



Mgr Jacek Głębcki



Mgr inż. Grzegorz Pietruszewski

Poznań - Splendor Systemy Informacyjne

KONCEPCJA TWORZENIA WYKAZU CZASOPISM DO CELÓW PUNKTACJI JAKO REZULTAT ANALIZY ZAŁĄCZNIKA DO KOMUNIKATU MNISW Z 19 WRZEŚNIA 2012

Abstract

On Sept. 19, 2012 the Polish Ministry of Science and Higher Education published a list of journals as a basis for evaluation of scientific activity.

This paper presents:

- description of the data delivered by the Polish Ministry of Science and Higher Education,
- specification of the algorithms used to analysing the data,
- results of the analysis: the list of errors, lacks and ambiguities,
- discussion of the results and their sources,
- a concept of creating journal lists, based on ISSN-L key.

Streszczenie

Artykuł przedstawia opis danych zawartych w wykazie czasopism punktowanych opublikowanym przez MNiSW 19 września br., omawia środki informatyczne wykorzystane do analizy danych oraz podaje wykaz wykrytych błędów, braków i sytuacji niejednoznacznych. Autorzy wskazują na wady zastosowanej metody tworzenia wykazu oraz brak należytych procedur weryfikacyjnych, przedstawiając własną koncepcję wykorzystującą numer ISSN-L.

Wstęp

Zgodnie z par. 14 pkt. 2 Rozporządzenia Ministra Nauki i Szkolnictwa Wyższego z dnia 13 lipca 2012 r. w sprawie kryteriów i trybu przyznawania kategorii naukowej jednostkom naukowym [1] (dalej – Rozporządzenie) w dniu 19 września 2012 r. opublikowano komunikat w sprawie wykazu czasopism naukowych wraz z liczbą punktów przyznawanych za publikacje w tych czasopismach [2] (dalej Komunikat) wraz z załącznikiem zawierającym wykaz (dalej w treści publikacji Wykaz).

W myśl par. 14 pkt. 3 Rozporządzenia wykaz czasopism naukowych składa się z:

- 1) części A – zawierającej liczbę punktów za publikacje w czasopismach naukowych posiadających współczynnik wpływu Impact Factor (IF), znajdujących się w bazie Journal Citation Reports (JCR) [3] (dalej Lista A).
- 2) części B – zawierającej liczbę punktów za publikacje w czasopismach naukowych nieposiadających współczynnika wpływu Impact Factor (IF) [4] (dalej Lista B).
- 3) części C – zawierającej liczbę punktów za publikacje w czasopismach naukowych znajdujących się w bazie European Reference Index for the Humanities (ERIH) [5] (dalej Lista C).

Na serwerze internetowym Ministerstwa Nauki i Szkolnictwa Wyższego dostępna jest także część Z, niebędąca załącznikiem do Komunikatu, zawierająca wykaz tych czasopism naukowych zgłoszonych do oceny, którym przyznano zero punktów [6] (dalej Lista Z).

Działania zmierzające do automatyzacji gromadzenia, przetwarzania i udostępniania informacji o szkolnictwie wyższym były podejmowane jeszcze w latach 90. przez Komitet Badań Naukowych, czego efektem była skomputeryzowana wersja ankiety jednostek naukowych opracowywana i przetwarzana przez Centralny Ośrodek Badawczo-Rozwojowy Aparatury Badawczej i Dydaktycznej COBRABiD Oddział w Poznaniu, a wypełniana przez szkoły wyższe i instytuty badawcze.

W roku 2002 zadanie opracowywania programu komputerowego zostało przejęte przez Ośrodek Przetwarzania Informacji. Oprogramowanie dostosowywane na bieżąco do aktualnie obowiązujących przepisów było rozbudowywane także pod względem informatycznym.

Ważną zaletą była możliwość importowania danych z zautomatyzowanych systemów bibliograficznych, zarówno w plikach tekstowych o określonej strukturze, jak i w formacie xml. Ankieta jednostki OPI w ostatniej wersji udostępnionej w czerwcu 2010 uzyskała wysoki poziom i pozwalała na sprawną realizację zadań stawianych przez ministerstwo nauki przed uczelniami wyższymi w zakresie raportowania informacji o dorobku naukowym.

Istotnym usprawnieniem było zaimplementowanie w oprogramowaniu wykazu czasopism punktowanych, a także opracowanie mechanizmu automatycznej punktacji publikacji na podstawie identyczności tytułu czasopisma w dostarczanym opisie

bibliograficznym i tytułu czasopisma zawartym na wbudowanym wykazie czasopism punktowanych.

Wymagało to od projektantów lokalnych systemów bibliograficznych zapewnienia mechanizmu automatycznej wymiany tytułu (lub skrótu) czasopisma stosowanego w opisach bibliograficznych na tytuł obowiązujący w wykazie czasopism punktowanych. W przypadku niemożności zapewnienia takiej procedury na poziomie informatycznym, do dyspozycji osób opracowujących dane była opcje grupowej identyfikacji tytułów czasopism bezpośrednio ankiecie jednostki.

Wykaz czasopism punktowanych złożony jedynie z tytułu czasopisma i przypisanych czasopismu punktów nie był jednak rozwiązaniem wystarczającym do bezbłędnej identyfikacji źródła publikacji.

