



Uniwersytet
Wrocławski



PRÉKRAČUJEME HRANICE
PRÉKRAČAZAMY GRANICE
2014–2020



EVROPSKÁ UNIE / UNIA EUROPEJSKA
EVROPSKÝ FOND PRO REGIONÁLNÍ ROZVOJ
EUROPEJSKI FUNDUSZ ROZWOJU REGIONALNEGO

Uniwersytet Wrocławski i Uniwersytet Palackiego w Ołomuńcu wspólnie realizują projekt
współfinansowany przez Unię Europejską

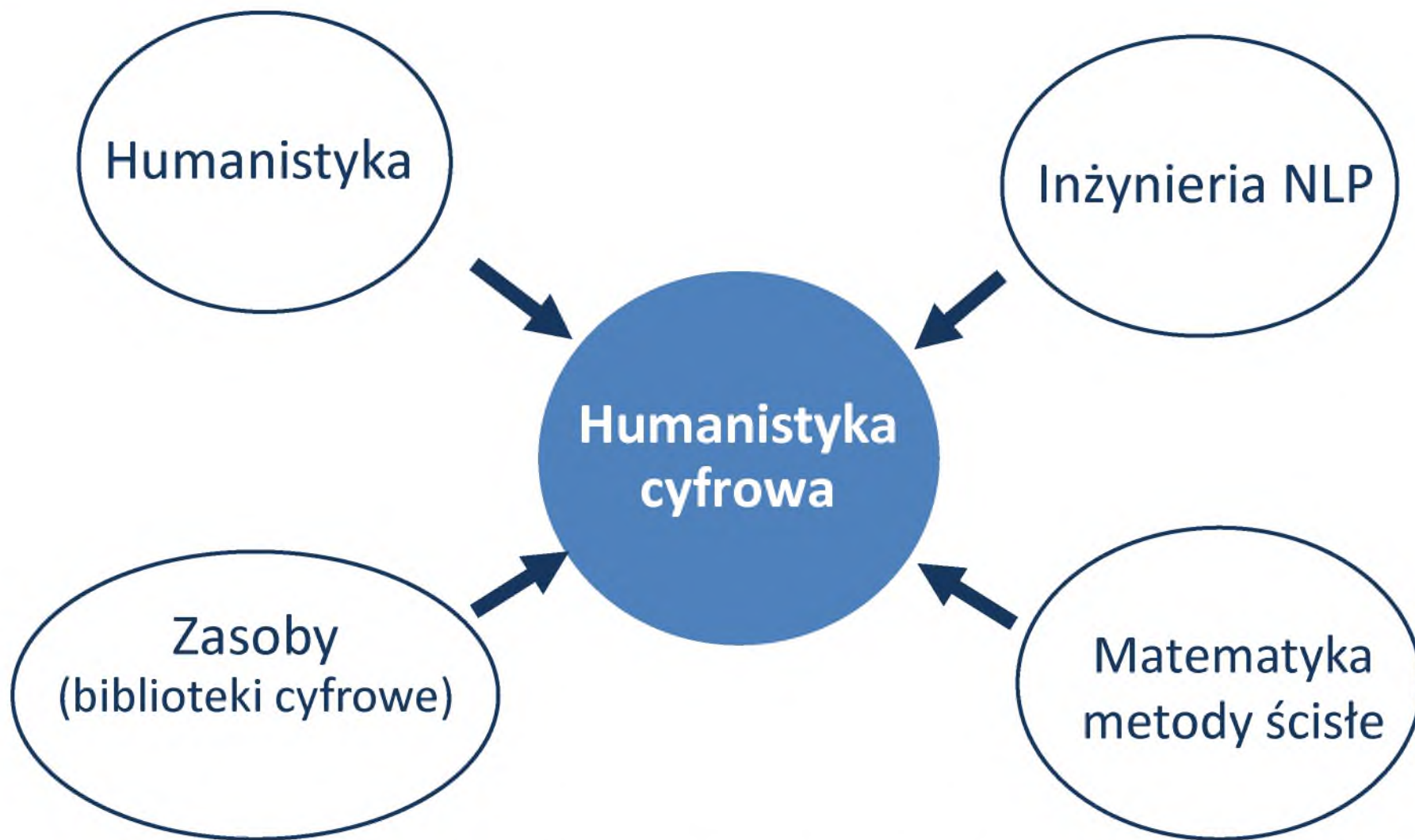
HUMANISTYKA CYFROWA DLA PRZYSZŁOŚCI

Humanistyka cyfrowa z punktu widzenia językoznawstwa: moda, metoda, teoria czy nowy paradygmat?

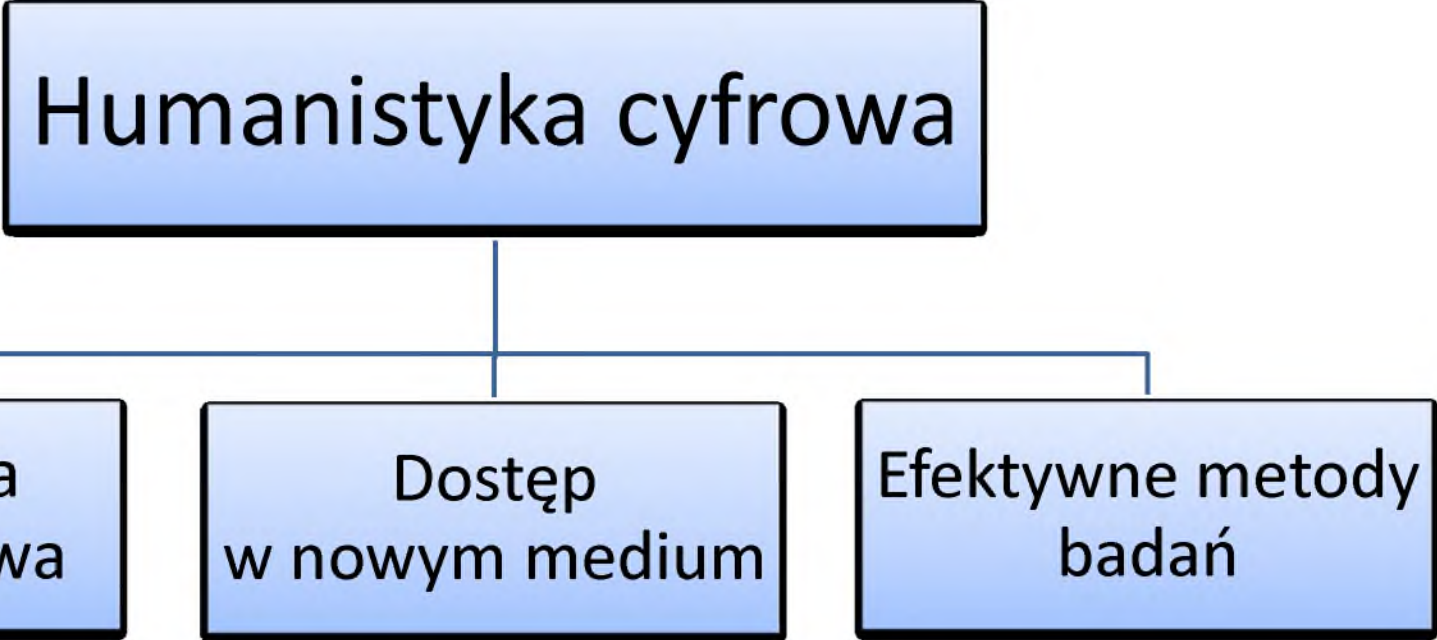
Prof. Adam Pawłowski (IINiB)



GENEZA: SKĄD PRZYCHODZIMY I DOKĄD ZMIERZAMY?



Humanistyka cyfrowa



```
graph TD; A[Humanistyka cyfrowa] --- B[Ochrona dziedzictwa]; A --- C[Dostęp w nowym medium]; A --- D[Efektywne metody badań];
```

Ochrona
dziedzictwa

Dostęp
w nowym medium

Efektywne metody
badań

- scjentystycznie: teoria i definicje
 - słowo o definiowaniu u definicjach
 - definicje zwięzłe
 - definicje enumeracyjne
- przez egzemplia (rewia mody, case studies)
- przez monitoring, zgodnie ze znaną heurystyką potoczności:

Czy większość może się mylić?

Zbiór praktyk poznawczych wraz z zasobami wiedzy i informacji, będący efektem przeniesienia do sfery cyfrowej i rozszerzenia (ilościowego, jakościowego) praktyk poznawczych humanistyki opartej na druku.

- brak opozycji tradycja / nowoczesność
- brak wartościowania
- brak myślenia teleologicznego (nieuchronne prawa rozwoju...)

Badanie dynamiki zmian można prowadzić przez analizy użyć w sieci WWW.

- n-gram Viewer
- Google Trends
- serwisy monitoringowe

Frekwencja deskryptora "humanities" w zasobach indeksowanych przez Google (1900-2000) (dostęp 30.11.2015)

Graph these comma-separated phrases: ☐ case-insensitive

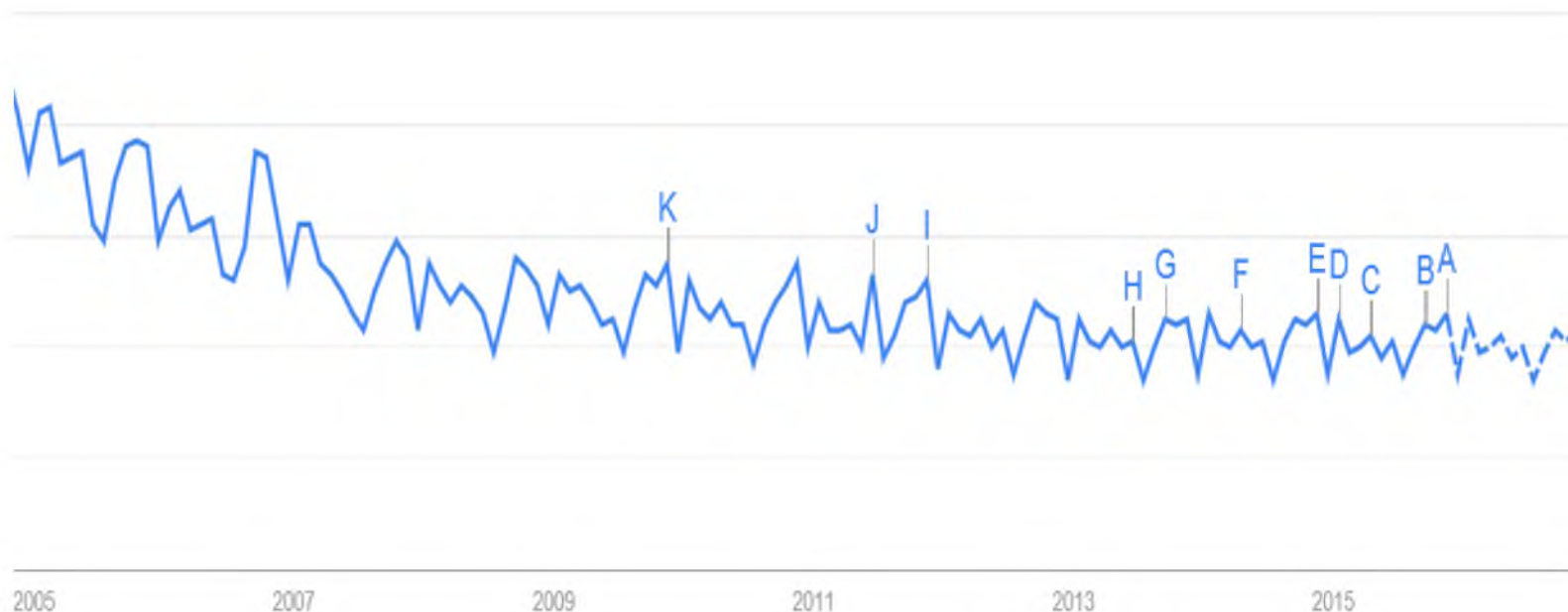
between and from the corpus with smoothing of [Search lots of books](#)



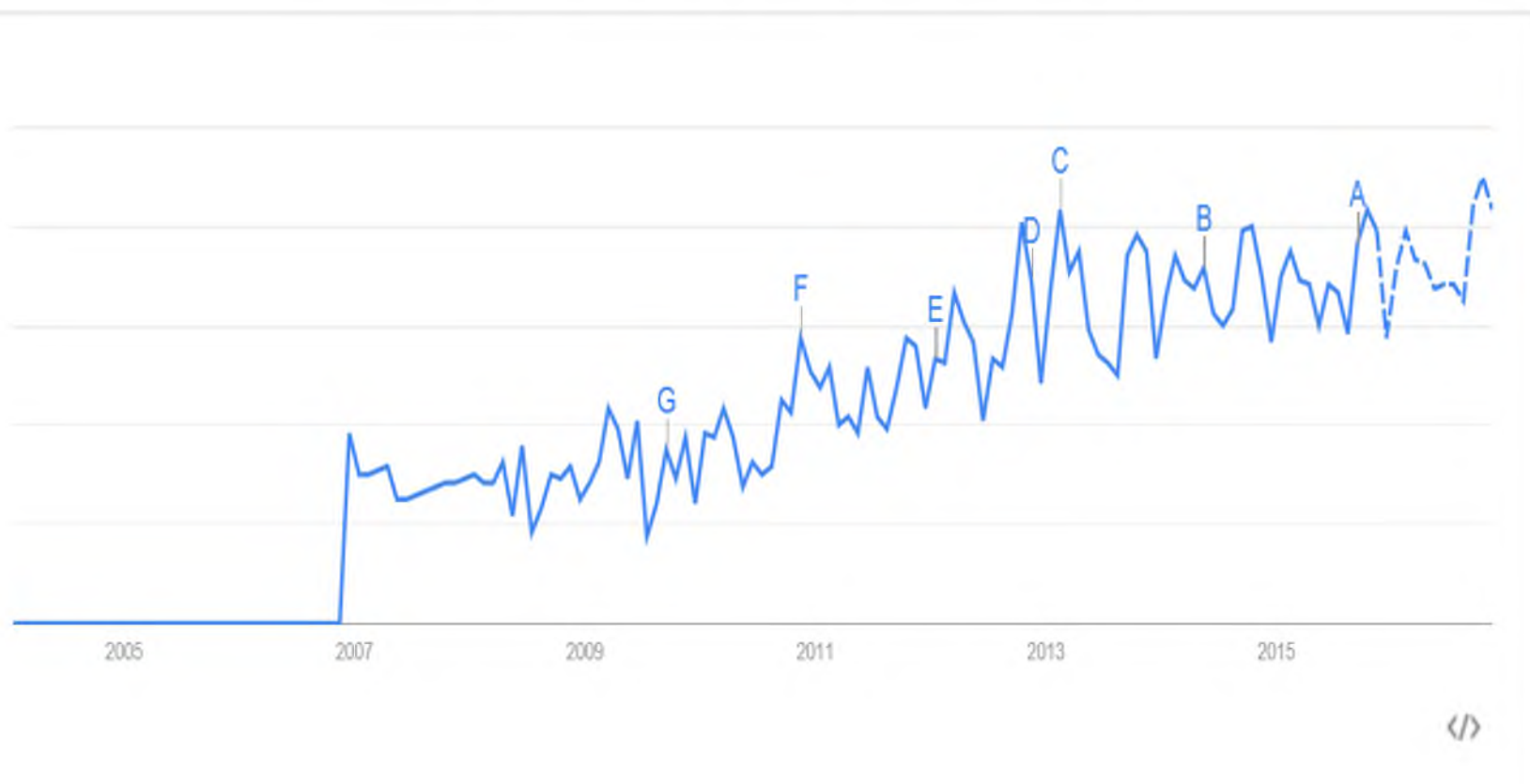
Frekwencja deskryptora "humanities" w zasobach indeksowanych przez Google (1970-2015) (dostęp 30.11.2015)



Frekwencja deskryptora "humanities" w zasobach indeksowanych przez Google (po 2005, dostęp 30.11.2015)

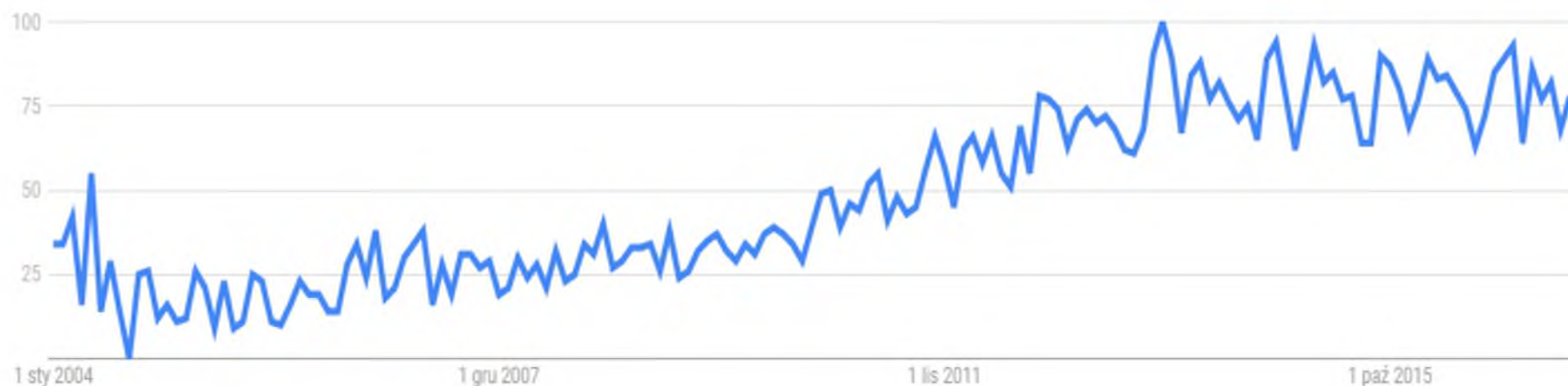


Frekwencja frazy "digital humanities" w zasobach indeksowanych przez Google (od 2005, dostęp 30.11.2015)



Frekwencja deskryptora "digital humanities" w zasobach indeksowanych przez Google (2004-2017) (dostęp 25.05.2017)

Zainteresowanie w ujęciu czasowym ?



Frekwencja frazy "humanistyka cyfrowa" w zasobach indeksowanych przez Google (dostęp 25.06.2017)

Zainteresowanie w ujęciu czasowym ?



