

# ROZWÓJ METOD MAPOWANIA DOMEN NAUKOWYCH I POTENCJAŁ ANALITYCZNY W NIM ZAWARTY

Veslava Osińska  
Uniwersytet M. Kopernika w Toruniu

*Wizualizacja informacji, mapowanie domen naukowych, analiza współcytowań,  
sieci społeczne*

## Historia wizualizacji domen naukowych

Mapowanie i wizualizacja domen naukowych jako metodyka badań zdobywa coraz większe zainteresowanie wśród specjalistów informacji naukowej i naukoznawstwa. Doprowadziły do tego co najmniej dwa czynniki. Pierwszy to cyfryzacja zasobów piśmiennictwa naukowego i innych źródeł wiedzy i co za tym idzie, przyrost i rozwój bibliograficznych baz danych dostępnych online, takich jak np.: ISI Web of Knowledge, Scopus, Google Scholar. Drugim jest nieliniowy wzrost mocy obliczeniowej komputerów, wykorzystywanie superkomputerów i technologii rozproszonych [6, 21].

Aby lepiej zrozumieć strukturę i dynamikę rozwoju poszczególnych działów nauki oraz znaleźć sposób identyfikacji w nich trendów tematycznych, naukowcy analizują literaturę naukową i ścieżki cytowań poszczególnych publikacji. Analiza cytowań jest znana od dawna z bibliometrii. Termin **wizualizacja dyscyplin wiedzy** lub **nauki** (*KDViz – Knowledge Domain Visualization*) wymyślony został przez Ch. Chena w 2001 r., redaktora czołowego czasopisma w zakresie wizualizacji informacji *Information Visualization*<sup>1</sup> (powszechny skrót: *InfoViz*). Ponieważ metody te prowadzą do generowania map graficznych, równolegle używana jest inna nazwa: **mapowanie nauk** (*Mapping Science*) lub rzadziej naukografia (*scientography*) [8]. Tę ostatnią wprowadził w 1960 roku E. Garfield - założyciel Instytutu Informacji Naukowej w Filadelfii (*ISI*). Na przestrzeni ostatniego 10-lecia badania wizualizacji domen wiedzy zostały rozszerzone do zadań wyszukiwania informacji [10].

Nad rozwojem technik i metod wizualizacji pracują interdyscyplinarne zespoły informatyków, programistów od projektowania interakcji człowiek – komputer (*Human Computer Interface – HCI*), analityków, naukoznawców, grafików i plastyków. Warto tu wspomnieć o projekcie Wirtualne Studio Wiedzy (*Virtual*

<sup>1</sup> <http://www.palgrave-journals.com/ivs/index.html>.

*Knowledge Studio – VKS*)<sup>2</sup>, który niedawno powstał w Holenderskiej Akademii Nauk i wspiera naukowców nauk humanistycznych i społecznych w Holandii w opracowywaniu nowych rozwiązań naukowych. Celem jest tworzenie metafo-rycznych modeli struktur nauk humanistycznych i społecznych. Główną cechą takiego centrum jest integracja wizualizacji i analizy na tle współpracy badaczy nauk społecznych, humanistycznych oraz informatyków.

O atrakcyjności takich graficznych zasobów wiedzy dla odbiorców przesąd-zaają nie tylko walory estetyczne, lecz także aktualność, potencjał analityczno-merytoryczny oraz dostępność online. Dlatego nie bez powodu wśród prakty-ków *KDViz* istnieje tendencja do szybkiej publikacji wyników badań w sieci. Na przykład, wirtualna wystawa *Places@Spaces*<sup>3</sup> jest ogólnościatowym projektem, ukierunkowanym na studiowanie infrastruktury nauki w skali globalnej w celu zademonstrowania osiągnięć w dziedzinie mapowania informacji o roku 1998. Portal redagowany jest przez K. Börner, *School of Library and Information Science, Indiana University*, której zespół prowadzi zaawansowane badania nad metodami wizualizacji dziedzin nauki, obszarów współpracy naukowców oraz bibliotek cyfrowych. Zamieszczane tam na bieżąco mapy domenowe, kartograficzne i pojęciowe mają na celu inspirowanie interdyscyplinarnych dys-kusji, umożliwienie komunikacji pomiędzy działalnością człowieka i postępem naukowym. Coraz więcej w sieci udostępnia się aplikacje do wizualizacji wła-snych, niekoniecznie akademickiego przeznaczenia, danych w standardowych formatach (np. ASCII, CSV), zachęcając dociekliwych użytkowników do testów rozmaitych konfiguracji graficznych, od wykresów bąbelkowych i rozproszonych (*skater chart*) począwszy, na chmurkach tekstowych kończąc (*text cloud*). Takim narzędziem jest eksplorator autorstwa IBM *ManyEyes*<sup>4</sup>, gdzie użytkownicy mają możliwość nie tylko wizualizacji własnych danych przygotowanych w od-powiednich formatach, lecz i wytypowania najlepszych projektów.

