

Wyszukiwarka PELCRA dla danych NKJP

Piotr Pężik

14.1. O wyszukiwarce

Wyszukiwarka PELCRA¹ to jedna z dwóch wyszukiwarek internetowych dostępnych dla danych NKJP. Narzędzie to pozwala na szybkie i wygodne przeszukiwanie zasobów zgromadzonych na potrzeby Narodowego Korpusu Języka Polskiego. Wyszukiwarka jest oparta na składni zapytań korpusowych, która z jednej strony oferuje funkcjonalność porównywalną z opcjami wyszukiwania dostępnymi w standardowych narzędziach korpusowych, a z drugiej umożliwia szczególnie skuteczne wyszukiwanie pojedynczych wyrazów, wariantów morfologicznych i semantycznych oraz elastycznych wielowyrazowych kolokacji w obszernych zbiorach danych NKJP. W czasie powstawania niniejszej publikacji wyszukiwarka pozwalała na bardzo szybkie przeszukiwanie ponad 1500 milionów słów tekstowych znajdujących się w ogólnej puli danych NKJP. Dla przykładu, niezapisane w pamięci podręcznej serwera zapytanie o wszystkie wystąpienia różnych form rzeczownika *brama*, występującego w ponad 54 tysiącach kontekstów korpusowych zajmuje niecałe 0,2 sekundy. Co ważne, wyszukiwarka PELCRA zawsze podaje całkowitą liczbę kontekstów wystąpień wyrazów, lub fraz pasujących do danego zapytania, nawet dla bardzo częstych wyrazów lub fraz.

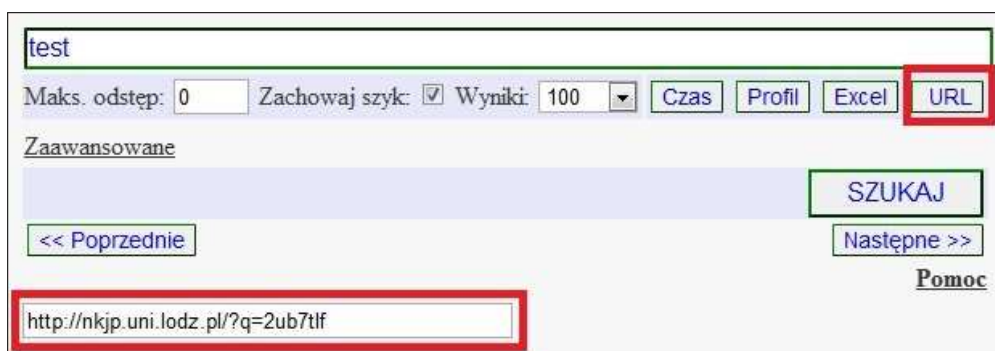
¹ PELCRA to akronim nazwy zespołu badawczego działającego od 1997 roku w Instytucie Anglistyki Uniwersytetu Łódzkiego – zob. <http://pelcra.ia.uni.lodz.pl>. Opisana w tym rozdziale wyszukiwarka dla danych NKJP, zwana po prostu *wyszukiwarką PELCRA*, dostępna jest pod adresem <http://nkjp.uni.lodz.pl/>.

Inną wyszukiwarką dostępną od wczesnych etapów projektu jest internetowa wersja silnika Poliqarp (Janus i Przepiórkowski 2007)², która została przystosowana do przeszukiwania danych NKJP. Do niewątpliwych zalet tej wyszukiwarki należy jej bogata składnia, pozwalająca formułować między innymi zaawansowane zapytania uwzględniające kryteria morfosyntaktyczne, które wykorzystują anotację lingwistyczną całego korpusu. Szczegółowy opis funkcji Poliqarpa dla NKJP zostały opisane na stronach pomocy tej wyszukiwarki (<http://nkjp.pl/poliqarp/help/pl.html>). Składnia Poliqarpa została szczegółowo opisana w Przepiórkowski 2004.

14.2. Skrócone odsyłacze

Zanim przejdziemy do omawiania poszczególnych funkcji wyszukiwarki PEL-CRA, warto wprowadzić opcję generowania skompresowanych odsyłaczy do wyników. Aby ułatwić użytkownikom odtwarzanie wyników wysyłanych do wyszukiwarki zapytań można za pomocą przycisku URL stworzyć krótki odsyłacz do bieżącego ekranu zapytania. Odsyłacz zostaje wyświetlony tuż pod oknem wyszukiwania (zob. rys. 14.1).

Rysunek 14.1. Przykład skróconego odsyłacza



W skompresowanym odsyłaczu zakodowane są wszystkie informacje o wybranych opcjach wyszukiwania. Po jego kliknięciu wyświetlany zostanie nie tylko ekran zapytania, ale też wyniki, które zwraca dane zapytanie. Taki skompresowany odsyłacz można łatwo zapisać, zamieścić w publikacji, lub przesłać pocztą elektroniczną. Przy większości przykładów omawianych poniżej podano bezpośredni skrócony odsyłacz do ekranu wyników pasujących do danego zapytania, dzięki czemu czytelnik może sprawdzić działanie zapytań.

² W ramach NKJP Poliqarp był rozwijany i utrzymywany przez Jakuba Wilka.

14.3. Składnia zapytań w przykładach

14.3.1. Wyszukiwanie dokładnych dopasowań pojedynczych wyrazów

Aby wyszukać wystąpienia danego słowa w korpusie, należy je wpisać w szerokim polu tekstowym na górze formularza zapytania. Po kliknięciu przycisku Szukaj wyświetlą się wystąpienia tego słowa w zindeksowanym korpusie. Na przykład po wpisaniu wyrazu *tymianek*³ powinny się ukazać konteksty zawierające jego dokładne dopasowania (rys. 14.2)⁴.

Rysunek 14.2. Przykład dokładnego dopasowania dla wyrazu *tymianek*

The screenshot shows a search interface with the following elements:

- Search input field containing "tymianek".
- Advanced options: "Maks. odstęp: 0", "Zachowaj szyk: ", "Wyniki: 100", and buttons for "Czas", "Profil", "Excel", and "URL".
- A "Zaawansowane" section with a "SZUKAJ" button.
- Navigation buttons: "<< Poprzednie" and "Następne >>".
- A "Pomoc" link.
- Summary text: "Przeszukiwany zbiór zawiera 1.225.343.686 słów. Znaleziono 370 akapitów pasujących do zapytania w 0.013s."
- A table of 10 search results, each with a green plus icon in the right column.