Trochę terminologii



- korpusy
- serwisy monitoringowe



- Clarin-PL
- Clarin-EU
- inne



- leksykony
- słowosieci
- tezaury
- korpusy
- biblioteki cyfrowe

- **Zasoby tradycyjne**
 - *archiwa i biblioteki*
 - *ręczne zbieranie danych z natury*
- **Zasoby współczesne (prymarnie cyfrowa)**
 - *korpusy*
 - *repozytoria biblioteki cyfrowe*
 - *cyfrowe narzędzia badawcze*

*Zasoby i narzędzia są po stronie konsorcjów (np. CLARIN),
teorie , pomysły i zastosowania po stronie użytkownika.*



Jak **h**umanistyka **c**yfrowa usiłuje sprostać zalewowi informacji:

- *korzysta z inżynierii języka (NLP)*
- *przez multimodalność i obecność infografik*
- *przez badania ilościowe*
- *dzięki metodom **big data** opartych na tzw. wielkich infrastrukturach naukowych*



*Dzięki szybkości i skali badań HC i NLP tworzą nowe obiekty
badań empirycznych.*

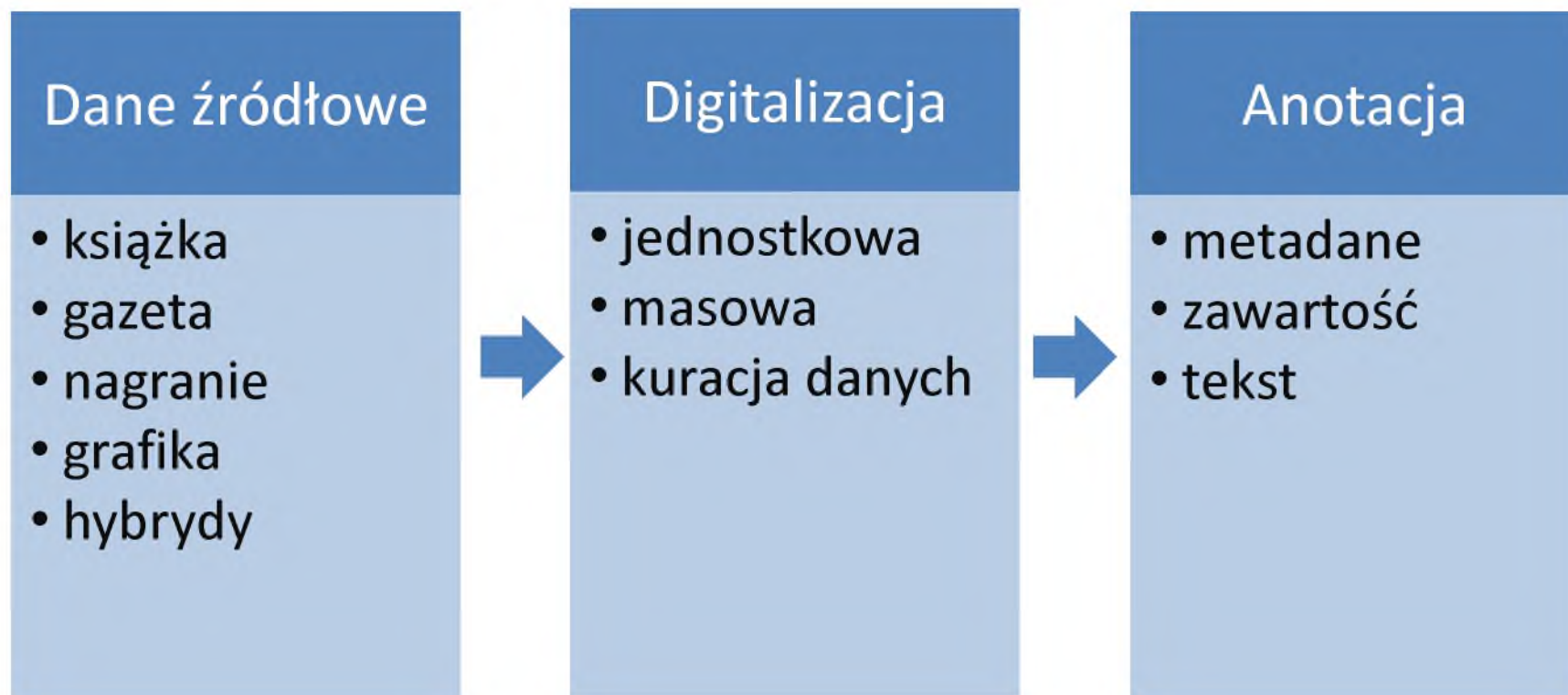
*Wcześniej obiekty takie istniały w wyobraźni uczonych – jako
przedmioty dociekań teoretycznych lub fantastyki naukowej.*



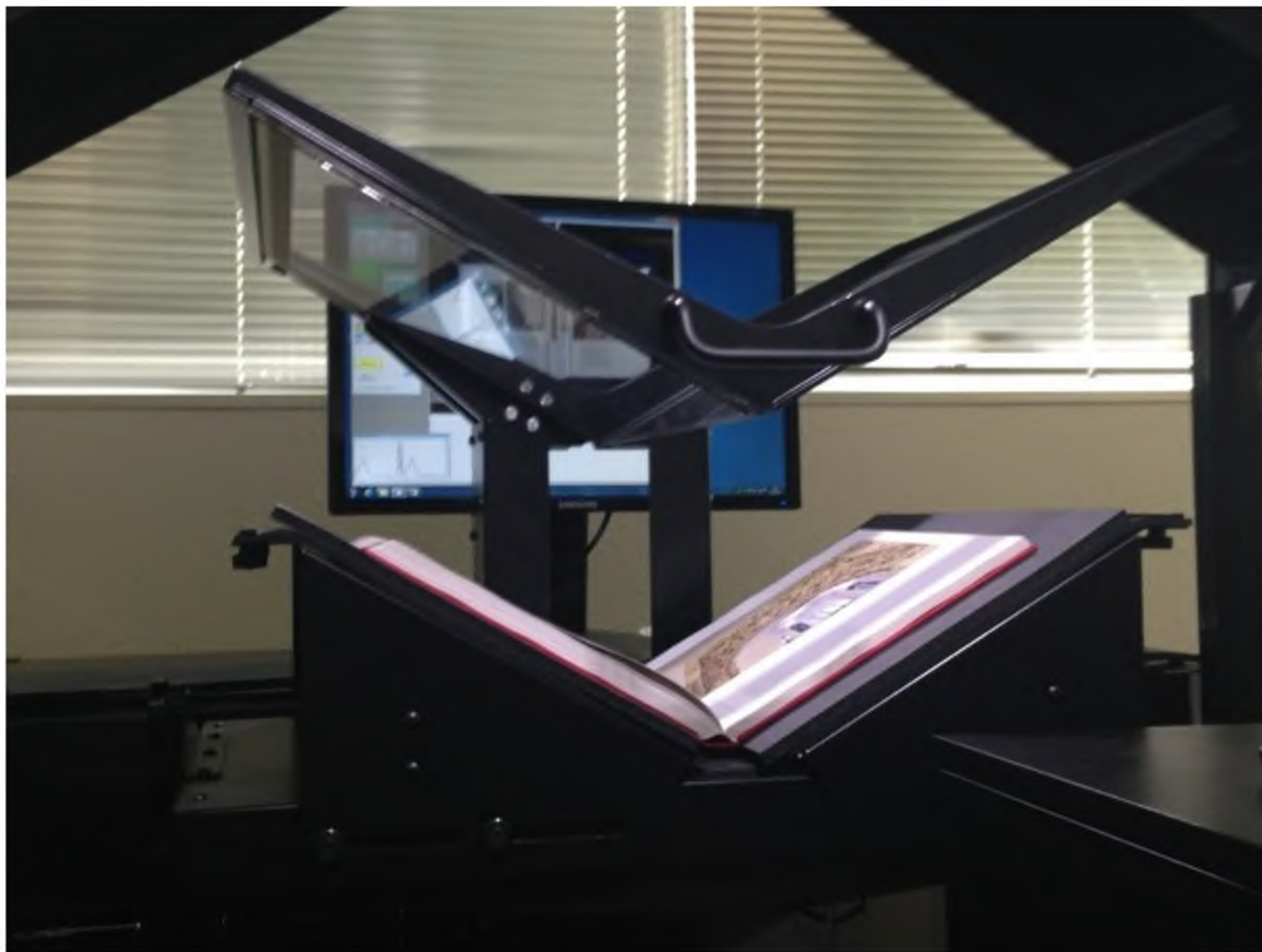
Uniwersytet
Wrocławski

JAK DZIAŁA HUMANISTYKA CYFROWA?

DH workflow (1)



Digitalizacja





Między Gdynią a Gdańskiem, wojska radzieckie doszły do wybrzeża Morza Bałtyckiego

Otoczone ugrupowanie Niemców rozbite na dwie części

Komunikat

Rodzickiego Biura Informacyjnego z dnia 23-go marca

[illegible]

48. wieża frontu wazygi do siemow ponad 16 słonecznych żelaznicy i alufazw.

Na Skaplo, na północny azchid, od RACH
KZA, wieża znowa, w wyższej wieży obfite
rych, szelwity miano JERNIA i szelwity miano
wasi HEDERSDOFF, LINDENGOFF,
PERDORF, HIGCHORNO, BAKENRE, K
SCHENGOFF, SCHLECHTERHO, RACHW
KREYER, HONOFF, 22 marca, wieża 16
PONTENKREDO, 22 marca, wieża 16
ponad 166 słonecznych żelaznicy i alufazw.

22 marca na wszystkich frontach miano
miano 156 słonecznych żelazw. W
22 marca, wieża 16, 22 marca, wieża 16
szelwity miano 156 słonecznych żelazw.

Pola przed zatoką Frisch-Haf
usłane tysiącami hitlerowskich trupów

Jak donosi Radiotelewizyjne Biuro Informacji, w najbliższych tygodniach 2-ga Formacja Białoruskich Komandosów wyprawi się na terenach, gdzie żyją białoruskie społeczności, które zostały zniszczone przez hitlerowskie Niemcy. W czasie misji Białoruskie Komandosy będą odwiedzać i odwiedzać pod koniec wojny białoruskich mieszkańców, którzy nie mieli czasu na powrót do swoich domów.

W sierpniu i na początku lata przyszłego białoruskie komandosy będą odwiedzać białoruskie wioski, które zostały zniszczone przez hitlerowskie Niemcy. W czasie misji Białoruskie Komandosy będą odwiedzać i odwiedzać pod koniec wojny białoruskich mieszkańców, którzy nie mieli czasu na powrót do swoich domów.

Gwałtowny cios radzieckich bohaterów

Stachyris and vittatus are common here, also

W kierunku na Gólszok, jak
dawniej, Radzieckie Biuro Infor-
macyjne, wieśka Armii Czerw-
nej, kandydaci na stano-
wisko radzieckiego i arty-
lerzysty przy ogniu węgla-
kowi lotnictwa przesłano-
wano. (Ciekawostka)

Nieprerzinali ponow w toke
wally wełłine strone. Wollne
dudeckie polakow w stado
dane ponow 4000 alomach
infirmary i alomach. Na dno
moch i alomach. Na dno

Na północny zachód od Gdańska rozciąga się zachodnie wybrzeże. W ciągu dnia odbywa się 10 koncertów symfonicznych i orkiestrowo-wokalnych, wspaniałych słowno-musycznych i śpiewnych i tańca - mackowatech.

Następnie winna, zdaniem wyznawcy, być w klasach polonistycznych i skierować do polonistycznych przedmiotów Zappal, i

[illegible]

W szkole Górnickiej istnieje własny zespół artystyczny. W tym roku w konkursie w kategorii "Wzrost i rozwój" zdobył on pierwsze miejsce. W konkursie "Wzrost i rozwój" zdobył on pierwsze miejsce. W konkursie "Wzrost i rozwój" zdobył on pierwsze miejsce.

„Sławną jest droga żołnierzy, którzy przeszli od Wołgi do Odry“

Depesza marszałka Żukowa do chłopów jerszowskiego rejonu
saratowskiego obwodu

plaskiel. W ten sposób zostały przeważnie przez wodniczek dotychczas nadmienione linie kolonizacji Gdansk — Gdynia, wyprzedzających zaś grupa wojsk została rozbita na dwie mniejsze grupy: 1. Grupa wojsk, która została skierowana do Gdyni, 2. Grupa wojsk, która została skierowana do Gdansk.

Celulozowa arka nohowa
Inicjator XVII Zjazdu Partii
gł. jasnokowickiego rejonu, so-
łectwskiego obwodowego, zebrał
w ciągu trzech dni 170 tysięcy rubli
kasy samodzielnego, które podar-
ował dla inicjatorów i dowódców
1-go Frontu Białoruskiego. Po-
tęga kolchozistów podzieliła ob-
rotowy wyprawy 15 ton maszyn
wyprodukowanych przez pa-
ństwo.

[illegible]

Walki na Śląsku

Armia Czerwona kroczy
niepowstrzymanie naprzód

Na Marko, na polnocy, szedł od Kuchborska, wozika, co
dostawiał słomę, opiół nępczynie-
kowi i szepnął: wazod słomowy
Włokan, łona, łonakowi żenił
Czerwoni, szedłszy, młodziem-
stwo Kuchborskich, łona, na
graniec Niemiec z Czechami
wzięł.

Rudziory pszczołny i żółty,
ci wyprężył nępczynie, a młoda
z Jermian, wzięła kółkowską
w którym szedłszy, na trzy waz-
ne linie kółkowskie i kłosa, waz-

...Tych dwóch podobieństw kochanek depens, podjętych przez dwudzięty I-go Fundus Miasteczka, Marzenna, Związek Miasteczka Złotowa. Oni, 1944).

Dołączony jest do machówki.
Palać nie będą zabójstwa nierz-
wypowiedzieli braci, którzy bez
względnego prawnego na fionie
prawy.

Pytanie nie o nasze życie. Na-
ze bawie życie nie, pójde.
Wiem, że jest, że jest, że jest,
które pójde do Włoch do Ody-
ry i schodzą, które natężył
czytelnik w naszym kierunku lego-

wych mianem w pory.
Ciepło nie dla bliznie. Zrobił
współnie, ale natężył nie jak
machówki.

Dopiero można zobaczyć na
zobaczyć, na których kochanku
nie stępił wyrażone powieści
współnie, ale natężył nie jak
machówki, ale natężył nie jak
machówki, ale natężył nie jak
machówki.

Anotacja

Trudności z indeksacją zawartości

To nie karnawał
Ale tańczyć chcę
I będę tańczył z nią po dzień
To nie zabawa
Ale bawię się
Bezsenne noce senne dnieee

To nie kochanka
Ale śpiam z nią
Choć śmieją ze mnie się i drwią

Taka zmęczona [ona]
I pijana wciąż [ona]
Dlatego nie
Nie pytaj więcej mnie

Dlaczego jestem z nią
(Nie pytaj mnie)
Dlaczego z inną nie
(Nie pytaj mnie)
Dlaczego myślę że
Że nie ma dla mnie innych miejsc
(Nie pytaj mnie)
co ciągle widzę w niej

(Nie pytaj mnie)
Dlaczego w innej nie
(Nie pytaj mnie)
Dlaczego ciągle chcę
Zasypiać w niej i budzić się

Te brudne dworce
Gdzie spotykam ją
Te tłumy które cicho klną
Ten pijak który mruczy coś przez sen
Że póki my żyjemy ona żyje też

Nie pytaj mnie
Nie pytaj mnie
Co widzę w niej

(Nie pytaj mnie)
co ciągle widzę w niej
(Nie pytaj mnie)
Dlaczego w innej nie
(Nie pytaj mnie)
Dlaczego ciągle chcę
Zasypiać w niej i budzić się

Trudności z indeksacją zawartości: koreferencja desygnatu "Polska"



To nie karnawał
Ale tańczyć chcę
I będę tańczył z **nią** po dzień
To nie zabawa
Ale bawię się
Bezsenne noce senne dnieee

To nie **kochanka**
Ale śpiam z **nią**
Choć śmieją ze mnie się i drwią

Taka zmęczona [**ona**]
I pijana wciąż [**ona**]
Dlatego nie
Nie pytaj więcej mnie

Dlaczego jestem z **nią**
(Nie pytaj mnie)
Dlaczego z inną nie
(Nie pytaj mnie)
Dlaczego myślę że
Że nie ma dla mnie innych miejsc
(Nie pytaj mnie)
co ciągle widzę w **niej**

(Nie pytaj mnie)
Dlaczego w innej nie
(Nie pytaj mnie)
Dlaczego ciągle chcę
Zasypiać w **niej** i budzić się

Te brudne dworce
Gdzie spotykam **ją**
Te tłumy które cicho klną
Ten pijak który mruczy coś przez sen
Że póki my żyjemy **ona** żyje też

Nie pytaj mnie
Nie pytaj mnie
Co widzę w **niej**

(Nie pytaj mnie)
co ciągle widzę w **niej**
(Nie pytaj mnie)
Dlaczego w innej nie
(Nie pytaj mnie)
Dlaczego ciągle chcę
Zasypiać w **niej** i budzić się

Anotacja

Tagowanie

<http://ws.clarin-pl.eu/tager.shtml>

Pierwsze wzmianki o Via Regia pochodzą z 1252, kiedy pojawia się pod nazwą „Strata Regia” (Droga Królewska) w dokumentach margrabiego Marchii Miśnieńskiej Henryka III.

```
<?xml version="1.0" encoding="UTF-8"?><!DOCTYPE chunkList SYSTEM "ccl.dtd"><chunkList> <chunk id="ch1" type="p"> <sentence id="s1"> <tok> <orth>Pierwsze</orth> <lex disamb="1"><base>pierwsza</base><ctag>subst:pl:nom:f</ctag></lex> </tok> <tok> <orth>wzmianki</orth> <lex disamb="1"><base>wzmianka</base><ctag>subst:sg:gen:f</ctag></lex> </tok> <tok> <orth>o</orth> <lex disamb="1"><base>o</base><ctag>prep:acc</ctag></lex> </tok> <tok> <orth>Via</orth> <lex disamb="1"><base>via</base><ctag>prep:nom</ctag></lex> </tok> <tok> <orth>Regia</orth> <lex disamb="1"><base>Regia</base><ctag>ign</ctag></lex> </tok> <tok> <orth>pochodzą</orth> <lex disamb="1"><base>pochodzić</base><ctag>fin:pl:ter:imperf</ctag></lex> </tok> <tok> <orth>z</orth> <lex disamb="1"><base>z</base><ctag>prep:gen:nwok</ctag></lex> </tok> <tok> <orth>1252</orth> <lex disamb="1"><base>1252</base><ctag>num:pl:nom:m1:rec</ctag></lex> </tok> <ns/> <tok> <orth>,</orth> <lex disamb="1"><base>,</base><ctag>interp</ctag></lex> </tok> <tok> <orth>kiedy</orth> <lex disamb="1"><base>kiedy</base><ctag>comp</ctag></lex> </tok> <tok> <orth>pojawia</orth> <lex disamb="1"><base>pojawiać</base><ctag>fin:sg:ter:imperf</ctag></lex> </tok> <tok> <orth>się</orth> <lex disamb="1"><base>się</base><ctag>qub</ctag></lex> </tok> <tok> <orth>pod</orth> <lex disamb="1"><base>pod</base><ctag>prep:acc:nwok</ctag></lex> </tok>
```

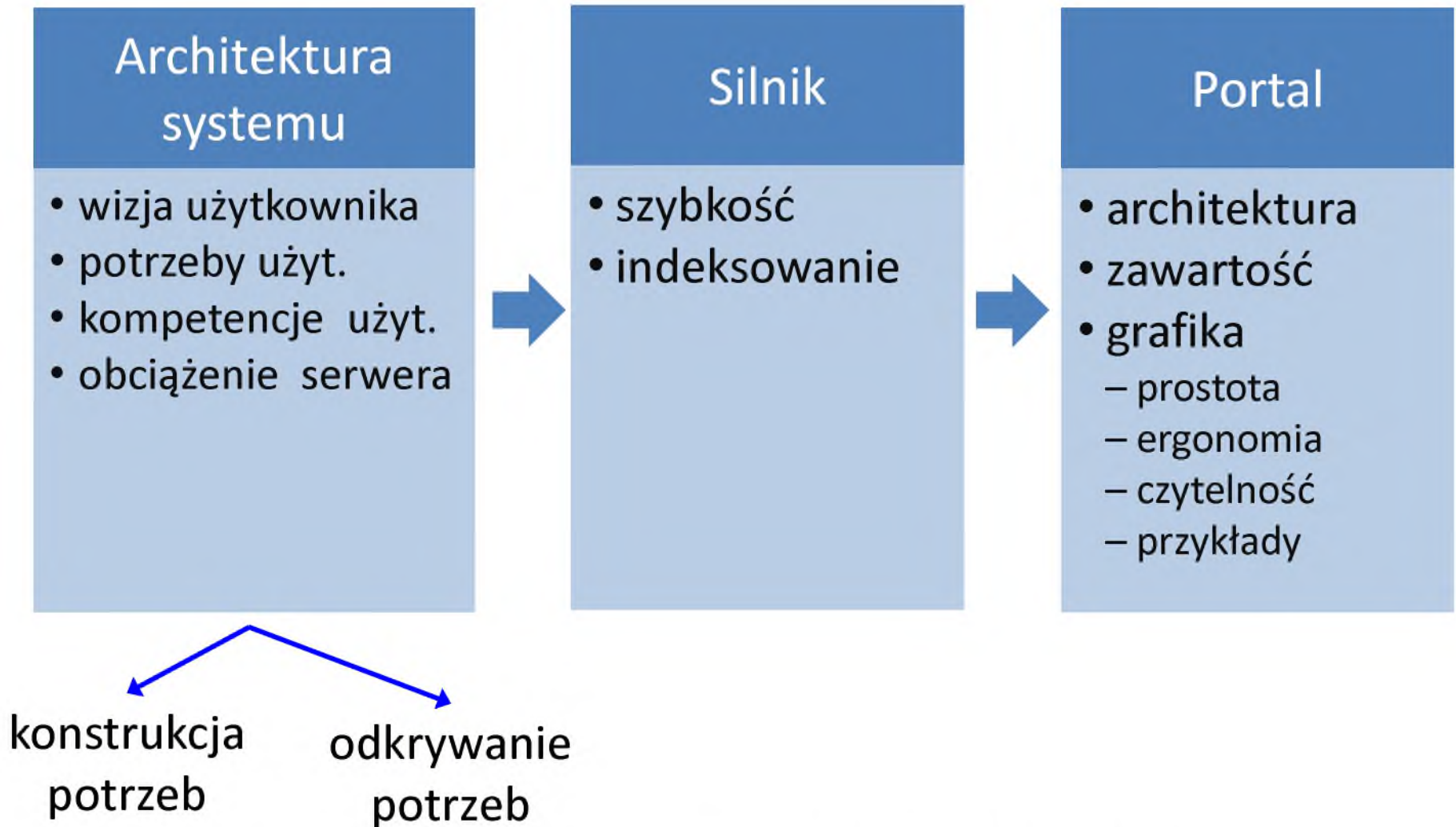


```
"date": "1952-03-28T23:00:00Z",
  "period": "d",
  "medium": "paper",
  "title_a": "Mocarstwa zachodnie wypowiadają się przeciwko
    traktatowi pokojowemu i za dalszą remilitaryzacją Trizonii",
  "text_id": "Bkrm",
  "utt_tagged": "<s id='3'>
<wts>
<wt id='\" seq='1'>
  <w>Nota</w>
  <ps>
    <p l='nota' a='noun' c='subst' r='1'>subst:sg:nom:f</p>
  </ps>
</wt>
<wt id='\" seq='2'>
  <w>mocarstw</w>
  <ps>
    <p l='mocarstwo' a='noun' c='subst' r='1'>subst:pl:gen:n</p>
  </ps>
</wt>
<wt id='\" seq='3'>
  <w>zachodnich</w>
  <ps>
    <p l='zachodni' a='adj' c='adj' r='1'>adj:pl:gen:m3:pos</p>
  </ps>
</wt>
<wt id='\" seq='4'>
  <w>odrzuca</w>
  <ps>
    <p l='odrzucać' a='verb' c='fin' r='1'>fin:sg:ter:imperf</p>
  </ps>
</wt>
```

