

# R o z d z i a ł 1

## WIZUALIZACJA INFORMACJI – OBSZAR BADAŃ INFORMACJI NAUKOWEJ

### 1.1. Geneza wizualizacji, pojęcie i historia

Wizualizacja jako swoista kompilacja *nauki i sztuki* od dawna jest obiektem zainteresowania naukowców, informatyków, grafików i wszystkich, którzy używają technologii komputerowych. Początkowo analizowana jako proces reprezentacji wyrazów i pojęć w formie wizualnej, rozszerzyła swój zakres zastosowań do narzędzi i metod interpretacji danych graficznych oraz generowania obrazów na podstawie zbiorów danych wielowymiarowych.

Na przestrzeni wieków, wizualne reprezentacje takie jak rysunki na starożytnych budowlach, mapy, reprezentacje geometrii euklidesowej a później diagramy statystyczne rozwijały się, aby wspomagać myślenie i wyobraźnię. Na tym podłożu wyrosły dyscypliny zajmujące się graficznym obrazowaniem: kartografia, kreślarstwo, projektowanie urządzeń technicznych za pomocą technik typu CAD<sup>1</sup>, grafika, drukarstwo, projektowanie etykiet czy wizualizacja danych statystycznych. Dzisiaj graficzna prezentacja danych jest już pełnoprawną częścią produkcji telewizyjnej i filmowej. Wystarczy spojrzeć na współczesne filmy, których nieodłącznym elementem są migające komunikaty, kolorowe okienka, wykresy, nie wspominając o efektach specjalnych. Nowoczesne środki grafiki komputerowej pozwalają na estetyczną, przystępną, przejrzystą, a często zaskakującą prezentację danych, informacji,

---

<sup>1</sup> CAD (*Computer Aided Design*) – projektowanie wspomagane komputerowo. Programy typu CAD służą do obliczeń inżynierskich, rysunków konstrukcyjnych, przedstawiania rysowanych elementów w perspektywie itp. Por. *Słownik pojęć komputerowych*. Pod red. V. Illingworth i J. Daintitha. Warszawa: Świat Książki 2004, s. 44.

czy wiedzy. Powstał nowy termin określający *najbardziej udany związek słowa z obrazem*<sup>2</sup> – **infoGRAFika** czyli grafika informacyjna.

Problem czy „wizualizacja jest bardziej grafiką czy nauką” ciągle powraca w momentach pojawienia się nowych trendów w metodach wizualizacji bądź zaawansowanego programu wizualizacyjnego. Z kombinacji tych dwóch podstawowych perspektyw wyłania się kilka podejść do wizualizacji. Można ją badać w ramach tradycji artystycznej projektowania graficznego. Z kolei w obszarze zastosowań nauk komputerowych może ona określać algorytmy wyświetlania danych. W konstruktywnym przybliżeniu systemów symbolicznych wizualizacja może być również przydatna jako część semiotyki. Nowoczesne podejście, bazując na teorii poznania wykorzystuje w regułach projektowania aktualną wiedzę o ludzkim systemie percepcji wzrokowej.

Wizualizację w praktycznych zastosowaniach naukowych i informatycznych nazwano **wizualizacją naukową** (ang. *scientific visualization*) lub **wizualizacją danych** (ang. *data visualization*). Uformowała się ona w odpowiedzi na zapotrzebowanie świata nauki na wydajną prezentację danych eksperymentalnych w formie graficznej i późniejszą sprawną ich analizę. Od kiedy naukowcy uznali wizualizację za dziedzinę wyłaniającą się bezpośrednio z nauk komputerowych powstało dużo prac metodologicznych definiujących tę dyscyplinę badawczą poprzez jej cele. Wizualizacja naukowa jest procesem reprezentacji danych wyjściowych w postaci obrazów, który ma pomóc zrozumieć znaczenie wyników eksperymentu. Podstawowym celem takiego procesu, co podkreśla się w wielu pracach<sup>3</sup> jest „pogłębianie wiedzy” (z angielskiego: *insight*; inne tłumaczenie to: „wnikliwe poznawanie”) poprzez mapowanie danych do podstawowych figur graficznych: prostokątów, kwadratów, okręgów itp. Ma on również polepszać możliwości poznawcze człowieka w trakcie badania, podejmowania decyzji i eksploracji. Wizualizacja wspiera naukowców w udowadnianiu i obalaniu hipotez naukowych, w odkrywaniu nowych zjawisk fizycznych i symulacji pomiarów eksperymentalnych.

Z najnowszych definicji, które można znaleźć w literaturze naukowej, warto przytoczyć następującą, ze względu na jej aktualność, przejrzystość i ścisłość. *Wizualizacja jest to więcej niż metoda komputerowa. Wizualizacja jest to proces, w którym dane, informacja i wiedza przekazywane są formie wizualnej. Zaangażowane są w tym momencie następujące elementy: komputer do przetwarzania informacji, ekran – do prezentacji oraz ludzki mózg – do jej percepcji i analizy.*<sup>4</sup> Jest to proces pozwalający odbiorcy na obserwację, przeglądanie oraz zrozumienie informacji. Wizualizacja zatem to zarówno sam akt tworzenia graficznej reprezentacji danych, jak i proces logicznej analizy jej treści<sup>5</sup>.

<sup>2</sup> M. Burns, T. Bitner: *Sztuka informowania*. Digit online [on-line] 2003, nr 6 [dostęp 19 maja 2009]. Dostępny w World Wide Web: [http://www.digit.pl/artykuly/34291\\_6/sztuka.informowania.html](http://www.digit.pl/artykuly/34291_6/sztuka.informowania.html).

<sup>3</sup> I. Niskanen. *An interactive ontology visualization approach for the domain of networked home environments* [on-line]. Oulu: Julkaisija-Utgivare, 2007 [dostęp 19 maja 2009]. Dostępny w World Wide Web: <http://www.vtt.fi/inf/pdf/publications/2007/P649.pdf>; E. R. Tufte. *Envisioning Information*. Connecticut, USA: Graphics Press 1990, s. 12-51; W. J. Yurcik. *Scientific Visualization* [on-line]. BookRags [dostęp 19 maja 2009]. Dostępny w World Wide Web: <http://www.bookrags.com/research/scientific-visualization-csci-03/scientific-visualization-csci-03.html>.

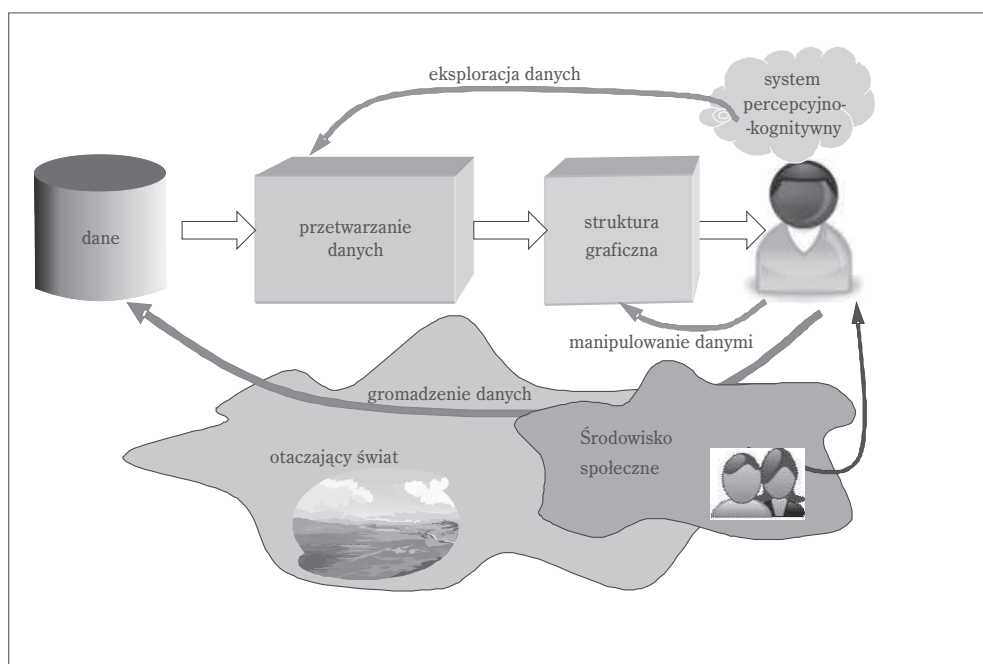
<sup>4</sup> *What is Visualization?* [on-line]. Infovis. Information Visualization Resources [dostęp 19 maja 2009]. Dostępny w World Wide Web: <http://infovis.org/>.

<sup>5</sup> I. Niskanen, dz.cyt.

Jak zatem wygląda proces wizualizacji? Według C. Ware, dyrektora *Data Visualization Research Lab* i autora monografii dotyczącej wizualizacji informacji składa się on z czterech podstawowych etapów z trzema pętlami oddziaływania zwrotnego<sup>6</sup>:

- skompletowanie zbioru danych;
- przetwarzanie danych, które prowadzi do ich konwersji w formę zrozumiałą dla człowieka;
- użycie algorytmów graficznych do wyświetlenia obrazu wizualizacji na ekranie;
- włączenie ludzkiego systemu percepcyjno-kognitywnego.

Analitik postępuje w następującej kolejności: kolekcjonuje lub mierzy serie danych i wprowadza je w stadium obróbki, która włącza konwersję formatów i filtrowanie, po czym następuje mapowanie danych zgodnie z koncepcją wizualizacji oraz renderowanie<sup>7</sup> w przypadku przestrzeni trzymiarowych (Rysunek 1). Końcowy obraz ma zaktywować ludzkie systemy wizualny i kognitywny – czyli ostatni etap procesu wizualizacji informacji. Trzy pętle zwrotne to: zbieranie, manipulowanie i eksploracja danych. Najdłuższą pętlą jest kolekcjonowanie danych, kiedy naukowiec lub analitik gromadzi interesujące go informacje ze środowiska. To ostatnie jest źródłem danych, *gdy środowisko społeczne kompleksowo determinuje obiekt kolekcjonowania* – zaznacza C. Ware<sup>8</sup>. Manipulowanie danymi oznacza zakres kontroli prezentowania wyników odbiorcy, np. wybór kolorystyki, rozdzielczości obrazu lub transformacji geometryczno-przestrzenne obiektów badanych. Dogłębne zbadanie danych można przeprowadzić, jeśli cofniemy się do etapu przetwarzania, aby wybrać stosowną metodę transformacji – pętla eksploracji danych.



**Rysunek 1.** Diagram procesu wizualizacji

Źródło: Na podst. C. Ware. *Information Visualization: Perception for Design*. Wyd. 2. San Francisco: Morgan Kaufmann, 2004, s. 4.

<sup>6</sup> C. Ware. *Information Visualization: Perception for Design*. Wyd. 2. San Francisco: Morgan Kaufmann, 2004, s. 4-5.

<sup>7</sup> *Renderowanie* (ang. *rendering*) lub obrazowanie w grafice 3D: komputerowa analiza modelu danej sceny i utworzenie na jej podstawie dwuwymiarowego obrazu wyjściowego w formie statycznej lub w formie animacji. Por. *Słownik pojęć komputerowych...*, s. 271.

<sup>8</sup> C. Ware, dz. cyt., s. 5.

Inni badacze tej problematyki także podejmowali się wykonania schematyzacji procesu wizualizacji. E. Tufte w swym dziele *Readings in Information Visualization. Using vision to think* (1992) z perspektywy badań neuropsychologicznych i kognitywistycznych prowadzonych w ciągu ostatnich 15 lat stworzył diagram, który ilustruje etapy przekształcenia wierszy danych (ang. *raw data*) do postaci graficznej. Pierwszy etap włącza konwersję danych do tabel lub macierzy, drugi – reprezentację graficzną, w trzecim dokonuje się transformacji obrazu końcowego.

Rozważmy zatem, jak funkcjonowała ta dyscyplina badawcza w okresach przedkomputerowym i ogólnej dostępności komputerów. Badacze już od XII wieku potrafili wypracować metody graficznego odwzorowywania elementów rzeczywistości: system układu współrzędnych, wszelkiego rodzaju wykresy (liniowy, strumieniowy, rozrzutu), mapy konturowe (1594 r.), kolorowe (1741 r.) warstwowe i stereogramy (1896 r.). Mimo że mapy wykonywano odręcznie i powstanie kartografii datuje się na rok 1664, to komputery zrewolucjonizowały tę dziedzinę. Dzisiejsze mapy wysokiej jakości o zastosowaniu komercyjnym wykonuje się przy pomocy programów CAD, GIS (*Geographic Information System*) lub w innych systemach specjalistycznych.

Wizualizacja naukowa wykonywana za pomocą komputera rozwijała się równolegle z grafiką komputerową. W połowie lat osiemdziesiątych XX w. zaawansowane procesy informatyczne, takie jak symulacje na superkomputerach, dostarczały tak dużych ilości danych, iż wzmogło to konieczność poszukiwania nowych narzędzi i algorytmów do kompleksowej wizualizacji. Rok 1986 rozpoczął etap integracji grafiki komputerowej i nauk komputerowych. Przyczyniła się do tego Narodowa Fundacja Nauki w USA (*National Science Foundation*), która poleciła instytucjom naukowym nabywać oprogramowanie graficzne oraz odpowiedni sprzęt komputerowy. Raport z pierwszych warsztatów wizualizacji naukowej rok później głosił:

*Naukowcy potrzebują alternatywy dla liczb. Wykorzystanie obrazu zamiast liczb jest technicznie realne dzisiaj i jutro stanie się niezbędnym rekwizytem wiedzy. Celem obliczeń naukowych jest poszukiwanie, nie wyliczenie. Dowiedziono, że 50% neuronów w mózgu odpowiada za widzenie. Wizualizacja w naukowych obliczeniach pomaga całą neuronową maszynę puścić w ruch<sup>9</sup>.*

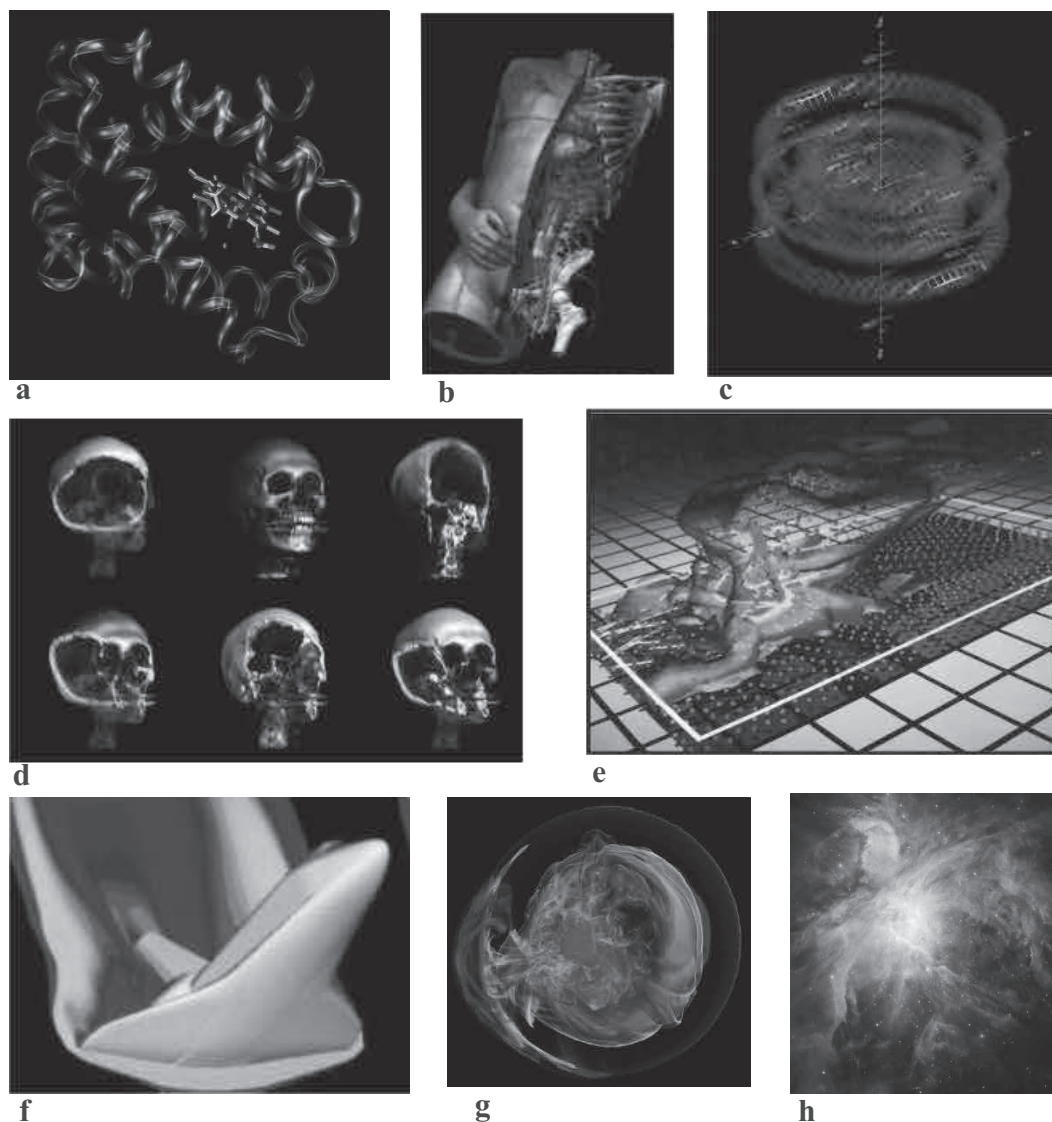
Dzisiaj już doskonale wiemy, że dwie półkule mózgu funkcjonują w różny sposób. Lewa pomaga w obliczeniach analitycznych, komunikacji werbalnej i operowaniu symbolami abstrakcyjnymi. Prawa odpowiada za przestrzenne, intuicyjne lub holistyczne myślenie. Aktywuje się w trakcie oceny kompleksowości sytuacji. Graficzne reprezentacje stymulują właśnie tę część mózgu. Używając takiego przybliżenia naukowcy otrzymują całościowy obraz danych. W późniejszych etapach do wykrywania pewnych anomalii we wzorach danych stosuje się metody bardziej analityczne.

