



CLARIN-PL

**Tomasz Naskręt,
Marcin Pol**

**Wrocław University of Technology
Institute of Informatics
G4.19 Research Group
marcin.pol@pwr.edu.pl**

2016-11-09



**Politechnika
Wroclawska**

CLARIN-PL

Common Language Resources and Technology Infrastructure



Polish language



Users: Collegium Civitas, Warsaw

Problem: Semantic text classification for sociology research

Initially:

- Text document classification according to the manually annotated examples

Finally:

- Whole system from corpus gathering to tuning machine learning methods for the semantic classification of text snippets

POLISH

1. dwa
2. dwie
3. dwoje
4. dwóch
5. dwaj
6. dwiema
7. dwom
8. dwoma
9. dwojga
10. dwojgu
11. dwojgiem
12. dwójka
13. dwójki
14. dwójkę
15. dwójką
16. dwójce
17. dwójko

ENGLISH

1. Two
2. Second

Polish language



NLP Web Services for Polish

Part of the SyNaT project developed at Wrocław University of Technology

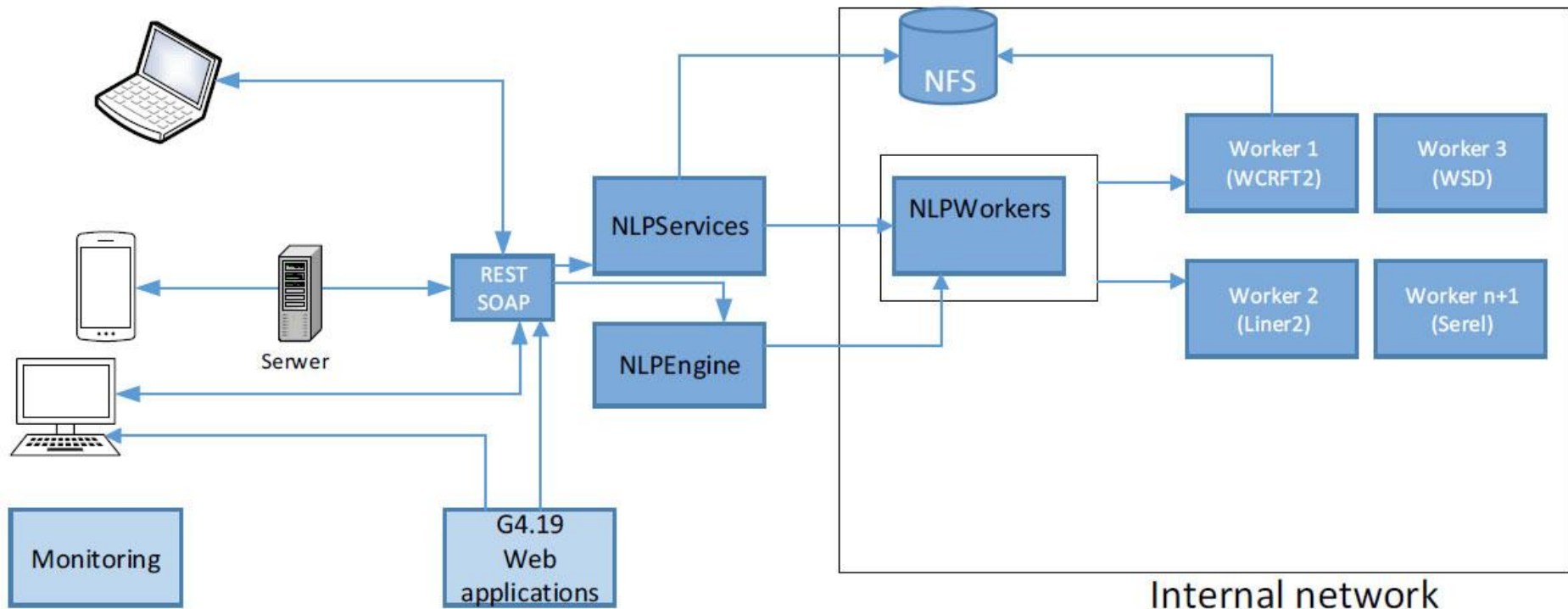


Wrocław University of Technology

Home	Morpho-WS	Tagger-WS	Chunker-WS	ChunkRel-WS	NER-WS	Serel-WS
NLP Web Services for Polish						
Plain text	W XIX wieku na wyspie mieszkał sultan Abdullah z Peraku, zesłany tu przez Brytyjczyków.					
↓						
Morpho-WS	Performs tokenisation and morphological analysis of Polish text using Morfeusz SGJP and MACA toolkit. For full tagging (no ambiguity) please use Tagger-WS.	W XIX wieku na wyspie mieszkał sultan Abdullah z Peraku, zesłany tu przez Brytyjczyków .				
↓						
Tagger-WS	Performs tokenisation and morphosyntactic tagging of Polish text using WCRFT tagger.	W XIX wieku na wyspie mieszkał sultan Abdullah z Peraku, zesłany tu przez Brytyjczyków .				
↓						
Chunker-WS	Recognises boundaries of noun phrases, verb phrases and adjective phrases in Polish text.	W XIX wieku na wyspie mieszkał sultan Abdullah z Peraku, zesłany tu przez Brytyjczyków .				
↓						
ChunkRel-WS	Prototype service for recognition of inter-chunk syntactic relations (object, subject, copula).	W XIX wieku na wyspie mieszkał sultan Abdullah z Peraku, zesłany tu przez Brytyjczyków .				
↓						
NER-WS	Recognises 56 fine-grained categories of proper names in Polish texts.	W XIX wieku na wyspie mieszkał sultan Abdullah z Peraku, zesłany tu przez Brytyjczyków .				