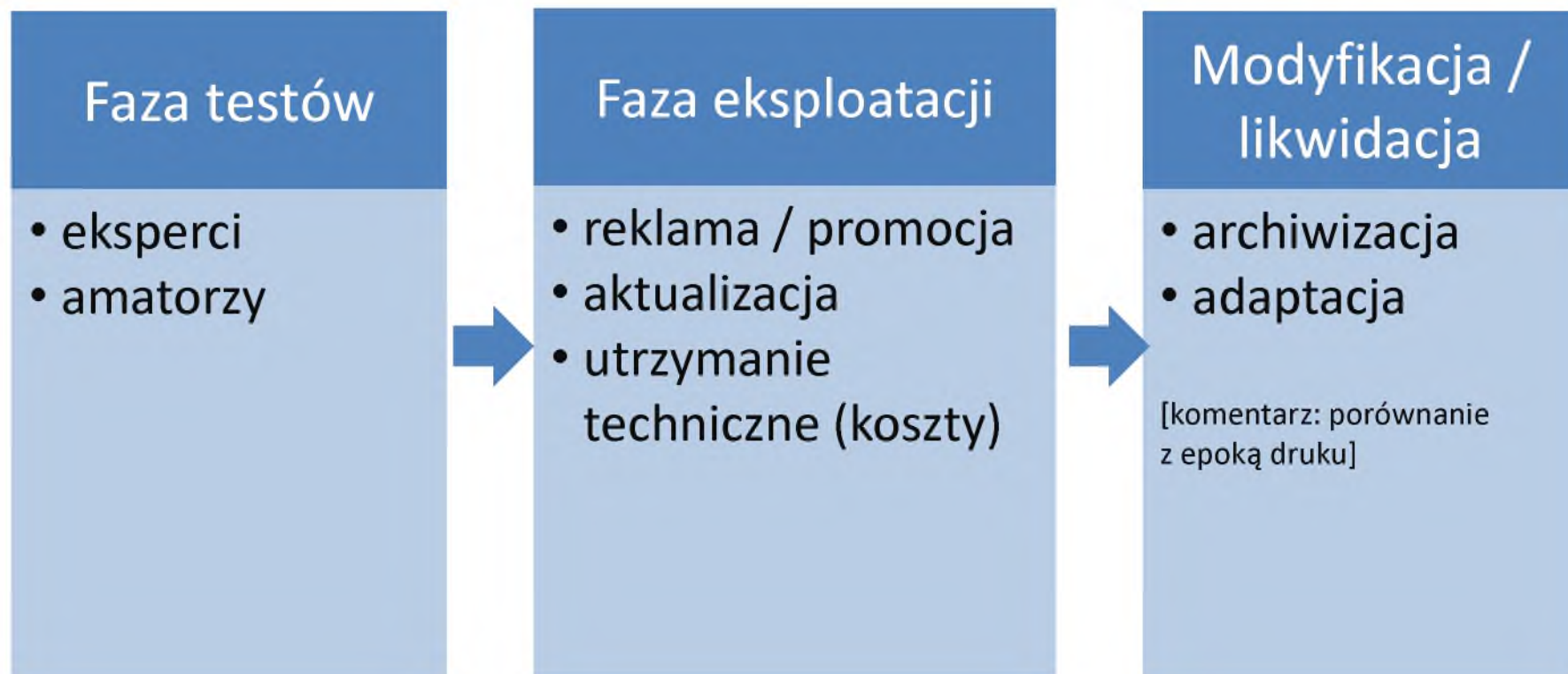
```
<wt id='\" seq='5'>
  <w>propozycje</w>
  <ps>
    <p l='propozycja' a='noun' c='subst'
r='1'>subst:pl:acc:f</p>
  </ps>
</wt>
<wt id='\" seq='6'>
  <w>radzieckie</w>
  <ps>
    <p l='radziecki' a='adj' c='adj'
r='1'>adj:pl:acc:f:pos</p>
  </ps>
</wt>
<wt id='\" seq='7'>
  <w>.</w>
  <ps>
    <p l='.' a='punct' c='interp' r='1'>interp</p>
  </ps>
</wt>
</wts>
</s>",
  "utt": "Nota mocarstw zachodnich odrzuca
propozycje radzieckie .",
  "title_j": "Trybuna Ludu",
  "style": "press",
  "id": "eaWbe",
  "seq": 3,
  "status": "1_obieg",
  "authors": [
    "" ],
```

flexeme	abbreviation	base form	example
noun	<i>subst</i>	singular nominative	<i>profesor</i>
depreciative form	<i>depr</i>	singular nominative form of the corresponding noun	<i>profesor</i>
main numeral	<i>num</i>	inanimate masculine nominative form	<i>pięć, dwa</i>
collective numeral	<i>numcol</i>	inanimate masculine nominative form of the main numeral	<i>pięć, dwa</i>
adjective	<i>adj</i>	singular nominative masculine positive form	<i>polski</i>
ad-adjectival adjective	<i>adja</i>	singular nominative masculine positive form of the adjective	<i>polski</i>
post-prepositional adjective	<i>adjp</i>	singular nominative masculine positive form of the adjective	<i>polski</i>
predicative adjective	<i>adjc</i>	singular nominative masculine positive form of the adjective	<i>zdrowy, ciekawy</i>
adverb	<i>adv</i>	positive form	<i>dobrze, bardzo</i>
non-3rd person pronoun	<i>ppron1 2</i>	singular nominative	<i>ja</i>
3rd-person pronoun	<i>ppron3</i>	singular nominative	<i>on</i>
pronoun <i>SIEBIE</i>	<i>siebie</i>	accusative	<i>siebie</i>
non-past form	<i>fin</i>	infinitive	<i>czytać</i>
future <i>BYĆ</i>	<i>bedzie</i>	infinitive	<i>być</i>
agglutinate <i>BYĆ</i>	<i>aglt</i>	infinitive	<i>być</i>
I-participle	<i>praest</i>	infinitive	<i>czytać</i>
imperative	<i>impt</i>	infinitive	<i>czytać</i>
impersonal	<i>imps</i>	infinitive	<i>czytać</i>
infinitive	<i>inf</i>	infinitive	<i>czytać</i>
contemporary adv. participle	<i>pcon</i>	infinitive	<i>czytać</i>
anterior adv. participle	<i>pant</i>	infinitive	<i>czytać</i>
gerund	<i>ger</i>	infinitive	<i>czytać</i>
active adj. participle	<i>pact</i>	infinitive	<i>czytać</i>
passive adj. participle	<i>ppas</i>	infinitive	<i>czytać</i>
winien	<i>winien</i>	singular masculine form	<i>powinien, rad</i>
predicative	<i>pred</i>	the only form of that flexeme	<i>warto</i>
preposition	<i>prep</i>	the non-vocalic form of that flexeme	<i>na, przez, w</i>
coordinating conjunction	<i>conj</i>	the only form of that flexeme	<i>oraz</i>
subordinating conjunction	<i>comp</i>	the only form of that flexeme	<i>że</i>
particle-adverb	<i>qub</i>	the only form of that flexeme	<i>nie, -że, się</i>
abbreviation	<i>brev</i>	the full dictionary form	<i>rok, i tak dalej</i>
bound word	<i>burk</i>	the only form of that flexeme	<i>trochu, oścież</i>
interjection	<i>interj</i>	the only form of that flexeme	<i>ech, kurde</i>
punctuation	<i>interp</i>	the only form of that flexeme	<i>! , , , [,]</i>
alien	<i>xxx</i>	the only form of that flexeme	<i>cool , nihil</i>
unknown form	<i>ign</i>	the only form of that flexeme	

DH workflow (2)



DH workflow (3)





Uniwersytet
Wrocławski

PRZYKŁADY NARZĘDZI I BADAŃ HUMANISTYKI CYFROWEJ



chronopress.clarin-pl.eu/

[słowosieć](#)

frazeo.pl/

www.google.pl/trends/

books.google.com/ngrams/

www.deutschestextarchiv.de

spraakbanken.gu.se/

Słownosieć języka polskiego

WordnetLoom 2.0.3

File Edit Other Window Help

Candidates Synsets

Lexical units

Search:

Parts of speech:

Domain:

Relations:

Search

Lexical units:

film 1 (artefact)

film 3 (artefact)

film 4 (artefact)

film 3D 1 (artefact)

film akcji 1 (artefact)

film aktorski 1 (artefact)

film amatorski 1 (artefact)

film animowany 1 (artefact)

film antywojenny 1 (artefact)

film autorski 1 (artefact)

film awanturniczy 1 (artefact)

film barwny 1 (artefact)

film dla dorosłych 1 (artefact)

film dokumentalny 1 (artefact)

film dramatyczny 1 (artefact)

film drogi 1 (artefact)

film dźwiękowy 1 (artefact)

film długometrażowy 1 (artefact)

film fabularny 1 (artefact)

film fotograficzny 1 (artefact)

film historyczny 1 (artefact)

film interaktywny 1 (artefact)

Items: 57

General memory usage: 14 MB

film #1

produkcja 1

człowiek 1

ambony 4

animowany 1

archiwalny 2

19 ...

film 2

ekranizacja 1

film 3D 1

film amatorski 1

film autorski 1

31 ...

movie 1

coming attraction 1

kinoteka 1

anamorfoza 1

5 ...

Preview

Synset Properties

Lexical units in synset:

film 1 (artefact)

obraz 6 (artefact)

Comment:

n/d

Lexical unit relations

film 1 (artefact)

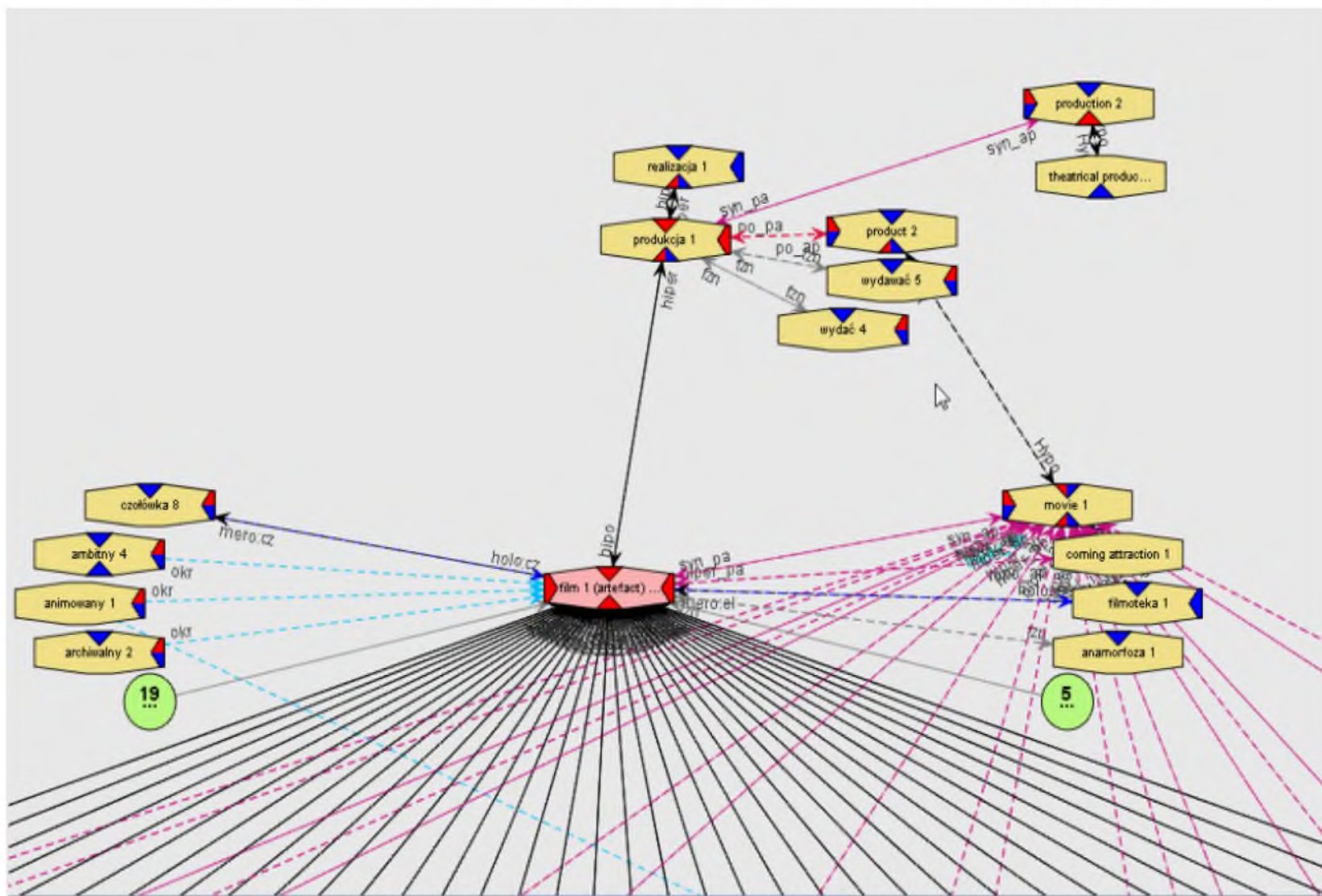
From

To

agents_with_hidden_predica

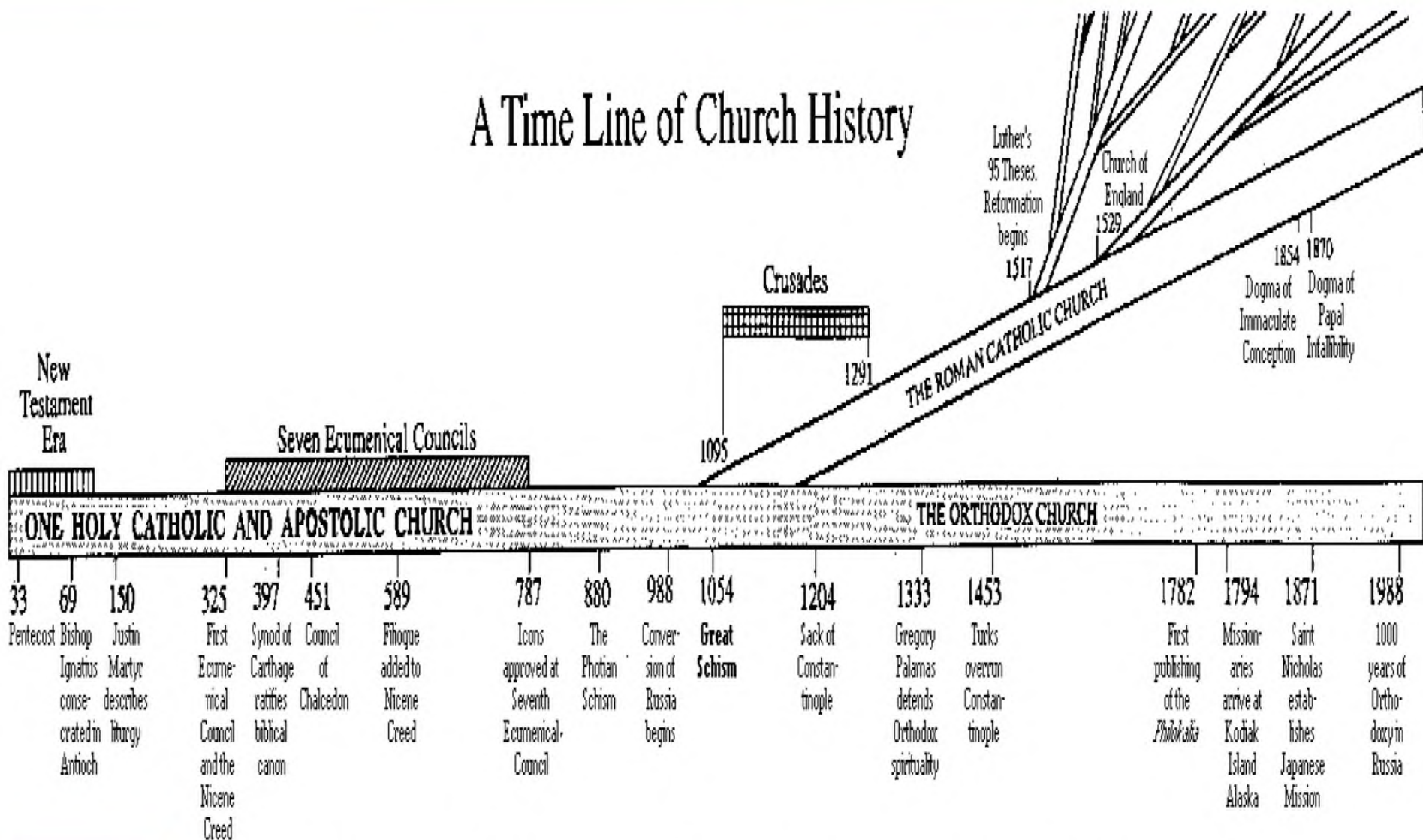
filmowiec 1 (person)