## Proces wizualizacji

### A. Elementy składowe

W ponad 10-letniej historii wizualizacji informacji na czoło wysunęło się kilka zespołów badawczych specjalizujących się w metodach wizualizacji wie-dzy i nauki. Najbardziej znaczące osoby w *Infoviz*, kierujące tymi zespołami i występujące w większości źródeł cytowań, zostały wymienione w poprzednim rozdziale.

W każdej analizie wykorzystującej mapy wizualizacji zawarte są trzy podsta-wowe składniki: obiekty analizy, relacje pomiędzy nimi oraz metoda mapowania. Wcześniejsze metody wykorzystywały najprostsze grafy, gdzie węzły (*nodes*) reprezentowały obiekty badań, a krawędzie (*edges, links*) – związki pomiędzy nimi. Im bliżej usytuowane były obiekty, tym bardziej były do siebie podobne pod względem wybranej miary (kierunki badań, źródła cytowań, wspólni auto-

<sup>2</sup> <http://virtualknowledgestudio.nl/aboutvks/>.

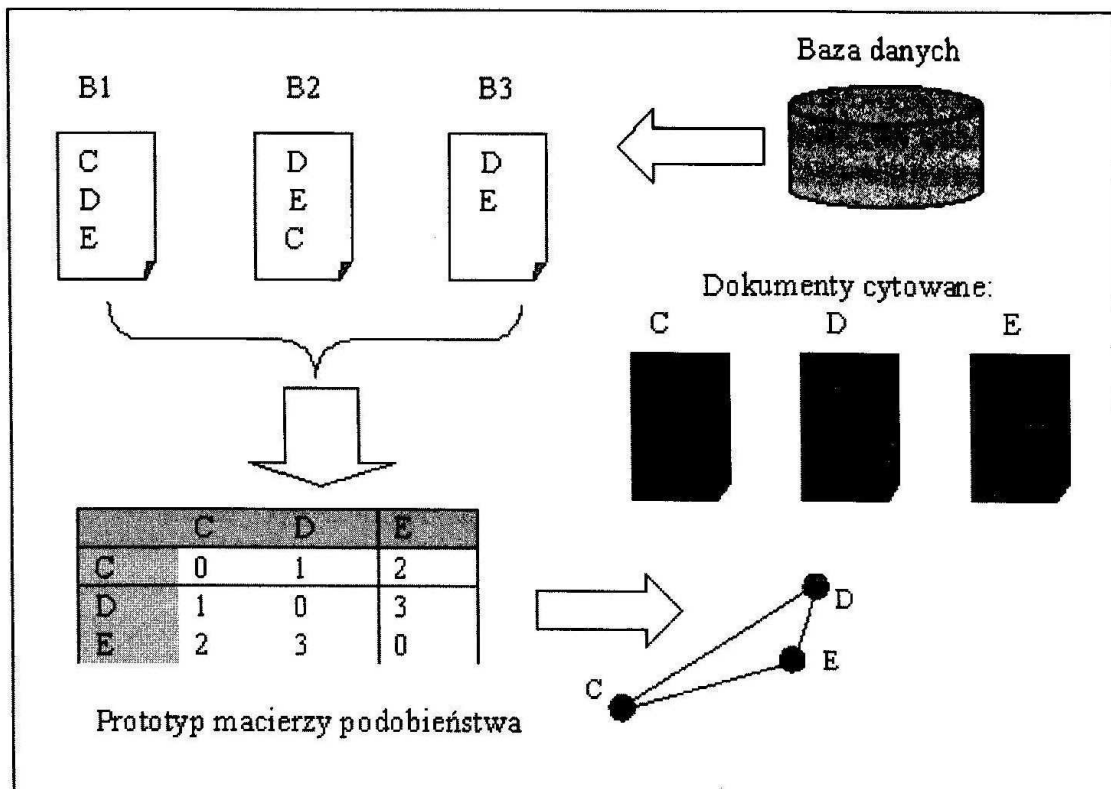
<sup>3</sup> <http://scimaps.org>.

<sup>4</sup> <http://manyeyes.alphaworks.ibm.com/manyeyes/>.

rzy itp.). Później zaczęto opierać się na algorytmach reprezentujących wyniki w postaci bardziej kognitywnej, bo dwu- lub trójwymiarowej sieci, z wyeksponowaniem różnych poziomów abstrakcji. Krótka historia *KDViz* i interdyscyplinarny charakter tej dyscypliny badawczej utrudniają systematykę metodologii badań.

Ponieważ tradycyjne podejście do wyboru danych analizy nie zmienia się od lat, można spróbować podzielić sposoby mapowania ze względu na obiekty analizy i miary podobieństwa.

