujemy ją znakiem_ podkreślenia).

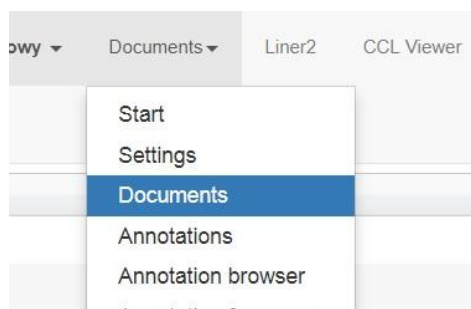
Annotation sets				Annotation subsets		Categories			
Id	Name	Owner	Access	Id	Name	Symbolic name	Display name	Description	Default visibility
63	Zdrobnienia	Aleksandra	public	134	Nazwy pospolite	imiona_meskie	męskie		Visible
				133	Nazwy własne	imiona_zenskie	żeńskie		Visible

Delete corpora

Usuwanie korpusu

W tej zakładce możemy usunąć korpus. Po wybraniu opcji **Delete corpora** przestanie on być dostępny dla użytkownika w sekcji "Private corpora". Usunięte zostaną wszystkie oznaczenia, flagi, anotacje, a użytkownicy, którym korpus został udostępniony, również stracą do niego dostęp. Usunięcie korpusu jest **nieodwracalne**.

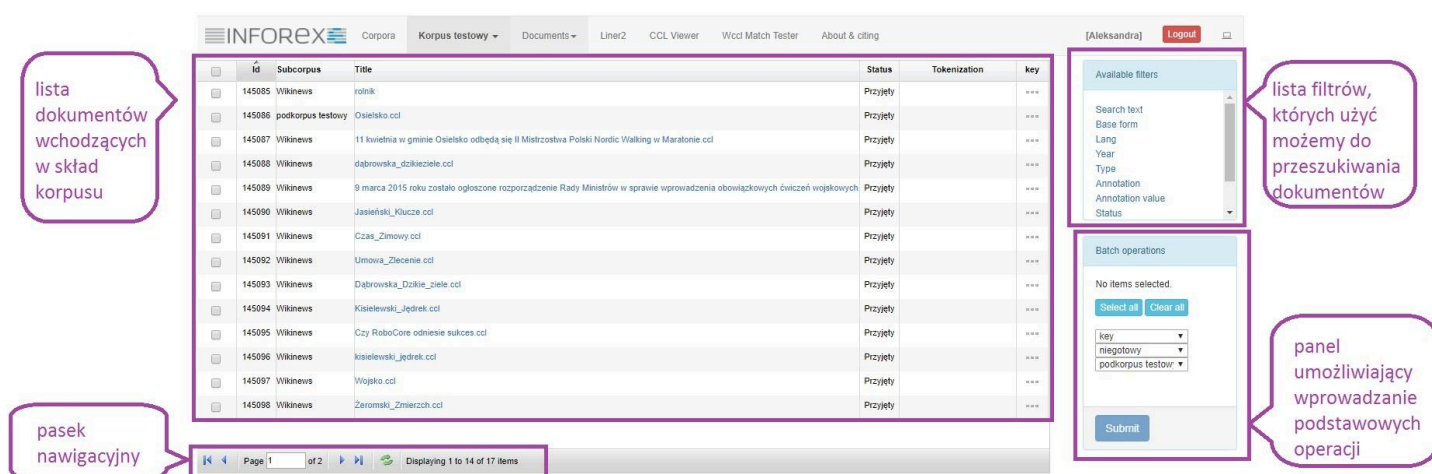
Oznaczanie dokumentów w korpusie



Użytkownik Inforexa może przeglądać dokumenty, które wchodziły w skład korpusu. Aby to zrobić, należy przejść do sekcji Documents.

Widok listy dokumentów:

Główną część sekcji zajmuje lista dokumentów, które wchodziły w skład korpusu. Do każdego dokumentu przypisana jest liczba porządkowa (Id). W tym miejscu możemy sprawdzić, do którego podkorpusu przypisany jest wybrany dokument, a także jaki jest status jego opracowania.

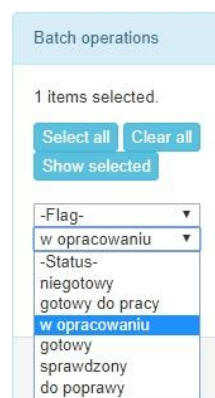


The screenshot shows the Inforex application interface. The main part is a table of documents. To the right, there are panels for filters and batch operations. Callouts point to specific elements:

