

---

# ProBIT – prospektywna metoda tworzenia trawersowalnych indeksów cytowań a współczesne problemy organizacji przestrzeni informacji w tradycyjnych bibliograficznych bazach danych

Anna Małgorzata Kamińska

*Biblioteka Główna Politechniki Śląskiej w Gliwicach*

---

## Abstrakt

**Cel/Teza:** Artykuł przedstawia autorską metodę budowy bibliograficznych baz danych, sprzyjającą sprawnemu gromadzeniu danych bibliograficznych o wysokiej jakości, które bez dalszych przekształceń stanowią źródło zaawansowanych analiz bibliometrycznych (również tych, które nie ograniczają się jedynie do prostych zestawień ilościowych wyliczanych na podstawie danych zgromadzonych w układzie relacyjnym, ale takich które wymagają trawersowania po ścieżkach grafów reprezentujących sieć cytowań artykułów, czy współpracy między ich autorami).

**Koncepcja/Metody badań:** Autorka, stojąc przed problemem zgromadzenia znacznego wolumenu danych bibliograficznych do realizacji badań własnych, zapoznała się z szeregiem gotowych narzędzi pozwalających na rejestrowanie danych bibliograficznych, jednak żadne z nich nie oferowało oczekiwanej ergonomii, dbałości o jakość gromadzonych danych ani sposobu ich gromadzenia w układzie umożliwiającym prowadzenie na nich dalszych badań bibliometrycznych. Wyciągając wnioski z negatywnych doświadczeń nabytych podczas pracy z tymi narzędziami, zaproponowała własną metodę, którą przedstawia z wykorzystaniem podstawowych notacji języka UML.

**Wyniki i wnioski:** Przedstawiona metoda okazała się skuteczna i pozwoliła zgromadzić z autopsji w stosunkowo krótkim czasie bardzo dobrej jakości dane o ponad 36 tys. artykułów i ponad 22 tys. autorów, które bezpośrednio posłużyły do realizacji badań bibliometrycznych, które zostaną przedstawione w osobnej pracy.

**Zastosowania praktyczne:** W pierwszej części artykułu zaprezentowano zagrożenia, jakie niesie z sobą gromadzenie danych bibliograficznych w tzw. tradycyjnym układzie. Na konkretnych przykładach, dla wybranych krajowych bibliograficznych baz danych wykazano, że zagrożenia te materializują się w postaci występowaniu szeregu anomalii, które zostały przez autorkę skategoryzowane, a naprzeciw którym wychodzi zaproponowana metoda. Dzięki jej zastosowaniu możliwe jest podniesienie jakości danych gromadzonych w bibliograficznych bazach danych oraz obniżenie kosztów ich gromadzenia poprzez usprawnienie całego procesu.

**Oryginalność/Wartość poznawcza:** W polskim piśmiennictwie bibliologii i informatologii rzadko podejmowana jest tematyka funkcjonowania bibliograficznych baz danych, najczęściej badacze ograniczają się do definiowania „zakresów informacyjnych przetwarzanych danych”. Autorka zwraca uwagę, że na osiągnięty wynik końcowy (jakość danych, koszt ich pozyskania i możliwości dalszego przetwarzania) w równym stopniu wpływ ma również układ, w jakim dane są gromadzone (model informacyjny) oraz odpowiednia organizacja procesu gromadzenia.

## Słowa kluczowe

Bibliograficzne bazy danych. ProBIT. Metoda prospektywna. Przestrzeń informacji. Trawersowalny indeks cytowań.

Otrzymano: 19 lutego 2017. Zrecenzowano: 23 czerwca 2017. Zaakceptowano: 30 sierpnia 2017.

## 1. Wprowadzenie

Dzięki mechanizmom systemów zarządzania bibliograficzne bazy danych umożliwiają gromadzenie, przechowywanie, przetwarzanie, wyszukiwanie i udostępnianie informacji. Ważnymi elementami bibliograficznych baz danych są informacje opisujące cytowania pomiędzy poszczególnymi jednostkami bibliograficznymi. Dzięki tym informacjom możliwe jest wsparcie szeregu działań, spośród których wymienić należy chociażby takie jak:

- ewaluacja prac naukowych,
- ewaluacja dorobku naukowego autora;
- ewaluacja potencjału instytucji afiliujących;
- ewaluacja rangi czasopism i innych wydawnictw naukowych;
- badania nad rozwojem dziedzin nauki i powstawaniem nowych ich obszarów;
- wzbogacenie standardowych mechanizmów wyszukiwawczych OPAC (Woźniak-Kasperek, 2011) o możliwość podpowiadania dokumentów podobnych, gdzie miary podobieństwa mogą bazować na metodach powiązań bibliograficznych, współcytowań i innych (Marszakowa-Szajkiewicz, 2009).

Chociaż pierwsze opinie o istnieniu kognitywnego związku między pracą cytowaną i cytującą datuje się na 1955 r. (Garfield, 1955), to zainteresowanie badaniami tych związków ciągle rośnie i to zarówno pod względem ilościowym (włączanie w badawcze obszary coraz to nowych dziedzin nauki), jak i jakościowym (opracowywanie nowych miar analiz bibliometrycznych oraz bibliometrycznych metod badawczych). Sprzyja temu niewątpliwie rozwój technologii informacyjnych, powszechna komputeryzacja coraz to nowych dziedzin życia, w tym ekosystemu bibliotek oraz ciągły wzrost dostępnych mocy obliczeniowych. Wraz z gwałtownym rozwojem metod badania powiązań w sieciach społecznościowych (ang. *social network analysis*) (Al et al., 2012), przeniesienie na grunt bibliometrii i interpretowanie uzyskanych tą drogą wyników, wydają się być ograniczone jedynie wyobraźnią poszczególnych naukowców (Kretschmer & Kretschmer, 2007).

Jak zauważają badacze krajowi i zagraniczni, w tym Władysław Kolasa, obszernie wyjaśniając to na wielu przykładach, możliwość prowadzenia takich badań oraz wiarygodność ich wyników, zależą przede wszystkim od dostępności i jakości gromadzonych danych, a ta pozostawia wiele do życzenia szczególnie w przypadku bibliograficznych baz danych obejmujących nauki humanistyczne, gdyż te będąc bardziej niż nauki ścisłe czy techniczne spolaryzowane językowo i mając bardziej lokalny charakter współpracy naukowej, są indeksowane przez globalne serwisy w dużo mniejszym stopniu, co prowadzi do powstawania lokalnych bibliograficznych baz danych (Kolasa, 2011). Choć trudno się z tą opinią nie zgodzić, to można zaryzykować postawienie szerszych hipotez. Po pierwsze, przypuszczać można, że niska jakość danych wynika przede wszystkim z lokalnego charakteru gromadzonych informacji bibliograficznych, a w mniejszym stopniu z dziedziny, w ramach której te informacje są gromadzone. Duże i globalne korporacje posiadające

odpowiednie środki na weryfikację jakości rekordów oraz nadążanie za rozwojem technologii, stoją na uprzywilejowanej pozycji wobec dostawców lokalnych (krajowych) baz danych, które często zaprojektowane wiele lat temu, są eksploatowane po dni dzisiejsze bez koniecznych ulepszeń. Wśród wrywkowej kontroli jakości i spójności większości bibliograficznych baz danych dostępnych w Polsce, a realizowanej w ramach badań prowadzonych na potrzeby niniejszych rozważań, stwierdzono problemy, które dotyczą w tym samym stopniu baz humanistycznych jak i technicznych. Przedstawione w dalszej części artykułu przykłady ograniczono do bazy BazTech (<http://yadda.icm.edu.pl/baztech/>), która jest jedną z najbardziej reprezentatywnych baz bibliograficznych z zakresu nauk technicznych oraz bazy CYTBIN (<http://www1.bg.us.edu.pl/bazy/cytbins/>), indeksującej prace z dziedziny bibliologii i informatologii. Jako drugą hipotezę można sformułować przypuszczenie, że możliwość realizacji bardziej zaawansowanych badań bibliometrycznych nie jest warunkowana jedynie jakością gromadzonych danych, ale również organizacją tych danych w konkretnych systemach zarządzania bazami danych. Wiele metod analiz bibliometrycznych operuje na ścieżkach cytowań, tak więc organizacja informacji opisującej bibliografię załącznikową w postaci par „jednostka cytowana – jednostka cytująca” jest niewystarczająca (nawet gdy dane te odzwierciedlają prawdziwe informacje) bez możliwości uwzględnienia faktu, że jednostka cytująca może być jednostką cytowaną dla innych jednostek cytujących.

