

KRZYSZTOF SZAFRAN

Warszawa

Ku elektronicznej wersji *Słownika polszczyzny XVI wieku*

W niniejszym artykule przedstawione zostaną ogólne zarysy projektu, którego celem jest stworzenie elektronicznej wersji *Słownika polszczyzny XVI wieku*. Pierwsza część tekstu stanowi wprowadzenie do tematyki słowników elektronicznych. Przedstawimy w niej, między innymi, argumenty przemawiające za elektroniczną słowników oraz typowe przykłady słowników elektronicznych. W części drugiej wskażemy prace niezbędne dla realizacji ostatecznego celu projektu.

Słowniki elektroniczne – wprowadzenie

Ogromny postęp technologiczny ostatnich lat niezwykle wyraźnie daje się zaobserwować w dziedzinie wymiany i rozpowszechniania informacji. Do niedawna podstawowym nośnikiem używanym do tych celów był papier. Obecnie coraz częściej słyszy się głosy o zmierzchu tradycyjnej książki. Niezależnie od tego, czy są one uzasadnione, nie można nie zauważyć zalet innych nośników i technologii przechowywania i przekazywania informacji. Najważniejsze z nich to oczywiście nośniki i technologie cyfrowe (zwane też niekiedy komputerowymi czy elektronicznymi). Najpopularniejsza wśród nich jest obecnie płyta CD.

Zalet stosowania nośników cyfrowych do przechowywania, a technologii cyfrowej do tworzenia różnego rodzaju publikacji, w szczególności słowników, jest wiele.

- Słownik w wersji komputerowej jest znacznie mniejszy i wygodniejszy w użyciu. Może zawierać dane z wielu tradycyjnych woluminów, a nawet z wielu różnych słowników. Wykorzystanie Internetu pozwala na sięganie do słowników, których fizycznej kopii nie musimy mieć w domu. Możliwe jest wręcz przezroczyste¹ korzystanie z sieci rozproszonych słowników elektronicznych, to znaczy takich, których poszczególne elementy udostępniane są przez różne ośrodki i fizycznie znajdują się w różnych miejscach.
- Słownik elektroniczny daje możliwości użytkowe, których nie dają słowniki drukowane. Wyposażenie go w odpowiednie oprogramowanie, np. analizator mor-

¹ Niewidoczne na poziomie użytkowym.

fologiczny, umożliwia znacznie efektywniejsze posługiwanie się nim. Może on być też wzbogacony o informacje niemożliwe do zamieszczenia w tradycyjnym słowniku, np. dźwiękowe czy filmowe.

- Dużymi zaletami słownika elektronicznego są: potencjalnie niższa cena i łatwiejsza dostępność (zwłaszcza w przypadku wydawnictw unikalnych albo zabytkowych). Cena płyty CD jest wielokrotnie (o kilka rzędów wielkości) niższa niż cena książki. Zapisane na płycie informacje znacznie łatwiej poddają się kopiowaniu.
- Słownik elektroniczny jest też zapewne trwalszy niż tradycyjny, choć – ze względu na relatywnie krótki okres stosowania komputerów – nie można tego sądu formułować kategorycznie. Oczywiście nośnik papierowy w miarę upływu czasu ulega nieodwracalnemu niszczeniu, natomiast nośniki cyfrowe są wielokrotnie od papieru trwalsze. Nie jest to jednak równoznaczne z trwałością zapisu na takim nośniku, być może ulega on wcześniej uszkodzeniu. Jednak łatwość kopiowania zapisu i przenoszenia go na inne nośniki sprawia, że – przy rozsądnej polityce archiwalnej – słownik elektroniczny może być znacznie lepiej przechowywany i chroniony niż wydrukowany w wielu kopiach na papierze.

Są też oczywiście pewne mankamenty. Słownik elektroniczny wymaga posiadania odpowiedniego komputera i przełamania pewnych przyzwyczajeń co do sposobu pracy ze słownikiem. To jednak niewielka cena, którą trzeba zapłacić za potencjalnie wysokie zyski.