Do wzmożonego zainteresowania wizualizacją już od końca lat dziewięćdziesiątych ubiegłego wieku przyczyniło się kilka czynników. Dostępność cenowa komputerów, kolorowych monitorów i wydajnych kart graficznych dało początek rozpowszechnieniu grafiki prezentacyjnej i oprogramowaniu do tego służącemu. Popularyzacja Linuxa oraz programów Open Source pozwoliła na rozwój technik

<sup>9</sup> X. Berenguer: *The synthetic image as language*. Temes de Disseny [on-line] 1991, nr 5 [dostęp 19 maja 2009]. Dostępny w World Wide Web: <http://tdd.elisava.net/coleccion/5/berenguer-en>.



wizualizacyjnych i programów graficznych w środowisku naukowym, gdzie wzrosło też zapotrzebowanie na zaawansowaną wizualną analizę obszernych kolekcji danych. I wreszcie, błyskawicznie taniejące monitory LCD i karty graficzne z akceleratorem 3D wspomogły dalszy rozwój metod wizualizacji przestrzennej.



**Rysunek 2.** Przykłady nowoczesnej wizualizacji naukowej: **a)** dynamiki molekuł; **b)** wnętrza ludzkiego ciała; **c)** chmury elektronowej; **d)** tomografia komputerowa czaszki; **e)** klimatycznej; **f)** dynamiki cieczy; **g)** symulacja wybuchu gwiazdy supernowej; **h)** międzygwiazdowego pyłu

**Źródła:** a, b, d, e, g) Kwan-Liu Ma. *Introduction to Visualization* [on-line]. DOE Office of Science Homepage [dostęp 19 maja 2009]. Dostępny w World Wide Web: [http://www.er.doe.gov/ascr/Research/scidac/intro\\_datavis.pdf](http://www.er.doe.gov/ascr/Research/scidac/intro_datavis.pdf); c) Help programu *MATLAB*; f) *Orbiter model* [on-line]. NASA [dostęp 19 maja 2009]. Dostępny w World Wide Web: [http://www.nasa.gov/multimedia/imagegallery/image\\_feature\\_431.html](http://www.nasa.gov/multimedia/imagegallery/image_feature_431.html); h) *Wnętrze Wielkiej Mgławicy w Orionie (M42)* [on-line]. *teleskopy.pl* [dostęp 19 maja 2009]. Dostępny w World Wide Web: <http://www.teleskopy.pl/obiektymglawicowe.html>.

Nowoczesne technologie wizualizacyjne interesowały nie tylko naukowców, lecz także środowiska komercyjne. Informacja biznesowa potrzebowała rzetelnego przedstawienia danych, np. w organizowaniu, identyfikacji i komunikacji trendów

rynku i danych konsumenta. Wizualizacja naukowa wykorzystywana jest w astronomii do wizualizacji obiektów i zjawisk z kosmosu, których nie sposób obserwować bezpośrednio: czarnych dziur, fal grawitacyjnych i kolizji gwiazd neutronowych (Rysunek 2g, h). W fizyce technologie wizualizacyjne są obecne, kiedy bada się dynamikę płynów (*Computational Fluid Dynamics – CFD*) (Rysunek 2f) lub konstruuje się mapy rozkładu gęstości elektronów (Rysunek 2c). Fascynujące są ilustracje przedstawiające naukowe procesy w sposób zrozumiały dla zwykłego czytelnika które możemy obejrzeć w czasopismach naukowo-popularnych np. „Wiedza i Życie”, „Świat Nauki”. W medycynie i biologii technik wizualizacji używa się, najczęściej aby zobrazować procesy biologiczne np. rozwój komórki, dynamikę molekuł, funkcjonowanie ludzkich narządów (Rysunek 2a, b), wspomagając diagnostykę medyczną lub w przygotowaniu operacji chirurgicznych. Godną uwagi jest tomografia komputerowa, metoda diagnostyczna obrazowania ludzkiego ciała, wykorzystująca osiągnięcia wizualizacji naukowej (Rysunek 2d). Ogólnie znanym przykładem jest wizualizacja danych meteorologicznych, którą możemy obserwować w prognozach pogody – Rysunek 2e.

Zaadoptowanie metod wizualizacyjnych w różnych sferach nauki i aktywności człowieka spowodowało, że oprócz pierwotnego przeznaczenia do analizy zbiorów danych naukowych, wykształciły się inne ścisłe kierunki zastosowań wizualizacji: edukacja, komunikacja, informacja naukowa i organizacja wiedzy. **Wizualizacja edukacyjna** (ang. *Educational Visualization*) używa przede wszystkim symulacji komputerowej w celu zaprezentowania zjawisk lub tematów, których nie jest prosto zaobserwować, np. strukturę atomu, wnętrze ludzkiego ciała albo życie prehistoryczne dinozaurów<sup>10</sup>. W nienaukowym odniesieniu wizualizacja jest jednym ze sposobów komunikacji międzyludzkiej. Od sztuki jaskiniowej, znaków, ideogramów, alfabetu, książki, obrazów, fotografii aż po strony WWW – taką ścieżką podążała historia **komunikacji wizualnej**<sup>11</sup>, którą w Wikipedii określa się jako *komunikację idei poprzez wizualne wyświetlanie informacji*<sup>12</sup>.

Jeśli wizualizacja danych jest skoncentrowana wokół danych mierzalnych, takich jak wyniki medycznych badań ludzkiego ciała lub dane geograficznych systemów informacyjnych, to **wizualizacja informacji** zajmuje się danymi nierzeczywistymi czyli np. tekstem lub strukturami hierarchicznymi. W myśl ogólnej definicji, wizualizacja informacji jest wizualną prezentacją przestrzeni informacyjnych i struktur w celu ułatwienia ich szybkiego przyswojenia i zrozumienia<sup>13</sup>. W rzeczywistej (nie abstrakcyjnej) reprezentacji informacji wykorzystywana jest wiedza o naturalnej zdolności człowieka do szybkiego rozpoznawania obrazów. Jednak nie każdą informację da się sprowadzić do jej bezpośredniej interpretacji w świecie fizycznym.

**Infovis** (popularnie używany skrót w literaturze naukowej i biznesowej od ang. *Information Visualization*; tak również nazywają się corocznie odbywające się kon-

<sup>10</sup> I. Niskanen, dz. cyt., s. 40-45.

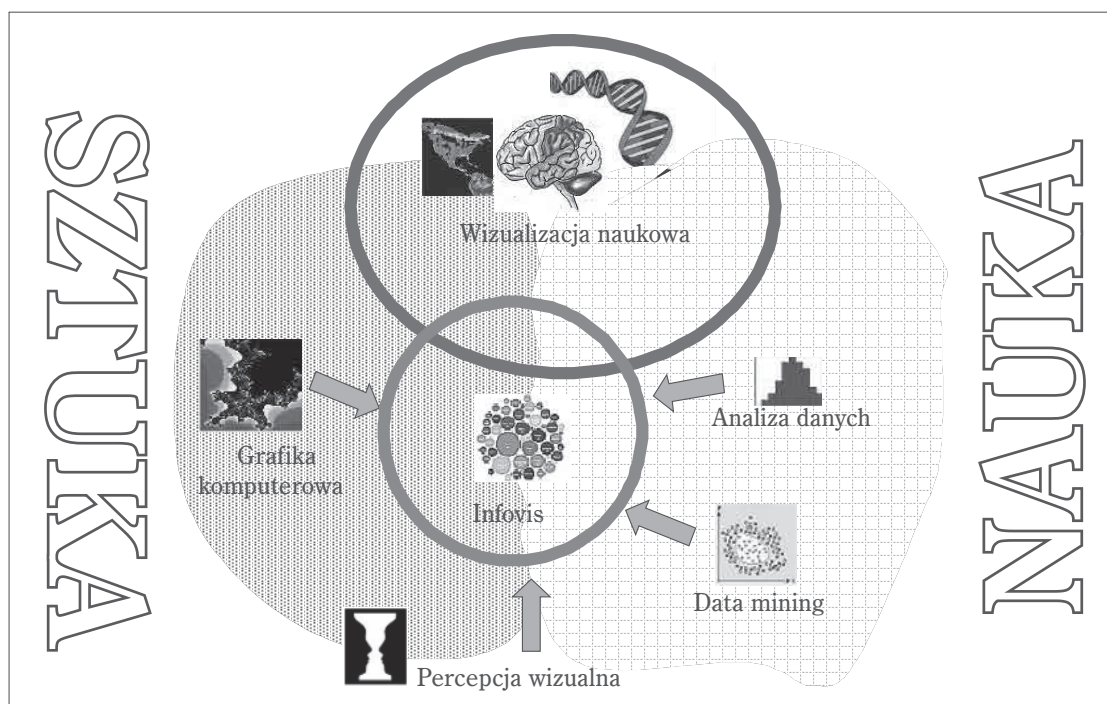
<sup>11</sup> *The History of Visual Communication* [on-line]. Sabanci University, Istanbul [dostęp 19 maja 2009]. Dostępny w World Wide Web: [http://www.citrinitas.com/history\\_of\\_viscom/](http://www.citrinitas.com/history_of_viscom/).

<sup>12</sup> *Visual Communication*. W: *Wikipedia. The Free Encyclopedia* [on-line] [dostęp 19 maja 2009]. Dostępny w World Wide Web: [http://en.wikipedia.org/wiki/Visual\\_communication](http://en.wikipedia.org/wiki/Visual_communication).

<sup>13</sup> N. D. Gershon, S. G. Eick. *Guest Editors' Introduction: Information Visualization. The Next Frontier*. Journal of Intelligent Information Systems. 1998, Vol. 11 (3), s. 199-201.

ferencje pod patronatem *IEEE*<sup>14</sup>) jest to dyscyplina, która poszukuje nowych metafor graficznych w celu przedstawienia informacji nie mającej naturalnej i oczywistej reprezentacji. Należy nadmienić, że wizualizacja informacji jest stosunkowo młodą dyscypliną badawczą o zaledwie 10-letniej historii, intensywnie się rozwija i jest przedmiotem rozważania w niniejszej pracy. Infovis wykorzystuje osiągnięcia takich pokrewnych mu dziedzin jak: wizualizacja naukowa czy eksploracja danych, interakcja człowiek-komputer (*Human Computer Interaction*<sup>15</sup>), percepcja wizualna (*Vision, Visual Perception*) i grafika komputerowa. Wzajemne relacje pomiędzy wymienionymi dyscyplinami przedstawiono na Rysunku 3.

Znamiennym cytatem o Infovis jest: *Oko wypatruje podobne obiekty, aby je porównać, dokonuje ich analizy pod różnym kątem z różnej perspektywy, aby dopasować ich elementy składowe* (S. Feldman)<sup>16</sup>. Wizualizacja może być jednym z etapów procesu analitycznego, jeśli pozwala na szybkie wykrycie związków pomiędzy poszczególnymi cechami lub nieprawidłowych wartości tych cech. Taka analiza wizualna koncentruje się na procesach rozumowania i odkrywania sensu danych. Techniki wizualizacji są stosowane jako jedna ze skuteczniejszych form eksploracji danych (ang. *data mining*); mogą one w niektórych przypadkach wykryć więcej korelacji niż klasyczne metody statystyczne.



**Rysunek 3.** Infovis i dyscypliny pokrewne

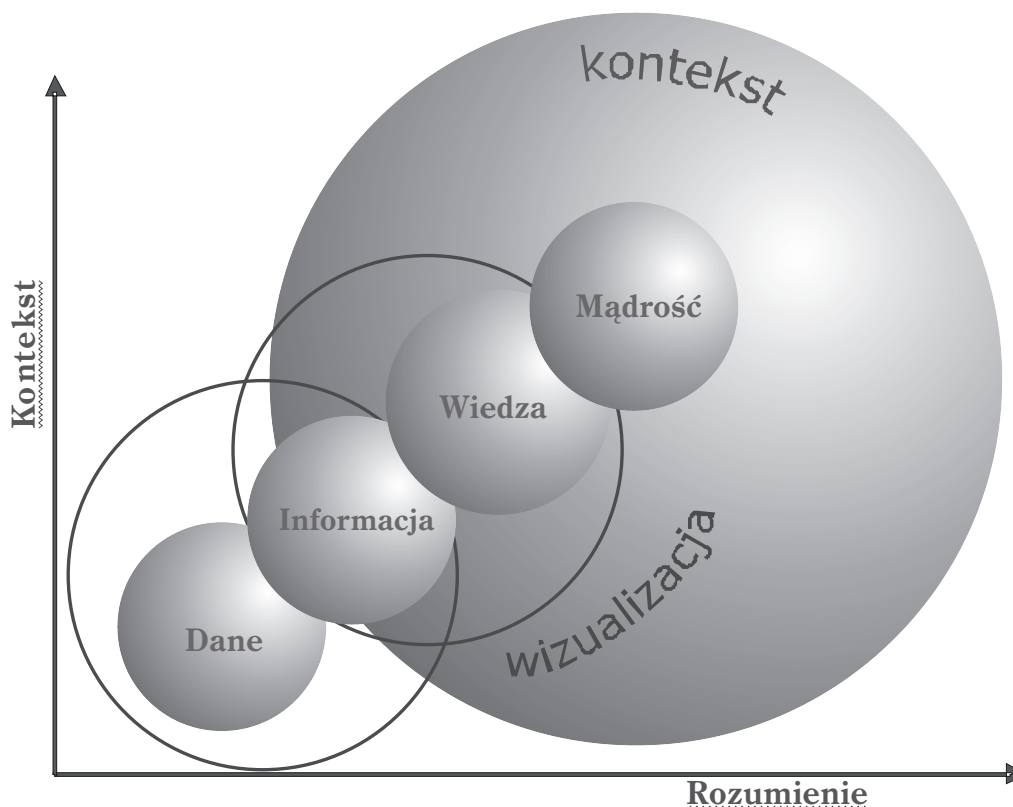
Źródło: Opracowanie własne.

