

Krzysztof NAROJCZYK
Uniwersytet Warmińsko-Mazurski w Olsztynie
Instytut Historii i Stosunków Międzynarodowych

**Nowe perspektywy elektronicznych wydań materiałów źródłowych.
Uwagi na marginesie cyfrowej edycji *Herbarza polskiego*
Adama Bonieckiego**

Elektroniczne publikacje wydawnictw źródłowych mają już u nas kilkuletnią tradycję. Pionierskie na gruncie polskim zasługi ma w tym względzie Biblioteka Kórnicka PAN, która w 1995 roku wydała na CD-ROM-ie bardzo starannie opracowane materiały historyczno-genealogiczne do dziejów szlachty wielkopolskiej XV–XX w., znane jako Teki Dworzaczka (druga wersja w roku 1997)¹. Stopniowo problematykę elektronicznych publikacji źródłowych podjęły także inne ośrodki naukowe, co zaowocowało kilkoma bardzo interesującymi wydawnictwami na CD-ROM-ach². Powszechnie podkreśla się bardzo wysoką użyteczność badawczą tego typu edycji źródłowych, jednakże podstawową trudnością w ich rozwoju jest ogromna pracochłonność konwersji tradycyjnych materiałów do postaci elektronicznej. W zależności od charakteru źródła, zakładanego celu publikacji oraz dostępnych środków technicznych tworzy się różnego rodzaju elektroniczne kopie dokumentów oryginalnych. Najprostszą ich formą jest uzyskiwana drogą digitalizacji kopia graficzna źródła (faksymile). Metoda ta jest jedyną możliwą do zastosowania w odniesieniu do materiałów kartograficznych, ikonograficznych itp.³ W przypadku dokumentów tekstowych możliwe jest odczytanie tekstu oryginalnego i wprowadzenie go do komputera w postaci tzw. tekstu elektronicznego (e–textu). Wydaje się, że rola tego typu edycji będzie stale wzrastała, tradycyjne wydania „papierowe” materiałów źródłowych osiągnęły bowiem już często taki stopień nasycenia adnotacjami, że dalsze ich wzbogacanie groziłoby utratą czytelności źródła. Równocześnie wyczerpał się już praktycznie katalog możliwości adnotowania w druku szczególnych cech tekstu (liczba stosowanych krojów czcionek i ich atrybutów nie może przekraczać możliwości percepcyjnych czytelnika). Edycje elektroniczne wolne są od tych ograniczeń, umożliwiając praktycznie nieograniczony zakres adnotowania, i to na wielu poziomach równocześnie. W ten sposób możliwe jest wielowarstwowe odczytywanie tego samego źródła w zależności od zainteresowań badawczych czytelnika, np. w badaniach historycznych, literackich lub językoznawczych.

¹ *Teki Dworzaczka. Materiały historyczno-genealogiczne do dziejów szlachty wielkopolskiej XV–XX w.*, CD # 1, v. 1.2.0 (1997) dla Windows.

² Polskie wydawnictwa elektroniczne ze wszystkich dziedzin rejestruje *Bibliografia Dokumentów Elektronicznych*, wkładka do „Przewodnika Bibliograficznego”, 2001, nr 27, nr 1 styczeń–czerwiec 2001; nr 50, nr 2 lipiec–grudzień 2001; 2002, nr 27, nr 1 styczeń–czerwiec 2002; nr 27, nr 2 lipiec–grudzień 2002.

³ Zob. np.: J. McGann, *The Rossetti Archive and Image-Based Electronic Editing*, 1995, <<http://www.engl.virginia.edu/courses/enwr101/f96/38/imagebas.htm>>, maj 2003.