Nic zatem dziwnego, że coraz częściej dodatkiem do tradycyjnej (papierowej) wersji książki jest płyta CD zawierająca jej wersję elektroniczną. Dotyczy to zwłaszcza książek takich jak słowniki.

Czym jest słownik elektroniczny

Wspólnym elementem wydawnictw określanym terminem² *słownik elektroniczny* jest umieszczenie materiału słownikowego na współczesnym nośniku komputerowym oraz dołączenie do niego oprogramowania umożliwiającego korzystanie z tego materiału w sposób nie mniej wygodny niż ze słownika w wersji papierowej.

Konkretne rozwiązania mogą znacznie różnić się między sobą. W niektórych przypadkach oprogramowanie dostarcza bardzo wyrafinowanych funkcji i bardzo znacznie rozszerza zakres możliwości wykorzystania materiału słownikowego.

Wielu wydawców używa terminu *elektroniczna wersja słownika* na określenie płyty CD (niekiedy DVD) stanowiącej dodatek do wersji książkowej słownika. Płyta taka zwykle zawiera materiał identyczny jak książka, wzbogacony jednak o jakieś programy narzędziowe.

² Niedawno wydany *Inny słownik języka polskiego PWN* [1] w haśle **słownik** podaje: **Słownik to książka, która zawiera...** Hasło **elektroniczny** tego słownika mówi natomiast, że: **Elektroniczny to taki, który ma związek z elektroniką.** W haśle tym jednak nie wymienia się słownika elektronicznego ani książki elektronicznej.

Coraz częściej pojawiają się również wydawnictwa wyłącznie płytowe, nie odpowiadające jednemu konkretnemu słownikowi tradycyjnemu i nie będące, siłą rzeczy, załącznikiem do konkretnej książki.

Najprostszym możliwym wariantem słownika elektronicznego jest elektroniczna kopia (ang. electronic facsimile). W postaci takiej kopii – graficznego obrazu poszczególnych tomów – wydany został ostatnio sławny *Słownik Geograficzny Królestwa Polskiego* [15]. Ponieważ każdy tom zapisany tu został w jednym, dużym pliku, korzystanie ze słownika nie jest najwygodniejsze.

Polskiemu użytkownikowi słowników najlepiej znane są dwa przykłady elektronicznych wydawnictw leksykograficznych.

Pierwszy, elektroniczna wersja *Słownika języka polskiego pod red. W. Doroszewskiego* [4, 11] jest rozwiązaniem stosunkowo prostym, ale mimo to w pełni funkcjonalnym. Jest to, w zasadzie, prawie dokładny graficzny obraz słownika papierowego. Modyfikacje polegają na włączeniu haseł z suplementu do głównej listy haseł, uwzględnieniu errat oraz zamieszczeniu dodatkowego indeksu. Dołączone oprogramowanie umożliwia użytkownikowi wyszukiwanie haseł za pomocą indeksu oraz tworzenie zakładek i notatek. Nic dziwnego, że ta wersja słownika zdobyła uznanie odbiorców i do dzisiaj znajduje się w ofercie wydawcy.

Jeszcze większą popularność uzyskało drugie dzieło, znacznie bardziej wyrefinowane. *Komputerowy słownik języka polskiego PWN* [9, 10] zawiera materiał z dwu słowników tradycyjnych, a mianowicie ze *Słownika języka polskiego PWN* [12] oraz ze *Słownika wyrazów obcych* [14]. Oprogramowanie słownika, wzbogacone o moduł analizy morfologicznej, umożliwia automatyczne przeszukiwanie artykułów hasłowych pod kątem występujących w nim słów. W Edycji 1998, słownik wzbogacony został o nagrania wymowy wyrazów obcych i o wymowie trudnej (nietypowej). Szczegółowy opis możliwości tego słownika zawiera, między innymi, praca [2]. Niestety od pewnego czasu słownik ten nie jest dostępny na rynku.