Więcej informacji o inicjatywach bibliograficznych baz danych realizowanych na obszarze Polski, stanie tych projektów na 2009 r. i planowanych naówczas kierunkach ich rozwoju, uzyskać można ze sprawozdania z konferencji dotyczącej bibliograficznych baz danych (Drabek, 2009), natomiast ocenę stopnia ich dojrzałości i możliwości wykorzystania przez pracowników i jednostki szkół wyższych znaleźć można w (Garczyńska, 2013).

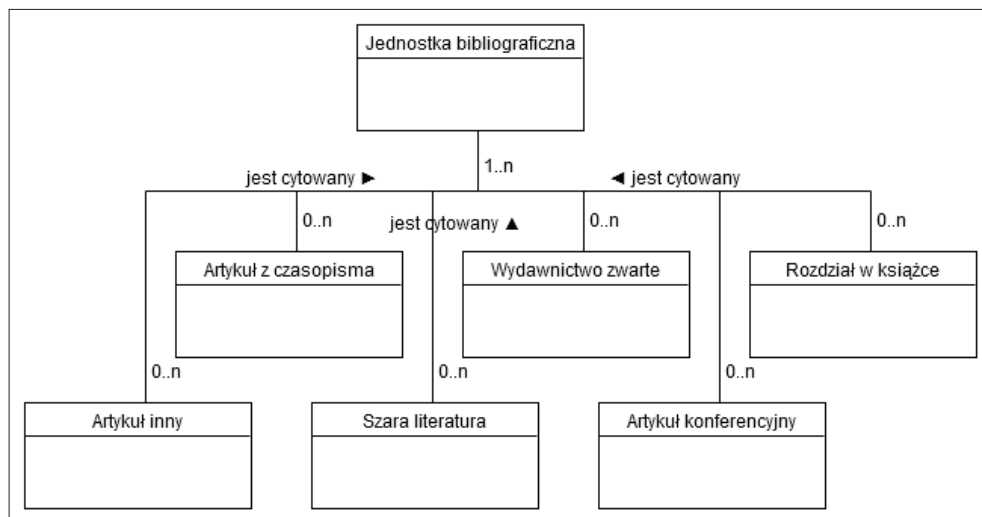
Niniejszy artykuł opisuje problemy, z jakimi styka się większość instytucji udostępniających bibliograficzne bazy danych na terenie Polski, skupiając się na perspektywie konceptualno-technicznej, a dokładniej na architekturze przestrzeni informacyjnej oraz organizacji pracy osób obsługujących systemy bibliograficznych baz danych. Po przedstawieniu i omówieniu tych problemów, w dalszej części artykułu, zaproponowana zostanie metoda organizacji informacji bibliograficznych gromadzonych w bazach, która w połączeniu ze zmianą organizacji pracy osób rejestrujących te informacje, zapewni dużo lepszą jakość gromadzonych danych oraz gromadzenie ich w układzie umożliwiającym realizację zaawansowanych badań bibliometrycznych wprost na zgromadzonych danych w formie relacyjnej. Skuteczność tej metody znalazła praktyczne potwierdzenie podczas realizacji badań własnych, na potrzeby których, rezygnując z wykorzystania narzędzi gotowych, takich jak np. te, które przedstawiono w osobnym opracowaniu (Kamińska, 2017d), zaimplementowano środowisko z zastosowaniem przedstawianej metody. Dane zgromadzone w ramach tych badań posłużyły do realizacji szeregu analiz bibliometrycznych od prostych zestawień ilościowych, poprzez zastosowanie metody powiązań bibliograficznych i metody współcytowań, aż po badania z zastosowaniem szeregu metod analiz sieci społecznościowych.

## 2. Problemy organizacji przestrzeni informacji w tradycyjnych bibliograficznych bazach danych

Większość inicjatyw bibliograficznych baz danych realizowanych na terenie Polski opiera się na tradycyjnym modelu struktury danych przedstawionym za pomocą diagramu klas języka UML na rysunku 1. Omówienie tej wizualnej techniki komunikacyjnej i propozycję jej wykorzystania na gruncie zagadnień bibliologii i informatologii autorka zaproponowała w pracy pt. *O rozwoju graficznych języków komunikacji* (Kamińska, 2017b). Bardziej obszernie omówienie poszczególnych technik języka UML znaleźć można w monografii samych jego twórców (Booch et al., 2002).

Z modelu wynika, że nadrzędnym elementem tradycyjnej organizacji przestrzeni informacyjnej jest jednostka bibliograficzna, która podlega opisowi. Poza innymi elementami składowymi jej opisu (jak np. tytuł, typ nośnika, autorzy i wiele innych), które nie mają jednak istotnego wpływu na prowadzone tu rozważania, głównym elementem leżącym u podstaw koncepcji bibliograficznych baz danych jest bibliografia załącznikowa.

Rys. 1. Model organizacji danych w tradycyjnych bibliograficznych bazach danych

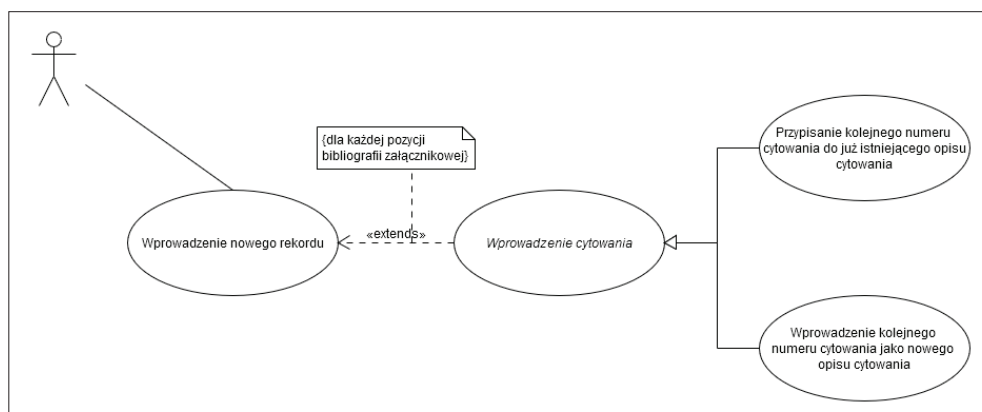


W tradycyjnym modelu opisy bibliograficzne zawarte w bibliografii załącznikowej są gromadzone w składnicach danych oddzielnie dla poszczególnych form wydawniczych (np. artykuł z czasopisma, wydawnictwo zwarte). Na przedstawionym diagramie wyróżniono sześć przykładowych form, choć w konkretnych bazach danych liczba tych kategorii może być różna. Potrzeba wyróżniania różnych kategorii wynika z faktu, że w ramach różnych kategorii dokumentuje się różne cechy dokumentów, co z kolei dyktowane jest potrzebą jak najdokładniejszego opisu jednostek poszczególnych typów. Jak wynika z diagramu, każda jednostka danej kategorii (tj. obiekt klasy Artykuł z czasopisma, obiekt klasy Wydawnictwo zwarte itd.), może być przyporządkowana (tę relację asocjacji możemy nazwać „jest cytowany”) do jednej lub wielu jednostek bibliograficznych. Analizując relacje w odwrotnym kierunku widzimy z kolei, że każda opisywana jednostka bibliograficzna,

w swojej bibliografii załącznikowej, może zawierać zero lub wiele jednostek poszczególnych form wydawniczych. Taka organizacja danych, oparta jedynie na związkach asocjacji, choć na pierwszy rzut oka sprawia wrażenie porządku i dobrej organizacji przestrzeni danych, ma szereg wad i niesie ze sobą wiele zagrożeń, które omówione zostaną w dalszej części artykułu.