Na rynku elektronicznych publikacji źródłowych poczesne miejsce zajmują cyfrowe opracowania materiałów wydanych już wcześniej w tradycyjnej technologii „papierowej”. Do grupy tej zalicza się udostępniona na CD-ROM-ie w 2002 r. elektroniczna wersja *Herbarza polskiego*, wydanego przez Adama Bonieckiego w latach 1899–1913⁴. Dzieło to nie jest klasycznym herbarzem, chociaż zawiera graficzną prezentację wymienianych w toku narracji herbów. Jego zasadniczą treść stanowią informacje o historii rodzin szlacheckich i osób do nich należących. Jest to więc raczej genealogia szlachty polskiej. Zamierzeniu temu podporządkowany jest także układ haseł, uporządkowanych według kolejności alfabetycznej nazwisk, a nie herbów. Każde hasło zawiera artykuł na temat historii określonej rodziny. Bazą źródłową *Herbarza* był obszerny zestaw dokumentów drukowanych i rękopiśmiennych, starannie zweryfikowanych przez Bonieckiego pod kątem wiarygodności i dokładności opisów poszczególnych osób oraz ich powiązań rodzinnych, a także pełnionych przez nie urzędów i funkcji publicznych. Według pierwotnego założenia, *Herbarz* miał w części pierwszej zawierać historię poszczególnych rodzin szlacheckich dawnej Rzeczypospolitej, w części drugiej zaś spisy urzędników w ujęciu chronologicznym. Niestety tego ambitnego projektu nie udało się zrealizować. Przed śmiercią Boniecki zdołał opublikować trzynaście tomów części I — do hasła *Lasocki*. Kolejne trzy tomy — do hasła *Makomascy* — wydał jego współpracownik Artur Reiski. Pomimo że nie został ukończony, *Herbarz Polski* ze względu na bardzo rzetelną bazę źródłową stanowi do dziś podstawowe źródło do badań nad genealogią i heraldyką szlachty polskiej. Wszystkich badaczy zajmujących się tą problematyką niewątpliwie ucieszył zatem fakt opublikowania jego wersji cyfrowej. Podstawą tej edycji było pierwsze wydanie, w którym poprawiono ewidentne błędy drukarskie i ujednolicono konwencje wydawnicze. Zasadniczą zmianą było wprowadzenie treści erraty bezpośrednio do tekstu oryginalnego. W efekcie np. tom *Uzupełnień i sprostowań* rozplątał się w tomie I i II. Zabieg ten w niczym nie naruszył warstwy informacyjnej dzieła, a podyktowany był przede wszystkim chęcią poprawienia funkcjonalności w dostępie do danych.

Międzynarodowym standardem elektronicznego kodowania i wymiany dokumentów (norma ISO-8879) jest od 1986 r. metajęzyk SGML (Standard Generalized Markup Language). Jego faktyczne upowszechnienie nastąpiło wraz z pojawieniem się w internecie technologii World Wide Web, która wykorzystywała początkowo najbardziej znaną aplikację SGML — język HTML (Hyper Text Markup Language). Jako platforma programowa edycji posłużył Extensible Markup Language (XML) — otwarty standard (zgodny z SGML) języka adnotowania dla sieci WWW, który stopniowo zastępuje dotychczas stosowany HTML. Wykorzystano specyfikację Extensible Style Language (XSL) tego języka, służącą głównie do tworzenia arkuszy stylów prezentacyjnych. Wybór tego standardu jest w pełni uzasadniony zarówno względami technologicznymi, jak i funkcjonalnymi — każdy internauta intuicyjnie potrafi obsługiwać napisane w nim aplikacje. Warunkiem uruchomienia programu jest zainstalowanie na komputerze dowolnej przeglądarki WWW, obsługującej standard plików XML/XSL, np. MS Internet Explorer 6.0, Netscape Navigator 7.0, Mozilla 1.0 (lub nowsze wersje tych programów). Po włożeniu płyty CD-ROM do czytnika program uruchamia się automatycznie i na wstępie oferuje użytkownikowi wybór języka interfejsu: polski lub angielski. Istotne jest to, że aplikacja nie instaluje w systemie żadnych swoich składników, które potem wymagać by mogły deinstalacji. Po wybraniu języka zgłasza się ekran początkowy, zawierający kartę tytułową oraz hipertekstowe odnośniki do poszczególnych opcji programu. W wydawnictwach tradycyjnych adnotacje metatekstowe, opisując poszczególne fragmenty tekstu właściwego, ułatwiają alinearne dostępowanie do jego warstwy informacyjnej, aczkolwiek sam tekst pozostaje nadal strukturą linearną. W edycjach elektronicznych funkcję tradycyjnych odsyłaczy (ang. *reference*) spełniają faktyczne połączenia (ang. *link*), umożliwiające automatyczne przechodzenie pomiędzy różnymi fragmentami informacji. W ten sposób powstaje tekst o zupełnie nowej — alinearnej, drzewiastej strukturze, którą tworzą zagnieżdżone, uporządkowane hierarchicznie elementy-objekty, takie jak rozdziały, akapity, przypisy, ilustracje, aneksy, cytaty itd. Autor elektronicznej wersji *Herbarza* w pełni wykorzystał technologię hipertekstu. Wybranie jednego z linków-odsyłaczy dostępnych na ekranie powodu-