1.	mężczyzny zmarzurza; Jest taki mróz. Nie jesteś dla mnie	tymianek	ni róża, Ani też	Wiersze	+
2.	o zawartości. Kwiat lipy, skrzyp polny, macierzanka,	tymianek	, mniszek lekar	Kariera na trzy karp...	+
3.	Wściekłość rozsądzała go niczym ogrodniczka zbyt gęsto posiany	tymianek	.	Wilc... i śmierć ban...	+
4.	lekarski, kosaciec, kolender, kora chinowca, majeranek, szalwia,	tymianek	, cynamon, run	Vinum sacrum et prof...	+
5.	dorodne poziomki, błękitne fiołki, białe konwalie i fioletowy	tymianek	. Miasto tonie	Twierdze kresowe Rze...	+
6.	+ cynamon koper, bazylia	tymianek	rozmaryn	Makrobiotyka w polsk...	+
7.	używamy cebuli, czosnku, przypraw ziołowych (cząber, estragon,	tymianek	, kolendra, koz	Makrobiotyka w polsk...	+
8.	prażymy na patelni, następnie dodajemy zioła przyprawowe (tymianek	, majeranek, es	Makrobiotyka w polsk...	+
9.	mąkę, aby uzyskać bardziej stałą konsystencję. Dodajemy również	tymianek	i sól. Po wyrob	Makrobiotyka w polsk...	+
10.	mąką i wszystko razem krótko przesmażamy. Następnie dodajemy sól,	tymianek	i zestawiamy z	Makrobiotyka w polsk...	+

Ogólna liczba kontekstów pasujących do zapytania jest podawana bezpośrednio nad tabelą wyników. Wyniki można *posortować* według dopasowania (ma to sens w przypadkach opisywanych poniżej), lub też według *lewego* albo *prawego* słowa w konkordancji. Możliwe jest też określenie *maksymalnej liczby wyników* pojawiających się na stronie. Przechodzenie między kolejnymi stronami wyników umożliwiają przyciski << *Poprzednie* oraz *Następne* >>. Informacje o tekście, z którego pochodzi dany cytat, a także szerszy kontekst wystąpienia można uzyskać poprzez kliknięcie symbolu zielonego kółka z plusem w ostatniej kolumnie danego wiersza wyników.

³ Zob. <http://nkjp.uni.lodz.pl/?q=yhfmprx>.

⁴ Dopasowaniem nazywamy tu wystąpienie słowa lub frazy w korpusie, które pasuje do zapytania użytkownika.

14.3.2. Wyszukiwanie dokładnych dopasowań fraz

Aby wyszukać frazę dokładnie pasującą do zapytania, należy ją wpisać w pole zapytania, zaznaczyć opcję *Zachowaj szyk* oraz określić maksymalny *odstęp*⁵ między wyrazami wartością 0 (rys. 14.3).

Rysunek 14.3. Przykład dokładnego dopasowania frazy *dobra wola*

The screenshot shows a search engine interface with the search term "dobra wola" entered in the search bar. Below the search bar, the "Zachowaj szyk" option is checked, and the "Maks. odstęp" is set to 0. The search results are displayed in a table with 10 entries, each showing a snippet of text and the search term "dobra wola" highlighted in blue. The results are sorted by relevance, with the top result being "ofiarować... Bo jedna jest tylko człowiecza zasługa: dobra wola".

Rank	Snippet	Search Term	Additional Info
1.	ofiarować... Bo jedna jest tylko człowiecza zasługa:	dobra wola	. I jedna jest ty Błogostawiona wina
2.	"Anim ci jo nie twój, aniś ty nie moja, jak mnie przenocujesz,	dobra wola	twoja". Między Panem a Pleba...
3.	jest mniej, lecz nie oznacza to, że terror zastąpiła obopólna	dobra wola	i współpraca. F Finał klasycznej Eur...
4.	awanturnictwem. Nie chcemy awanturnictwa. Ale entuzjazm i	dobra wola	są jedynie wek Mój wybór : rzeczy m...
5.		Dobra wola	w każdej, ale t Dopóki milczy Ukrain...
6.	poranienia, załamania, nawet niewierności, jeśli tylko istnieje	dobra wola	obu stron, są d Niezwykły pontyfikat...
7.	wszystkie te nasze obrony, które - gdyby nie stojąca za nimi	dobra wola	- byłyby praw Po szkodzie? Przed s...
8.	osiąga pewność, że nie dobroć zewnętrzna, lecz jego własna "	dobra wola	" jest zasadą w W krainie schorowane...
9.	oblężenia i spoczęły na krajobrazie za oknem. - Pańska	dobra wola	jest w tym prze Mojra
10.	uzyska pan może wgląd we własną istotę i wówczas pańska	dobra wola	będzie musiał Mojra

14.3.3. Ortograficzne symbole wieloznaczne

Składnia obsługuje kilka rodzajów symboli wieloznacznych. Dwa podstawowe symbole wieloznaczne, tj. * (0 lub więcej dowolnych znaków) oraz ? (jeden dowolny znak) umożliwiają ortograficzne rozszerzenie terminu zapytania. Na przykład zapytanie *tymian**⁶ zwraca dopasowania *tymianek*, *tymiankowy*, ale też *Tymiankach*. Z kolei zapytanie *osobliw?* zwraca dopasowania wyrazów *osobliwy*, *osobliwa*, *osobliwą* itd.

14.3.4. Wyszukiwanie fleksyjne słownikowe

W językach bogatych fleksyjnie ortograficznie rozszerzone zapytanie może zwracać mało dokładne konkordancje, w których warianty fleksyjne są przemieszane z derywatami należącymi do innej kategorii części mowy.

⁵ Zob. <http://nkjp.uni.lodz.pl/?q=yhvcekc>.

⁶ Zob. <http://nkjp.uni.lodz.pl/?q=62t4z2f>.

Dlatego w wyszukiwarkach tworzonych dla korpusów polszczyzny bardzo przydatna jest możliwość wyszukiwania fleksyjnego. Opisywana tu wyszukiwarka obsługuje prosty, ale bardzo przydatny rodzaj wyszukiwania fleksyjnego, tzn. wyszukiwanie fleksyjne słownikowe (na podstawie słownika Morfologik, <http://morfologik.blogspot.com/>). Aby automatycznie rozszerzyć zapytanie o warianty fleksyjne zadanej formy podstawowej (np. rzeczownika w mianowniku, rodzaju męskim w liczbie pojedynczej), należy na końcu takiej formy dodać symbol podwójnej gwiazdki (**). Na przykład zapytanie `tymianek**` może zwrócić zbiór dopasowań ukazany na rys. 14.4.

Rysunek 14.4. Przykład wyszukiwania fleksyjnego słownikowego dla zapytania `tymianek**`

The screenshot shows a search interface with the following elements:

- Search input field containing `tymianek**`.
- Options: Maks. odstęp: 0, Zachowaj szyk: , Wyniki: 100.
- Buttons: Czas, Profil, Excel, URL.
- Section: Zaawansowane.
- Buttons: SZUKAJ, << Poprzednie, Nastepne >>, Pomoc.
- Summary: Przeszukiwany zbiór zawiera 1,225,343,686 słów. Znaleziono 760 akapitów pasujących do zapytania w 0.015s.
- Results list (lines 1, 64, 65, 66, 67, 68):
 - 1. mężczyzny zamurza; Jest taki mróz. Nie jesteś dla mnie **tymianek** ni róża, Ani też "cz
 - 64. mały kotlecik przykryty ananase i nie wiedzieć czemu posypany **tymiankiem**, do tego bardzo py
 - 65. Amatorzy baraniny łatwo się zgodzą, że tutejsza pieczeń z **tymiankiem** w sosie kminkowy
 - 66. Henri Matisse pachnie Matką, wiosną i **tymiankiem**. I on uratuje miast
 - 67. pod jego nogami szeleściły zeschnięte badyle mięty, ostów i **tymianków**. Nad głową wisiał
 - 68. się w pachnące zarośla eukaliptusów, mimozy i krzaczastego **tymianku**. Przez mały moste

14.3.5. Wyszukiwanie wariantów

Składnia wyszukiwarki umożliwia również formułowanie zapytań zawierających warianty morfologiczne, zbiory synonimów lub nawet antonimy określone przez autora zapytania. Użycie symbolu `|` spowoduje, że dopasowane zostaną wystąpienia dowolnego z wyrazów w danej grupie wariantów. Na przykład zapytanie `tymianek**|bazyliia**|czosnek**`⁷ zwróci wystąpienia dowolnego z tych trzech wyrazów. W tym wypadku dopasowane zostaną również ich odmiany, ze względu na użyty symbol podwójnej gwiazdki (rys. 14.5).