- lista dokumentów wchodzących w skład korpusu (points to the document table)
- lista filtrów, których użyć możemy do przeszukiwania dokumentów (points to the Available filters panel)
- panel umożliwiający wprowadzanie podstawowych operacji (points to the Batch operations panel)
- pasek nawigacyjny (points to the navigation bar at the bottom)

id	Subcorpus	Title	Status	Tokenization	key
145085	Wikinevs	rolnik	Przyjęty	---	---
145086	podkorpus testowy	Osielsko.ccl	Przyjęty	---	---
145087	Wikinevs	11 kwietnia w gminie Osielsko odbędą się II Mistrzostwa Polski Nordic Walking w Maratonie.ccl	Przyjęty	---	---
145088	Wikinevs	dąbrowska_dzikieziele.ccl	Przyjęty	---	---
145089	Wikinevs	9 marca 2015 roku zostało ogłoszone rozporządzenie Rady Ministrów w sprawie wprowadzenia obowiązkowych ćwiczeń wojskowych	Przyjęty	---	---
145090	Wikinevs	Jasieński_Klucze.ccl	Przyjęty	---	---
145091	Wikinevs	Czas_Zimowy.ccl	Przyjęty	---	---
145092	Wikinevs	Umowa_Zeczenie.ccl	Przyjęty	---	---
145093	Wikinevs	Dąbrowska_Dzikie_ziele.ccl	Przyjęty	---	---
145094	Wikinevs	Kisielewski_Jedrek.ccl	Przyjęty	---	---
145095	Wikinevs	Czy RoboCore odniesie sukces.ccl	Przyjęty	---	---
145096	Wikinevs	kisielewski_jedrek.ccl	Przyjęty	---	---
145097	Wikinevs	Wojtko.ccl	Przyjęty	---	---
145098	Wikinevs	Zaromski_Zmierzh.ccl	Przyjęty	---	---

Informacje te możemy w prosty sposób zmienić za pomocą **panelu** znajdującego się w prawym dolnym rogu sekcji. Umożliwia on **przypisanie wybranego z listy dokumentu do innego podkorpusu**, a także zmianę jego statusu (na np. "w opracowaniu") i oflagowania. W celu wprowadzenia zmian zaznaczamy pole wyboru: przy wybranym dokumencie z listy i zatwierdzamy zmiany w dolnym panelu. Dzięki opcjom **Select all** **Clear all**, a



The image shows a close-up of the 'Batch operations' panel. It displays '1 items selected.' and buttons for 'Select all', 'Clear all', and 'Show selected'. Below these is a dropdown menu for '-Flag-' with options: 'w opracowaniu', '-Status-', 'niegotowy', 'gotowy do pracy', 'w opracowaniu' (highlighted), 'gotowy', 'sprawdzony', and 'do poprawy'.

także możliwości filtrowania możemy szybko zmienić informacje dla dużej liczby dokumentów. Operacje te potwierdzamy przyciskiem **Submit**.

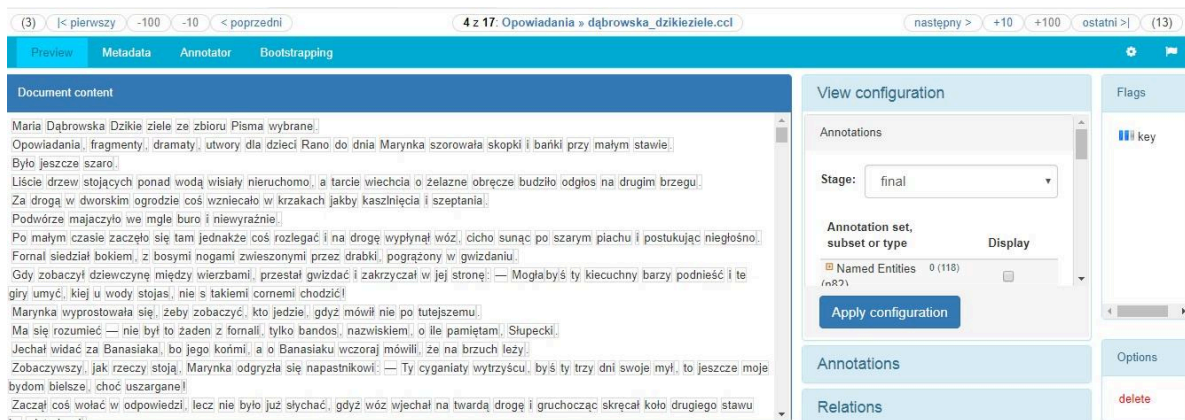


Pomiędzy kolejnymi listami dokumentów przemieszczamy się za pomocą przycisków na **pasku nawigacyjnym**, znajdującym się u dołu ekranu.

W sekcji **Documents** użytkownik może przeszukiwać dokumenty wchodzące w skład korpusu za pomocą filtrów dostępnych w panelu **Available filters**, który znajduje się po prawej stronie. Należy wybrać jedną z dostępnych kategorii wyszukiwania lub wprowadzić szukany tekst (opcje: Search text, Base form). Różne filtry wyszukiwania mogą się na siebie nakładać. Aby je usunąć, klikamy **cancel**. Zobacz: [Przeglądanie anotacji za pomocą filtrów w sekcji Documents](#)

Widok zawartości dokumentu

Aby wyświetlić zawartość dokumentu, wystarczy kliknąć wybrany plik z listy dokumentów w sekcji **Documents**. Uzyskamy w ten sposób dostęp nie tylko do jego treści, ale także do panelu konfiguracji opcji oznaczania:



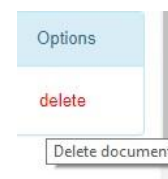
Treść dokumentu wyświetla się z podziałem na **tokens** - fragmenty tekstu w jasnoszarej ramce (najmniejsze jednostki w tekście, które można anotować). Możemy wprowadzić dodatkową opcję wyodrębnienia poszczególnych zdań (View configuration / Display every sentence separately).

