

Barokowa polszczyzna w Internecie,

czyli jak powstają

Elektroniczny słownik języka polskiego XVII i XVIII w.

oraz

Elektroniczny korpus tekstów polskich z XVII i XVIII w.

Włodzimierz Gruszczyński

Instytut Języka Polskiego PAN

i SWPS Uniwersytet Humanistycznospołeczny

e-mail: wgruszczyński@swps.edu.pl

wlodekiewa@poczta.onet.pl



Po co taki słownik?

Co to znaczy?

Lecz, że mu usta stękaia i ięczą,
Dla tego plagi nieznośne mię dręczą,
Dla tego mię BOG dotyka i drażni,
Ze znieść cierpliwie trudno iego łażni.
Ná moje bole, prawie iest ząwziętem,
Obie mi nogi twardym ścisnął pętem,
I wszystkie ścieszki bodzcami otoczył,
Bym z nich bez szwanku nigdzie nie
wyskoczył.

ChrośJob 123.



Co znaczy wyraz *bodziec*?

- W słowniku współczesnej polszczyzny, np. USJP PWN:
 1. *fizjol.* «czynnik wywołujący reakcję organizmu»
 - Bodziec chemiczny, mechaniczny, świetlny.
 - Bodziec wewnętrzny, zewnętrzny.
 2. *książk.* «okoliczność, zdarzenie, czynnik zachęcający do działania; podnieta, zachęta, pobudka»
 - Bodźce finansowe, materialne.
 - Dodać, dostarczyć komuś bodźca do pracy.
 - Stać się bodźcem do ożywionej działalności społecznej.
- Zajrzyjmy do e-SXVII
 - http://xvii-wiek.ijp-pan.krakow.pl/pan_klient
 - <http://sxvii.pl/>



znajdź hasła zaczynające się od ▾

szukaj

☐ przeszukaj także hasła w indeksie

☒ a fronte ☐ a tergo

wtyczki wyszukiwania dla przeglądarek Firefox i Microsoft Internet Explorer

A B C Ć D E F G H I J K L Ł M N O Ó P R S Ś T U V W X Y Z Ź Ż

ELEKTRONICZNY SŁOWNIK JĘZYKA POLSKIEGO XVII I XVIII WIEKU



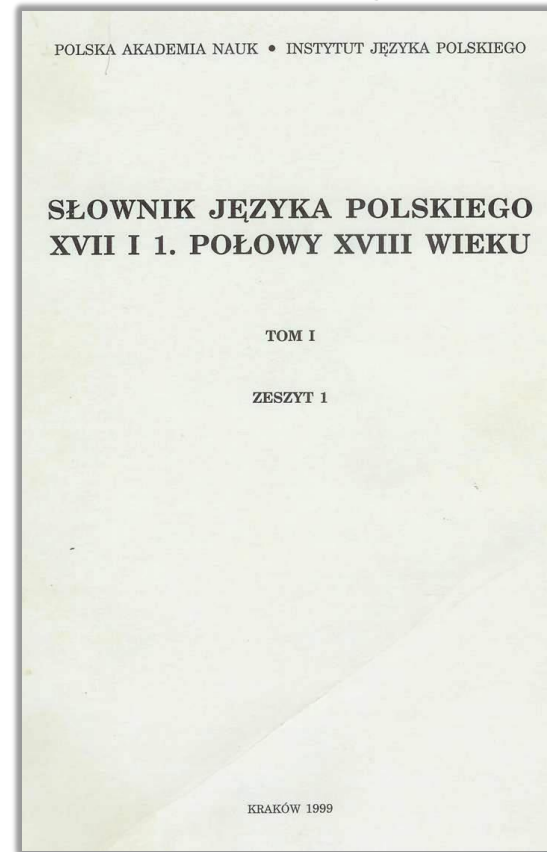
Jak powstawał słownik w XX w.?

- **1954-1993:**
 - opracowanie koncepcji,
 - ustalenie kanonu źródeł,
 - ekscerpcja, w wyniku której powstało prawie 3 mln kart z cytatami.