⁷ Zob. <http://nkjp.uni.lodz.pl/?q=p9knmh>.

Rysunek 14.5. Przykład wyszukiwania wariantów morfologicznych dla zapytania tymianek**|bazyli**|czosnek*

Przeszukiwany zbiór zawiera 1,225,343,686 słów. Znaleziono 6,335 akapitów pasujących do zapytania w 0,022s. Bieżąca strona zawiera 100 wyników.

1.	fasoli, co wymaga chwili artystycznego skupienia. Oliwa, czosnek,	bazyli	, konieczne cząber... pomidory
2.	pociągnął swym wrażliwym, długim na dwa cale nosem. - Piołun,	bazyli	, szalwia, anyżek... Cynamon?
3.	hostii? Urozmaicić wzorem kuchni regionalnych, trochę z papryką, z	bazyli	, albo wegetariańską? Przycho
10.	Marek wciągnął nosem woń przypraw. Lubił kuchnię bogatą w	bazyli	, czosnek i ziarna jałowca. Lub
11.	Ligoniowa dosypała do garnka z zupą garść suszonej	bazyli	.
12.	- Uwważaj! - ostrzega ją żartobliwie - podobno od wążania	bazyli	lęgną się w głowie skorpiony!
13.	z jakim patrzą sobie w oczy. Morze jest niebieskie, kwiaty	bazyli	wazonie ciemnofioletowe, a
14.	zapachu, okropnie mnie męczył, a wszystkie kobiety masowo kupowały	czosnek	w sklepiek więziennym. Jeśli s
15.	można było raz na tydzień zamówić obwarzanki, machorkę i	czosnek	. Czosnek zabijał zapach i sma
96.	szyjach pozawieszali. Niektórzy, osobiwie niewiasty, całe główki	czosnku	pozatykali sobie, gdzie jeno m
97.	mężczyzny zmarmurza; Jest taki mróz. Nie jesteś dla mnie	tymianek	ni róża, Ani też "czuła pod mie
98.	o zawartości. Kwiat lipy, skrzyp polny, macierzanka,	tymianek	, mniszek lekarski - dobrze zna
99.	gromadzą miód, a ten świeży miód — czytamy — "pachnie	tymiankiem	": "redolentque thymo fragrant
100.	pod jego nogami szeleściły zeszcłe badyle mięt, ostów i	tymianków	. Nad głową wisiało gęsto ugw

14.3.6. Rozszerzenie ortograficzne na początku wyrazu

Wyszukiwarka obsługuje również zapytania z „gwiazdką” na początku wyrazu. Na przykład zapytanie *filetow* zwróci wystąpienia wyrazów *sfiletować*, *odfiletować* oraz *wyfiletować*. Z kolei zapytanie *essa** zwróci wszystkie odmiany występujących w słowniku morfologicznym wyrazów zakończonych przyrostkiem *-essa*, czyli na przykład *stewardessa*, *poetessa*, *hostessa*.

14.3.7. Dopasowywanie elastycznych związków frazeologicznych poprzez wyszukiwanie kontekstowe

Składnia wyszukiwarki umożliwia szczególnie wygodne wyszukiwanie wielowrazowych związków frazeologicznych, które często cechują się luźnym szykiem wyrazów. Aby wyszukać kolokacje rzeczowników *łza* oraz/lub *łezka* z rzeczownikami *oko* w żądanym kontekście we wszystkich odmianach tych wyrazów, należy sformułować zapytanie *łza**|łezka**__oko*** (grupy wariantów są tu rozdzielone potrójnym podkreślnikiem). Maksymalny odstęp między terminami zapytania możemy dla przykładu określić wartością 2, przy czym zaznaczenie

opcji *Zachowaj szyk* ograniczy liczbę dopasowań do kontekstów, w których wyrazy występują w kolejności ich podania w zapytaniu.

Podobne, choć nieco bardziej uściślone zapytanie `łza**|łezka**__oko**_ -__kręcić**| zakręcić**`⁸ może zwrócić zbiór wyników ukazany na rys. 14.6.

Rysunek 14.6. Przykład dopasowywania elastycznych związków frazeologicznych poprzez wyszukiwanie kontekstowe

65.	w Warszawie, również w pobliżu siedziby rządu.	Łza się w oku kręci	. Gazeta kosztowa
66.	czy to prawda. Kto wiem, może tak... Jeśli tak, to tylko	łza się w oku kręci	na wspomnienie d
67.	ci ludzie którzy dostają nowy domek się cieszą to aż	łza się w oku kręci	i cieszę się razem
82.	na oknie. Od zaduchu dojrzwania i skwaru	łzy kręciły się w oczach	, rozum ciemniał
83.	I powiem panu jeszcze: mnie, staremu, nawet	łzy się w oczach zakręciły	, bo mi się jakoś,
84.	o Dobrym Pasterzu i kwitnącej winnicy, że ojcu	łzy się w oczach zakręciły	, a i Szczęsnego v
85.	że ze mnie mięczak ale autentycznie aż mi się	łzy w oczach zakręciły	podczas tej rozm
86.	Alicji	łzy zakręciły się w oczach	. Ileż to razy usi
87.	Krysi	łzy zakręciły się w oczach	. Żawołała z wyrz
88.	Owszem, również tobie na ułamek sekundy	łzy zakręciły się w oczach	, ale to tylko dlat
89.	W	oczach jej zakręciły się łzy	. Wysłała na balk
90.	się mu podchmielone stoliki. Poczuł jak w	oczach kręca mu się łzy	, lecz nim oślept
91.	spojrzała na niego nienawistnie, usta jej drżały, w	oczach kręciły się łzy	. Schwyciła torebi

Warto zwrócić uwagę na to, że w niektórych wyszukiwarkach korpusowych dopasowanie tak elastycznego związku frazeologicznego wymagałoby sformułowania co najmniej kilku osobnych zapytań dla poszczególnych wariantów.

14.4. Sortowanie

Opcje sortowania oraz grupowania wyników dostępne są w zaawansowanym formularzu wyszukiwania. Zbiory wyników można sortować dwustopniowo (np. najpierw według źródła, a następnie daty) według następujących kryteriów:

1. Dopasowanie (środek). Sortowanie według dopasowania ułatwia analizę konkordancji wariantów ortograficznych i morfologicznych. Na przykład posortowanie wyników zapytania `ręka**` według dopasowania podzieli

⁸ Zob. <http://nkjp.uni.lodz.pl/?q=p33wg5>.

- konkordancje na podzbiory zawierające wystąpienia różnych form rzeczownika *ręka*.
2. Lewy lub prawy kontekst. Sortowanie według kontekstu umożliwia prostą analizę najczęstszych kolokacji pozycyjnych występujących w zbiorze wyników.
 3. Źródło. Źródłem w wypadku tekstów gazetowych jest tytuł gazety (ale nie pojedynczego artykułu), a w wypadku książek ich tytuł.
 4. Data publikacji.
 5. Kanał, np. prasa, książka, Internet, nagrania języka mówionego.

Warto podkreślić, że sortowane są tylko zbiory wyników (maks. 10 000 na raz), a nie wszystkie wystąpienia w korpusie.