Dwa dalsze przykłady dotyczą ważnych, historycznych słowników narodowych udostępnianych potencjalnym użytkownikom nie bezpośrednio w postaci płyt, ale za pośrednictwem sieci komputerowej.

Pierwszy, to projekt o nazwie *The Electronic Scottish National Dictionary* (eSND), celem którego jest stworzenie elektronicznej, internetowej wersji 10 tomowego słownika współczesnego języka szkockiego³ obejmującego słownictwo od roku 1700 do chwili obecnej [7]. Słownik ten ukazał się w latach 1931–76 w 40 zeszytach, z których ostatni zawierał suplement do wcześniejszych tomów.

Osiągnięcie zamierzonego celu wymagało zastosowania wieloetapowej procedury, której pierwszym krokiem było skanowanie, czyli przeniesienie obrazu słownika z nośnika papierowego na elektroniczny. Graficzny obraz stron słownika jest (o ile mi wiadomo prace te nie zostały zakończone) następnie poddawany obróbce z zastosowaniem programów rozpoznających tekst (zwanym w skrócie OCR – Optical

³Język szkocki jest (bogata w dialekty regionalne) formą języka angielskiego używaną (przede wszystkim) w Szkocji.

Character Recognition). Uzyskane w ten sposób dane w formacie dokumentu typu MS Word poddawane są niezwykle żmudnej korekcie i konwersji do formatu XML z uwzględnieniem zaleceń TEI (patrz dalej). Konwersja z zachowaniem zaleceń TEI okazała się jednak niebanalna. Konieczne były pewne rozszerzenia i modyfikacje tego standardu.

Drugi przykład to praca mogąca być załączkiem sieci łączącej wszystkie ważniejsze słowniki okresów historycznych i dialektów języka niemieckiego. Celem pracy [5] było zrealizowanie cyfrowej sieci (*Mittelhochdeutsche Wörterbücher Online*) obejmującej wszystkie dostępne słowniki średniowiecznej niemczyzny z okresu 1050-1350. Słowniki te ze względu na ich budowę oraz fakt, że zawierają słownictwo z różnych źródeł, w praktyce używane są łącznie.

Dane ze słowników zostały zakodowane zgodnie ze standardem TEI (znowu z pewnymi problemami). Dostęp do tych danych zrealizowany jest poprzez bazę danych. Pozwala to na wyszukiwanie według określonych kryteriów. Przeszukiwanie może dotyczyć wszystkich słowników tworzących sieć albo zostać ograniczone do niektórych z nich. Użytkownik korzystający z tej sieci ma do dyspozycji rozbudowany interfejs graficzny.

Zalecenia TEI oraz SGML/XML

TEI (The Text Encoding Initiative, por. <http://www.tei-c.org>) jest grupą osób i instytucji, której głównym celem jest rozwijanie zastosowań komputera opartych na językach SGML/XML (Standard Generalized Markup Language/Extensible Markup Language) dla różnych projektów w zakresie nauk humanistycznych.

Język XML (An Extensible Markup Language), będący faktycznie uproszczeniem języka SGML, jest metajęzykiem adiustacyjnym (znacznikowym) używanym do reprezentowania tekstów, przede wszystkim w wersji elektronicznej.

W każdym tekście nie będącym przypadkowym ciągiem słów czy bajtów da się wyróżnić elementy oddające jego strukturę. Co więcej, w tym samym tekście, w zależności od potrzeby, można wyodrębnić wiele różnych, niezależnych struktur. Np. powieść możemy podzielić na tomy, rozdziały, akapity, zdania – podział ten odzwierciedla logiczną strukturę tekstu. Jednocześnie, po wydrukowaniu, ten sam tekst da się podzielić np. na voluminy czy strony – taki podział oddaje strukturę fizyczną. Ponieważ obie te struktury są niezależne, ich elementy mogą się przecinać, np. zdanie może zaczynać się na jednej stronie, a kończyć na następnej.