Słownosieć języka polskiego



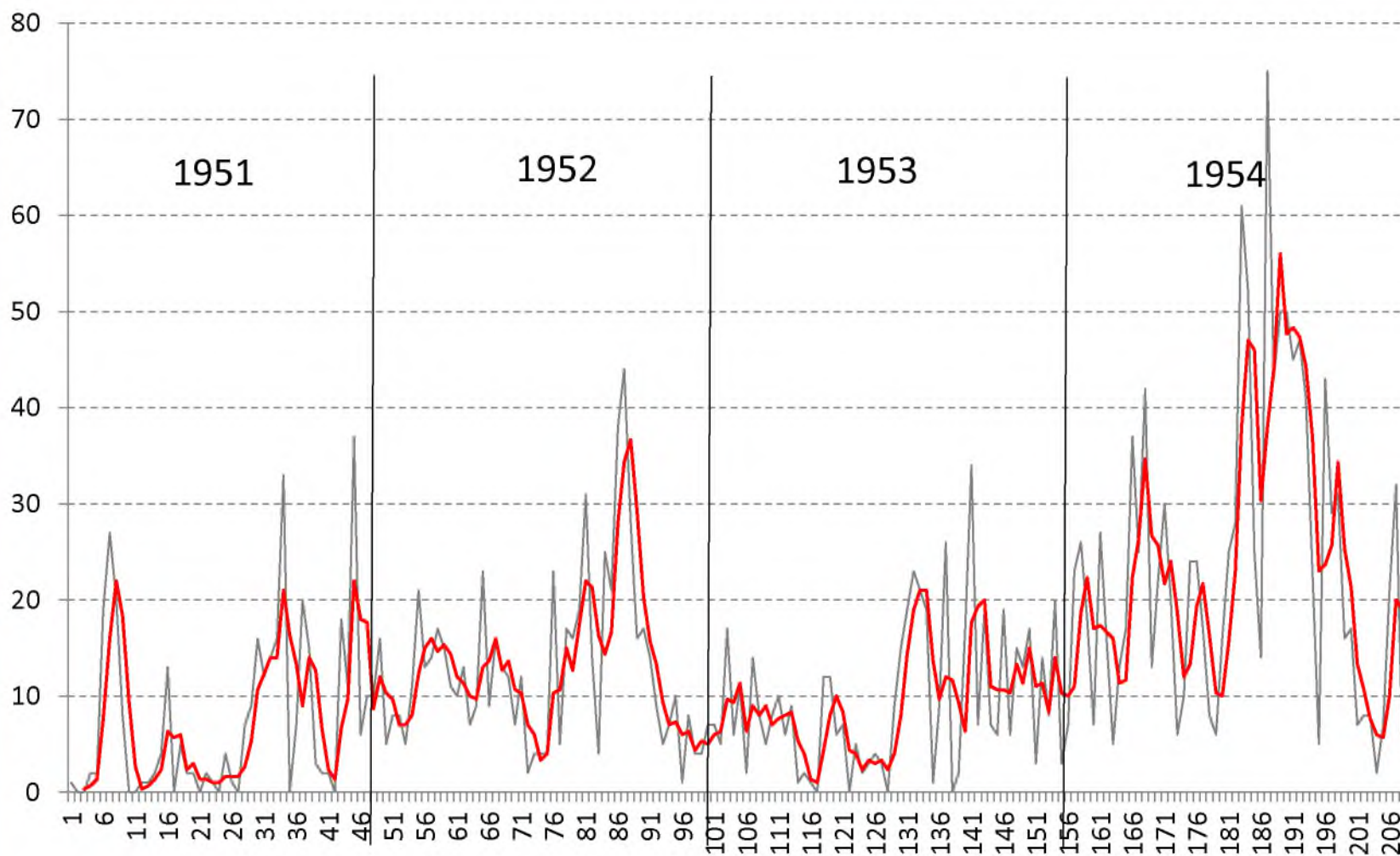


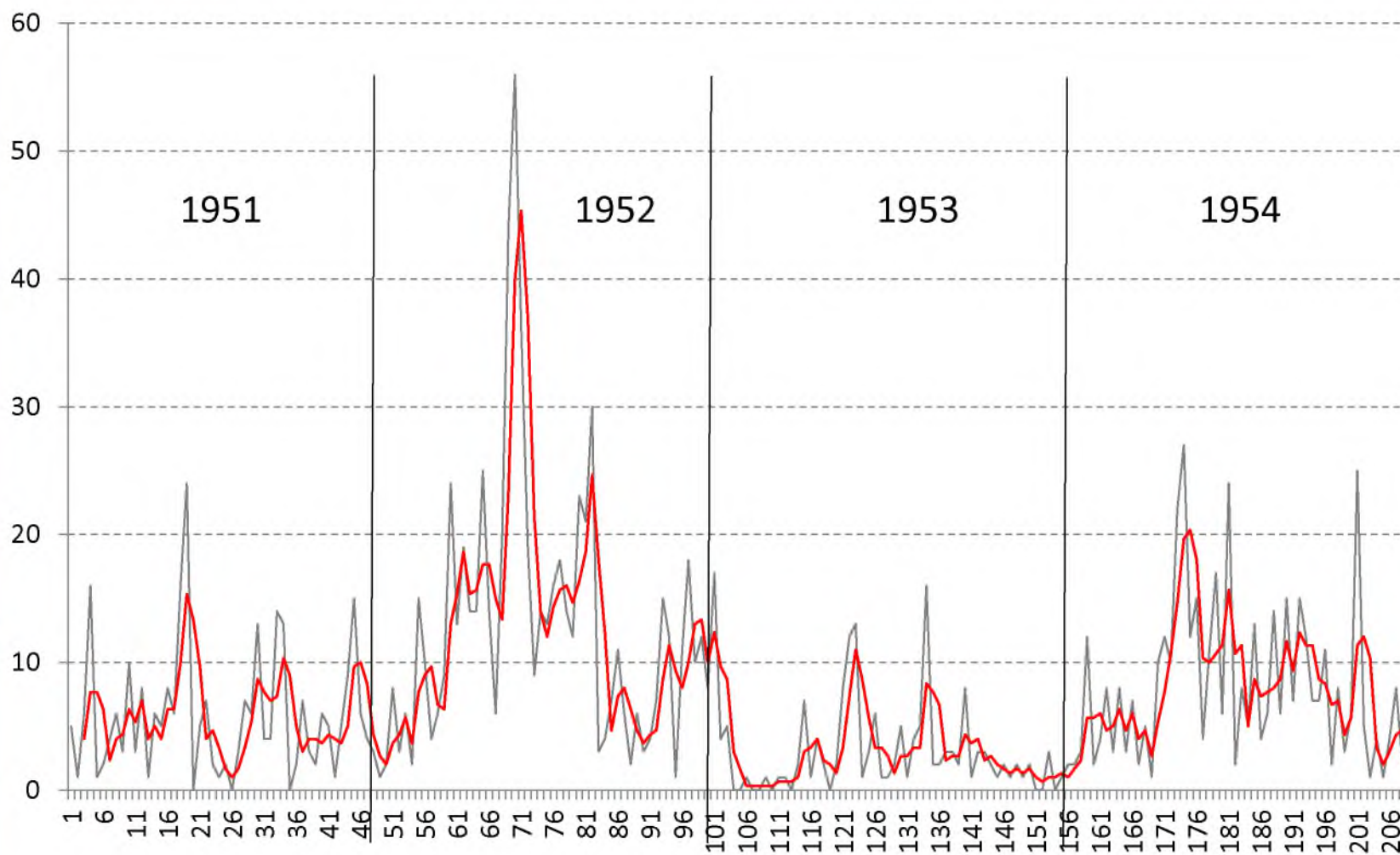
A Time Line of Church History





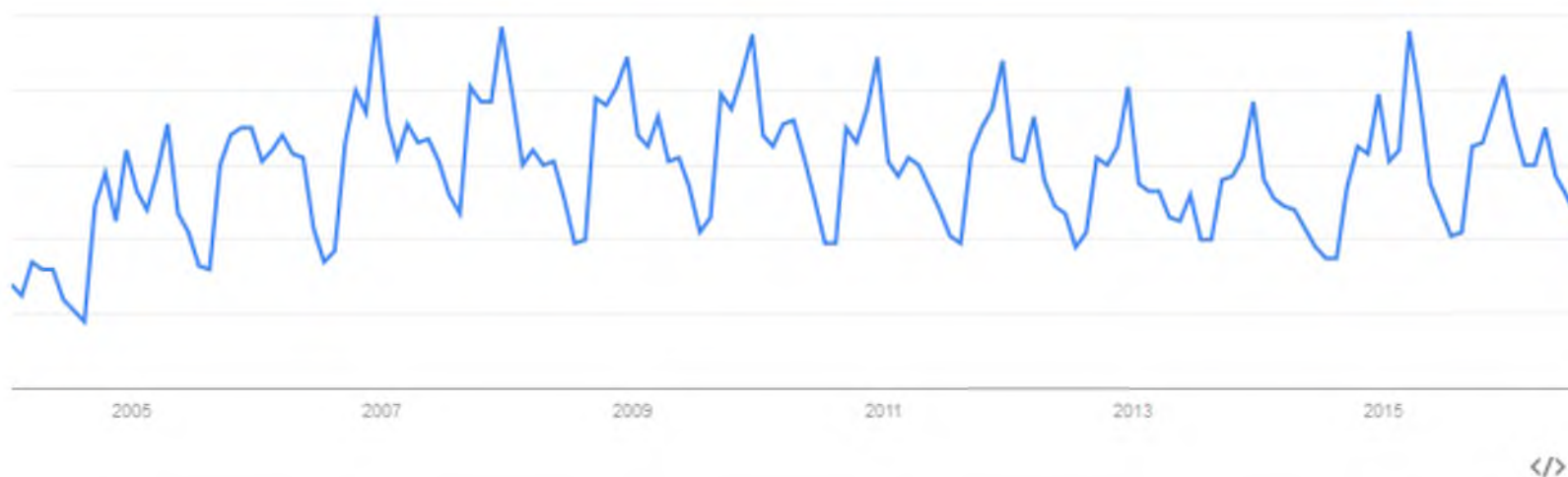
OSCYLACJE PERIODYCZNE: POLE SEMANTYCZNE ROLNICTWA







Zapytania w wyszukiwarce Google



maj ✕

październik ✕

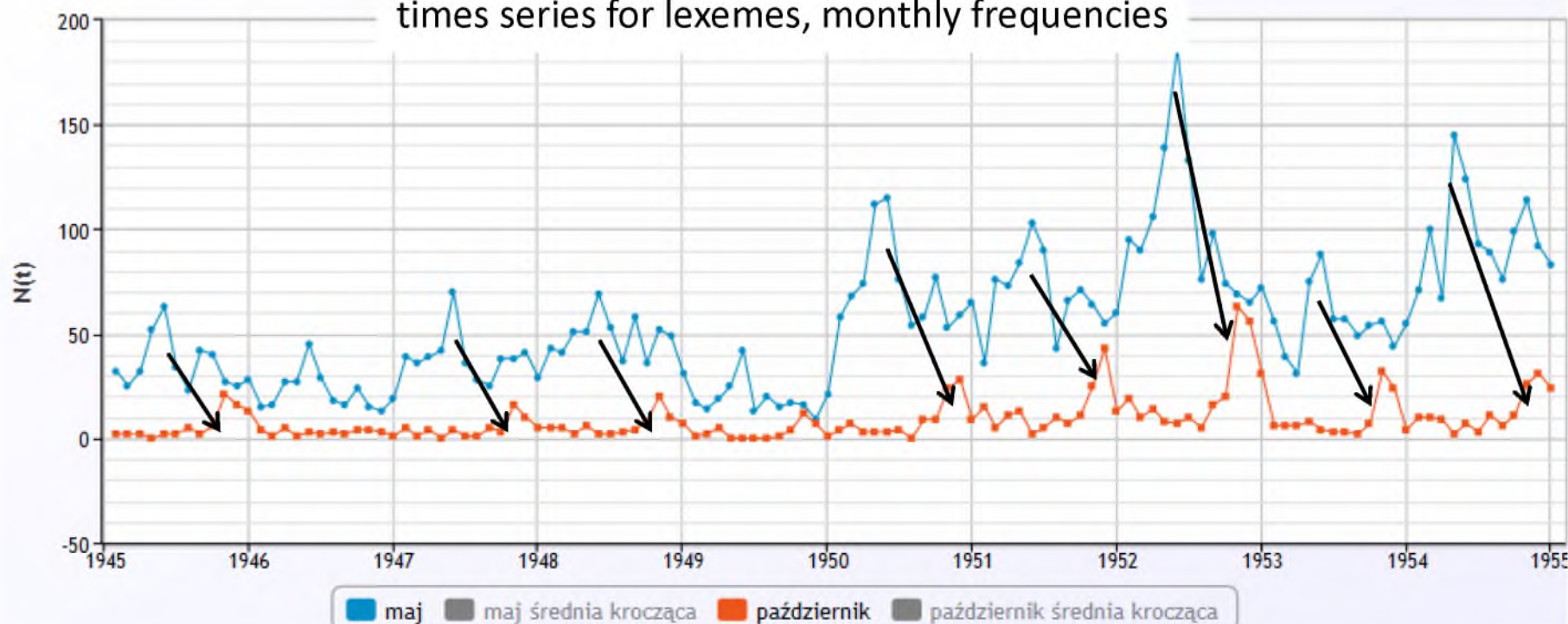
enter lemma(s)



Szereg czasowy dla słów

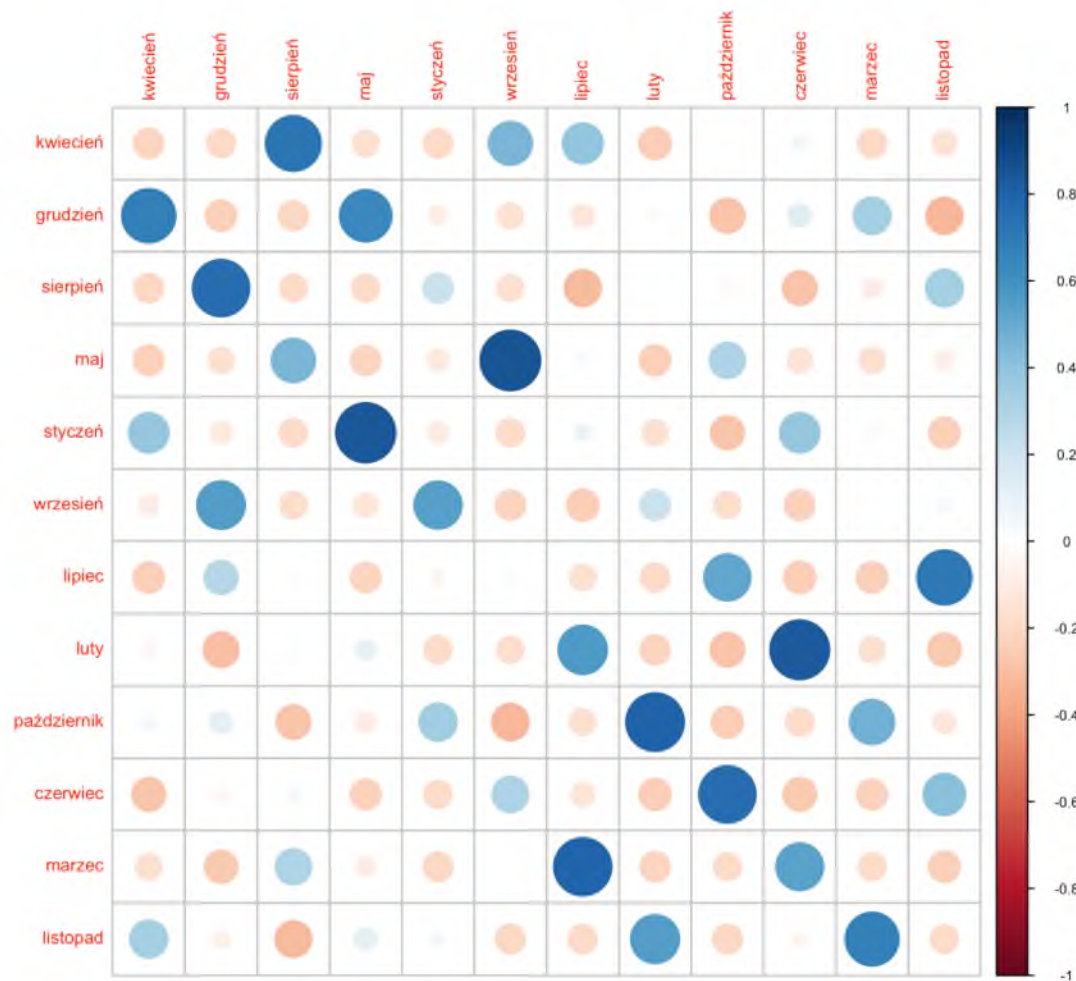


times series for lexemes, monthly frequencies



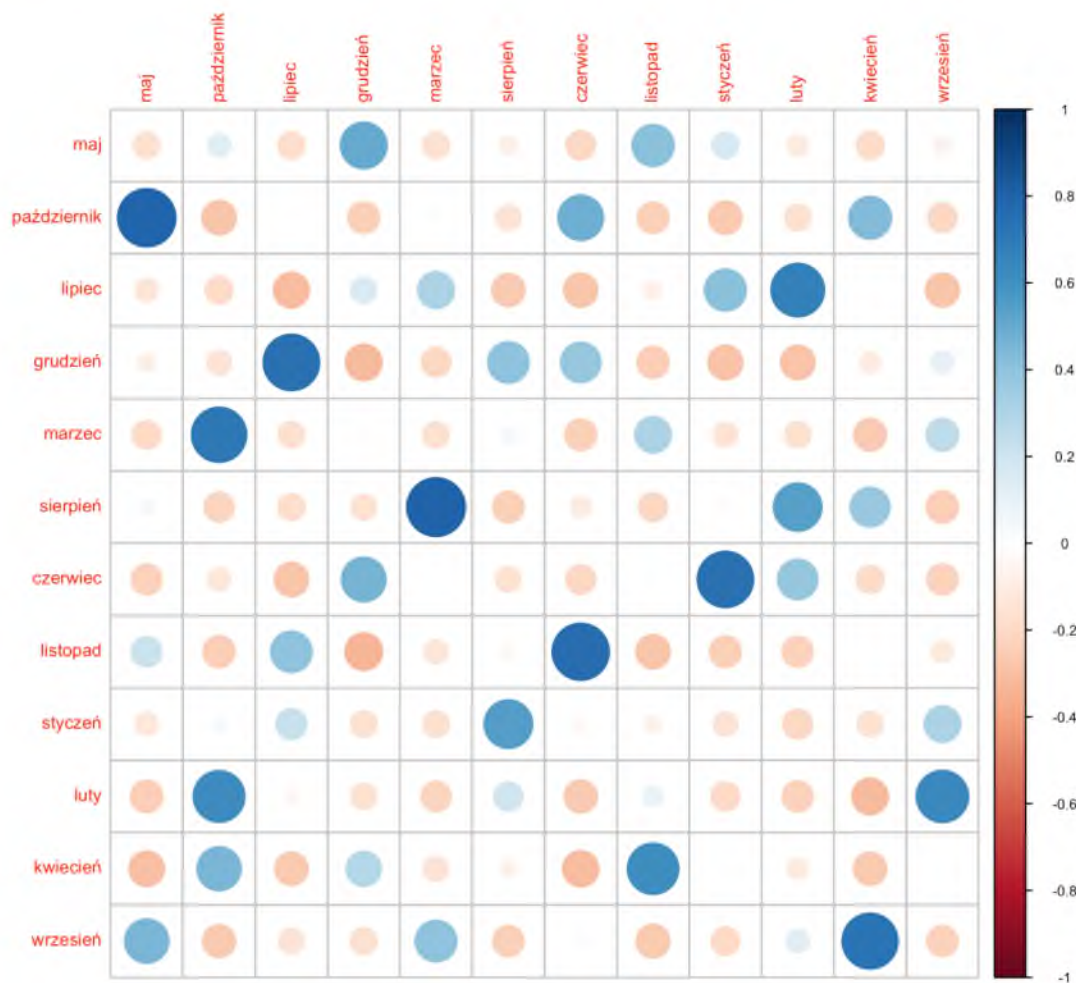


MATRYCA KORELACJI NAZW MIESIĘCY ($k=-4$)





MATRYCA KORELACJI NAZW MIESIĘCY ($k=5$)





Cykl epok czyli literaturoznawcza mniemanologia stosowana



<http://corpus.byu.edu>

CORPUS OF HISTORICAL AMERICAN ENGLISH

400 MILLION WORDS, 1810-2009 [OFFLINE DATA]

EMAIL

PASSWORD

(HELP) [LOG IN](#) ([REGISTER](#))

DISPLAY

CLICK ON BARS FOR CONTEXT

CLICK ON COLUMN HEADINGS FOR FREQUENCY IN SUB-SECTION

[COMPARE](#) [?](#) [SIDE](#)

[LIST](#)

[CHART](#)

[KWIC](#)

[COMPAR](#)

[SEARCH](#)

[STRING](#)

[WORD](#)

[COLLO](#)

[POS](#)

[LIST](#)

[RANDO](#)

SECTION	1810	1820	1830	1840	1850	1860	1870	1880	1890	1900	1910	1920	1930	1940	1950	1960	1970	1980	1990	2000
FREQ	0	21	19	44	23	18	15	21	37	28	84	89	171	142	188	198	178	377	143	62
PER MIL	0.00	3.03	1.38	2.74	1.40	1.06	0.81	1.03	1.80	1.27	3.70	3.47	6.95	5.83	7.66	8.26	7.47	14.89	5.12	2.10

SEE ALL
YEARS
AT ONCE

SECTION
1940

TOKENS
142

SIZE
24,347,955

PER MILLION
5.83

KEYWORD IN CONTEXT DISPLAY

[Help / information / contact](#)

SECTION: 1940 (142)

PAGE: << < 1 / 2 > >>

SAMPLE: 100

CLICK FOR MORE CONTEXT

☐ [?]

1	1940	FIC	Harpers	A B C	Barowski was an important Pole; one of the farmers who had been settled by Warsaw on former German estates in the Corridor. He was a smart, powerful man
2	1940	FIC	Harpers	A B C	boat with me. Her friend was on it -- she had written Dunya from Warsaw and in her letter had made remarks about Poland; Dunya did not want to
3	1940	FIC	Harpers	A B C	Gillian. The captain introduced us on the first night out. She came from Warsaw , after a trip through the Balkans, and she was worried about her Polish
4	1940	FIC	Harpers	A B C	her that Dunya had been hurt by something she, Jill, had written from Warsaw but she only smiled. By then I had fallen for her like a ton
5	1940	FIC	Harpers	A B C	of us was thinking of the real Dunya any more. Poland was through; Warsaw fell and then Gdynia, and it would have been quite possible to make inquiries
6	1940	MAG	Time	A B C	Weygand, who in 1920 helped the Poles to crush the Red Army offensive against Warsaw . Under General Weygand in Syria today is a French Army of some 20 di
7	1940	MAG	Time	A B C	, long record (he joined the Foreign Service in 1919, has served at Warsaw , Berne, Tokyo, Constantinople, Sydney) prevent Jay Pierrepont Moffat from being
8	1940	MAG	Time	A B C	LADY IN WARTIME-E. M. Delafield-Harper (\$2.50). # A travelogue caught in Warsaw , a clever novelist at home in England, might be expected to present diverge
9	1940	MAG	Time	A B C	hopeless September.They do. # Julien Bryan's film of the " siege " of Warsaw was released some weeks ago. His book contains 48 photographs, some of them
10	1940	MAG	Time	A B C	, some of them of scenes omitted from the film. The only cameraman in Warsaw at the time, Bryan tells in words part of what he saw and how

TREND: ENTROPIA JAKO ZMIENNA LOSOWA

Dane dla języka chińskiego, okres ok. 45 lat, pochodzące z dokumentu opisującego stan państwa. Źródło: [...]

