Ekstrakcja terminologii dziedzinowej program TermoPL

Małgorzata Marciniak, Agnieszka Mykowiecka, Piotr Rychlik

Instytut Podstaw Informatyki PAN
mm@ipipan.waw.pl,
agn@ipipan.waw.pl,
rychlik@ipipan.waw.pl

Warszawa 5 lipca 2017

Zadanie

Cel ekstrakcji terminologii:

wydobycie specyficznej terminologii z tekstów dotyczących wybranej dziedziny.

Zastosowania:

- tworzenie słowników dziedzinowych;
- tworzenie zasobów do tłumaczenia tekstów;
- anotacja dokumentów i wspomaganie wyszukiwania odpowiedzi na pytania;
- przydatne przy streszczaniu dokumentów;
- wstępny krok przy opracowywaniu ontologii dziedziny;
- ...

Przykład

Plik Edycja Widok Historia Zakładki Narzędzia Pomoc

CLARIN2015 Submission 6

https://easychair.org/conferences/submission.cgi?submission=2427016;a=95

easychair login

Często odwiedzane Bezpłatna usługa poc... Dostosuj łącza file:///C:/Documents... Windows Media Windows

For all questions related to processing your submission you should contact the conference organizers. Click here to see information about this conference.

All **reviews sent to you** can be found at the bottom of this page.

Paper 6	
Title:	TermoPL tool for automatic extraction of Polish terminology
Submission	
Author keywords:	terminology extraction Polish C-value
EasyChair keyphrases:	term candidate (70), automatic terminology extraction (63), base form (60), noun phrase (60), nap gen (50), terminology extraction (50), narodowy korpus j ezyka (40), orthographic form (40), simplified form (40)
Abstract:	The paper presents the TermoPL tool created to extract terminology from domain corpora in Polish. The program extracts term candidates with help of a simple grammar recognizing noun phrases and uses the C-value method to rank them. The program accepts as input morphologically annotated and disambiguated domain texts and creates a list of terms, the top part of which comprise domain terminology.
Time:	Jul 13, 10:44 GMT

Realizacja zadania

- Zgromadzenie tekstów dziedzinowych.
- Wstępna analiza lingwistyczna — tagowanie (przypisanie formy podstawowej, części mowy oraz charakterystyki morfologicznej).
- Identyfikacja fraz — kandydatów na terminy.
- Szeregowanie fraz.
- Selekcja fraz.

Program Termo.PL

Opracowany w ramach projektu Clarin.Pl

- Java Runtime Environment w wersji 7 lub nowszej;
- Wymaga Morfeusza 2 do wygenerowania formy podstawowej z uproszczonej formy;
- Wymaga otagowanego i ujednoznacznionego korpusu danych w jednym z formatów:
  - NKJP;
  - XCES;
  - XML-owe wyjście z Korpusomatu (src/morph.xml);
  - zapis uproszczony: token # lemat # tag.
- na wyjściu: lista uporządkowanych terminów (w uproszczonych formach lub zrekonstruowanych formach podstawowych wraz z formami znalezionych fraz).

Adresy internetowe

TermoPL adresy internetowe

- http://ws.clarin-pl.eu/termopl.shtml
- http://zil.ipipan.waw.pl/TermoPL

Adresy narzędzi

- Morfeusz http://sgjp.pl/morfeusz/
- Kuźnia http://zil.ipipan.waw.pl/Kuźnia
- Korpusomat http://korpusomat.nlp.ipipan.waw.pl

Przykład tagowania

```
<tok>
<orth>prawa</orth>
<lex><base>prawa</base><ctag>subst:sg:nom:f</ctag></lex>
<lex><base>prawa</base><ctag>subst:sg:voc:f</ctag></lex>
<lex><base>prawo</base><ctag>subst:pl:acc:n</ctag></lex>
<lex><base>prawo</base><ctag>subst:pl:nom:n</ctag></lex>
<lex><base>prawo</base><ctag>subst:pl:voc:n</ctag></lex>
<lex disamb="1"><base>prawo</base><ctag>subst:sg:gen:n</ctag></lex>
<lex><base>prawy</base><ctag>adj:sg:nom:f:pos</ctag></lex>
<lex><base>prawy</base><ctag>adj:sg:voc:f:pos</ctag></lex>
</tok>
<tok>
<orth>autorskiego</orth>
<lex><base>autorski</base><ctag>adj:sg:acc:m1:pos</ctag></lex>
<lex><base>autorski</base><ctag>adj:sg:acc:m2:pos</ctag></lex>
<lex><base>autorski</base><ctag>adj:sg:gen:m1:pos</ctag></lex>
<lex><base>autorski</base><ctag>adj:sg:gen:m2:pos</ctag></lex>
<lex><base>autorski</base><ctag>adj:sg:gen:m3:pos</ctag></lex>
<lex disamb="1"><base>autorski</base><ctag>adj:sg:gen:n:pos</ctag></lex>
</tok>
```

Co rozumiemy pod pojęciem terminu?

Definicja słownikowa

Wyraz albo połączenie wyrazowe o specjalnym, konwencjonalnie ustalonym znaczeniu naukowym lub technicznym; (Doroszewski)

Definicja robocza

Fraza rzeczownikowa, która w tekstach dziedzinowych występuje dostatecznie często by przypuszczać, że opisuje pojęcie istotne dla dziedziny. Częstość tej frazy w tekstach spoza dziedziny jest niższa.

Struktura gramatyczna terminów w języku polskim

- rzeczownik, akronim lub skrót rzeczownika:
  - *podatek*, *angiografia*,
  - *PKB*, *USG*
  - *ust.*(awa),
- rzeczownik z przymiotnikiem (który wystąpił po lub rzadziej przed rzeczownikiem):
  - *stosunki gospodarcze*,
  - *granulocyty obojętnochłonne*;
- sekwencja rzeczownika z rzeczownikiem w dopełniaczu:
  - $udar_{n,nom}$ $mózgu_{n,gen}$;
  - $kodeks_{n,nom}$ $pracy_{n,gen}$;
- kombinacja powyższych dwóch struktur:
  - $europejski_{adj}$ $rynek_{n,nom}$ $usług_{n,gen}$ $finansowych_{adj}$,
  - *wodonercze niewielkiego stopnia dolnego układu podwójnego nerki prawej*;

Struktura gramatyczna terminów w języku polskim

- fraza rzeczownikowa modyfikowana frazą przyimkową:
  - *wierzytelność podatnika wobec skarbu państwa*,
  - *podatek dochodowy od osoby fizycznej*;
  - *poziom hormonów we krwi*;
- można uwzględnienić koordynację:
  - *bezsporna i wymagalna wierzytelność podatnika wobec skarbu państwa*,
  - *zapalenie mózgu i rdzenia*,
  - *oddział alergologii, endokrynologii i pediatrii ogólnej*.

Gramatyka (domyślna w programie)

$$
\begin{aligned}
NPP &: \$NAP \quad NAP_GEN^*; \\
NAP[agreement] &: AP^* \quad N \quad AP^*; \\
NAP_GEN[case = gen] &: NAP; \\
AP &: ADJ \mid ADJA \quad DASH \quad ADJ \mid PPAS; \\
N[pos = subst, ger]; & \\
ADJ[pos = adj]; & \\
ADJA[pos = adja]; & \\
PPAS[pos = ppas]; & \\
DASH[form = \text{"-"}]; &
\end{aligned}
$$

Wykluczenie niektórych słów/fraz

Terminy nie powinny składać się ze:

- słów wskazujących na określenie czasu, jak np: *miesiąc, dzień*;
- nazwy dni i miesięcy, np: *styczeń, poniedziałek*;
- przymiotników wymagających kontekstu do interpretacji np: *inny, niektóry, jakiś, pewien*.

Należy wykluczyć przyimki złożone:

- *[w kierunku] zapalenia nerek* —> *kierunek zapalenia nerek*;
- *[pod postacią] podatku VAT* —> *postać podatku VAT*;
- *[pod kątem] diagnostyki obrazowej* —> *kąt diagnostyki obrazowej*;
- *[pod kątem] prostym* —> *kąt prosty*.

Wykluczenie niektórych słów/fraz (cd)

Należy wykluczyć frazy ogólne, czyli takie które występują w wielu typach tekstów:

- frazy, zwroty służące do budowania narracji: *punkt widzenia*, zasadnicze pytanie, prosty wniosek
- frazy ogólne: duża liczba, gruba warstwa, niski poziom

Szeregowanie terminów

Dla każdej frazy kandydackiej p liczona jest wartość C-value:

$$C-value(p) = \begin{cases} l(p) * (freq(p) - \frac{1}{r(LP)} \sum_{lp \in LP} freq(lp)), & if \;\; r(LP) > 0, \\ l(p) * freq(p), & if \;\; r(LP) = 0 \end{cases}$$

p — rozważana fraza,
LP — zbiór fraz zawierających p,
$r(LP)$ — liczba różnych fraz w LP,
$l(p) = log_2(length(p))$, jeśli p ma długość 1 to przyjmujemy stałą np: $l(p) = 0.1$;

referencja bibliograficzna

Frantzi, K., Ananiadou, S., Mima, H.: Automatic recognition of multi-word terms: the C-value/NC-value method. Int. Journal on Digital Libraries 3 (2000) 115–130

Identyfikacja fraz

	pojedyncza	mnoga
nom	*przewlekły nieżyt żołądka*	*przewlekłe nieżyty żołądka*
gen	*przewlekłego nieżytu żołądka*	*przewlekłych nieżytów żołądka*
dat	*przewlekłemu nieżytowi żołądka*	*przewlekłym nieżytom żołądka*
acc	*przewlekły nieżyt żołądka*	*przewlekłe nieżyty żołądka*
inst	*przewlekłym nieżytem żołądka*	*przewlekłymi nieżytami żołądka*
loc	*przewlekłym nieżycie żołądka*	*przewlekłych nieżytach żołądka*

Wykorzystujemy uproszczoną formę podstawową:

- *przewlekły nieżyt żołądka* –> *przewlekły nieżyt żołądek*;
- *ostra niewydolność nerek* –> *ostry niewydolność nerka*.