14.5. Grupowanie

Pewnych problemów przy analizowaniu wyników konkordancji z dużych korpusów nastroczają powtórzenia wystąpień częstych wyrazów w tych samych gazetach, książkach lub też w tekstach z tego samego okresu. Często użytkownika korpusu interesują przykłady użycia danego wyrazu lub frazy w różnych gazetach, tekstach, latach, podczas gdy zbiory nieogrupowanych wyników mogą zawierać nadmiar przykładów z jednego źródła.

Opcja grupowania wyników umożliwia określenie maksymalnej liczby wyników z danego roku, źródła lub tekstu. Widać to na ukazanym na rys. 14.7 przykładzie zapytania, które z danego źródła zwraca maksymalnie trzy konkordancje na jednym ekranie wyników. Po wybraniu kryterium grupowania i określeniu maksymalnej liczby wyników, wyświetlone zostają co najwyżej trzy wystąpienia dopasowania w danej gazecie lub książce.

14.6. Metadane

W zaawansowanym formularzu zapytań możliwe jest także zawężenie wyszukiwania do wystąpień dopasowań w tekstach o zadanym typie funkcjonalnym, tytule lub też dacie publikacji. Domyślnie w polu metadanych musi wystąpić jedno lub więcej z podanych słów kluczowych, ale poprzedzając słowo kluczowe operatorem AND, możemy wymusić jego wystąpienie⁹. W polach metadanych można także stosować rozszerzenie ortograficzne oraz dowolnie zagnieżdzać

⁹ Zob. <http://nkjp.uni.lodz.pl/?q=4c3axmp>.

Rysunek 14.7. Przykład grupowania konkordancji

ręka**
 Maks. odstęp: 0 Zachowaj szyk: Wyniki: 100 Czas Profil Excel URL
 Ukryj opcje
 Sortowanie: 1: źródło 2: środek
 Grupowanie: źródło 3

1,225,343,686 słów. Znaleziono 365,631 akapitów pasujących do zapytania w 1,588s.

Wzrost	Wzrost	Wzrost	Wzrost	Wzrost	Wzrost
zapropony przez prezydenta Reagana, żeby otrzymać z jego	rąk	odznaczenie za z	Abecadło Miłosza	+	
Nie mam wiadomości z pierwszej	ręki	, coś słyszałem i	Abecadło Miłosza	+	
a: nie będę ich tutaj przywoływać, choć mógłbym. Biorę do	ręki	listę członków K	Abecadło Miłosza	+	
mu nie opowiadał, że jest ładna. Mama mówiła, że ma dwie	ręce	, nogi do ziemi, i	Anna Dymna - ona to ...	+	
adała o swych nie spełnionych namiętnościach. Przesuwając	rękami	po ciele od nóg c	Anna Dymna - ona to ...	+	
Obyczajowy obrazek, powstający pod	ręką	nestorki polskie	Anna Dymna - ona to ...	+	
zgasila wzrok, rozluźniła mięśnie, spowodowała drżenie	rąk	i odebrała sen. Z	Błogosławiona wina	+	
uszczone, że Ona Matką rodzaju ludzkiego i tylko przez Jej	ręce	Bóg zmieni grze	Błogosławiona wina	+	
których muszą brać udział jako gospodarz. Gdybym nie miał	ręki	czy nogi, mógłby	Błogosławiona wina	+	
i płakać, kiedy tyk śpiewano. Róża zwiesiła głowę, opuściła	ręce	, melodia płynęła	Cudzoziemka	+	
sobą a Michałem Róża postawiła skrzypce. Trzymali się za	ręce	, nic nie mówili.	Cudzoziemka	+	
alenie wzruszający, z głową zadartą w stronę okna. Śpiewał,	ręką	dawał znaki... R	Cudzoziemka	+	
s, że muszą iść swoją drogą, a on - że wymknęłam mu się z	rąk	.	Gwiazdy mają czerw...	+	
ej z frędzlami, siedzący po turecku, z mikrofonem w prawej	ręce	. Rysunek ten, w	Gwiazdy mają czerw...	+	
zo szedł (był już bardzo chory), stanął w świetle i wyciągnął	rękę	, żeby mnie pogł	Gwiazdy mają czerw...	+	
Niewielka jest nadzieja, że póki książka trafi do	rąk	Czytelnika, w Ci	Hongkong dla Chin?	+	
skich, którzy handel ten do niedawna utrzymywali w swych	rękach	. Tak więc istniał	Hongkong dla Chin?	+	
lowaniu się w Pekinie Mandżurowie rozpoczęli rządy silnej	ręki	. Stworzyli uprzy	Hongkong dla Chin?	+	

warunki wystąpienia terminów. Na przykład wpisanie w polu *Tytuł źródła* warunku gazeta AND (Lubuska OR Wrocławska) ograniczy wyszukiwanie do tekstów z „Gazety Lubuskiej” oraz „Gazety Wrocławskiej”.

14.7. Wyrazy kontekstowe

Pewne możliwości ujednoznaczniania wyników zapytania daje opcja określania wyrazów kontekstowych, które mogą lub nie powinny wystąpić w tym samym akapicie, którym znaleziono dopasowanie zapytania. Przypuśćmy, że szukamy wystąpień wyrazu *połączenie* w sensie *połączenie telefoniczne* i że chcemy automatycznie odsiać wszystkie konkordancje, które zawierają wyraz *kolejowy* albo frazę z *Internetem*. W tym celu wystarczy wpisać w polu *Wymagane wyrazy kontekstowe* zapytanie *zamiejscow* OR telefoniczn**, a w polu *Niedopuszczalne wyrazy kontekstowe* zapytanie: *"z Internetem" OR kolejow**. W zwróconych wynikach

powinny się w ten sposób znaleźć głównie wystąpienia rzeczownika *połączenie* w znaczeniu *połączenie telefoniczne*.

14.8. Analiza rejestru

Teksty NKJP opatrzone są informacją o typie funkcjonalnym, dzięki czemu możliwe jest sprawdzenie frekwencji występowania danego wyrazu lub frazy w różnych rejestrach języka. Aby wygenerować wykres słupkowy obrazujący frekwencję danego wyrazu lub frazy, wystarczy kliknąć przycisk *Profil* po wpisaniu zapytania. Na przykład po wpisaniu zapytania *zważywszy na* i kliknięciu przycisku *Profil* (rys. 14.8) wygenerowany zostaje wykres słupkowy podobny do wykresu ukazanego na rys. 14.9, z którego wynika, że fraza *zważywszy na...* pojawia się najczęściej w danych „quasi-mówionych”, na przykład w sprawozdaniach stenograficznych Sejmu RP.