Uchwalona 18 marca 2011 r. ustawa o zmianie ustawy Prawo o szkolnictwie wyższym, ustawy o stopniach naukowych i tytule naukowym oraz o stopniach i tytule w zakresie sztuki oraz o zmianie niektórych innych ustaw dodała do ustawy Prawo o szkolnictwie wyższym art. 34a, który zobowiązuje ministra nauki i szkolnictwa wyższego do prowadzenia Systemu Informacji o Szkolnictwie Wyższym [7].

7 kwietnia 2011 ministerstwo nauki podjęło decyzję o dofinansowaniu projektu systemowego w ramach Priorytetu IV Programu Operacyjnego Kapitał Ludzki pt.: "Stworzenie systemu informacji o szkolnictwie wyższym" (Działanie 4.1 Wzmocnienie i rozwój potencjału dydaktycznego uczelni oraz zwiększenie liczby absolwentów kierunków o kluczowym znaczeniu dla gospodarki opartej na wiedzy, Poddziałanie 4.1.3 Wzmocnienie systemowych narzędzi zarządzania szkolnictwem wyższym) [8].

Zakres planowanego systemu o nazwie POL-on jest o wiele szerszy niż dotychczasowe funkcje realizowane przez oprogramowanie Ankieta Jednostki OPI i składa się z dwóch zadań:

1) opracowania projektu systemu informatycznego i jego wdrożenia, monitoringu procesu jakości kształcenia, zbierania i przechowywania danych, udostępniania publikacji naukowych

2) analiza publikacji naukowych i technologicznych, ocena dorobku naukowego [10].

Doceniając rolę, jaką ma pełnić nowoczesny system informatyczny i zdając sobie sprawę, że rzetelność, kompletność i terminowość przekazywania danych przez uczelnie będą miały kluczowe znaczenie [11], autorzy artykułu dokonali analizy opublikowanego Wykazu w celu:

- zapewnienia możliwości przekazywania danych do systemu POL-on z baz bibliograficzno-bibliometrycznych tworzonych przez uczelnie i instytuty naukowe w systemie Expertus

- stwierdzenia, czy zawartość Wykazu pozwala na jednoznaczną identyfikację czasopisma i niezawodne przypisanie publikacjom naukowym liczby punktów zawartej na Wykazie

- w razie negatywnej odpowiedzi na poprzednie pytanie – zaproponowania właściwej metody opracowywania takich wykazów.

Material i metody

Wykorzystano Wykaz opublikowany w postaci plików pdf, zawierających Listę A, Listę B, Listę C oraz dodatkowo, niewchodzącą w skład Wykazu, ale dostępna na serwerze ministerstwa Listę Z.

Pliki pdf zawierają tabele, w których każdy wiersz składa się z czterech kolumn, zatytułowanych kolejno: lp. (czyli liczba porządkowa), tytuł czasopisma, nr ISSN, liczba punktów za publikację w czasopiśmie naukowym.

Dane zapisano jako tekst i przeprowadzono konwersję do postaci bazy danych przy użyciu operacji programistycznych.

Powstała baza danych zawierała rekordy składające się z następujących pól:

- kod listy (A, B, C, Z)
- numer pozycji na liście
- tytuł czasopisma
- ISSN
- liczba punktów

Baza danych liczyła łącznie 16525 rekordów, z czego na poszczególne listy przypadało:

- a) na listę A (dane z Listy A) : 10230 rekordów
- b) na listę B (dane z Listy B) : 1798 rekordów
- c) na listę C (dane z Listy C) : 4329 rekordów
- d) na listę Z (dane z Listy Z) : 168 rekordów

Liczba rekordów w każdej kategorii odpowiada liczbie pozycji na Wykazie, składających się na Wykaz poszczególnych list (Lista A, B, C) oraz Listy Z spoza Wykazu.

W celu weryfikacji zgodności zawartości otrzymanej bazy danych z danymi źródłowymi, utworzono zwrrotnie plik pdf i uzyskaną zawartość tekstową porównano z zawartością tekstową pliku źródłowego stosując standardowe narzędzia porównywania plików tekstowych dostępne w systemach operacyjnych.

Opracowano oprogramowanie pomocnicze zawierające funkcje pozwalające na:

a) porównywanie zawartości pól 'tytuł czasopisma' w celu zbadania identyczności tytułów w różnych rekordach z uwzględnianiem wielkości znaków w porównywanych ciągach (dalej procedura F1)

b) porównywanie zawartości pól 'tytuł czasopisma' w celu zbadania identyczności tytułów w różnych rekordach bez względu na wielkość znaków (dalej procedura F2)

c) wyszukiwaniu rekordów, które w polu tytuł czasopisma zawierają słowa, w których pominięto polskie lub obce znaki diakrytyczne (dalej procedura F3)

d) wyszukiwaniu rekordów, które w polu tytuł czasopisma zawierają słowa powsta-

le z innych słów na skutek permutacji liter (dalej procedura F4)

e) porównywanie zawartości pól 'ISSN' w celu zbadania identyczności numerów ISSN w różnych rekordach wraz z badaniem identyczności z wyłączeniem separatora (dalej procedura F5)

f) badanie poprawności struktury numerów ISSN (dalej procedura F6)

g) badanie poprawności numerów ISSN poprzez weryfikację cyfry kontrolnej (dalej procedura F7)