Problemy z uproszczoną formą podstawową

Taką sama uproszczoną formę podstawową mają:

- frazy w liczbie mnogiej i pojedynczej np. *zapalenie ucha* i *zapalenie uszu*, uproszczona: *zapalenie ucho*;
- przymiotniki w różnych stopniach (mały, mniejszy) np. *miednica mała* (częściej *mała miednica* — opisuje rozmiar) podczas gdy *miednica mniejsza* (określenie anatomiczne), uproszczona: *miednica mały*;
- pozytywne i zanegowane imiesłowy przymiotnikowe . *powiększony*/*niepowiększony* mają formę podstawową *powiększyć*$_{inf}$;
- gerundia i imiesłowy maja bezokoliczniki jako formy podstawowe:
  - *usunięcie*$_{ger}$ *kamienia*$_{subst:gen}$ — operacja,
  - *usunięty*$_{ppas}$ *kamień*$_{subst:nom}$ — opis kamienia,

  forma uproszczona: *usunąć*$_{inf}$ *kamień*$_{subst}$.

Konteksty

planowa	*operacja*	*przepukliny*	*pachwinowej*	*lewostronnej*
	operacja	*przepukliny*	*pachwinowej*	*lewostronnej*
planowa	*operacja*	*przepukliny*	*pachwinowej*	
	operacja	*przepukliny*	*pachwinowej*	
		przepuklina	*pachwinowa*	*lewostronna*
	lewostronna	*przepuklina*	*pachwinowa*	
		przepuklina	*pachwinowa*	*prawostronna*
		przepuklina	*pachwinowa*	*obustronna*
	prawostronna	*przepuklina*	*pachwinowa*	
	uwięźnięta	*przepuklina*	*pachwinowa*	*prawostronna*

Liczenie kontekstów

Metody liczenia kontekstów (ograniczamy do jednego słowa):

1. liczba różnych kontekstów liczona po obu stronach razem;
2. suma różnych kontekstów po obu stronach;
3. maksimum z kontekstów liczonych z lewej i prawj strony osobno.

Konteksty dla frazy: *przepuklina pachwinowa*:

1. ‘operacja’–‘lewostronny’, ‘operacja’–[pusty], [pusty]–‘lewostronny’, ‘lewostronny’–[pusty], [pusty]–‘prawostronny’, [pusty]–‘obustronny’, ‘prawostronny’–[pusty], ‘uwięźnięty’–‘prawostronny’;
2. ‘operacja’, ‘lewostronny’, ‘prawostronny’, ‘obustronny’, ‘uwięźnięty’;
3. ‘operacja’, ‘lewostronny’, ‘prawostronny’, ‘uwięźnięty’ (lewych o jeden więcej).

Problem uciętych fraz

Przykłady frazy o silnym powiązaniu słów:

- w medycynie: *pęcherzyk żółciowy*, *jama brzuszna*, *staw kolanowy*;
- w ekonomii: *papiery wartościowe*, *fundusz inwestycyjny*;
- w angielskim: *contact lens*.

Gramatycznie poprawne zagnieżdżone frazy:

- *[zapalenie pęcherzyka] żółciowego*;
- *[USG jamy] brzusznej*;
- *[operacja lewego stawu] kolanowego*;
- *[giełda papierów] wartościowych*;
- *[uczestnik funduszu] inwestycyjnego*;
- *[soft contact] lens*.

NPMI – Normalised Pointwise Mutual Information

$$NPMI(x,y) = \left(ln \ \frac{p(x,y)}{p(x)p(y)} \right) \Big/ - ln \ \ p(x,y)$$

Where:

- ‘x y’ jest bigramem składajacym się z lematów tokenów x i y,
- p(x,y) jest prawdopodobińśtwem bigramu ‘x y’ w korpusie,
- p(x), p(y) jest prawdopodobieństwem unigramów ‘x’ i ‘y’ w korpusie .

referencja bibliograficzna

Gerlof Bouma, 2009, *Normalized (pointwise) mutual information in collocation extraction.*, w: *Proceedings of the Biennial GSCL Conference 2009*, strony 31—40.

Przykład

infekcja górnych dróg oddechowych

Noun_j Adj_i Noun_i Adj_i

infekcja | górnych dróg | oddechowych

infekcja górny droga oddechowy

bigram	NPMI
infekcja górny	0.66
górny droga	0.79
droga oddechowy	0.95

Porównanie dwóch metod

Poprawne gramatycznie podfrazy				Podfrazy z wykorzystaniem NPMI			
'*infekcja*'	'*górny*'	'*droga*'	'*oddechowy*'	'*infekcja*'	'*górny*'	'*droga*'	'*oddechowy*'
infekcja	*górnych*	*dróg*	*oddechowych*	*infekcja*	*górnych*	*dróg*	*oddechowych*
infekcja	*górnych*	*dróg*		—			
infekcja				*infekcja*			
	górne	*drogi*	*oddechowe*		*górne*	*drogi*	*oddechowe*
	górne	*drogi*		—			
		drogi	*oddechowe*			*drogi*	*oddechowe*
		drogi				*drogi*	

Preferowanie podziału na dwie frazy rzeczownikowe

prawidłowa$_{adj}$ *mikroflora*$_{noun}$ *górnych*$_{adj}$ *dróg*$_{noun}$ *oddechowych*$_{adj}$ —> *prawidłowa mikroflora* oraz *górne drogi oddechowe*

częste$_{adj}$ *infekcje*$_{noun}$ *górnych*$_{adj}$ *dróg*$_{noun}$ *oddechowych*$_{adj}$ —> *częste* modyfikuje całą frazę *infekcje górnych dróg oddechowych*

Modyfikacja:

- szukamy najsłabszej pozycji pozwalającej podzielić frazę na dwie podfrazy rzeczownikowe;
- jeśli różnica pomnięcy nasłabszym miejscem podziału a tym dzielącym na dwie frazy rzeczownikowe jest mniejsza od ustalonego progu to preferujemy podział na dwie frazy rzeczownikowe.

Analiza kontrastywna terminologii

Cel

Na podstawie porównania wyników ekstrakcji terminologii dla dwóch korpusów mają być wskazane frazy:

- bardziej specyficzne dla innej dziedziny (porównanie z terminologią wydobytą z innego korpusu dziedzinowego)
- terminy ogólne np." *własny sposób*, *lewa strona*, *trudne zadanie* (porównanie z korpusem języka ogólnego).

Metody porównań

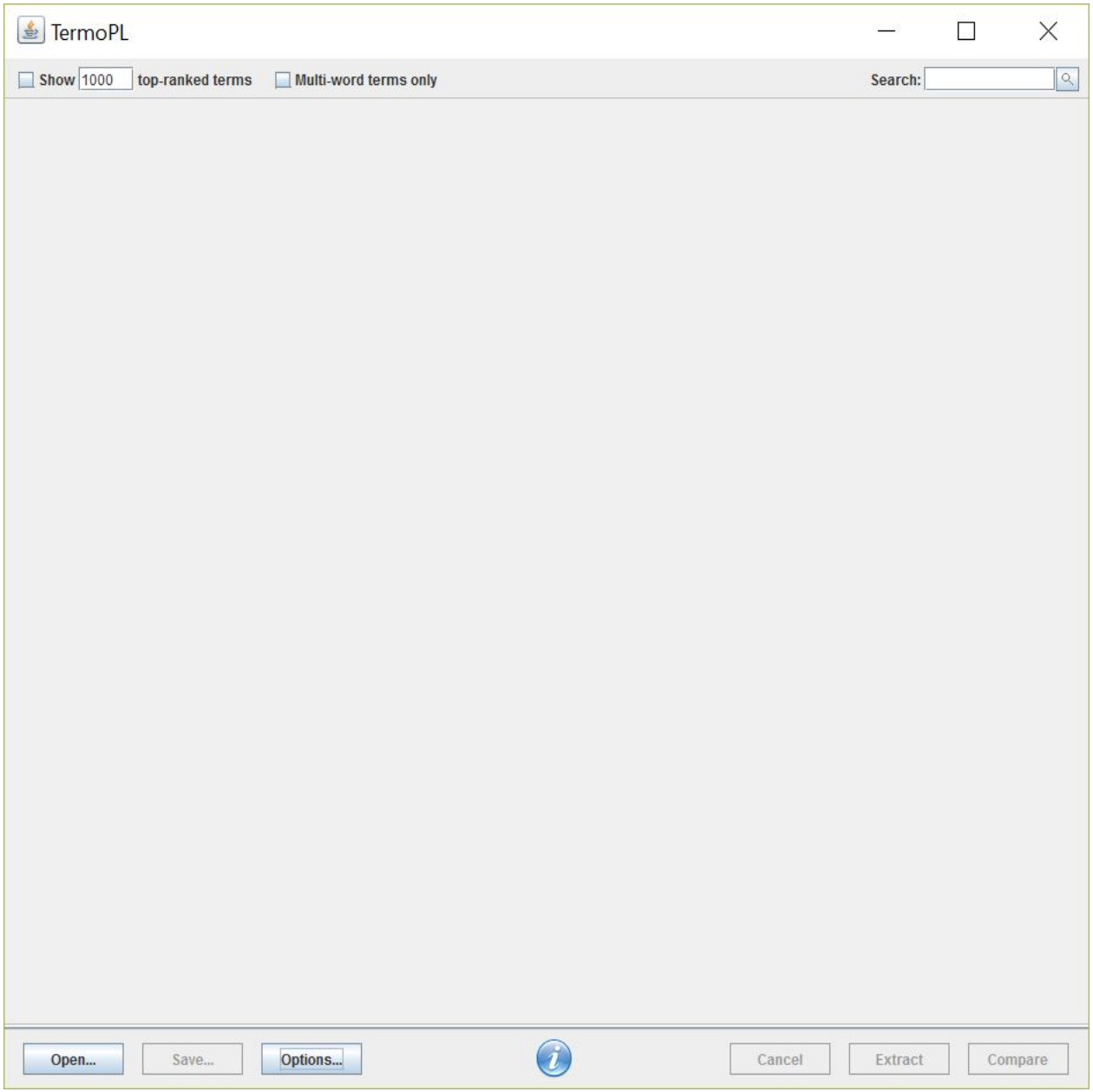
Zaimplementowane metody wykorzystują:

- Log-Likelihood (LL – logarytm wiarygodności): na ile różni się częstość konkretnego terminu w dwóch porównywanych korpusach;
- Term Frequency Inverse Term Frequency (TFITF): łączy częstość występowania w korpusie dziedzinowym z odwrotną częstością występowania w korpusie ogólnym (liczoną jako stosunek wielkości korpusu do częstości badanego terminu);
- Contrastive Selection of Multi-Word Terms(CSmw): dla terminów wielowyrazowych, uwzględnia zarówno częstość występowania pełnych terminów, ale też częstość występowania słów stanowiących element główny badanej frazy.