Wspólne cytowanie autorów (*ACA – Author Co-citation analysis*) lub artykułów (*DCA – document co-citation analysis*) to metoda używana w naukometrii od lat 70., pierwsza, na której opierały się technologie wizualizacyjne [16]. Dwa dokumenty (autorzy) są wspólnie cytowane, jeśli równocześnie występują w wykazie bibliograficznym trzeciego dokumentu. Podobieństwo pomiędzy dokumentami jest obliczane na podstawie częstości ich wspólnych cytowań. Uproszczony schemat wizualizacji na podstawie źródeł cytowań przedstawia rys. 1. Wizualizacja za pomocą *DCA* charakteryzuje dynamikę domeny wiedzy, natomiast *ACA* pozwala na wnioskowanie o strukturze intelektualnej obszaru badawczego [2, 5]. O wiele rzadziej za miarę podobieństwa przyjmuje się wspólne pozycje bibliograficzne w dwóch różnych dokumentach – tzw. bibliografie sprzężone (*bibliographic coupling*).



Rys. 1. Schemat analizy i mapowania wspólnego cytowania dokumentów. Dokumenty DE mają największe podobieństwo, dokumenty CD – najmniejsze.

Wspólne występowanie pojedynczych słów lub sekwencji w tytułach, abstraktach lub treści publikacji. Mogą to być słowa kluczowe, deskryptory tema-

tyczne, terminy główne, hasła przedmiotowe lub najważniejsze wyrażenia i frazy w tekście. Nie ma tu potrzeby używania programów do analizy częstości występowania i współwystępowania słów, można to bowiem wykonać za pomocą albo własnych aplikacji, albo ogólnie dostępnych narzędzi, chociażby arkusza kalkulacyjnego EXCEL. Takim graficznym reprezentacjom opartym na wspólnej terminologii trochę na wyrost nadano nazwę mapy semantycznej, ponieważ w założeniu ma ona pomagać w zrozumieniu kognitywnej struktury danej domeny naukowej. Typowym przykładem autentycznej mapy semantycznej jest wizualna reprezentacja tezaursusa (Visual Thesaurus<sup>5</sup>). W wizualizacji słów kluczowych istotny jest algorytm wyszukujący semantycznie podobne terminy, w przypadku różnotematycznej bazy danych występuje wysoka przypadkowość.

W mapowaniu nauk używa się także kolekcji poklasyfikowanych dokumentów. Liczba wspólnych artykułów naukowych w przypisanych kategoriach tematycznych przekłada się na podobieństwo semantyczne ich tematów. Mapowanie dokumentów na podstawie wspólnych klas lub kategorii tematycznych daje najbardziej prawidłowy obraz organizacji przedmiotowej danej domeny naukowej. Jednak prac wykorzystujących tę metodę jest niewiele [12, 17], zapewne z powodu braku danych o klasyfikacji w bazach bibliograficznych.

Jak się można domyślić, większość jednostek analizy pochodzi z bibliograficznych baz danych, dlatego przeważająca część publikacji na temat wizualizacji domen naukowych to zmapowane kolekcje danych baz przedmiotowych ISI Web of Knowledge. W pracach [16, 17] podano przykład zmapowania uniwersum klasyfikacyjnego w dziedzinie nauk komputerowych oraz zaproponowano metodę modernizacji pierwotnego schematu klasyfikacji ACM CCS. Bibliograficzne dane dokumentów naukowych nie wystarczą, jeśli obiekty analizy wyłaniają się z treści artykułów. Oprócz dostępu do pełnotekstowych zasobów naukowych niezbędne jest również zastosowanie odpowiednich algorytmów przetwarzania języka naturalnego.

Na etapie rozwoju Web 2.0 popularne jest mapowanie zasobów wiedzy pochodzących z ogólnodostępnych serwisów sieciowych. Na przykład mapa artykułów angielskiej Wikipedii wygenerowana została przy użyciu miary rejestrującej wspólne kategorie artykułów [11]. Aktywność użytkowników serwisów specjalistycznych również inspirowała do analizowania i mapowania takich danych jak blogi i fora internetowe, logi użytkowników [1] itd. Dotychczasowe metody wizualizacji informacji zostały z powodzeniem zaadaptowane w środowisku sieciowym, gdzie strony internetowe z pewnym przybliżeniem można traktować jako artykuły naukowe, a hiperłącza jako powiązania cytowań. Takie podejście wspomogło szybki rozwój webometrii [21].

## **B. Etapy procesu wizualizacji**

Z praktycznego punktu widzenia model mapowania informacji tekstowej – a z taką mamy do czynienia w analizie treści publikacji naukowych – obejmuje następujące etapy [2,4]:

---

<sup>5</sup> [www.visualthesaurus.com](http://www.visualthesaurus.com)

Wybór odpowiedniej bazy danych: bibliograficzne bazy danych oparte na modelu meta, a więc zawierające metadane, np. tytuł, autora, datę publikacji, słowa kluczowe. Dodatkowym atutem są informacje bibliometryczne, np. liczba cytowań, współpracownicy autora, miara oddziaływania czasopisma itp., ułatwiające ustalenie związków pomiędzy badanymi obiektami.

Wytypowanie i obróbka jednostek analizy, którymi mogą być artykuły, czasopisma, autorzy, słowa kluczowe, określniki tematyczne.

Wybór stosownej miary podobieństwa, np. wspólne pary cytowań, wspólne kategorie, współautorzy itp. oraz obliczenie częstości współwystępowania wraz z wyznaczeniem progu zliczeń przypadków, poniżej którego analiza staje się niemiarodajna.