<sup>14</sup> IEEE (Institute of Electrical and Electronics Engineers) – Instytut Inżynierów Elektryków i Elektroników – jedna z głównych organizacji skupiająca informatyków – praktyków. Ustala standardy konstrukcji urządzeń elektronicznych. Por. Concise Encyclopedia of Computer Science. Ed. by E. D. Reilly. Chichester, UK: Wiley, 2004, s. 395.

<sup>15</sup> Interdyscyplinarną naukę zajmującą się projektowaniem interfejsów użytkownika oraz badaniem i opisywaniem zjawisk związanych z używaniem systemów komputerowych przez ludzi.

<sup>16</sup> J. Luther, M. Kelly, D. Beagle: *Visualize This*. Library Journal [on-line] 2005, nr 3 (1) [dostęp 19.05.2009]. Dostępny w World Wide Web: <http://www.libraryjournal.com/article/CA504640.html>.

Niektórzy badacze rozpatrują wizualizację informacji w kontekście zarządzania wiedzą jako stymulator jej zrozumienia. Wówczas definicja w nawiązaniu do tej koncepcji brzmi<sup>17</sup>: Infovis - jest to *proces uwewnętrznienia* (ang. *internalization*) *wiedzy poprzez percepcję informacji*. W efekcie rozumowanie można zinterpretować jako continuum, które rozciąga się od danych pierwotnych do mądrości poprzez informację i wiedzę, „zespolone” w procesie wizualizacji – por. Rysunek 4. Diagram ten zawiera cztery koncepcyjne koła: dane, informację, wiedzę i wreszcie mądrość jako umiejętność praktycznego wykorzystywania posiadanej wiedzy. Dane figurują jako czyste fakty, w oderwaniu od kontekstu.



**Rysunek 4.** Koncept zamiany danych wejściowych w wiedzę i mądrość

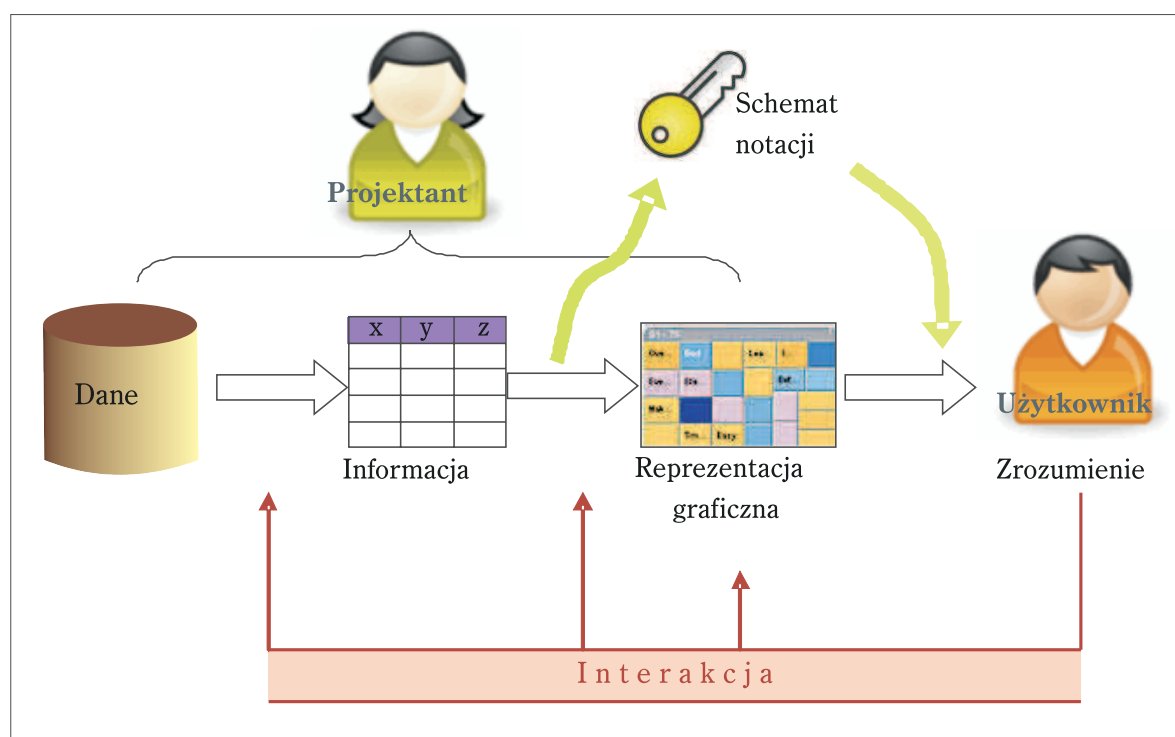
Źródło: Na podst. J. C. Dürsteler. *Diagrams for Visualisation*. The digital magazine of InfoVis.net [on-line] 2007, nr 186 [dostęp 19 maja 2009]. Dostępny w World Wide Web: <http://www.infovis.net/printMag.php?num=186&lang=2>.

Dane przekształcają się w informację pod warunkiem że przenoszą informację, którą zdołamy zrozumieć i która jest dla nas wartościowa. Informacja – to są dane wzbogacone o kontekst; zamienia się ona w kolejny element conceptualny – wiedzę. W słowniku termin „wiedza” jest określony jako „zasób zorganizowanych informacji zaopatrzone w szeroki kontekst” oraz „wiedzę nabywa się poprzez badanie lub doświadczenie”. Wizualizacja jako konstrukcja w umyśle rozpościera się poza percepcję sensoryczną, przez co zbliża się do wiedzy. Staje się intelektualnym ujęciem obiektów. Zrozumieć oznacza otoczyć, zawierać, uwewnętrznzić.

<sup>17</sup> J. C. Dürsteler. *Diagrams for Visualisation*. The digital magazine of InfoVis.net [on-line] 2007, nr 186 [dostęp 19.05.2009]. Dostępny w World Wide Web: <http://www.infovis.net/printMag.php?num=186&lang=2>.



W oparciu o takie abstrakcyjne budulce jak fenomen percepcji i wiedzę postaramy się odtworzyć pełniejszy diagram procesu wizualizacji informacji. Propozycja autora pracy (Dürsteler 2008) włącza nowe aspekty. Po pierwsze, wprowadzono wizualny język kodowania – schemat notacji, aby ułatwić użytkownikowi wnioskowanie. Po drugie, wizualizację potraktowano dwójako: zarówno jako proces jej tworzenia, jak i analizy prezentowanych treści graficznych. Na Rysunku 5 widzimy trzy fazy przejściowe: dane → informacja, informacja → reprezentacja wizualna, reprezentacja wizualna → zrozumienie. Na pierwszym etapie – konwersji danych w informację, wykonywane są trzy czynności: gromadzenie i przechowywanie danych, obróbka danych oraz organizacja danych stosownie do ich znaczenia, np. konstruowanie tabel lub macierzy. W fazie transformacji informacji do jej graficznej reprezentacji ważna jest możliwość późniejszej identyfikacji struktury graficznej i percepcyjnej od strony użytkownika. Pomocnym elementem może tu się okazać schemat notacji – taki język wizualny, przedstawiający strumienie informacji za pomocą symboli graficznych.



**Rysunek 5.** Diagram procesu Infovis

Źródło: Na podst. J. C. Dürsteler. *Diagrams for Visualisation*. The digital magazine of InfoVis.net [on-line] 2007, nr 186 [dostęp 19 maja 2009]. Dostępny w World Wide Web: <http://www.infovis.net/printMag.php?num=186&lang=2>.

W końcowej fazie wizualizacja ma zmusić użytkownika do percepcji, a następnie do wnioskowania. A zatem projektant interfejsu musi wziąć pod uwagę wpływ takich czynników jak: percepcja wizualna, psychologia kognitywna, lingwistyka i inne. Odbiorca powinien mieć możliwość interakcji z programem, np. modyfikować formę organizacji danych, generować nowe dane lub struktury i manipulować parametrami reprezentacji graficznej.

Infovis w swoim założeniu ma służyć przede wszystkim użytkownikowi, zwiększać jego zdolności percepcyjne – to jest główna różnica w porównaniu z zadaniami wizualizacji naukowej. Właśnie teraz, kiedy dociera do nas nadmiar informacji, w różnym stopniu przydatnej oraz czytelnej dla zainteresowanego odbiorcy, priorytetowe znaczenie ma skuteczna metoda wizualizacji. Ma ona zapewnić użytkownikowi pełną kontrolę nad procesem wyszukiwania informacji (ang. *Information Retrieval* – *IR*). W obliczu problemów przed jakimi staje społeczeństwo informacyjne, aplikacje Infovis mają być projektowane z uwzględnieniem trzech zasad<sup>18</sup>. Po pierwsze efektywny interfejs programu z zaimplementowanymi algorytmami wizualizacyjnymi ma wykonywać realizację takich zadań jak: obserwacja, wyszukiwanie, nawigacja, rozpoznanie, filtrowanie danych, oraz rozumienie i odkrywanie ukrytych wzorów korelacji. Projektanci Infovis za cel stawiają ergonomię systemów informacyjnych. Wiedza o działaniu ludzkiej percepcji i tego w jaki sposób zachodzi przetwarzanie obrazu pomaga w formułowaniu reguł wyświetlania informacji, które są niezbędne w budowaniu interfejsów graficznych. Kolejny podrozdział przybliży problematykę badań neuropsychologicznych i kognitywistycznych.

Drugą zasadą jest trzymanie się walorów estetycznych w prezentacji informacji. Wizualizacja wykształciła się po części z grafiki komputerowej, w związku z czym potrafiła zaadoptować nowoczesne środki technologii graficznych. Ładna i przejrzysta wizualizacja skutecznie przykuwa uwagę odbiorcy.

I wreszcie ważną rolę w programie odgrywa komunikacja z użytkownikiem. Funkcja interaktywna, a o niej tu mowa, stwarza możliwość manipulowania danymi na bieżąco. A parametry, na które użytkownik może wpłynąć, to między innymi: zakres danych, rozdzielczość, kąt widzenia i perspektywa (3D), sposób sortowania, filtrowania, kolorystyka obrazu itp.

Projektowanie graficzne ewoluowało skokowo, równolegle z rozwojem technologii komputerowych<sup>19</sup>. Ta dynamika znajduje odwzorowanie w przebiegu zmian w stosowanych metodach wizualizacji. Dzisiaj ogólną tendencją jest projektowanie interfejsów 3D. Można spróbować wymienić przyczyny powodujące, że wizualizacja (włącznie z Infovis) w XXI wieku przybiera nowe kompleksowe oblicze:

- zaawansowany sprzęt komputerowy i rozwój technologii informatycznych;
- wzrastające rozdzielczość i złożoność struktur danych;
- rozwój grafiki 3D;
- rozwój gier komputerowych i technologii rzeczywistości wirtualnej;
- postępy w robotyce i systemach sztucznej inteligencji;
- zaawansowany stan badań nad ludzką percepcją;
- społeczny Web 2.0;
- używanie semantyki i zapowiedź semantycznego WWW.

Współczesne aplikacje do wizualizacji można podzielić na trzy grupy według przeznaczenia i typów danych: matematyczne dla danych numerycznych (*Matlab*, *Mathematica*), naukowe dla danych przestrzennych (*IDL*, *Insight*, *Arcgis*) i informacji naukowej dla danych symbolicznych, np. ze stron internetowych. W sieci moż-

<sup>18</sup> J. C. Dürsteler. *InfoVis Diagram...*

<sup>19</sup> W. Osińska: *Dynamika historycznego rozwoju stron WWW*. Biuletyn EBIB [on-line] 2007, nr 7 (88) [dostęp 19 maja 2009]. Dostępny w World Wide Web: <http://www.ebib.info/2007/88/>.

na bez większego wysiłku znaleźć imponującą ilość darmowych aplikacji Infovis. Do ich historycznego przeglądu przeznaczony został podrozdział 1.2. Wizualizacja informacji cieszy się stałym zainteresowaniem nie tylko informatyków, grafików, specjalistów informacji naukowej, lecz także użytkowników Internetu. Świadczą o tym liczne czasopisma elektroniczne (np. *Infovis*, *Information Visualization*, *International Journal of Human-Computer Studies*), a także portale sieciowe z zasobami oferujące źródła programów, moduły, dyskusje i programy on-line do testowania sposobów wizualizacji. Dzisiejszym trendem w Infovis jest skierowanie na masowego odbiorcę. Pod hasłami „infografika dla mass” i „wiele oczu” (*Infovis for Masses*<sup>20</sup>, *Many Eyes*<sup>21</sup>) odbyła się konferencja Infovis w roku 2007 w Sacramento, USA.

Eksperci Infovis, np. Ch. Chen<sup>22</sup> – główny redaktor czasopisma *Information Visualization* i wiodący specjalista w wizualizacji zasobów bibliotek cyfrowych, sygnalizują konieczność powstania rozwiniętej teorii wizualizacji. Infovis jest młodą subdyscypliną, która rozwija się niejednorodnie. Literatura bibliograficzna przedstawia szerokie spektrum zaimplementowanych metod lub technik, gdzie trudno o solidną systematyzację. Brakuje jednolitej teorii, aby móc dokonać ewaluacji danego schematu reprezentacji. Wobec tego na razie podstawowym wymogiem jest zapewnienie aplikacjom, opisanych powyżej funkcji w interakcji z użytkownikiem, do którego należy ocena końcowa.

## Kolor i kształt w projektowaniu informacji

Kolor jest niezbędny w wizualizacji informacji, ponieważ monochromatyczne kodowanie danych jest nieefektywnym wykorzystaniem zasobów percepcyjnych człowieka. Kolorowa wizja udowodniła swoją przydatność w procesach ewolucji. Na przykład pomagała ujawniać kamuflaż, wynajdywać użyteczne obiekty, odczytywać świeżość pożywienia itp. Taka naturalna rola koloru sugeruje, że poprawniej byłoby traktować go jako atrybut obiektu niż jego pierwotną charakterystykę, przez co idealnie się nadaje zadań kategoryzacji i etykietowania. U podstaw percepcji kolorów leży trójkromatyczna teoria, według której informacja o kolorze przenoszona jest w trzech kanałach. Przyczyna tkwi w budowie ludzkiego oka: otóż światłoczułe receptory siatkówki, nazywane czopkami są trzech rodzajów. Każdy ma inną charakterystykę widmową, czyli reaguje na światło z innego zakresu barw. Pierwszy rodzaj reaguje głównie na światło czerwone (ok. 700 nm), drugi na światło zielone (ok. 530 nm) i ostatni na światło niebieskie (ok. 420 nm). Wyróżnienie tych trzech rodzajów czopków wpłynęło na opracowanie modelu kolorów RGB. Oprócz ludzi takim trójkromatycznym widzeniem charakteryzują się niektóre naczelne. Badania neuropsychologiczne postrzegania barw przez ludzi i małp wykazały ciekawe właściwości i zależności. Do zapamiętania łatwiejsze są kolory zbliżone do „idealnych”,

<sup>20</sup> Strona konferencji InfoVis 2007: *InfoVis 2007 Welcome* [on-line] [dostęp 19 maja 2009]. Dostępny w World Wide Web: <http://conferences.computer.org/infovis/infovis2007/>.

<sup>21</sup> Portal „publicznej wizualizacji”: *Many Eyes* [on-line] [dostęp 19 maja 2009]. Dostępny w World Wide Web: <http://manyeyes.alphaworks.ibm.com/manyeyes/>.

<sup>22</sup> *Chaomei Chen's Homepage* [on-line] [dostęp 19 maja 2009]. Dostępny w World Wide Web <http://www.pages.drexel.edu/~cc345/>.

za które się uważa następujące: czerwony, zielony, niebieski, żółty, różowy, morski, pomarańczowy, fioletowy oraz biały i czarny. W kodowaniu kolorowej informacji kluczowym może się okazać nie tylko kontrast, lecz i luminancja. Z perspektywy wizualizacji danych obrazy lub mapy powinny być konstruowane w palecie kanałów „kolorów przeciwnych”<sup>23</sup>: czerwony-zielony albo żółty-niebieski. Da się też zauważyć trend projektowania współczesnych interfejsów aplikacji infowis w dwóch gamach kolorów: cieplej lub zimnej.