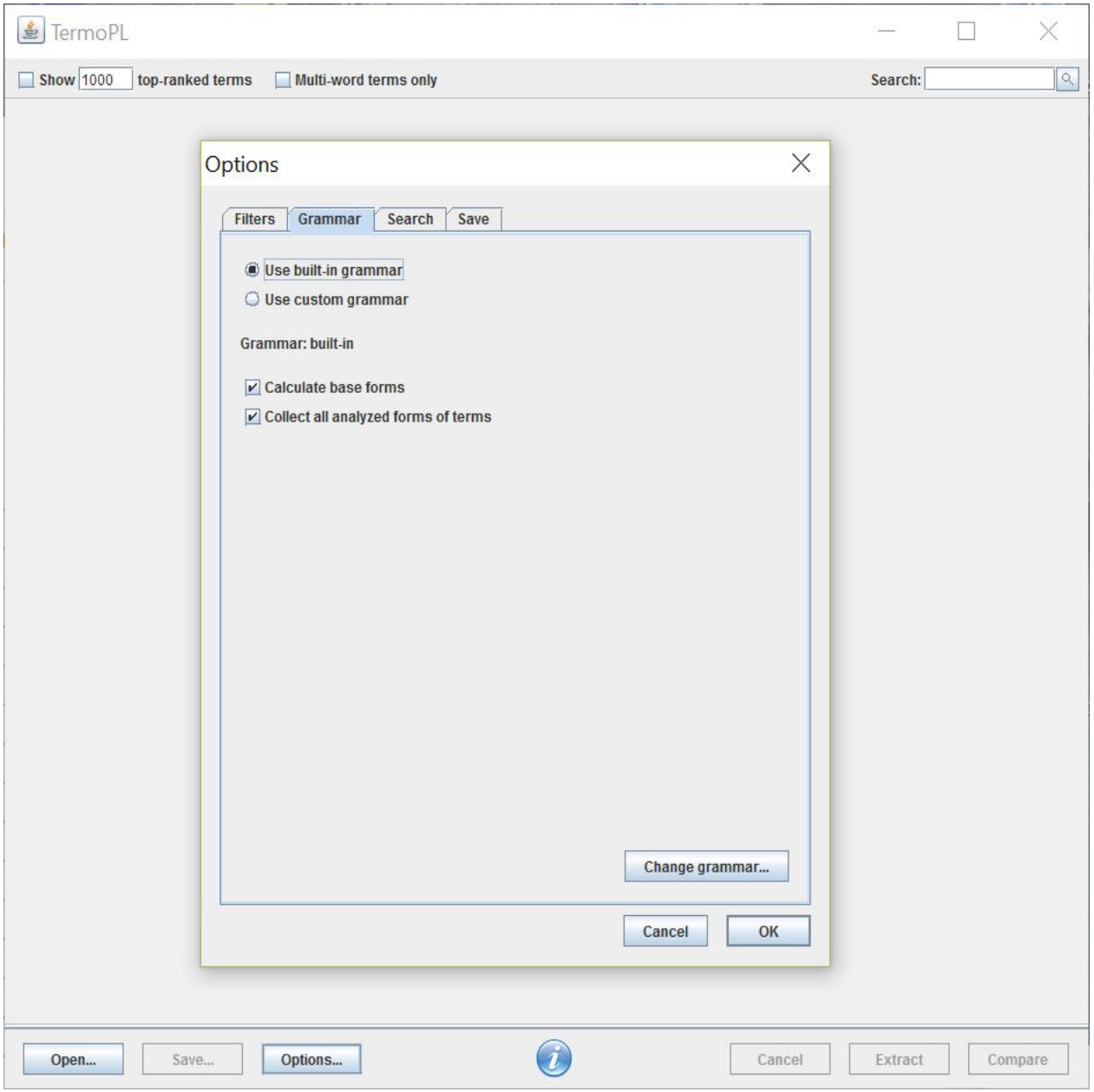
Proszę ściągnąć i przetworzyć Korpusomatem

Tekst książki ze strony: `https://wolnelektury.pl`
Autor: Konrad Gliściński
Tytuł: Wszystkie prawa zastrzeżone