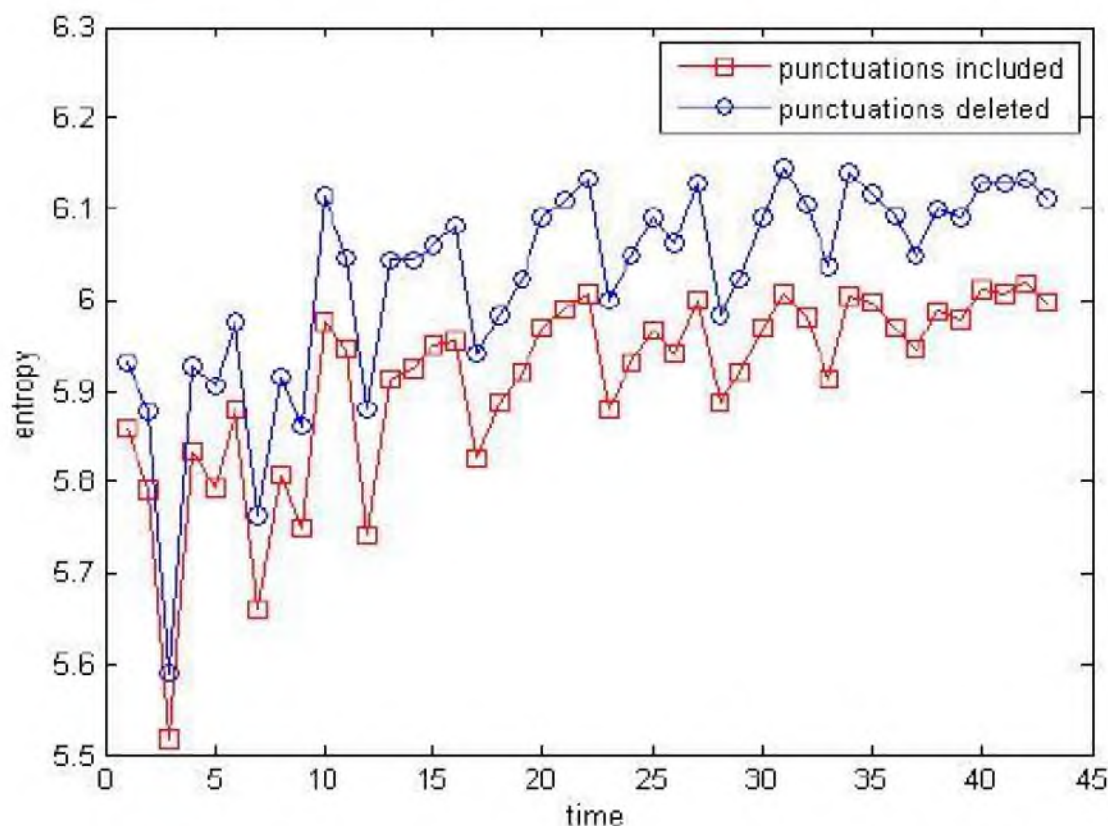
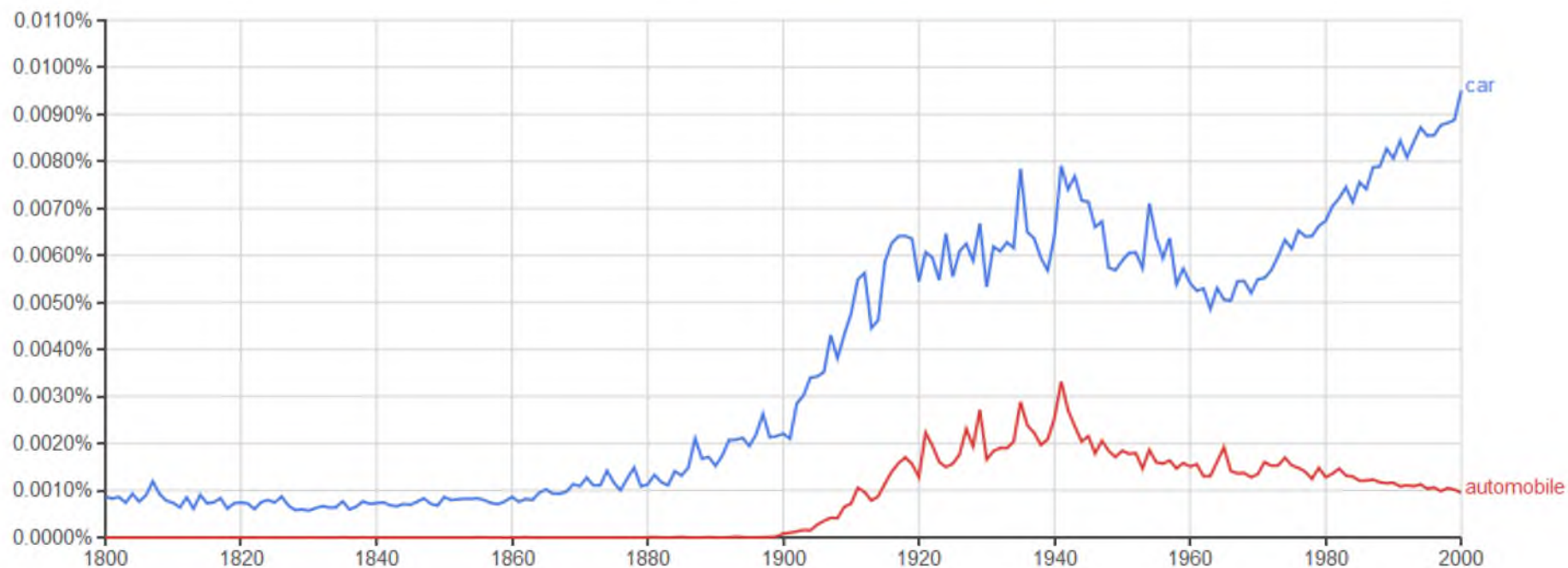


Figure 2. Scattered plot of the entropic process in time series

Google books Ngram Viewer

Graph these comma-separated phrases: ☐ case-insensitive
between and from the corpus with smoothing of [Search lots of books](#)



Google books Ngram Viewer

Graph these comma-separated phrases: ☐ case-insensitive
between and from the corpus with smoothing of [Search lots of books](#)





Google Ngram Viewer jako generator szeregów leksykalnych

Google books Ngram Viewer

Graph these comma-separated phrases: ☐ case-insensitive

between and from the corpus with smoothing of

[Search lots of books](#)





MODEL "KATASTROFY" W KORPUSIE CHRONOPRESS

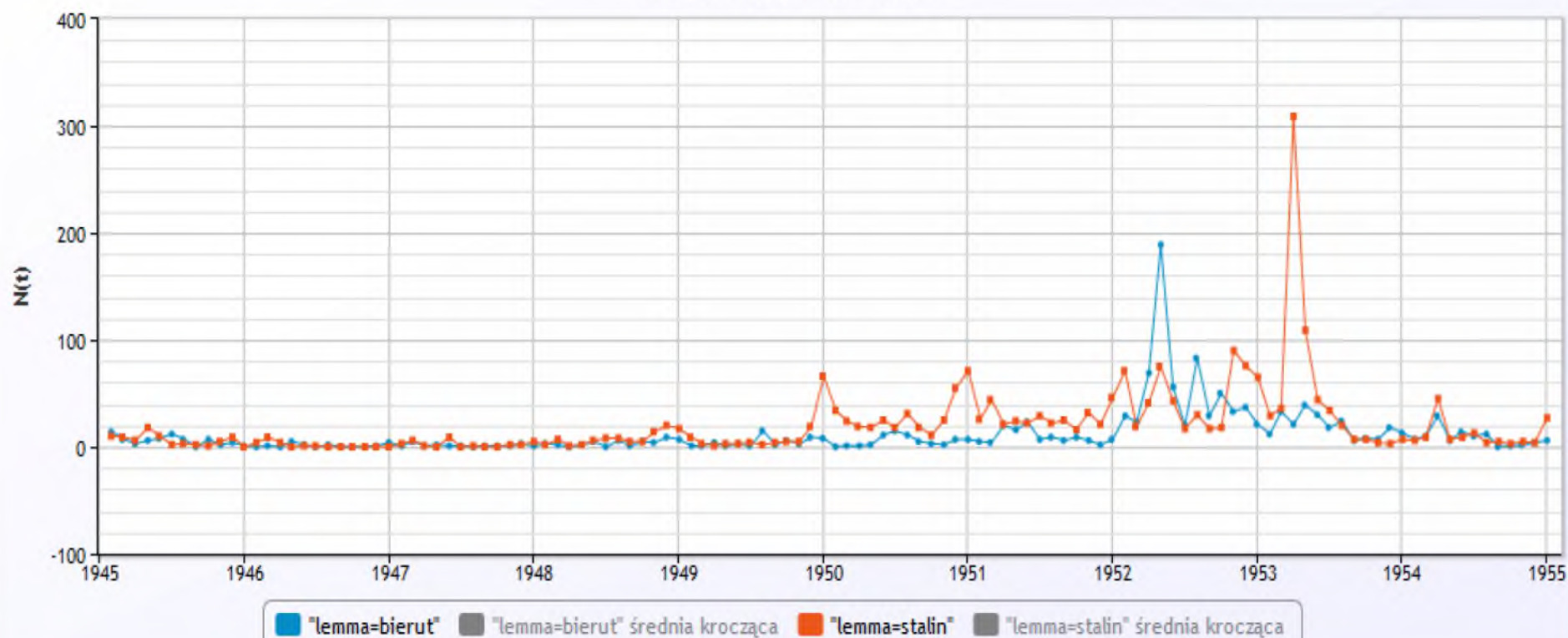
<lemma=bierut> x

<lemma=stalin> x

Query

Szereg czasowy dla słów

Częstości wystąpień w miesiącu



<lemma=chłop> x

<lemma=robotnik> x

<lemma=pracownik> x

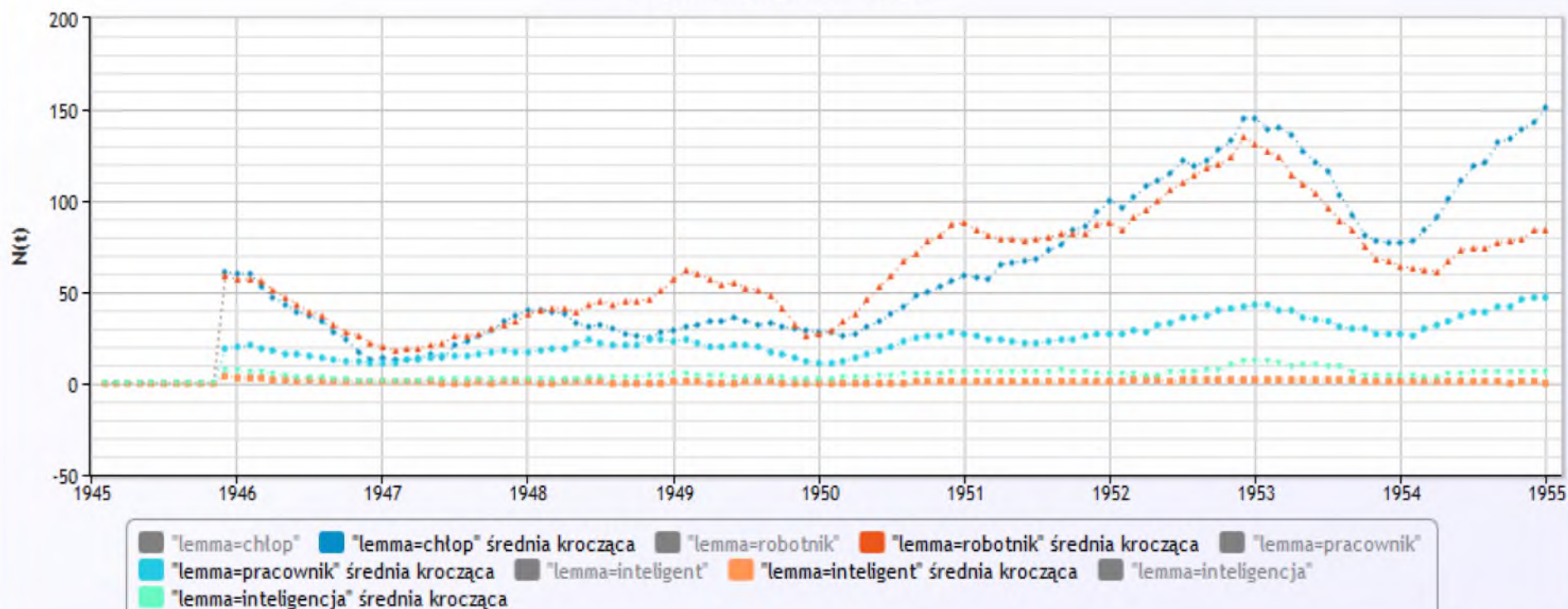
<lemma=inteligent> x

<lemma=inteligencja> x

Query

Szereg czasowy dla słów

Częstości wystąpień w miesiącu



Przykład "intelligentnej" biblioteki: Deutsches Textarchiv (DTA)

Anmelden (DTAQ)

DWDSdlexDBCLARIN





Texte ▼Projekt ▼Dokumentation ▼Impressum

☐ In den Titeldaten☒ Im Korpus☐ In der Dokumentation

Beispielanfragen: {Gott,Herr} && "das ewige Leben #2 geben" "Sp=ADJA Werkstoff|gn-asi" ehelichen with Sp=VVINF

Deutsches Textarchiv

GRUNDLAGE FÜR EIN REFERENZKORPUS DER NEUHOCHDEUTSCHEN SPRACHE

Das Deutsche Textarchiv stellt einen disziplinen- und gattungsübergreifenden Grundbestand deutschsprachiger Texte aus dem Zeitraum von ca. 1600 bis 1900 bereit. Die Textauswahl erfolgte auf der Grundlage einer von Akademiemitgliedern erstellten und ausführlich kommentierten, umfangreichen Bibliographie. In Ergänzung wurden einschlägige Literaturgeschichten und (Fach-)Bibliographien ausgewertet. Aus der Gesamtliste der auf diesem Wege ermittelten Titel wurde von der DTA-Projektgruppe ein hinsichtlich der repräsentierten Textsorten und Disziplinen ausgewogenes Korpus zusammengestellt ([weitere Informationen zur Textauswahl](#)).

Um den historischen Sprachstand möglichst genau abzubilden, werden als Vorlage für die Digitalisierung in der Regel die Erstausgaben der Werke zugrunde gelegt. Das elektronische Volltextkorpus des DTA ist über das Internet frei zugänglich und dank seiner Aufbereitung durch (computer-)linguistische Methoden schreibweisentolerant über den gesamten jeweils verfügbaren Bestand durchsuchbar. Sämtliche Texte stehen zum Download zur Verfügung. [mehr ...](#)

Das DTA in Zahlen

- 1 550 Werke
- 461 287 digitalisierte Seiten
- 110 229 535 fortlaufende Wortformen
- 771 088 217 Zeichen (Unicode)
- 764 weitere Werke in DTAQ

Das DTA am 12. November 2014



Am 12. November 1819 wurde Daniel Sanders in Strelitz (Mecklenburg) geboren. Sanders' Hauptwerk, das *Wörterbuch der deutschen Sprache mit Belegen von Luther bis auf die Gegenwart*, ist heute ein Standardwerk. Weiterhin wurde er durch seine heftige Kritik am Wörterbuchprojekt der Brüder Grimm bekannt, das er „in seiner ganzen Anlage und großentheils auch in seiner Ausführung durchaus verfehlt“ bezeichnete.



Sanders, Daniel: Aus der Werkstatt eines Wörterbuchschreibers. Plaudereien. Berlin, 1889.

Neue Werke im DTA

Lista tekstów w DTA

Anmelden (DTAQ) DWDS dlexDB CLARIN-D

D**T****A**

☐ in den Titeldaten ☒ im Korpus ☐ in der Dokumentation

Texte ▼ Projekt ▼ Dokumentation ▼ Impressum

Werke im Deutschen Textarchiv

Momentan sind 1551 Werke im Deutschen Textarchiv verfügbar.

gehe zu: Anfang ... 1-100 · 101-200 · 201-300 ... Ende

Abel, Heinrich Kaspar: Wohlerfahrner Leib-Medicus der Studenten. Leipzig, 1699.

Abschatz, Hans Assmann von: Poetische Übersetzungen und Gedichte. Leipzig, 1704.

Achenwall, Gottfried: Abriß der neuesten Staatswissenschaft der vornehmsten Europäischen Reiche und Republicken. Göttingen, 1749.

Adams, George: Versuch über die Electricität. Leipzig, 1785.

Adler, Alfred: Studie über Minderwertigkeit von Organen. Berlin u. a., 1907.

Adler, Emma: Die berühmten Frauen der französischen Revolution 1789-1795. Wien, 1906.

Frapan, Ilse: Bittersüß. Novellen. Berlin, 1891.

Frapan, Ilse [i. e. Ilse Akunian]: Flügel auf! Novellen. Berlin, 1895.

Albertinus, Aegidius: Der Landstörtzer: Gusman von Alfarche oder Picaro genannt. Bd. 1. München, 1615.

Alexis, Willibald: Ruhe ist die erste Bürgerpflicht oder Vor fünfzig Jahren. Bd. 1. Berlin, 1852.

Alexis, Willibald: Ruhe ist die erste Bürgerpflicht oder Vor fünfzig Jahren. Bd. 2. Berlin, 1852.

Alexis, Willibald: Ruhe ist die erste Bürgerpflicht oder Vor fünfzig Jahren. Bd. 3. Berlin, 1852.

Alexis, Willibald: Ruhe ist die erste Bürgerpflicht oder Vor fünfzig Jahren. Bd. 4. Berlin, 1852.

Alexis, Willibald: Ruhe ist die erste Bürgerpflicht oder Vor fünfzig Jahren. Bd. 5. Berlin, 1852.

Altenberg, Peter: Prödrömös. 2. Aufl. Berlin, 1906.

