

Janusz S. Bień

Elektroniczne indeksy fiszek słownikowych

1. Wstęp

Niniejszy komunikat opisuje niektóre z zasobów przedstawionych wcześniej w referacie¹. Chodzi mianowicie o indeksy do kartotek dwóch słowników: *Słownika języka polskiego XVII i pierwszej połowy XVIII wieku*, nazywanego przez mnie krótko słownikiem późnośredniopolskim, i *Słownika łaciny średniowiecznej w Polsce*; skoncentrujemy się tutaj na pierwszej z nich. Kartoteki znajdują się w bibliotece cyfrowej projektu „Repozytorium Cyfrowe Instytutów Naukowych”² w postaci dokumentów w formacie DjVu, przy czym każdy z nich odpowiada jednemu pudełku papierowych fiszek. Oczywiście możliwe jest wyświetlanie fiszek bezpośrednio z biblioteki, ale pokażemy tutaj, że korzystanie z indeksów stanowi interesującą alternatywę. Wcześniej podobnemu tematowi był poświęcony artykuł *Dygitalizacja kartotek słownikowych*³, ale zawarte w nim informacje techniczne się zdezaktualizowały.

Tekst niniejszy może być traktowany również jako uzupełnienie i uaktualnienie pod względem technicznym artykułu dotyczącego indeksu haseł

¹ J.S. Bień, *Elektroniczne indeksy leksykograficzne i djview4poliqarp*, seminarium *Przetwarzanie języka naturalnego*, Instytut Podstaw Informatyki PAN, Warszawa 2018, <https://www.slideshare.net/jsbien/jsb-i-linde181001ipi-117452985> [dostęp: 4.10.2018].

² <http://rcin.org.pl/>.

³ J.S. Bień, *Dygitalizacja kartotek słownikowych*, „Prace Filologiczne” 2013, LXIV, s. 25–34, <https://www.ceeol.com/search/article-detail?id=4783>.

do słownika Lindego⁴. Indeks ten jest obsługiwany za pomocą tego samego narzędzia co omawiane kartoteki fiszek, mianowicie przy użyciu programu djview4poliqarp, który od czasu złożenia wspomnianego artykułu do druku został wyposażony w różne nowe funkcje wspólne dla wszystkich indeksów tego rodzaju.

Trzeba koniecznie podkreślić, że wykorzystanie tego programu do indeksów fiszek nie zostało zaplanowane z góry, jego zasadnicze przeznaczenie było inne. W konsekwencji opisane dalej procedury mogą być uznane za nieeleganckie, ale wydają się godne uwagi ze względu na ich skuteczność.

2. Niezbędny sprzęt i oprogramowanie

Opisane dalej działania wymagają dysponowania dwoma programami: djview4poliqarp i djview4.

Program djview4poliqarp jest dostępny tylko dla komputerów 64-bitowych, może pracować w systemie Linux lub MS Windows – był testowany m.in. w systemie Debian GNU/Linux (wersja stabilna i testowa), a także w systemach Windows 7 i Windows 10. Program może również pracować na komputerach firmy Apple z systemem iOS, ale wersja taka nie jest udostępniana – zapotrzebowanie okazuje się znikome, a przygotowanie – kosztowne, bo wymaga zakupu przez autora odpowiedniej autoryzacji (jest dla tego alternatywa w postaci nieoficjalnego menedżera pakietów Homebrew, ale to rozwiązanie też ma swoje wady).

Program djview4poliqarp pobieramy z repozytorium <https://bitbucket.org/mru-dolf/djview-poliqarp/>. W wersji dla MS Windows nie wymaga instalacji, można go np. uruchomić z przenośnej pamięci USB (ale działanie programu zostawia ślad przez zapisywanie konfiguracji w tzw. rejestrze systemu). Jest również dostępna do pobrania wersja dla systemów linuksowych Debian i Ubuntu. Wykorzystanie systemu na innych platformach wymaga pobrania tekstu źródłowego i samodzielnej kompilacji.

Większość funkcji programu jest dostępna za pomocą różnych menu, które podają także skróty klawiaturowe do tych funkcji. Dokumentacja w formie pomocy i osobnego tekstu nie jest niestety w pełni aktualna. Dostęp do wcześniejszej wersji znajduje się w repozytorium <https://bitbucket.org/jsbien/ndt>.