Wyżej wymienione funkcje zrealizowano wg poniższego opisu:

a) F1 - zastosowano standardową procedurę porównywania ciągów znakowych (niewymagającą szczegółowego omówienia)

b) F2 - stosowano funkcję F1 po wcześniejszym sprowadzeniu obu porównywanych ciągów tekstowych do małych liter.

c) F3 - stosowano funkcję F1 po wcześniejszym sprowadzeniu wielkich liter w obu porównywanych ciągach tekstowych do odpowiadających im małych liter i sprowadzeniu znaków diakrytycznych do odpowiadających im liter łańcuchów

d) F4 - każde słowo pola 'tytuł czasopisma' we wszystkich rekordach sprowadzano do małych liter, usuwano powtórzenia liter i powstały ciąg porządkowano alfabetycznie; następnie dla każdego słowa z pola 'tytuł czasopisma' przeszukiwano to samo pole we wszystkich pozostałych rekordach w celu znalezienia identycznych ciągów znakowych; z wyników wykluczano te rekordy, w których wystąpiła równość ciągów znakowych przed opisaną operacją (tj. usunięciem powtórzeń i permutacją); z uwagi na wielojęzyczność tytułów czasopism oraz możliwość przekształcenia do tego samego ciągu dwóch różnych, poprawnych słów, metodę traktowano jedynie pomocniczo jako materiał do dalszej weryfikacji.

Przykłady słów w istocie różnych, a identycznych w wybranej metodzie to:

- "de" oraz "ed" w tytułach *Academia-Revista Latinoamericana de Administracion* oraz *Bollettino di Geofisica Teorica ed Applicata*

- „and” oraz „dna” w tytułach *Abacus-A Journal of Accounting Finance and Business Studies* oraz *Mitochondrial DNA*

W celu ograniczenia podobnych przypadków zawężono przeszukiwany zbiór słów do podzbioru słów o liczbie liter większej niż 5.

e) F5 - do pola ISSN stosowana funkcję F2 w celu nierozróżniania małej i wielkiej litery x na ostatnim znaku

f) F6 - stosowano kontrolę długości numeru issn (9 znaków łącznie z separatorem) oraz kontrolę typu danych tj. dla kolejnych znaków sprawdzano, czy pierwsze cztery znaki są cyframi, kolejny znak jest myślnikiem, kolejne trzy cyframi i ostatnia cyfra jest cyfrą lub wielką literą X

g) F7 - zastosowano standardowy algorytm sprawdzania cyfry kontrolnej polegający na obliczaniu różnicy liczby 11 i sumy iloczynów siedmiu kolejnych cyfr ISSN

poczynając od cyfry pierwszej, następnie obliczeniu reszty z dzielenia otrzymanego wyniku przez 11, na zakończenie otrzymany wynik porównywano z z cyfrą kontrolną, która powinna być równa reszcie z dzielenia (dla reszty od zera do dziewięciu) lub wielkiej literze X - dla reszty równej 10.

Algorytm zaimplementowano na podstawie wytycznych zawartych w pozycji [12].

Metoda pozwala na wykrycie transpozycji znaków i pojedynczych błędów w numerze ISSN. Istnieją jednakże kategorie błędów, których wykryć się nie da, ponieważ nie skutkują zmianą cyfry kontrolnej obliczonej wg algorytmu.

Trywialnym przykładem poprawnych i wykorzystanych numerów ISSN różniących się na dwóch pozycjach jest para numerów: 0033-2097 i 1033-0097, należących odpowiednio do czasopisma *Przegląd Elektrotechniczny* [13] oraz *Editions: the new monthly Australian review of books* [14] (nie występuje w Wykazie).

Mimo że wiele jednoczesnych błędów może nie zmienić wartości znaku kontrolnego, zastosowanie procedury F7 pozwala na znalezienie najbardziej podstawowych błędów.

Wyniki

Po przeprowadzeniu automatycznej analizy i weryfikacji uzyskano następujące wyniki:

I.

1. Lista A i C - duplikat

pozycja 10190 na liście A (15 punktów) oraz 4304 na liście C (10 punktów)

Zeitschrift für Pädagogik

Procedura: F1, F5, F6

Komentarz: ISSN na obu listach 0044-3247, wizualnie identyczny, ale w jednym przypadku separatorem jest 'en dash'

2. Lista C

pozycja 331

Archaiologikon Deltion

Procedura: F7

Komentarz: ISSN na Liście C: 0572-622X (a poprawny to 0570-622X)

3. Lista C

pozycja 585

Blake/An Illustrated Quarterly

Procedura: F7

Komentarz: ISSN na Liście C: 0160-628x (a poprawny to 0160-628X – zgodnie z issn.org ostatni znak musi być cyfrą lub wielką literą X)

4. Lista C

pozycja 1257

Ethnologia Balkanica. Journal for Southeast European Anthropology. Zeitschrift für die Anthropologie Südosteuropas. Journal d'anthropologie sud-est européen

Procedura: F7

Komentarz: ISSN na Liście C: 1111-0411 (a poprawny to 0111-0411)

5. Lista C

pozycja 1271

Etnologické rozpravy / Ethnological Disputes

Procedura: F6

Komentarz: ISSN na Liście C: 1335–5074 [jako separator 'en dash'], a poprawny ISSN to 1335-5074 [separatorem powinien być 'hyphen-minus']