Koniec XX wieku – początek edycji

- 1996:
 - wydanie zeszytu próbnego S-XVII
- 1999:
 - wydanie 1. zeszytu I tomu S-XVII



Pytanie z roku 2005:

**Czy i jak komputeryzować prace nad
*Słownikiem języka polskiego XVII
i 1. połowy XVIII wieku?***



Rok 2005 – początek rewolucji

- Zawieszenie wydawania wersji drukowanej.
- Opracowanie założeń słownika w wersji elektronicznej.
- Zatrudnienie informatyka, który stworzył cały system informatyczny *S-XVII*:
 - bazę danych,
 - formularza dla autorów haseł
 - program przekształcający dane z bazy na tekst artykułów hasłowych.



Postulaty sprzed 11 lat...

Jak przyspieszyć i ułatwić pracę nad słownikiem?

- Ułatwić dostęp do materiałów źródłowych!
- Rozwiązać problem fiszek (Ale jak?!).
- Przenieść jak najwięcej źródeł na nośnik elektroniczny!
- Stworzyć KORPUS tekstów średniopolskich.
- Stworzyć narzędzia ułatwiające pisanie artykułów hasłowych.



Wątpliwości sprzed 11 lat...

Problem fiszek – podstawowego źródła materiału językowego na obecnym etapie prac

- Istnieje ok. 3,2 mln fiszek
- Większość fiszek pisana jest ręcznie
- Pewna część to powielaczowe kopie fiszek przepisywanych na metalowych matrycach.
- Czy przenosić je na nośnik elektroniczny?



Pesymizm nieuzasadniony...

Skanowanie fiszek jest nieracjonalne

- ➡ czas potrzebny na zeskanowanie jednej fiszki i zapisanie pliku na dysku porównywalny jest z czasem przepisania jej na komputerze;
- ➡ skanowanie fiszek doprowadziłoby do tego, że ten sam cytat zapisany na różnych fiszkach, byłby wprowadzony do pamięci komputera wielokrotnie;
- ➡ tekst na fiszkach nie nadaje się do automatycznego rozpoznawania (OCR)



Marzenia o korpusie

Konieczność stworzenia
korpusu tekstów źródłowych
XVII- i XVIII-wiecznych i
automatyczne wygenerowanie
ich indeksów

- Korpus w pewnym stopniu zastąpi kartotekę, a w znacznym ją uzupełni.
- Korpus przyspieszy prace nad słownikiem, a w pewnym stopniu go zastąpi (!).