Na górze strony znajduje się pasek nawigacyjny, za pomocą którego możemy sprawnie przemieszczać się pomiędzy poszczególnymi dokumentami:



Usuwanie dokumentu:

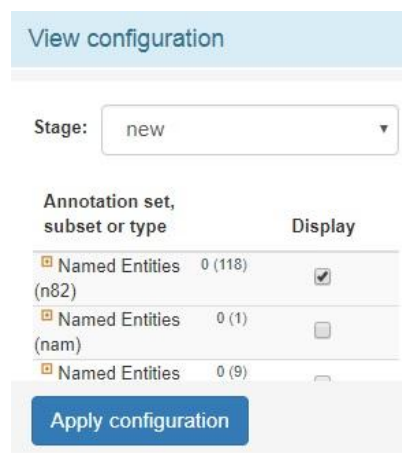
Konkretny dokument możemy usunąć z korpusu, wybierając opcję **delete** w prawym dolnym rogu.



Perspektywy

Preview

Podgląd treści dokumentu. Tu użytkownik może wprowadzić ustawienia dotyczące widoku dokumentu. W panelu **View configuration** poprzez zaznaczenie pola wyboru możemy określić, które zestawy kategorii mają być wyświetlane, a także zmienić status opracowania dokumentu. Wybór podzestawów i konkretnych kategorii jest możliwy po kliknięciu znaku **[+]** przy nazwie wybranego zestawu anotacji. W ten sam sposób wybieramy, które relacje między anotacjami mają być widoczne. Wszystkie wprowadzone ustawienia należy potwierdzić przyciskiem **Apply configuration**.



Edycja metadanych dokumentów

Metadata W sekcji metadanych użytkownik może przypisać wybranemu dokumentowi różne informacje. Zmienić możemy jego tytuł, autora, źródło, przypisać tekst **do innego podkorpusu**, edytować status opracowania (przyjęty, załącznik, odrzucony, odłożony, do usunięcia), wprowadzić datę (wydania, opublikowania lub utworzenia), a także zmienić format (prosty format tekstowy, xml, zapis premorph). Możemy także wprowadzić wcześniej zdefiniowane przez użytkownika metadane (Custom metadata) (więcej: sekcja Settings/Metadata). Wszystkie zmiany zapisujemy, klikając **Save**.

Common metadata	
Title	11 kwietnia w gminie Osielsko odbędą się II Mistrzostwa Polski Nordic Walking w Maratonie.ccl
Author	dspace
Source	dspace
Subcorpus	Wikinews
Status	Przyjęty
Date	0000-00-00 released, published or created
Format	plain

Custom metadata	
-----------------	--

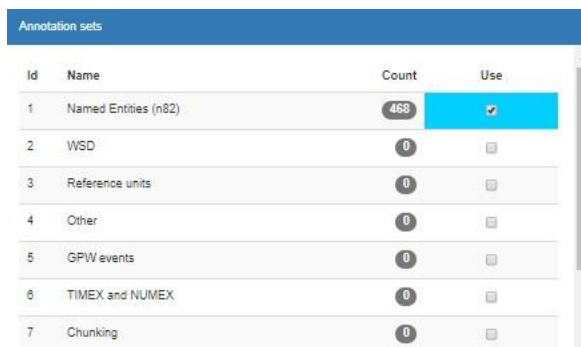
Anotowanie korpusu

sekcja Documents / dokument z listy/ zakładka Annotator

Annotator - znakowanie ręczne

Przygotowanie - wybór zestawu kategorii

W sekcji ustawień **Settings** w zakładce **Annotation sets** wybieramy zestaw kategorii, który chcemy wykorzystać przy znakowaniu korpusu, zaznaczając pole wyboru w tabeli. Wybieramy nie tylko z gotowych zestawów anotacji, ale także spośród uprzednio utworzonych samodzielnie w zakładce **Custom annotation sets**. W każdej chwili możemy zrezygnować z wybranych zestawów (ponownie klikając).



Id	Name	Count	Use
1	Named Entities (n82)	468	<input checked="" type="checkbox"/>
2	WSD	0	<input type="checkbox"/>
3	Reference units	0	<input type="checkbox"/>
4	Other	0	<input type="checkbox"/>
5	GPIW events	0	<input type="checkbox"/>
6	TIMEX and NUMEX	0	<input type="checkbox"/>
7	Chunking	0	<input type="checkbox"/>

Uwaga! Odłączenie warstwy nie spowoduje usunięcia wcześniejszych oznaczeń tekstu wprowadzonych w ramach danej warstwy.