Historycznie termin *markup* oznaczał znaki umieszczane w tekście roboczym przeznaczonym do sporządzenia składu, informujące drukarza o sposobie jego złożenia. Współczesne języki znacznikowe, takie jak XML, wykorzystują taki sam mechanizm. Tekst w nich reprezentowany zawiera dodatkowe elementy niosące informacje o jego strukturze i/lub wyglądzie.

Język XML charakteryzuje się kilkoma istotnymi cechami.

Po pierwsze, mocny nacisk położony został na opisowy charakter znaczników. XML koncentruje się na znaczeniu danych, a nie na ich wyglądzie. Taki sposób opisu tekstu pozwala na przetwarzanie go na wiele różnych sposobów, w zależności od potrzeb.

Po drugie, XML traktuje dokumenty tekstowe jako obiekty posiadające swój typ. Definicja typu dokumentu określa, jakie części i w jakiej kolejności muszą wystąpić w dokumencie oraz jaka jest struktura tych części. Przyjęcie takiej definicji pozwala na automatyczne sprawdzanie, czy dany dokument jest zbudowany poprawnie.

Po trzecie, niezwykle istotną cechą tekstu zapisanego w XML jest jego niezależność od środowika sprzętowego czy systemowego. Oznacza to możliwość przenoszenia dokumentu z jednego środowiska do innego bez straty informacji. Na przykład dokument przygotowywany w środowisku Windows może być łatwo przeniesiony w środowisko Uniksowe. Wszystkie znaki w dokumentach reprezentowanych w XML kodowane są w standardzie Unicode, który pozwala na reprezentowanie wszystkich znaków pisańskich używanych na świecie.

TEI opracowała międzynarodowy standard reprezentacji tekstów naukowych w języku SGML/XML znany jako *Zalecenia TEI* (The TEI Guidelines). Jest on dość powszechnie stosowanym w rozmaitych projektach związanych z elektronicznymi wersjami różnych dokumentów, w szczególności ważnych słowników narodowych.

W publikacjach na temat słowników elektronicznych można jednak znaleźć opinie, że w wielu przypadkach standard TEI nie jest wystarczający dla opisu struktury artykułu hasłowego. Dlatego też często twórcy słowników nie stosują się do niego ściśle. Natomiast większość z nich korzysta jednak z jakiegoś formalizmu znacznikowego. Najczęściej jest to XML/SGML.

Jak wyraźnie pokazuje przypadek *Angielsko-polskiego słownika elektronicznego XeLDA* opisany przez K. Głowińską i M. Wolińskiego [6] istotny jest dobrze przemyślany, strukturalny sposób reprezentowania materiału. Przy jego zachowaniu możliwa jest automatyczna konwersja do praktycznie dowolnego formatu, również SGML/XML. Jeśli jest to możliwe, lepiej jest oczywiście używać formalizmu standardowego.

Również polscy wydawcy liczący się na rynku słownikarskim wykorzystują przy tworzeniu słowników formalizmy znacznikowe. Jak wynika z informacji uzyskanej od kierownika redakcji słowników języka polskiego PWN M. Bańki, prawie wszystkie słowniki przygotowywane w redakcji tworzone są w formacie SGML lub XML.

Słownik polszczyzny XVI wieku

Jak wiadomo, *Słownik polszczyzny XVI wieku* rozpoczęty z inicjatywy M.R. Mayenowej, przygotowywany jest przez kierowaną obecnie przez profesora dra hab. Franciszka Peplowskiego Pracownię Słownika Polszczyzny XVI w., będącą jednostką organizacyjną Instytutu Badań Literackich Polskiej Akademii Nauk. Po reorganizacji – w chwili obecnej pracownię tworzą dwa zespoły redakcyjne: toruński i wrocławski.