Program djview4poliqarp ściśle współpracuje z programem djview4, który wymaga instalacji. Dla MS Windows instalator pobieramy z repozy-

⁴ J.S. Bień, *Elektroniczny indeks do słownika Lindego*, „Kwartalnik Językoznawczy” 2015, nr 3, s. 1–19, http://kwartjez.amu.edu.pl/teksty/teksty2015_3-4_23-24/Bien.pdf.

torium <http://djvu.sourceforge.net/djview4.html>, w systemach Linux odpowiedni pakiet jest z reguły dostępny w systemie. Po instalacji w ustawieniach należy włączyć opcje zaawansowane.

Gdyby program `djview4poliqrp` nie uruchamiał `djview4` w opisany dalej sposób, to należy sprawdzić ustawienia tego pierwszego (pole definiujące lokalizację `djview4`).

3. Pobranie indeksów fiszek

Indeks fiszek znajduje się w repozytorium <https://bitbucket.org/jsbien/rcin-index4djvu>. Najprościej pobrać całe repozytorium w postaci skompresowanego archiwum i rozpakować je na swoim komputerze.

Jest to typowe repozytorium informatyczne, można więc śledzić zmiany (zatwierdzone, czyli *commits*), można też zgłaszać problemy i pytania (*issues*). Są również dostępne mechanizmy pozwalające wspólnie ulepszać repozytorium, np. przez dodawanie i ulepszanie indeksów. Użytkownicy, którzy nie znają tych mechanizmów i nie mają ochoty się z nimi zapoznać, mogą zgłaszać swoje propozycje jako *issues*. Dla zmniejszenia ryzyka wandalizmu wymagane jest założenie konta w systemie i zalogowanie się.

4. Wyszukiwanie fiszek i tworzenie indeksów

Dla ustalenia uwagi skoncentrujemy się na fiszce «epoka, epocha» występującej na ilustracji w przytoczonym wcześniej artykule *Dygitalizacja kartotek słownikowych*⁵.

Korzystając z menu *Index*, otwieramy indeks fiszek słownika późnośredniopolskiego (plik `RCIN-fiszki-Sjp17w.csv`) i wybieramy pozycję odpowiadającą właściwemu pudełku, czyli «> Emundowani – E». Ponieważ indeks jest względnie mały (838 pozycji, aktualny indeks haseł do słownika Lindego ma ich 77 930), łatwo to zrobić za pomocą kółka myszy lub paska przewijania, klikając lewym klawiszem na wybraną pozycję. Jest to jednak dobra okazja do przeciwiczenia tzw. wyszukiwania przyrostowego: do pola *Find* wpisujemy kolejno: znak `>`, spację (prefiks ten powoduje, że w indeksie będą wyraźnie oddzielone pudełka od indywidualnych fiszek), literę *e* (nie ma potrzeby mniej wygodnego wpisywania *E*, bo wyszukiwanie ignoruje wielkość liter) itd. – patrzmy, jak zmienia się aktualna pozycja w miarę wpisywania, a kiedy podświetli się ta właściwa, zatwierdzamy ją przez

⁵ J.S. Bień, *Dygitalizacja kartotek słownikowych*.

wciśnięcie Enter. Spowoduje to wyświetlenie pierwszej fiszki tego pudełka w panelu głównym, sposób wyświetlania można zmienić za pomocą menu View Zoom lub stosując podane tam skróty klawiaturowe.

Jeśli panel korpusów (po lewej stronie) jest niewidoczny, należy go wyświetlić za pomocą menu View, po czym wybrać panel wyników graficznych. Następnie, klikając dwukrotnie na hasło w indeksie, powodujemy, że fiszka wyświetla się również w okienku wyników graficznych, w którym klikamy potem środkowym klawiszem myszy (konwencja ta nawiązuje do otwierania linku w nowym oknie w przeglądarkach WWW), dzięki czemu fiszka wyświetli się dodatkowo w programie djview4 (w razie potrzeby możemy to powtórzyć i oglądać nasz dokument w kilku egzemplarzach djview4 jednocześnie – czasami okazuje się to przydatne).

Dzięki użyciu djview4 mamy teraz dostęp do informacji o liczbie stron w dokumencie i dodatkowej funkcji przejścia do strony o wskazanym numerze, możemy więc użyć tzw. wyszukiwania binarnego – bardzo szybkiego sposobu polegającego na wybieraniu środkowego elementu z coraz mniejszego przedziału. Warto przy tym korzystać z kalkulatora.