Rysunek 14.8. Generowanie profilu dla zapytania *zważywszy na*

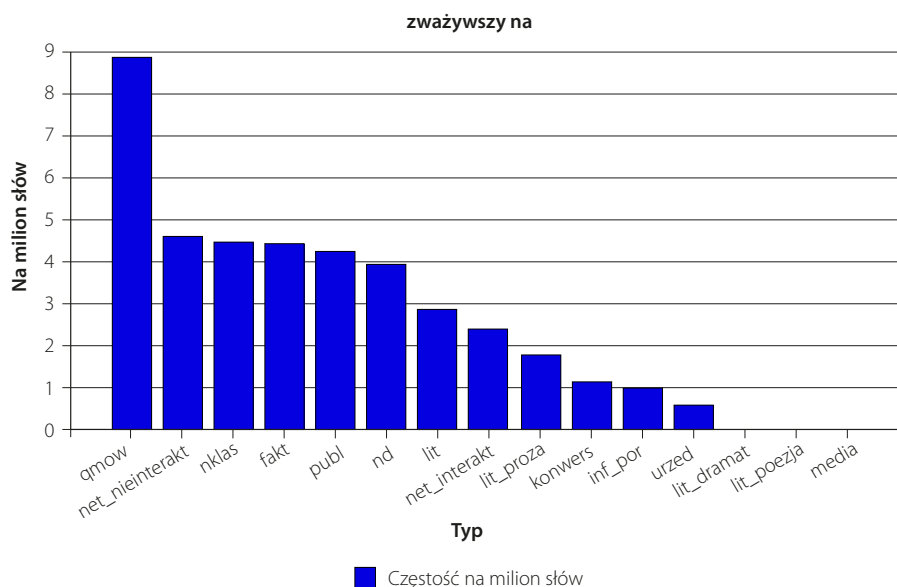
The screenshot shows a search interface with the following elements:

- Search input field containing the text "zważywszy na".
- Buttons for search options: "Maks. odstęp: 0", "Zachowaj szyk: Wyniki: 100", "Czas", "Profil" (highlighted with a red box), "Excel", and "URL".
- A link labeled "Zaawansowane".
- A large "SZUKAJ" button at the bottom right.

Poniżej wykresu częstości w różnych typach tekstów generowany jest wykres słupkowy częstości danego wyrazu lub frazy w „kanałach” publikacji uwzględnionych w taksonomii NKJP (rys. 14.10). Objaśnienia skrótów typów funkcjonalnych używanych w NKJP zawiera tab. 14.1.

14.9. Szeregi czasowe

Zasoby NKJP są bardzo zróżnicowane nie tylko ze względu na gatunek lub typ funkcjonalny tekstów, ale również z uwagi na czas ich powstania. Chociaż NKJP nie jest w zamierzeniu korpusem diachronicznym, w którym różne okresy czasu są równomiernie reprezentowane, to jednak dostępność informacji o dacie powstania lub pierwszej publikacji tekstu stwarza możliwości analizy frekwencji form językowych w zależności od czasu ich użycia. Analiza taka ukazuje, iż niektóre słowa, frazy, idiomy, nazwy własne i zwroty zyskują znacznie na popularności w krótkim czasie, odzwierciedlając tym samym nośność danego tematu w dyskursie publicznym.

Rysunek 14.9. Profil występowania dla zapytania *zważywszy na*

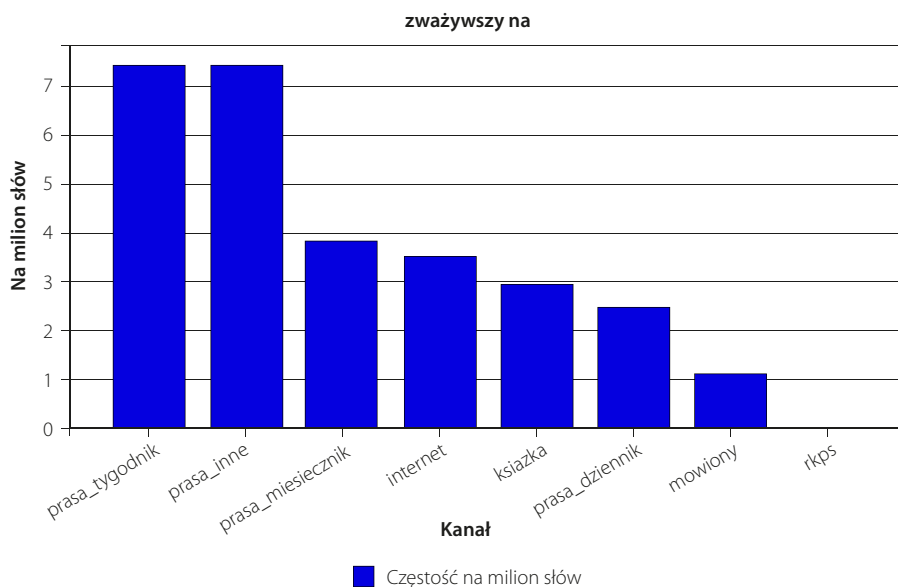
Wyszukiwarka PELCRA NKJP umożliwia wydobycie tego typu informacji o profilu diachronicznym słowa, lub frazy w bardzo prosty sposób. Po wpisaniu dowolnego zapytania w składni wyszukiwarki, należy kliknąć przycisk *Czas*. Po chwili poniżej formularza wyszukiwania powinien się pojawić wykres szeregu czasowego wraz z tabelą, na podstawie której został wygenerowany. Na przykład aby sprawdzić popularność słów *moherowy* lub *moher* we wszystkich odmianach w ostatnich 20 latach, należy wpisać zapytanie *moher**|moherowy***, a następnie kliknąć przycisk *Czas* w formularzu zapytania (rys. 14.11).

Jak widać na wygenerowanym w ten sposób diagramie, popularność tych wyrazów wyraźnie wzrosła w latach 2005/2006 (rys. 14.12).

Odpowiednie zapytanie¹⁰ o wystąpienia tych wyrazów po 2005 roku ukazuje przyczynę tego wzrostu frekwencji. *Moher* i *moherowy* beret nabrały w tym czasie metonimicznego znaczenia i zaczęły funkcjonować jako pejoratywne określenie pewnej grupy społecznej.

¹⁰ Zob. <http://nkjp.uni.lodz.pl/?q=yzsxvo5>.

Rysunek 14.10. Wykres słupkowy częstości danego wyrazu lub frazy w „kanałach”



Rysunek 14.11. Zapytanie moher**|moherowy**

moher**|moherowy**

Maks. odstęp: 0 Zachowaj szyk: Wyniki: 100 Czas Profil Excel URL

Zaawansowane

SZUKAJ

14.10. Pobieranie wyników w postaci arkuszy kalkulacyjnych

Widoczne na stronie wyniki wyszukiwania można pobrać z dodatkowymi metadanyami w postaci arkusza kalkulacyjnego, po kliknięciu przycisku *Excel*. Dzięki temu, użytkownik może dla własnych potrzeb sortować i edytować wyniki wyszukiwania. Arkusze z wynikami mają rozszerzenie *.xml* i należy je otwierać bezpośrednio z programu Microsoft Excel, Open Office lub Libre Office. Zeszyt wyników zawiera dwa arkusze. W arkuszu *Wyniki* można znaleźć konkordancje z podstawowymi metadanyami (rys. 14.13).

Tabela 14.1. Typy funkcjonalne tekstów w taksonomii NKJP

Skrót	Objaśnienie
publ	publicystyczne
net-interakt	internetowe interaktywne (np. fora, blogi z komentarzami, listy dyskusyjne)
net-nieinterakt	internetowe nieinteraktywne (np. strony domowe, blogi bez komentarzy)
nd	naukowo-dydaktyczne
qmow	quasi-mówione
fakt	literatura faktu
urzed	urzędowe
lit	literatura
inf-por	informacyjno-poradnikowe
nklas	inne
lit-poezja	poezja
media	mówione medialne
lit-proza	proza
konwers	mówione konwersacyjne
lit-dramat	dramat