Stworzenie czytelnego rozkładu danych, dającego się zmapować na przestrzeń 2D lub 3D. Posługując się skalarnym modelem obliczania (liczony iloczyn skalarny jako podstawowa miara podobieństwa) wartości podobieństwa dla każdej pary obiektów, uzyskuje się macierz podobieństwa. Wartości macierzy odwzorowują odległości topologiczne pomiędzy danymi w wybranej przestrzeni; za pomocą odpowiednich algorytmów grupujących analizuje się powstające klastry obiektów. <http://www.ebib.info/publikacje/matkonf/mat19/osiewalska.Php> – 10. Można też zastosować wektorowy model przestrzeni informacyjnej, np. analizę ukrytych grup semantycznych (*Latent Semantics Analysis – LSA*) [15]. Ponieważ na tym etapie wielowymiarowy charakter danych nie pozwala na ich zobrazowanie, następnym krokiem musi być redukcja wymiarów do dwóch lub trzech. Używa się tu technik statystycznych, takich jak analiza czynnikowa, analiza głównych składowych, skalowanie wielowymiarowe (*Multidimensional Scaling - MDS*). Do inteligentnego rozmieszczania wielowymiarowych danych wykorzystuje się również sztuczne sieci neuronowe, mapy samoorganizujące się (*Self Organizing Map – SOM*) lub algorytmy klasteryzacji [2, 5, 16].

Zastosowanie aplikacji do wizualizacji i wyświetlenie informacji w sposób umożliwiający użytkownikowi wygodną obserwację. Współczesne interfejsy wizualizacyjne zapewniają przejrzystą i interaktywną eksplorację danych, umożliwiając szerokokątne przeglądanie, powiększanie, filtrowanie, wyszukiwanie i wyszczegółowienie wybranych fragmentów danych.

Walidacja statystyczna w celu ewaluacji i doboru odpowiednich parametrów początkowej miary podobieństwa. Nieznaczna zmiana warunków początkowych może całkowicie zmodyfikować końcowy rozkład danych. Przy całej różnorodności metod i technik wizualizacji informacji problemem pozostaje ewaluacja końcowego rozkładu wizualizacji. Dlatego do oceny wyników wizualizacji podchodzi się empirycznie: sprzężenie zwrotne „parametry początkowe – wyniki” pozwala na wygenerowanie najbardziej dokładnej mapy wizualizacji.

## Oprogramowanie do wizualizacji

Proces wizualizacji wymaga rozwiązań programistycznych na co najmniej dwóch poziomach. Dysponując bazą danych, analityk staje przed problemem ekstrakcji relewantnych informacji, takich jak cytowania, słowa kluczowe, abstrakty artykułów w badanej kolekcji. Następnie należy umiejętnie obliczyć przy-

padki współwystępowania jednostek analizy i wypełnić macierz podobieństwa tymi wartościami, po uprzednim przeskalowaniu i normalizacji (konieczna jest automatyzacja tej procedury). Odpowiedniego programu wymaga mapowanie danych na przestrzeń obserwacji, a przedtem redukcja wymiarów. Tworzenie takiego oprogramowania i jego wdrożenie jest czasochłonne, dlatego większość badaczy *KDViz* korzysta z gotowych programów lub modułów softwarowych typu *Open Source* [3,6,20].

Do obróbki danych bibliograficznych i bibliometrycznych korzysta się z oprogramowania:

- Bibexcel<sup>6</sup> – narzędzie mające wiele zastosowań bibliometrycznych autorstwa szwedzkiego naukowca O. Perrsona, założyciela dziedziny BIN w Szwecji. Akceptowalny format danych – *Dialog*,

- CiteSpace<sup>7</sup> – aplikacja stworzona przez Ch. Chena do rozmaitych analiz bibliometrycznych oraz wizualizacji wyników. Współpracuje z formatami danych pobieranych ze strony Web of Science; aktualnie opracowywane formaty baz Scopus i Google Scholar.

Oprogramowanie do mapowania i wizualizacji:

- CiteSpace,

- Pajek<sup>8</sup> – słoweński darmowy, ceniony przez profesjonalistów program do analizy i wizualizacji obszernych sieci danych,

- UCINET<sup>9</sup> – program dostępny w wersji trial do konstruowania i analizy społecznościowej sieci świata nauki; preferowane są dane macierzowe,

- XLSTAT<sup>10</sup> – komercyjne rozszerzenie programu MS Excel o funkcje analizy statystycznej i wizualizacji danych,

- Permap<sup>11</sup> – darmowy, nieduży program do mapowania metodą skalowania wielowymiarowego; szczególnie polecany w obszarach zastosowań psychometrii i socjometrii,

- Aplikacje do typowej analizy statystycznej dużych zbiorów danych: Statistica, Origin, Matlab, Mathematica.

Z powyższej listy wyróżnić należy ciągle rozwijany CiteSpace – kompletne narzędzie obsługujące wszystkie etapy procesu wizualizacji: od ekstrakcji danych do walidacji wynikowych konfiguracji graficznych.