Możemy uzyskać wizualizacje wielowymiarową posługując się różnymi mapami kolorów. Kolory są dobrze rozpoznawalne i klasyfikowane w korze mózgowej człowieka i dzięki właściwej manipulacji ich parametrami – nasyceniem, kontrastem i głębią – możemy uzyskać bardzo dobre własności percepcyjne prezentowanych wizualizacji. Na korzyść koloru świadczą wyniki ewaluacji percepcji klastrów, gdzie za pomocą mapowania kolorem udało się zwiększyć liczbę wymiarów obserwowanych danych.

W symbolizowaniu danych zamiast dobrze znanych kropek i kółek z powodzeniem używa się różnego rodzaju **glify**<sup>24</sup>. W zależności od ilości takich symboli w zbiorze i czasu wyświetlanej sceny, jesteśmy w stanie zapamiętać od trzech do siedmiu jednostek.

Atrybuty i zmienne graficzne

Tabela 1.

Zmienna graficzna	Wymiarowość
Pozycja przestrzenna	3 wymiary: X,Y,Z
Kolor	3 wymiary zgodnie z teorią tróchromatyczności
Kształt	2 ub 3 w zależności od otoczenia
Orientacja	3 wymiary, w zależności od układu współrzędnych
Tekstura powierzchni	3 wymiary: orientacja, rozmiar i kontrast
Ruch	2 lub 3, może być przydatna faza
Miganie	1 wymiar (używa się wspólnie ze zmienną ruchu)

Źródło: Na podst. C. Ware. *Information Visualization: Perception for Design*. Wyd. 2. San Francisco: Morgan Kaufmann, 2004, s. 183.

Badania nad detekcją glifów świadczą, że powinno się je rozmieszczać w sąsiedztwie własnym, lecz z dala od innych obiektów. Uważa się, że ludzie postrzegają obrazy świata, sprowadzając je do określonych symboli, gdzie znaczenie ma np. kolor i kształt. Eksperymentalnie wykryto cechy istotne w projektowaniu informacji, i które są przydatne w wizualizacji kategoryzacji i klasyfikacji obiektów to:

- forma: orientacja, szerokość i rozmiar linii; *współliniowość*<sup>25</sup>, krzywizna, grupowanie przestrzenne, rozmycie, znakowanie, liczebność;
- kolor: jaskrawość, jasność;

<sup>23</sup> M. L. Hurvich, D. Jameson. *An opponent-process theory of color vision*. Psychological Review 1957, Vol. 64 (6), s. 384-404.

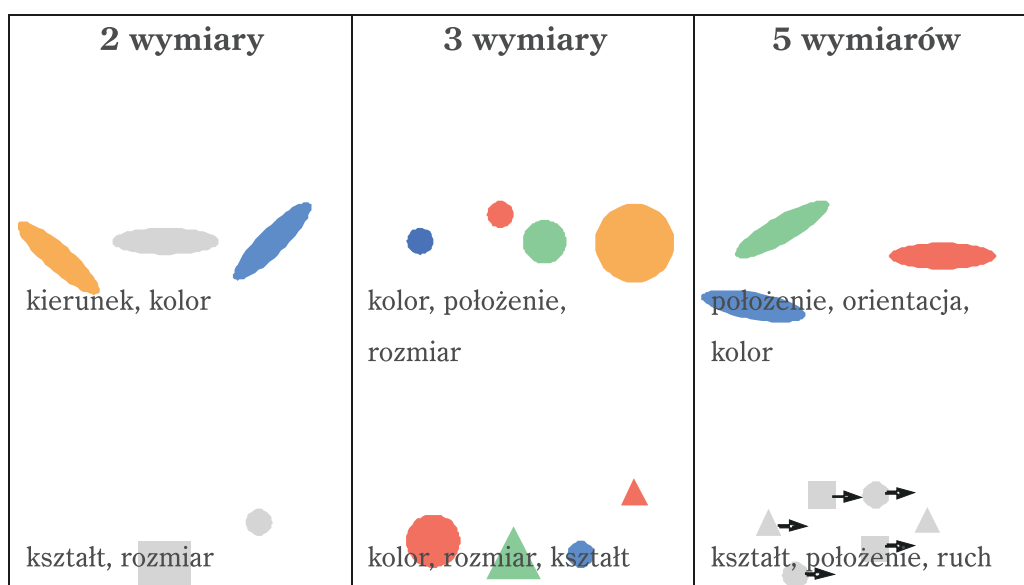
<sup>24</sup> *Glif* – obiekt graficzny, w typografii określany jako kształt przedstawiający w określonym kroju pisma konkretny graf lub symbol.

<sup>25</sup> *Współliniowość* – cecha, która determinuje czy dane punkty leżą do jednej krzywej.



- ruch: kierunek, migotanie;
- pozycja przestrzenna: 2D, głębokość stereoskopowa, wypukłość/wklęsłość.

Widzimy, iż do mapowania danych wielowymiarowych oprócz koloru mogą także służyć orientacja, rozmiar, pozycja elementów, tekstura, prosty ruch. Są to niskopoziomowe kanały informacji, działające osobno. Problem jest zwykle związany z wymiarem mapowanych danych. Tabela 1 prezentuje podstawowe atrybuty graficzne, które się stosuje w projektowaniu glyfów. Jak efektywnie zakodować dane wielowymiarowe najlepiej zobaczyć na praktycznych przykładach. Rysunek 6 ilustruje glyfy kodowane według zadanej ilości cech; opisane są również atrybuty mapujące zmienne.



**Rysunek 6.** Przykłady glyfów kodowanych według zadanej ilości cech

Źródło: Na podst. C. Ware. *Information Visualization: Perception for Design*. Wyd. 2. San Francisco: Morgan Kaufmann, 2004, s. 181.

Wizualizacja wielowymiarowych danych staje się jednym z ważniejszych czynników decydujących o właściwym zrozumieniu danych. Percepcja wizualna informacji złożonej posługuje się najbardziej naturalnymi drogami przetwarzania w ludzkim mózgu, wykorzystując również najstarszy i z tego względu najbardziej stabilny limbiczny szlak przetwarzania informacji.

## 1.2. Nowoczesne metody oraz przykłady wizualizacji informacji

Stosowane w aplikacjach techniki wizualizacyjne uwarunkowane są przede wszystkim rodzajem informacji, jej wymiarem i poziomem abstrakcji. Do klasycz-

nych technik zalicza się podstawowe formaty wizualne pośredniczące w przekazywaniu wiedzy, np. diagramy, listy, wykresy, tabele i macierze.

Dokładniejsze przyjrzenie się metodom wizualizacji warto jest rozpocząć od określenia rodzajów informacji, jaką ludzie potrafią zinterpretować. Poniżej zostaną scharakteryzowane wyróżniane typy informacji wraz z przykładami ich prototypów przestrzeni informacyjnej<sup>26</sup>:

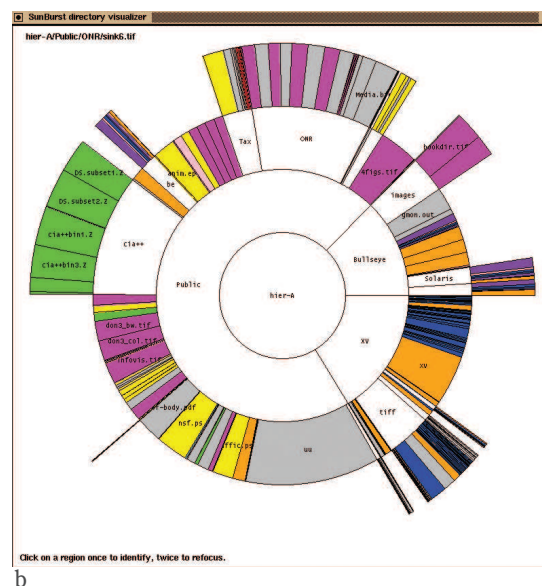
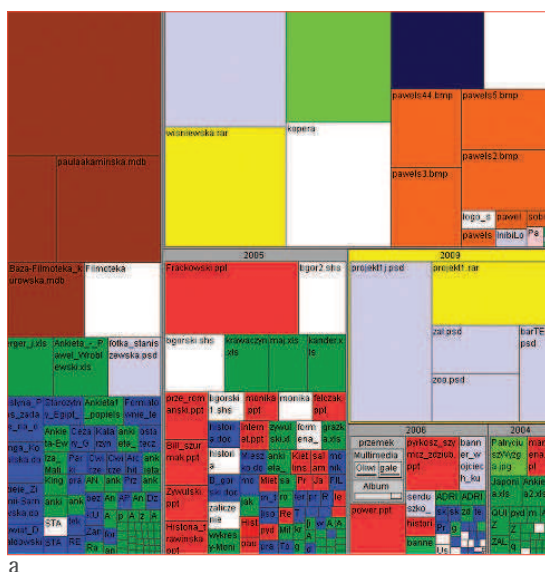
- Liniowa: listy alfabetyczne, chronologiczne, tabele, kody programów;
- Hierarchiczna: drzewa klasyfikacji, hierarchie struktury plików/katalogów;
- Sieci: topologie sieciowe, struktury **grafów**<sup>27</sup>, sieci semantyczne;
- Wielowymiarowa: dane kompleksowe, metadane, takie jak: typ, rozmiar, autor dokumentu itp.;
- Przestrzenie wektorowe: reprezentacja dokumentów za pomocą macierzy liczb w zagadnieniach wyszukiwania informacji (*Information Retrieval*),
- Przestrzenne: mapy topologiczne, obrazy 2D lub 3D, modele w systemach CAD.

Najprostszym typem informacji jest informacja liniowa, składająca się z sekwencji liczb i cyfr. Dane w postaci różnego rodzaju list i tabel, powszechne w historii piśmienniczej i obliczeniowej ludzkiej działalności, znane są jeszcze z czasów starożytnych. Znaki alfanumeryczne trudno jest przedstawić w innej formie niż tekst, np. graficznej. Nie przeszkadzało to, aby w latach dziewięćdziesiątych XX w. inżynierowie wiodących koncernów programistycznych poszukiwali nowych, na miarę ówczesnego rozwoju technologicznego, rozwiązań wizualizacji danych liniowych. Wartości liczbowe w tabelach zastąpili oni odpowiednią ilością kolorowych pikseli, w ten sposób powstawały kolorowe spektra, przedstawiające zależności co najwyżej dwóch wartości. Za przykłady mogą posłużyć wizualizacje wyników sondaży ankiet i statystyk odwiedzin portali sieciowych.

Informacja hierarchiczna jest najliczniejszą, wytypowaną grupą danych, ponieważ większość współczesnej informacji interpretowana jest poprzez struktury hierarchiczne. Hierarchia jest obecna w organizacji systemów katalogów i plików, bibliotecznych systemach klasyfikacji, danych genealogicznych, a również w definicjach klas języków programowania zorientowanego obiektowo. Hierarchiczne struktury drzewiaste najczęściej są prezentowane za pomocą **dendrogramów** (z greckiego: *dendron* – drzewo, *gramma* – rysować). Dendrogram w istocie przypomina rozgałęzione drzewo z tą różnicą, że korzeń (element główny) umieszczony jest zazwyczaj u samej góry, a liście (elementy najniższego poziomu) – na samym dole drzewa. Obiekty w dendrogramie łączone ze sobą za pomocą relacji pod- i nadrzędnych (ang. *parent-child*).

<sup>26</sup> V. Osińska: *Przybliżenie semantyczne w wizualizacji informacji w Internecie i bibliotekach cyfrowych*. Biuletyn EBIB [on-line] 2006, nr 7 (77) [dostęp 19 maja 2009]. Dostępny w World Wide Web: <http://www.ebib.info/2006/77/osinska.php>.

<sup>27</sup> *Grafy* – w matematyce: struktury składające się z wierzchołków i krawędzi; są wykorzystywane powszechnie w algorytmice.



Browse for: <input type="text" value="information technology"/> All of the keywords <input type="button" value="v"/> All languages <input type="button" value="v"/> Browse <input type="button" value="fewer options"/> Display captions in <input type="button" value="English"/> <input type="button" value="v"/>									
<input type="checkbox"/> At least 10,000 records <input type="checkbox"/> At least 1,000 records <input type="checkbox"/> At least 100 records <input type="checkbox"/> At least 10 records <input type="checkbox"/> At least 1 record <input type="checkbox"/> No records									
0	1	2	3	4	5	6	7	8	9
Computer science, information & general works	Philosophy & psychology	Religion	Social sciences	Language	Science	Technology	Arts & recreation	Literature	History & geography
00	01	02	03	04	05	06	07	08	09
Computer science, knowledge & systems	Bibliographies	Library & information sciences	Encyclopedias & books of facts	[Unassigned]	Magazines, journals & serials	Associations, organizations & museums	News media, journalism & publishing	Quotations	Manuscripts & rare books
000	001	002	003	004	005	006	007	008	009
Computer science, information & general works	Knowledge	The book	Systems	Data processing & computer science	Computer programming, programs & data	Special computer methods	[Unassigned]	[Unassigned]	[Unassigned]

**Rysunek 8.** Interfejs aplikacji on-line autorstwa *OCLC DeweyBrowser* do przeszukiwania zbiorów bibliotecznych. Kolorowe komórki tabel są odpowiednikami klas i podklas trzech poziomów klasyfikacji KDD.

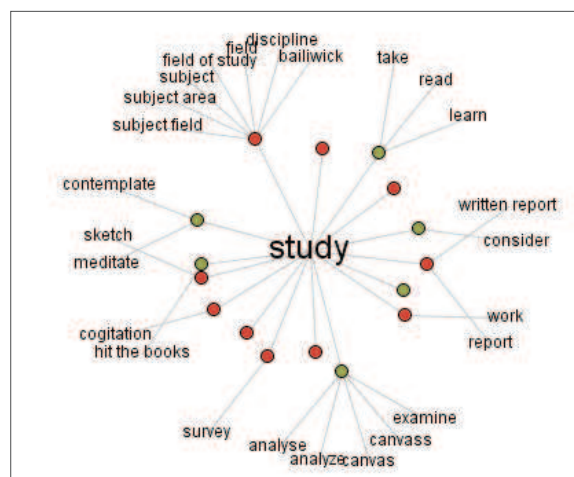
**Źródło:** V. Osińska. *Przybliżenie semantyczne w wizualizacji informacji w Internecie i bibliotekach cyfrowych*. Biuletyn EBIB [on-line] 2006, nr 7 (77) [dostęp 19 maja 2009]. Dostępny w World Wide Web: <http://www.ebib.info/2006/77/osinska.php>.

Poziomy struktury hierarchicznej w innych projektach przedstawiano również w postaci oddzielnych tabel. Korporacja *OCLC (On-line Computer Library Center)* nadzorująca rozwój klasyfikacji Dewey'a udostępniła na swoim portalu eksperymentalne oprogramowanie pod nazwą *DeweyBrowser*<sup>31</sup>, które umożliwia użytkownikom wyszukiwanie i przeglądanie zasobów bibliotecznych zorganizowanych zgodnie z klasyfikacją KDD. W tej aplikacji używa się hierarchii tabel do reprezentacji trzech górnych poziomów klasyfikacji. Rysunek 8 ilustruje rzut ekranowy programu w odpowiedzi na zapytanie „Information Technology”.

Informacja o liczebności zbiorów w każdej z klas i podklas przekazywana jest za pomocą kolorów. W odróżnienie od dendrogramu, w sieciowych strukturach powiązania istnieją nie tylko w kierunku góra-dół, lecz także pomiędzy węzłami równorzędnymi. Powszechnie obserwujemy, że wiele domen rzeczywistości takich jak systemy danych geograficznych przedstawia się za pomocą węzłów i wektorów, czyli grafów. Liczne przykłady można znaleźć w aplikacjach sieciowych: hiperłącza w dokumentach WWW, mapy powiązań wyrazów bliskoznacznych w tezaurusach, relacje pomiędzy tabelami w bazach danych, algorytmy, procesy technologiczne i logistyczne, struktury organizacyjne firm, scenariusze lekcyjne itp.