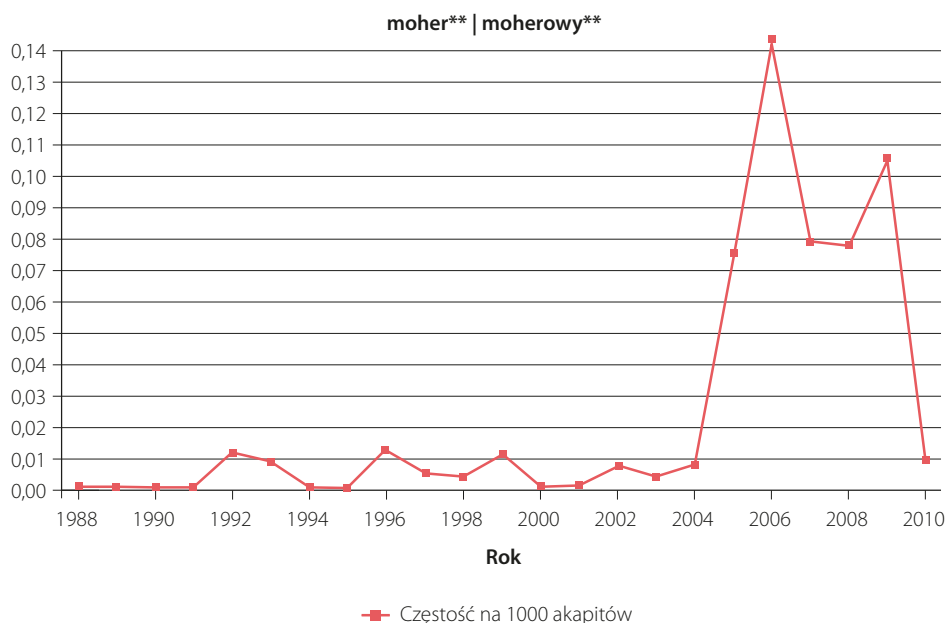
Warto zauważyć, że kolumna *left_word* zawiera słowo występujące bezpośrednio po lewej stronie dopasowania, dzięki czemu wyniki można sortować według lewego kontekstu. W arkuszu *Podsumowanie* znajdują się informacje o zapytaniu i zbiorze wyników.

14.11. Wyszukiwanie kolokacji

Korpusy językowe zawierają cenne informacje o łączliwości słów. Czasem typowe kolokacje danego wyrazu można wydobyć przez zwykłe posortowanie konkordancji po lewej lub prawej stronie. Badanie kolokacji przez sortowanie konkordancji może jednak okazać się kłopotliwe w wypadku często występujących słów. Na przykład różne odmiany rzeczownika *niebo* występują kilkanaście tysięcy razy w zrównoważonym podkorpusie NKJP. Ręczne przejście wszystkich jego wystąpień w celu ustalenia najczęstszych kolokacji przymiotnikowych tworzonych z tym rzeczownikiem byłoby co najmniej niepraktyczne. Kolokator¹¹ to moduł automatycznej ekstrakcji kolokacji zaimplementowany w wyszukiwarce PELCRA NKJP, który znacznie ułatwia to zadanie. Narzędzie to jest dostępne w menu na górze głównej strony wyszukiwarki.

¹¹ Zob. <http://www.nkjp.uni.lodz.pl/collocations.jsp>.

Rysunek 14.12. Profil diachroniczny dla zapytania moher** | moherowy**



14.11.1. Ekstrakcja kolokacji pojedynczych wyrazów

Aby wyszukać lewostronne kolokacje przymiotnikowe rzeczownika *niebo* w różnych odmianach, należy najpierw sformułować odpowiednie zapytanie o ośrodek kolokacji, którym w tym wypadku jest wyraz *niebo*. W tym celu, używając opisanej powyżej składni, wpisujemy zapytanie `niebo**` do pola tekstowego ośrodka kolokacji, tak jak to ukazano na ilustracji poniżej. Dwie gwiazdki na końcu wyrazu oznaczają, że chodzi nam o wszystkie odmiany tego rzeczownika (rys. 14.14).

Kolejnym krokiem jest określenie kryteriów kolokacji. Ponieważ chcemy wyłuskać z korpusu kolokacje przymiotnikowe, z listy *Części mowy* wybieramy opcję *Przym./Imiesł.*, która uwzględnia przymiotniki i imiesłowy przymiotnikowe. Opcje *Kontekst z lewej* oraz *Kontekst z prawej* określają liczbę sąsiadujących z zadanym ośrodkiem kolokacji wyrazów, które mają być rozpatrywane jako część potencjalnych kolokacji.

Ze względu na złożoność obliczeniową ekstrakcji kolokacji, wyszukiwarka PELCRA chwilowo może jednorazowo w ciągu kilku sekund przeanalizować do 50 000 kontekstów wystąpień danego ośrodka kolokacji. Kolokacje wyrazów występujących w korpusie częściej niż 50 000 razy można wydobyć stopniowo, klikając przycisk *Następne*.

Rysunek 14.13. Wyniki w formacie arkusza kalkulacyjnego Excel (XML)

	A	B	C	D	E	F	G
1	#	Left	Match	Right	pubDate	channel	title_mono
2		1 i starał się wyjść z	błędnego koła	prawd względnych i "	1996	#kanal_ksiazk	Legendy nowo
3		2: j jest analiza tego	błędnego koła	historycznego. Czerpa	1995	#kanal_ksiazk	Między Panem
4		3 ten sposób obwód	błędnego koła	zostaje zamknięty.	1997	#kanal_ksiazk	Final klasyczn
5		4 wrócił do początku	błędnego koła	.	1957	#kanal_ksiazk	Kolumbowie –
6		5 raz doświadczałem	błędnego koła	na skutek okazywane	2005	#kanal_ksiazk	Harar
7		6 nii jest przerwanie	błędnego koła	pesymizmu oraz ozyw	2002	#kanal_ksiazk	Choroba dyplo
8		7: zartykułować bez	błędnego koła	, ponieważ percepcja j	1974	#kanal_ksiazk	Obecność mitu
9		8: indukcji; krytyka	błędnego koła	zawartego w próbach	1974	#kanal_ksiazk	Obecność mitu
10		9 reguły indukcji bez	błędnego koła	. Ale argument taki nie	1974	#kanal_ksiazk	Obecność mitu
11		10 reguły, nie uniknie	błędnego koła	naturalizmu. Nadawar	1974	#kanal_ksiazk	Obecność mitu
12		11 zeczności (lub bez	błędnego koła) rozumieć mit w jego	1974	#kanal_ksiazk	Obecność mitu
13		12 wiodują powstanie	błędnego koła	, z którego wyjść może	1996	#kanal_ksiazk	Podręcznik odi
14		13 tania mechanizmu	błędnego koła	. Używa się narkotyku	1974	#kanal_ksiazk	Melancholia

Rysunek 14.14. Zapytanie o ośrodek kolokacji

Kryteria ośrodka kolokacji:

niebo**

Maksymalny odstęp: Zachowaj szyk:

Kryteria kolokatu:

Część mowy: Kontekst z lewej: Kontekst z prawej:

Wielkość próbki: Min. współwystąpienia:

[Pomoc](#)

Po kliknięciu przycisku *Szukaj* należy odczekać kilkanaście sekund. Poniżej formularza zapytania powinna się ukazać tabela wyników, co ilustruje przedstawiony poniżej zrzut ekranu¹² (rys. 14.15).

Na górze tabeli wyników podana jest kolejno ogólna liczba wystąpień ośrodka kolokacji w korpusie, liczba przeanalizowanych kontekstów, oraz liczba potencjalnych kolokacji. Pierwsza kolumna tabeli wyników podaje liczbę porządkową kolokacji. W drugiej kolumnie wyświetlone są znormalizowane formy podstawowe kolokatów. Trzecia kolumna podaje konkretne kombinacje kolokacyjne dla wszystkich odmian formy podstawowej podanej w poprzedniej kolumnie, oraz

¹² Zob. <http://nkjp.uni.lodz.pl/?q=qg2gpv>.