Powyżej przedstawiony strukturalny opis przestrzeni tradycyjnych bibliograficznych systemów baz danych jest niepełny, gdyż nie opisuje w żaden sposób jego aspektów behawioralnych, z których dla niniejszych rozważań najistotniejszy jest aspekt interakcji użytkownika takiego systemu z graficznym interfejsem użytkownika (ang. GUI – *graphical user interface*). Aspekt ten ukazuje diagram przypadków użycia (ang. *use case diagram*) wyrażony w języku UML na rysunku 2.

Rys. 2. Model organizacji pracy dla rejestracji jednostek bibliograficznych w tradycyjnych bibliograficznych bazach danych



Modelowanie interakcji użytkownika (interesariusza) z systemem (niekoniecznie musi to być system informatyczny) jest dziedziną samą w sobie, na temat której powstało wiele obszernych opracowań, jak np. (Cockburn, 2004). Dla potrzeb tego badania (jako że mowa o systemach baz danych) przyjęto, że przypadek użycia (ciąg interakcji użytkownika z systemem) jest najmniejszą (atomową) jednostką ciągu czynności, skutek której może zostać utrwalony w składnicy danych. W tym przypadku będzie to wprowadzenie do bazy nowego rekordu jednostki bibliograficznej lub wprowadzenie informacji o pojedynczej pozycji bibliografii załącznikowej dla danej jednostki bibliograficznej. Systemy bibliograficznych baz danych udostępniają wiele innych funkcjonalności (np. możliwość korekty rekordu wcześniej wprowadzonego, edycję słowników typów rozpowszechniania, itp.), jednak nie mają one istotnego znaczenia z punktu widzenia niniejszych rozważań i dlatego zostały w modelu pominięte. Możliwość budowania ograniczonych modeli pozwalających przedstawić system z istotnej w danym momencie perspektywy, z zaniechaniem umieszczania w nim elementów, które mogłyby zaciemnić obraz, jest postrzegane jako dobra praktyka budowy modeli.

Przedstawiony diagram należy „czytać” następująco. Wprowadzany jest nowy rekord do systemu (sposób w jaki jest to robione – kolejność wypełniania pól, itp. zależne będą od konkretnego systemu i mogłyby być wyrażone poprzez opis scenariuszy przypadków

użycia, ale można od nich abstrahować w ramach niniejszych rozważań, podobnie jak można było abstrahować od zawartości informacyjnej poszczególnych klas na poprzednim diagramie). Po wpisaniu danych opisujących nowy rekord, można przejść do opisu jego bibliografii załącznikowej. Oznaczenie «extends» wyraża właśnie, że dany przypadek użycia może, ale nie musi, „rozszerzać” inny przypadek użycia. W tej sytuacji „rozszerzenie” nie nastąpi dla jednostek nie posiadających bibliografii załącznikowej. Notatka nad relacją «extends» oznacza natomiast, że rozszerzający przypadek użycia, będzie realizowany dla każdej z pozycji bibliograficznych z osobna. Rozszerzającym przypadkiem jest „Wprowadzenie cytowania”, które – jak widać na diagramie – opisane jest kursywą. Oznacza to, że ten przypadek użycia jest abstrakcyjny, a więc nie jest możliwe doprecyzowanie przebiegu czynności, które powinny być w jego ramach realizowane, gdyż w tej sytuacji czynności te mogą być realizowane na dwa różne, wzajemnie wykluczające się sposoby, co obrazuje zamknięta strzałka oznaczająca związek „specjalizacji”, poprowadzona z przypadków użycia opisujących odpowiednio wprowadzenie jako pozycję cytowaną istniejącego już wcześniej rekordu z przypisaniem mu numeru z listy bibliografii załącznikowej lub dopisanie nowego rekordu z przypisaniem mu numeru z listy bibliografii załącznikowej.

Przedstawiona powyżej tradycyjna organizacja przestrzeni informacji prowadzi często do gromadzenia danych redundantnych oraz może nie sprzyjać właściwej dbałości o jakość gromadzonych danych, co skutkuje powstawaniem licznych anomalii. Redundancja najczęściej zachodzi w sytuacjach, kiedy dana pozycja bibliografii załącznikowej staje się (lub była wcześniej) również opisywaną jednostką bibliograficzną. Mamy wtedy do czynienia z sytuacją, kiedy ta sama jednostka występuje w roli jednostki bibliograficznej oraz którejś z cytowanych form wydawniczych (dla potrzeby późniejszych rozważań oznaczono ten przypadek jako  $\alpha/\alpha$ ). W takich sytuacjach oprócz redundancji dochodzi również często do wewnętrznych sprzeczności zarejestrowanych informacji w przypadkach, kiedy wspólne części zakresu informacyjnego jednostki bibliograficznej oraz którejś z cytowanych form wydawniczych przechowują różne dane reprezentujące te same atrybuty (przypadek  $\beta/\beta$ ). Z trzecią anomalią mamy natomiast do czynienia w sytuacji, gdy w bazie danych istnieje już rekord opisujący daną cytowaną formę wydawniczą, natomiast operatorowi systemu nie uda się go odnaleźć i zamiast realizacji przypadku użycia „Przypisanie kolejnego numeru cytowania do już istniejącego opisu cytowania” zrealizuje „Wprowadzenie kolejnego numeru cytowania jako nowego opisu cytowania” (przypadek  $\gamma/\gamma$ ). Anomalia ta niesie ze sobą poważne, negatywne konsekwencje dla mechanizmu wyszukiwania jednostek, które cytują zadany artykuł, czyniąc jego działanie mało wiarygodnym. Dwóch z wcześniej wymienionych przypadków nie można uniknąć i wpisują się one wręcz w konstrukcję systemów o wyżej przedstawionej architekturze, jednak ostatni przypadek jest już zależny od rzetelności operatora, która powinna być wspierana przez taką konstrukcję graficznego interfejsu użytkownika, która będzie faworyzować „ponowne użycie” (ang. *reuse*) rekordu, a więc przypisanie istniejącego rekordu cytowanej formy do numeru pozycji bibliografii załącznikowej danej jednostki zamiast dopisywania nowego. Najbardziej oczywistą i najczęstszą (choć nie jedyną) motywacją każdego operatora systemu jest możliwość włożenia jak najmniejszego wysiłku i poświęcenia jak najmniejszej ilości czasu na realizację danej czynności, zatem interfejs graficzny powinien być tak skonstruowany, aby wyszukiwanie wśród rekordów istniejących i przypisanie ich do danego numeru bibliografii załącznikowej było znacznie mniej czasochłonne niż dopisanie rekordu nowego.



Niektóre systemy bibliograficznych baz danych zakładają wręcz zasilanie swoich składnic informacji jedynie poprzez dopisywanie nowych rekordów cytowań (przypadek użycia opcji „wprowadzenie kolejnego numeru cytowania jako nowego opisu cytowania”), pozostawiając porządkowanie zgromadzonych danych realizowanej z opóźnieniem operacji scalania rekordów. Należy tu zauważyć, że występujący często w literaturze termin „scalanie” używany w opisywanym kontekście, nie wydaje się do końca właściwy, gdyż sugeruje, że rekordy wynikowe zawierać będą sumę informacji pochodzącą z rekordów źródłowych, co najczęściej nie jest prawdą. Bardziej właściwym terminem opisującym ten proces wydaje się być „deduplikacja” (Ilyas & Chu, 2015). Owa deduplikacja, realizowana z opóźnieniem w stosunku do czasu zarejestrowania rekordów źródłowych, nie tylko zwiększa sumaryczny nakład pracy na utrzymanie danych w stanie „maksymalnie spójnym”, ale również zmniejsza skuteczność dbałości o jakość gromadzonych danych. I choć jej mechanizmy są i będą użyteczne w każdym systemie bibliograficznych baz danych, to należy dążyć do sytuacji, w której liczba rekordów „nią dotkniętych” będzie jak najmniejsza.