⁴ Adam Boniecki, *Herbarz Polski*, wydanie 2 elektroniczne, opr. Marek Jerzy Minakowski, CD-ROM, version 1.0.1, Kraków 2002.

je przejście do następnego ekranu–strony, z którego z kolei można przejść do kolejnych itd. itd. Powstaje w ten sposób, znana wszystkim internautom, krokowa struktura wyświetlanych stron, po której można poruszać się, przechodząc o jeden ekran wstecz lub o jeden do przodu. Niestety zastosowanie wyboru ekranu z rozwijalnej listy (dostępnej w opcjach *Wstecz do:* oraz *Dalej do:* każdej przeglądarki internetowej) jest mało praktyczne, ekrany nie mają bowiem własnych unikalnych nazw. Zawsze jednakże można przerwać nawigację po jednym kroku i powrócić do strony startowej programu, wybierając stosowny odnośnik dostępny na każdej wyświetlanej stronie. Podobnie jak ma to miejsce w wielu publikacjach tego typu, użytkownik ma możliwość wyświetlania na ekranie monitora wiernych kopii graficznych wszystkich stron źródła oryginalnego. Dostęp do nich realizowany jest sekwencyjnie poprzez kliknięcie myszką na odpowiednim odnośniku do określonego tomu i następnie do wybranej strony, umieszczonymi w osobnym polu z lewej strony ekranu. Możliwe jest tam także wybranie konkretnej rodziny szlacheckiej i natychmiastowe przejście do e–tekstu, zawierającego informacje o niej. Jeżeli w tomie *Uzupełnień i sprostowań* znajduje się korekta informacji na temat jakiejś rodziny, jest to sygnalizowane odpowiednim odnośnikiem, umożliwiającym natychmiastowe przejście do tej informacji. Zastosowana do skanowania rozdzielczość 483 x 700 pikseli pozwala bezproblemowo odczytać wszystkie karty, jednakże stosunkowo mała czcionka oraz pewna nieostrość konturów liter powoduje, że dłuższa praca z tym materiałem byłaby niewątpliwie męcząca dla wzroku. Istnieje naturalnie możliwość wydrukowania wszystkich oglądanych kart i uzyskania fizycznej kopii *Herbarza*.

Jak już wspomniano, faksymile, w przeciwieństwie do tekstu elektronicznego, można jedynie oglądać. Nie poddają się one przede wszystkim zautomatyzowanym operacjom przeszukująco–selekcyjnym. Dopiero zatem odczytanie i przekształcenie oryginalnego tekstu do postaci e–tekstu oraz adnotowanie (otagowanie) go pozwala w pełni wykorzystać wszystkie możliwości technologii cyfrowej.