Nasze pudełko ma 1778 fiszek, przejdźmy więc do tej z nr. 887. Jest to fiszka «ESENcjALNY (punkt)», którą nasza fiszka poprzedza, przechodzimy więc do środkowej fiszki z pierwszej połowy dokumentu – czyli nr 443 – «Erato O.», którą też szukana fiszka poprzedza, więc przechodzimy do środkowej fiszki w pierwszej ćwiartce dokumentu. Jest to fiszka nr 221 «EPIGRAMA», która poprzedza naszą, powinniśmy zatem przejść do środkowej w przedziale 221–443, czyli nr 331. Jest to «EPITAFIA», też poprzedzająca szukaną. Idziemy zatem do fiszki nr 390 «EPITAFIUM», a następnie nr 415 «EPITALAMIUM». Możemy tę procedurę kontynuować, ale pozostało nam do przejrzania już tylko 28 fiszek, zaleca się je po prostu przejrzeć po kolei. Po drodze znajdujemy nr 434 «EPOCHA» (błędnie umieszczoną między fiszkami «EPITYMOWY»), a na stronie 436 szukaną fiszkę «epoka, epocha».

Teraz możemy wrócić do djview4poliqarp i przygotować go do zapisania adresu znalezionej fiszki. Przy wciśniętym klawiszu CTRL i lewym przycisku myszy zaznaczamy na panelu głównym dowolny prostokąt – powoduje to otworenie panelu podręcznego, z którego wybieramy *Add index entry*. W polu *Entry* wpisujemy odpowiedni tekst, np. „epoka” (może być ono wstępnie wypełnione, jeśli zaznaczyliśmy tekst poddany automatycznemu rozpoznawaniu znaków), a zawartość pola URL kasujemy. Wracamy obecnie do djview4 i w analogiczny sposób zaznaczamy nagłówek fiszki, a w menu podręcznym wybieramy *Copy URL* (nie będzie tej opcji, jeśli wcześniej nie wybraliśmy funkcji zaawansowanych). Następnie wkle-

jamy zawartość schowka do pola URL w *djview4poliqarp* i wybieramy *OK* (możemy oczywiście wypełnić również jakoś pola *Comment* i *Description*). W panelu indeksu zobaczymy nowe utworzone przez nas hasło. Zmodyfikowany indeks możemy zapisać natychmiast, korzystając z menu *Index*, lub wybrać taką możliwość przy kończeniu pracy z programem.

Jeśli planujemy szukać w przyszłości jeszcze innych fiszek w tym pudełku, warto do indeksu wstawiać także wyniki pośrednie z jakimś komentarzem, np. numerem strony. Można też, zamiast wyliczać kolejne numery fiszek, podawać ich intuicyjne oszacowania, co tylko nieznacznie wydłuży liczbę kroków niezbędnych do znalezienia fiszki. Ponieważ ich liczba w poszczególnych pudełkach jest mniej więcej taka sama, można również odpowiednio wartości wyliczyć wcześniej i korzystać z nich wielokrotnie.

5. Fiszki a inne zasoby

Pokażemy niżej, jak znaną fiszkę skonfrontować z innymi przykładowymi zasobami dostępnymi w systemie *Poliqarp for DjVu*, aktualnie na serwerze Fundacji Języka Polskiego⁶. Szczegółowe wyjaśnienie użytych kwerend można znaleźć w opisach systemu *Poliqarp*⁷.

5.1. Korpus IMPACT

Z adnotacji na fiszce wiemy, że cytat pochodzi z *Nowych Aten*, które są jednym z utworów dostępnych w korpusie IMPACT⁸. Najbardziej charakterystycznym słowem cytatu wydaje się *Veteris*.

Najpierw wybieramy korpus jednowymiarowy i podajemy kwerendę *Veteris* (duża litera jest istotna). Otrzymujemy dwa trafienia, ale na skanach zaznaczone są całe akapity. Powtarzamy więc kwerendę na korpusie dwuwymiarowym – otrzymujemy również dwa trafienia, co oznacza, że żadne z wystąpień słowa nie jest przeniesione do nowej linii. W korpusie dwuwymiarowym lokalizacja trafienia jest podawana z dokładnością do wiersza, można więc łatwo sprawdzić, gdzie w tekście występuje cytat z fiszki. Mo-

⁶ <https://szukajwslownikach.uw.edu.pl/>.