Ansichten

Volltitelangaben (kurz)

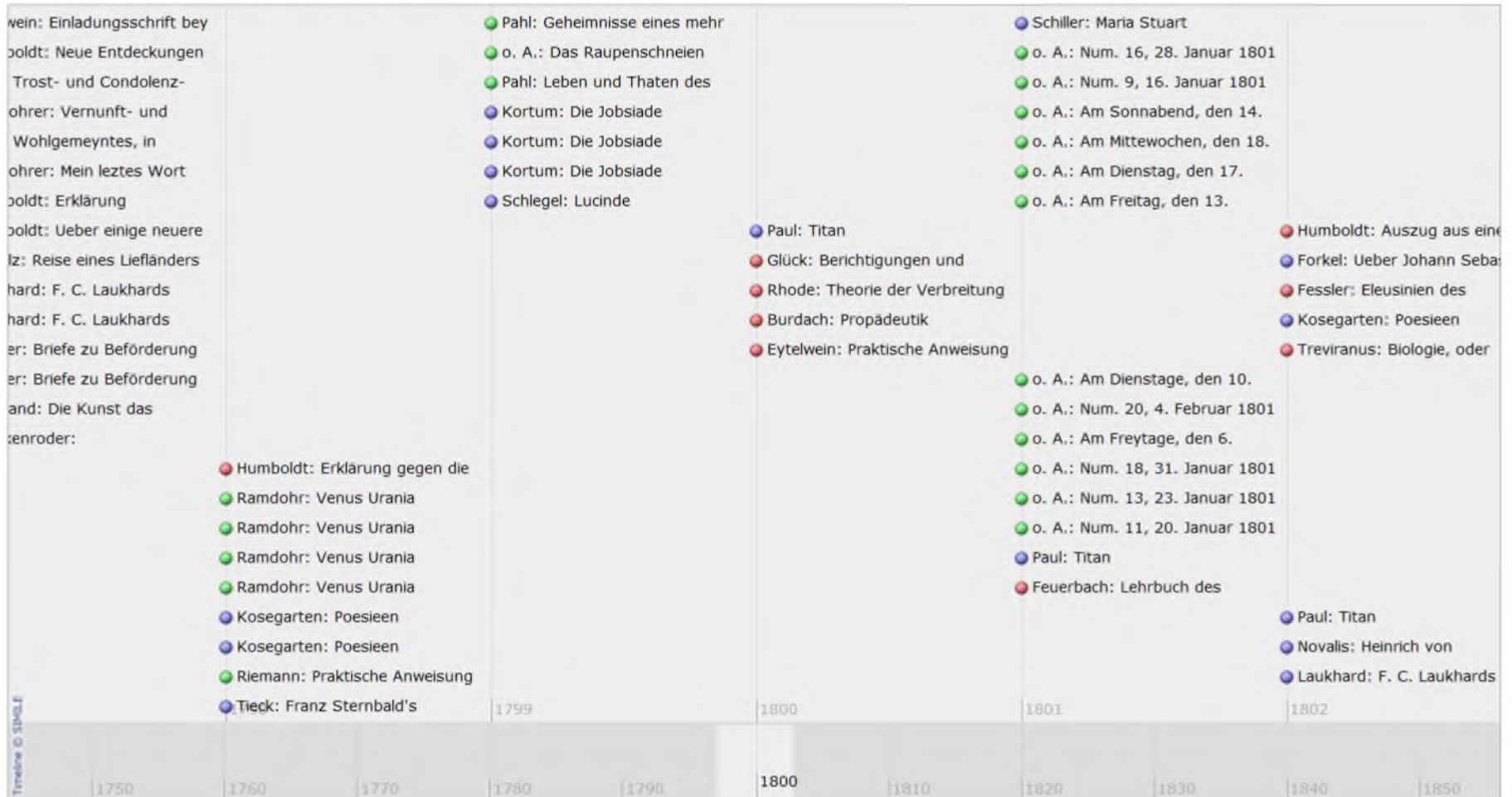
Volltitelangaben (lang)

Tabelle:

☒	AutorIn 1
☐	AutorIn 2
☐	AutorIn 3
☐	Kurztitel
☒	Titel
☐	Untertitel
☐	Bandbezeichnung
☐	Bandnummer
☐	Reihentitel
☐	Ausgabe
☐	Ort
☐	Verlag
☒	Jahr
☐	Schriftart
☐	Kategorie
☐	Unterkategorie
☐	Bibliothek
☐	im DTA seit
☐	DTA-Teilkorpus

Style, gatunki, epoki w DTA

● Belletristik
 ● Gebrauchsliteratur
 ● Wissenschaft



Przykład dzieła w inteligentnej bibliotece DTA

h (DTAQ)

DWDS dlexDB CLARIN-D

T

A

☐ in den Titeldaten ☒ im Korpus ☐ in der Dokumentation

Texte ▼ Projekt ▼ Dokumentation ▼ Impressum

Aufruf
an die Slaven.

von
einem russischen Patrioten

Michael Bakunin.
Mitglied des Slavencongresses in Prag.

Koethen.
Selbstverlag des Verfassers.
1848.

Bakunin, Michail Alexandrowitsch: Aufruf an die Slaven. Koethen, 1848.

BIBLIOGRAPHISCHE ANGABEN

URN:	urn:nbn:de:kobv:b4-30147-5
Titel:	Aufruf an die Slaven
Untertitel:	Von einem russischen Patrioten Michael Bakunin. Mitglied des Slavencongresses in Prag
Autor/in:	Michail Alexandrowitsch Bakunin (GND, ADB/NDB)
Erscheinungsjahr:	1848
Verlag:	Selbstverlag des Verfassers
Ort:	Koethen
Auflage:	1. Auflage

INFORMATIONEN ZUM WERK

Publikationstyp:	Monographie
Verfügbarkeit:	Text (TEI-XML-, HTML-, TCF-, E-Book-Fassung): Dieses Werk ist gemeinfrei. Weitere Informationen: Nutzungsbedingungen.
Schriftart:	Fraktur
Genre:	Gebrauchsliteratur :: Rede
im DTA seit:	2012-11-20 18:01:31
Korpus:	DTAE (DTA-Erweiterungen) – Wikisource

Suche im Werk

[Hilfe](#)

Ansichten für dieses Werk

- Text-Bild-Ansicht
- alle Faksimiles
- DTAQ (Qualitätssicherung)
- Lese- bzw. E-Book-Fassung

Download

XML (TEI P5) · HTML · Text
TCF (text annotation layer)
TCF (tokenisiert, serialisiert, lemmatisiert, normalisiert)

Metadaten

TEI-Header · CMD1 · Dublin Core

Statistiken

Scans:	36
Zeichen:	ca. 54 599
Tokens:	ca. 7 960
Oberflächentypes:	ca. 2 356

Wortwolken

KWIC – słowa w kontekście (DTA)

Treffer 1 - 10 von 12

KWIC · Neue Suche · Vorherige · Nächste · Hilfe

10 Treffer pro Seite Sortierung: Datum aufsteigend/absteigend · zufällig · Verlaufskurve



- 1: Bakunin, Michail Alexandrowitsch: Aufruf an die Slaven. Koethen, 1848. #4 [\[mehr\]](#)

Polen haben sie ins Verderben gestürzt, sie werden auch Euch ins Verderben stürzen.



- 2: Bakunin, Michail Alexandrowitsch: Aufruf an die Slaven. Koethen, 1848. #7 [\[mehr\]](#)

"Hinweg die Unterdrücker! "erscholl es wie aus Einem Munde ,"
den Bedrückten Heil, den **Polen**, den Italienern und Allen!



- 3: Bakunin, Michail Alexandrowitsch: Aufruf an die Slaven. Koethen, 1848. #8 [\[mehr\]](#)

Als, berührt vom Hauche der Revolution, Italiener, **Polen**,
Slaven, Deutsche, Magyaren, Wallachen, Die in Oesterreich und
Die in der Türkei, kurz Alle, welche bis dahin in heimischen
Ketten oder unter fremden Joche geächtzt hatten, sich
freudebebend erhoben!



- 4: Bakunin, Michail Alexandrowitsch: Aufruf an die Slaven. Koethen, 1848. #8 [\[mehr\]](#)

Aufgelöst erklärte die Revolution aus ihrer Machtvollkommenheit
die Despoten-Staaten - aufgelöst das preußische Reich, das sie
die ihm zugefallenen polnischen Landestheile aus seinem
Staatsverbande entlassen hieß, - aufgelöst Oesterreich, dieses
aus den verschiedenartigsten Nationalitäten durch List, Gewalt
und Verbrechen zusammengeknüttelte Ungethüm, - aufgelöst das
türkische Reich in welchem kaum siebenhunderttausend

Fahren Sie über die einzelnen
Tokens mit der Maus, um folgende
Informationen zu sehen:

u: Originaltext, UTF-8-kodiert
w: approximierter Latin-1-Text
v: CAB-normalisierte Wortform
l: Lemma (unflektierte Form)
p: Part-of-Speech-Analyse

Czytanie wspomagane w inteligentnej bibliotece (DTA)

Bild: 0019 : 19

<< vorherige Seite

nächste Seite >>

+ - 1:1 fit 52%

49

sten fehlschlügen, einzig die Armeen des Kaisers Nicolai, des Hauptes und Hüters aller volksverräterischen Complotte in Europa? — Oder seid Ihr selbst allein Eure Stärke, Ihr zwölf Millionen Slaven gegenüber einer Welt von Hassern und Feinden und ohne die Sympathie und den Beistand Eurer von Euch verlassenen und zurückgestoßenen Stammesbrüder in Rußland und Polen, dieser Eurer natürlichen Bundesgenossenschaft von sechzig Millionen, — Ihr, die Ihr schon jetzt nicht auf Euch selbst stehen zu können, sondern Euch auf die schwarzgelbe Camarilla und ihre Staatsstreiche stützen zu müssen meint?

Was werdet Ihr sein in Eurer Vereinsamung und Verlassenheit? Nichts! Was könntet Ihr sein im Vaterlande mit allen Euern Brüdern? Eine Riesenmacht von achtzig Millionen, ein gewaltiges Panier der Freiheit, die Freude und der Stolz der gesamten jugendlich erwachenden Menschheit.

Brüder! ich bin ein Russe, ich spreche zu Euch als Slave. Meine Gesinnungen, meine Gefühle, meine Gedanken habe ich Euch auf dem Congresse zu Prag offen dargelegt. Ihr wisst, daß ich, als Russe, das Heil meiner Landsleute nur in der Gemeinsamkeit mit allen übrigen Brüdern, nur in der Verbindung aller slavischen Völker als einer Föderation freier Stammgenossenschaften erblicke. Ihr wisst, daß ich mir die Beförderung dieses großen und heiligen Zwecks zur Aufgabe meines Lebens gemacht habe. Dies giebt mir ein Recht so zu Euch zu reden wie ich thue, denn in Euern Angelegenheiten bin ich bei meinen eigenen, Euere Sache ist die unsrige, Euer Heil unser Heil, Eure Ehre unsere Ehre, Eure Schmach unsre Schmach und Euer Verderben unser Verderben. Im

2*

fehlschlügen, einzig die Armeen des Kaisers Nicolai, des Hauptes und Hüters aller volksverräterischen Complotte in Europa? — Oder seid Ihr selbst allein Eure Stärke, Ihr zwölf Millionen Slaven gegenüber einer Welt von Hassern und Feinden und ohne die Sympathie und den Beistand Eurer von Euch verlassenen und zurückgestoßenen Stammesbrüder in Rußland und Polen, dieser Eurer natürlichen Bundesgenossenschaft von sechzig Millionen, — Ihr, die Ihr schon jetzt nicht auf Euch selbst stehen zu können, sondern Euch auf die schwarzgelbe Camarilla und ihre Staatsstreiche stützen zu müssen meint?

Was werdet Ihr sein in Eurer Vereinsamung und Verlassenheit? Nichts! Was könntet Ihr sein im Verbande mit allen Euern Brüdern? Eine Riesenmacht von achtzig Millionen, ein gewaltiges Panier der Freiheit, die Freude und der Stolz der gesamten jugendlich erwachenden Menschheit.

Brüder! ich bin ein Russe, ich spreche zu Euch als Slave. Meine Gesinnungen, meine Gefühle, meine Gedanken habe ich Euch auf dem Congresse zu Prag offen dargelegt. Ihr wisst, daß ich, als Russe, das Heil meiner Landsleute nur in der Gemeinsamkeit mit allen übrigen Brüdern, nur in der Verbindung aller slavischen Völker als einer Föderation freier Stammgenossenschaften erblicke. Ihr wisst, daß ich mir die Beförderung dieses großen und heiligen Zwecks zur Aufgabe meines Lebens gemacht habe. Dies giebt mir ein Recht so zu Euch zu reden wie ich thue, denn in Euern Angelegenheiten bin ich bei meinen eigenen, Euere Sache ist die unsrige, Euer Heil unser Heil, Eure Ehre unsere Ehre, Eure Schmach unsre Schmach und Euer Verderben unser Verderben. Im

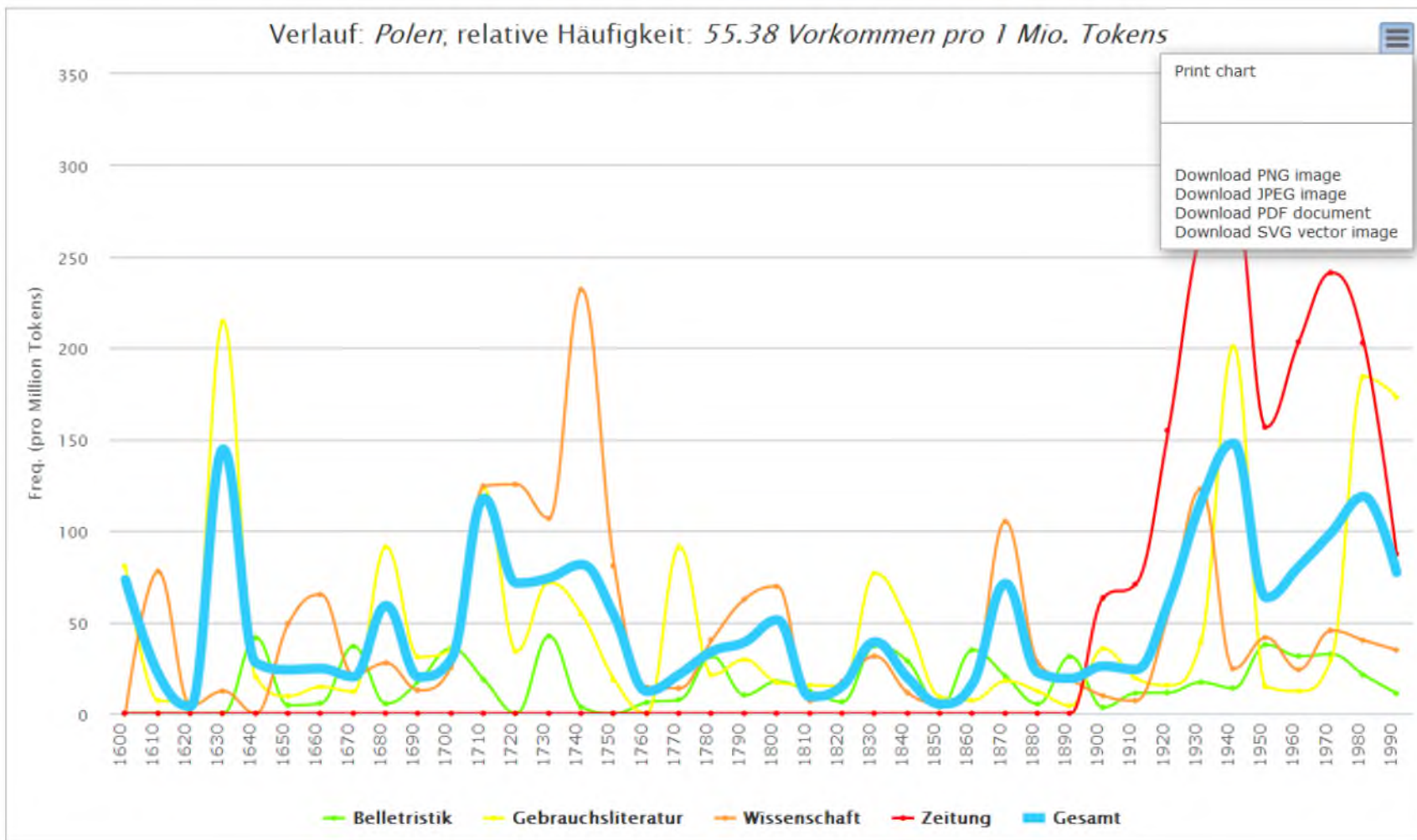
Szeregi leksykalne z Deutsches Textarchiv

Suchanfrage: Polen

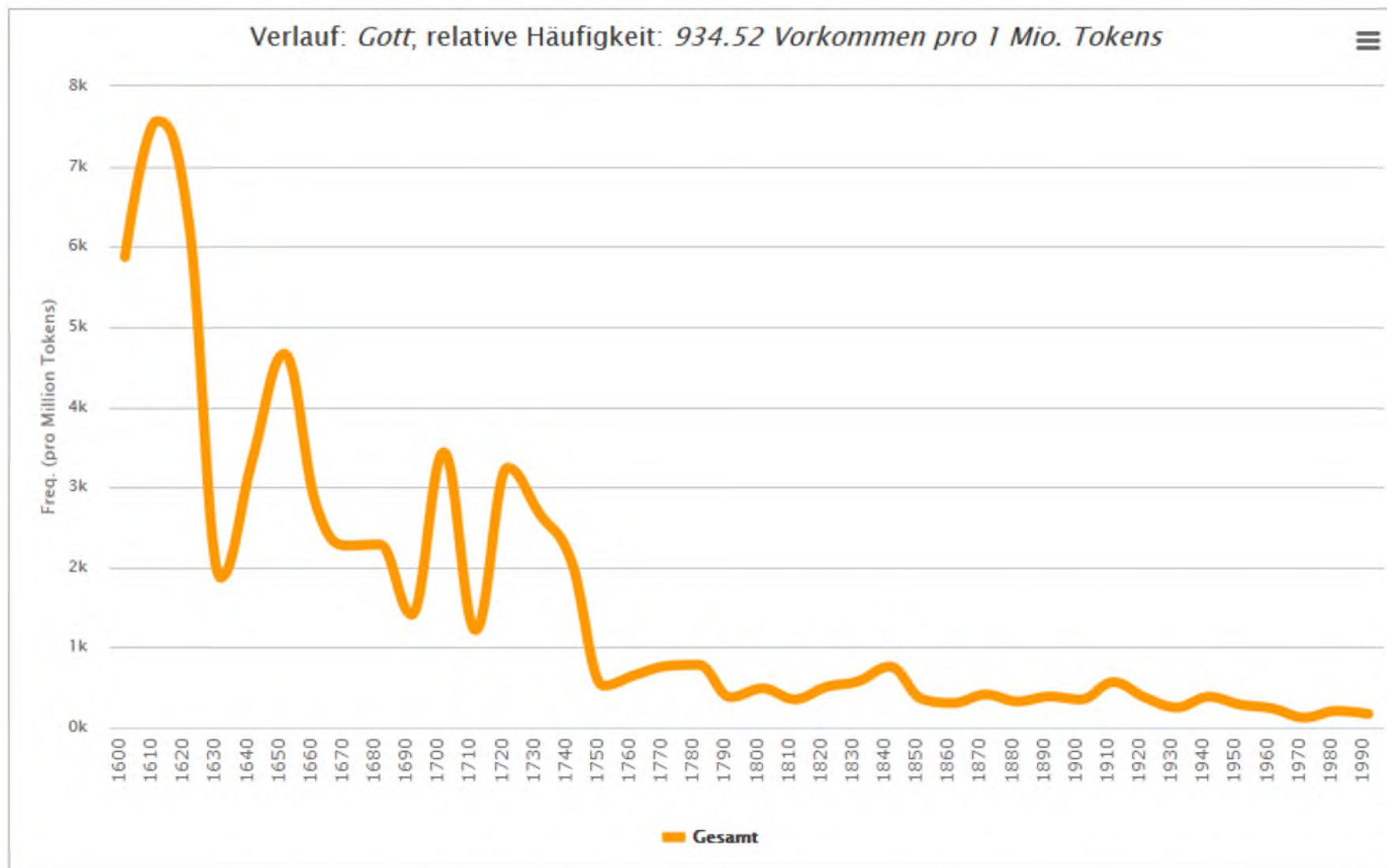
☒ Textkategorien anzeigen beta

alle Belege im DWDS

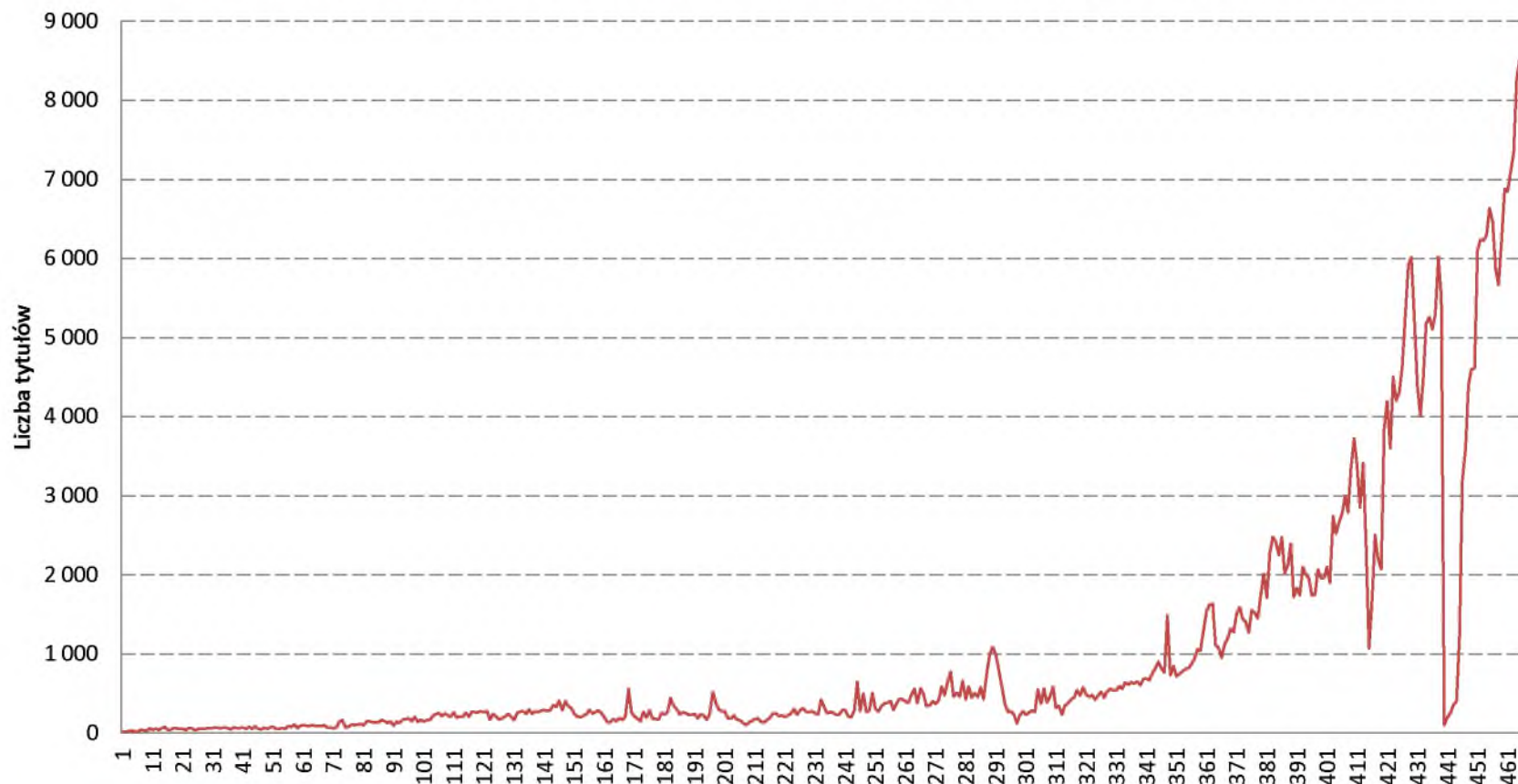
alle Belege im DTA



Szeregi leksykalne z Deutsches Textarchiv



Liczba tytułów książek wydawanych w Polsce w latach 1500-1965



Źródło danych: M. Czarnowska (1967), *Ilościowy rozwój polskiego ruchu wydawniczego*. Warszawa: Biblioteka Narodowa. Wykres: opracowanie własne.