Tekst w postaci elektronicznej może występować bądź jako tzw. tekst czysty (ang. *plain*), bądź jako tekst adnotowany (ang. *marked up*)⁵. W pierwszym wypadku nie zawiera on, poza elementarnymi wyróżnieniami tytułów i akapitów, praktycznie żadnego klucza semantycznego. Mamy do czynienia jedynie ze strumieniem symboli pisarskich i interpunkcyjnych oraz znaków końca wiersza lub akapitu. Nie zawiera on zatem żadnej wiedzy pozaźródłowej. Zadaniem wydawcy jest jedynie poprawne odczytanie i przekształcenie do postaci elektronicznej oryginalnego tekstu źródłowego. Tego typu wydania nie nastęrczają dzisiaj większych problemów technicznych, zwłaszcza w odniesieniu do materiałów dających się opracowywać oprogramowaniem automatycznie odczytującym pismo. Badacz uzyskuje w ten sposób doskonały materiał, który oprócz reprezentowania wiedzy źródłowej daje się opracowywać z zastosowaniem metod analizy ilościowej, konkordancji i kolokacji. Jednakże pełne odczytanie źródła w każdym wypadku wymaga możliwie szerokiego włączenia wiedzy pozaźródłowej. Tekst elektroniczny bezproblemowo umożliwia stosowanie wszystkich wypracowanych dotychczas w krytycznym edytorstwie źródłowym mechanizmów alinearnego docierania do określonych fragmentów samego tekstu, jak i informacji pozaźródłowej (spisy treści, indeksy, konkordancje, przypisy). Informacja o niej, nazywana metatekstem, zawarta jest w adnotacjach. W edycjach elektronicznych wprowadzenia metainformacji do tekstu dokonuje się przy pomocy znaczników (ang. *tag*), które mogą być następnie interpretowane przez oprogramowanie komputerowe w zależności od zapotrzebowania czytelnika. Standardowo otagowanie tekstu jest procesem bardzo pracochłonnym i złożonym, wymagającym od wydawcy najwyższych kwalifikacji w rozumieniu zarówno kompetencji merytorycznych (wszechstronna znajomość przedmiotu badań), jak i kompetencji badawczych (zespół umiejętności nabywanych w wyniku doświadczenia i praktyki). Dlatego wielkie edycje tego rodzaju wymagają pracy wysoko wykwalifikowanego zespołu edytorskiego i należą jeszcze do rzadkości ze względu na trudną do przełamania barierę kosztów.

⁵ Interesujące uwagi o problemach adnotowania tekstu formułują: C. M. Sperberg–McQueen, C. Huitfeldt, A. Renear, *Meaning and interpretation of markup*, 2000, <<http://www.w3.org/People/cmsmcq/2000/mim.html>>, maj 2003.; C. M. Sperberg–McQueen, D. Dubin, C. Huitfeldt, A. Renear, *Drawing inferences on the basis of markup*, 2002, <<http://www.w3.org/People/cmsmcq/2002/EML2002Sper0518.final>>, maj 2003.

Stosuje się dwa podstawowe rodzaje adnotowania tekstu. Rodzaj pierwszy, nazywany proceduralnym lub preskryptywnym, określa jedynie wygląd tekstu, nic nie mówiąc o jego wartości semantycznej. Wskazuje, jakie operacje na tekście należy wykonać, aby nadać mu pożądaną formę typograficzną (np. krój i wielkość czcionki oraz jej atrybuty, wcięcia akapitów itd.). Adnotowanie tego typu wykorzystywane jest np. w języku HTML. Drugi rodzaj adnotowania — określane jako deklaratywny lub deskryptywny — wychodzi z zupełnie odmiennego założenia logicznego. Tagi określają tu nie wygląd, a wskazaną przez wydawcę wartość semantyczną oznaczonego fragmentu tekstu (np. tytuł, podtytuł, tekst zasadniczy, przypis, akapit, aneks itd., ale też np. imię, nazwisko, miejscowość, urząd, godność, datę itd.). Ten sam fragment może być oznaczony praktycznie kilkunastoma (kilkudziesięcioma) znacznikami, nadającymi mu różną wartość semantyczną. Taki tryb adnotowania stwarza ogromne możliwości manipulowania zidentyfikowanymi przez wydawcę elementami tekstu, poprzez np. wyszukiwanie, zestawianie, porównywanie, łączenie w określone zespoły, agregowanie, indeksowanie, dokonywanie rozmaitych transformacji i przekształceń w oparciu o dowolny, wcześniej zaprojektowany algorytm. Możliwe jest więc automatyczne generowanie baz danych, czyli zamknięcie ściśle wyselekcjonowanej pod określonym kątem części wiedzy źródłowej w ramach sformalizowanej struktury⁶. Jak już wspomniano, platformą programową aplikacji jest język XML, który wykorzystuje adnotowanie deklaratywne. Stworzyło to możliwość zaoferowania użytkownikowi elektronicznej wersji *Herbarza* podstawowego zestawu standardowych operacji na tekście. Tekst ów został odczytany i rozpoznany oprogramowaniem OCR oraz otagowany, co zamieniło go w pewnym sensie w tekstową bazę danych. To z kolei umożliwiło poindeksowanie tej bazy w oparciu o kilka najważniejszych pod względem informacyjnym kluczy: nazwiska, miejscowości, herby, urzędy i godności ziemskie. Standardowe indeksy umożliwiają natychmiastową lokalizację i przejście do wybranej informacji. Jednakże autor aplikacji poszedł zdecydowanie dalej. Wykorzystując mechanizmy baz danych, wykonał on, niejako za użytkownika, kilka podstawowych operacji sortująco-agregujących. Podniosły one na tyle walory informacyjne *Herbarza*, że śmiało można mówić o radykalnej zmianie jakościowej w wykorzystywaniu tego dzieła do badań naukowych.