Należy tutaj również wspomnieć, że oprócz wyżej wymienionych czynników sprzyjających powstawaniu anomalii, a leżących po stronie konstrukcji systemu lub rzetelności operatora, istnieje również czynnik od nich niezależny. Są to błędy w materiałach źródłowych popełniane przez samych autorów lub też powstałe w procesie redakcji i druku. Zachodzi wtedy pytanie, które informacje bibliograficzna baza danych powinna uważać za prawdziwe? Te przedstawiane przez egzemplarze drukowane, czy też te, które miał na myśli autor i których poprawność można sprawdzić w innych źródłach? Oczywiście jest, że z punktu widzenia skuteczności realizacji badań bibliometrycznych i wiarygodności uzyskanych tą drogą wyników pozytywnej odpowiedzi należy udzielić na drugie pytanie. Takie też stanowisko przyjęła autorka podczas realizacji swoich badań. Nic nie stoi jednak na przeszkodzie, a nawet wydaje się właściwe, aby docelowa konstrukcja systemów bibliograficznych baz danych przechowywała oba warianty informacji.

### 3. Przykłady problemów zidentyfikowanych w wybranych ogólnodostępnych bibliograficznych bazach danych

W rezultacie wrywkowej kontroli jakości i spójności danych gromadzonych w bibliograficznych bazach danych opisujących obszary nauki, a realizowanej w ramach prezentowanych tu badań, potwierdzono problemy wcześniej opisane oraz stwierdzono, że dotyczą one w podobnym stopniu baz humanistycznych jak i technicznych. Przedstawione w dalszej części rozdziału przykłady ograniczono do bazy BazTech oraz bazy CYTBIN, której zawartość stanowiła również podstawy do innych badań autorskich dotyczących rozwoju metod bibliometrycznych (Kamińska, 2017c), wizualizacji wyników uzyskanych z ich pomocą (Kamińska, 2017e), poprawy jakości gromadzonych danych bibliograficznych (Kamińska, 2017a) oraz zastosowania metod analizy sieci społecznościowych w bibliometrii (Kamińska, 2018).

Rysunek 3 przedstawia okno systemu CYTBIN w przeglądarce WWW z informacją o opisywanej jednostce (czyli takiej, dla której gromadzona jest informacja o bibliografii załącznikowej), zaś rysunek 4 przedstawia okno z informacją o artykule zarejestrowanym w ramach cytowanych wydawnictw w formie wydawnictwa ciągłego. Jak widać, oba rekordy reprezentujące jeden byt logiczny zawierają sprzeczne informacje o czasopiśmie

i stronach, na których znajduje się artykuł. Przykład ten ilustruje występowanie anomalii typu/ $\alpha$ / oraz/ $\beta$ /. Faktycznie artykuł, którego dotyczą przedstawione rekordy ukazał się w „Bibliotekarzu” nr 7–8 na stronach 10–14, a nie na stronach 30–34. Widzimy więc, że oba rekordy zawierają błędne informacje oraz reprezentują oddzielne fizyczne jednostki danych.

Rys. 3. Okno systemu CYTBIN wyświetlające informacje o przykładowej opisywanej jednostce

<b>Autor:</b>	Wolff K.
<b>Rodzaj publikacji:</b>	Artykuł
<b>Tytuł:</b>	Biblioteka publiczna w opinii społecznej
<b>Czasopismo:</b>	Przegląd Biblioteczny
<b>Rok publikacji:</b>	2010
<b>Tom:</b>	78
<b>Numer:</b>	7-8
<b>Strony:</b>	10-14
<b>Symbol klasyfikacji tematycznej:</b>	3.12 - Biblioteki publiczne
<b>Liczba cytowań:</b>	0
<b>Liczba instytucji sprawczych:</b>	1
<b>Słowa kluczowe:</b>	Bibliotekoznawstwo, Biblioteka publiczna, Jakość usług bibliotecznych, Czytelnik, Badanie ankietowe
<b>Słowa kluczowe ang.:</b>	Library and information science, Public library
<b>Abstrakt:</b>	Opisano wynik ankiety przeprowadzonej w 2008r. przez Instytut Książki i Czytelnictwa BN. Była ona

Źródło: Baza CYTBIN, dostęp: 6.12.2016


Rys. 4. Okno systemu CYTBIN wyświetlające informacje o przykładowym artykule dla wydawnictw ciągłych

<b>CYTOWANIE - WYDAWNICTWO CIĄGŁE</b>	
<b>Autorzy:</b>	Wolff K.
<b>Rodzaj nośnika:</b>	D
<b>Tytuł:</b>	Biblioteka publiczna w opinii społecznej
<b>Czasopismo:</b>	Bibliotekarz
<b>Rok publikacji:</b>	2010
<b>Numer:</b>	7-8
<b>Strony:</b>	30-34

Źródło: Baza CYTBIN, dostęp: 6.12.2016



Rys. 5. Okno systemu BAZTECH wyświetlające informacje o przykładowym artykule

Artykuł - szczegóły	
Czasopismo	<b>Biuletyn Wojskowej Akademii Technicznej</b>
Tytuł artykułu	<b>Aerozele węglowe otrzymywane z prekursora rezorcynowo-furfuralowego</b>
Autorzy	<a href="#">Kiciński, W.</a>
Treść / Zawartość	
Warianty tytułu	EN Carbon aerogels obtained from resorcinol-furfural precursors
Języki publikacji	PL
Abstrakty	<p>PL Przedstawiono nową metodę otrzymywania aerozeli węglowych poprzez karbonizację organicznych aerozeli rezorcynowo-furfuralowych. Zele rezorcynowo-furfuralowe otrzymywano w wyniku katalizowanej kwasowo kondensacji rezorcyny i furfuralu prowadzonej w wodno-metanolowym roztworze soli nieorganicznych. Następnie sól z sieci żelu wymywano i oczyszczony materiał suszono. Suszenie tradycyjnymi metodami nie prowadziło do zniszczenia struktury porowatej żelu dzięki czemu możliwe było uzyskanie aerozeli organicznych bez konieczności stosowania warunków nadkrytycznych ekstrakcji rozpuszczalnika ze struktury żelu. Karbonizacja aerozeli organicznych pozwalała uzyskać mikro-makroporowate lub mikro-mezo-makroporowate materiały węglowe o strukturze mikropianek i powierzchni właściwej z przedziału 250-580 m<sup>2</sup>/g. Otrzymane aerozele organiczne i węglowe badano przy użyciu wielu technik, m.in. SEM, XRD, niskotemperaturowej adsorpcji azotu, analizy termogravimetrycznej i elementarnej.</p> <p>EN Acidic condensation of resorcinol with furfural carried out in water-methanol solution of inorganic salts (NaCl or NH<sub>4</sub>ClO<sub>4</sub>) leads to organic R-F gel. After removing the salt from the gel network, the purified gel can be dried under ambient conditions without significant shrinkage, so the supercritical drying can be avoided, and an organic aerogel can be obtained through a simpler procedure. By changing the H<sub>2</sub>O/MeOH ratio and/or the inorganic salt amount, the structure of the created gel can be controlled in the nano and micron-level and as a result polymeric (nano-sized) or colloidal (micron-sized) aerogels can be produced. Carbonization of the colloidal aerogels yields bimodal micro-macroporous carbonaceous materials with the structure of microfoam and low density, while carbonization of the polymeric ones yields bimodal, micro and meso-macroporous carbonaceous materials with high density and mechanical strength. Specific surface area of the carbon aerogels ranged from 250 to 580 m<sup>2</sup>/g. The obtained organic and carbon aerogels were investigated by the means of SEM, XRD, elemental analysis, N<sub>2</sub> adsorption, and TG.</p>
Słowa kluczowe	<p>PL <a href="#">synteza żol-żel</a> <a href="#">aerozel węglowy</a> <a href="#">aerozel koloidalny</a> <a href="#">polikondensacja</a></p> <p>EN <a href="#">sol-gel synthesis</a> <a href="#">carbon aerogel</a> <a href="#">colloidal structure</a> <a href="#">polycondensation</a></p>
Wydawca	<a href="#">Wojskowa Akademia Techniczna im. Jarosława Dąbrowskiego</a>
Czasopismo	<a href="#">Biuletyn Wojskowej Akademii Technicznej</a>
Rocznik	<a href="#">2009</a>
Tom	<a href="#">Vol. 58, nr 4</a>
Strony	197--221
Opis fizyczny	Bibliogr. 57 poz., tab., wykr.
Twórcy	<p>autor <a href="#">Kiciński, W.</a></p> <p>Wojskowa Akademia Techniczna, Wydział Nowych Technologii i Chemii, Instytut Chemii, 00-908 Warszawa, ul. S. Kaliskiego 2</p>
Bibliografia	<p>[1] R. W. Pekala, F. M. Kong, Polym. Prep., 30, 1, 1989, 221.</p> <p>[2] R. W. Pekala, D. W. Schaefer, Macromolecules, 26, 1993, 5487.</p> <p>[3] S. A. Al-Muhtaseb, J. A. Ritter, Adv. Mater., 15, 2, 2003, 101.</p> <p>[4] K. Brodzik, M. Stolarski, J. Walendziewski, Wiadomości Chemiczne, 7-8, 2004, 58.</p>

Źródło: Baza BAZTECH, dostęp: 6.12.2016

Jak już wspomniano, na podstawie danych gromadzonych w takim układzie nie jest możliwe budowanie ścieżek cytowań stanowiących podstawę zaawansowanych analiz bibliometrycznych.