Analiza big data

(źródło: prace dr. Macieja Maryla)

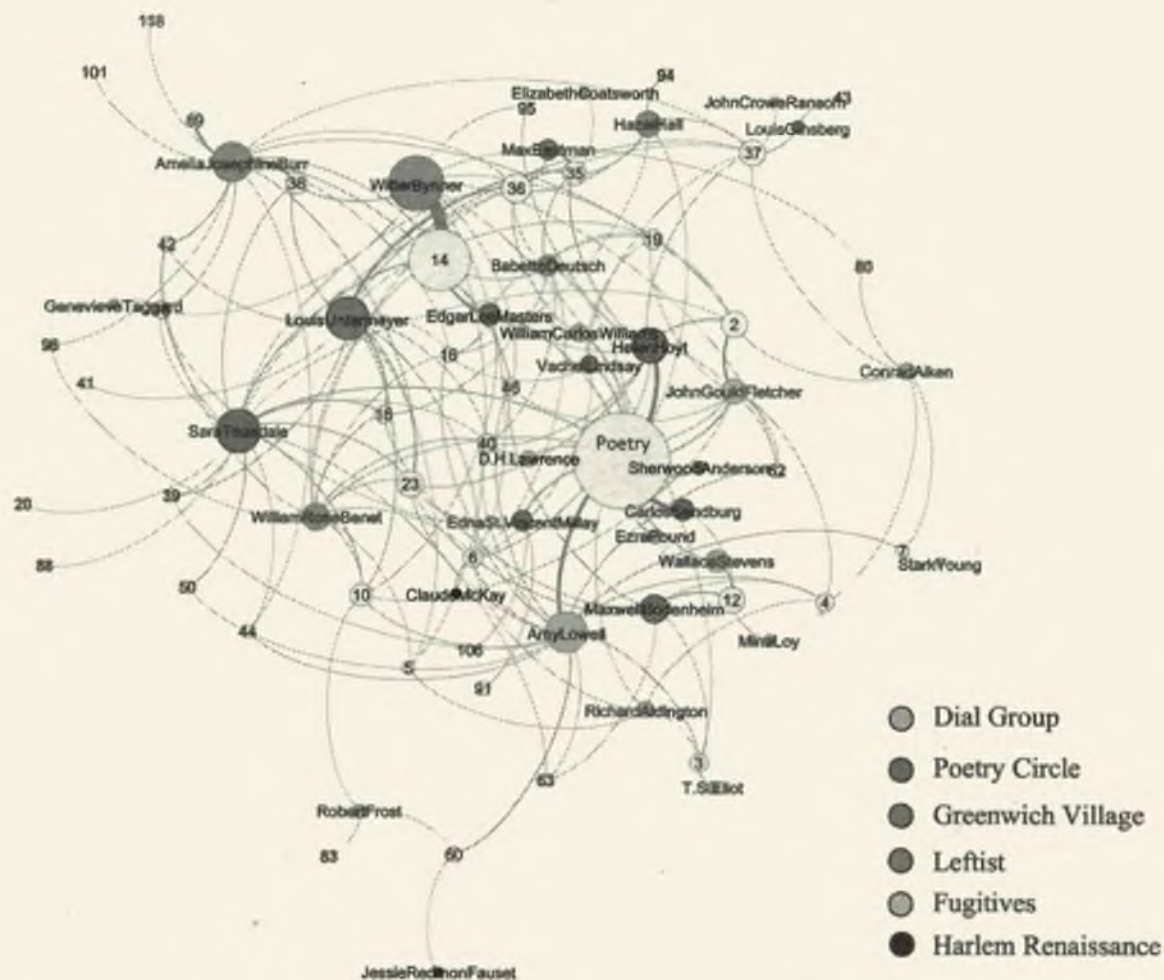
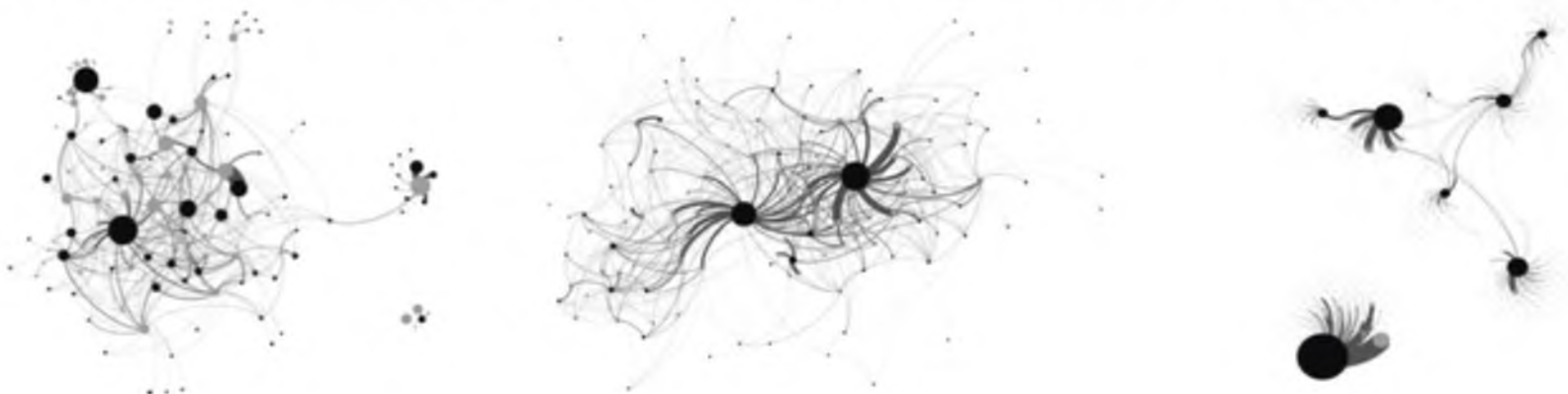


Figure 1. American poets, 1917–18

Analiza big data

(źródło: prace dr. Macieja Maryla)

Fig. 1 Journal-Poet affiliation networks for the U.S. (1918-1922), Japan (1928-1931), and China (1915-1945). (© Hoyt Long)



100 najczęściej cytowanych autorów w "Tekstach Drugich"
(grafika: Wordle)



źródło: badania dr. Macieja Maryla

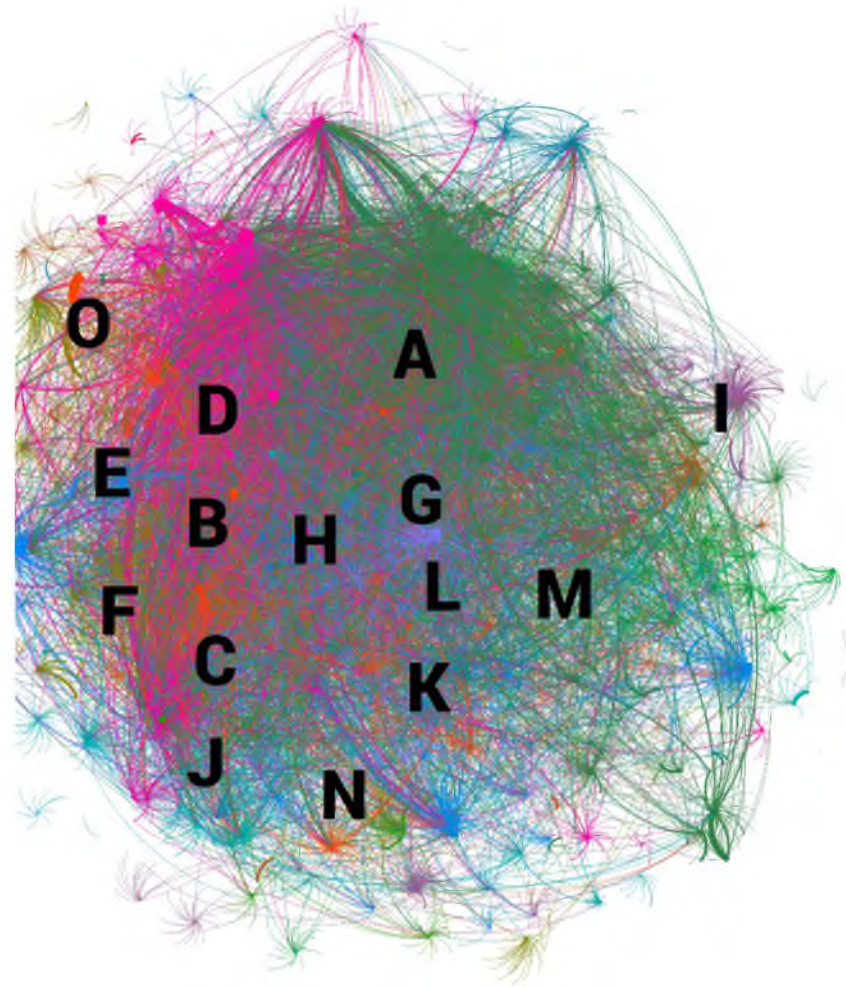
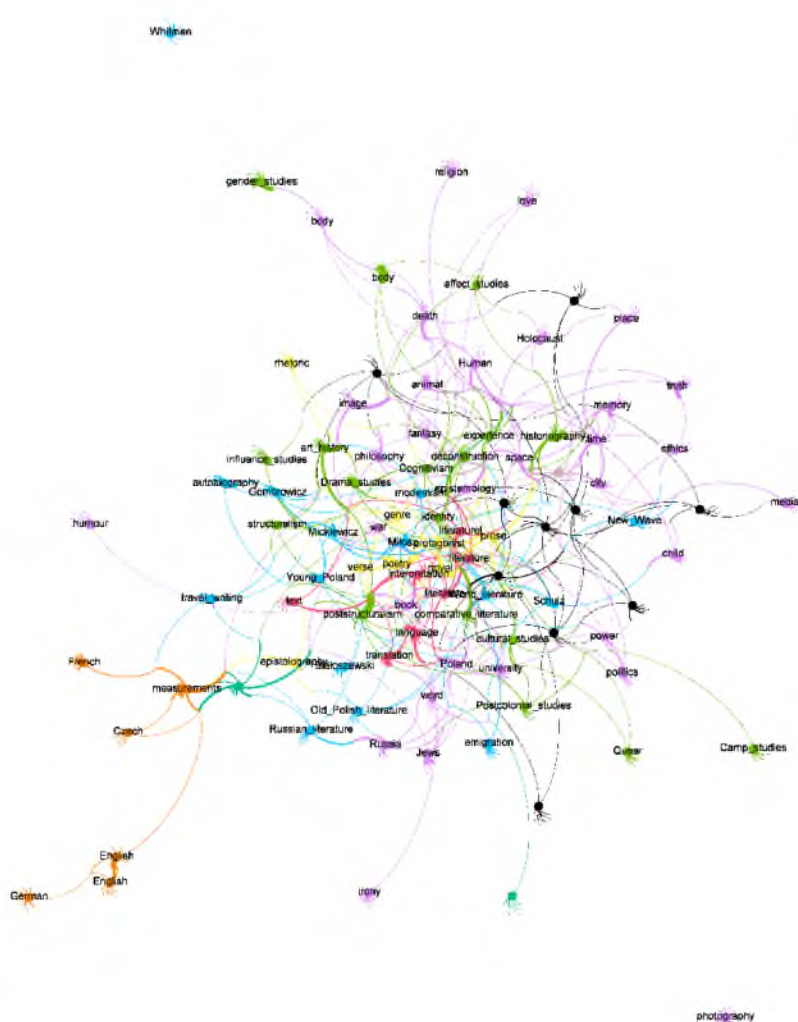
(źródło: prace dr. Jana Rybickiego)

(źródło: prace dr. Jana Rybickiego)



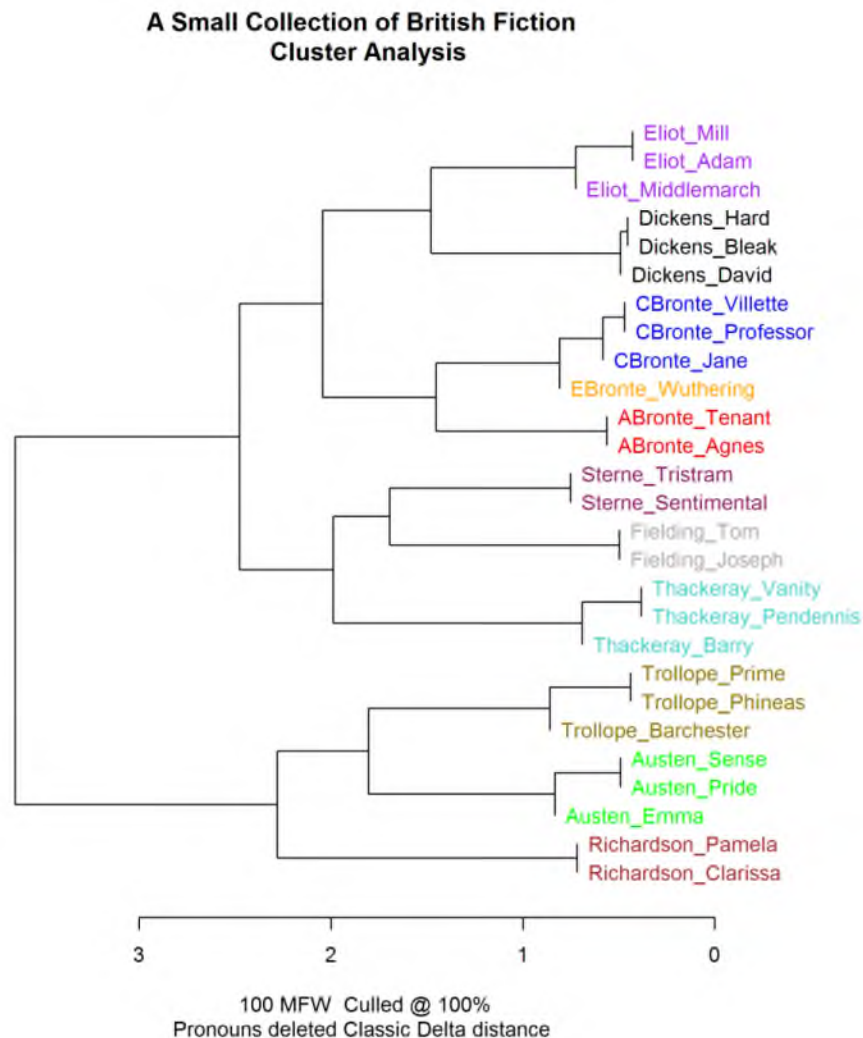
Modelowanie przestrzenne literatury

(źródło: prace dr. Jana Rybickiego)



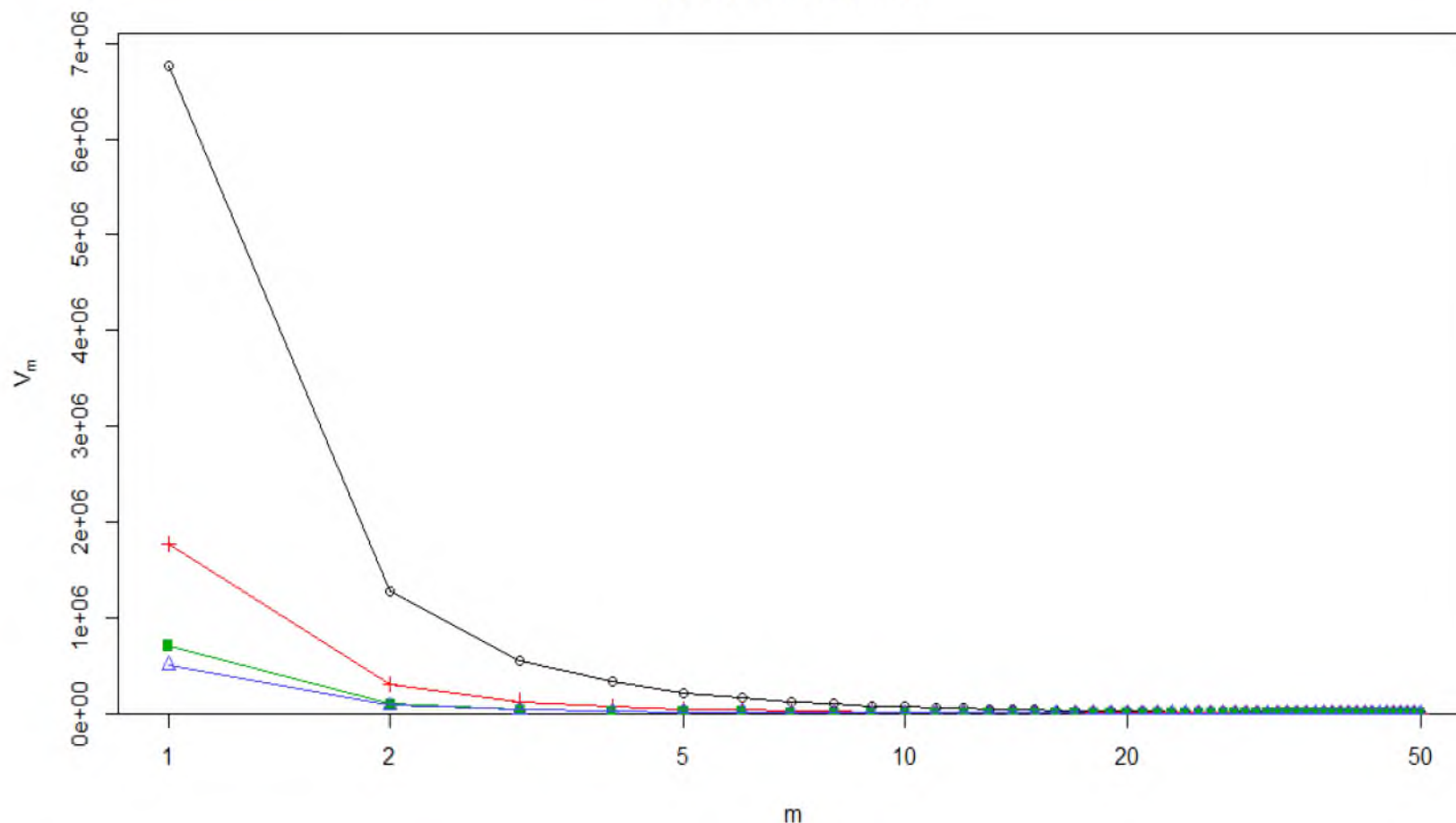
Klasyfikacja dzieł literackich w formie dendrytu

(źródło: badania dr. hab. Macieja Edera)