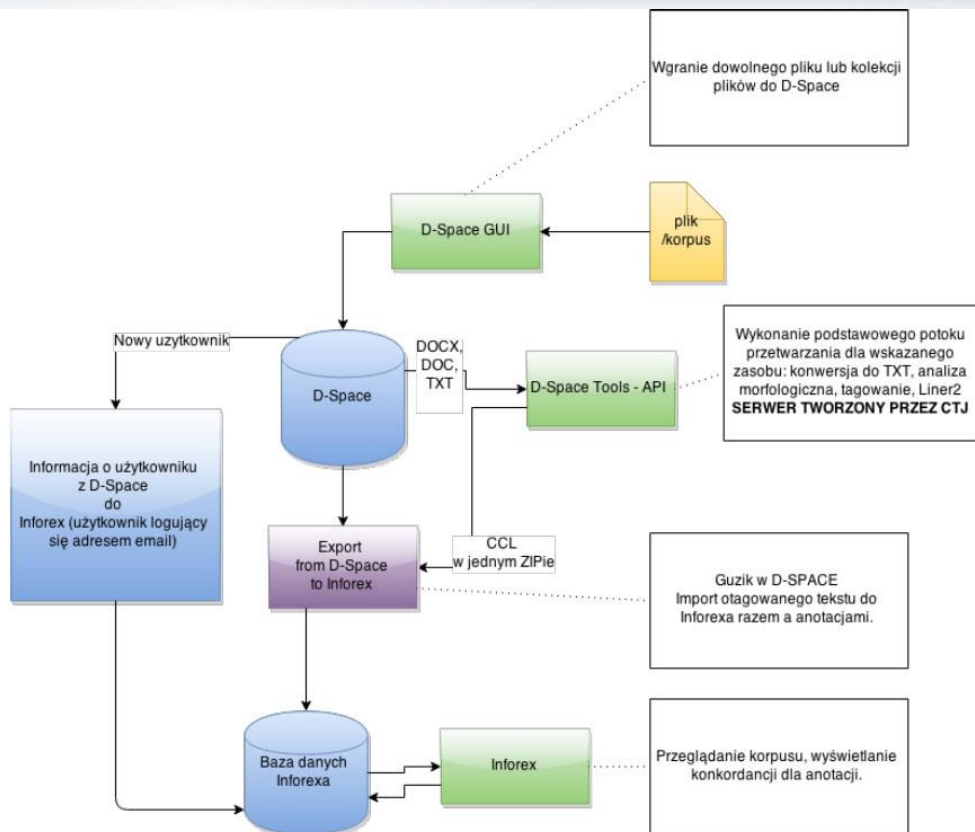


Services workflow





D-Space workflow



D-Space process files & export



“ Please use the following text to cite this item or export to a predefined format:

[VIEW BIBTEX](#) [VIEW CMDI](#)

Smith, Donald, 2015, *Example*, CLARIN-PL digital repository, <http://hdl.handle.net/11321/325>.



[CLARIN-PL](#)

Authors	Smith, Donald
Date issued	2015-11-24
Type	corpus
Language(s)	English
Description	ausid ulasdh uladsh sad
Publisher	Helios
Subject(s)	example
Collection(s)	CLARIN-PL

Files in this item

[Add to Archive](#)

[Process files to CCL](#)

This item is [Distributed under Creative Commons](#) and licensed under:
[IS PAS corpora license](#)

Name	W2 - QUERY 2016.pdf
Size	899.61 KB



[CLARIN-PL Repository Home](#) / [View Item](#)

Teksty reklam TVP ABC

“ Please use the following text to cite this item or export to a predefined format:

[VIEW BIBTEX](#) [VIEW CM](#)

Ka, Emilia, 2016, *Teksty reklam TVP ABC*, CLARIN-PL digital repository, <http://hdl.handle.net/11321/242>.