Wybór trybu znakowania

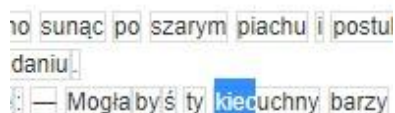
W panelu **Working mode** możemy określić tryb pracy nad danym dokumentem, a więc to, czy będzie on widoczny dla wszystkich osób, którym został udostępniony korpus, czy też wgląd w postępy pracy będzie zastrzeżony i dostępny tylko wybranym użytkownikom.

Working mode

- public annotations
- agreement

Wprowadzanie anotacji

Uwaga! Użytkownik może znakować tylko w ramach tych warstw, które zostały określone jako widoczne. Wyświetlanie anotacji definiujemy w ustawieniach widoku dokumentu (zob. [View configuration](#)).

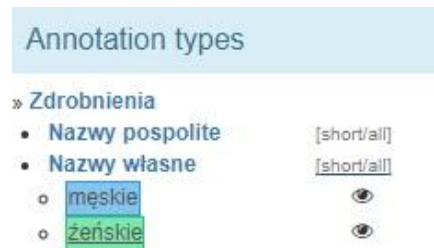


to słońce po szarym piachu i postu
daniu.
— Mogła byś ty kiecuchny barzy

Zaznaczamy wybrany fragment w tekście - wystarczy zaznaczenie kilku liter w obrębie szarej ramki, którą wydzielone są tokeny - najmniejsze jednostki w tekście, które można anotować.

W panelu **Annotation Types** wybieramy odpowiednią kategorię, którą chcemy oznaczyć w

treści dokumentu i klikamy w nią. Uwaga! Niektóre kategorie widoczne są dopiero po kliknięciu opcji [short/all].



Flagowanie dokumentu

Po zakończeniu znakowania treści dokumentu zaznaczamy status anotacji, ustawiając odpowiednią flagę:



Edytowanie anotacji

Użytkownik może zmienić przypisaną wcześniej kategorię. W tym celu należy w wybranym dokumencie kliknąć w anotację, którą chcemy edytować i w oknie **Annotation details** wybrać opcję **change**, wprowadzić zmianę, zaznaczając właściwą kategorię.

Uwaga: opcja **change** umożliwia edycję tych kategorii, które zostały określone jako widoczne.

Usuwanie anotacji



Opcja usuwania anotacji dostępna jest w oknie Annotation details. Po wybraniu jej użytkownik będzie musiał potwierdzić decyzję w oddzielnym oknie. Należy pamiętać, że usunięcie anotacji jest nieodwracalne:



Add relation: aby móc określić rodzaj relacji między anotacjami, należy:

1. zdefiniować rodzaj relacji, jaki będziemy oznaczać (zob. [View configuration](#));
2. oznaczyć przynajmniej dwie anotacje;

Bootstrapping - weryfikacja automatycznej anotacji

Wybór zestawu kategorii

Bootstrapping to nazwa perspektywy, która daje nam wgląd w opcje weryfikowania anotacji, które zostały dodane w sposób automatyczny. Pozwala użytkownikowi określić, czy anotacja jest prawidłowa czy błędna.

Anotowanie automatyczne odbywa się w ramach gotowych zestawów:

- nazw własnych (Named Entities) - 3 zestawy różniące się ilością rozpoznawanych kategorii
- wyrażeń temporalnych (Temporal Expressions) - 2 zestawy.

Weryfikacja

Aby zweryfikować automatycznie rozpoznane kategorie, należy przejść do zakładki Bootstrapping, a następnie określić, który zestaw anotacji chcemy sprawdzić, klikając w jego nazwę panelu [View configuration](#):

View configuration			
Annotation set	New	Final	Discarded
Named Entities (nam)	147	0	0
Named Entities (top9)	132	0	0
Named Entities (n82)	122	0	0
Temporal Expressions (1 class)	46	0	0
Temporal Expressions (4 classes)	41	0	0

Type	Text	Later	Accept	Discard	Change to
nam_org	Maria	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	-
	Dabrowska	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	-
	Dzikie	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	-
nam_liv	Marynka	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	-
nam_liv	Fornał	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	-
nam_liv	Słupecki	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	-
nam_liv	Banasiaka	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	-
nam_liv	Banasiaku	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	-
nam_liv	Marynka	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	nam_adj
nam_liv	Marynka	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	nam_eve
nam_liv	Marynka	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	nam_fac
nam_liv	Marynka	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	nam_liv
nam_liv	Dionizy	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	nam_loc
nam_liv	Dionizemu	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	nam_num
nam_loc	Marynka	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	nam_org
nam_liv	Bogu	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	nam_oth
nam_liv	Marynka	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	nam_pro
nam_liv	Dionizego	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	-

Dla każdej przypisanej anotacji użytkownik może wybrać jedną z opcji weryfikowania: Later, Accept, Discard, Change to.