Podstawowy z punktu widzenia pierwotnych funkcji *Herbarza* jest indeks nazwisk. Tu jednak czeka użytkownika pewna niespodzianka. Niemożliwe jest, czego należałoby oczekiwać, bezpośrednio wskazanie i przeskok do poszukiwanego nazwiska–hasła w oparciu o alfabetyczną listę. Dostępne są natomiast, wspomniane wyżej, efekty agregacji dokonanych przez autora. Wskazanie odnośnika opisanego jedną z liter alfabetu (od a do ż) powoduje wyświetlenie ekranu z alfabetyczną (sortowaną według alfabetu angielskiego) listą wszystkich odnotowanych w dziele nazwisk, zaczynających się od tejże litery. Nie jest ona tożsama z alfabetycznym zestawieniem haseł–artykułów. Obok znajdują się linki do wszystkich kolejnych haseł *Herbarza*, w których dane nazwisko jest wymieniane. Zatem poszukując informacji o dowolnej rodzinie lub konkretnej osobie, wskazujemy odpowiedni link i uzyskujemy dostęp do strony z listą wszystkich nazwisk zaczynających się od tych liter. Na wielu stronach przewijanie okienka w celu odnalezienia poszukiwanej rodziny jest mało wygodne, np. na literę K zaczyna się 3718 nazwisk. Można jednak posłużyć się, standardowym dla wielu aplikacji, mechanizmem wyszukiwania określonej frazy na aktywnej stronie: *Znajdź na tej stronie... Ctrl+F*. W ten sposób stosunkowo łatwo lokalizujemy na liście poszukiwane nazwisko. Obok umieszczone są alfabetycznie linki do haseł–artykułów *Herbarza*, gdzie jest ono przynajmniej raz wzmiankowane. Często jest ich kilkadziesiąt. Ręczne sporządzenie takich zestawień wymagałoby za każdym razem uważnego przewertowania wszystkich 17 tomów *Herbarza* i wynotowania potrzebnych danych, co poza jednostkowymi przypadkami jest mało realne. Wybranie dowolnego hasła powoduje przejście do standardowej strony, zawierającej pełne elektroniczne opracowanie wybranego artykułu. To, co otrzymuje badacz, jest naprawdę imponujące. U góry ekranu znajdują się podstawowe informacje o lokalizacji hasła w obrębie dzieła (tom, strona, ew. informacja o uzupełnieniach), linki nawigacyjne umożliwiające skok o jedno hasło do przodu lub wstecz oraz link do zeskanowanej strony oryginału. Następnie prezentowany jest pełny tekst artykułu. Można go automatycznie przeszukiwać

⁶ Zagadnienia związane z komputerową analizą źródeł historycznych szczegółowo omawia Rafał T. Prinke, *Fontes ex machina. Komputerowa analiza źródeł historycznych*, Poznań 2000, passim.