⁷ Zob. np. A. Przepiórkowski, *Korpus IPI PAN. Wersja wstępna*, Warszawa 2004, http://nlp.ipipan.waw.pl/~adamp/Papers/2004-corpus/book_pl.pdf [dostęp: 14.12.2018].

⁸ J.S. Bień, *The IMPACT project Polish Ground-Truth texts as a DjVu corpus*, „Cognitive Studies | Études Cognitives” 2014, 14, s. 75–84, <https://ispan.waw.pl/journals/index.php/cs-ec/article/view/cs.2014.008>.

żemy to wystąpienie dodać do indeksu jako hasło w analogiczny sposób, jak robiliśmy to poprzednio.

Nasuwa się naturalne pytanie, jakie jeszcze wystąpienia słowa *epocha* znajdują się w *Nowych Atenach* i ogólnie w korpusie. Podajemy zatem kwerendę „epocha/i” (/i oznacza ignorowanie rozróżnienia dużych i małych liter). Otrzymujemy osiem trafień, które najlepiej przejrzeć w oknie metadanych – do następnego lub poprzedniego trafienia przechodzimy przyciskami *Next* i *Previous*. Kliknięcie w tym oknie na trafienie wyświetla je w *djview4* – jest to sposób wywołania tego programu alternatywny do poznanego wcześniej. Jeśli jednak chcemy obejrzeć to trafienie na skanie w głównym panelu, to musimy przejść do widoku wyników tekstowych lub graficznych i tam dwukrotnie kliknąć na odpowiedni wiersz lub okienko. Jeśli któreś trafienie chcemy zapisać w indeksie, postępujemy podobnie, jak to było opisane wcześniej.

W celu znalezienia wszystkich form fleksyjnych należy użyć kwerendy „epoch./i” (słowa sześcioliterowe zaczynające się od *epoch*, czyli m.in. pozostałe przypadki – 13 trafień) i ewentualnie „eposze/i” (zero trafień). Można potem samodzielnie ocenić, czy przykład, który znalazł się w kartotece słownika późnośredniopolskiego, jest najbardziej reprezentatywny.

Jeśli interesują nas wszystkie wystąpienia, można kompletny wynik kwerendy zapisać w pliku i potem przekształcić go na indeks. Możemy również wyświetlić i zapisać wynik, który jest wierniejszy oryginałowi, bo zachowuje oryginalne ligatury typu *Æ*.

Zainteresowany Czytelnik może powtórzyć to wyszukiwanie w dostępnym od 2017 roku korpusie barokowym⁹ i porównać wyniki. Korpus znajduje się pod adresem <http://korba.edu.pl> i też zawiera *Nowe Ateny*; sugerowana postać kwerend to `[orth_lc="epoch."]` i `[base="epoch"]` – korpus wykorzystuje system MTAS¹⁰, którego składnia kwerend różni się od tej w systemie Poliqarp.

⁹ Zob. np. R. Bronikowska i in., *The Use of Electronic Historical Dictionary Data in Corpus Design*, „Studies in Polish Linguistics” 2016, Volume 11, Issue 2, <http://www.ejournals.eu/SPL/2016/Issue-2/art/7035/>.

¹⁰ M. Brouwer, H. Brugman, M. Kemps-Snijders, *MTAS: A Solr/Lucene based Multi Tier Annotation Search solution*, [w:] *Selected papers from the CLARIN Annual Conference 2016, Aix-en-Provence, 26–28 October 2016, CLARIN Common Language Resources and Technology Infrastructure*, Linköping University Electronic Press, Linköpings Universitet, s. 19–37, <http://www.ep.liu.se/ecp/136/002/ecp17136002.pdf> [dostęp: [14.12.2018].

5.2. Słownik polszczyzny XVI wieku

Wybieramy korpus słownika i podajemy kwerendę EPOCHACHA (duże litery są istotne). Otrzymujemy zero trafień, czyli takie hasło w słowniku nie istnieje (hasła są pisane w całości dużymi literami). Możemy dodatkowo wykonać kwerendy opisane wyżej, pobierając je z historii – również otrzymujemy zero trafień.