Pierwszy tom słownika [13] ukazał się drukiem w nakładzie 1000+180 egzemplarzy w 1966 roku. Tomy I-XXII wydawane były we Wrocławiu w Zakładzie Narodowym im. Ossolińskich. Począwszy od tomu XXIII słownik wydawany jest w Warszawie, w Wydawnictwie IBL PAN. Pełniejszą informację dotyczącą poszczególnych tomów (np. zakres haseł oraz nazwiska redaktorów) znaleźć można na stro-

nie internetowej Instytutu Badań Literackich (<http://www.ibl.waw.pl/>). Do chwili obecnej ukazało się 31 tomów słownika. Tom XXXI, zawierający hasła do **przemieść** ukazał się wiosną 2004 roku. Zakładając, że liczba haseł rozpoczynających się od poszczególnych liter alfabetu polskiego w polszczyźnie XVI wieku jest zbliżona do tej w polszczyźnie współczesnej, słownik nieznacznie przekroczył półmetek.

Jak można przeczytać we wstępie do tomu I, prace nad słownikiem rozpoczęły się w Instytucie Badań Literackich PAN w roku 1949. Do składania tom oddano w 1964 roku. Pierwsze tomy przygotowane zostały do druku metodą tradycyjnego składu gorącego.

Od pewnego czasu (konkretnie od tomu XXII, który został wydany w roku 1995) kolejne tomy są w Pracowni nie tylko opracowywane, ale również składane i przekazywane do druku w postaci plików komputerowych. Wykorzystywany był do tego edytor MiniCyfroset 1.0 oraz system fotoskładu Cyfroset 2.6. Niestety zachowały się tylko pliki zawierające tomy XXIV, XXVII, XXVIII, XXIX i XXX.

Począwszy od tomu XXXI słownik opracowywany jest i składany za pomocą nowszego systemu o nazwie KOMBI.

Komputerowa wersja Słownika

Jak widać, *Słownik polszczyzny XVI w.* jest wydawnictwem bardzo specyficznym. Ze względu na swój charakter tworzony jest z bardzo dużym nakładem pracy i kosztów. Z drugiej strony wydawany jest w niewielkim nakładzie, co powoduje, że jest dość trudno dostępny. Ponadto większość tomów istnieje wyłącznie w postaci książkowej. Z tych względów warto chyba rozważyć możliwość digitalizacji tego monumentalnego wydawnictwa.

Poszczególne tomy słownika przygotowywane były różnymi technikami, co oczywiście ma wpływ na zakres i rodzaj ewentualnych prac niezbędnych dla uzyskania słownika w wersji cyfrowej.

Pierwszym, podstawowym celem wszelkich prac powinno być niedopuszczenie do utraty rezultatów dotychczas wykonanej pracy. Oznacza to stworzenie kopii kompletu informacji merytorycznej zawartej w słowniku w postaci plików komputerowych zapisanych na nośniku cyfrowym. Przez komplet informacji rozumieć należy wszelkie elementy tekstu słownika niosące ze sobą jakiegokolwiek znaczenia, począwszy od struktury i zawartości poszczególnych elementów artykułów hasłowych, a na pełnej informacji typograficznej kończąc.

Dobrze byłoby (nie jest to wymóg bezwzględny), aby postać, w jakiej te informacje są zapamiętane, była w miarę możliwości standardowa i nadawała się do łatwego przetwarzania komputerowego, najlepiej za pomocą istniejących typowych i powszechnie używanych narzędzi programistycznych,

Wydaje się, że do realizacji tego celu, czyli zapisu danych, najodpowiedniejszy byłby język XML. Należy rozważyć również celowość dostosowania użytego formatu do zaleceń TEI.

Dalszą sprawą będzie udostępnianie informacji zawartych w słowniku. Tutaj istnieje wiele różnych możliwości. Najprostszy wariant to udostępnianie poszczególnych

tomów w postaci plików, zapisanych w formacie pdf z tak zwanymi zakładkami, zawierających tekst w kształcie zbliżonym do wersji drukowanej.

Zakres prac do wykonania dla różnych tomów Słownika nie będzie jednakowy – w zależności od tego, których tomów będzie dotyczył. Tomy te można wstępnie podzielić na następujące cztery kategorie: ukończone i istniejące tylko w formie papierowej, ukończone i istniejące zarówno w formie papierowej jak też w wersji elektronicznej, właśnie przygotowywane oraz planowane.