oraz kopiować w całości lub w części do schowka Windows, co daje możliwość użycia do dalszych poszukiwań dowolnych komputerowych narzędzi i programów analitycznych. Dalej generowane jest w formie tabelarycznej zestawienie wszystkich osób wzmiankowanych w tekście. Poszczególne kolumny zawierają kolejno (o ile informacje są dostępne) numer kolejnej osoby, adekwatny fragment tekstu oryginalnego gdzie jest ona wymieniana, płeć, imię/nazwisko, pełnione urzędy i godności oraz datację. Dodatkowo autor aplikacji zastosował kolorowe wyróżnienia wierszowe, oznaczając różowym tłem małżonków członków rodziny, tłem niebieskim zaś osoby jedynie wzmiankowane (nie spokrewnione). Układ taki znakomicie ułatwia śledzenie losów rodziny i odnajdywanie konkretnych osób. Poniżej tabeli kolejno umieszczone są alfabetyczne zestawienia:

- nazwisk małżonków danej rodziny,
- nazwisk osób jedynie wzmiankowanych,
- urzędów i godności dzierzonych przez osoby wymienione w artykule,
- rozkładu geograficznego wymienionych urzędów i godności,
- miejscowości wymienionych w artykule,
- źródeł cytowanych w artykule.

Oprócz powyższych elementów na ekranie dostępna jest mapka przedrozbiorowej Rzeczypospolitej z konturami wszystkich województw i zaznaczonym czerwonymi symbolami gwiazdki rozkładem geograficznym wymienionych urzędów i godności. Kliknięcie myszką na odpowiednim województwie powoduje przejście do kolejnej opcji programu, dostępnej także ze strony startowej.

Opcją tą jest, generowany w ujęciu geograficznym, szczegółowy spis wszystkich wzmiankowanych w *Herbarzu* urzędów i godności oraz osób je piastujących. Po wyborze określonego województwa wyświetlany jest ekran zawierający na wstępie listę miejscowości z tegoż województwa, w stosunku do których dało się przypisać choć jeden urząd lub godność. Obok każdej nazwy miejscowości znajduje się alfabetyczne zestawienie wszystkich skorelowanych z nią dostojenstw. Poniżej lista ta jest wzbogacona o szczegółowe wyliczenie osób wymienionych w *Herbarzu*, które kolejno pełniły owe urzędy wraz z podaniem dostępnej datacji. Kliknięcie myszką na dowolnej osobie powoduje przeskoczenie do odpowiedniego fragmentu tekstu elektronicznego, gdzie jest ona wymieniana.

Następną opcją programu dostępną ze strony startowej jest indeks miejscowości. Zawiera on, podobnie jak indeks nazwisk, zestaw linków oznaczonych kolejnymi literami alfabetu. Kliknięcie na którymś z nich powoduje wyświetlenie ekranu z listą wszystkich wymienionych w *Herbarzu* miejscowości, których nazwa zaczyna się od wskazanej litery. Obok nazw umieszczone są w formie odnośników zestawienia nazwisk skojarzonych z daną miejscowością. Umożliwiają one naturalnie, po kliknięciu na nie, natychmiastowy dostęp do odpowiedniego fragmentu tekstu elektronicznego.

Ostatnią opcją osiągalną ze strony startowej jest indeks herbów. Wybranie tej opcji wyświetla listę wszystkich 598 herbów wymienianych przez Bonieckiego. Przy każdym znajduje się zestawienie pieczętujących się nim rodzin szlacheckich. Kliknięcie na dowolnym nazwisku umożliwia oczywiście automatyczne przejście do adekwatnego hasła–artykułu.

Wydaje się, że omawiana publikacja jest zwiastunem swoistego przełomu w wydawaniu elektronicznych wersji źródeł historycznych, które były wcześniej opublikowane w tradycyjnej „papierowej” formie. Już samo zeskanowanie i przekształcenie oprogramowaniem OCR do postaci e–tekstu łącznie 6504 stron *Herbarza* musi budzić szacunek i uznanie. Jednakże operacje tego typu nie są niczym nowym. Także zakres udostępnionych opcji przeszukiwawczych nie odbiega od standardowych możliwości tekstu adnotowanego. Pytaniem zasadniczym było tylko, jak autor zdołał otagować tak obszerny materiał tekstowy. Okazało się, że zastosował on opracowany przez siebie algorytm automatycznego lub półautomatycznego tagowania tekstu. Może to zapowiadać prawdziwą rewolucję w technice adnotowania, skutkującą radykalnym zmniejszeniem niezbędnego nakładu czasu. Dość powiedzieć, że praca nad *Herbarzem* zajęła autorowi, według jego zapewnień, zaledwie około roku, co wszystkich znających to zagadnienie od strony praktycznej musi wprawić w zdumienie. Otagował on łącznie ponad milion słów, wyodrębnił nazwy 42 921 miejscowości, 26 370 nazwisk szlacheckich, 598 herbów. Oczywiście automatyczne adnotowanie nie pozbawione jest ewidentnych błędów i pomyłek, co związane jest najczęściej z trudnościami analizowania słów występujących w różnych for-