<sup>31</sup> *OCLC DeweyBrowser Beta v.1.0*. OCLC Online Computer Library Center [on-line] [dostęp 19.05.2009]. Dostępny w World Wide Web: <http://deweybrowser.oclc.org/ddcbrowser/wcat>.





**Rysunek 9.** Mapa tezausa dla wyrazu „study” w programie *VisualThesaurus* autorstwa firmy *ThinkMap*. Kolory punktów czerwony, niebieski i żółty wskazują rzeczowniki, czasowniki i przymiotniki odpowiednio.

**Źródło:** *Thinkmap Visual Thesaurus* [on-line] [dostęp 19.05.2009]. Dostępny w World Wide Web: <http://www.visualthesaurus.com/>.

Rozwiązania map grafopodobnych wykorzystują interfejsy programów edukacyjnych. Oprogramowanie *Visual Thesaurus*<sup>32</sup> jest zintegrowanym słownikiem i tezaurem w zakresie języka angielskiego. Interaktywny interfejs pozwala użytkownikowi na naukę poprzez eksplorację wyników zapytania. Na Rysunku 9 przedstawiony jest zrzut ekranowy wersji on-line programu. Diagram przedstawia mapę powiązań wyrazów bliskoznacznych dla słowa „study”. Kolory kółek czerwony, niebieski i żółty są zarezerwowane do oznakowania rzeczowników, czasowników i przymiotników odpowiednio. Linie ciągłe łączą wyrazy - synonimy. Według замыслу autorów, studenci i uczniowie za pomocą tego narzędzia mogą nie tylko nauczyć się nowych słów i pojęć, lecz ulepszyć swoje umiejętności czytania, pisanie i komunikacji.

**Informacja wielowymiarowa** stanowi najodpowiedniejszy zasób danych dla badań nad strukturami semantycznymi. Metadane niosą informację o danych dokumentu i jednocześnie zawarte są w samym dokumencie. Według standardu *Dublin Core*<sup>33</sup> do metadanych należą informacje o tytule, autorze, wydawcy dokumentu, słowach kluczowych, opisie, języku itp. Dokumenty WWW przechowują te parametry w polach meta, opisywanych za pomocą znaczników.

Wraz z sukcesem wyszukiwarki Google firmy komercyjne intensywnie rozwijające oprogramowanie wizualizacyjne, takie jak: *KartOO*<sup>34</sup>, *Grokker*<sup>35</sup>, *AquaBrowser*<sup>36</sup>, *Vivisimo*<sup>37</sup> w celu pozyskania nowych klientów, zaczęły profilować swoje produk-

<sup>32</sup> Wersja demo dostępna pod adresem: : <http://www.visualthesaurus.com> [dostęp 19 maja 2009].

<sup>33</sup> *Dublin Core (Dublin Core Metadata Element Set, DC)* – ogólny standard metadanych do opisu zasobów (np. bibliotecznych). DC definiuje 15 prostych elementów, np. tytuł, autor, słowa kluczowe, data itp. Wykorzystywany jest w Documnet Type Definition, formatach RDF, XML. Por. *Dublin Core*. W: *Wikipedia. The Free Encyclopedia* [on-line] [dostęp 19 maja 2009]. Dostępny w World Wide Web: [http://pl.wikipedia.org/wiki/Dublin\\_Core](http://pl.wikipedia.org/wiki/Dublin_Core).

<sup>34</sup> Wyszukiwarka dostępna pod adresem: [www.kartoo.com](http://www.kartoo.com) [dostęp 19 maja 2009].

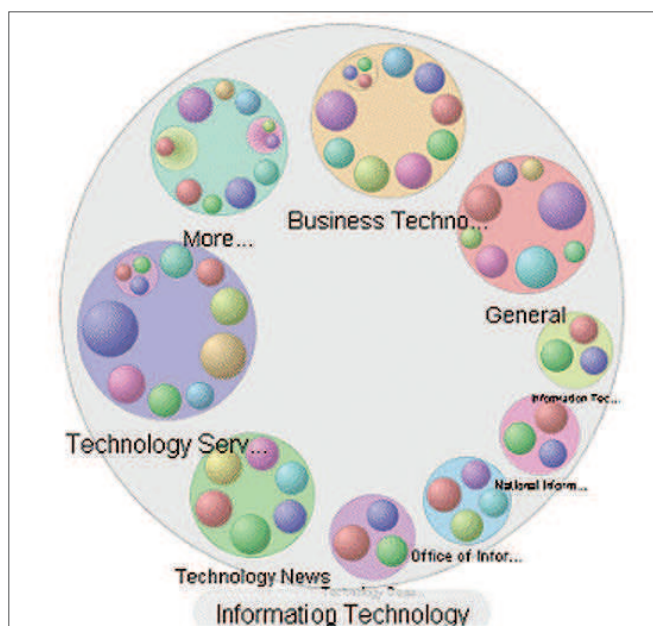
<sup>35</sup> Wyszukiwarka dostępna pod adresem: <http://www.grokker.com/> [dostęp 19 maja 2009].

<sup>36</sup> Wyszukiwarka *AquaBrowser* dostępna pod adresem: [www.medialab.nl/](http://www.medialab.nl/) [dostęp 19 maja 2009].

<sup>37</sup> *Vivisimo search engine* dostępny pod adresem: <http://vivisimo.com/> [dostęp 19 maja 2009].

ty w kierunku integracji zadań wyszukiwania i nawigacji<sup>38</sup>. W zależności od koncepcji autorów i zastosowanych metod wizualizacji użytkownik może zapoznać się nie z listą rankingową, lecz z wielowymiarową przestrzenią nawigacyjną. Zgodnie z założeniem większej swobody w nawigacji, powinien on również mieć możliwość kolekcjonowania wyselekcjonowanych elementów. Tu można przytoczyć analogię do koszyka zakupów w sklepie internetowym. W takich wielowymiarowych mapach odrębne znaczenie przyjmują kolor, kształt, rozmiar, pozycja oraz połączenia obiektów.

Firma *Groxis*, działająca od 2001 r. zaprojektowała wyszukiwarkę z graficznym interfejsem *Grokker*, której w mediach przepowiadano konkurencyjną przyszłość wobec Google. Aplikacja ta korzysta z baz danych serwisów Yahoo, *ACM Digital Library*<sup>39</sup> i/lub *Amazon Books*<sup>40</sup>. Kolorowe koła wewnątrz innych kół (mogą to być też kwadraty) są odpowiednikami klas i podklas (Rysunek 10). Przy najechaniu myszką na obiekt w polu objaśnienia wyświetlane są metadane dla wybranego zasobu, takie jak tytuł, autor, czas utworzenia itp. Użytkownicy mogą posortować wyniki według dziedziny oraz zachować je do późniejszego wykorzystania. Zastosowanie filtrów powoduje zawężenie wyników. Ciekawostką jest to, że *Grokker* jest w stanie pokategoryzować pliki z naszego dysku według zawartości, pomijając informację o przynależności do folderów.



**Rysunek 10.** Wygenerowana mapa skojarzonych tematycznie obszarów z wyrażeniem „Information Technology” w wyszukiwarce *Grokker*.

Źródło: *Information Technology* [on-line]. Grokker, Inc. [dostęp 19 maja 2009]. Dostępny w World Wide Web: [http://live.grokker.com/grokker.html?query=information%20technology&OpenSearch\\_Yahoo=true&Wikipedia=true&numResults=250](http://live.grokker.com/grokker.html?query=information%20technology&OpenSearch_Yahoo=true&Wikipedia=true&numResults=250).

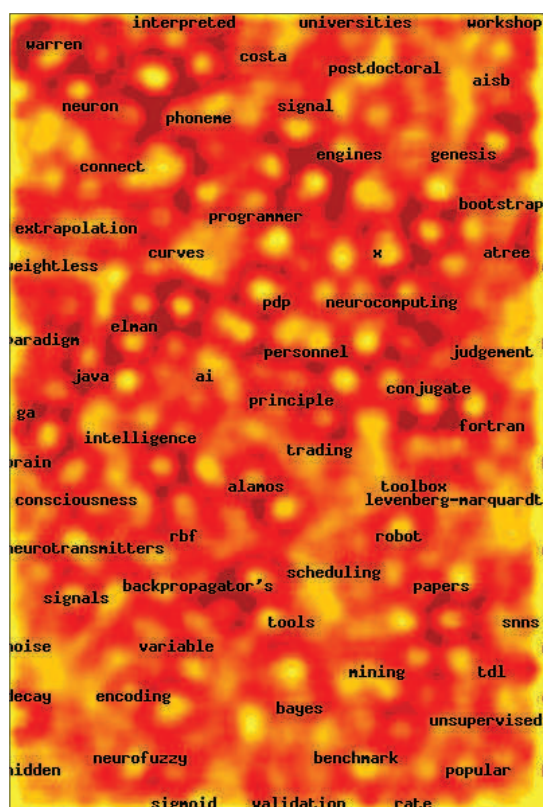
Niejednorodna struktura bardzo licznych zasobów w sieci wymaga od systemów wizualizacji umiejętności wykrywania i reprezentowania złożoności tych danych. Pierwszym krokiem do wizualnej analizy dużych zbiorów danych jest automatyczna klasteryzacja zgodnie z miarą ich podobieństwa. W tym podejściu używa się staty-

<sup>38</sup> J. Luther, M. Kelly, D. Beagle, dz. cyt.

<sup>39</sup> Strona dostępna pod adresem: <http://www.acm.org/dl> [dostęp 19 maja 2009].

<sup>40</sup> Strona dostępna pod adresem: <http://www.amazon.com> [dostęp 19 maja 2009].

styczno-lingwistycznych algorytmów, uczenia się maszynowego i sztucznych sieci neuronowych, aby na bieżąco określić tematyczne kategorie zasobów. W wizualizacji grupowanych obiektów bardzo przydatne są tak zwane mapy samoorganizujące się (ang. *Self Organizing Maps* - *SOM*), rodzaje sztucznych sieci neuronowych o szerokim zastosowaniu. *SOM* zostały rozwinięte od roku 1982 przez T. Kohonena; przez co znane też są pod nazwą sieci Kohonena. Rysunek 11 ilustruje semantyczną mapę grup listy dyskusyjnej comp.ai.neural.nets, wygenerowanej na portalu grupy fińskich naukowców *WebSOM*<sup>41</sup>. Przy modulowaniu reprezentacji semantycznych w zadaniach filtrowania i wyszukiwania informacji wykorzystywany jest wektorowy model przestrzeni wielowymiarowej



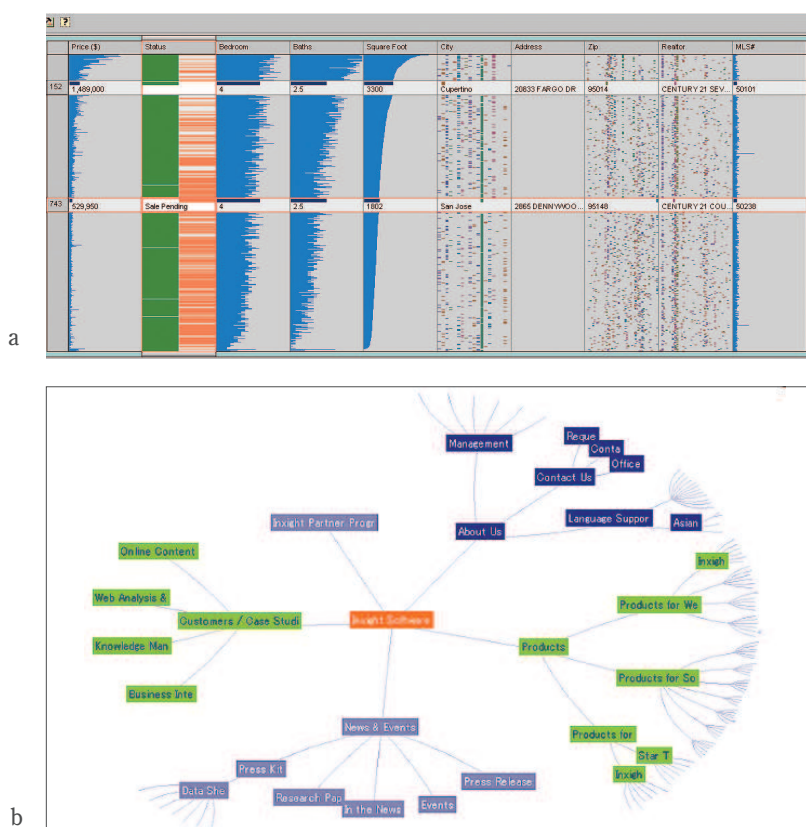
Rysunek 11. Grupowanie mapy *WebSOM* grupy dyskusyjnej comp.ai.neural.nets

Źródło: *WEBSOM map* [on-line] [dostęp 19 maja 2009]. Dostępny w World Wide Web: <http://websom.hut.fi/websom/comp.ai.neural-nets-new/html/root.html>.

(*Vector Space Modelling* – *VRM*)<sup>42</sup>. Zagadnienia informacji wielowymiarowej w oparciu o przestrzeń wektorową opisuje **semantyka wektorowa** (ang. *vectorial semantics*). Dokumenty są przedstawiane w sposób formalny przy użyciu **wektorów cech**, za które mogą nam posłużyć np. słowa kluczowe, sekwencje słów, odległość pomiędzy wyrazami, występowanie spójników, topologia obiektów w dokumentach, formaty i rozmiary plików itp. Procedurę tworzenia modelu przestrzeni wektorowej można podzielić na trzy etapy. Pierwszym jest indeksowanie dokumentów i wyłonienie słów oddających treść dokumentu.

<sup>41</sup> *WEBSOM – A novel SOM-based approach to free-text mining* [on-line] [dostęp 19 maja 2009]. Dostępny w World Wide Web: <http://websom.hut.fi/websom/>.

<sup>42</sup> I. Larsen. *Vector Space Modeling* [on-line]. THOR Center for Neuroinformatics [dostęp 19 maja 2009]. Dostępny w World Wide Web: <http://eivind.imm.dtu.dk/thor/projects/multimedia/textmining/node5.html>.



**Rysunek 12.** Strategie powiększenia przestrzeni informacyjnej firmy *Inxight*:

- a) *TableLens* – wybrany fragment danych jest powiększany podczas eksploracji całej tabeli;  
b) *StarTree* – projekcja hiperboliczna *fish-eye* na wybrany fragment danych.

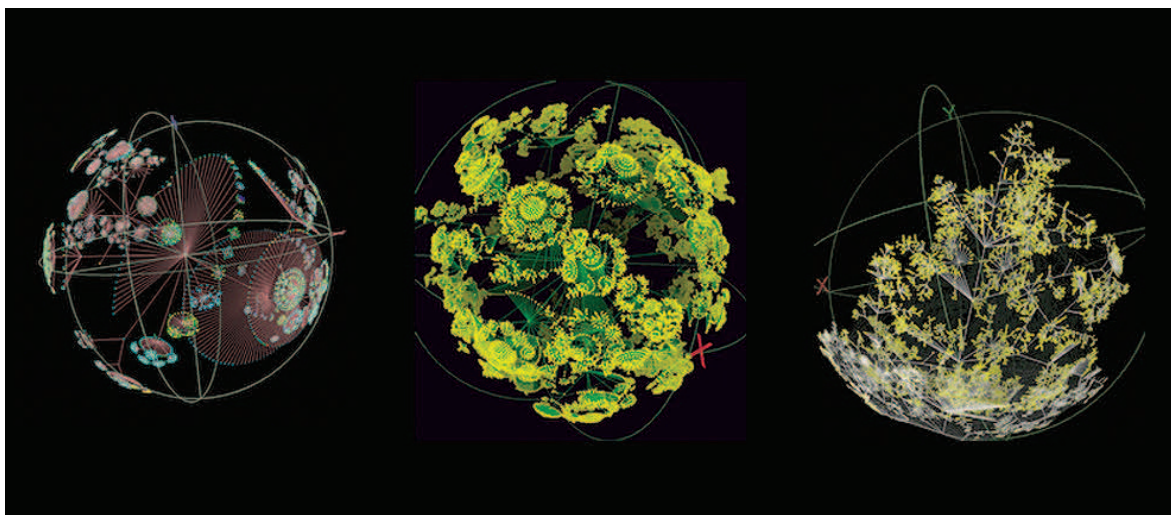
**Źródło:** *Inxight* [on-line] [dostęp 31 sierpnia 2008]. Dostępny w World Wide Web: <http://www.inxight.com> (adres aktualny do grudnia 2008 r.).