## Przykład wizualizacji domeny naukowej

Spektrum mapowania nauk jest bardzo szerokie: nauki ścisłe, nauki medyczne, społeczne i humanistyczne. Te ostatnie stanowią prawdziwe wyzwanie dla ekspertów *KDViz*, bazy Web of Science bowiem indeksują nieznaczną część wszystkich światowych zasobów w zakresie nauk humanistycznych [18], również prace, gdzie się podejmuje mapowania nauk komputerowych i inżynierii

<sup>6</sup> <http://www8.umu.se/inforsk/Bibexcel/>.

<sup>7</sup> <http://cluster.cis.drexel.edu/~cchen/citespace/>.

<sup>8</sup> <http://vlado.fmf.uni-lj.si/pub/networks/pajek/>.

<sup>9</sup> <http://www.analytictech.com/ucinet6/ucinet.htm>.

<sup>10</sup> <http://www.xlstat.com/>.

<sup>11</sup> <http://www.ucl.ac.uk/~rbh8900/old%20web%20page.html>.

są bardzo nieliczne [17]. Nie jest tajemnicą, że w bazach ISI ciągle **przeważają** źródła amerykańskie, co nie odwzorowuje aktualnej struktury świata nauki, **gdzie** ostatnio znaczący jest wkład europejski, a szczególnie chiński. Skala materiału badanego, tj. liczba czasopism wpływa na rozpiętość dziedzinową mapowanej wiedzy. Na przykład, na podstawie 2 tys. czasopism zwizualizowano 32 dyscypliny i 140 specjalności naukowe [4], natomiast zobrazowanie całościowej nauki wymagało użycia 16 tys. publikacji i materiałów konferencyjnych, co dało ponad 7 mln publikacji za okres 2001-2005<sup>12</sup>.

Interdyscyplinarny charakter dziedziny Bibliotekoznawstwo i Informacja Naukowa (BIN) powoduje, iż jest ona najczęściej wybierana do wizualizacji. Badania struktury BIN przeprowadza się od lat, w podobny sposób używając źródeł; na podstawie analizy najważniejszych 12-tu czasopism anglojęzycznych z tego zakresu:

- Annual Review of Information Science and Technology,
- Electronic Library,
- Information Processing & Management,
- Information Technology and Libraries,
- Journal of Documentation,
- Journal of Information Science,
- Journal of the American Society for Information Science,
- Journal of the American Society for Information Science and Technology
  - JASIST,
- Library & Information Science Research,
- Library Resources & Technical Services,
- PROGRAM-Automated Library and Information Systems,
- Scientometrics.

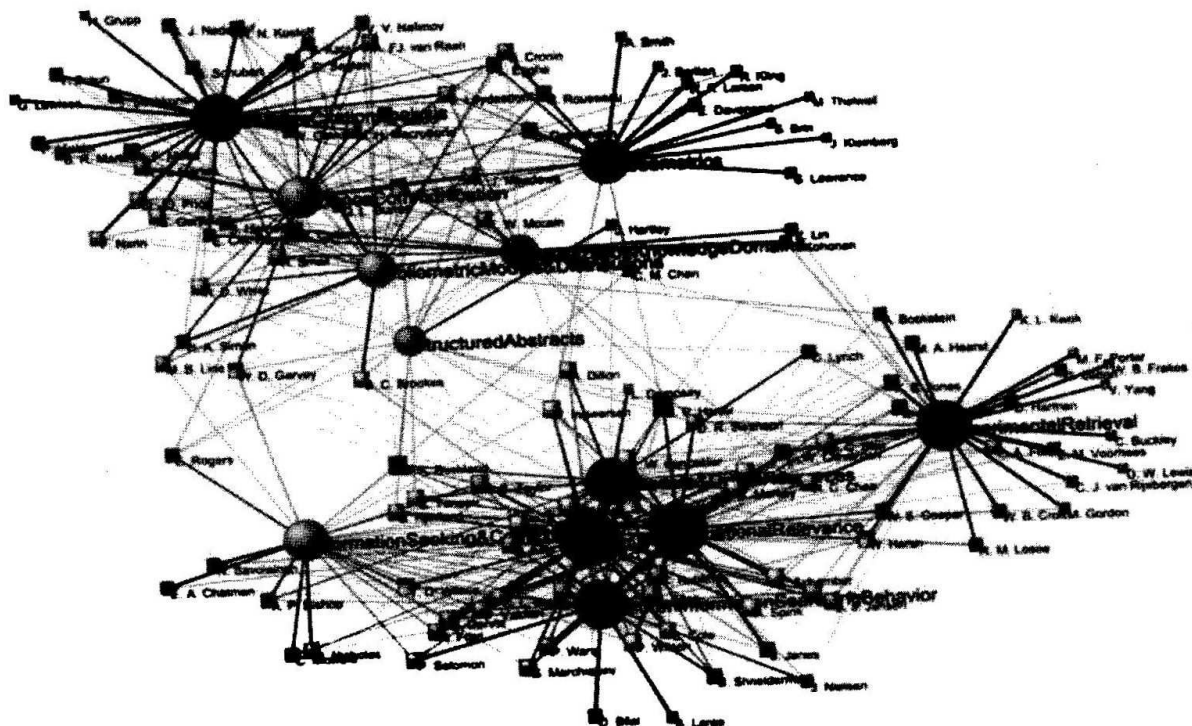
Spośród wielu wizualizacji BIN wybraliśmy przykład najbardziej czytelny i kompletny, a jednocześnie aktualny opisany w pracy [21]. Jest to wynik zmapowania literatury naukowej z lat 1996-2005 przy zastosowaniu analizy współcytowań autora (ACA), wzbogaconej o algorytmy rotacji ortogonalnej (rys. 2).