Podobne problemy można zauważyć w bazie BazTech, której przykładowe okno (przełądarki WWW) przedstawiono na rysunku 5.

Informacje o tym samym artykule, ale występującym w ramach bibliografii załącznikowej przedstawia rysunek 6. Widać tutaj wyraźnie, że dane na rysunku 5 różnią się od danych na rysunku 6 z pozycji [10] w atrybutach opisujących strony, czy numery czasopisma (choć reprezentują te same wartości, to są to jednak inne dane), co świadczy o występowaniu

anomalii typu/ $\alpha$ / oraz/ $\beta$ /. Na rysunku 6 „wklejono” również fragmenty opisów bibliograficznych dla innych artykułów (pozycje [61] i [79]).

Również tutaj widać różnice w zapisie, co świadczy o redundancji przechowywanych danych, czyli anomalii typu/ $\gamma$ /.

Rys. 6. Okno systemu BAZTECH wyświetlające informacje o bibliografii załącznikowej

Czasopismo	<b>Przegląd Włókienniczy - Włókno, Odzież, Skóra</b>	
Tytuł artykułu	<b>Nowy kierunek zastosowań aerożeli w balistyce</b>	
Autorzy	Litwa, P.	
Warianty tytułu	EN	New fields of applications of aerogels in ballistic
Języki publikacji	PL	
Abstrakty	EN	The paper presents a literature review considering methods of obtaining the carbon aerogels by polycondensation from resorcinol-formaldehyde (RF) precursor and discuss of research results for RF and crosslinked aerogels. The result of own, preliminary research of soft inserts for bullet- and fragment proof vests are presented. They clearly indicate that use of aerogel is effective and increase the ballistic protection. The article includes also through characterization of ability to absorb impact energy during static and dynamic compression tests as well as the ways to control the size and shape of nanoporous. Structure and porosity of porous materials have a significant impact on their properties.
Słowa kluczowe	PL	aerozele balistyka
Wydawca	Wydawnictwo SIGMA-NOT	
Czasopismo	Przegląd Włókienniczy - Włókno, Odzież, Skóra	
Rocznik	2012	
Tom	nr 12	
Strony	23--28	
Opis fizyczny	Bibliogr. 33 poz., fot., rys., tab.	
Twórcy	autor	Litwa, P. Instytut Technologii Bezpieczeństwa MORATEX w Łodzi
Bibliografia	<ol style="list-style-type: none"> <li>Fricke J., Tillotson T: Aerogels: production, characterization and applications, Thin Solid Films 297 (1997); 212 - 223.</li> <li>Zhang R., Li W., Li K.: Effect of concentration of reactants on porosity of hydrogels, organic and carbon aerogels, Microporous and Mesoporous Materials 72 (2004); 167 - 173.</li> <li>Feng J., Feng H., Jiang Y.: Ultrahigh density carbon aerogels with low thermal conductivity up to 2000 °C. Materials Letters 59 (2005); 1035 - 1038.</li> </ol>	
[61]	Kiciński W.: Aerozele węglowe otrzymywane z prekursora rezorcynowo- -furfuralowego. Biul. WAT 58 (4) (2009) 197÷221.	
[79]	W. Kiciński, Aerozele węglowe otrzymywane z prekursora rezorcynowo-furfuralowego, Biul. WAT, 58, 4, 2009, 197-221.	
	<ol style="list-style-type: none"> <li>Yan J., Li S., Yan L. i wsp.: Compressive behaviors and morphological changes of resorcinol-formaldehyde aerogel at high strain rates, Microporous and Mesoporous Materials 133 (2010); 134 - 140.</li> <li>Mirzaeian M., Hall P. J.: Preparation of controlled porosity carbon aerogels for energy storage in rechargeable lithium oxygen batteries, Electrochimica Acta 54 (2009); 7444 - 7451.</li> <li>Czakkel O., Szekeley E., Koczka B.: Drying of resorcinol-formaldehyde gels with CO<sub>2</sub>medium, Microporous and Mesoporous Materials 148 (2012); 34-42.</li> </ol>	
	10. Kiciński W.: Aerozele węglowe otrzymywane z prekursora rezorcynowo-furfuralowego, Biuletyn Wojskowej Akademii Technicznej, Vol. 58, 4(2009); 197-221.	

Źródło: Baza BAZTECH, dostęp: 6.12.2016

Omawiane problemy, choć do tej pory nie tak dokładnie skategoryzowane i powiązane z przyczynami ich powstawania, wynikającymi wprost z przyjętej organizacji przestrzeni informacyjnych, dostrzegane były już od jakiegoś czasu. Ich negatywne skutki rzutują wprost na podstawowe funkcje takich systemów (wyszukiwanie informacji) czyniąc je często niedoskonałymi i mało wiarygodnymi. Ograniczona wiarygodność zgromadzonych danych oraz przechowywanie ich w układzie tradycyjnym, mogą utrudniać, a nawet uniemożliwiać prowadzenie wiarygodnych i bardziej zaawansowanych analiz bibliometrycznych.

Władysław Kolasa, dla potrzeb realizacji bibliometrycznych badań z dziedziny humanistyki, zaproponował metodę „retrospektywnego indeksu cytowań”, którą z powodzeniem zastosował do zgromadzenia potrzebnych mu danych badawczych (Kolasa, 2011). Należy zauważyć, że, jak można domyślać się z opisu zamieszczonego w przytoczonym artykule, celem tej metody nie było zbudowanie „operacyjnej” bibliograficznej bazy danych, która po wdrożeniu mogłaby na bieżąco być zasilana nowymi strumieniami informacji, a jedynie

pozyskanie materiału badawczego o z góry założonym zakresie, drogą „kolejnych przekształceń danych” danych źródłowych pobranych z bazy bibliograficznej, a następnie ich transformacji i migracji do innych składnic danych. Stąd też zapewne wzięła się sama nazwa tej metody. Należy tutaj jednak zauważyć, że, mimo iż niektórzy badacze zaliczają to rozwiązanie do jednej grupy funkcjonalnej np. z bazą CYTBIN (Seweryn & Swoboda, 2014), to powstały „produkt” trudno już nazwać bibliograficzną bazą danych w powszechnym znaczeniu tego określenia, gdyż stanowi on jedynie materializację jej zawartości w wyznaczonym przez badacza okresie. Dodatkowo w bazie tej nie były gromadzone informacje o pozycjach cytowanych niebędących jednocześnie pozycjami cytującymi, gdyż zdaniem autora tej metody wyniki badań bibliometrycznych bez takich informacji będą zbliżone do wyników badań opartych na informacji pełnej.