Na drugim etapie zachodzi ważenie słów indeksowanych, czyli określenie, w jakim stopniu termin jest ważny dla dokumentu w odniesieniu do zapytania. Na koniec ustalana jest pozycja rankingowa dokumentu na liście odpowiedzi.

W procesach przeglądania i wyszukiwania danych dużą rolę w aplikacji odgrywa przestrzeń eksploracyjna, która jest ograniczona oknem monitora. Zapotrzebowanie na wyjawienie oraz śledzenie szczegółów (niski poziom informacyjny) na obrazie koliduje z informacją kontekstową wysokiego poziomu. Użytkownik zazwyczaj wymaga zapewnienia obu poziomów. Problem ten występuje pod nazwą **focus+context**.

Znane rozwiązania to: użycie wielu widoków w osobnych oknach (*multiply view*), powiększenie/pomniejszenie fragmentów obiektów (ang. *zooming*, np. technika *table lens*) oraz zniekształcenia geometryczne (rybie oko, ang. *fish-eye*) zaprezentowane odpowiednio na Rysunkach 12a, b. Geometrycznym sposobem na rozciągnięcie obszaru eksploracji jest reprezentacja hierarchicznych struktur w przestrzeni hiperbolicznej. Pierwszymi aplikacjami, które wykorzystały technikę **fish-eye** były przeglądarki hiperboliczne. Przestrzeń Euklidesową zastępuje się hiperboliczną, którą rzutuje się na kolisty obszar widzenia. Ten mechanizm zapewnia więcej miejsca na wizualizację hierarchii (obwód koła rośnie wykładniczo z promieniem, co oznacza, że ze wzrostem odległości mamy eksponencjalne powiększenie przestrzeni).



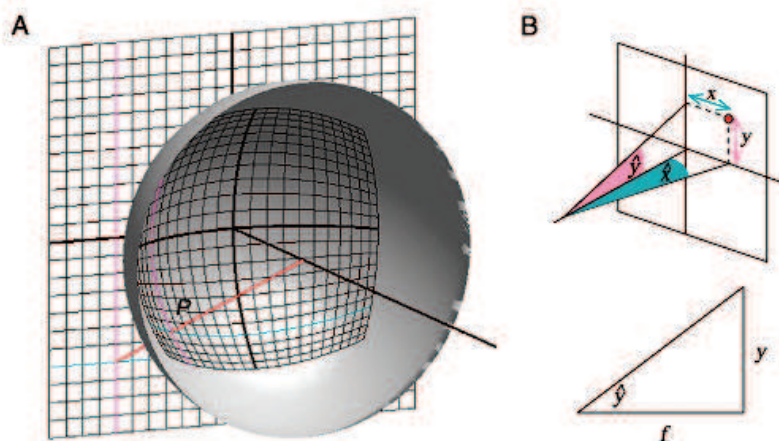


**Rysunek 13.** Wizualizacje sieci z ilością węzłów ponad 50 tys. w programie *Walrus*

**Źródło:** *Walrus – Gallery: Visualization & Navigation* [on-line] CAIDA, the Cooperative Association for Internet Data Analysis [dostęp 19 maja 2009]. Dostępny w World Wide Web: <http://www.caida.org/tools/visualization/walrus/gallery1/>.

W automatycznej klasteryzacji pozycjonowanie dokumentów zachodzi w kierunku dół-góra: od najniższego poziomu do najwyższego. Próba wizualizacji takiej struktury w przestrzeni trójwymiarowej prowadzi do umieszczenia głównego węzła w centrum, a węzłów podrzędnych we wszystkich kierunkach wokół środka (program *Walrus*<sup>43</sup>, dla którego na Rysunku 13 podane są przykłady wizualizacji). Narzędzie *Walrus* jest otwartym oprogramowaniem do interaktywnej wizualizacji zorientowanych grafów o dużej ilości węzłów w przestrzeni trzywymiarowej. Poprzez zastosowanie zniekształcenia *fisheye* zapewnienia jednocześnie wyświetlanie zarówno szczegółów, tak i całego kontekstu. Charakterystyczną obecnie tendencją w projektowaniu graficznym jest korzystanie z przestrzeni 3D. Znajdujemy sporo argumentów przemawiających za ostateczną dominacją systemów wizualizacji przestrzennej. Naturalnym jest stwierdzenie, że żyjemy w świecie trójwymiarowym i nasz mózg jest przystosowany do interakcji właśnie w trzech wymiarach. Obraz pierwotny odwzorowywany na sferycznej siatkówce (Rysunek 14) również posiada cechy kulistej struktury 3D. Wydaje się więc uzasadnione wykorzystywanie sferycznych metod wizualizacji danych dostosowanych do naturalnych predyspozycji ludzkiego aparatu wizualnego. Jednak jak do tej pory wizualizacja dwuwymiarowa pozostaje ciągle najprostszym sposobem przedstawiania wyników, gdyż jest zazwyczaj umieszczana na dwuwymiarowej kartce papieru lub podobnej płaszczyźnie monitora komputerowego.

<sup>43</sup> *Walrus – Graph Visualization Tool* [on-line] CAIDA, the Cooperative Association for Internet Data Analysis [dostęp 19.05.2009]. Dostępny w World Wide Web: <http://www.caida.org/tools/visualization/walrus/>.



**Rysunek 14.** Reprezentacja obrazu na siatkówce. (A) Mapowanie powierzchni na sferze geometrycznej układu siatkówki. (B) Sposób konwersji obrazu sferycznego do płaszczyzny projekcyjnej.

Źródło: J. C. A. Read, B. G. Cumming. *Does depth perception require vertical-disparity detectors?* Journal of Vision 2006, Vol. 6 (12), A. 1, s. 1327.

Tabele z dużą ilością kolumn zawierających liczby są zazwyczaj zupełnie nieczytelne dla użytkownika – dopiero ich prezentacja na wykresach w postaci zbioru punktów lub linii staje się istotnym elementem procesów kognitywnych zachodzących w mózgu. I chociaż wykresy takie dobrze wykorzystują własności kory wzrokowej człowieka, która posiada wyspecjalizowane obszary do analizy krawędzi – poziomych, pionowych i pochyłych – to już jednak wizualizacje przedstawiające rozproszone zbiory punktów bez ich wcześniejszej konglomeracji stają się często nic nie znaczącym szumem. Tak więc zwiększenie przestrzeni o jeden wymiar znacząco rozciąga zakres informacji wizualnej. Kolejnym powodem zainteresowania wizualizacją przestrzenną, to zaawansowanie technologii graficznych 3D, które stają się coraz mniej kosztowne. Systemy rzeczywistości wirtualnej cieszą się dużą popularnością wśród miłośników gier komputerowych. Nie bez znaczenia jest również efekt estetyczny, który wywołują umiejętnie zilustrowane sceny 3D.

Aktualne prace nad rozwojem metod wizualizacji oprócz doskonalenia aplikacji 3D przewidują także wprowadzenie wymiaru czasowego i tym samym stworzenie dynamicznej eksploracji, wspomagającej badanie zmian w zasobach informacyjnych, np. w piśmiennictwie, bibliotekach cyfrowych, serwisach sieciowych.

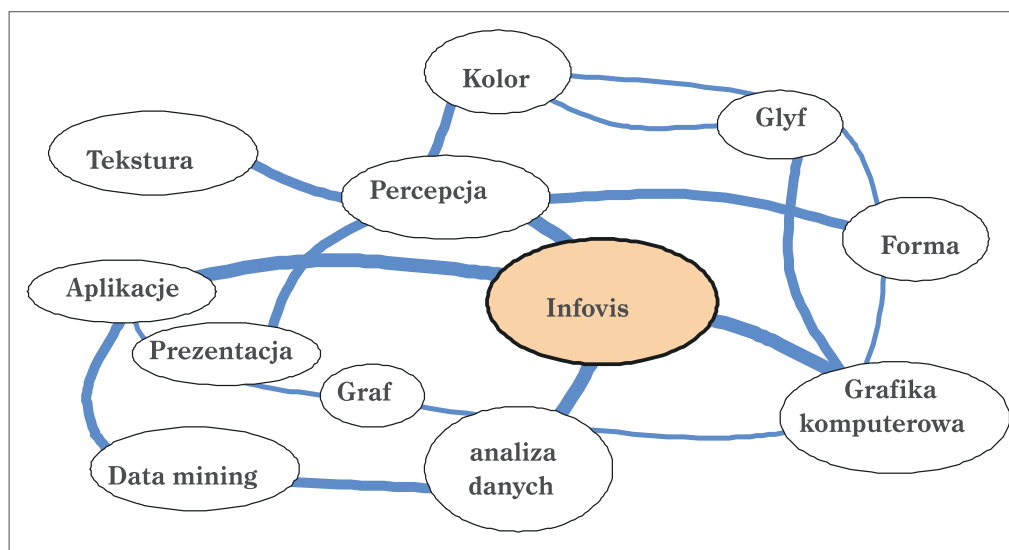
W mnogości opisanych sposobów graficznych reprezentacji, technik i coraz to nowych idei nie jest łatwo o porządną systematykę tych metod. Przedstawione w tym rozdziale funkcjonalne podejście odwołuje się do rodzaju informacji, wykrycia: czy to liniowa, hierarchiczna, sieciowa albo wielowymiarowa informacja? Na obecnym etapie rozwoju metodologii Infovis da się wyodrębnić główne modele wizualizacji:

- grafy, czyli węzły i linki;
- algorytmy wypełnionej przestrzeni takie jak treemap, mapy SOM;
- krajobrazy informacji, często używane przez naukowców w latach dziewięćdziesiątych ubiegłego wieku.

A zatem, metody wizualizacji danych ewoluowały od interfejsów programów z minimalną ilością elementów graficznych, wykorzystujących zagnieżdżone drze-

wa klasyfikacji, tabele oraz wykresy dwuwymiarowe, przez diagramy relacji między dokumentami przy użyciu takich abstrakcyjnych kształtów jak koła, kwadraty, linie oraz łącza, do przeglądarek hiperbolicznych i geoprzestrzennych map z włączoną osią czasu.

Próba wyselekcjonowania głównych typów informacji nie oznacza, że nie możemy spotkać w życiu przykładów kombinowanych albo zmieniających typów danych. Ponieważ w świecie zachodzą dynamiczne procesy, a zatem informacja też bezustannie się zmienia, pomiędzy pierwotnie równorzędnymi elementami niosącymi informację też mogą powstawać relacje hierarchiczne. W analizie informacji płynącej z otaczającego nas świata dążymy do sklasyfikowania występujących w rzeczywistości obiektów w grupy – klasy. Czynimy to na podstawie wspólnych ich cech (wygląd, przeznaczenie, pochodzenie itp.) lub zachowań (co obiekt może wykonać?). Obiekty – w trakcie poznawania coraz większej ich – ilości grupujemy w klasy, a następnie klasy nadrzędne. Dążenie do hierarchizacji elementów informacji jest więc naturalnym objawem, który sygnalizuje potrzebę mapowania przeszukiwanych wyników. Mapy, które obecnie są najpopularniejszym rozwiązaniem w sieciowych projektach wizualizacji informacji, znalazły szerokie spektrum zastosowań, przez co dzielimy je na kartograficzne, koncepcyjne (semantyczne) i domenowe<sup>44</sup>. Na koniec w ramach autodoświadczenia w percepcji map dla czytelnika zamieszczony został przykład mapy koncepcyjnej Infovis (Rysunek 15).



Rysunek 15. Mapa koncepcyjna Infovis

Źródło: Na podst. C. Ware. *Information Visualization: Perception for Design*. Wyd. 2. San Francisco: Morgan Kaufmann, 2004, s. 368.

### 1.3. Wizualizacja informacji w poszukiwaniu strategii mapowania nauk (Mapping Science)

W obliczu integracji wielu dyscyplin naukowych wykrywanie obszarów zainteresowania różnych dyscyplin naukowych staje się praktyką, która może przyczynić się do stymulacji badań interdyscyplinarnych. Aby lepiej zrozumieć strukturę i dy-

<sup>44</sup> *Exhibit Purpose and Goals* [on-line]. *Places@Spaces: Mapping Science* [dostęp 19 maja 2009]. Dostępny w World Wide Web: <http://scimaps.org/>.

namikę rozwoju poszczególnych działów nauki oraz znaleźć sposób wykrycia w nich trendów tematycznych, naukowcy posługują się wskaźnikiem „przesunięcia granic naukowych” (ang. *scientific frontiers*) To pole badawcze, luźno zdefiniowane jako **wizualizacja dyscyplin wiedzy lub nauki** (ang. *Knowledge Domain Visualisation – KDVis*), jest przedmiotem badań od zaledwie dziewięciu lat (Börner i in. 2003, Chen 2003). Ponieważ stosowane tutaj metody prowadzą do generowania map graficznych, równolegle istnieje inne nazewnictwo: **mapowanie nauk** (ang. *Mapping Science*)<sup>45</sup>, lub rzadziej używana: **naukografia** (ang. *Scientography*)<sup>46</sup>. Ostatnie pojęcie zostało wprowadzone w 1960 r. przez Eugene Garfielda, założyciela Instytutu Filadelfijskiego (*Institute of Scientific Information – ISI*, obecnie znanego jako *Thomson Scientific*)<sup>47</sup>. Na swoim portalu<sup>48</sup> umieścił on szereg artykułów, poświęconych problemom i przykładom naukografii (Garfield 1998). Mapowanie nauki jako metodologia wyrosła z tradycji bibliometrycznych. Wykorzystuje ona najbardziej znane bibliograficzne bazy danych instytutu *ISI*, czyli indeksy cytowań poszczególnych artykułów publikowanych w czasopismach naukowych (*ISI Citation Index*), jak również sumaryczny indeks liczby i dynamiki cytowań wszystkich artykułów w danym czasopiśmie *Journal Citation Reports – JCR*. W ten sposób mapowane są nauki ścisłe, nauki społeczne oraz nauki humanistyczne<sup>49</sup>. Przegląd historyczny oraz opinie, w jaki sposób nauka była mapowana od dziesięcioleci, przedstawiono szczegółowo w serii artykułów<sup>50</sup>.

Do klasycznych technik mapowania zalicza się przestrzenną reprezentację danych bibliometrycznych (autorów, dokumentów, czasopism, kategorii) w oparciu analizę wspólnych cytowań autorów (ang. *author co-citation analysis*) lub czasopism (ang. *document co-citation analysis*). Metoda ta zainicjowana została w latach siedemdziesiątych XX w. w pracach Marshakowej<sup>51</sup>, Small i Griffitha<sup>52</sup> oraz Garfielda<sup>53</sup>. Dwa dokumenty są wspólnie cytowane, jeśli równocześnie występują w wykazie bibliograficznym trzeciego dokumentu. Podobieństwo pomiędzy dokumentami

<sup>45</sup> Więcej: *Science Frontiers 1976–* [on-line] [dostęp 19 maja 2009]. Dostępny w World Wide Web: <http://www.science-frontiers.com/>.

<sup>46</sup> E. Garfield: *Scientography: Mapping the tracks of science*. Current Contents: Social & Behavioural Sciences 1994, nr 7(45), s. 5-10.

<sup>47</sup> *Science – Thompson Reuters* [on-line] [dostęp 19 maja 2009]. Dostępny w World Wide Web: <http://scientific.thomson.com/>.

<sup>48</sup> *Eugene Garfield, Ph. D. Home Page* [on-line] [dostęp 19 maja 2009]. Dostępny w World Wide Web: <http://garfield.library.upenn.edu/>.

