

# Warsztat nr 1

*Gromadzenie, udostępnianie i anotowanie korpusów*

## Publikowanie zasobów w repozytorium DSpace

### Czas

25 minut

### Cel

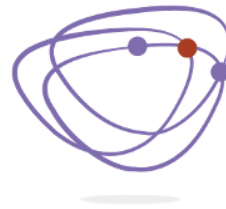
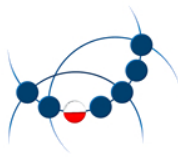
Zapoznanie Użytkownika z systemem DSpace, służącym jako repozytorium danych przechowywanych w ramach CLARIN-PL.

### Przydatne materiały

1. [Wideo \(youtube\)](#)
2. [Instrukcja](#)

### Do samodzielnego wykonania

1. Rejestracja w systemie DSpace ([link](#))
2. Pobranie i rozpakowanie testowego korpusu ([link](#))
3. Przeglądanie plików w testowym korpusie
4. Zdeponowanie i opisanie testowego korpusu (*Start a new submission*, [link](#))
5. Wyeksportowanie przetworzonego testowego korpusu do systemu Inforex ([link](#))
6. Przeglądanie zaimportowanych dokumentów z testowego korpusu w systemie Inforex



## Zarządzanie korpusami i anotacja tekstów w systemie Inforex

### Czas

35 minut

### Cel

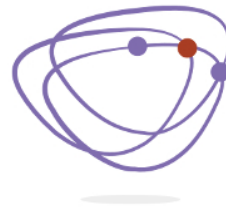
Zapoznanie uczestników z podstawowymi funkcjami systemu do zarządzania korpusami i ich anotacji - Inforex

### Przydatne materiały

1. [Wideo](#)
2. [Instrukcja](#)

### Do samodzielnego wykonania

1. Logowanie do systemu ([link](#))
2. Flagowanie korpusu
3. Zarządzanie użytkownikami i ich uprawnieniami
4. Tworzenie podkorpusów
5. Opisywanie metadanymi
6. Oznakowanie wybranego dokumentu słowami kluczowymi
7. Generowanie listy frekwencyjnej przymiotników w wybranym podkorpusie



## Gromadzenie korpusów ze źródeł internetowych

### Czas

15 minut

### Cel

Zapoznanie Użytkownika z możliwością gromadzenia korpusów ze źródeł internetowych z wykorzystaniem narzędzia *CorpoGrabber* oraz przeglądanie pobranych zasobów w systemie *Inforex*.

### Przydatne materiały

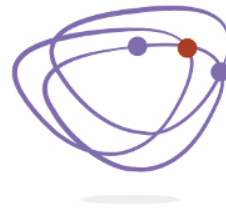
1. [Wideo \(youtube\)](#)
2. [Instrukcja](#)

### Do samodzielnego wykonania

1. Wybranie w systemie *Inforex* korpusu, który był wyeksportowany z systemu *DSpace*
2. Przejście do perspektywy *Tasks*
3. Wybranie strony internetowej zawierającej dokumenty tekstowe (strony z informacjami, encyklopedie, blogi, inne)
4. Skopiowanie adresu URL
5. Kliknięcie w przycisk *New CorpoGrabber Task* w perspektywie *Tasks*
6. Wklejenie skopiowanego adresu i wciśnięcie przycisku *OK*

**Uwaga!** Niektóre strony mają blokadę uniemożliwiającą pobranie zawartości w taki sposób, dlatego proces może zakończyć się niepowodzeniem. W zależności od liczby uczestników może powstać kolejka oczekujących żądań, dlatego trzeba będzie chwilę poczekać, aż proces pobierania się rozpocznie.

7. Przeglądanie kilku pobranych dokumentów



## Przeszukiwanie korpusów za pomocą systemu NoSketch Engine

### Czas

15 minut

### Cel

Zapoznanie uczestników z podstawowymi funkcjami systemu do przeszukiwania korpusów tekstowych - NoSketch Engine

### Przydatne materiały

1. [Wideo](#)
2. [Instrukcja](#)

### Do samodzielnego wykonania

1. Przejście po adres usługi przeszukiwania korpusów opartej na NoSketch Engine ([link](#))
2. Wyszukanie w KPWr wystąpień słowa *dom* wraz z ich kontekstem (konkordancja)
3. Wygenerowanie z KPWr listy frekwencyjnej czasowników
4. Porównanie częstości występowania słowa *dom* w dwóch podkorpusach KPWr
5. Wyszukanie kolokacji słowa *dom* w korpusie KPWr