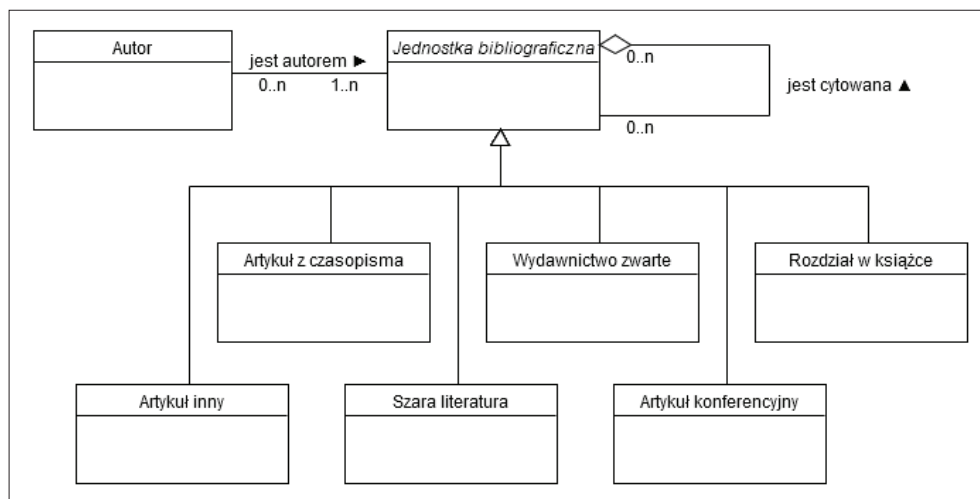
Autorka stała również przed problemem zgromadzenia obszernego wolumenu danych bibliograficznych, których pozyskanie możliwe było tylko z autopsji. Próby wykorzystania gotowych narzędzi od ogólnodostępnych pakietów biurowych do zaawansowanych systemów bibliotek cyfrowych nie zakończyły się zachęcającymi wynikami, ale dostarczyły wielu przemyśleń co do warunków niezbędnych do efektywnego wprowadzania danych o wymaganej jakości. Również próby automatyzacji tego procesu z zastosowaniem mechanizmów OCR okazały się mało skuteczne z powodu różnic w formatach i braku standaryzacji zapisów bibliografii. Stąd ostatecznie autorka przystąpiła do rozwiązania problemu zgromadzenia dużego zbioru danych w sposób odmienny. Nie zakładając z góry, że strumienie napływających danych ustaną w konkretnym momencie oraz chcąc zgromadzić pełną informację o danych zawartych w bibliografiach załącznikowych (co było konieczne chociażby do realizacji badań metodą powiązań bibliograficznych i współcytowań), zaproponowała metodę tworzenia Prospektywnego Bibliograficznego Indeksu Trawersowalnego – ProBIT. Dzięki tej metodzie zbudowana baza bibliograficzna może być zasilana w sposób ciągły (równoległe z realizacją badań bibliometrycznych), a ponieważ gromadzi pełny zakres danych bibliograficznych w układzie relacyjnym, może służyć jednocześnie jako podstawowy system informacyjny. Dodatkowo w dowolnym momencie cyklu życia systemu możliwe jest wykonywanie analiz bibliometrycznych, nawet tych bardziej zaawansowanych i wymagających trawersowania po ścieżkach cytowań, bez potrzeby dodatkowych migracji danych. Nazwę swoją metoda bierze stąd, że (posługując się nomenklaturą algebry relacji używaną w pracy wyżej wspomnianego autora) „klucze relacji” pomiędzy jednostkami cytującymi a cytowanymi określane są już w momencie pierwszego zaistnienia informacji o danym cytowaniu jednostki cytowanej i powinny być one użyte w przyszłości, w przypadku pojawienia się innych źródeł cytujących tę samą jednostkę cytowaną.

#### **4. Propozycja metody tworzenia prospektywnego trawersowalnego indeksu bibliograficznego**

Koncepcyjny model klas opisujący organizację przestrzeni informacyjnej dla przedstawianej metody zamieszczono na rysunku 7. U jego podstaw leży obiektowe postrzeganie rzeczywistości, nakazujące dostrzeganie pewnych abstrakcji w postaci związków generalizacji-specjalizacji (a nie tylko związków asocjacji, jak ma to miejsce w modelu tradycyjnym).

Widzimy tutaj, że opisywana jednostka bibliograficzna jest klasą abstrakcyjną co oznacza, że bez jej doprecyzowania („ukonkretnienia”) jej byty nie mogą zaistnieć samodzielnie, tj. jako konkretne obiekty (ang. *concrete objects*). Klasy potomne (np. Artykuł z czasopisma, Wydawnictwo zwarte itp.) dziedziczą wszystkie właściwości tej klasy bazowej, a jednocześnie rozszerzają ją o atrybuty specyficzne dla tych klas.

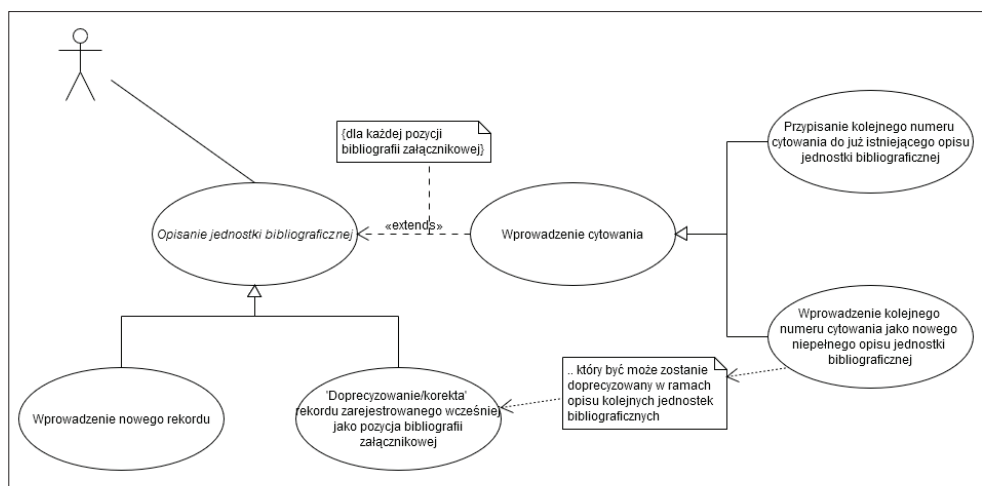
Rys. 7. Model klas w metodzie prospektywnej



Z modelu klas opracowanego dla metody prospektywnej wynika, że potencjalnie opisywaną jednostką bibliograficzną może być jednostka o dowolnej formie wydawniczej. Klasa jednostki bibliograficznej jest związana sama ze sobą relacją agregacji o nazwie „jest cytowana”, a kardynalności tej relacji zdefiniowano na „0..n”. Oznacza to, że bibliografia załącznikowa danej (konkretnej) jednostki bibliograficznej stanowi agregat (zbiór, który w szczególnym przypadku może być zbiorem pustym) innych (konkretnych) jednostek bibliograficznych, zaś dana jednostka bibliograficzna może wchodzić w skład wielu takich agregatów. Tak określony model, mimo że możliwy do zaimplementowania w postaci relacyjnej, może reprezentować całą sieć (a dokładniej skierowany graf) cytowań, czyniąc struktury danych zgodnie z nim zaimplementowane łatwo trawersowalnymi. Dodatkowo wyróżniono na diagramie klasę Autor, która związana jest z jednostką bibliograficzną relacją (0..n,1..n) co oznacza, że każdy autor rejestrowany w systemie powinien być autorem przynajmniej jednej jednostki bibliograficznej, zaś dana jednostka bibliograficzna nie musi mieć przypisanego autora (np. przypadki dokumentów z nieznanym autorem) lub może mieć przypisanego jednego lub więcej autorów. Taki zapis sugeruje, że autorzy są samodzielnymi bytami w systemie (a nie jedynie atrybutami tekstowymi jednostki bibliograficznej), zaś relacja „jest autorem” stanowi tylko wskazanie od konkretnej jednostki bibliograficznej do konkretnych autorów. W ogólnych rozważaniach takich samodzielnnych bytów pojawić się może więcej np. instytucja sprawcza, wydawca, itp., jednak powoływanie ich na potrzeby opisu przedstawianej metody zaciemniłoby tylko obraz. Próby przedstawienia całego zakresu informacyjnego systemu bibliograficznej bazy danych mogącego stanowić „model referencyjny” będą przedmiotem rozważań dalszych prac badawczych.