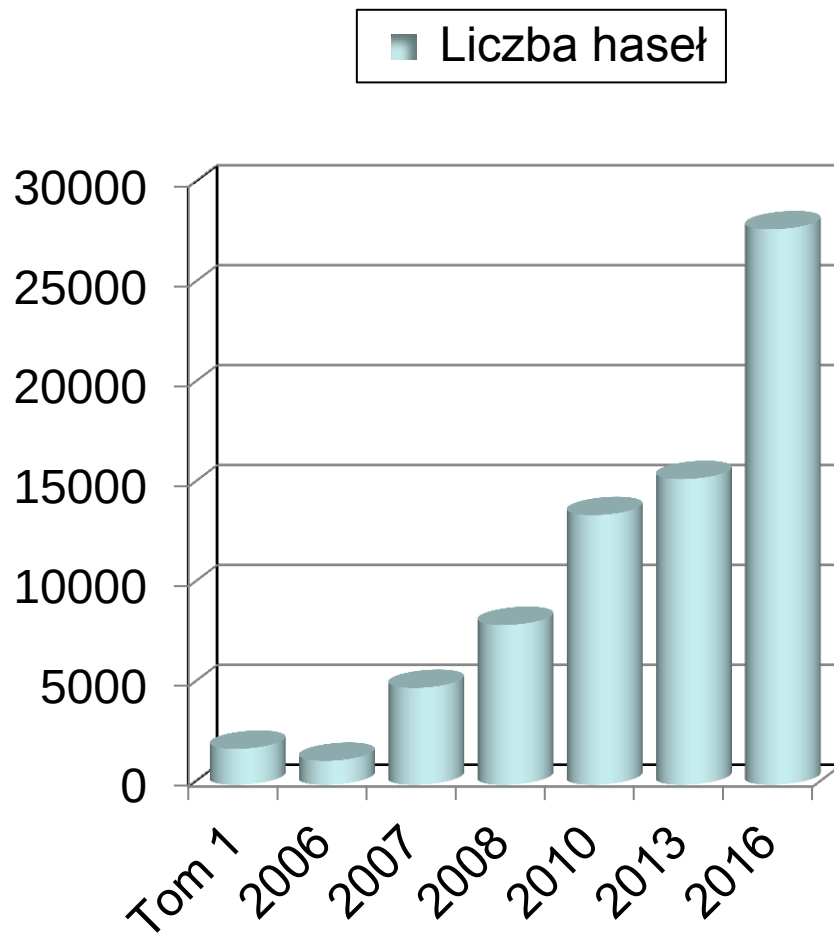


Trudne początki ☹️

- Konieczność zmian wieloletnich przyzwyczajeń członków zespołu.
- Opór najstarszych członków zespołu i części środowiska językoznawczego.
- Czasowe zahamowanie prac.



Opłacało się 😊



Typowy artykuł hasłowy

BACIK rzecz. m

Słowniki: SXVI, Kn, T, L (bez cyt.), SWil, SW, SJP notują

Słownik SStp nie notuje

Formy gramatyczne:

lp M. bacik; D. bacika; B. bacik; Ms. baciku;

lm M. baciki

>>łódka, szalupa<<:

- Łódź/ łódka máła/ wielka/ dłubána/ szyta/ barká/ bat/ **bacik**/ łozá.
Nauicula, nauigiolum [...] Scapha [...] Phasellus [...]. *Kn* 367.
- A gdysmy pod niektory máły wysep przypłynęli/ który zowią Klaudá/
ledwiesmy mogli **bacik** zâtrzymâć. *BG Dz* 27, 16.
- A gdy żeglarze myśliłi z okrętu uciec/ y spuścili **bacik** ná morze/ chcąc
rzekomo od przodku okrętu zârzucâć kotwice/ Rzekł Páweł
Setnikowi y żołnierzom: Jesli ci nie zostaną w okręcie/ wy zâchowani
być nie możecie. Tedy żołnierze obcięli powrozy u **baciká** y dopuścili
mu odpâść. *BG Dz* 27, 30-32.
- Lodz/ łódká málá/ bárrká [!]/ **bácik**. *SzyrDict* 134.
- Przy onym okręcie/ w którym on iácháł/ był ná zâdzi uwiązány bacik
ieden máły ná powrozie. *StarKaz II*, 47.
- Nie wielki ale piekny Ogród y Kanały na których **Baciki** ładne były.
SobTDzien 8.
- **Bacik**. kleines Boot, ein Nachen. nacelle, ésquif. *T* 17.

Autor: [PK](#)



A może dodać ilustracje?

POLSKA AKADEMIA NAUK | INSTYTUT JĘZYKA POLSKIEGO
SŁOWNIK JĘZYKA POLSKIEGO XVII I 1. POŁOWY XVIII WIEKU

o słowniku | lista haseł | zaawansowane wyszukiwanie | kwerendy szerokość: 600 pl

znajdź hasła **zaczynające się od**

☐ przeszukaj także hasła w indeksie ☒ a fronte ☐ a tergo
 wtyczki wyszukiwania dla przeglądarek Firefox i Microsoft Internet Explorer
 A B C Ć D E F G H I J K L Ł M N O Ó P R S Ś T U V W X Y Z Ź Ż

sposób prezentacji hasła ☒ obszerny ☐ zwięzły

poprzednie hasło [BIEDRZECHEC](#) następne hasło [BIEDRZEŃCOWY](#)

ARTYKUŁ HASŁOWY W TRAKCIE OPRACOWANIA
 Data ostatniej modyfikacji: 02.09.2010

BIEDRZENIEC *rzecz. m*
 Warianty fonetyczne: BIEDRZENIEC, BIEDRZYNIEC
 Słowniki notują
 Formy gramatyczne:

	<i>lp</i>
M.	biedrzeniec biedrzyńiec
D.	biedrzyńcu biedrzeńcu biedrzeńca
B.	biedrzeniec
N.	biedrzeńcem

bot. » roślina zielna o właściwościach leczniczych; *Pimpinella* «:

- Biedrzeniec/ Rospikámień/ Rozdział 18. Tragoselinum. Pimpinella. ... własny jest troiaki. Wielki/ Mály albo sámicá. Trzeci namnieyszy kryspowaty/ albo kędzierzawy. *SyrZiel* 62.
- Wodká z Biedrzeńcu. *SyrZiel* 71.
- Syrop z Biedrzeńcá. Bywa też Syrop wyciągniony z biedrzeńcu korzeniá suchego/ tym sposobem/ iáko sie powiedziało o czynieniu Syropu z Cynámonu. *SyrZiel* 73.
- Są ktorzy ten Koziorožnik twierdzą być Biedrzeńcem. Drudzy Rospikámieniem/ który Łácinnicy Saxifragiam Hircinam zowią. Drudzy omylnym Dyptanem. Ale ci wszyscey w swym mniemaniu się mylą. *SyrZiel* 1346.
- Biedrzeniec/ Tragoselinum [...] quasi hircinum apium. [...] Pimpinella. [...] Saxifragiae hircina. [...] Duae eius species maior et minor. *Kn* 26.
- Pimpinella vera, Saxifragia quibusdam. Biebenell. Biedrzeniec/ Rospikámień. *GuldOn* 14.
- Biedrzyńcu uwiercie troche oskrobanego a pieprz[em] potraszawszy smiotaną ugniesc y na bolacy zamb przylozyc. *GrodzMisc* 98.



Tzw. hasła podwójne

- Jeżeli formy tekstowe nie pozwalają na jednoznaczne odtworzenie formy hasłowej, tworzymy hasła z dwiema formami hasłowymi.



Tzw. hasła podwójne

- Jeżeli w materiale są poświadczone dwa paradygmaty wariantywne pod względem rodzaju:

ABRYS rzecz. m, ***ABRYSA** rzecz. ż

Formy: lp M. abrys; D. abrysu; B. abrys // abrysę; N. abrysem; Ms. abrysie; ~ lm M. abrysy; D. abrysów; C. abrysom; B. abrysy; N. abrysami; Ms. abrysach.

1. » plan, projekt, zapis; układ, zarys «
2. » szkic, rysunek «
3. » wizerunek, obraz «



Tzw. hasła podwójne

- Dla rzeczowników mających wariantywną budowę morfologiczną (np. temat fleksyjny poszerzony o przyrostek tematowy lub bez tego przyrostka).

ARYJAN, ARYJANIN *rzecz. m*

Formy: *lp* M. aryjan // aryjanin; D. aryjana; C. aryjanowi; *Ms.* aryjaninie; *W.* aryjaninie // aryjanie; ~ *lm* M. aryjani // aryjanowie // arianie; D. aryjanów; C. aryjanom; B. aryjanów; N. aryjany.



Tzw. hasła podwójne

Dla niektórych rzeczowników zapożyczonych z łaciny, zakończonych tam np. na *-us*, *-um*, *-or*:

AMIJANTUS, *AMIJANT *rzecz. m*

Formy: lp M. amijantus; Ms. amijancie.

» minerał, odmiana azbestu «



Problem wariacji

- Báldekin złoty rozpięty na sznurze Wisiał drogiemi tkány
iedwabiámi. *TasKochGoff* 422.
- Báldakin/ v. Ciennik. *Kn* 13.
- Insignia zaś victoriae przed nogi I. K. M. v Báldáchinu Zołnierstwo
rzuczało. *MerkPol* 271.
- Nad nią baldachim nieoszaczowany Z klejnotów wszytek noszono
ubrany. *BorzNaw* 49.
- Ten Pan [król] [...] ktorego więcej znamy na koniu, niż na Thronie,
pod Namiotami niż Báldákinámi. *LubJMan* 93.
- Cesarze pod baldachynami Stoią, pieknemi świecąc ozdobami.
OblJasGór 151v.
- Sequebantur tedy damy, wprzód stateczne, znaczne senatorów
żony, wszytko godnie wprzód przed ołtarzem kortezyją, potym
przed księstwem czynią, i klękają na ziemi niedaleko beldekina.
BillTDiar 77v.
- Na śródku Dziedzińca przed Pałacem 4. kolumny, Baldakin
bogaty axamitny karmazynowy galonem y frandzlą bogatą
przyozdobiony wspierały [...]. *GazPol* 1735/48, 2 nlb.
- Pani szczęścia [...] Pod Baldachinem Siedzi złotolitym. *DružZbiór* 232.
- Z Królem JM pod beldakiem siedziała księżna matka królewska i
legat papieski z posłem cesarskim Szanffgociem. *ChrapDiar* II, 536.
- Baldachium tysz nagotowali byli Raycy Tabinowe, pozłociste, który
[!] mieli niść nad nim sami, ale dla niepogody dali pokoy. *GollPam* 24-
25.