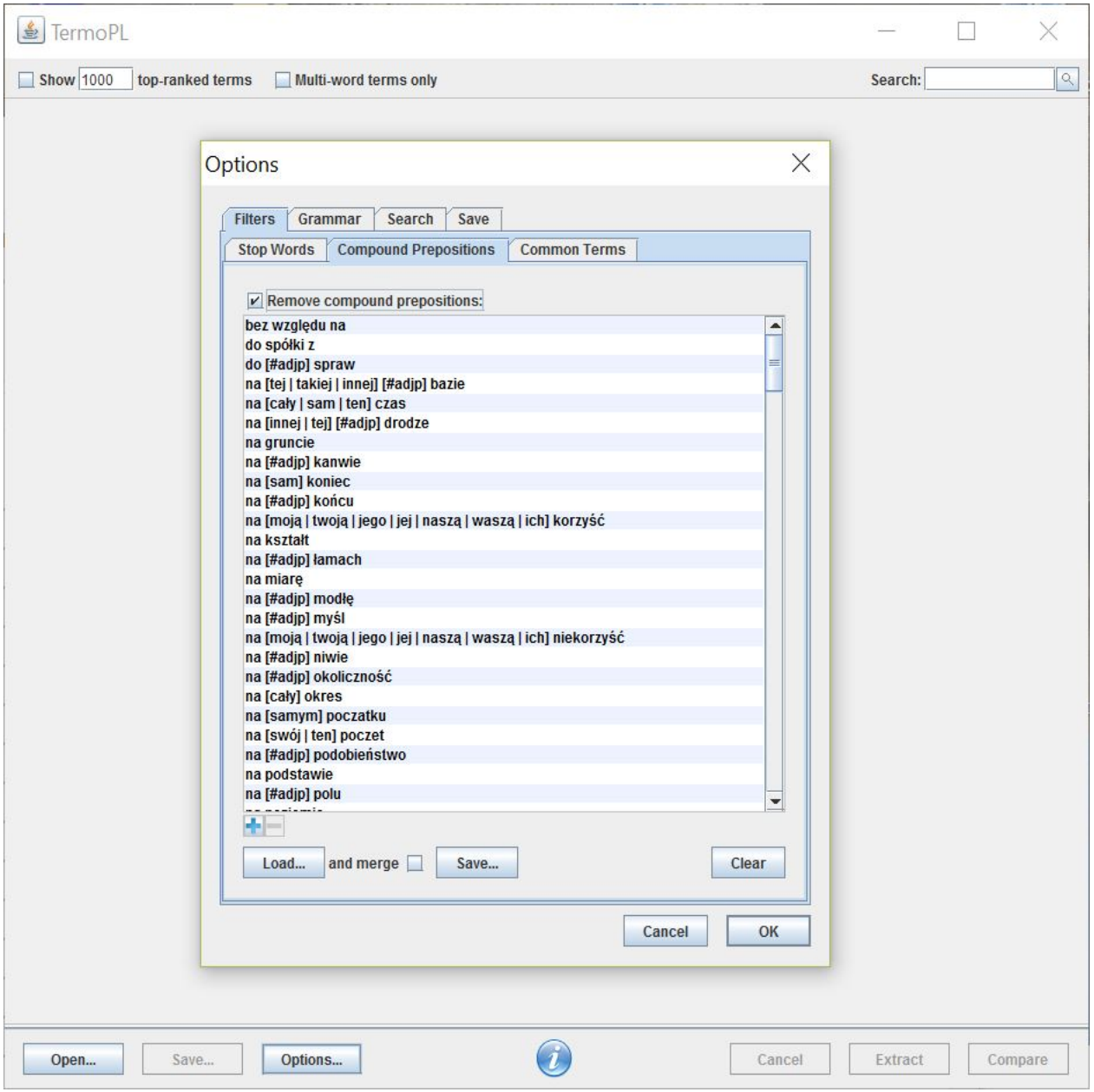
Otwieramy TermoPL



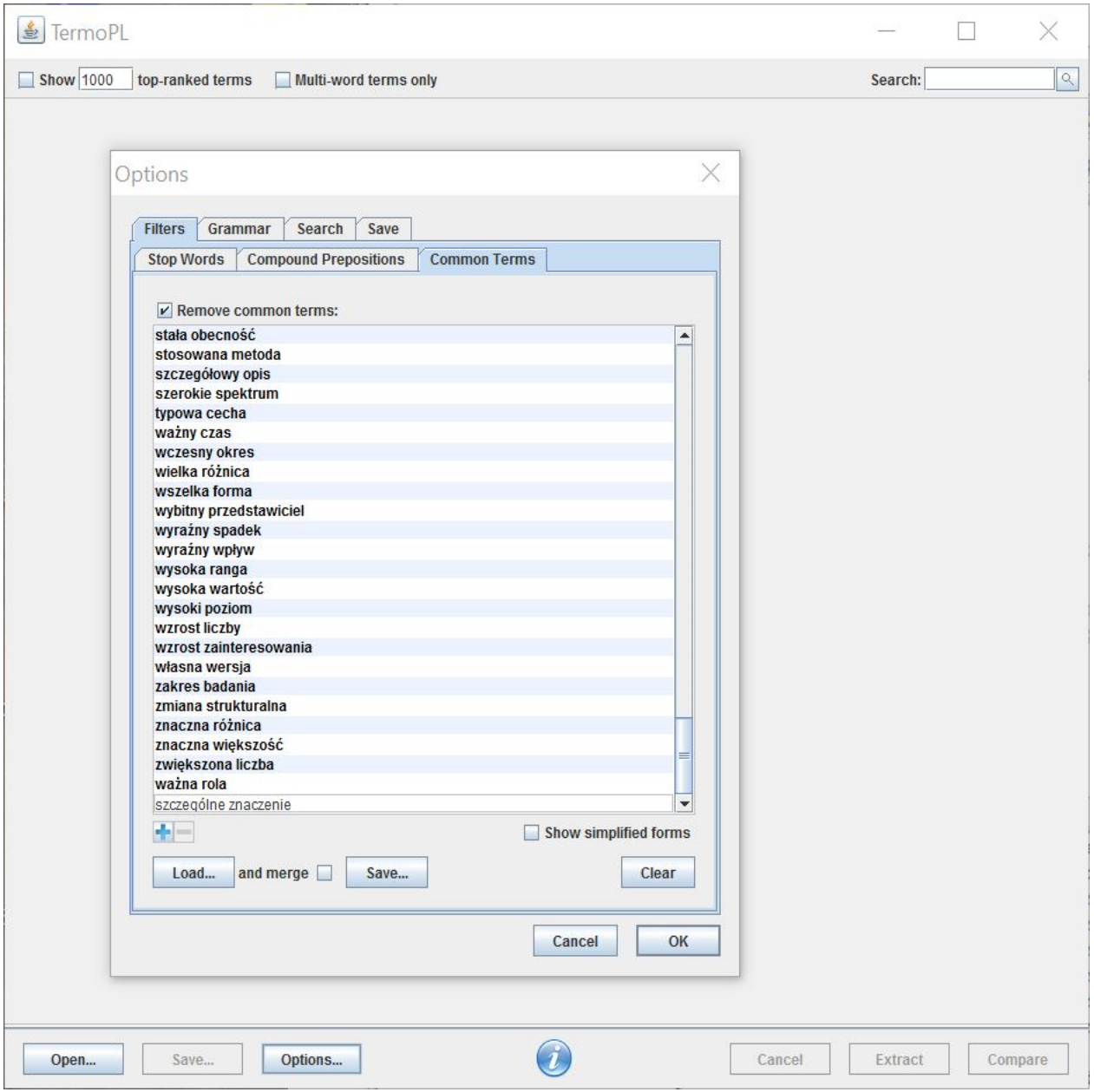
Opcje – gramatyka



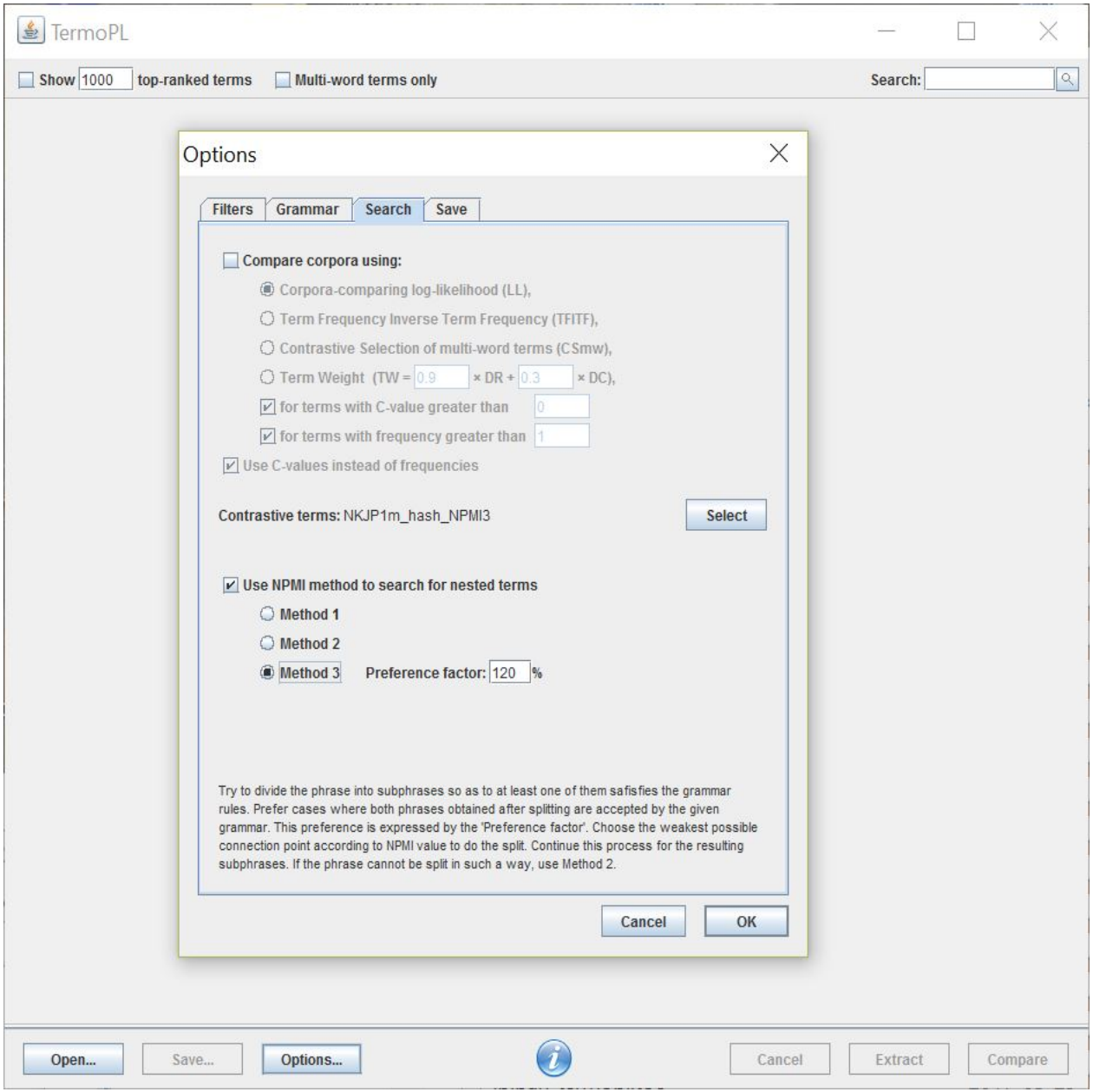
Filtry – przyimki złożone



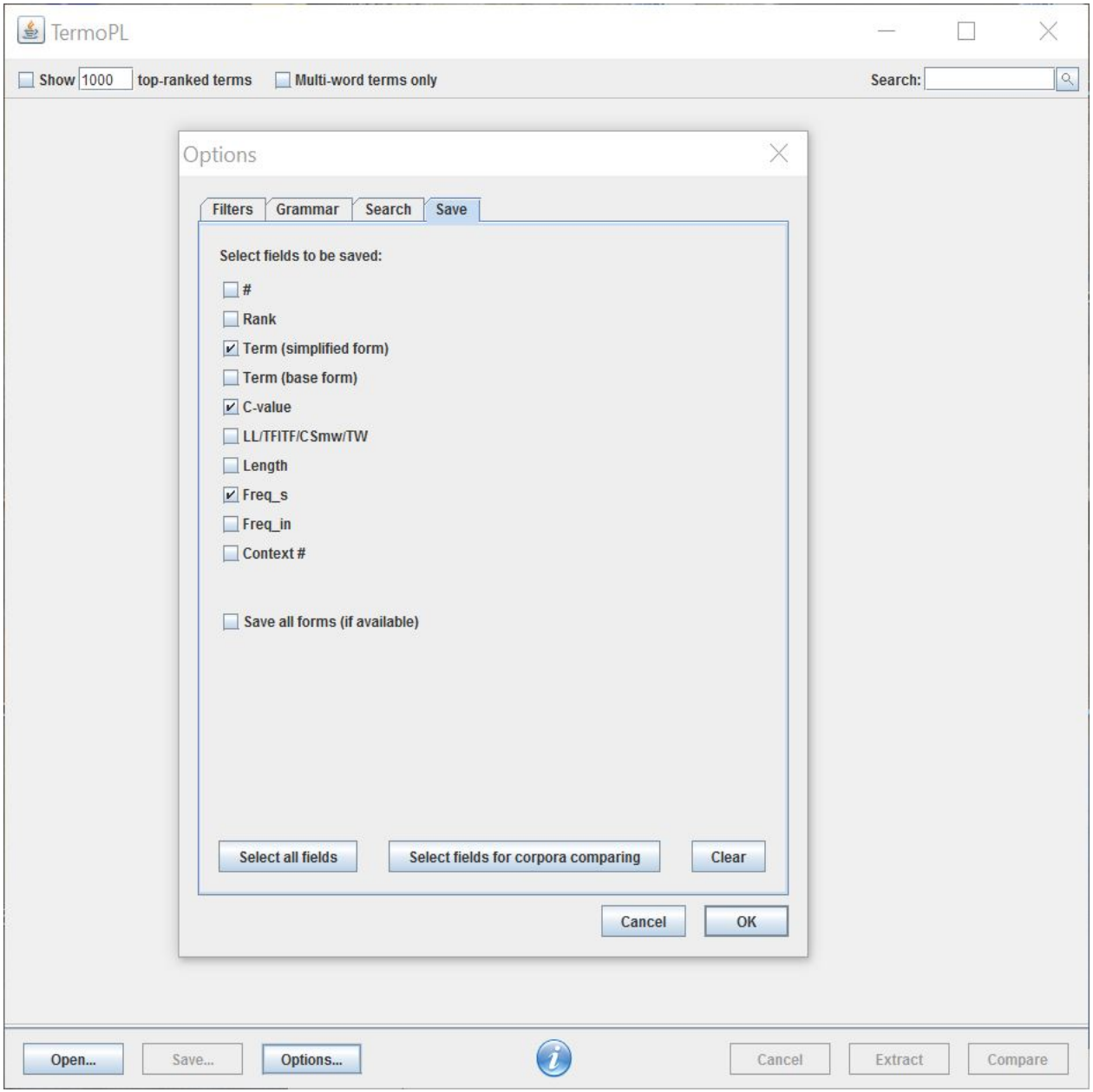
Filtry – dodawanie nowych



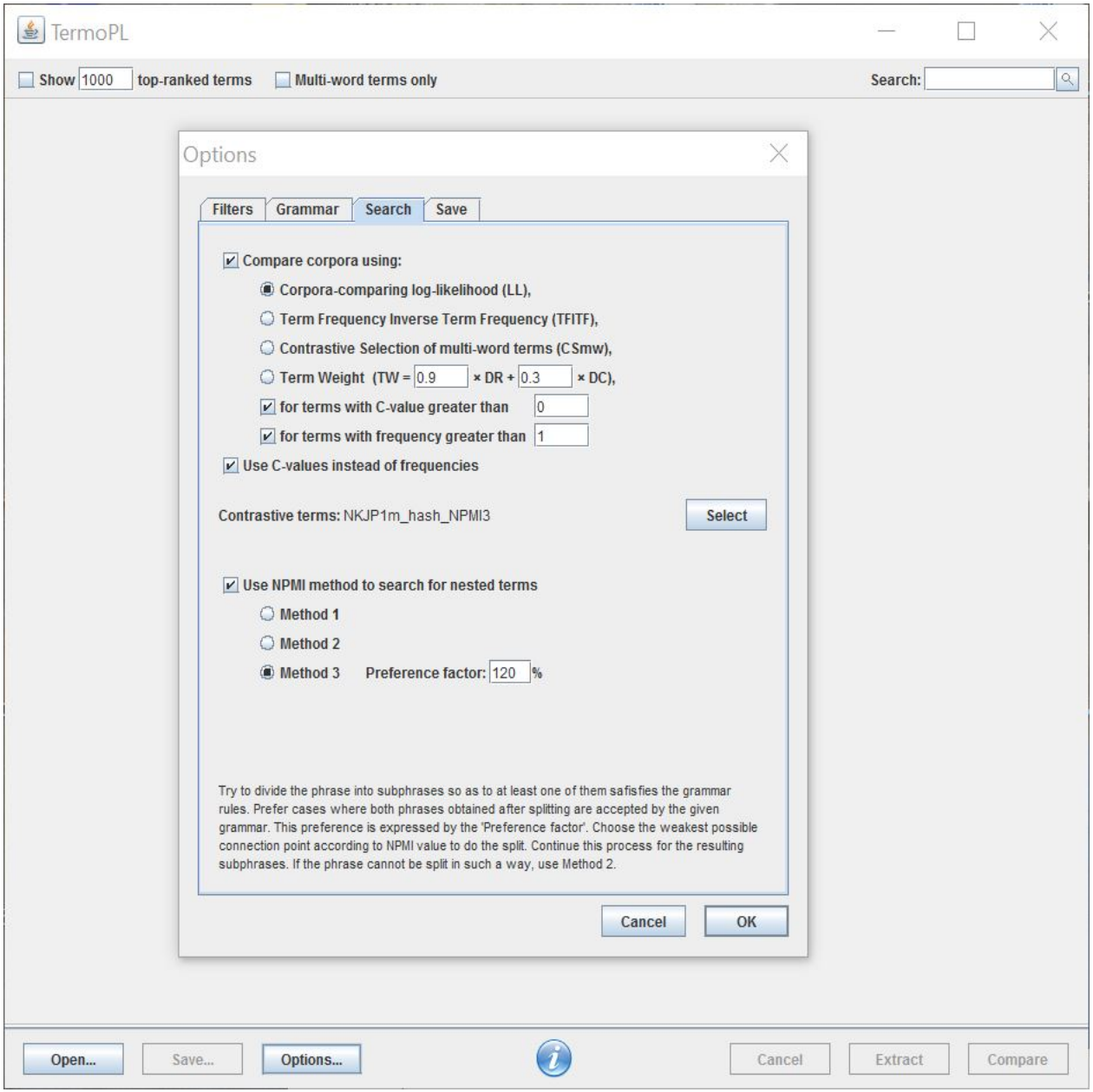
Podfrazy


TermoPL
Show 1000 top-ranked terms
Multi-word terms only
Search:
Options
Filters
Grammar
Search
Save
Compare corpora using:
Corpora-comparing log-likelihood (LL),
Term Frequency Inverse Term Frequency (TFITF),
Contrastive Selection of multi-word terms (CSmw),
Term Weight (TW = 0.9 × DR + 0.3 × DC),
for terms with C-value greater than 0
for terms with frequency greater than 1
Use C-values instead of frequencies
Contrastive terms: NKJP1m_hash_NPMI3
Select
Use NPMI method to search for nested terms
Method 1
Method 2
Method 3
Preference factor: 120 %
Try to divide the phrase into subphrases so as to at least one of them safisfies the grammar rules. Prefer cases where both phrases obtained after splitting are accepted by the given grammar. This preference is expressed by the 'Preference factor'. Choose the weakest possible connection point according to NPMI value to do the split. Continue this process for the resulting subphrases. If the phrase cannot be split in such a way, use Method 2.
Cancel
OK
Open...
Save...
Options...
Cancel
Extract
Compare