W wyniku analizy współcytowań zidentyfikowano główne specjalności badawcze. Największe na mapie węzły to:

- wyszukiwanie informacji,
- metadane i zasoby cyfrowe,
- Webometria,
- analiza cytowań (naukometria, bibliometria),
- badanie zachowań użytkownika,
- wizualizacja domen naukowych (ACA, analiza współcytowań),
- wyszukiwanie doświadczalne (algorytmy, modele, systemy, ewaluacja),
- komunikacja naukowa,
- użyteczność sieciowa, projektowanie interfejsów,
- abstrakty strukturalne (piśmiennictwo naukowe).

Autorzy podkreślają, iż w porównaniu z wynikami z końca lat 90., czyli sprzed ekspansji technologii sieciowych i używania Google'a („googlowania”),

<sup>12</sup> R. Klavans, K. Boyack: *Maps of Science: Forecasting Large Trends in Science. Places@Spaces: Mapping Science*. [online]. [dostęp: 19.05.2009]. Dostępny w World Wide Web: <[http://www.scimaps.org/dev/big\\_thumb.php?map\\_id=164](http://www.scimaps.org/dev/big_thumb.php?map_id=164)>.



Rys. 2. Wizualizacja dziedziny Bibliotekoznawstwa i Informatyki Naukowej metodą ACA na podstawie zmapowania artykułów z 12 czasopism za okres 1996-2005 [21]

powstał klaster obejmujący zagadnienia Webometrii oraz klaster odnoszący się do *KDViz*, a reprezentacje przestarzałych obszarów badań, takie jak OPAC i Online Retrieval uległy rozmyciu. Wyniki mapowania wskazują też na rosnące znaczenie studiów użyteczności sieciowej i zachowań użytkowników. Widać też nowy temat badawczy: analiza strategii wyszukiwawczej dla dzieci (*Children's information searching behaviour*). Na mapie dystrybucji ognisk tematów badawczych zauważalne jest również szerzące się kognitywne podejście do rozwiązywania problemów w naukach o informacji.

### Podsumowanie: mapy wiedzy a tradycyjne analizy bibliometryczne

Jako że metodologia wizualizacji informacji, zaadaptowana do nowej w ostatnim 10-leciu dyscypliny badawczej jaką jest mapowanie nauki/wiedzy, wyrosła z bibliometrii i naukometrii, nie dziwi fakt korzystania z cytowań jako z podstawowego materiału badawczego. Oceny bibliometryczne bazują na cytowalności, uznawanej za obiektywną miarę jakości źródeł piśmienniczych. Potencjał informacji zawartej w indeksach cytowań od dawna wykorzystuje się do badań naukowych, co spowodowało, iż analiza cytowań stała się samodzielnym działem bibliometrii [14].

Miary wpływu czasopism (np. Impact Factor – IF, index Hirscha, Page-Rank), czy autora podawane są w postaci liczbowej, analiza sprowadza się więc do porównania w tabelach liczb, które uważane są za parametry świadczące o wydajności naukowej danego czasopisma/autora. Uzyskane w ten sposób wskaźniki efektywności są podstawą do tworzenia rankingów i często



decydują o skali dotacji budżetowych, stąd kontrowersje w świecie nauki dotyczące powiązań komercyjności przedsięwzięć naukowych z nadinterpretacją wskaźnika IF [7]. Tabela bibliometryczna pozwala na wyodrębnienie grupy najbardziej lub najmniej wydajnych czasopism/autorów, jednak z tradycyjnej analizy bibliometrycznej nie da się wywnioskować o relacjach pomiędzy badanymi jednostkami i ich podobieństwem, np. tematycznym. Problem ten jest rozstrzygany na poziomie analizy wielowymiarowej, badań właściwości dużych zbiorów danych i ich dynamiki przy zastosowaniu technik datamining. W artykule [14] zademonstrowano za pomocą dendrogramów możliwości analizy skupień, jednej z metod bibliometrii strukturalnej. Na podstawie danych współcytowań pogrupowane zostały polskie czasopisma ekonomiczne, w grupy podobne pod względem opisywanych tematów.