Warianty w artykułach hasłowych

BALDACHIUM *rzecz. n*

BALDEKIN *rzecz. m*

[warianty fonetyczne:]

**BALDEKIN, BALDACHIN, BALDAKIN,
BALDACHIM, *BALDACHYN, *BELDAKIN,
BELDEKIN, BALDECHIN**

BELDAK *rzecz. m*



Indywidualizmy, okazjonalizmy...

- ...czemuż się tedy nie zowiecie Apostolikámi/ Epistolarzámi/ Apokálipsikámi/ Prophetykámi/ Bibliakámi? Lacno odpowiedzą. Chcą zgoła iáko nadáley rozroźnić się od Kátholikow.

BirkNiedz 131.

- Nuż wąpić się nie może/ iż to iedenże beł u rożnych/ Bał/ Báál/ Beł/ Beel y Bellus. A że od waleczności tak beł názwany/ sáme słowá bellum, to iest walka y baleczność ábo waleczność/ także walił ábo bálił/ bił y bol wydaia. *DembWyw 64.*

- Poczęli byli mężnie chłopislausowie strzelać.

DembPrzew 67.



Frazeologizmy, połączenia terminologiczne

Skąd wiemy, że były „ustabilizowane” lub „częste”, jeśli mamy kilka wystąpień, a czasem tylko jedno?

- Na koniec i Belgrad forteca prawie niedobyta w akord poszedłszy, poddała się, prosząc o pokój Turczyna. *KronZak 200.*
- Dasz znosne kondycje staniemy w Accordzie Bez krwi rozlania; Turkom, tak powiedz y Ordzie. *PotWoj 19.*



Frazeologizmy, połączenia terminologiczne

- To początek dopiero tego coś nawarzył piwa które, że źle uwarzone przykre iest. Iak się ponim dochowasz kwasu tobie Samemu gorzką wgarle Stanie zołcią. *PasPam 196.*



Mogą pomóc dawne słowniki

- GALANTERIE [...] ZARAZIĆ się szpetną chorobą od niewiast, od wszeteczeństwa galanteryki nabyć, na aptekę zarobić. *DanKolaDyk II, 74.*
- Vervex sycophanta. BARANEK wilkiem podszyty niewiniątko ná pozor w samey rzeczy hultay wierutny. *DanKolaDyk II, 57.*
- Podbiiam komu bębenká/ v. Dodáię sércá ná złe. *Kn 734.*



Narzędzia, pomoce...

- Formularz do tworzenia haseł:

Edycja hasła

Słownik XVII wieku

xvii-wiek.jp-pan.krakow.pl/index.php

Formy hasłowe (1) Tylko odsyłacz rzecz.

BODZIEC

Warianty fonetyczne (4)

Słowniki

Formy gramatyczne 9 Uwagi:

Etymologia (0)

Znaczenia (5) + (niepewne)

÷ (0)

ı (0)

Użycie metajęzykowe (0)

Podhasła (0)

Odsyłacze (wypisać po przecinku)

Autor

BR-K

Wersje (1)

Informacje dodatkowe

Edytowana wersja

uwagi redakcyjne:

WG: (3.12.2013) dodałem znaczenie 2., zmieniłem kolejność znaczeń 4 i 5.
Dodałem przykłady użyc, formę fleksyjną (M.lm)

dziennik modyfikacji:

wlodek (04.12.2013 09:12:24)
wlodek (03.12.2013 23:12:33)

wybierz hasło nowe hasło usuń hasło podgląd zapisz i wybierz inne hasło zapisz

sprawdź poświadczenia indeks korpusu korpus źródła poczta wyloguj zgłoś błąd szerokość: 800

poprz. hasło BODARZ nast. hasło BODEJ

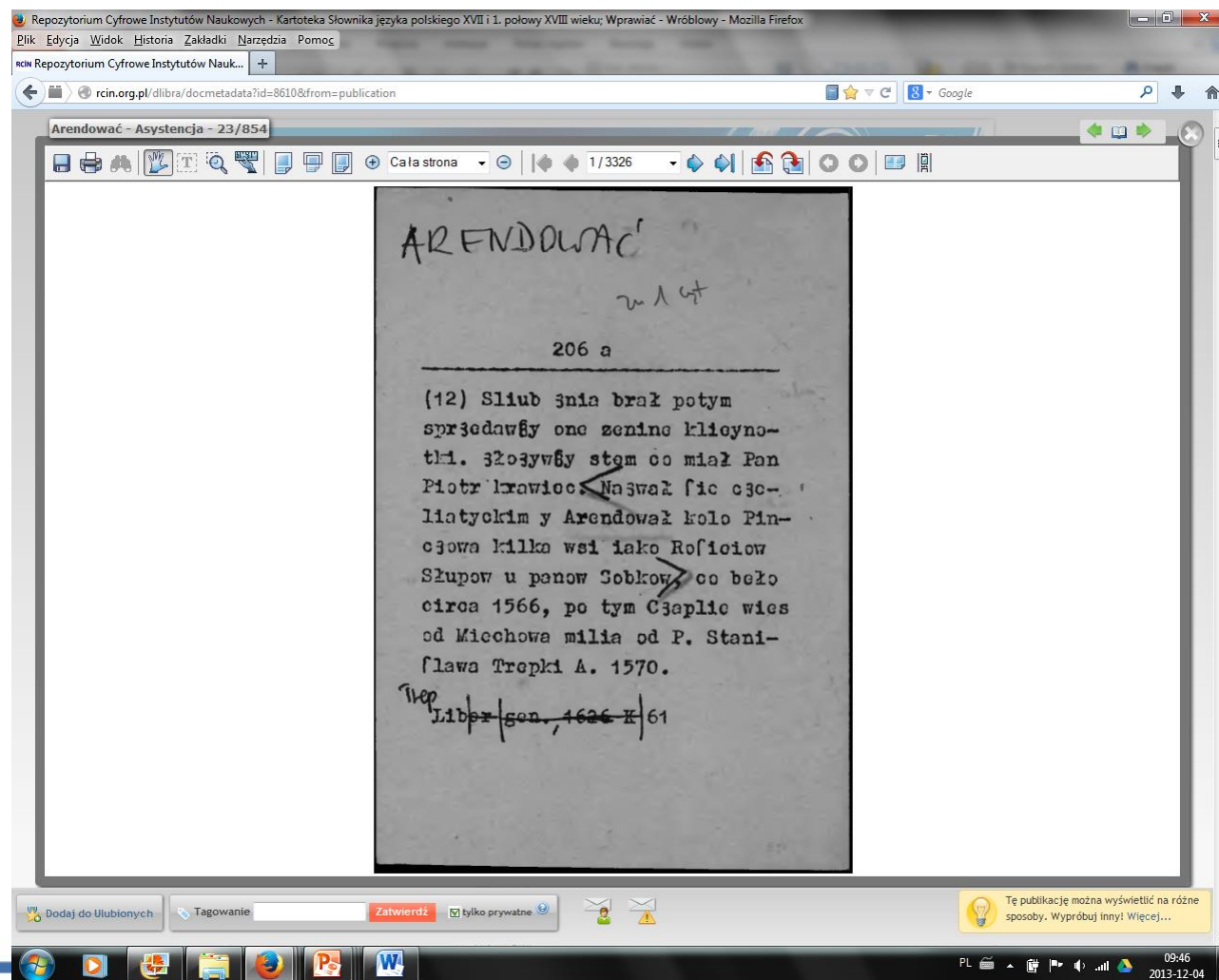


Fiszki w postaci elektronicznej

The screenshot shows a web browser window displaying the RCiN website. The page title is "Repozytorium Cyfrowe Instytutów Naukowych - Kartoteka Słownika języka polskiego XVII i 1. połowy XVIII wieku; Wprawiać - Wróblowy". The browser address bar shows "rcin.org.pl/dlibra/docmetadata?id=8610&from=publication". The website header includes the RCiN logo and navigation links: "STRONA GŁÓWNA", "KOLEKCJE", "ZAREJESTRUJ", "LOGOWANIE", and "KONTAKT". The main content area displays the title "Kartoteka Słownika języka polskiego XVII i 1. połowy XVIII wieku; Wprawiać - Wróblowy" and a list of publications under the heading "Struktura publikacji:". The list includes various entries such as "A1", "A2", "A3", "A4", "A5", "A6", "A7", "Aaron - Abo", "Abo (Abo; Albo; Albo-)", "Abo-Abundacja", "Aby 2", "Aby3 - Abzac", "Aby - Abowiem", "Acemetka - Aftować", "Aga - Alczbant", "Ale", "Ale 2", "Ale2 - Albowiem", "Aleat - Angustyja", "Ani", "Anihilacja - Apetytowy", "Apijski - Arendka", "Arendować - Asystencja", "Asystent - Aż1", and "Ażur". The left sidebar contains sections for "Wydanie" (with links