Możemy wykonać analogiczne kwerendy dla słowa EPOKA. Przekonamy się wówczas, że takie hasło w słowniku także nie istnieje – słowo pojawia się w tekście słownika, ale wszystkie wystąpienia pochodzą z objaśnień, a nie cytatów.

5.3. Słownik Lindego

Wykorzystując te same kwerendy dla słownika Lindego, przekonamy się, że słowo EPOCHACHA nie występuje w żadnej formie, jest natomiast hasło EPOKA. Jeśli chcemy, możemy je dodać do indeksu.

5.4. Słownik warszawski

Stosując wspomniane kwerendy, znajdziemy zarówno hasło EPOCHACHA, jak i EPOKA, to pierwsze ilustrowane cytatem – z pomocą wyszukiwarek Książki Google i Federacji Bibliotek Cyfrowych można ustalić, że cytat pochodzi z 1785 roku z dzieła *Nauka Prawa Przyrodzonego, Politycznego, Ekonomiki Politycznej, y Prawa Narodów* autorstwa Hieronima Strojnowskiego¹¹. Jest ono dostępne w formacie DjVu w Elbląskiej Bibliotece Cyfrowej, więc cytat ten też możemy dodać do indeksu (choć niedoświadczony użytkownik może mieć z tym problemy techniczne).

6. Zakończenie

Omawiane indeksy powstały przez adaptację danych przeznaczonych dla programu Maleks (omawianego we wspomnianym wcześniej artykule¹²), który z różnych powodów nie wszedł do użytku. Były one stworzone pół-automatycznie, zrobienie takich indeksów całkowicie ręcznie byłoby bardzo żmudne. Niestety odbywało się to w pośpiechu i cała procedura jest udoku-

¹¹ H. Strojnowski, *Nauka Prawa Przyrodzonego, Politycznego, Ekonomiki Politycznej, y Prawa Narodów*, Wilno 1785, Drukarnia Królewska przy Akademii, s. 230.

¹² J.S. Bień, *Dygitalizacja kartotek słownikowych*, *op.cit.*

mentowana tylko w prywatnych notatkach, z wyjątkiem pomocniczego programu fiszki, który jest udostępniony w repozytorium. Kluczowy problem polega na tym, że w Repozytorium Cyfrowym Instytutów Naukowych, jak we wszystkich bibliotekach cyfrowych używających systemu dLibra, zwykły użytkownik ma dostęp do adresu dokumentu DjVu dopiero wtedy – i to tylko pośrednio – gdy go sobie we właściwy sposób wyświetli.

Warto zwrócić uwagę, że z formalnego punktu widzenia indeksy są prostymi arkuszami kalkulacyjnymi i mogą być edytowane lub w inny sposób przetwarzane za pomocą odpowiednich programów. Można dzięki temu dodać wygodnie komentarze lub zmodyfikować hasła – np. hasło fiszki «epoka, epoka» rozbić na dwa: właściwe «epoka» i pomocnicze «epoka».

Korzystanie z fiszek może powodować pewne problemy prawne, o których mówiono¹³ i pisano¹⁴ wcześniej. Według mojej najlepszej wiedzy od tego czasu sytuacja nie uległa zmianie.

W opisanej procedurze w istotny sposób wykorzystujemy funkcję, która do programu `djview4poliqrp` została wprowadzona dopiero w 2018 roku w wersji 3.0, mianowicie możliwość „przenoszenia” haseł indeksu do panelu graficznego. Podstawowym przeznaczeniem tej funkcji jest wyświetlanie naraz wszystkich homonimów z indeksu do słownika Lindego¹⁵.

Przykładowy indeks widoczny na ilustracji jest umieszczony w repozytorium indeksów jako `epoka.csv`, uwzględnia on m.in. wyniki pośrednie wyszukiwania i wyniki kwerendy w korpusie IMPACT.

¹³ J.S. Bień, *Otwarty dostęp do zasobów lingwistycznych w praktyce. Czy na badanie polszczyzny potrzebne jest pozwolenie?*, [w:] *Język jest labiryntem ścieżek. Nowe kierunki i nowe zadania w badaniach nad językiem polskim*, Warszawa 2016, <https://www.slideshare.net/jsbien/otwarty-dostp-do-zasobw-lingwistycznych-w-praktyce> [dostęp: 4.10.2018].