Kolejny etap analiz wielowymiarowych – prezentacja wyników – daje nie tylko wgląd w związki, hierarchie i podobieństwo pomiędzy obiektami, lecz i obraz globalnej struktury danych, czasem też tendencje zmian, a co za tym idzie, możliwości prognozowania. Naukowcy wykorzystują materiał badawczy pochodzący nie tylko z indeksów cytowań, lecz coraz częściej z abstraktów i tekstu dokumentów. Eksploracja tekstu (*text mining*) w dużych zbiorach danych jest modnym przedmiotem badań w naukach komputerowych. Informatycy wypracowują metody, które mogłyby weryfikować tematyczną klasyfikację/kategoryzację badanych dokumentów. Niebagatelne znaczenie ma tu stały postęp w rozwoju technik prezentacji informacji. Cechą większości współczesnych aplikacji do wizualizacji, gdzie znajdują zastosowanie najnowsze odkrycia nauk kognitywnych, jest interakcja z użytkownikiem. To powoduje, iż analityk ma możliwość modernizacji konfiguracji danych wynikowych, posługując się empirycznie doбором parametrów wejściowych. Ważnym aspektem badawczym jest to, że mapy wizualizacyjne mogą służyć użytkownikowi jako interfejs systemu wyszukiwawczego [16].

Liczne prace w ostatnich latach w zakresie wizualizacji nauki/wiedzy udowadniają zasadność takich badań i przydatność wynikowych map nie tylko w naukoznawstwie i opracowaniach bibliometrycznych, lecz również w biologii, obrazowaniu medycznym i genetyce [3]. Warto tu wymienić możliwości analityczne, które niesie wizualizacja informacji i danych masowych. Mapowanie nauki umożliwia:

- określenie aktualnych trendów tematycznych i dominujących obszarów badawczych;
- poznanie bazy naukowej danej dziedziny oraz jej frontów badawczych, kierunki i zakresy integracji multidyscyplinarnych, identyfikację subdyscyplin;
- wykrycie społecznościowej (intelektualnej) struktury pola badań, włączającej zarówno aktywność autorów i wydawców, jak i osób zaangażowanych w tworzenie baz wiedzy; nakreślenie wzoru sieci współpracy pomiędzy naukowcami;
- analizę dorobku naukowego badaczy, instytucji naukowej, bazy naukowej danego kraju;
- wykrycie ewolucji za pomocą mapowania ciągłego (*longitudinal mapping*) [9] oraz prognozowanie przyszłych zachowań domeny naukowej poprzez wizualizację i modulowanie struktury;

– lepsze zrozumienie struktury, dynamiki dziedziny, wiedzy oraz kierunków rozprzestrzeniania się (dyfuzji) wiedzy. Naukowcy *Infoviz* są w stanie wykryć zmiany krytyczne w nauce, posługując się teorią paradygmatów [2]; może to wspomóc dokonanie odkryć naukowych;

– wyszukiwanie semantyczne informacji (za pomocą map semantycznych);  
– weryfikację i modernizację klasyfikacji przedmiotowej na podstawie semantycznej organizacji dokumentów na mapie wizualizacji.

Wizualizacja domen wiedzy jako odnowione i wzbogacone o techniki datamining podejście metodologii biblio- i naukometrycznych zawiera spory potencjał badawczy. Jak widać, wizualizacja zasobów naukowych decyduje o właściwym postrzeganiu i rozumieniu danych na wszystkich poziomach organizacji wiedzy. Twórcy *KDViz* podkreślają wieloperspektywiczność zastosowań metod wizualizacji: w wyszukiwaniu informacji, monitorowaniu trendów w nauce i technologiach, polityce finansowania ośrodków badań i konkretnych naukowców.