like Opis, Informacje, Struktura, Treść, etc.), "Języki opisu" (with a dropdown menu set to "polski" and a "Zmień" button), "Eksport metadanych" (with buttons for OAI-PMH, RDF, RIS, and BIBTEX), and "Ulubione pozycje" (with links for Dodaj do Ulubionych, Tagowanie, and a "Zatwierdź" button).



Fiszki w postaci elektronicznej



Ogromne biblioteki cyfrowe

Federacja Bibliotek Cyfrowych | Wyszukiwanie wg opisu - Mozilla Firefox

Plik Edycja Widok Historia Zakładki Narzędzia Pomoc

Repozytorium Cyfrowe Instytutów Na... Federacja Bibliotek Cyfrowych | Wysz... +

fbcpionier.net.pl/owoc

146 1 610 071

FEDERACJA BIBLIOTEK CYFROWYCH

Wyszukiwanie Przeglądanie Biblioteki cyfrowe

Wg opisu Wg identyfikatora Zaawansowane

Opis publikacji Szukaj

Dostępnych publikacji: 1 610 071

Planowanych publikacji: 41 776

[Zobacz nasze źródła danych](#)

Poszukujemy pracowników

2013-12-03

Poszukujemy pracowników na stanowiska:

- programista Java,
- specjalista Semantic Web,
- wdrożeniowiec aplikacji Java.

Zachęcamy do zapoznania się ze [szczegółowym ogłoszeniem](#).

Konferencja openGLAM 2013

2013-08-20

Zapraszamy 11-12 października 2013 do Zachęty – Narodowej Galerii Sztuki na konferencję „openGLAM 2013. Otwarte zasoby kultury” – pierwszą w Polsce konferencję na temat idei open GLAM – nowego sposobu myślenia o misji instytucji kultury i sposobach korzystania z jej zasobów.

Tworzymy program konferencji wspólnie – do 31 sierpnia czekamy na propozycje sesji, warsztatów i wystąpień. Już teraz prowadzimy też rejestrację uczestników konferencji (udział w konferencji jest bezpłatny).