<sup>49</sup> K. Börner, Ch. Chen, K.W. Boyack: *Visualizing Knowledge Domains*. W: Blaise Cronin (red.). *Annual Review of Information Science & Technology*. Medford, NJ: Information Today, Inc./American Society for Information Science and Technology 2003, Vol. 37, s. 180-185; F. Moya-Anegón i in. *A new technique for building maps of large scientific domains based on the cocitation of classes and categories*. *Scientometrics* 2004, Vol. 61, nr 1, s. 129-145; E.C.M. Noyons, H.F. Moed. *Combining Mapping and Citation Analysis for Evaluative Bibliometric Purposes: A Bibliometric Study*. *Journal of the American Society for Information Science* 1999, nr 50(2), s. 115-131; E. Garfield. *Essays/Papers on „Mapping the World of Science”* [on-line]. *Eugene Garfield, Ph. D. Home Page* [dostęp 19 maja 2009]. Dostępny w World Wide Web: <http://garfield.library.upenn.edu/mapping/mapping.html>.

<sup>50</sup> F. Moya-Anegón i in., dz. cyt., s. 130-133; Boyack K. W. i in. *Mapping the backbone of Science*. *Scientometrics* 2005, Vol. 64, nr 3, s. 353-354; E.C.M. Noyons, H.F. Moed. dz. cyt., s. 115-116.

<sup>51</sup> I. V. Marshakova: *A system of document connection based on references*. *Scientific and Technical Information Serial of VINITI* 1973, Vol. 6 (2), s. 3-8.

<sup>52</sup> H. Small. *Co-citation in the scientific literature: A new measurement of the relationship between two documents*. *Journal of the American Society of Information Science* 1973, Vol. 24 (4), s. 265-269; H. Small, B. C. Griffith. *The structure of scientific literatures I: Identifying and graphing specialities*. *Science Studies* 1974, nr 4, s. 17-40.

<sup>53</sup> E. Garfield: *Essays/Papers on „Mapping the World of Science”...*



jest obliczane na podstawie częstości ich wspólnych cytowań. W innej metodzie bibliografii sprzężonych (ang. *bibliographic coupling*) – dwie prace odwołują się w bibliografii do tych samych dokumentów, których ilość determinuje podobieństwo. Analiza może być oparta również o powiązane cytowania lub przypisy kilku autorów: *inter-citation*. Koncepcja prefiksów *co-* i *inter-* jest dobrze udowodniona w pracy White & McCain<sup>54</sup>. Pierwszy implikuje wspólne występowanie badanych jednostek, drugi – ich wzajemne powiązania. Analiza cytowań pozwala na rozpoznawanie dominujących obszarów badań w danej dziedzinie, jak również struktur społecznych wewnątrz i na zewnątrz środowisk naukowych. O aspekcie społecznym takiej analizy wielokrotnie pisał B. Hjørland – wybitny znawca problematyki organizacji wiedzy oraz traktujący analizę domenową jako jeden z podstawowych paradygmatów informacji naukowej<sup>55</sup>. Ścieżki koo-cytowań mogą także ujawniać kierunki integracji interdyscyplinarnych.

Od roku 2008 w celu zademonstrowania 10-letnich osiągnięć w dziedzinie mapowania informacji uruchomiono wirtualną wystawę *Places@Spaces: Mapping Science*<sup>56</sup>. Portal redagowany jest przez K. Börner, *School of Library and Information Science, Indiana University*, której zespół prowadzi zaawansowane badania nad metodologiami wizualizacji dziedzin nauki, obszarów współpracy naukowców oraz bibliotek cyfrowych. Jak podkreśla Ch. Chen<sup>57</sup> – główny inicjator badań nad *KDViz* – do rzetelnej analizy wyników potrzebne są połączone wysiłki specjalistów różnych dziedzin: od filozofii nauki i naukometrii do zarządzania wiedzą i eksploracją danych. Perspektywę interdyscyplinarną zapewni tylko współpraca wielu naukowców. Nauka jest społecznym fenomenem, na który składają się poglądy i obserwacje badaczy. Przedmiotem badań wizualizacji dyscyplin wiedzy są tak zwane sieci naukowe, składające się ze współpracy naukowej, powiązań cytowań w piśmiennictwie naukowym, frontów i zakresów badań oraz kierunków rozprzestrzeniania się (dyfuzji) wiedzy.

Nauka rozwija się za pomocą specyficznych metod naukowych nazywanych też **paradygmatami nauki**. Uchwycenie dynamiki naukowych paradygmatów można zbadać za pomocą analizy cytowań w literaturze naukowej, co jest szczegółowo opisywane w wiodących pracach na temat *KDViz*<sup>58</sup>. W książce *The Structure of Scientific Revolutions* opublikowanej w 1962 r. T. S. Kuhn dowodzi, że nauka nie jest jednostajnym, kumulatywnym pozyskiwaniem wiedzy<sup>59</sup>. Zamiast tego nauka *jest serią spokojnych okresów przerywanych przez gwałtowne intelektualne rewolucje, po których*

<sup>54</sup> H. D. White, K. W. McCain: *Visualization of literatures*. Annual Review of Information Science and Technology 1997, Vol. 32, s. 99-168.

<sup>55</sup> B. Hjørland, H. Albrechtsen: *Toward a new horizon in information science: domain analysis*. Journal of the American Society for Information Science (JASIS) 1995, nr 46, s. 400-425; B. Hjørland. *Domain analysis in information science: eleven approaches – traditional as innovative*. Journal of documentation 2002, nr 58, s. 422-462.

<sup>56</sup> *Exhibit Purpose and Goals...*

<sup>57</sup> Ch. Chen: *Mapping Scientific Frontiers. The Quest for Knowledge Visualization*. London: Springer-Verlag, 2003, s. 1-5.

<sup>58</sup> Tamże; Ch. Chen. *Information Visualization. Beyond the Horizon*. 2nd ed. London: Springer, 2006, s. 5.

<sup>59</sup> F. Pajares: *The Structure of Scientific Revolutions by Thomas S. Kuhn. Outline and Study Guide* [on-line]. Emory University. Division of Education Studies [dostęp 19 maja 2009]. Dostępny w World Wide Web: <http://www.des.emory.edu/mfp/Kuhn.html>.

*jeden koncepcyjny światopogląd jest zamieniany przez inny. W przedstawionym ujęciu rozwój nauki przebiega w kilku fazach:*

- W okresie przedparadygmatycznym nie ma między uczonymi zgody co do fundamentalnych podstaw danej dziedziny, dlatego ich badania nie są jednoznacznie ukierunkowane.
- W „normalnej” (dojrzałej) fazie - jakiejś szkole udaje się dokonać przełomu, który staje się obowiązującym paradygmatem, do którego przyłączają się inne szkoły.
- Następnie pojawiają się anomalie, których nie można rozwiązać w obrębie paradygmatu i które prowadzą do kryzysów. Następuje wówczas rewolucja naukowa i stary paradygmat zastępowany jest nowym.

Zgodnie z poglądami Kuhna, paradygmat kieruje wysiłkiem badawczym społeczności naukowych i jest tym kryterium, które najbardziej ściśle identyfikuje obszary nauk. Fundamentalnym argumentem Kuhna jest to, że dla dojrzałej nauki typową drogą rozwojową jest kolejne przechodzenie w procesie zmian dynamicznych od jednego do innego paradygmatu. Gdy ma miejsce zmiana paradygmatu, *świat naukowy zmienia się jakościowo i jest jakościowo wzbogacany przez fundamentalnie nowe zarówno fakty jak i teorie*<sup>60</sup>.

Badacze KDViz od lat z powodzeniem wykorzystują techniki wizualizacji informacji, aby wykryć zmiany krytyczne w rozwoju wiedzy naukowej. Teorie i metody mapowania zakresów badawczych nauk formowały się od dziesięcioleci, wykorzystując osiągnięcia wielu dyscyplin: filozofii i socjologii nauki, naukometrii, wyszukiwania informacji, wizualizacji informacji. Większość tradycyjnych metod polega na wykrywaniu wzorów sieci cytowań w dokumentach naukowych. Intelktualną strukturę społeczności naukowych można analizować na podstawie dynamiki tej sieci. Przy takim założeniu, obserwacja ewolucji dyscypliny naukowej jest możliwa poprzez modulowanie oraz wizualizację struktur cytowań. Wraz z nagromadzeniem piśmiennictwa naukowego badania nad siecią cytowań rozwinęły się od studiów pionierskich z wykorzystaniem technik niekomputerowych przed 1980 r. do zaawansowanych metod, bazujących na algorytmach statystycznych (od roku 2000).

Mapowanie nauk może pomóc w identyfikacji subdyscyplin lub obszarów badawczych i ich wzajemnych powiązaniach w ramach konkretnej dziedziny. Niektórzy utrzymują, iż analiza ewolucji sieci cytowań pozwala nawet na predykcję trendów badawczych oraz oszacowanie rozpiętości oddziaływań społecznych nauki w przyszłości<sup>61</sup>. Ponadto tworzone mapy nauk stymulują poznawanie współczesnego stanu wiedzy i mogą pomóc w dokonaniu odkryć naukowych.

Ch. Chen opisuje w jaki sposób zaobserwował dwa paradygmaty, przez które przeszła teoria superstrun<sup>62</sup>. Teoria strun jest jednym z najbardziej aktywnych tematów badawczych w fizyce teoretycznej. Tradycyjne pojęcie cząstki elementarnej zastąpiono struną, czyli drgającą wielowymiarową strukturą topologiczną czasami utożsamianą z odcinkiem linii lub kołem. Cząstki są strunami mającymi rozmiary

<sup>60</sup> *Struktura rewolucji naukowych*. W: Wapedia [on-line] [dostęp 19 maja 2009]. Dostępny w World Wide Web: <http://wapedia.mobi/pl/Paradygmat>.

<sup>61</sup> Ch. Chen: *Information Visualization...*, s. 5; I. Samoylenko, T.-C. Chao, W.-C. Liu, C. M. Chen. *Visualizing the scientific world and its evolution*. Journal of the American Society for Information Science and Technology 2006, Vol. 57 (11), s. 1461-1469.

<sup>62</sup> Ch. Chen: *Information Visualization...*, s. 157-171.

zblizone do długości Plancka (około  $10^{-35}\text{m}$ ), które wibrują z pewnymi ściśle określonymi częstotliwościami. Struny te cechuje supersymetria. Do analizy wybrano dane bibliograficzne, zawierające w temacie pracy wyrazy „string (superstring) theory”. Dalej zbadano częstotliwości występowania terminów, charakteryzujących daną teorię w różnych okresach jej rozwoju, np. „black holes”, „branes” oraz „supgravity”. Wyniki wykazały dwie krytyczne zmiany na skali czasu. Nawiązywały one do dwóch rewolucji paradygmatycznych w teorii superstrun: pierwsza w latach 1984-1985, kiedy rozwiązywano anomalie w supersymetrii wielowymiarowej oraz druga - zaczynając od 1994, kiedy podstawowe problemy zaczęły dotyczyć dualizmu i teorii strun. Warto zaznaczyć, iż doświadczalna weryfikacja tych podejść będzie prawdopodobnie możliwa już w 2009 r. dzięki nowym pomiarom satelity Planck wprowadzonej na orbitę w 2008 r.

Jak wspomniano wyżej podstawowe metody analizy bibliometrycznej skupiają się na wzajemnych i/lub wspólnych cytowaniach prac naukowych. W tym schemacie podobieństwo pomiędzy artykułami *i* oraz *j* wyznaczone jest za pomocą liczby dokumentów, powołujących się na obydwie źródła. Obiekty badane są sprowadzane do postaci macierzy podobieństwa (ang. *proximity/similarity matrix*)<sup>63</sup>, które następnie są konwertowane do map przestrzennych. Aby zmniejszyć wymiary danych stosuje się dwie popularne techniki statystyczne: analizę czynnikową oraz skalowanie wielowymiarowe, wykorzystane w danej pracy. Oprócz statystycznych coraz częściej wykorzystuje się algorytmy oparte o sztuczne sieci neuronowe. Metody te prowadzą do poprawienia widoczności ukrytych struktur, zawartych w macierzy danych.

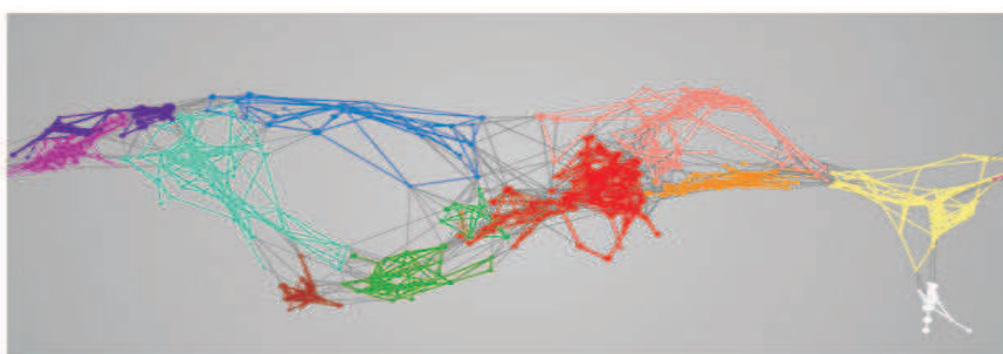
Dane wzajemnie powiązanych cytowań z czasopism rozmaitych dziedzin zgromadzano przy wykorzystaniu bazy danych *Science Citation Index* w okresie 1994-2001<sup>64</sup>. Autorzy na mapie zinterpretowali współczesną ewolucję świata naukowego, charakteryzującego się liniową strukturą i trzema głównymi kierunkami rozwoju: nauki matematyczno-fizyczne, nauki o życiu i nauki medyczne. Analiza cytowań pomiędzy czasopismami potwierdza semantyczne odseparowanie nauk matematyczno-fizycznych od innych domen, a badania w dwóch pozostałych domenach są powiązane, co wskazuje na wzajemne lansowanie odkryć w obu naukach. Większość map nauki generowana była raczej dla ograniczonych pól badawczych z niewielkich zbiorów danych, liczących setki lub tysiące węzłów w grafie. Takie prekursorskie wizualizacje ujawniały nie tylko relacje pomiędzy dokumentami i artykułami, lecz wspomagały detekcję najważniejszych badaczy w ramach konkretnej dziedziny, a także analizę struktury dyscypliny nauki i jej ewolucję. W późniejszym okresie gwałtownego wzrostu zasobów piśmiennictwa naukowego bazy danych liczyły kilka milionów rekordów – artykułów. Aby sprostać wymogom skalowalności, metodologia rozszerzyła się do algorytmów klastrowania, sieci neuronowych i map samoorganizujących się oraz technik kombinowanych.

<sup>63</sup> W bibliometrii macierze prawdopodobieństwa przedstawiają relacje między publikacjami na podstawie liczby ich cytowań. Model macierzowy opiera się na liczbowej reprezentacji dokumentów, co pozwala zmierzyć podobieństwo między nimi.

<sup>64</sup> Samoylenko I. i in., dz. cyt.

W ostatniej dekadzie mapowanie nauki wykonuje się na dużą skalę. Jednym z interesujących przykładów, gdzie wykorzystano indeks naukowych cytowań i reprezentującą kompletną strukturę aktualnej nauki, włączając nauki przyrodnicze i społeczne jest praca *Mapping the backbone of Science*<sup>65</sup>. Kolekcja danych liczyła ponad milion dokumentów i około siedmiu tysięcy czasopism. Do wygenerowania grafu użyto ośmiu miar podobieństwa pomiędzy czasopismami, różniące się dokładnością, skalowalnością algorytmu oraz czytelnością układu czyli stopniem klasteryzacji. Dzięki temu, że badane czasopisma są przypisane do konkretnych dyscyplin, można zrozumieć organizację aktualnej nauki na poziomie agregacyjnym.