Na rysunku 8 przedstawiono model przypadków użycia charakteryzujący metodę prospektywną. Podobnie jak w modelu tradycyjnym, podstawową czynnością jest opisanie jednostki bibliograficznej z tym, że tutaj, jako że czynność ta może być realizowana na dwa wykluczające się wzajemnie sposoby, jest ona reprezentowana przez abstrakcyjny przypadek użycia.

Rys. 8. Model organizacji pracy dla rejestracji jednostek bibliograficznych metodą prospektywną



Może on być realizowany poprzez wprowadzenie nowego rekordu opisu lub poprzez doprecyzowanie (a w niektórych przypadkach wręcz skorygowanie) informacji o już istniejącej jednostce bibliograficznej, która wcześniej zaistniała w systemie jako element bibliografii załącznikowej. W trakcie opisywania jednostki bibliograficznej następuje rejestrowanie informacji o poszczególnych pozycjach bibliografii załącznikowej. Różnica w stosunku do modelu tradycyjnego polega na tym, że operuje się tutaj nie na rekordach cytowań, ale na obiektach jednostek bibliograficznych. Możliwe jest więc przypisanie do konkretnego numeru pozycji bibliografii załącznikowej jednostki bibliograficznej wcześniej opisanej lub też zarejestrowanie nowej, która być może w trakcie realizacji kolejnych opisów jednostek bibliograficznych będzie doprecyzowana i stanie się jednostką cytującą – wyraża to odpowiednia notatka zamieszczona na modelu.

Dopowiedzenia wymaga wcześniejszy komentarz o doprecyzowaniu lub wręcz skorygowaniu informacji podczas realizacji przypadku użycia „Doprecyzowanie/korekta”. Wydaje się oczywiste, że informacje o jednostce bibliograficznej, którą właśnie opisujemy (a jeśli robimy to z autopsji, to „mamy ją właśnie w ręce”) są bardziej wiarygodne od informacji zamieszczonych o tej jednostce w bibliografiach załącznikowych innych jednostek. Dlatego jest to najlepszy moment na weryfikację wprowadzonych wcześniej informacji i ich ewentualną korektę. Sytuacja jest mniej oczywista w przypadku rejestrowania informacji o danej jednostce przy okazji realizacji przypadku użycia „Przypisanie kolejnego numeru cytowania do już istniejącego opisu jednostki bibliograficznej”, ale istniejący już w systemie opis tej jednostki „kandydującej” różni się nieco od opisu w bibliografii załącznikowej.

Jeśli opis tej jednostki bibliograficznej został wprowadzony bądź wcześniej skorygowany w ramach przypadku użycia „Opisanie jednostki bibliograficznej”, to dane te oczywiście możemy uznać za wiarygodne. W przeciwnym przypadku nie mamy pewności, czy zarejestrowany wcześniej opis jest bardziej wiarygodny od posiadanego aktualnie. Jest to więc dobry moment na sięgnięcie do innego, bardziej wiarygodnego źródła lub oznaczenie rekordu jako „podejrzanego” i wymagającego późniejszej weryfikacji. Tak czy inaczej zidentyfikowanie takich niezgodności w momencie ich rejestracji zawsze daje większą szansę na ich wyjaśnienie, gdyż mamy „przed sobą” materiał źródłowy, niż wykrywanie ich i korygowanie *post factum*.

Skuteczność wyszukiwania „kandydujących” rekordów spośród zarejestrowanych wcześniej zależy od przyjazności graficznego interfejsu użytkownika i efektywności jego „mechanizmów sugestii”, podpowiadających operatorowi możliwość ponownego użycia informacji o wcześniej zarejestrowanej jednostce bibliograficznej. Skuteczne mechanizmy nie powinny opierać się jedynie na wyszukiwaniu „dosłownym”, ale powinny implementować bardziej elastyczne algorytmy dopasowania wzorców, np. oparte na miarach podobieństw łańcuchów znakowych, jak np. metody zaproponowane w (Kamińska, 2017a).

Przedstawiony model metody prospektywnej jest modelem obiektowym (wykorzystuje związki generalizacji-specjalizacji). Natomiast większość popularnych systemów zarządzania bazami danych posługuje się modelami relacyjnymi, gdzie nie ma dosłownej implementacji wyżej wspomnianych związków. Zachodzi wobec tego konieczność transformacji modelu obiektowego do modelu relacyjnego. Zadanie to jest jak najbardziej możliwe, co więcej, znanych jest tutaj kilka podejść, których wybór powinien zależeć od wymagań funkcjonalnych i niefunkcjonalnych stawianych budowanemu systemowi (Keller, 1997).

Autorka, budując własną bibliograficzną bazę danych, stanęła przed koniecznością podjęcia wielu decyzji dotyczących architektury gromadzonych informacji, które ostatecznie doprowadziły do powstania logicznego modelu relacyjnego, a następnie modelu danych możliwego do wdrożenia na konkretnej platformie systemu zarządzania bazą danych. Tak powstały system, będący implementacją koncepcji przedstawionej w niniejszym artykule, dostarczył w relatywnie krótkim czasie materiału badawczego o wysokiej jakości. Oznacza to, że dzięki opisanej metodzie można nie tylko zapewnić wysoką jakość gromadzonych danych i rozpocząć na ich podstawie realizację zaawansowanych badań bibliometrycznych, ale również obniżyć ogólne koszty bieżącego zasilania tradycyjnych bibliograficznych baz danych.

## 5. Wnioski

Artykuł przedstawia problemy z jakimi współcześnie borykają się operatorzy krajowych, zarówno technicznych jak i humanistycznych, bibliograficznych baz danych gromadzących informacje w układzie tradycyjnym. Na konkretnych przykładach wykazano, że niespójności w danych oraz brak możliwości realizacji zaawansowanych badań bibliometrycznych, które bazowałyby na tych danych, nie wynika jedynie z niefrasobliwości operatorów zasilających te bazy, a przede wszystkim z przyjętej organizacji przestrzeni informacyjnej. Zmiana takiego stanu rzeczy nie będzie możliwa poprzez wprowadzanie kolejnych dyrektyw czy zaleceń dla wyżej wspomnianych operatorów, a jedynie poprzez reorganizację modelu danych



i oprogramowania na nim operującego, np. zgodnie z prospektywną metodą tworzenia trawersowalnych indeksów cytowań.

Zaproponowaną w artykule metodę, na tle innych opisywanych w piśmiennictwie badawczym czy implementowanych i stosowanych w bibliograficznych bazach danych/indeksach cytowań, wyróżniają ważne cechy:

- dane gromadzone są bezpośrednio w relacyjnej bazie danych, ale w układzie pozwalającym na łatwą reprezentację tych danych w postaci grafów cytowań o trawersowalnych krawędziach;
- system może realizować funkcje bibliograficznej bazy danych poprzez bieżące gromadzenie, składowanie, przetwarzanie i udostępnianie danych bibliograficznych;
- system może realizować funkcje indeksu cytowań dostępnego wprost z poziomu relacyjnej bazy danych, która może stanowić zarazem podstawowe środowisko analityczno-obliczeniowe;
- funkcje bibliograficznej bazy danych oraz indeksu cytowań mogą być realizowane współbieżnie, bez potrzeby „zamrażania” czy „materializacji” danych przed rozpoczęciem analiz;
- prospektywny sposób wprowadzania danych oszczędza czas operatora, a jednocześnie sprzyja gromadzeniu danych o bardzo dobrej jakości;
- brak potrzeby migracji danych w różnych formatach plików, pomiędzy różnymi platformami w tym brak potrzeby posługiwania się formatem MARC (w środowisku zbudowanym przez autorkę format MARCXML służył jedynie fakultatywnemu udostępnieniu danych dla innych systemów);
- opis pozycji cytujących oraz cytowanych przechowywany jest w jednej strukturze danych (tabeli), co pozwala uniknąć redundancji przechowywanych informacji.

Dalsze badania autorki w tych obszarach koncentrować się będą na opracowaniu ogólnych obiektowych i relacyjnych modeli referencyjnych, transformacji zastanych modeli danych do modeli umożliwiających zastosowanie metody prospektywnej oraz udoskonalaniu funkcji oceniającej podobieństwo rekordów bibliograficznych, która może znaleźć zastosowanie w implementacji przypadku użycia „Przypisanie kolejnego numeru cytowania do już istniejącego opisu jednostki bibliograficznej”.