6. Lista C

pozycja 2261

Kunst + Architektur in der Schweiz

Procedura: F6

Komentarz: ISSN na Liście C: 1421-086x (a poprawny to 1421-086X)

7. Lista C

pozycja 2564

Mitteilungsblatt des Instituts für soziale Bewegungen

Procedura: F4, F6

Komentarz: ISSN na Liście C: 10173-2471 (a poprawny to 0173-2471)

ponadto błąd w tytule - powinno być Mitteilungsblatt des Instituts für soziale Bewegungen

8. Lista C

pozycja 2655

Naamkunde. Tijdschrift voor naamkunde in het Nederlands taalgebied

Procedura: F7, F1

Komentarz: ISSN na Liście C: 0167-5357 (a poprawny ISSN to 0167-5257);

na pozycji 2654 identyczny tytuł z prawidłowym ISSN, zatem musi być to duplikat; oba ocenione na 10 punktów

9. Lista C

pozycja 4150

Viking (Norsk arkeologisk arbok)

Procedura: F6

Komentarz: ISSN na Liście C: 0332-608x (a poprawny to: 0332-608X)

10. Lista C

pozycja 4222

World of Music

Procedura: F6

Komentarz: ISSN na Liście C: 0043–8774 [jako separator 'en dash'], a poprawny ISSN to 0043-8774 [jako separator powinien być 'hyphen-minus']

uwaga: na pozycji 4223 zapis World of Music (The) dotyczący tego samego czasopisma

11. Lista A

pozycja 8160

Phlebologie-Annales Vasculaires

uwaga: brak numeru ISSN, na Liście A podany jako ****_****

wymaga wyjaśnień

12. Lista C

pozycja 3438

Romance Studies Alternate Ed 12

uwaga: na Liście C - brak ISSN

wymaga wyjaśnień

13. Lista C

pozycja 3044

Przegląd Archeologiczny

Procedura: F4

14. Lista C

pozycja 3772

Studia z dziejów Rosji i Europy Środkowo-Wschodniej

Procedura: F4

15.

wiele innych przypadków typu „przegląd”, „przemysł”, które pomijamy

Procedura: F4

II.

Ponadto w poniższej tabeli zebrano przypadki identyczności tytułów przy różnych ISSN, często różnej punktacji. Wyniki otrzymano dzięki zastosowaniu przynajmniej jednej z procedur F1, F2, F3.

Wybrane przypadki omawiamy w części Dyskusja, pozostałe wymagają dalszej analizy.

lista	pozycja	tytuł	ISSN	punkty
A	8402	Poznan Studies in Contemporary Linguistics	1897-7499	15
C	3027	Poznan Studies in Contemporary Linguistics	0137-2459	10

C	2592	Modern Italy	1469-9877	10
C	2593	Modern Italy	1353-2944	12
C	2550	Millennium. Tijdschrift voor middeleeuwse studies.	1782-3404	10
C	2551	Millennium. Tijdschrift voor middeleeuwse studies.	0922-0887	12
A	7003	Logopedics, Phoniatics, Vocology	1401-5439	20
C	2422	Logopedics, Phoniatics, Vocology	1651-2022	12
C	2348	Lettre de Zuylen et du Pontet SINCE 2006: Cahiers Isabelle de Charriere / Belle de Zuylen Papers	0920-9468	10
C	2349	Lettre de Zuylen et du Pontet SINCE 2006: Cahiers Isabelle de Charriere / Belle de Zuylen Papers	1872-7832	10
A	6940	Learning Media and Technology	1743-9892	30
C	2324	Learning Media and Technology	1358-1651	12
A	6935	LEARNING & MEMORY	1072-0502	35
C	2321	Learning & Memory	0143-7534	14
A	6564	JOURNAL OF THE AUDIO ENGINEERING SOCIETY	1549-4950	25
C	2163	Journal of the Audio Engineering Society	0004-7554	10
A	6480	Journal of Studies on Alcohol and Drugs	1937-1888	30
C	2157	Journal of Studies on Alcohol and Drugs	0096-882X	14
A	5645	JOURNAL OF FORENSIC PSYCHIATRY & PSYCHOLOGY	1478-9949	25
C	2063	Journal of Forensic Psychiatry & Psychology	1478-9957	10
C	2016	Journal of Asian and African Studies	0021-9096	12
C	2017	Journal of Asian and African Studies	0387-2807	12
A	4033	IBERICA	1139-7241	15
C	1727	Iberica	0153-0364	10
A	3953	HISTORY WORKSHOP JOURNAL	1363-3554	40
C	1688	History Workshop Journal	0309-2984	14
C	1453	Freiburger Zeitschrift für Philosophie und Theologie	1422-4380	10
C	1454	Freiburger Zeitschrift für Philosophie und Theologie	0016-0725	10
A	3384	EYE	0950-222X	30
C	1343	Eye	0960-779X	10