Kiedy błędne anotacje zostaną odrzucone a właściwe zaakceptowane, należy potwierdzić weryfikację, wybierając **Save verification** na dole listy anotacji:



Przeglądanie anotacji po weryfikacji

Po sprawdzeniu automatycznie rozpoznanych kategorii możemy sprawdzić, jak wyglądają zaakceptowane przez nas anotacje. Przechodzimy do perspektywy *Annotator*. Należy pamiętać, że będziemy mogli przeglądać wyłącznie te kategorie, które zostały zaznaczone w panelu konfiguracji jako widoczne. W ten sam sposób możemy przeanalizować pozostałe dokumenty (pasek nawigacyjny u góry ekranu).

Przeglądanie anotacji

Użytkownik może przeglądać anotacje w korpusie na kilka różnych sposobów, w zależności od tego, jakich informacji potrzebuje:

1. [Filtry w sekcji Documents](#) - jeśli chcemy zlokalizować daną kategorię i przypisane do niej anotacje wśród dokumentów w korpusie.
2. [Sekcja Annotations](#) - jeśli interesuje nas zestawienie liczbowe przypisanych do poszczególnej kategorii anotacji, z możliwością szybkiego przejścia do określonego zastosowania w korpusie.
3. [Sekcja Annotation Browser](#) - jeśli chcemy przyjrzeć się wprowadzonym anotacjom w kontekście, w jakim występują w tekście, a także eksportować informacje do formatu CSV.

Filtry w sekcji Documents

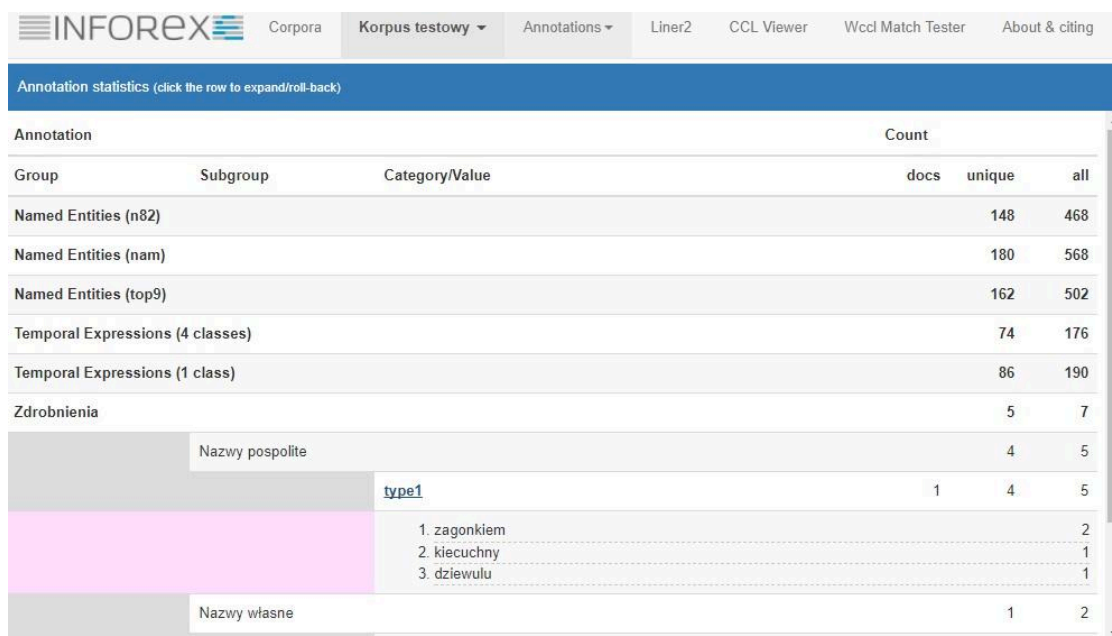


Po prawej stronie listy wszystkich dokumentów, zawartych w korpusie, znajduje się panel umożliwiający ich filtrowanie. Za jego pomocą możemy wprowadzić sprecyzowane kryteria wyszukiwania anotacji, wybierając filtr [Annotation](#). Początkowo aktywne są wszystkie kategorie anotacji (ich wybór oznaczony jest zielonym kolorem). Aby wybrać jedną z nich, klikamy w nazwę kategorii (np. `name_location`). Jeśli chcemy zaznaczyć dwie lub więcej, klikamy znak `[+]` po lewej stronie. Aby usunąć filtr, wybieramy opcję `cancel`.

Możemy także wybrać inny filtr: [Annotation value](#), za pomocą którego wprowadzamy typ anotacji i wpisujemy jej konkretną wartość.

Annotations:

W sekcji **Annotations** użytkownik może przeglądać przypisane do poszczególnych kategorii anotacje. Znajduje się w niej pełna lista anotacji, które znajdują się w korpusie z podziałem na: zestawy, podzestawy i kategorie:



The screenshot shows the INFOREX interface with the 'Annotations' tab selected. Below the navigation bar, there is a section titled 'Annotation statistics (click the row to expand/roll-back)'. The main table has columns for 'Annotation', 'Group', 'Subgroup', 'Category/Value', and 'Count' (with sub-columns for 'docs', 'unique', and 'all').