Na koniec warto zauważyć, że z przedstawionych treści wynika, iż możliwości informacyjne czy analityczne danych systemów bibliograficznych nie zależą jedynie od „zakresów gromadzonych danych”, na których opisie najczęściej poprzestają badacze charakteryzując dane rozwiązanie. Potencjał niedopuszczania do powstawania opisanych wcześniej anomalii oraz możliwości realizacji zaawansowanych obliczeń bibliometrycznych wynika natomiast przede wszystkim z architektury składowanych informacji w poszczególnych systemach, których opis mogłaby ułatwić zaproponowana w artykule notacja. A ponieważ trudno w literaturze znaleźć relewantne informacje na temat opisu poszczególnych systemów z tej perspektywy, mogłyby one prawdopodobnie stanowić wkład w rozwój badań nad tą dziedziną.

## Bibliografia

- Al, U.; Taşkın, Z.; Düzyol, G. (2012). Use of Social Network Analysis in Bibliometric Researches. In: *Information on E-motion. Proceedings. BOBCATSSS 2012, 20th International Conference on Information Science, 23–25 January*. Amsterdam: Bock+Herchen Verlag, 40–44.
- Booch, G.; Rumbaugh, J.; Jacobson, I. (2002). *UML przewodnik użytkownika*. Warszawa: Wydaw. Naukowo-Techniczne.
- Cockburn, A. (2004). *Jak pisać efektywne przypadki użycia*. Warszawa: Wydaw. Naukowo-Techniczne.
- Drabek, A. (2009). Konferencja „Bibliograficzne bazy danych: kierunki rozwoju i możliwości współpracy”, Bydgoszcz, 27–29 maja 2009 roku. *Bibliotheca Nostra*, 2(18), 116–121.
- Garczyńska, M. (2013). Bibliograficzne bazy danych w ocenie pracowników i jednostek szkół wyższych w Polsce [online]. *Materiały konferencyjne EBIB*, 24, [06.12.2016], [http://open.ebib.pl/ojs/index.php/Mat\\_konf/article/view/24/164](http://open.ebib.pl/ojs/index.php/Mat_konf/article/view/24/164).
- Garfield, E. (1955). Citation Indexes for Science: A New Dimension in Documentation through Association of Ideas. *Science*, 122(3159), 108–111.
- Kretschmer, H.; Kretschmer, T. (2007). A New Centrality Measure for Social Network Analysis Applicable to Bibliometric and Webometric Data. *COLLNET Journal of Scientometrics and Information Management*, 1(1), 1–7.
- Ilyas, I. F.; Chu X. (2015). Trends in Cleaning Relational Data: Consistency and Deduplication. *Foundations and Trends in Databases*, 5(4), 281–393.
- Kamińska, A. M. (2017a). Miary podobieństw łańcuchów znakowych a deduplikacja rekordów w bibliograficznych bazach danych. *Przegląd Biblioteczny*, 4 [w druku].
- Kamińska, A. M. (2017b). O rozwoju graficznych języków komunikacji. *Zagadnienia Informatyki Naukowej*, 2(110) [w druku].
- Kamińska, A. M. (2017c). Potencjał bibliometryczny bibliograficznej bazy danych CYTBIN w świetle prostych i złożonych wskaźników analitycznych. *Bibliotheca Nostra* [w druku].
- Kamińska, A. M. (2017d). Tam, gdzie zaczyna się bibliometria, czyli jak pozyskać materiał analityczny z autopsji [online]. *Biuletyn EBIB*, 3(173), [28.08.2017], <http://open.ebib.pl/ojs/index.php/ebib/article/view/534/698>.
- Kamińska, A. M. (2017e). Wizualizacje wybranych wskaźników bibliometrycznych na przykładzie bibliograficznej bazy danych CYTBIN. *Toruńskie Studia Bibliologiczne*, 2 (19) [w druku].
- Kamińska, A. M. (2018). Zastosowanie metod analizy sieci społecznościowych w bibliometrii i webometrii. Miary i narzędzia. *Nowa Biblioteka. Usługi, technologie informacyjne i media*, 2 (29) [w druku].
- Kolasa, W. M. (2011). Retrospektywny indeks cytowań w humanistyce. Koncepcja, metoda, zastosowanie. *Przegląd Biblioteczny*, 79(4), 466–486.
- Marszakowa-Szajkiewicz, I. (2009). *Badania ilościowe nauki. Podejście bibliometryczne i webometryczne*. Poznań: Uniwersytet im. Adama Mickiewicza.
- Seweryn, A.; Swoboda, I. (2014). Cytowania w polskich bazach bibliograficznych. *PTINT Praktyka i Teoria Informatyki Naukowej i Technicznej*, 1(85), 3–20.
- Keller, W. (1997). Mapping Objects to Tables: A Pattern Language. *Proceedings of the 1997 European Conference on Pattern Languages of Programming (EURO PLoP'97)* [online], Software Research and Industry, Dirk Riehle, [06.12.2016], <http://dirkriehle.com/community-service/hillside-group/europlop-1997/p13final.pdf>.
- Woźniak-Kasperek, J. (2011). *Wiedza i język informacyjny*. Warszawa: Wydawnictwo SBP.

---

## ProBIT – the Prospective Method of Creating the Traversable Citation Indices versus the Contemporary Issues in Information Space Organization in Traditional Bibliographic Databases

### Abstract

**Purpose/Thesis:** The article presents an original method of building bibliographic databases, conducive to the efficient gathering of high quality bibliographic data which, without any further transformations, can serve as a source of advanced bibliometric analysis (i.e. not limited to simple summaries calculated on the basis of data collected in a relational form, but those that need traversing the edges of the graph representing the networks of article citations and cooperation among their authors).

**Approach/Methods:** Encountering a problem with gathering a significant volume of bibliographic data to conduct her research, the author evaluated a number of tools used to collect the bibliographic data, but none of them offered the expected ergonomics, sufficient care about the quality of the data collected, or data model enabling further advanced bibliometric analysis. Drawing conclusions from the negative experience gained while working with these tools, the author proposed her own original method, which is presented using basic UML notation.

**Results and conclusions:** The presented method has proved effective and helped the author – in a relatively short time – to gather data of very good quality on more than 36 thousand articles and over 22 thousands of authors. These data were used directly for the implementation of bibliometric research, which will be discussed in a separate work.

**Practical implications:** First part of the article presents risks brought by gathering bibliographic data in so-called traditional form. Using examples from selected national bibliographic databases the author proved that these risks appear as a number of anomalies that have been categorized by the author, and which can be prevented by the proposed method. The method in question enablesto increase the quality of data collected in the bibliographic databases and reduce the cost of data gathering by improving the entire process.

**Originality/Value:** Polish researchers in the area of book studies rarely pursue the subject of constructing bibliographic databases and usually limit this issue to defining only 'sets of entities and attributes' being processed. The author argues that the final result (the quality of data, the total cost of data collecting, the possibility of further analysis) is also affected by the data model (following which the data are collected) and the appropriate organization of the collecting process.

### Keywords

Bibliographic database. Information space. ProBIT. Prospective method. Traversable citation index.

---

*Dr ANNA MAŁGORZATA KAMIŃSKA jest pracownikiem Biblioteki Głównej Politechniki Śląskiej w Gliwicach i absolwentką Wydziału Filologicznego Uniwersytetu Śląskiego w Katowicach. W 2016 r. obroniła rozprawę doktorską pt. Informacja naukowa o górnictwie w świetle wydawnictw ciągłych uczelni technicznych w Polsce (1945–1989) na Wydziale Filologicznym Uniwersytetu Śląskiego. Jej zainteresowania naukowe skupiają się wokół trzech uzupełniających się obszarów: informatologia, graficzne języki komunikacji oraz wizualizacja informacji.*

*Kontakt z autorką:*

*Anna.Kaminska@polsl.pl*

*Biblioteka Główna Politechniki Śląskiej*

*ul. Kaszubska 23*

*44–100 Gliwice*