Ponieważ większość dotychczas wydanych tomów dostępna jest tylko na papierze, na wstępie konieczne jest wykonanie kopii poszczególnych stron w postaci graficznej (zeskanowanie), a następnie przetworzenie uzyskanego obrazu programami rozpoznawania pisma (OCR). Programy tego typu zwykle wymagają ‘przyuczenia’ do rozpoznawania konkretnych tekstów. Może się okazać, że proces uczenia się programu trzeba będzie przeprowadzić więcej niż raz, np. z powodu przejścia z tradycyjnej techniki druku na technikę fotoskładu. Uzyskane pliki tekstowe będą musiały zostać poddane wnikliwej korekcie i konwersji do postaci ostatecznej.

Dla tomów, dla których istnieją pliki komputerowe zawierające zasadniczy tekst w formacie Cyfrosetu, konieczne jest opracowanie odpowiedniego programu konwertującego np. na XML. Podobnie dla tomu przygotowanego za pomocą systemu KOMBI konieczne jest opracowanie konwertera.

Dla tomów planowanych możliwe są co najmniej dwa warianty postępowania.

W wariantie pierwszym kolejne tomy byłyby przygotowywane i składane aktualnie używanym systemem fotoskładu. Uzyskane tym sposobem pliki byłyby konwertowane do postaci odpowiedniej dla słownika elektronicznego.

W wariantie drugim tomy słownika byłyby przygotowywane w postaci plików tekstowych z wykorzystaniem powszechnie dostępnych, bezpłatnych narzędzi programistycznych [3, 8] w ustalonym formacie, który można bez większych problemów konwertować zarówno do postaci akceptowanej przez nowoczesne maszyny poligraficzne, jak też do postaci przydatnej dla potrzeb słownika elektronicznego. Zapewne możliwe byłoby redagowanie słownika od razu w postaci plików XML-owych, czyli w praktyce przygotowywanie go od razu w wersji elektronicznej.

Każdy z tych wariantów ma oczywiście swoje zalety i wady. Warto zauważyć, że wszystkie prace dotyczące tomów już wydanych, tomu przygotowywanego obecnie oraz tomów przyszłych w wariantie pierwszym mogą być realizowane niezależnie od pracy Pracowni Słownika Polszczyzny XVI w. Jedyne warunki to zachowywanie i udostępnianie wszelkich plików komputerowych z materiałami poszczególnych tomów.

Wnioski

Idea elektronicznej wersji Słownika polszczyzny XVI w. nie jest nowa. Myślał o niej już profesor Woronczak. Niedawno wznowiono dyskusje robocze. Wstępne rozmowy z profesorem Pełowskim prowadzili prof. Saloni, członek Komitetu Redakcyjnego, i prof. Bień.

Prace praktyczne dotyczące zagadnień związanych z elektroniczną wersją Słownika polszczyzny XVI wieku dopiero się rozpoczęły i mają charakter eksperymentów sprawdzających trafność naszych przewidywań.

Uczestnik seminarium magisterskiego *Narzędzia i metody przetwarzania tekstów* odbywającego się w Instytucie Informatyki UW pan Jacek Fedoryński wykonał wstępną wersję konwertera kodu Cyfrosetu, którą wstępnie przetestował na części materiału tomu XXIX. Uzyskane wyniki są bardzo obiecujące.

Używany do przygotowywania kolejnych tomów Słownika system KOMBI pozwala na zapisanie plików słownika w specyficznym dla siebie formacie znacznikowym. Znaczniki te, co prawda, nie opisują struktury artykułu hasłowego, tylko sposób jego złożenia, ale wstępna analiza tego formatu pokazuje, że możliwe jest dokonanie konwersji tak zapisanych plików do wygodniejszej dla naszych celów postaci. To, co stosunkowo łatwo dałoby się uzyskać, to wspomniana już wcześniej kopia elektroniczna w formacie pdf.