C	1320	European Medieval Drama	1378-2274	10
C	1321	European Medieval Drama	1287-7484	12
A	3252	EUROPEAN JOURNAL OF PUBLIC HEALTH	1101-1262	35
C	1316	European Journal of Public Health	1210-7778	12
C	1145	Educational Philosophy and Theory	0013-1857	12
C	1146	Educational Philosophy and Theory	1469-5812	12
A	2811	EAST EUROPEAN POLITICS AND SOCIETIES	0888-3254	25
C	1116	East European Politics and Societies	1533-8371	12
C	996	Dacoromania	1582-4438	10
C	997	Dacoromania	0342-0299	10
C	923	Contemporary Theatre Review	1048-6801	14
C	924	Contemporary Theatre Review	1477-2264	10
A	2107	CLINICAL NEUROPSYCHOLOGIST	1385-4046	25
C	860	Clinical Neuropsychologist	0920-1637	12
A	1907	Chemistry Education Research and Practice	1109-4028	20
C	825	Chemistry Education Research and Practice	1756-1108	14
A	1859	CHAOS	1054-1500	40
C	820	CHAOS	0108-4453	10
A	1849	CEREBRAL CORTEX	1047-3211	40
C	817	Cerebral Cortex	1566-6816	14
C	613	Book History	1098-7371	14
C	614	Book History	1529-1499	12
C	518	Axiomathes	1572-8390	12
C	519	Axiomathes	1122-1151	12
C	506	Australasian Victorian Studies	1325-2585	10
C	507	Australasian Victorian Studies	1327-8746	10
C	473	AS Mediatijdschrift	1377-5197	10
C	474	AS mediatijdschrift	0773-5855	10
A	3080	EPILEPSIA	0013-9580	35
C	1205	Epilepsia	0356-598X	14
A	3624	Gender, Work and Organization	0968-6673	25
C	1475	Gender, Work and Organization	1468-0432	14
C	1582	Helikon	0017-999X	10
C	1583	Helikon	0017-9981	10
C	1490	Germanic Review (The)	0016-8890	10
C	1491	Germanic Review (The)	1930-6962	12
A	7739	Nordic Psychology	1901-2276	15
C	2737	Nordic Psychology	0029-1463	10

A	7737	Nordic Journal of Music Therapy	0803-9828	15
C	2735	Nordic Journal of Music Therapy	0809-8131	12
A	7555	NATURE REVIEWS NEUROSCIENCE	1471-0048	50
C	2674	Nature Reviews Neuroscience	1471-003X	14
C	2960	Philosophica	0379-8402	12
C	2961	Philosophica	0872-4784	10
B	1306	Rozprawy Komisji Językowej	0076-0390	5
B	1307	Rozprawy Komisji Językowej	0084-2990	4
C	3404	Rivista di Letteratura Italiana	1724-0638	10
C	3405	RIVISTA DI LETTERATURA ITALIANA	0392-825X	10
C	3396	Rivista di cultura classica e medioevale	1724-062X	10
C	3397	Rivista di cultura classica e medioevale	0035-6085	10
C	3200	Research in Drama Education	1470-112X	12
C	3201	Research in Drama Education	1356-9783	12
C	3106	Quaderni di storia	1722-5043	10
C	3107	Quaderni di Storia	0391-6936	10
C	3944	The Journal of the Society for American Music	1752-1963	12
C	3945	The Journal of the Society for American Music	1752-1971	12
A	9660	TALANTA	0039-9140	40
C	3863	Talanta	0165-2486	10
A	9560	STUDIA PSYCHOLOGICA	0039-3320	15
C	1459	Studia Psychologica	1642-2473	4
B	1392	Studia Europejskie	2080-3567	3
B	1393	STUDIA EUROPEJSKIE	1428-149X	8
A	215	ADDICTION RESEARCH & THEORY	1606-6359	25
C	59	Addiction Research & Theory	1476-7392	12
C	70	Aevum	1827-787X	10
C	71	Aevum	0001-9593	10
C	196	Anglo-Saxon England	1474-0532	14
C	197	Anglo-Saxon England	0263-6751	12
C	245	Anthropological Journal of European Cultures	0960-0604	12
C	246	Anthropological Journal of European Cultures	1755-2923	12
C	254	Anthropology Today	0268-540X	12
C	255	Anthropology today	1467-8322	12
C	268	Antropologi Indonesia	1693-167X	12

C	269	Antropologi Indonesia	1693-6086	12
A	913	ARCHIVE FOR MATHEMATICAL LOGIC	1432-0665	20
C	383	Archive for Mathematical Logic	0933-5846	12
C	506	Australasian Victorian Studies	1325-2585	10
C	507	Australasian Victorian Studies	1327-8746	10
C	518	Axiomathes	1572-8390	12
C	519	Axiomathes	1122-1151	12
A	9973	URBAN EDUCATION	0042-0859	25
C	4112	Urban Education	0143-9154	12

Dyskusja

Zastosowanie procedur kontrolnych ujawniło następujące błędy:

- wadliwe numery ISSN (błędy w strukturze)
- błędne numery ISSN (błędy sumy kontrolnej)
- duplikaty polegające na umieszczeniu tego samego czasopisma na różnych liściach z różną wartością punktową
- duplikaty czasopisma na tej samej liście
- inne usterki: literówki

Wszystkie powyższe błędy mają krytyczne znaczenie przy przetwarzaniu Wykazu w sposób automatyczny, a większość z nich także przy analizach nieautomatycznych (nieefektywnych z uwagi na wielkość badanego zbioru).

Analiza ujawnia także problemy, których – na bazie opublikowanych dokumentów – rozwiązać się nie da, choć można to było zrobić na etapie tworzenia bazy danych.