Annotation	Group	Subgroup	Category/Value	Count		
				docs	unique	all
Named Entities (n82)					148	468
Named Entities (nam)					180	568
Named Entities (top9)					162	502
Temporal Expressions (4 classes)					74	176
Temporal Expressions (1 class)					86	190
Zdrobnienia					5	7
		Nazwy pospolite			4	5
			type1	1	4	5
			1. zagonkiem			2
			2. kiecuchny			1
			3. dziewulu			1
		Nazwy własne			1	2

Z lewej strony tabeli znajdują się nazwy zestawów kategorii, które użytkownik zatwierdził (w panelu ustawień: Settings) jako widoczne w korpusie. Jeśli któryś z automatycznie rozpoznawanych zestawów został oznaczony przez użytkownika jako niewidoczny, w tabeli przed nazwą pojawi się informacja: *HIDDEN*.

<i>HIDDEN::Named Entities (n82)</i>	148	469
<i>HIDDEN::Named Entities (nam)</i>	180	568
<i>HIDDEN::Named Entities (top9)</i>	162	502

Po prawej stronie znajduje się zestawienie liczbowe przypisanych do poszczególnej kategorii anotacji (z wyróżnieniem: liczby dokumentów zawierających dany typ anotacji - kolumna **docs**; liczby unikalnych anotacji danego typu - **unique**; liczby wszystkich anotacji danego typu w korpusie - **all**).

Category/Value	docs	unique	all
timex	16	86	190
1. nocy			20
2. dzisiaj			6
o dąbrowska_dzikieziele.ccl			
o Dąbrowska_Dzikie_ziele.ccl			
3. dziś			6
o dąbrowska_dzikieziele.ccl			
o Dąbrowska_Dzikie_ziele.ccl			
4. rano			5
o dąbrowska_dzikieziele.ccl			
o Dąbrowska_Dzikie_ziele.ccl			
o Jasieński_Klucze.ccl			
o jasieński_klucze.ccl			
o Żeromski_Zmierzch.ccl			
5. noc			5

Jeśli chcemy rozszerzyć widok w tabeli o poszczególne podgrupy i kategorie, wystarczy kliknąć w konkretny jej wiersz. Wybranie danej kategorii wyświetli przynależące do niej anotacje, a także dane liczbowe. Po kliknięciu konkretnej

wartości anotacji otrzymamy informacje o dokumentach, w których dana anotacja występuje. Jeśli następnie wybierzemy dany dokument, zostaniemy przeniesieni do sekcji Documents i widoku jego zawartości, gdzie będziemy mogli zlokalizować konkretną anotację.

Lista Common filters:

Widok sekcji Annotations możemy filtrować za pomocą panelu Common filters, znajdującym się po prawej stronie.

Common filters	
Subcorpus:	gromada , Wikinews , Opowiadania , podkorpus testowy , all
Status:	Nieznany , Przyjęty , Załącznik , Odrzucony , Odłożony , Do usunięcia , all

Przeglądanie anotacji w kontekście w sekcji Annotation browser

Zakładka **Annotation browser** to jeszcze jeden mechanizm ułatwiający przeglądanie anotacji w korpusie. Użytkownik może za jej pomocą bardzo dokładnie określić, które spośród anotacji chce przeglądać, definiując odpowiednio ich status, kategorię, a także wybierając konkretną formę występującą w tekście lub formę bazową (lemat).

Annotation stage - w tym panelu określamy, jaki status mają mieć przeglądane przez nas anotacje:

New - określa status anotacji, które zostały dodane do korpusu w sposób automatyczny.

Final - anotacje zweryfikowane i zaakceptowane przez użytkownika.

Discarded - anotacje odrzucone przez użytkownika.

Annotation stage	
1894	new
16	final

Annotation types - po wybraniu statusu wyświetlają się wszystkie kategorie anotacji z podziałem na zestawy, które są określone tym statusem. Po lewej stronie w kolumnie znajduje się informacja o ilości anotacji danego typu/kategorii.

Annotation types	
2	t3_date
1	t3_time
Zdrobnienia	
2	imiona_zenskie

Text forms/Lemmas - po wybraniu kategorii możemy zdecydować, czy interesuje nas konkretna forma wyrazowa czy forma podstawowa (lemat) we wszystkich wariantach jej użycia.

Text forms	Lemmas
2 zagonkiem	2 zagonek
1 dziewulu	1 dziewula
1 kiecuchny	1 kiecuchna

Kontekst

Po wprowadzeniu poszczególnych opcji wyszukiwania, w tabeli po prawej stronie pojawi się podgląd anotacji wraz z fragmentem treści dokumentu (kontekstem), a

także podstawowymi informacjami: numerem dokumentu, z którego pochodzą (kliknięcie numeru przeniesie nas do widoku zawartości dokumentu - perspektywa *Preview*), źródłem anotacji (anotowanie ręczne lub automatyczne) i statusem.