Rysunek 14.15. Ekstrakcja kolokacji pojedynczych wyrazów – niebo**

#	Kolokacja	Pasujące współwystąpienia	Ogółem	Chi ²
1.	rozgwieździć	rozgwieźdzone__niebo (30), rozgwieźdzonym__niebem (13), rozgwieźdzonemu__niebu (1), niebem__rozgwieźdzonym (1),	61	1,291,956.44
2.	rozgwieźdzony	rozgwieźdzone__niebo (30), rozgwieźdzonym__niebem (13), rozgwieźdzonemu__niebu (1), niebem__rozgwieźdzonym (1),	61	1,291,956.44
3.	bezczmurny	bezczmurne__niebo (23), bezczmurnym__niebie (22), bezczmurne__niebie__bezczmurnym (3), nieba__bezczmurnego (2), niebu__be	72	969,869.4
4.	goły	gołym__niebem (202), gołe__niebo (4), gołego__nieba (2), niebi	209	539,774.91
5.	wygwieździć	wygwieźdzone__niebo (7), wygwieźdzonym__niebem (3), niebo__	14	494,209.52
6.	gwiazdzisty	niebo__gwiazdziste (22), gwiazdziste__niebo (11), gwiazdzistego__nieba__gwiazdzistego (3), niebem__gwiazdzistym (3), gwiazdziste	58	449,726.09
7.	zachmurzyć	zachmurzone__niebo (17), zachmurzonym__niebie (7), niebo__z zachmurzonego__nieba (2), zachmurzonym__niebem (1),	34	317,931.32
8.	pochmurny	pochmurne__niebo (10), pochmurnego__nieba (9), pochmurnym__niebie__pochmurnym (1), nieba__pochmurnego (1),	29	201,454.9
9.	jasne	jasnego__nieba (106), jasnym__niebie (3), nieba__jasnym (1),	110	189,997.74
10.	siódmy	siódmym__niebie (86), siódme__niebo (16), siódmego__nieba (7)	109	105,407.01
11.	 błękitny	błękitne__niebo (27), błękitnego__nieba (18), niebo__błękitne (1), błękitnemu__niebu (1), nieba__błękitnymi (1), niebu__błękitne	65	95,748.27
12.	granatowy	granatowe__niebo (13), granatowego__nieba (10), granatowym__niebie__granatowe (1),	38	77,201.19

liczebności poszczególnych kombinacji. Po kliknięciu liczebności w osobnym oknie wyświetlane są konkordancje danej kombinacji. Pozwala to zweryfikować wyniki grupowania odmian do formy podstawowej. W czwartej kolumnie ukazano ogólną liczebność wszystkich form, która jest sumą form wszystkich kombinacji. Ostatnia kolumna podaje wartość chi kwadrat, która określa istotność statystyczną danej kolokacji. Właśnie według tej wartości sortowane są potencjalne kolokacje. Sortowanie kolokacji według zwykłej liczebności współwystąpień słów obniżałoby czytelność wyników, ze względu na dużą liczbę częstych słów, które tworzą z zadaniem wyrazem związki składniowe, a nie kolokacyjne.

Jak widać, do typowych kolokacji rzeczownika *niebo* można zaliczyć takie frazy jak: *gołe niebo*, *rozgwieźdzone niebo*, *bezczmurne niebo*, *siódme niebo* itd., co chyba pozostaje w zgodzie z intuicją leksykalną użytkowników polszczyzny. Ciekawe są także informacje o preferencjach frazeologicznych wyłaniających się z liczebności niektórych form, np. *rozgwieźdzone*, *wygwieźdzone*, *gwieździste*, *gwiazdziste niebo*.

Warto pamiętać, że wyszukiwarka nie zawsze jest w stanie rozszerzyć zapytanie o formy pokrewne morfologicznie. Na przykład jeżeli dla ośrodka kolokacji zdefiniowanego jako VAT** nie zwrócono żadnych wyników, to warto użyć zwykłego rozszerzenia ortograficznego, stosując zapytanie z jedną, a nie dwiema gwiazdkami, czyli VAT*.

14.11.2. Ekstrakcja złożonych kolokacji

Wyszukiwarka kolokacji umożliwia także badanie wielowyrzowych ośrodków kolokacji. Na przykład, aby wyszukać kolokacje występujące z czasownikiem *dojść* i przyimkiem *do*, można sformułować zapytanie ukazane na rys. 14.16¹³.

Rysunek 14.16. Zapytanie dla kolokacji frazy *dojść** do*

Kontekst kolokacyjny ustawiono w tym przypadku na dwa słowa z prawej strony dopasowania. Jak widać na poniższym zrzucie ekranu, najbardziej istotne statystycznie kolokacje rzeczownikowe zwrócone przez powyższe zapytanie to między innymi *dojść do skutku/wniosku/porozumienia/przekonania* (rys. 14.17).

14.11.3. Jak rozumieć wartość chi kwadrat?

Potencjalne kolokacje są obecnie sortowane według wartości testu statystycznego chi kwadrat, który dość precyzyjnie określa jeden z aspektów łączliwości frazeologicznej.

Podana w ostatniej kolumnie tabeli wyników wartość chi kwadrat wyraża prawdopodobieństwo tego, że częstotliwość współwystępowania ośrodka kolokacji z danym wyrazem w korpusie nie jest przypadkowa. Dokładniej wyrażają to wartości prawdopodobieństwa przypisane do wartości chi kwadrat dla jednego stopnia swobody przedstawione w tab. 14.2.

Tabela 14.2. Typowe dla polszczyzny mówionej kombinacje segmentów wyrazowych

Chi kwadrat	2,706	3,841	5,024	6,635	10,828
Istotność statystyczna	0,90	0,95	0,975	0,99	0,999

¹³ Zob. <http://nkjp.uni.lodz.pl/?q=oh3c5n>.

Rysunek 14.17. Ekstrakcja złożonych kolokacji – dojsć** do

#	Kolokacja	Pasujące współwystąpienia	Ogółem	Chi ²
1.	skutek	dojdzie do ___ skutku (144), doszła do ___ skutku (105), doszedł do ___ (25), dojdą do ___ skutku (14), doszłaby do ___ skutku (11), doszedł do ___ skutku (1),	501	1,256,018.83
2.	wniosek	doszedł do ___ wniosku (400), doszli do ___ wniosku (265), doszedł do ___ wniosku (125), doszłam do ___ wniosku (72), dojdzie do ___ wniosku (25), dojdziemy do ___ wniosku (23), dojdiesz do ___ do ___ wniosku (5), doszłaś do ___ wniosku (5), doszliby do ___ wnio (2), dojsć do ___ wniosków (1), doszedł do ___ wnioski (1), doszła do do ___ wniosków (1), doszli do ___ wniosków (1), dojdź do ___ wnios	1711	1,003,276.97
3.	porozumieć	dojsć do ___ porozumienia (178), doszło do ___ porozumienia (54), do do ___ porozumienia (43), doszły do ___ porozumienia (38), dojdzie do ___ porozumienia (24), doszła do ___ porozumienia (10), dojsć do (2), dojsć do ___ porozumienia (2), doszedłem do ___ porozumienia do ___ porozumienia (1), doszłaby do ___ porozumienia (1),	529	487,055.59
4.	zderzyć	doszło do ___ zderzenia (168), dojdzie do ___ zderzenia (1), dojsć do	170	273,322.09
5.	kolizja	doszło do ___ kolizji (129), dojsć do ___ kolizji (2), dojdzie do ___ kc	133	146,936.19
6.	szarpanina	doszło do ___ szarpaniny (27),	27	67,370.06
7.	rękoczyn	doszło do ___ rękoczynów (12), doszłoby do ___ rękoczynów (6), doj	24	58,876.79
8.	tragedia	doszło do ___ tragedii (69), dojsć do ___ tragedii (32), dojdzie do ___	121	53,072.95
9.	władza	dojsć do ___ władzy (64), doszedł do ___ władzy (60), dojsć do ___ (20), doszli do ___ władzy (19), dojdzie do ___ władzy (19), dojsć do do ___ władzy (7), dojdzie do ___ władzy (6), doszło do ___ władzy doszliby do ___ władzy (1), dojdę do ___ władzy (1), dojsć" do ___	327	41,964.19
10.	rozłam	doszło do ___ rozłamu (22), dojsć do ___ rozłamu (6), dojdzie do ___	32	39,135.17