Najnowsze mapy wynikające z badań *KDViz* regularnie są publikowane na portalu *Places@Spaces* w sekcji *Domain Maps*<sup>66</sup>. R. Klavans i K. Boyack użyli metryki opartej o bibliografię łączone oraz wektory słów kluczowych ponad siedmiu milionów publikacji oraz szesnastu tysięcy czasopism i materiałów konferencyjnych za okres 2001-2005<sup>67</sup> (p. Rysunek 16). Za przestrzeń docelową wybrano powierzchnię sfery, po czym na potrzeby prezentacji dane odwzorowano metodą walcową równokątną<sup>68</sup>. Zbieżność z metodologią niniejszej rozprawy doktorskiej jest przypadkowa, natomiast jest dowodem słuszności przedstawionej strategii. Na podstawie zmian powiązań pomiędzy poszczególnymi obszarami danych R. Klavans i K. Boyack podjęli się prognozowania struktury nauki w najbliższej przyszłości.



matematyka, fizyka	biotechnologia	nauki o mózgu
chemia	nauki o Ziemi	nauki o zdrowiu
nauki komputerowe	biologia	nauki społeczne
inżynieria	nauki medyczne	nauki humanistyczne

**Rysunek 16.** Dwuwymiarowa mapa nauki zrobiona na podstawie mapowania bibliografii łączonych i słów kluczowych ponad 7 milionów publikacji za okres 2001-2005

**Źródło:** R. Klavans, K. Boyack. *Maps of Science: Forecasting Large Trends in Science* [on-line]. *Places@Spaces: Mapping Science* [dostęp 19 maja 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).

<sup>65</sup> K. W. Boyack, i in. dz. cyt..

<sup>66</sup> *Browse map. Overview* [on-line]. *Places@Spaces: Mapping Science* [dostęp 19 maja 2009]. Dostępny w World Wide Web: <http://www.scimaps.org/browse/>.

<sup>67</sup> R. Klavans, K. Boyack: *Maps of Science: Forecasting Large Trends in Science* [on-line]. *Places@Spaces: Mapping Science* [dostęp 19 maja 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).

<sup>68</sup> *Odwzorowanie walcowe równokątne* – jeden z rzutów kartograficznych, zwany też *odwzorowaniem Mercatora*. Południkom i równoleżnikom odpowiadają odcinki, kąty między nimi są zachowane, przy czym najmniejsze zniekształcenia powstają blisko równika, największe – na biegunach... Por. *Odwzorowanie walcowe równokątne*. W: *Wikipedia. Wolna Encyklopedia* [on-line] [dostęp 19 maja 2009]. Dostępny w World Wide Web: [http://pl.wikipedia.org/wiki/Odwzorowanie\\_walcowe\\_r%C3%B3wnok%C4%85tne](http://pl.wikipedia.org/wiki/Odwzorowanie_walcowe_r%C3%B3wnok%C4%85tne).



Widzimy, że wizualizacja domen wiedzy sprowadza się głównie do mapowania literatury naukowej w wybranej przestrzeni semantycznej bądź tematycznej. W procesie tworzenia reprezentacji domen wiedzy najczęściej wyróżniamy następujące etapy:

- wybór odpowiedniego do postawionego problemu źródła danych;
- wybór jednostek analizy oraz ekstrakcja niezbędnych danych;
- wybór właściwej miary podobieństwa (w literaturze są opisywane osiem podstawowych miar<sup>69,70</sup>;
- tworzenie graficznego układu danych przy użyciu wybranych algorytmów;
- eksploracja wygenerowanej mapy w odpowiedzi na pierwotne zapytania.

Informację semantyczno-strukturalną wydobywa się także przy pomocy wspólnie występujących sekwencji słów w dwóch lub więcej artykułach (**co-words**), słów kluczowych (**co-keywords**) lub deskryptorów (**co-descriptors**). Na alternatywne jednostki analizy wybiera się między innymi: kategorie czasopism<sup>71</sup>, autorów, słowa kluczowe, źródła itp. i w zależności od potrzeb użytkowników generowane są mapy o różnym stopniu zawartości tematycznych treści naukowych. W mapowaniu dyscyplin naukowych zaadoptowane zostały współczesne technologie grafiki komputerowej. Na wspomnianym portalu *Places@Spaces* swoje artystyczne wizje naukowego świata zamieszczają również graficy. Rysunek 17 ilustruje hipotetyczny model ewolucji nauki, wykonany przez D. Zellera<sup>72</sup>, uniwersum nauki zostało przedstawione w kształcie meteoru, a obszary badawcze – jako tunele ewoluujące od środka.

Zainteresowanie tą tematyką może przebiegać równolegle z chęcią zrozumienia jak funkcjonują firmy naukowo-technologiczne i jak nimi zarządzać. Inicjatywa mapowania współczesnej wiedzy podejmowana jest również w sektorze komercyjnym przy wykorzystaniu danych bibliometryczno-biznesowych. Celem takich projektów jest też upublicznianie wiedzy naukowej i technologii informacyjno-komunikacyjnych. Aktualne badania KDViz uwzględniają zapotrzebowanie na uproszczone mapy domen o wystarczającej zawartości informacyjnej dla wszystkich niespecjalistów danej dziedziny. Obok projektów wizualizacji dyscyplin naukowych, w niekonwencjonalny sposób przedstawiane są również aktualne trendy w ruchu i aktywności użytkowników w sieci, na przykład: Wikipedia, blogi, technologie WWW. To tylko udowadnia obserwowany rozwój wszelkich form interakcji między użytkownikami w sieci Web 2.0. Pierwsza semantyczna mapa angielskiej Wikipedii (Rysunek 18) wygenerowana została przy użyciu metryki, rejestrującej wspólne kategorie artykułów<sup>73</sup>. Widoczna tu jest logiczna kompozycja klastrów, a mapa udowodniła sprawne zarządzanie strukturą kategorii tematycznych, pomimo dużego zróżnicowania autorów: zarówno botów tak i ludzi.

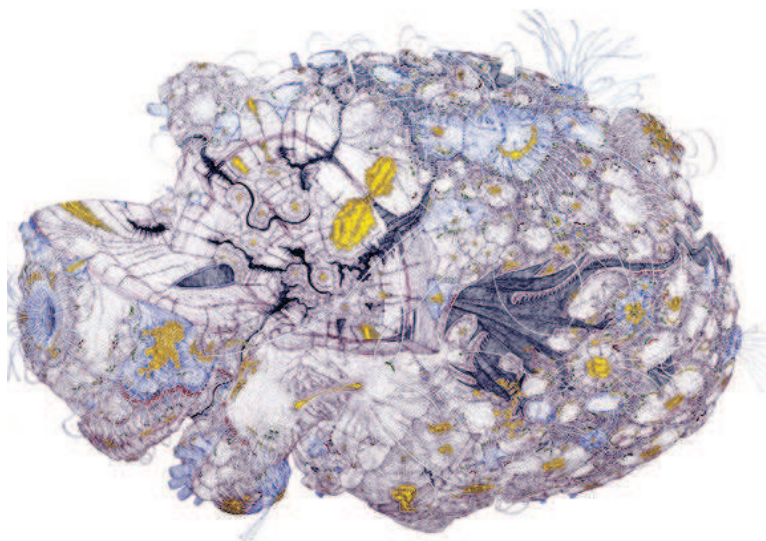
<sup>69</sup> K. W. Boyack, dz. cyt.

<sup>70</sup> R. Klavans, K.W. Boyack: *Identifying a better measure of relatedness for mapping science*. Journal of the American Society for the Information Science and Technology 2005. Vol. 57, nr 2, s. 251-263.

<sup>71</sup> F. Moya-Anegón i in., dz. cyt..

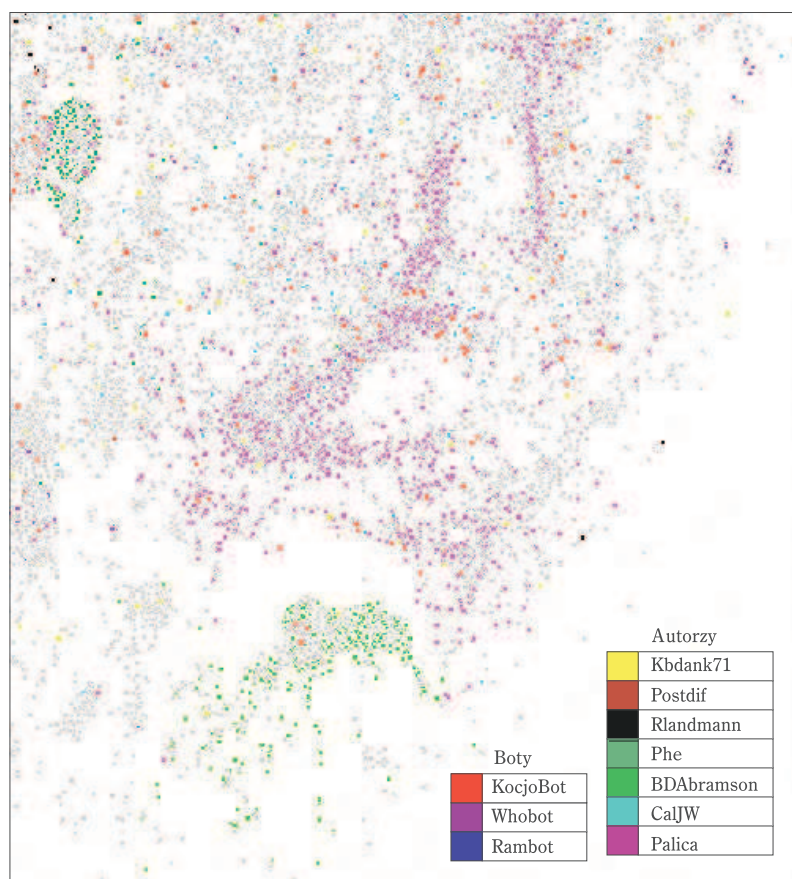
<sup>72</sup> D. Zeller: *Hypothetical Model of the Evolution of Science* [on-line]. *Places@Spaces: Mapping Science* [dostęp 19 maja.2009]. Dostępny w World Wide Web: [http://www.scimaps.org/dev/map\\_detail.php?map\\_id=163](http://www.scimaps.org/dev/map_detail.php?map_id=163).

<sup>73</sup> T. Holloway, M. Bozicevic, K. Börner: *Analyzing and Visualizing the Semantic Coverage of Wikipedia and Its Authors*. Wyd. specjalne: Understanding Complex Systems. Complexity 2007. Vol. 12, nr 3, s. 30-40.



**Rysunek 17.** Hipotetyczny model ewolucji nauki oczyma grafika. Uniwersum nauki zostało przedstawione w kształcie meteoru. Każdego roku pojawia się nowa warstwa; niebieski kolor wskazuje na nowe dyscypliny, brązowe – na ustabilizowane; każdy „trąbkopodobny” obszar badawczy ewoluuje od środka.

**Źródło:** D. Zeller. *Hypothetical Model of the Evolution of Science* [on-line]. *Places@Spaces: Mapping Science* [dostęp 19 maja 2009]. Dostępny w World Wide Web: [http://www.scimaps.org/dev/map\\_detail.php?map\\_id=163](http://www.scimaps.org/dev/map_detail.php?map_id=163).



**Rysunek 18.** Wizualizacja rodzaju autorstwa artykułów angielskiej Wikipedii

**Źródło:** : T. Holloway, M. Bozicevic, K. Börner: *Analyzing and Visualizing the Semantic Coverage of Wikipedia and Its Authors*. Wyd. specjalne: *Understanding Complex Systems*. Complexity 2007, Vol. 12, nr 3, s. 38.







**Rysunek 20.** Fragment mapy nauki, wygenerowanej na podstawie danych o logach użytkowników w serwisach specjalistycznych (kolorowe kółka reprezentują czasopisma, poklasyfikowane według klasyfikacji Art & Architecture Thesaurus; podpisy – ich grupy tematyczne).

**Źródło:** J. Bollen i in. *Clickstream Data Yields High-Resolution Maps of Science*. PLoS ONE [on-line] 2009, no. 4(3) [dostęp 19 maja 2009] Dostępny w World Wide Web: <http://www.plosone.org/article/info:doi/10.1371/journal.pone.0004803>.

Rozwiązania dynamicznych naukogramów pozwalają na obserwowanie ewolucji badań w ramach konkretnej dyscypliny. W wyniku eksploracji możliwa jest analiza trendów w rozwoju przyszłych kierunków badań w dowolnym regionie świata oraz co jest istotne – ich wzajemne porównanie. Wzory powstające na globalnych mapach dyscyplin i specjalności badawczych potwierdzają wzrastającą interdyscyplinarność nauki. Przypomnieć tu należy, iż według Kuhna, twórcy koncepcji paradygmatycznej nauki, typowi naukowcy nie są „obiektywnymi i niezależnymi myślicielami, lecz konserwatystami”, gdyż stosują wiedzę zgodnie z dyktatem wyuczonej przez nich teorii. W związku z tym można takie globalne mapy wykorzystać do weryfikacji stopnia twórczości i innowacyjności w podejmowaniu decyzji danej grupy autorów. W poszukiwaniu nowych rozwiązań problemów badawczych analiza piśmiennictwa naukowego nie powinna ograniczać się do tradycyjnie określonego zakresu danej dziedziny, ani do przetestowanych i uznanych w naukometrii i webometrii metodologii. Warto tu, jako przykład innowacyjnego podejścia, zacytować pracę z roku 2009<sup>78</sup>, gdzie do mapowania informacji pochodzących z sieciowych portali wykorzystano statystykę logów użytkowników – Rysunek 20. Studiowanie źródeł z tematycznie odległych nauk może przyczynić się zarówno do odnalezienia zaskakujących odpowiedzi, jak i do poszerzenia własnych horyzontów.

Widzimy zatem, iż mapowanie nauki nie jest nową procedurą bibliometryczną: powstała ona w latach siedemdziesiątych XX w. i intensywnie się rozwijała. Przyczynili się do tego głównie pomysłodawcy sprawdzonych i dzisiaj uważanych za podsta-

<sup>78</sup> J. Bollen i in. *Clickstream Data Yields High-Resolution Maps of Science*. PLoS ONE [on-line] 2009, Vol. 4, no. 3 [dostęp 19 maja 2009] Dostępny w World Wide Web: <http://www.plosone.org/article/info:doi/10.1371/journal.pone.0004803>.



wowe, metodologii w tej dziedzinie, jak np.: E. Garfield, H. Small, B. Griffith, I. Marshakova, B. Hjørland oraz H. White. Lecz do dalszych postępów w mapowaniu nauki potrzebne było podejście zorientowane na użytkownika. Postępy w technologiach informatycznych i badaniach kognitywnych wyznaczyły nowoczesne reguły projektowania interfejsów programów komputerowych. Wizualizacja (a także estetyka prezentacji) informacji stała się obecnie najistotniejszym ogniwem, decydującym o właściwym postrzeganiu i rozumieniu danych dla wszystkich poziomów końcowego użytkownika. Wizualizacja domen wiedzy jako nowe/odnowione podejście w mapowaniu nauki powstało na progu 3-go tysiąclecia. Znajdujący się w czołówce badaczy *KDViz*: Ch. Chen i K. Börner, podkreślają wieloperspektywiczność zastosowań metod wizualizacji: w wyszukiwaniu informacji, monitorowaniu trendów w nauce i technologiach, polityce finansowania badań i konkretnych naukowców.

Przyszłość nauki rysuje się na podstawie ścisłej współpracy ekspertów dziedzinowych w kwestiach teoretycznego i praktycznego wykorzystania narzędzi wizualizacyjnych w badaniach interdyscyplinarnych. Naukowcy mogą „spojrzeć z lotu ptaka na panoramę nauki”, aby jej użyć do eksploracji obszarów zainteresowań, do komunikowania wyników i do anonsowania odkryć naukowych. Początkujący badacze i studenci znajdą tu punkt startowy do weryfikacji własnego mentalnego obrazu świata naukowego. Podjęta tematyka nabiera szczególnej wagi w momencie dyskusji, na ile proponowana metoda może przydać się w wizualizacji nauk komputerowych i im pokrewnych w rozdziale podsumowującym.