[CLARIN-PL](#)

Authors	Ka, Emilia
Date issued	2016-01-27
Type	corpus
Language(s)	Polish
Description	teksty reklam emitowane na kanale TVP ABC miedzy lipcem 2014 a styczniem 2015
Publisher	Emilia
Subject(s)	reklama

[Show full item record](#)

Files in this item

[Add to Arch](#)

[Make sense frequency list](#)

[Export to MeWeX](#)

[Export to Inforex](#)

wosedon.pl/show/wlw_cd_20161107_235926_187169

Włodzisław Włodarczyk

Lista frekwencyjne znaczeń

źródło: wlw_cd_20161107_235926_187169

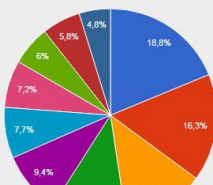
top lista
N:

część mowy
 rzeczownik
 czasownik
 przysłówek

stop lista
 brak
 domyślna
 Twoja

Wpisuj każde słowo w osobnej linijce

Lp	Znaczenie	Część mowy	Definicja	Przykłady użycia	Liczba wystąpień
1	narrator 1 (os)	rzecz.	«opowiadacz»		228
2	być 4 (cst)	czas.	«egzystować, istnieć»		198
3	kobieta 1 (os)	rzecz.	«dorosły człowiek płci żeńskiej; pani, kobiecina, kobiecisko»	Mężczyzna jest bardziej zmysłowy od kobiety.	152
4	nowy 4 (jak)	przym.	«taki, którego wcześniej nie było»	To jest zupełnie nowy problem, dwadzieścia lat temu nikt się nad tym nie zastanawiał.	139
5	głos 16 (z)	rzecz.	«dźwięk, jaki towarzyszy jakiemś zdarzeniu, bodziec dźwiękowy; którego źródłem nie jest aparat głosowy żywego stworzenia; np. głos wybuchu, dźwięk, odgłos»		114
6	mieć 1 (cst)	czas.	«charakteryzować się, cechować się, posiadać, odznaczać się, nosić»		94
7	kobięcy 1 (jak)	przym.	«typowy dla kobiety; np. kobieca sukienka, kobiety wdzięk; damski»	Zakładę sukienkę, jest znacznie bardziej podkreśla figurę i w ogóle jest urocza.	87
8	to 2 (cst)	czas.	«być»		73
9	mężczyzna 1 (os)	rzecz.	«dojrzały płciowo osobnik płci męskiej; pan»		71
10	który 1 (jak)	przym.	«jaki, jakiś, jakowyś, któryś»		58



D-Space additional CMDI



Item submission

https://clarin-pl.eu/dspace/handle/11321/4/submit/1631151e6b72274d5a271a033f464f63126e195f.continue

Home Repository Partners Contact

CLARIN-PL Repository Home / Item submission

Search

What can you do?

DEPOSIT CITE

Browse

> All of the Repository

My Account

Logout

Profile

Submissions

Context

> Edit Collection

> Item Mapper

Item submission

Basic Info Who's involved Describe Upload **5. Additional CMDI** License Review Complete

Please attach cmdi files to uploaded files

File	Size	Description	Uploaded CMDI File	Upload CMDI	Action
Architektura.odt	367.09 KB	1		Upload	Create

< Previous Save & Exit Next >

D-Space additional CMDI from prof.



Item submission

https://clarin-pl.eu/dspace/handle/11321/4/submit/1631151e6b72274d5a271a033f464f63126e195f.continue

Home Repository Partners Contact

CLARIN-PL Repository Home / Item submission

Item submission

Basic Info Who's involved Description

Create cmdi file from profile

Profile:

Resource

Create

Please attach cmdi files to uploaded files

File	Size	Description	Uploaded CMDI File	Upload CMDI	Action
Architektura.odt	367.09 kB	1		Upload	Create

< Previous Save & Exit Next >

What can you do?

DEPOSIT CITE

Browse

> All of the Repository

My Account

Logout

Profile

Submissions

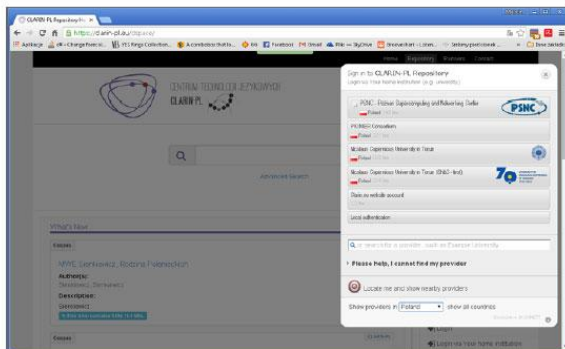
Context

> Edit Collection

> Item Mapper



D-Space plan: users proxy



Sign in to **CLARIN-PL Repository**
Login via Your home institution (e.g. university)

PSNC - Poznan Supercomputing and Networking Center
Poland 145 km

PIONIER Consortium
Poland 228 km

Nicolaus Copernicus University in Torun
Poland 235 km

Nicolaus Copernicus University in Torun (Shib3 - test)
Poland 235 km

Local authentication

Clarin.eu website account
777 km

▶ **Please help, I cannot find my provider**

Locate me and show nearby providers

Show providers in Poland ▼ show all countries

DiscoJuice © UNINETT 1

Thank you very much for your attention