Jeżeli więc wartość testu chi kwadrat podana w tabeli wyników wynosi 10,828, to z matematycznego punktu widzenia istnieje tylko jedna szansa na tysiąc, że dane dwa wyrazy występują w zaobserwowanych kontekstach zupełnie przypadkowo. Innymi słowy, prawdopodobieństwo tego, że liczba współwystąpień wynika tylko i wyłącznie z ogólnej częstości występowania pojedynczych wyrazów wynosi 0,001.

W obecnej wersji wyszukiwarki wyświetlane są wyniki o liczebności współwystąpień ≥ 5 , oraz o wartości testu chi kwadrat $\geq 3,841$. Oczywiście częstość współwystępowania wyrazów nie jest tylko funkcją ich łączliwości frazeologicznej i dlatego niektórych z współwystąpień wyrazów z wysoką wartością chi kwadrat nie można uznać za związki frazeologiczne.

14.12. Dostęp programistyczny

Wyszukiwarka PELCRA dla NKJP obsługuje także zapytania programistyczne przez protokół HTTP. Najlepiej to ilustruje skrypt napisany w języku Python¹⁴

¹⁴ Pełna wersja skryptu dostępna jest pod adresem: <http://www.nkjp.uni.lodz.pl/getConcord.py.jsp>.

(wydr. 14.1). Skrypt ten wysyła zapytanie do serwera i otrzymuje wyniki konkordancji w prostym formacie XML (wydr. 14.2). Możliwe jest również automatyczne pobieranie wyników konkordancji we wspomnianym powyżej formacie Microsoft Excela (XML)¹⁵.

Wydruk 14.1. Fragment skryptu napisany w języku Python

```
1 #coding=utf-8
2 import urllib
3 import random
4 servlet="http://nkjp.uni.lodz.pl/NKJPSpanSearchXML"
5 #Zapytanie w składni PELCRA NKJP
6 query="pleść** bzdura**"
7 #Maks. odstęp między słowami
8 span=2
9 #Zachowujemy szyk? true|false
10 preserve_order="false"
11 #Od którego wyniku zaczynamy?
12 offset=0
13 #od 1 do 5000 na raz. Wartości > 5000 są przycinane.
14 limit=50
15
16 #Inne parametry użyte w żądaniu HTTP poniżej... Zob.
17 http://www.nkjp.uni.lodz.pl/getConcord.py.jsp
18 params = urllib.urlencode({'query': query, 'offset': offset,
19 'span': span, 'sort': sort, 'second_sort': 'srodek', 'limit':
20 limit, 'groupBy':groupBy, 'groupByLimit':groupByLimit,
21 'preserve_order':preserve_order, 'dummystring':dummystring,
22 'sid':sid, 'm_date_from':m_date_from, 'm_date_to':m_date_to,
23 'm_styles':m_styles,
24 'm_channels':m_channels, 'm_title_mono':m_title_mono,
25 'm_title_mono_NOT':m_title_mono_NOT,
26 'm_paragraphKWs_MUST':m_paragraphKWs_MUST,
27 'm_paragraphKWs_MUST_NOT':m_paragraphKWs_MUST_NOT})
28
29 f = urllib.urlopen(servlet, params)
30 print f.read()
```

Na razie nie stosujemy dodatkowych ograniczeń w automatycznym dostępie HTTP, ale mogą się one pojawić w wypadku nadużyć. Należy pamiętać, iż wszelkie formy komercyjnego wykorzystania wyszukiwarki wymagają uzyskania licencji od NKJP.

¹⁵ Adres serwletu to <http://nkjp.uni.lodz.pl/NKJPSpanSearchExcelXML>.

Wydruk 14.2. Wyniki konkordancji w formacie XML

```

1 <?xml version="1.0" encoding="UTF-8"?>
2 <results type="concordance">
3 <index_size>1225343686</index_size>
4 <!--Words in subcorpus-->
5 <total_hits>808</total_hits>
6 <query_time_s>0.034</query_time_s>
7 <concordance>
8
9 <line><count>1</count>
10 <left><![CDATA[... że wreszcie udało mi się uciec choć na
    chwilę z tej okropnej Aleksandrii, brudnej i cuchnącej,
    gdzie nie można przejść dziesięciu jardów, nie napotkawszy
    tubylca załatwiającego się na skraju drogi. Cóż to za
    ]]></left><match><![CDATA[bzdury
    plecie ]]></match><right><![CDATA[ się o Wschodzie! Czar
    Orientu! Slumsy w Wembley mają mniej więcej tyle samo
    uroku. A osławione piramidy wyglądają po prostu jak hałdy ż
    wiru... No cóż, to obowiązek żony oficera towarzyszyć mu
    wszędzie, ale przysięgam, że jeśli miano by go wysłać do
    Indii, zażądam, by wystąpił z armii. W końcu ja też
    zarabiam – powiedzmy sobie szczerze: moje dochody są
    większe niż jego apanaże... ]]></right>
11 <pubDate>1997</pubDate>
12 <channel>#kanal_ksiazka</channel>
13 <domain>#typ_lit</domain>
14 <title_mono><![CDATA[Znak Anny ]]></title_mono>
15 <title_a><![CDATA[ ]]></title_a>
16 </line>
17
18 </concordance>
19 <!-- itd. -->
20 </results>

```

14.13. Wyszukiwarka dla danych mówionych

W puli danych NKJP znajduje się obecnie ponad 2 miliony słów języka mówionego zarówno *medialnych*, jak też *konwersacyjnych*, które omówiono w oddzielnym rozdziale tego podręcznika. Wszystkie te dane można przeszukiwać w głównej wyszukiwarce NKJP, po wybraniu odpowiednich opcji typu i stylu tekstów.

Niemniej jednak, dzięki osobnej wyszukiwarce dla danych mówionych można wygodniej przeszukiwać, sortować i wyświetlać konwersacyjną część korpusu. Możliwe jest na przykład sortowanie konkordancji według wieku, wykształcenia lub też płci mówiącego¹⁶.

14.14. Dalsze informacje

Aktualne informacje na temat wyszukiwarki podawane są na jej stronie pomocy, pod adresem: <http://nkjp.uni.lodz.pl/help.jsp>. Autor wyszukiwarki prosi o zgłaszanie uwag dotyczących działania wyszukiwarki na adres piotr.pezik@gmail.com.

¹⁶ Zob. <http://www.nkjp.uni.lodz.pl/spoken.jsp>.