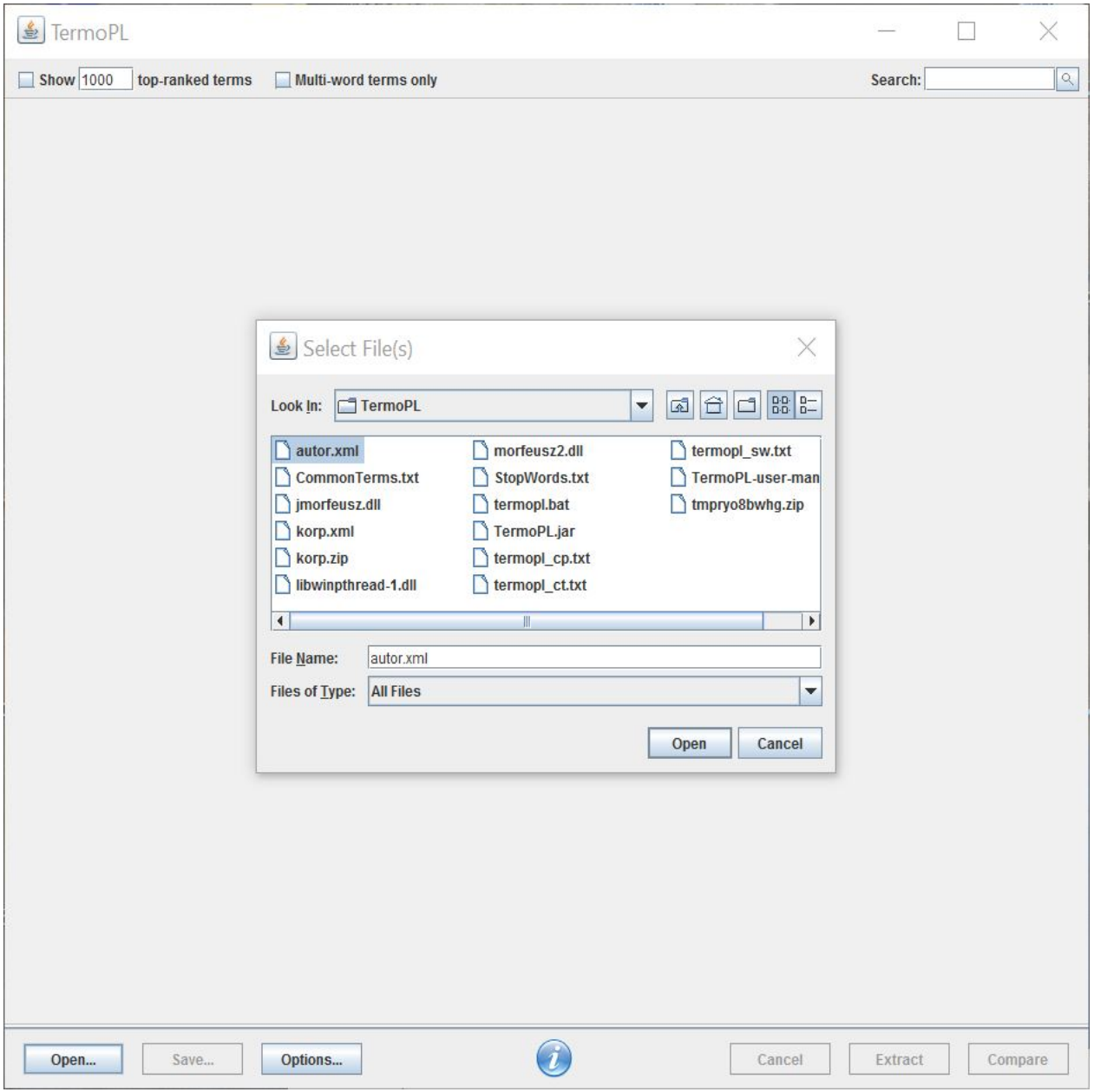
Zapis wyników



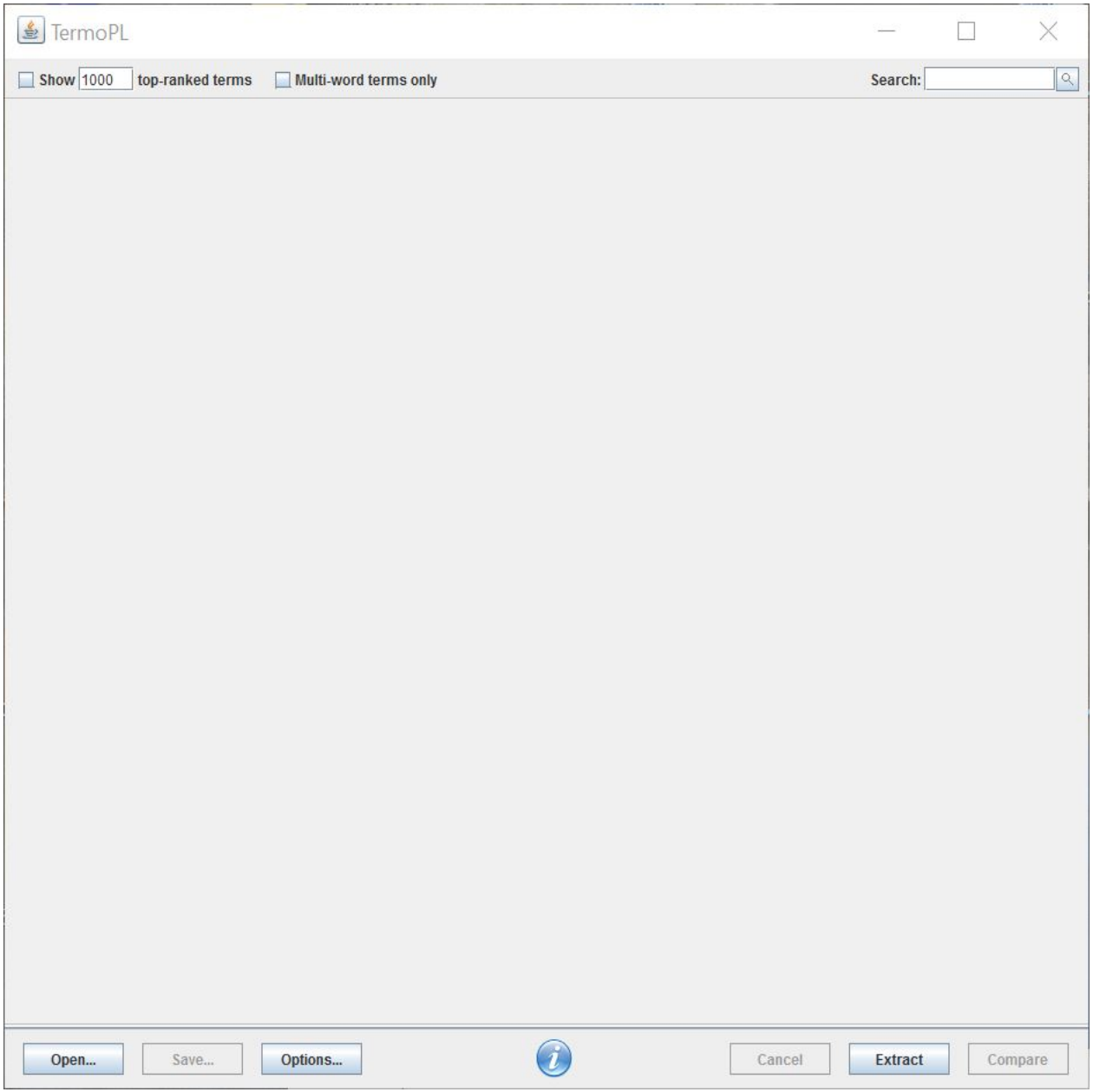
Porównanie z innym korpusem



Wybór korpusu do analizy



Ekstrakcja


TermoPL
Show 1000 top-ranked terms
Multi-word terms only
Search:
Open...
Save...
Options...
Cancel
Extract
Compare

Wynik ekstrakcji

TermoPL - autor_NPMI3

Show 1000 top-ranked terms Multi-word terms only Search:

#	Rank	Term	C-value	Length	Freq_s	Freq_in	Context #
1	1	prawo autorskie	926,13	2	929	528	184
2	2	prawo	319,75	1	3201	2436	701
3	3	prawo własności	276,24	2	278	141	80
4	4	utwór muzyczny	107,45	2	110	74	29
5	5	wyłączne prawo	98,73	2	101	84	37
6	6	autor	81,04	1	813	369	141
7	7	publiczne wykonywanie	80	2	84	80	20
8	8	utwór	76,7	1	770	502	169
9	9	prawo własności intelektualnej	76,55	3	50	34	20
10	10	własność	75,23	1	757	626	132
11	11	Statut Anny	75,15	2	77	24	13
12	12	prawo wyłączne	73,72	2	75	37	29
13	13	monopol autorski	69,21	2	71	43	24
14	14	własność literacka	63,05	2	65	39	20
15	15	konwencja berneńska	62,78	2	64	28	23
16	16	prawo autora	61,93	2	63	16	15
17	17	własność intelektualna	56,12	2	64	63	8
18	18	dobra niematerialne	53,78	2	55	11	9
19	19	prawo naturalne	48,27	2	50	19	11
20	20	książka	47,49	1	477	260	123
21	21	prawo osobiste	47,44	2	49	25	16
22	22	twórca	46,91	1	471	195	102
23	23	ochrona	42,89	1	431	298	143
24	24	publiczne wykonywanie utworów muzycznych	41,33	4	27	19	3
25	25	stan rzeczy	41	2	42	2	2
26	26	ustawa	38,97	1	392	190	84
27	27	wydawca	38,7	1	389	171	84
28	28	model biznesowy	38,33	2	40	20	12
29	29	wydawca muzyczny	38	2	39	4	4
30	30	przemysł muzyczny	35,74	2	37	24	19
31	31	system przywilejów	35	2	36	8	8
32	32	dzieło	34,44	1	346	191	122
33	33	przywilej	34,35	1	346	241	98
34	34	XIX wiek	34,2	2	37	14	5
35	35	utwór dramatyczny	34,09	2	36	21	11
36	36	interes publiczny	33,91	2	35	12	11
37	37	publiczne wykonywanie utworów	33,81	3	24	16	6
38	38	system prawa autorskiego	33,28	3	22	5	5
39	39	system	30,59	1	308	239	114

Open... Save... Options... Cancel Extract Compare

Wyłącznie wielowyrazowe

TermoPL - autor_NPMI3

Show 1000 top-ranked terms ☑ Multi-word terms only Search:

#	Rank	Term	C-value	Length	Freq_s	Freq_in	Context #
1	1	prawo autorskie	926,13	2	929	528	184
2	3	prawo własności	276,24	2	278	141	80
3	4	utwór muzyczny	107,45	2	110	74	29
4	5	wyłączne prawo	98,73	2	101	84	37
5	7	publiczne wykonywanie	80	2	84	80	20
6	9	prawo własności intelektualnej	76,55	3	50	34	20
7	11	Statut Anny	75,15	2	77	24	13
8	12	prawo wyłączne	73,72	2	75	37	29
9	13	monopol autorski	69,21	2	71	43	24
10	14	własność literacka	63,05	2	65	39	20
11	15	konwencja berneńska	62,78	2	64	28	23
12	16	prawo autora	61,93	2	63	16	15
13	17	własność intelektualna	56,12	2	64	63	8
14	18	dobra niematerialne	53,78	2	55	11	9
15	19	prawo naturalne	48,27	2	50	19	11
16	21	prawo osobiste	47,44	2	49	25	16
17	24	publiczne wykonywanie utworów muzycznych	41,33	4	27	19	3
18	25	stan rzeczy	41	2	42	2	2
19	28	model biznesowy	38,33	2	40	20	12
20	29	wydawca muzyczny	38	2	39	4	4
21	30	przemysł muzyczny	35,74	2	37	24	19
22	31	system przywilejów	35	2	36	8	8
23	34	XIX wiek	34,2	2	37	14	5
24	35	utwór dramatyczny	34,09	2	36	21	11
25	36	interes publiczny	33,91	2	35	12	11
26	37	publiczne wykonywanie utworów	33,81	3	24	16	6
27	38	system prawa autorskiego	33,28	3	22	5	5
28	40	prawo majątkowe	30,27	2	32	19	11
29	42	common law	29,25	2	33	15	4
30	44	organizacja zbiorowego zarządzania	28,53	3	19	6	6
31	48	prawo autorskie majątkowe	25,36	3	17	6	6
32	49	monopol eksploatacyjny	25	2	26	6	6
33	50	Stany Zjednoczone	23,92	2	25	13	12
34	51	XVIII wiek	23,5	2	26	10	4
35	53	dzieło sztuki	23	2	24	8	8
36	55	rynek wydawniczy	22,14	2	24	13	7
37	56	kompozycje muzyczne	22	2	23	3	3
38	57	ochrona prawna	21,91	2	23	12	11
39	59	prawo zakazowe	21	2	22	9	9

Open... Save... Options... Cancel Extract Compare

Zebrane formy

TermoPL - autor_NPMI3

Show 1000 top-ranked terms Multi-word terms only Search:

#	Rank	Term	C-value	Length	Freq_s	Freq_in	Context #
1	1	prawo autorskie	926,13	2	929	528	184
2	2	prawo	319,75	1	3201	2436	701
3	3	prawo własności	276,24	2	278	141	80
4	4	utwór muzyczny	107,45	2	110	74	29
5	5	wyłączne prawo	98,73	2	101	84	37
6	6	autor	81,04	1	813	369	141
7	7	publiczne wykonywanie	80	2	84	80	20
8	8	utwór	76,7	1	770	502	169
9	9	prawo własności intelektualnej	76,55	3	50	34	20
10	10	własność	75,23	1	757	626	132
11	11	Statut Anny	75,15	2	77	24	13
12	12	prawo wyłączne	73,72	2	75	37	29
13	13	monopol autorski	69,21	2	71	43	24
14	14	własność literacka	63,05	2	65	39	20
15	15	konwencja berneńska	62,78	2	64	28	23
16	16	prawo autora	61,93	2	63	16	15
17	17	własność intelektualna	56,12	2	64	63	8
18	18	dobra niematerialne	53,78	2	55	11	9
19	19	prawo naturalne	48,27	2	50	19	11
20	20	książka	47,49	1	477	260	123
21	21	prawo osobiste	47,44	2	49	25	16
22	22	twórca	46,91	1	471	195	102
23	23	ochrona	42,89	1	431	298	143
24	24	publiczne wykonywanie utworów muzycznych	41,33	4	27	19	3
25	25	stan rzeczy	41	2	42	2	2
26	26	ustawa	38,97	1	392	190	84
27	27	wydawca	38,7	1	389	171	84
28	28	model biznesowy	38,33	2	40	20	12
29	29	wydawca muzyczny	38	2	39	4	4
30	30	przemysł muzyczny	35,74	2	37	24	19
31	31	system przywilejów	35	2	36	8	8
32	32	dzieło	34,44	1	346	191	122
33	33	przywilej	34,35	1	346	241	98
34	34	XIX wiek	34,2	2	37	14	5