## Bibliografia

1. Bollen J. [et al.]: *Clickstream Data Yields High-Resolution Maps of Science*. *PLoS ONE*. 2009 vol. 4 no. 3. [online]. [dostęp: 31.08.2010]. Dostępny w World Wide Web: <<http://www.plosone.org/article/info:doi/10.1371/journal.pone.0004803>>.
2. Börner K., Chen Ch., Boyack K. W.: *Visualizing Knowledge Domains*. W: B. Cronin (red.): *Annual Review of Information Science & Technology*. 2003 vol. 37, pp. 179-255.
3. Boyack K. W. [et al.]: *Domain visualization using VxInsight for science and technology management*. „Journal of the American Society for Information Science and Technology” 2002, no. 53(9), pp. 764-774.
4. Boyack K. W. [et al.]: *Mapping the backbone of Science*. „Scientometrics” 2005 vol. 64, no. 3, pp. 351-374.
5. Chen Ch.: *Information Visualization. Beyond the Horizon*. 2nd ed. London, 2006.
6. Chen Ch. [et al.]: *The structure and dynamics of cocitation clusters: A multiple-perspective cocitation analysis*. „Journal of the American Society for Information Science and Technology” 2010 vol. 61(7), pp. 1386-1409.
7. *European Association of Science Editors statement on impact factors*. [online]. [dostęp: 31.08.2010]. Dostępny w World Wide Web: <[http://www.ease.org.uk/statements/EASE\\_statement\\_on\\_impact\\_factors.shtml](http://www.ease.org.uk/statements/EASE_statement_on_impact_factors.shtml). Retrieved, 2009-03-25>.
8. Garfield E.: *Scientography: Mapping the tracks of science*. „Current Contents: Social & Behavioural Sciences” 1994, no. 7(45), pp. 5-10.
9. Garfield E.: *Essays/Papers on Mapping the World of Science*. Eugene Garfield, Ph. D. Home Page. [online]. [dostęp: 31.08.2010]. Dostępny w World Wide Web: <<http://garfield.library.upenn.edu/mapping/mapping.html>>.
10. Hjørland B.: *Domain analysis in information science: eleven approaches – traditional as innovative*. „Journal of documentation” 2002, no. 58, pp. 422-462.
11. Holloway T. [et al.]: *Analyzing and Visualizing the Semantic Coverage of Wikipedia and Its Authors*. Wyd. specjalne: *Understanding Complex Systems. Complexity*. 2007, vol. 12, no. 3, pp. 30-40.
12. Moya-Anegón F. [et al.]: *A new technique for building maps of large scientific domains based on the cocitation of classes and categories*. „Scientometrics” 2004, vol. 61, no. 1, pp. 129-145.
13. Osiewalska A.: *Mierniki oceny czasopism i naukowców*. EBIB Biuletyn Elektroniczny Biuletyn Informacyjny Bibliotekarzy. 2008, nr 11(99). [online]. [dostęp: 31.08.2010]. Dostępny w World Wide Web: <<http://www.ebib.info/2008/99/a.php?osiewalska>>.
14. Osiewalska A.: *Bibliografie czasopism naukowych Biblioteki Głównej UEK jako źródło danych dla analiz bibliometrycznych*. W: *Bibliograficzne bazy danych: kierunki rozwoju*

- i możliwości współpracy*. Ogólnopolska konferencja naukowa z okazji 10-lecia bazy danych BazTech, Bydgoszcz 27-29 maja 2009 r. [online]. [dostęp: 31.08.2010]. Dostępny w World Wide Web: <<http://www.ebib.info/publikacje/matkonf/mat19/osiewalska.php>>
15. Osińska V.: *Przybliżenie semantyczne w wizualizacji informacji w Internecie i bibliotekach cyfrowych*. EBIB Elektroniczny Biuletyn Informacyjny Bibliotekarzy. 2006, nr 7(77). [online]. [dostęp: 31.08.2010]. Dostępny w World Wide Web: <<http://www.ebib.info/2006/77/osinska.php>>.
  16. Osińska V.: *Wizualizacja i wyszukiwanie dokumentów*. Warszawa 2010.
  17. Osińska V., Bala P.: *New Methods for Visualization and Improvement of Classification Schemes: The Case of Computer Science*. „Knowledge Organization” 2010, no. 37, pp. 157-172.
  18. Persson O., Danell R.: *How to use Bibexcel for various types of bibliometric analysis*. W: *Celebrating scholarly communication studies: A festschrift for Olle Persson*. Special volume of the e-zine of the ISSI. Pod red. F. Åström, R. Danell, B. Larsen and J. W. Schneider. 2009 vol. 05-S pp. 9-24. [online]. [dostęp: 12.11.2010]. Dostępny w World Wide Web: <<http://www8.umu.se/inforsk/Bibexcel/ollepersson60.pdf>>.
  19. Sivertsen G.: *Publication patterns in all fields 2009*. W: *Celebrating scholarly communication studies: A festschrift for Olle Persson*. Special volume of the e-zine of the ISSI. Pod red. F. Åström, R. Danell, B. Larsen and J. W. Schneider. 2009 vol. 05-S pp. 55-60. [online]. [dostęp: 12.11.2010]. Dostępny w World Wide Web: <<http://www8.umu.se/inforsk/Bibexcel/ollepersson60.pdf>>.
  20. Thelwall M.: *A Webometric Analysis of Olle Persson*. W: *Celebrating scholarly communication studies: A festschrift for Olle Persson*. Special volume of the e-zine of the ISSI. Pod red. F. Åström, R. Danell, B. Larsen and J. W. Schneider. 2009 vol. 05-S, pp. 61-71. [online]. [dostęp: 12.11.2010]. Dostępny w World Wide Web: <<http://www8.umu.se/inforsk/Bibexcel/ollepersson60.pdf>>.
  21. Zhao D., Strotmann A.: *Information Science during the first Decade of the Web: An enriched author cocitation analysis*. „Journal of the American Society for Information Science and Technology” 2008, vol. 59(6), pp. 916-937.

## Summary

The article describes a new research method, popular among bibliometric and scientometric specialists, i.e. science visualization. Genesis, technologies, and stages of the visualization process are presented, with an example of matrix map for library and information science for the years 1996-2005. Research potential of this method were discussed, in comparison with bibliometric methods. The article's aim is to encourage Polish scientists, interested in studies regarding science visualization, hardly popular in our country. The author describes mapping tools popular among visualization specialists, like Open Source, and their usability and technological features.