Niestety prace związane z tomami istniejącymi wyłącznie na papierze wymagają znacznych nakładów finansowych. Mamy jednak nadzieję, że trudność tę uda się w przyszłości pokonać.

Cały czas prowadzone są rozmowy i konsultacje w gronie osób zainteresowanych projektem. Mamy nadzieję, że będą one zbliżać wszystkich biorących w nich udział do ostatecznego celu, jakim jest elektroniczna wersja Słownika polszczyzny XVI wieku.

Summary

The article is concerned with the project to develop an electronic (digital) version of the well known ‘Dictionary of the 16th century Polish’.

In the first part of the text we discuss advantages of the digital versions of the dictionaries. We mention some interesting examples concerning both Polish and other languages dictionaries.

In the second part we focus on the practical aspects of digitalization of the specific dictionary, i.e. ‘Dictionary of the 16th century Polish’.

Bibliografia

- [1] Bańko M., (red.). 2000. *Inny słownik języka polskiego PWN*. Wydawnictwo Naukowe PWN, Warszawa.
- [2] Bień J.S., Linde-Usiekniewicz J. 1998. Elektroniczne słowniki języka polskiego. *Postscriptum* nr 24–25 (zima ’97 – wiosna ’98), s. 17–25.
- [3] Bień J.S. 2002. Gnu Emacs 21 i L^AT_EX2 ϵ – piszemy artykuł naukowy. In *Proceedings of the XIII European T_EX Conference, April 29–May 3, 2002, Bachotek, Poland*, pp 105–111.
- [4] Doroszewski W. 1958–1969. *Słownik języka polskiego PAN* pod red. W. Doroszewskiego. Wiedza Powszechna – PWN 1958–1969 (Przedrukowany przez PWN w 1997 r.), przedruk elektroniczny – ISBN 83-01-12321-4.

- [5] Fournier J. 2001. New Directions in Middle High German Lexicography: Dictionaries Interlinked Electronically. *Literary and Linguistic Computing*, Vol. 16, No. 1, pp 99–111.
- [6] Głowińska K., Woliński M. 2000. Angielsko-polski słownik elektroniczny XeLDA. *Acta Universitatis Nicolai Copernici. Studia Slavica 5, zeszyt 343*, s. 119–124
- [7] Grant W., Murison D. (eds.). 1931–1976. *The Scottish National Dictionary*. Scottish National Dictionary Association, Edinburgh.
- [8] Knuth D.D. 2000. *Computers & Typesetting Volumes A-E*. Addison-Wesley 1984–1986. Dostępne również jako „Millenium Edition”, Addison-Wesley.
- [9] *Komputerowy słownik języka polskiego*. 1996. Redaktor prowadzący J. Linde-Usiekiewicz. Wydawnictwo Naukowe PWN, Warszawa. ISBN 83-01-12232-3.
- [10] *Komputerowy słownik języka polskiego*. Edycja 1998, 1998. Wydawnictwo Naukowe PWN, Warszawa. ISBN 83-01-12504-7.
- [11] Saloni Z. 1997. Słownik Doroszewskiego w wersji elektronicznej. *Poradnik Językowy*, 1997, z. 5, s. 64–65.
- [12] *Słownik języka polskiego PWN* pod red. M. Szymczaka, t. I–III, Warszawa 1995, wydanie I.
- [13] *Słownik polszczyzny XVI wieku*. Tom I, A-Bany. 1966. Instytut Badań Literackich PAN, Zakład Narodowy im. Ossolińskich, Wrocław-Warszawa-Kraków (i tomy następne).
- [14] *Słownik wyrazów obcych*. Wydanie nowe, PWN, Warszawa 1995.
- [15] Sulimierski F., Chlebowski B., Waleski W. (eds.). 1880–1902. *Słownik Geograficzny Królestwa Polskiego*. Electronic facsimile edition, conceived and edited by Rafał T. Prinke, with additional materials by W.F. Hoffman. Polish Genealogical Society of America, Chicago, USA, 2002.