Forms:

prawa autorskiego[45,199], Prawo autorskie[23,0], Prawa autorskie[14,2], prawa autorskie[84,18], praw autorskich[48,277], prawo autorskie[78,14], prawem autorskim[22,5], prawami autorskimi[29,2], prawach autorskich[13,2], prawie autorskim[40,2], prawom autorskim[2,0], prawu autorskiemu[1,1], Prawa Autorskiego[0,5], Prawie Autorskim[1,0], prawie Autorskim[1,0], Prawo Autorskie[0,1]

Open... Save... Options... Cancel Extract Compare

Fraza „Owa możliwość rozszczepienia uprawnień”

TermoPL - autor_NPMI3

Show 1000 top-ranked terms ☑ Multi-word terms only Search:

#	Rank	Term	C-value	Length	Freq_s	Freq_in	Context #
226	223	wyłączne prawo sprzedaży	6,34	3	4	0	0
227	223	istota prawa własności	6,34	3	4	0	0
228	223	pojęcie własności literackiej	6,34	3	4	0	0
229	223	przemysł wydawnictw muzycznych	6,34	3	8	4	1
230	223	prawo mechanicznej reprodukcji	6,34	3	5	1	1
231	225	publiczne odtwarzanie	6,33	2	9	8	3
232	229	ancien régime	6,25	2	8	7	4
233	235	duża możliwość rozszczepiania prawa	6	4	3	0	0
234	235	specyficzna właściwość prawa autorskiego	6	4	3	0	0
235	235	Owa możliwość rozszczepiania uprawnień	6	4	3	0	0
236	235	twórczość muzyczna	6	2	7	4	4
237	235	regulacja prawna	6	2	7	2	2
238	235	monopol prawny	6	2	7	3	3
239	235	miejsca publiczne	6	2	6	0	0
240	235	porozumienie TRIPS	6	2	7	4	4
241	235	ustawodawstwo krajowe	6	2	6	0	0
242	235	rozpowszechnianie utworów	6	2	7	2	2
243	235	własność rzeczy	6	2	6	0	0
244	235	władza świecka	6	2	7	1	1
245	235	płyty gramofonowe	6	2	7	4	4
246	235	publiczne wykonanie	6	2	7	2	2
247	235	dystrybucja książek	6	2	6	0	0
248	235	niczyja zgoda	6	2	7	1	1
249	235	prawo podmiotowe	6	2	7	5	5
250	235	przedmiot ochrony	6	2	6	0	0
251	235	nowy wynalazek	6	2	7	1	1
252	235	sprzedaż książek	6	2	7	1	1
253	235	Józef Górski	6	2	6	0	0
254	235	pobieranie pożytków	6	2	6	0	0
255	235	prawo natury	6	2	7	1	1
256	235	centrum systemu	6	2	7	3	3
257	235	największy zysk	6	2	7	2	2
258	235	interes społeczny	6	2	7	1	1
259	235	historyczny punkt	6	2	6	0	0

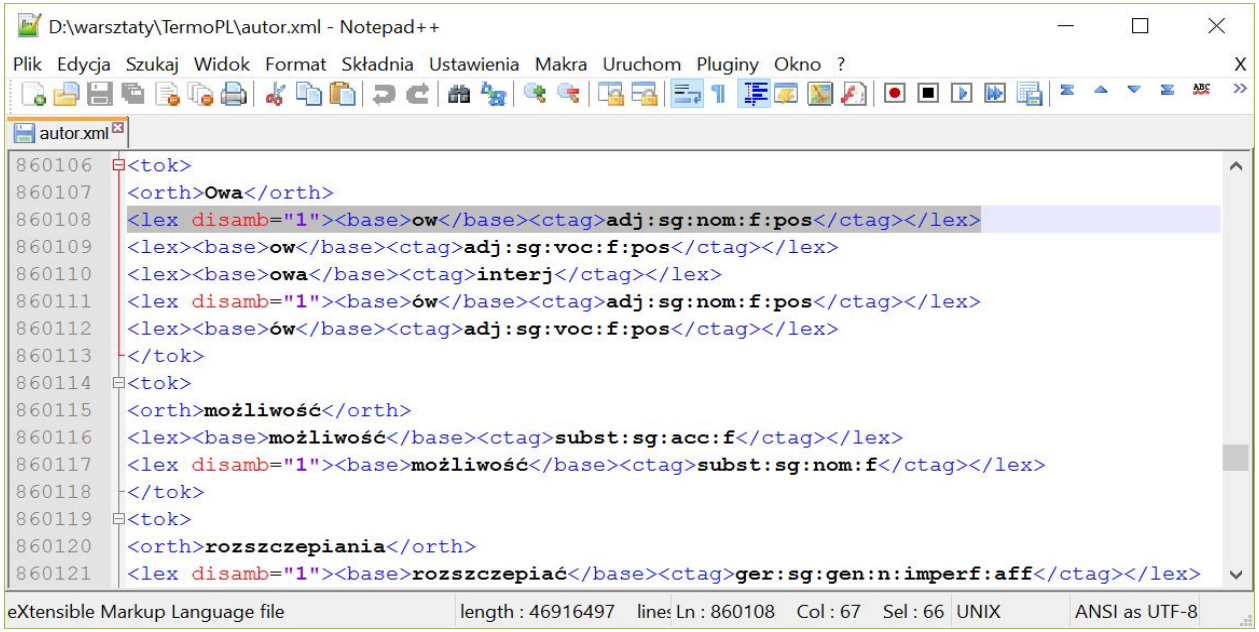
Forms:

Owa możliwość rozszczepiania uprawnień[3,0]

Open... Save... Options... Cancel Extract Compare

OW

Proszę dodać *'ow'* do listy stop words



By wrócić do możliwości ponownego ekstrahowania trzeba jeszcze raz wczytać tekst (klawisz Open)

Terminy bez ‘Owa’

TermoPL - autor_NPMI3

Show 1000 top-ranked terms ☑ Multi-word terms only Search: możliwość

#	Rank	Term	C-value	Length	Freq_s	Freq_in	Context #
331	270	członkowie gildii	5	2	7	2	1
332	270	interesa osobiste	5	2	6	2	2
333	270	cena książek	5	2	5	0	0
334	270	działalność twórcza	5	2	8	3	1
335	276	trwanie praw autorskich	4,75	3	3	0	0
336	276	koniec XVIII wieku	4,75	3	4	1	1
337	276	zakres przyznawanych praw	4,75	3	3	0	0
338	276	zasada domeny publicznej	4,75	3	4	1	1
339	276	prowadzona działalność gospodarcza	4,75	3	4	2	2
340	276	prawo wyłączności druku	4,75	3	3	0	0
341	276	natura praw autorskich	4,75	3	3	0	0
342	276	koncepcja własności intelektualnej	4,75	3	3	0	0
343	276	okres życia twórcy	4,75	3	3	0	0
344	276	fakt stworzenia dzieła	4,75	3	3	0	0
345	276	teoria własności Locke'a	4,75	3	3	0	0
346	276	zwolennik prawa własności	4,75	3	3	0	0
347	276	autorowie utworów dramatycznych	4,75	3	3	0	0
348	276	model biznesowy oparty	4,75	3	3	0	0
349	276	możliwość rozszczepiania uprawnień	4,75	3	3	0	0
350	276	naruszenie prawa autorskiego	4,75	3	3	0	0
351	276	koncepcja praw naturalnych	4,75	3	4	1	1
352	276	pismo wszelkiego rodzaju	4,75	3	4	1	1
353	276	konstrukcja praw zależnych	4,75	3	3	0	0
354	276	ochrona praw osobistych	4,75	3	4	1	1
355	276	grunt prawa autorskiego	4,75	3	3	0	0
356	276	rozciągnięcie praw autorskich	4,75	3	3	0	0
357	276	mechaniczne powielanie tekstu	4,75	3	4	2	2
358	276	utwory żyjących autorów	4,75	3	3	0	0
359	276	letni okres ochrony	4,75	3	3	0	0
360	276	prawo publiczne wykonywania	4,75	3	3	0	0
361	276	właściciele praw autorskich	4,75	3	3	0	0
362	276	prawo własności artystycznej	4,75	3	4	1	1
363	276	wystawianie utworów dramatycznych	4,75	3	4	1	1
364	276	ekonomia funkcjonowania rzeczy	4,75	3	3	0	0

Forms:

możliwość rozszczepiania uprawnień[3,0]