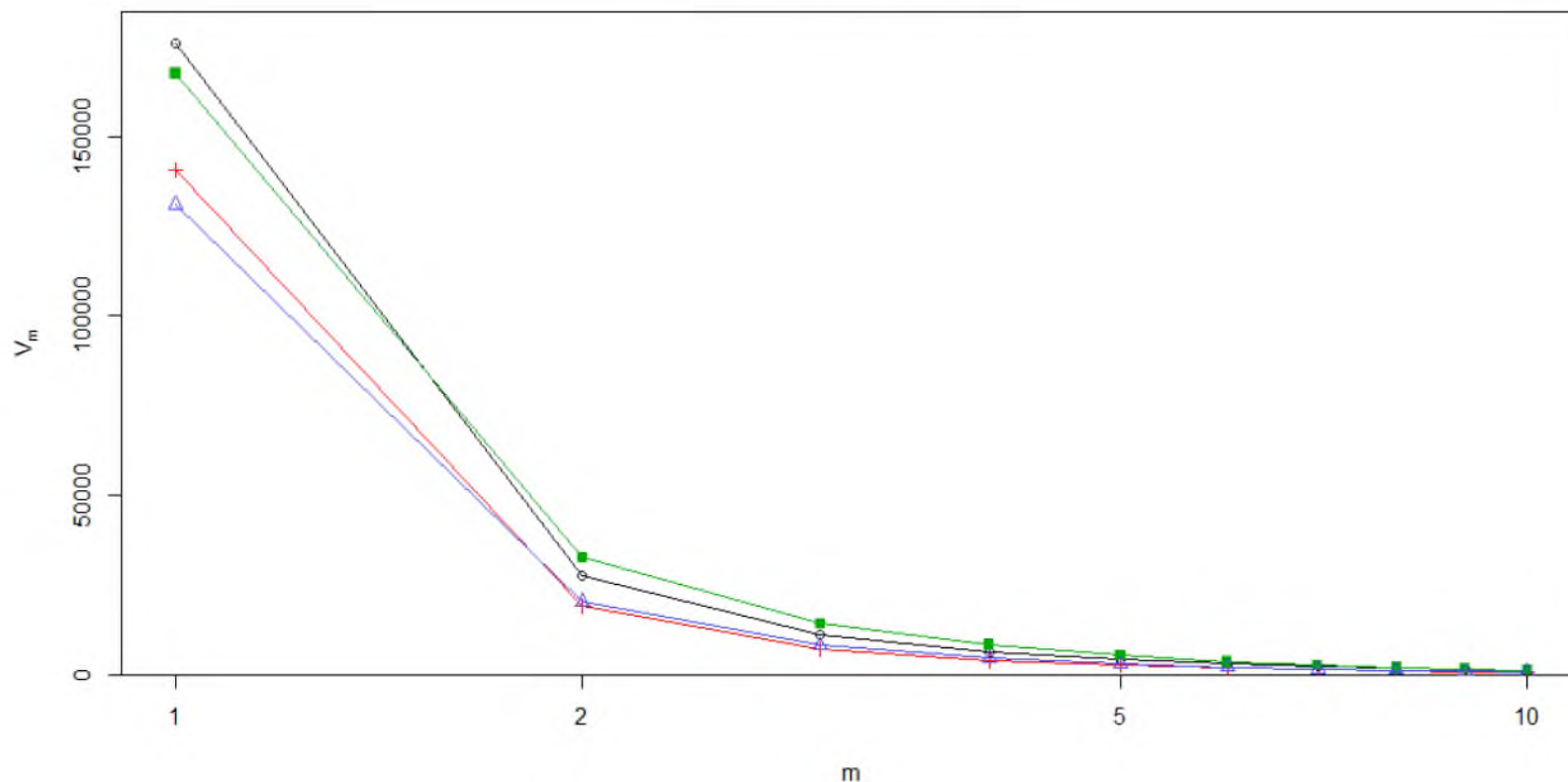
Frequency Spectrum



Linia czarna – rzeczownik, czerwona – przymiotnik, zielona – liczebnik, niebieska – czasownik



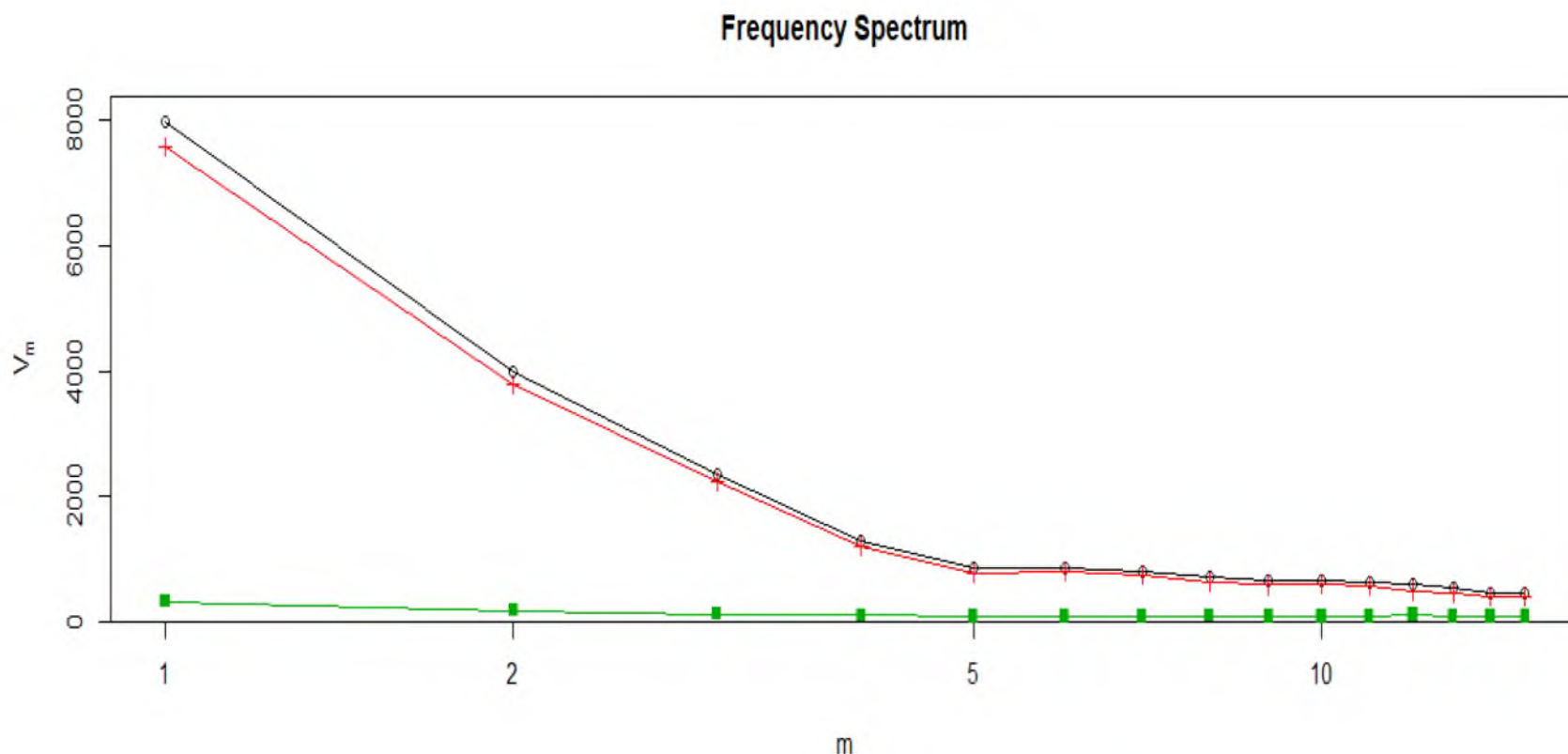
Frequency Spectrum



Linia czarna – przysłówek, zielona – wykrzyknik, czerwona – nierozpoznane, niebieska – kublik



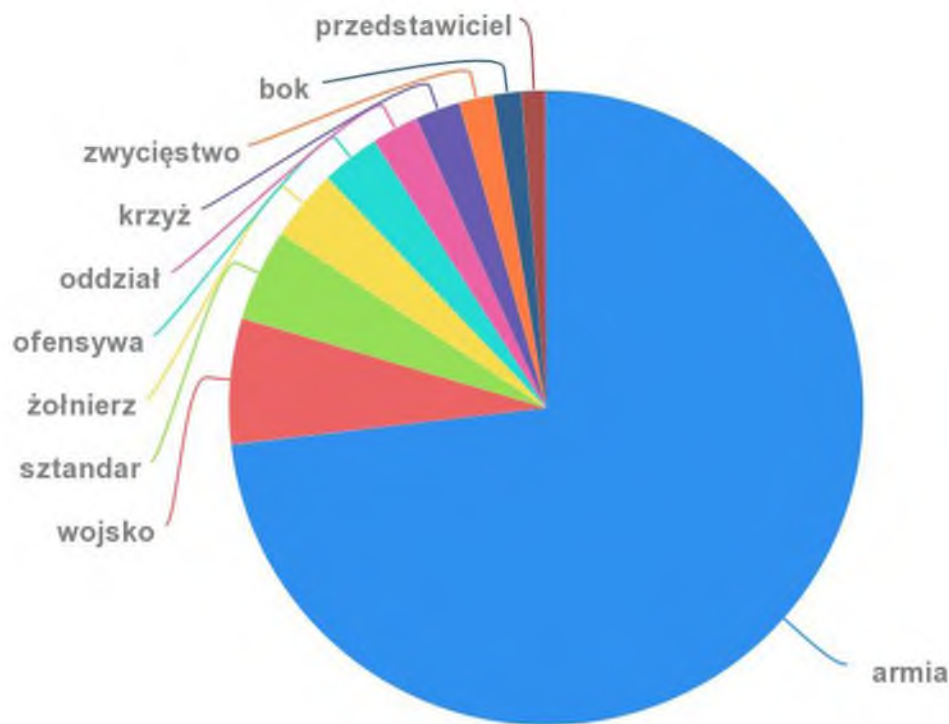
Rozkład widma częstości części mowy: nazwy pospolite a nazwy własne



Linia czarna – rzeczowniki, **czzerwona** – rzeczowniki **pospolite**,
zielona – nazwy własne

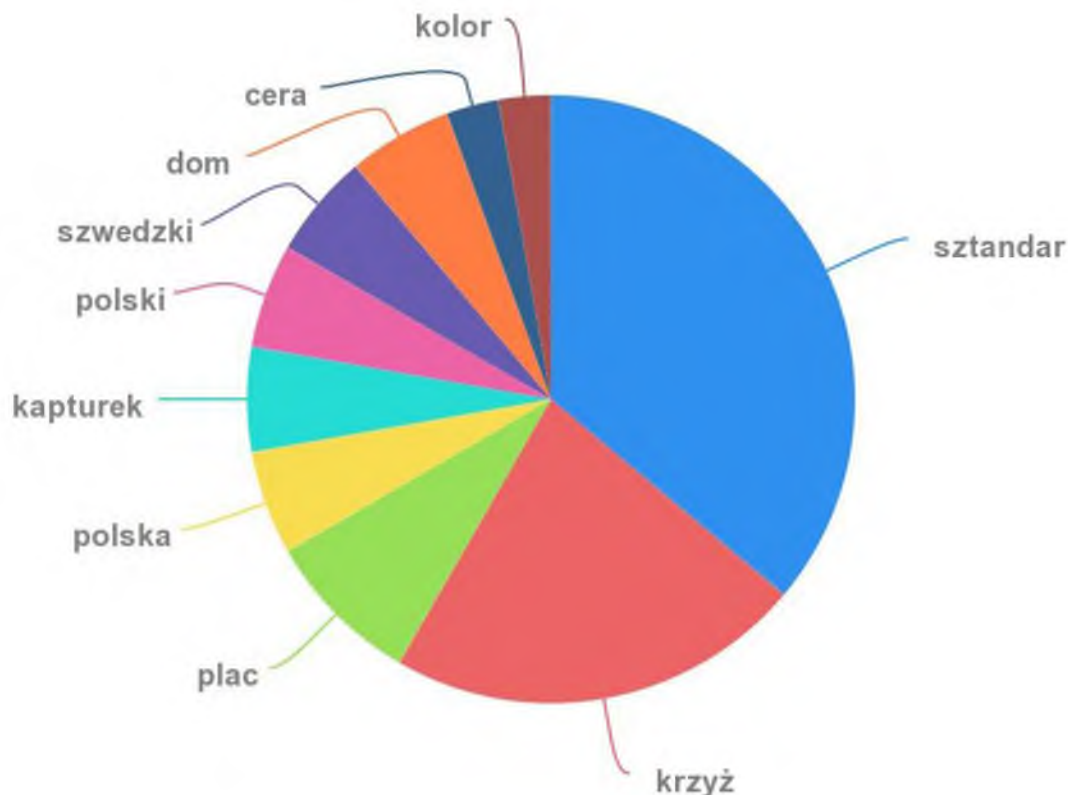
Profil leksemu "czerwony" (1945)

(prasa polska, źródło: ChronoPress)



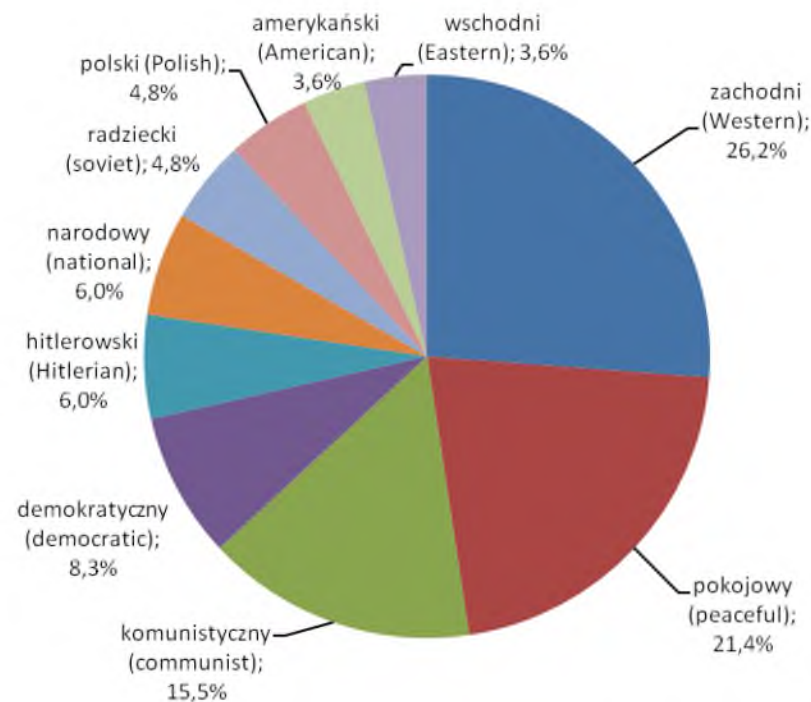
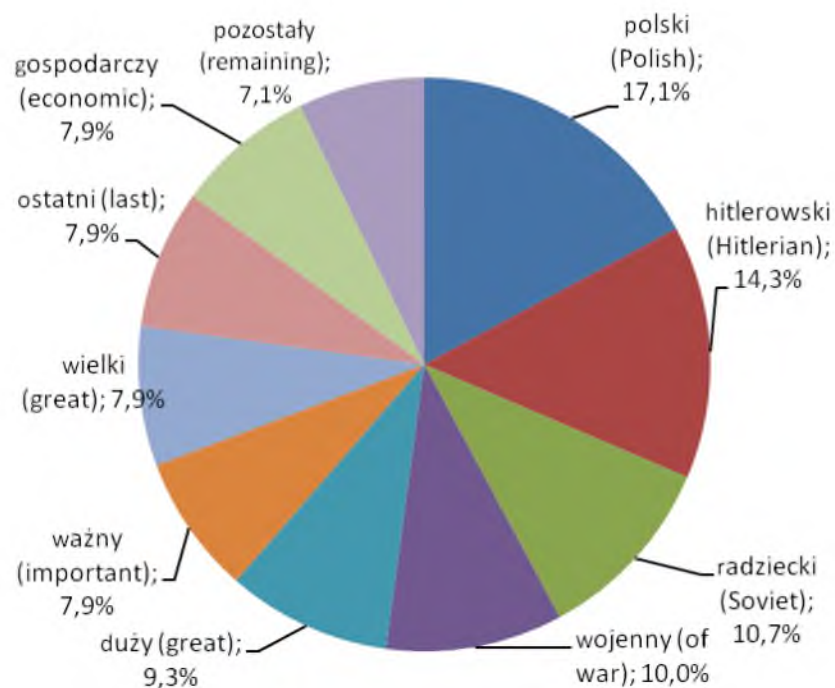
Profil semantyczny leksemu "czerwony" (1954)

(prasa polska, źródło **ChronoPress**)



Profil semantyczny leksemu "Niemiec" (1945 i 1954)

(prasa polska, źródło **ChronoPress**)





Uniwersytet
Wrocławski

PRZYKŁADY NARZĘDZI I BADAŃ HUMANISTYKI CYFROWEJ

I. Perspektywa poznawcza (naukowa)

- 1) Silne przesunięcie w stronę skrajnego empiryzmu na niekorzyść teorii (dominacja indukcji nad dedukcją)
- 2) Wiara w to, że w nauce jakość wiedzy może być pochodną ilości danych
- 3) Zasada emergencji obiektów badań (w rozumieniu teorii systemów)
- 4) Interdyscyplinarność: HC przypomina zjazd rodzinny, gdzie spotykają się dyscypliny, które o sobie zapomniały
- 5) Powrót sporu o prymat języka: prymarny kod modelujący całość procesów poznawczych czy jeden z kodów?

II. Perspektywa metodologiczna

- 1) Badania wielkich zbiorów (Big Data)
- 2) Filozofia Linked Open Data: synteza formatów danych tekstowych, graficznych, czasowych i geoprzestrzennych
- 3) Bezprecedensowa łatwość transmisji i udostępniania
- 4) Konieczność uwzględniania podczas projektowania aspektów prawnych (praw autorskich, ochrony wizerunku)
- 5) Bezprecedensowa rola metadanych
- 6) Zaawansowane metody AI i analizy danych (topic modelling, taksonomie, sieci neuronowe)
- 7) Wyzwaniem jest trwałość zasobów

III. Perspektywa społeczno-komunikacyjna

- 1) HC jest kolektywna, a nie indywidualna
- 2) Zasięg projektów HC prawie zawsze międzynarodowy
- 3) Współpraca sieciowa (networking), odbywa się w ramach konsorcjów lub innych sformalizowanych struktur
- 4) Dominacja komunikacji sieciowej przez specjalne aplikacje nad komunikacją bezpośrednią (F2F)
- 5) Produkty HC mają postać dynamicznych infrastruktur, a nie gotowych publikacji
- 6) Częste wykorzystanie kolektywnej pracy wolontariuszy nadzorowanych przez profesjonalistów (*crowdsourcing*)

III. Perspektywa społeczno-komunikacyjna

- 7) Odejście od linearnej lektury na rzecz hipertekstowości (w produktach i w myśleniu odbiorców...)
- 8) Bezprecedensowa rola grafiki i multimediiów w komunikacji naukowej (humanistyka jakby wypiękniała)
- 9) Podniesienie prestiżu humanistyki przez jej cyfrowość i powiązanie z dyscyplinami o bardziej ugruntowanej pozycji
- 10) W dobie internacjonalizmu praktyk akademickich i w świecie nauki zdominowanej przez angielski wzrosła rola humanistyk i języków narodowych

IV. Perspektywa krytyczna

- 1) Dzisiejsza humanistyka cyfrowa jest **antyhumanistyczna**:
 - prymat technologii (i inżynierów) nad treścią humanistyki
 - człowieka zastępują programy AI
- 2) Dzisiejsza HC jest **arogancka** (odwrócenie relacji w transmisji wiedzy)
- 3) Dzisiejsza HC jest **imperialna**: zagarnia kolejne dyscypliny i wszystkie dające się zdigitalizować obszary rzeczywistości
- 4) Dzisiejsza HC jest obarczona piętnem **ignorancji**, ponieważ badacze bezrefleksyjnie przekraczają granice uprawianych przez siebie dyscyplin

IV. Perspektywa krytyczna

- 5) Dzisiejsza HC jest nastawiona na digitalizację lub tworzenie zasobów i infrastruktur, a nie rozwiązywanie problemów poznawczych
- 6) Multimedialność i interdyscyplinarność HC podważają lingwocentryzm humanistyki tradycyjnej
- 7) HC zaciera bezrefleksyjnie granice między dyscyplinami i dziedzinami
- 8) HC narzuca myślenie kategoriami big data – nie zawsze jest ono konieczne, a czasem wiedzie na manowce.

DZIĘKUJĘ ZA UWAGĘ

Uwaga:

żaden fragment niniejszej prezentacji ekranowej (tekst, grafika, logotypy) nie może być powielany lub rozpowszechniany w żadnej formie i w żaden sposób bez uprzedniego zezwolenia jego twórcy. Wszelkie znaki graficzne, nazwy własne, logotypy i inne dane są chronione prawem autorskim i należą do ich właścicieli.