Contexts		
Left	Annotation	Right
Pleć też bardzo lubiła. Położyła się bokiem w bruździe, między	zagonkiem	a zagonkiem i wyskubywała z ciepłych grudek dzikie ziele i r
lubiła. Położyła się bokiem w bruździe, między zagonkiem a	zagonkiem	i wyskubywała z ciepłych grudek dzikie ziele i niepotrzebne

Jeśli przeglądamy anotacje jeszcze niezatwierdzone, możemy szybko przejść z tabeli do perspektywy *Bootstrapping*, gdzie będziemy mogli zweryfikować anotacje. Wystarczy kliknąć status *new* po prawej stronie tabeli.

	Source	Stage
li liczącej	bootstrapping	new
li liczącej	bootstrapping	new
stało to, czego	bootstrapping	new
zedł blisko. Powiedzia	bootstrapping	new
cki. — To z nim	bootstrapping	new
i nich wyglądać	bootstrapping	new

Eksportowanie wybranych anotacji do CSV

Export selected annotations to CSV

Anotacje widoczne w tabeli możemy eksportować do formatu CSV i zapisać na dysku komputera. Wygenerowany arkusz będzie zawierał dokładnie te same informacje, co zawartość tabeli:

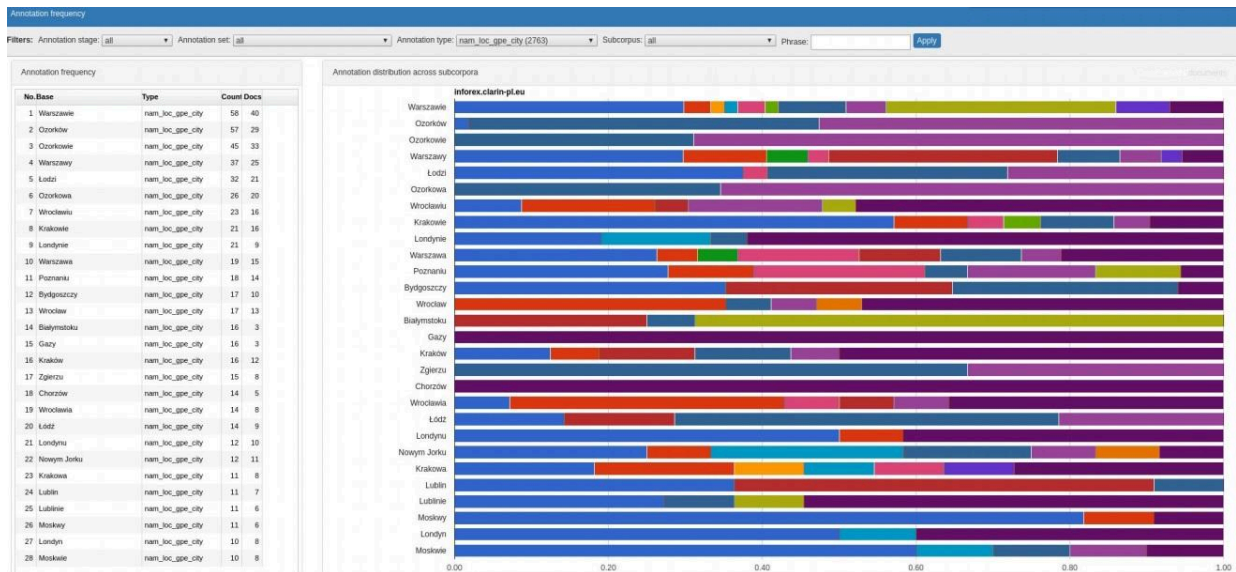
	A	B	C	D	E	F	G	H
1	Document	Annotation	Type	Left context	Annotation	Right context	Stage	Source
2	145088	1807762	type1	Pleć też bard	zagonkiem	a zagonkiem i w	final	user
3	145088	1807761	type1	lubiła. Położy	zagonkiem	i wyskubywała z	final	user

Annotation frequency

W zakładce *Annotation frequency* użytkownik może zapoznać się z danymi dotyczącymi częstotliwości występowania anotacji oraz ich dystrybucji w ramach poszczególnych podkorpusów. Dane te zaprezentowane są w czytelny sposób za pomocą wykresu paskowego i tabeli po lewej stronie.

Lista frekwencyjna anotacji jest zależna od wprowadzonych **wartości filtrowania**. Uzupełniamy pola: *annotation stage*, *annotation set*, *annotation type*, *subcorpus* w zależności od zakresu danych, jaki objąć ma wykres. W tym miejscu możemy sprawdzić np., ile spośród wszystkich anotacji w korpusie dotyczy nazw miast, a także, ile razy konkretna nazwa

występuje w korpusie, w ilu dokumentach się pojawia, w którym podkorpusie (wykres paskowy) najczęściej.



Możemy także w polu Phrase wprowadzić konkretną wartość anotacji w celu wygenerowania informacji o częstotliwości jej występowania w korpusie.