W części Wyniki w punkcie II zebrano pozycje o identycznych tytułach (bez uwzględniania wielkości znaków). Szczegółowego rozpoznania wymaga podwójna obecność tego samego czasopisma, gdy jeden zapis dotyczy wersji drukowanej (p-ISSN), a drugi wersji on-line (e-ISSN).

Przykład czasopisma zamieszczonego podwójnie (w pierwszym przypadku p-ISSN, w drugim e-ISSN):

A	8402	Poznan Studies in Contemporary Linguistics	1897-7499	15
C	3027	Poznan Studies in Contemporary Linguistics	0137-2459	10

Przykład różnych czasopism o tym samym tytule:

A	1859	CHAOS	1054-1500	40
C	820	CHAOS	0108-4453	10

Warto zauważyć, że tytuły mogą być zapisane identycznie lub różnić się wyłącznie wielkością znaków. W tym drugim przypadku najczęściej tytuł zapisany wielkimi literami pochodzi z listy A, a literami małymi z listy C. Ponadto taka sytuacja dotyczy także tytułów czasopism zagranicznych.

Na Wykazie występują także błędy w tytułach czasopism sugerujące identyczność tytułu z zupełnie innym czasopismem, spowodowane np. opuszczeniem istotnej części tytułu.

Przykład:

A	3252	EUROPEAN JOURNAL OF PUBLIC HEALTH	1101-1262	35
C	1316	European Journal of Public Health	1210-7778	12

W rzeczywistości czasopismo z pozycji 1316 Listy C tj. z ISSN 1210-7778 nosi tytuł: Central European Journal of Public Health.

Zagadnienie uzyskiwania baz danych łączących zasoby kilku innych baz źródłowych jest wspólne dla szerokiej klasy problemów przetwarzania danych i występuje często w praktyce codziennej.

Zawsze należy trafnie wybrać źródło danych pierwotnych, biorąc pod uwagę kompletność i wiarygodność danych, a do wszystkich danych pobieranych z baz dodatkowych, w tym z systemów zgłoszeniowych, stosować właściwe algorytmy identyfikacyjne i weryfikacyjne.

Zadania, w których w naturalny sposób można przyjąć jeden z elementów danych jako identyfikator należą do znacznie łatwiejszych. Pozostałe nie zawsze mogą być rozwiązane w sposób w pełni automatyczny.

Autorzy skonfrontowali zagadnienie będące przedmiotem analizy z problemami, które rozwiązywali przy okazji prac nad innymi systemami informatycznymi, skupiając się na następujących:

- a) system obsługi tworzenia wielotomowej encyklopedii "Encyclopaedia Britannica - edycja polska" [15]
- b) opracowanie publikacji zawierającej wykaz substancji biologicznie czynnych w środkach ochrony roślin [16]
- c) identyfikacji obiektów cyfrowych i wykorzystania zewnętrznych narzędzi linkujących w bazach danych tworzonych i udostępnianych w systemie Expertus [17]

W systemie obsługi tworzenia polskiej edycji Encyklopedii Britanniki oprócz konieczności opracowania polskiej wersji haseł pochodzących z oryginalnej edycji encyklopedii konieczne było uzupełnienie siatki haseł o takie hasła dotyczące Polski, które nie wystąpiły w edycji oryginalnej.

Rozmiar danych (ok. 100 000 haseł), brak naturalnego identyfikatora, możliwość występowania haseł jednobrzmiących pozwala zaliczyć do zadanie do problemów o złożoności nieporównywalnej z analizowanym.

W pracach nad opracowaniem pierwszego w Polsce zestawu informacji obejmującego wykaz substancji biologicznie czynnych i odpowiadających im klas chemicznych występujących w środkach ochrony roślin, głównym problemem było wyeliminowanie duplikatów wynikających z używania nazw synonimicznych i ustalenie jednoznacznej nazwy angielskiej, co pozwalało na skorzystanie z klasyfikacji stosowanej w źródłach zagranicznych w celu określenia przynależności do określonej grupy chemicznej.

W zagadnieniach dotyczących identyfikacji obiektów cyfrowych w systemie Expertus, związanych z wykorzystywaniem zewnętrznych narzędzi linkujących w sposób zapewniający tworzenie linków do pełnego tekstu dostępnego on-line bez względu na to, czy opis bibliograficzny dotyczy wersji print czy online, skutecznym rozwiązaniem okazało się wykorzystanie ISSN linkującego.

ISSN linkujący (ISSN-L), zgodnie z ISO 3297:2007 oraz PN-ISO 3297:2010 [18] został przewidziany właśnie jako rozwiązanie problemu wynikającego z istnienia różnych ISSN dla różnych form udostępniania czasopisma [19] i doskonale nadaje się do roli jednoznacznego identyfikatora w tworzeniu wykazów czasopism. Opis zastosowania ISSN-L znajduje się także w [20].

Tego typu uporządkowane podejście zostało zastosowane w rozwiązaniach zagranicznych, np. w szwajcarskim systemie SERVAL na Wydziale Biologii i Medycyny Uniwersytetu w Lozannie w roku 2008 [21].