Więcej informacji: <http://openglam2013.pl/>

Więcej wiadomości z FBC | [Zobacz wiadomości z bibliotek cyfrowych](#)



Wyszukiwarki

Wyszukiwarka Poliqarp - Mozilla Firefox

Plik Edycja Widok Historia Zakładki Narzędzia Pomoc

Wyszukiwarka Poliqarp

poliqarp.wbl.klf.uw.edu.pl/pl/

- Język interfejsu
Polish Zmień
- O wyszukiwarce
- Dostępne teksty:
 - J. Karłowicz, A. Kryński, W. Niedźwiedzki. Słownik języka polskiego. Warszawa 1900-1927
 - S. Bąk, M. R. Mayenowa, F. Peplowski (red.). Słownik polszczyzny XVI wieku. Wrocław — Warszawa, 1966-???? (praca w toku)
 - M. Samuel Bogumił Linde. Słownik języka polskiego (2. wydanie). Lwów 1854-1861.
 - B. Chlebowski, F. Sulimierski, W. Walewski (red.), Słownik Geograficzny Królestwa Polskiego i innych krajów słowiańskich, Warszawa 1880-1902
 - Korpus GT projektu IMPACT (wersja 1-w.), 1570-1756
 - Korpus GT projektu IMPACT (wersja 2-w.), 1570-1756
- Ustawienia
- Podstawowa pomoc

Udostępniona przez [Katedrę Lingwistyki Formalnej](#) wyszukiwarka pozwala na wygodne przeszukiwanie tekstów w formacie DjVu. Wykorzystuje ona specjalnie zmodyfikowaną wersję systemu Poliqarp, stworzonego w Instytucie Podstaw Informatyki PAN i aktualnie rozwijanego na potrzeby [Narodowego Korpusu Języka Polskiego](#); w konsekwencji składnia zapytań jest identyczna jak w korpusie narodowym. Modyfikacji systemu dokonał [Jakub Wilk](#), który również skonwertował większość udostępnianych tekstów na format wymagany przez system. Konceptę wykorzystania wyszukiwarki Poliqarp do tekstów w formacie DjVu sformułował [Janusz S. Bień](#). Była ona prezentowana m.in. w publikacji „Dygitalizacja i komputeryzacja słowników...”, a także w publikacjach angielskojęzycznych dostępnych w [Bibliotece Cyfrowej KLF UW](#). Prace nad wyszukiwarką są finansowane przez grant MNiSzW nr N N519 384036 „Narzędzia dygitalizacji tekstów na potrzeby badań filologicznych”. Zachęca się użytkowników wyszukiwarki do korzystania z list pocztowych: [listy komunikatów](#) i [listy dyskusyjnej](#). Dokumentacja techniczna wyszukiwarki znajduje się [tutaj](#).



Korpus! (in statu nascendi)



- Projekt badawczy „Elektroniczny korpus tekstów polskich XVII i XVIII w. (do 1772 r.)”
- Finansowanie projektu: Narodowy Program Rozwoju Humanistyki na lata: 2013-2017
- Jednostka koordynująca: IJP PAN.
- Współpraca: IPI PAN.
- Kierownik: Włodzimierz Gruszczyński.
- Kryptonim: KorBa (KORpus BARokowy).



Najważniejsze założenia

- Chronologiczne rozszerzenie Narodowego Korpusu Języka Polskiego.
- Planowana objętość: 12 mln segmentów
 - mała w porównaniu z NKJP (300 mln w korpusie zrównoważonym),
 - relatywnie duża jak na korpus historyczny (np. tylko 1 mln w szwedzkich podkorpusach tekstów z okresu *Nysvenska*, tzn. 1520-1850).
- Znakowanie:
 - strukturalne, umożliwiające jak najdokładniejszą lokalizację cytatu,
 - socjolingwistyczne (informacje o autorze, wydawcy, tłumaczu),
 - stylistyczno-genologiczne,
 - tematyczne,
 - językowe (wtręty obcojęzyczne z dokładnością do języka),
 - morfosyntaktyczne (ręczne w podkorpusie 0,5-milionowym).



Korbowe narzędzia (dzięki współpracy z IPI PAN!)

- Wczytywacz,
- Anotatornia 2,
- Ameba,
- Korbeusz (postarzony Morfeusz),
- Poliqarp 2.



KorBa – perspektywa udostępnienia

- KorBa 0.1 będzie dostępna w internecie jeszcze w bieżącym roku.
- W wersji 0.1 możliwe będzie najprostsze przeszukiwanie (segmenty, ich części, ich sekwencje).
- Wersja 1.0 będzie udostępniona nie później niż w końcu 2017 r.



Dziękuję za uwagę!