mach gramatycznych. Algorytm niekiedy błędnie rozpoznaje i klasyfikuje niektóre słowa, np. zaliczył do kategorii miejscowości określenia: Tyzenhauzenowi, Voluminach, Wzniesienia. Również rozpoznawanie poszczególnych osób nie zawsze jest precyzyjne, np. fragment tekstu dotyczący hetmana Jana Karola Chodkiewicza: „Jan Karol, często w aktach wprost Janem lub Karolem zwany...” automat zinterpretował tak, jakby wymieniano w nim trzy różne osoby (pozycja 113 do 115). Łącznie hetman wymieniany jest w opracowanym automatycznie tekście jako sześć różnych osób. Zatem nie ma mowy o czymś, co jest immanentną cechą poprawnego, zaawansowanego tagowania a mianowicie pełnej identyfikacji elementów tekstu i wyodrębnienia ich regularnej, ujednocionej formy. Badacz może wtedy wydać oprogramowaniu polecenie *Szukaj elementu X* i podać jego regularną formę. W odpowiedzi otrzyma wszystkie wystąpienia tego elementu w tekście, niezależnie od użytej pisowni czy formy gramatycznej. Możliwe jest także wtedy odnalezienie np. określonych osób lub miejscowości, nawet jeżeli w tekście nie są wymieniane *explicite*. Zastosowane w *Herbarzu* tagowanie automatyczne z oczywistych względów ogranicza się do rozróżniania kilku (kilkunastu) pojęć podstawowych typu: imię, nazwisko, miejscowość, herb, urząd, godność itp. Biorąc pod uwagę skalę przedsięwzięcia, autor uzyskał zadziwiająco poprawność identyfikacji tych elementów, aczkolwiek nadal jeszcze kilkadziesiąt tysięcy tagów wymaga korekty. Jednakże charakter popełnionych pomyłek jest najczęściej oczywisty i łatwy do uchwycenia, zawsze bowiem dostępny jest dla porównania odpowiedni fragment tekstu. Posługiwanie się elektronicznym *Herbarzem* nie rodzi zatem groźby popełnienia błędów przez badacza. Wręcz przeciwnie otrzymuje on do dyspozycji komfortowe środowisko pracy zaopatrzone w szereg narzędzi ułatwiających przeszukiwanie pełnotekstowej bazy danych, co stwarza całkowicie nowe, dotychczas praktycznie niedostępne, pole badawcze. Oczywiście kwestią do dyskusji jest taki lub inny kształt graficzny samej aplikacji oraz konkretne przyjęte w niej rozwiązania. Wydaje się, że w ostatecznej wersji *Herbarza* celowe byłoby wprowadzenie możliwości wyboru przez użytkownika dowolnego hasła wprost z alfabetycznej listy oraz mechanizmów pozwalających na samodzielne definiowanie zapytań do bazy. Niemniej jednak już obecnie trzeba uznać, że udostępniona badaczom wstępna wersja *Herbarza* stanowi bardzo ważny krok w rozwoju elektronicznych wydań polskich źródeł historycznych. Edycja ta jednoznacznie udowodniła, że pomimo całej złożoności form gramatycznych języka polskiego możliwe jest zaprojektowanie algorytmów tagowania automatycznego, które osiągają bardzo interesujące rezultaty praktyczne. Można zatem oczekiwać, że wraz z postępem badań nad takimi algorytmami realnie stanie się przełamanie bariery czasu i kosztów, która skutecznie blokowała dotychczas wydawców w rozwijaniu elektronicznych edycji materiałów źródłowych.