Zastosowano następującą procedurę:

- jako dane pierwotne potraktowano zasoby udostępniane komercyjnie przez Międzynarodowe Centrum ISSN, traktując ISSN-L jako identyfikator podstawowy
- pozostałe wersje ISSN posłużyły jako elementy pozwalające na lokalizację czasopisma w bazach danych niestosujących ISSN-L (np. Journal Citation Reports)

Przykład zastosowania ISSN-L w Cistin Current Research Information System In Norway [22] podany jest w wykazie [23].

Międzynarodowe Centrum ISSN udostępnia dane w różnych zakresach i formatach, takich jak:

- pełny rejestr
- dane pozwalające ustalić odpowiedniość ISSN – ISSN-L
- określanie poprawnego ISSN dla wskazanych czasopism

Dane dotyczące polskich czasopism przekazuje do międzynarodowego rejestru Biblioteka Narodowa będąca Narodowym Ośrodkiem ISSN.

Wnioski

1. Podstawą tworzenia Wykazu i pierwotnym źródłem danych powinny być dane

- pozyskane z Międzynarodowego Centrum ISSN.
2. Kluczem identyfikującym czasopismo powinien być ISSN-L, natomiast ISSN przypisane innym wersjom czasopisma (np. wersji drukowanej, wersji elektronicznej) także powinny być przechowywane w bazie, ale muszą być traktowane wyłącznie jako element pozwalający na lokalizację czasopisma w innych bazach, które nie wyróżniają numeru ISSN-L.
 3. System aplikacyjny dla wydawców czasopism powinien zapewniać jednoznaczną identyfikację czasopisma poprzez numer ISSN dowolnej wersji czasopisma.
 4. Do całości zgromadzonych danych należy stosować procedury kontrolne, obowiązkowo przynajmniej na ostatnim etapie poprzedzającym opublikowanie Wykazu, co zagwarantuje wykrycie błędów, które mogłyby powstać w procesie przetwarzania danych na skutek usterek obsługującego systemu informatycznego lub czynnika ludzkiego
 5. Wszelkie dostrzeżone nieścisłości i braki dotyczące czasopism polskich powinny być konsultowane z Biblioteką Narodową i za jej pośrednictwem zgłaszane do Międzynarodowego Centrum ISSN.
 6. Opublikowany Wykaz, który obecnie składa się z liczby porządkowej czasopisma (w danej części wykazu), tytułu, numeru ISSN (bez określenia, której wersji czasopisma dotyczy), liczby punktów, powinien zawierać wszystkie numery ISSN przypisane odpowiednim wersjom czasopisma oraz numer ISSN-L.
 7. Wykaz powinien zawierać także informacje pomocnicze, które pozwoliłyby na identyfikację czasopisma także tym jednostkom, które w lokalnych bazach danych rejestrują publikacje w sposób uproszczony tj. bez numerów ISSN. Do takich danych należą na pewno: wydawca, miejsce wydania, rok pierwszego wydania, ale zagadnienie należy skonsultować z doświadczonymi bibliotekarzami.

Bibliografia

[1] Rozporządzenie Ministra Nauki i Szkolnictwa Wyższego z dnia 13 lipca 2012 r. w sprawie kryteriów i trybu przyznawania kategorii naukowej jednostkom naukowym

Dziennik Ustaw RP z 1.08.2012 poz. 877

http://www.bip.nauka.gov.pl/_gAllery/19/31/19319/poz._877.pdf

[2] Komunikat Ministra Nauki i Szkolnictwa Wyższego z dnia 17 września 2012 r. w sprawie wykazu czasopism naukowych wraz z liczbą punktów przyznawanych za publikacje w tych czasopismach

http://www.bip.nauka.gov.pl/_gAllery/20/10/20107/20120917_Komunikat_-_lista_czasopism.pdf

[3] Załącznik do komunikatu Ministra Nauki i Szkolnictwa Wyższego z dnia 17 września 2012 r.

Część A wykazu czasopism naukowych: czasopisma naukowe posiadające współczynnik wpływu Impact Factor (IF), znajdujące się w bazie Journal Citation Reports (JCR) wraz z liczbą punktów przyznawanych za publikację w tych czasopismach

http://www.nauka.gov.pl/fileadmin/user_upload/Finansowanie/finansowanie_nauki/Lista_czasopism/20120918_WYKAZ_CZASOPISM_NIEPUNKTOWANYCH.pdf