¹⁴ J.S. Bień, *Otwarty dostęp do zasobów lingwistycznych w praktyce*, „Prace Filologiczne” 2017, LXXI, s. 23–31, <https://www.ceeol.com/search/article-detail?id=602323>.

¹⁵ Por. J.S. Bień, *Elektroniczny indeks do słownika Lindego*.



Rys. 1. Przykładowy indeks (patrz <https://bitbucket.org/jsbien/rcin-index4djuv/downloads/epocha-IMPACT1.png>)

Bibliografia

- Bień J.S., *Dygitalizacja kartotek słownikowych*, „Prace Filologiczne” 2013, LXIV, s. 25–34, <https://www.ceeol.com/search/article-detail?id=4783>.
- Bień J.S., *Elektroniczne indeksy leksykograficzne i djview4poliqarp*, seminarium *Przetwarzanie języka naturalnego*, Instytut Podstaw Informatyki PAN, Warszawa 2018, <https://www.slideshare.net/jsbien/jsb-i-linde181001ipi-117452985> [dostęp: 4.10.2018].
- Bień J.S., *Elektroniczny indeks do słownika Lindego*, „Kwartalnik Językoznawczy” 2015, nr 3, s. 1–19, http://kwartjez.amu.edu.pl/teksty/teksty2015_3-4_23-24/Bien.pdf.
- Bień J.S., *Otwarty dostęp do zasobów lingwistycznych w praktyce*, „Prace Filologiczne” 2017, LXXI, s. 23–31, <https://www.ceeol.com/search/article-detail?id=602323>.
- Bień J.S., *Otwarty dostęp do zasobów lingwistycznych w praktyce. Czy na badanie polszczyzny potrzebne jest pozwolenie?*, [w:] *Język jest labiryntem ścieżek. Nowe kierunki i nowe zadania w badaniach nad językiem polskim*, Warszawa 2016, <https://www.slideshare.net/jsbien/otwarty-dostp-do-zasobw-lingwistycznych-w-praktyce> [dostęp: 4.10.2018].
- Bień J.S., *The IMPACT project Polish Ground-Truth texts as a DjVu corpus*, „Cognitive Studies | Études Cognitives” 2014, 14, s. 75–84, <https://ispan.waw.pl/journals/index.php/cs-ec/article/view/cs.2014.008>.
- Bronikowska R. i in., *The Use of Electronic Historical Dictionary Data in Corpus Design*, „Studies in Polish Linguistics” 2016, Volume 11, Issue 2, <http://www.ejournals.eu/SPL/2016/Issue-2/art/7035/>.
- Brouwer M., Brugman H., Kemps-Snijders M., *MTAS: A Solr/Lucene based Multi Tier Annotation Search solution*, [w:] *Selected papers from the CLARIN Annual Conference 2016, Aix-en-Provence, 26–28 October 2016, CLARIN Common Language Resources and Technology Infrastructure*, Linköping University Electronic Press, Linköpings Universitet, s. 19–37, <http://www.ep.liu.se/ecp/136/002/ecp17136002.pdf> [dostęp: 14.12.2018].
- Przepiórkowski A., *Korpus IPI PAN. Wersja wstępna*, Warszawa 2004, Instytut Podstaw Informatyki, PAN, http://nlp.ipipan.waw.pl/~adamp/Papers/2004-corpus/book_pl.pdf [dostęp: 14.12.2018].
- Strojnowski H., *Nauka Prawa Przyrodzonego, Politycznego, Ekonomiki Politycznej, y Prawa Narodów*, Wilno 1785, Drukarnia Królewska przy Akademii.

Electronic indexes for dictionary slips (fiches)

SUMMARY

Dictionary slips (fiches) for several important Polish historical dictionaries are publicly available in a digital library as DjVu documents. An open source program is described which allows to browse those documents conveniently and to create indexes of selected slips.

Key words: lexicography, dictionary slips, digitization, DjVu.

O Autorze

Prof. dr hab. Janusz S. Bień - emeryt (wcześniej Katedra Lingwistki Formalnej Wydziału Neofilologii Uniwersytetu Warszawskiego); informatyk i lingwista (z wykształcenia matematyk); aktualne zainteresowania: dygitalizacja dawnych tekstów polskich, w tym słownika Lindego i traktatu Parkosza, historia pisowni polskiej.
E-mail: jsbien@uw.edu.pl