Phrase:

Statystyki

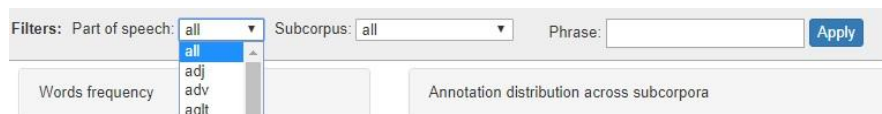
W zakładce Statistics użytkownik znajdzie podstawowe dane statystyczne dotyczące korpusu: liczbę dokumentów, słów, znaków i tokenów wchodzących w skład poszczególnych podkorpusów:

Corpus statistics				
Number of words in accepted documents. Word is a sequence of characters matchig following regex " <code>(\pL \pM \pN)+</code> " (according to http://www.regular-expressions.info/unicode.html).				
Subcorpus	Documents only accepted	Words	Characters (no whitespaces)	Tokens
Opowiadania	5	23.154	97.244	23.026
podkorpus testowy	1	191	881	186
Wikinews	11	6.008	29.276	6.523
TOTAL	17	29.353	127.401	29.735

Uwaga! Jako słowo system wyróżnia ciąg znaków (oddzielonych separatorem lub znakiem interpunkcyjnym), niezależnie od wielkości liter, obecności znaków diakrytycznych czy numerycznych. Liczba znaków natomiast wyliczana jest z pominięciem spacji.

Words frequency

Aby wygenerować **listę frekwencyjną słów**, wybieramy część mowy z dostępnych w rozwijanej liście klas gramatycznych (zastosowano [tagset NKJP](#)), a także podkorpus. Możemy także wprowadzić dowolne wyrażenie w pole wyszukiwania. Wszelkie wytyczne zatwierdzamy Apply:



Z lewej strony znajduje się tabela, w której zawarte są jednostki realizujące wcześniej określony zakres wyszukiwania. Możemy w tym miejscu dowiedzieć się, ile razy w korpusie i w ilu różnych dokumentach występuje dane wyrażenie. Wykres paskowy po prawej stronie ułatwia nam zapoznanie się ze statystyką występowania danego słowa w obrębie całego korpusu z podziałem na poszczególne podkorpusy. Użytkownik może zmienić widok wykresu przyciskiem w jego prawym górnym rogu: **Count: words/documents** tak, by wyświetlał dane liczbowe dotyczące słów lub dokumentów.

Words frequency			
No.	Base	Count	Docs
13	żaden	18	2
14	głupi	18	4
15	wielki	17	5
16	tamten	17	5
17	ki	16	4
18	cały	15	5
19	dziki	14	2
20	młody	12	4
21	biały	11	3
22	mały	11	5
23	inny	11	3

Page 2 of 32 Displaying



Za pomocą przycisków w dolnej części ekranu możemy eksportować listę frekwencyjną słów do formatu CSV. Eksportować możemy całą listę słów danego korpusu (z wyszczególnionym podziałem na podkorpusy)

Export frequency distribution to a CSV file

lub bieżącą z wprowadzonymi kryteriami.

Export current frequency list to a CSV file

Menadżer zadań

W zakładce Tasks użytkownik może sprawdzić listę wykonanych przez program zadań, sprawdzić ich status i poszczególne fazy ([show details](#)). W tym miejscu możemy także zlecić wykonanie nowego zadania.

Date and time	Task	Documents	Export	Details
2017-08-30 12:36:05	liner2	17	A	show details
2017-07-30 19:40:31	dspace_import	17	A	show details

Dodawanie dokumentów do korpusu

W zakładce **Add document** możemy w prosty sposób wprowadzić nowy dokument do korpusu. Jego zawartość wklejamy w pustym miejscu. Tekst automatycznie zostanie rozdzielony na akapity.

Common metadata	
Title	dokument1
Author	author
Source	create
Subcorpus	podkorpus testowy
Status	Przyjęty
Date	2017-08-30 released, published or created
Format	plain

W panelu po prawej stronie wprowadzamy opis dokumentu: tytuł, autora, opisujemy źródło, format, datę powstania, a także decydujemy, do którego podkorpusu będzie przypisany. Po wprowadzeniu wszystkich informacji klikamy **Create document**. Możemy dodać kolejny dokument lub wprowadzić zmiany (link) w dokumencie już dodanym.

i Info The document was added. [Edit the content](#) or add another one.

Przesyłanie plików

Zakładka Upload documents

W tym miejscu możemy przesłać do Inforexa zestaw plików. Poszczególne dokumenty powinny być w formacie **txt**. Przesyłamy je skompresowane do formatu **ZIP**.

Jeśli chcemy przesłać pliki wraz z metadanymi, powinniśmy zapisać informacje w formacie **ini** według wytycznych podanych w ramce.

Upload a set of txt files.

Zip file Nie wybrano pliku

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"
```

Subcorpus

Split into subcorpora based on the file prefix: **SUBCORPUS-DOCUMENT_NAME.txt**

Możemy zaznaczyć, do którego korpusu przypisane mają zostać pliki.

Jeśli podczas tworzenia plików wykorzystamy następujący format ich nazywania: SUBCORPUS-DOCUMENT_NAME.txt, automatycznie zostaną one przypisane do odpowiednich (określonych w nazwie) podkorpusów.