- [4] Załącznik do komunikatu Ministra Nauki i Szkolnictwa Wyższego z dnia 17 września 2012 r.
Część B wykazu czasopism naukowych : czasopisma naukowe nieposiadające współczynnika wpływu Impact Factor (IF) wraz z liczbą punktów przyznawanych za publikację w tych czasopismach
http://www.bip.nauka.gov.pl/_gAllery/20/10/20109/20120917_CZESC_B_WYKAZU_CZASOPISM_NAUKOWYCH.pdf
- [5] Załącznik do komunikatu Ministra Nauki i Szkolnictwa Wyższego z dnia 17 września 2012 r.
Część C wykazu czasopism naukowych: czasopisma naukowe znajdujące się w bazie European Reference Index for the Humanities (ERIH) wraz z liczbą punktów przyznawanych za publikację w tych czasopismach
http://www.bip.nauka.gov.pl/_gAllery/20/11/20110/20120917_CZESC_C_WYKAZU_CZASOPISM_NAUKOWYCH.pdf
- [6] Czasopisma naukowe nieujęte w wykazie czasopism naukowych ogłoszonym przez Ministra Nauki i Szkolnictwa Wyższego w komunikacie z dnia 17 września 2012 r. w sprawie wykazu czasopism naukowych wraz z liczbą punktów przyznawanych za publikację w tych czasopismach
http://www.nauka.gov.pl/fileadmin/user_upload/Finansowanie/finansowanie_nauki/Lista_czasopism/20120918_WYKAZ_CZASOPISM_NIEPUNKTOWANYCH.pdf
- [7] Ustawa z dnia 18 marca 2011 r. o zmianie ustawy – Prawo o szkolnictwie wyższym, ustawy o stopniach naukowych i tytule naukowym oraz o stopniach i tytule w zakresie sztuki oraz o zmianie niektórych innych ustaw
Dz. U. z 2011 r. Nr 84, poz. 455, Nr 112, poz. 654.
<http://isap.sejm.gov.pl/DetailsServlet?id=WDU20110840455>
- [8] Informacja dotycząca dofinansowania projektu systemowego w ramach Priorytetu IV Programu Operacyjnego Kapitał Ludzki pt.: "Stworzenie systemu informacji o szkolnictwie wyższym"
<http://www.nauka.gov.pl/ministerstwo/aktualnosci/aktualnosci/kategoria/2096/artikel/informacja-dotyczaca-dofinansowania-projektu-systemowego-w-ramach-priorytetu-iv-programu-operacyjnego/>
- [9] Odpowiedź ministra nauki i szkolnictwa wyższego na interpelację nr 23523 w sprawie planowanego wprowadzenia od nowego roku akademickiego elektronicznego rejestru szkolnictwa wyższego POL-on
<http://orka2.sejm.gov.pl/IZ6.nsf/main/076C46BC>
- [10] Ogłoszenie dotyczące wyboru Partnerów do projektu systemowego PO KL
<http://www.nauka.gov.pl/finansowanie/fundusze-europejskie/program-operacyjny-kapital-ludzki/projekty/projekty-systemowe/projekty-systemowe-informacje/projekty-systemowe-informacje/artikel/ogloszenie-dotyczace-wyboru-partnerow-do-projektu-systemowego-po-kl/>
- [11] Odpowiedź sekretarza stanu w Ministerstwie Nauki i Szkolnictwa Wyższego - z upoważnienia ministra - na interpelację nr 6478 w sprawie problemów z przesyłaniem zgłoszeń studentów przez niektóre uczelnie do nowego systemu informacji akademickiej POL-on
<http://www.sejm.gov.pl/Sejm7.nsf/InterpelacjaTresc.xsp?key=0A5A92D9>
- [12] ISSN Manual 2012
<http://issn.org/files/issn/Documentation/Manuels/ISSNManual2012-EN.pdf>
- [13] Przegląd Elektrotechniczny 0033-2097 – strona internetowa
<http://pe.org.pl/index.php>
- [14] Editions : the new monthly Australian review of books 1033-0097
<http://catalogue.nla.gov.au/Record/1159654>
- [15] Britannica – edycja polska / red. Wojciech Wolarski. T. 1–49 (wraz z suplementem). Poznań : Kurpisz, 1997–2005. ISBN (całość) 978-83-86600-54-0
- [16] Czaplicki Edward, Podgórska Barbara , Głębocki Jacek: Substancje biologicznie czynne i zawiera-

jące je środki ochrony roślin dopuszczone do obrotu i stosowania w Polsce w 1999 roku. Poznań : Instytut Ochrony Roślin, 2000. 105 s. ISBN 978-83-913860-1-9

[17] Głębocki Jacek, Pietruszewski Grzegorz: Zasoby bibliograficzne istotnym elementem bazy wiedzy. - Forum Bibl. Med. 2011 R. 4 nr 1, s. 469-473

<http://bg.umed.lodz.pl/images/stories/wydawnictwa/forum/ForumNr7.pdf>

[18] Polska Norma PN-ISO 3297:2010, Informacja i dokumentacja - Międzynarodowy znormalizowany numer wydawnictw ciągłych (ISSN). Polski Komitet Normalizacyjny, 2010 (na podstawie normy ISO 3297:2007, zastępującej normę ISO 3297:1998).

<https://sklep.pkn.pl/?a=show&m=product&pid=566428>

[19] Reynolds Regina Romano: „ISSN: Linking Data and Metadata”. NASIG NISO Preconference Phoenix, June 4, 2008

http://www.niso.org/news/events/2008/metadata08/agenda/reynolds_metadata08.pdf

[20] Rowińska Magdalena: ISSN - jednoznaczna identyfikacja czasopisma.

W: Katalogowanie wydawnictw ciągłych w katalogu NUKAT : materiały, Toruń, 23-24.05.2012

<http://kpbc.umk.pl/dlibra/docmetadata?id=73640&from=publication>

[21] Godel Sylvie, de Kaenel Isabelle, Iriarte Pablo: Web Services for Bibliometrics.

http://archivesic.ccsd.cnrs.fr/docs/00/54/02/89/PDF/EAHIL_2010_Web_services_for_Bibliometrics_20100615.pdf

[22] Cristin Current Research Information System In Norway

<http://www.cristin.no/>

[23] Częstkowy wykaz czasopism w systemie Cristin

<http://www.cristin.no/konsortieavtaler/avtaler/IOPscience%20extra%20titles%20for%20ABM.pdf>