Open... Save... Options... Cancel Extract Compare

Formy uproszczone

TermoPL - autor_NPMI3

Show 1000 top-ranked terms ☑ Multi-word terms only Search: możliwość

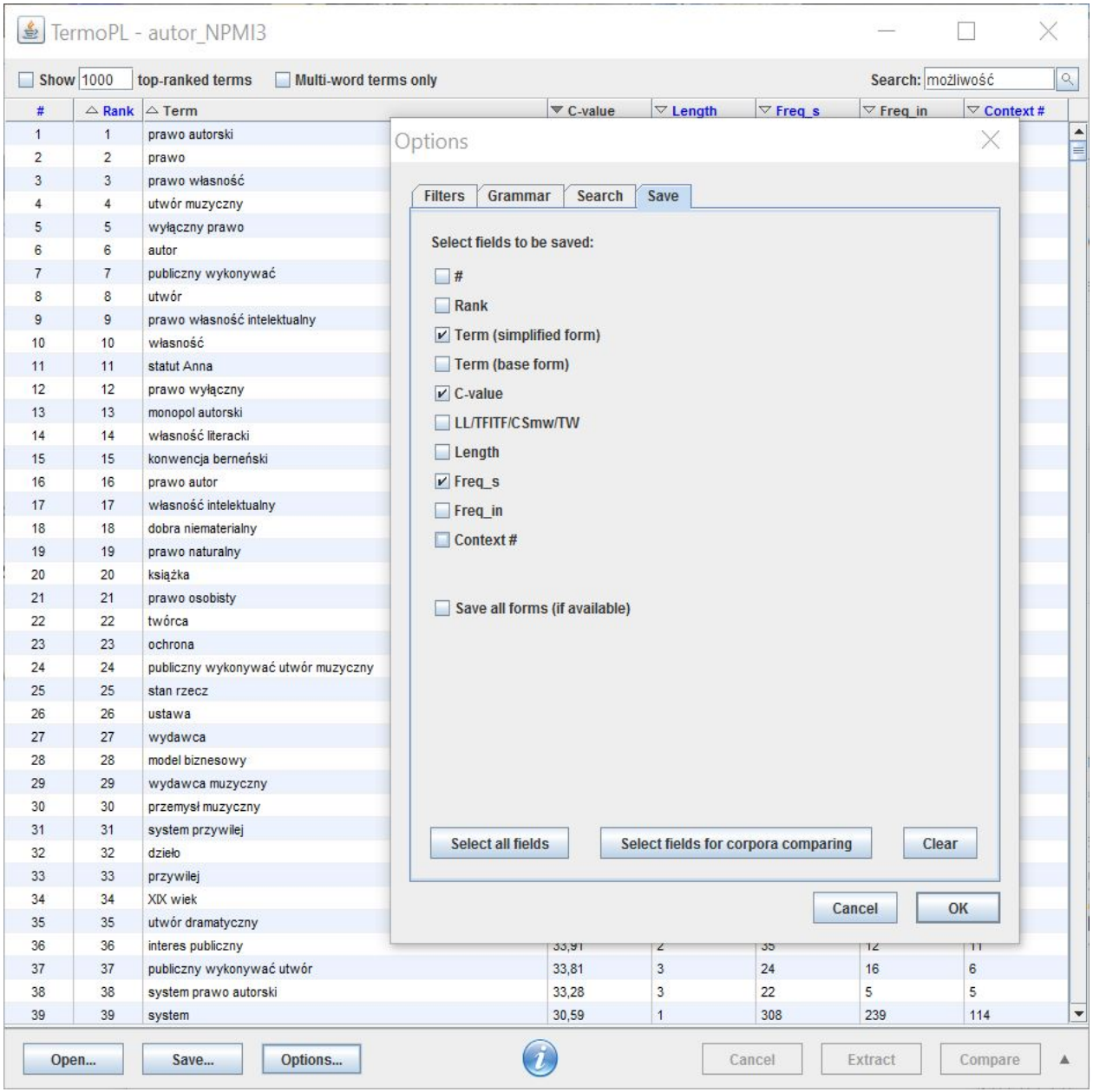
#	Rank	Term	C-value	Length	Freq_s	Freq_in	Context #
1	1	prawo autorski	926,13	2	929	528	184
2	3	prawo własność	276,24	2	278	141	80
3	4	utwór muzyczny	107,45	2	110	74	29
4	5	wyłączny prawo	98,73	2	101	84	37
5	7	publiczny wykonywać	80	2	84	80	20
6	9	prawo własność intelektualny	76,55	3	50	34	20
7	11	statut Anna	75,15	2	77	24	13
8	12	prawo wyłączny	73,72	2	75	37	29
9	13	monopol autorski	69,21	2	71	43	24
10	14	własność literacki	63,05	2	65	39	20
11	15	konwencja berneński	62,78	2	64	28	23
12	16	prawo autor	61,93	2	63	16	15
13	17	własność intelektualny	55,14	2	64	62	7
14	18	dobra niematerialny	53,78	2	55	11	9
15	19	prawo naturalny	48,27	2	50	19	11
16	21	prawo osobisty	47,44	2	49	25	16
17	24	publiczny wykonywać utwór muzyczny	41,33	4	27	19	3
18	25	stan rzecz	41	2	42	2	2
19	28	model biznesowy	38,33	2	40	20	12
20	29	wydawca muzyczny	38	2	39	4	4
21	30	przemysł muzyczny	35,74	2	37	24	19
22	31	system przywilej	35	2	36	8	8
23	34	XIX wiek	34,2	2	37	14	5
24	35	utwór dramatyczny	34,09	2	36	21	11
25	36	interes publiczny	33,91	2	35	12	11
26	37	publiczny wykonywać utwór	33,81	3	24	16	6
27	38	system prawo autorski	33,28	3	22	5	5
28	40	prawo majątkowy	30,27	2	32	19	11
29	42	common lawa	29,25	2	33	15	4
30	44	organizacja zbiorowy zarządzanie	28,53	3	19	6	6
31	48	prawo autorski majątkowy	25,36	3	17	6	6
32	49	monopol eksploatacyjny	25	2	26	6	6
33	50	Stany zjednoczyć	23,92	2	25	13	12
34	51	XVIII wiek	23,5	2	26	10	4

Forms:

prawa autorskiego[45,199], Prawo autorskie[23,0], Prawa autorskie[14,2], prawa autorskie[84,18], praw autorskich[48,277], prawo autorskie[78,14], prawem autorskim[22,5], prawami autorskimi[29,2], prawach autorskich[13,2], prawie autorskim[40,2], prawom autorskim[2,0], prawu autorskiemu[1,1], Prawa Autorskiego[0,5], Prawie Autorskim[1,0], prawie Autorskim[1,0], Prawo Autorskie[0,1]

Open... Save... Options... Cancel Extract Compare

Plik do porównywania terminów



Porównanie z NKJP

TermoPL - autor_NPMI3 vs. NKJP1m_hash_NPMI3 [LL]

Show 1000 top-ranked terms Multi-word terms only Search: możliwość

#	Rank	Term	C-value	LL	Length	Freq_s	Freq_in
1	1	prawo autorski	926,13	1503,41	2	929	528
2	2	prawo	319,75	371,34	1	3201	2435
3	3	prawo własność	276,24	443,56	2	278	141
4	4	utwór muzyczny	107,45	-	2	110	74
5	5	wyłączny prawo	98,73	-	2	101	84
6	6	autor	81,04	91,85	1	813	369
7	7	publiczny wykonywać	80	-	2	84	80
8	8	utwór	76,7	106,42	1	770	502
9	9	prawo własność intelektualny	76,55	-	3	50	34
10	10	własność	75,22	105,54	1	757	623
11	11	statut Anna	75,15	-	2	77	24
12	12	prawo wyłączny	73,72	118,3	2	75	37
13	13	monopol autorski	69,21	-	2	71	43
14	14	własność literacki	63,05	-	2	65	39
15	15	konwencja berneński	62,78	-	2	64	28
16	16	prawo autor	61,93	-	2	63	16
17	17	własność intelektualny	55,14	-	2	64	62
18	18	dobra niematerialny	53,78	-	2	55	11
19	19	prawo naturalny	48,27	-	2	50	19
20	20	książka	47,49	34,79	1	477	260
21	21	prawo osobisty	47,44	-	2	49	25
22	22	twórca	46,91	60,86	1	471	195
23	23	ochrona	42,89	38,99	1	431	298
24	24	publiczny wykonywać utwór muzyczny	41,33	-	4	27	19
25	25	stan rzecz	41	35,34	2	42	2
26	26	ustawa	38,97	15,21	1	392	190
27	27	wydawca	38,7	56,68	1	389	171
28	28	model biznesowy	38,33	-	2	40	20
29	29	wydawca muzyczny	38	-	2	39	4
30	30	przemysł muzyczny	35,74	-	2	37	24
31	31	system przywilej	35	-	2	36	8
32	32	dzieło	34,44	35,78	1	346	190
33	33	przywilej	34,35	51,75	1	346	241
34	34	XIX wiek	34,2	27,12	2	37	14
35	35	utwór dramatyczny	34,09	-	2	36	21
36	36	interes publiczny	33,91	52,67	2	35	12
37	37	publiczny wykonywać utwór	33,81	-	3	24	16
38	38	system prawo autorski	33,28	-	3	22	5

Open... Save... Options... Cancel Extract Compare

Wielowyrazowe terminy posortowane po różnicy

TermoPL - autor_NPMI3 vs. NKJP1m_hash_NPMI3 [LL]

Show 1000 top-ranked terms ☑ Multi-word terms only Search: możliwość

#	Rank	Term	C-value	LL	Length	Freq_s	Freq_in
1	1	prawo autorski	926,13	1503,41	2	929	528
2	3	prawo własność	276,24	443,56	2	278	141
3	12	prawo wyłączny	73,72	118,3	2	75	37
4	36	interes publiczny	33,91	52,67	2	35	12
5	40	prawo majątkowy	30,27	39,4	2	32	19
6	25	stan rzecz	41	35,34	2	42	2
7	51	XVIII wiek	23,5	31,42	2	26	10
8	63	wielki Brytania	20	29,95	2	21	2
9	64	stacja radiowy	19,8	29,63	2	21	12
10	70	rzecz materialny	18,5	27,52	2	20	6
11	34	XIX wiek	34,2	27,12	2	37	14
12	53	dzieło sztuka	23	26,29	2	24	8
13	57	ochrona prawny	21,91	25,95	2	23	12
14	75	wolność prasa	17,5	25,9	2	19	6
15	72	podmiot uprawniony	18	24,6	2	19	6
16	63	XX wiek	20	18,95	2	21	3
17	107	śmierć autor	13	16,81	2	13	0
18	125	wielki rewolucja	11,6	16,42	2	14	12
19	110	dzieło literacki	12,6	16,19	2	14	7
20	119	koniec koniec	12	15,27	2	12	0
21	149	następca prawny	10	13,88	2	10	0
22	137	własność prywatny	11	12,41	2	12	4
23	112	XVI wiek	12,5	12,41	2	14	6
24	172	określony czas	9	12,3	2	10	3
25	172	wyłączny korzystać	9	12,3	2	11	10
26	172	potrzeba ochrona	9	12,3	2	10	7
27	149	źródło dochód	10	12,23	2	12	8
28	160	ochrona prawo autorski	9,51	12,1	3	7	2
29	156	nowy technologia	9,8	11,93	2	11	6
30	176	prawo rzymski	8,67	11,78	2	10	4
31	147	czas trwać	10,33	11,45	2	12	10
32	173	światowy organizacja handel	8,98	11,3	3	7	4
33	188	prawo prywatny	8	10,73	2	9	3
34	136	wielki rewolucja francuski	11,09	10,72	3	9	6
35	582	unia europejski	1	10,26	2	3	2
36	169	światowy organizacja	9,17	10,15	2	11	11
37	188	polski ustawa	8	9,25	2	10	4
38	128	wszelki rodzaj	11,29	9,19	2	13	12

Open... Save... Options... Cancel Extract Compare