

# Organizacja wiedzy w cyfrowym świecie: wpływ rewolucji cyfrowej na ewolucję metod i narzędzi organizacji wiedzy<sup>1</sup>

Barbara Sosińska-Kalata

ORCID 0000-0002-4511-4701

*Katedra Informatologii, Wydział Dziennikarstwa, Informacji i Bibliologii,  
Uniwersytet Warszawski*

## Abstrakt

**Cel/Teza:** Celem artykułu jest przeanalizowanie zmian zachodzących w organizacji wiedzy pod wpływem rewolucji cyfrowej oraz identyfikacja kluczowych czynników, które wyznaczają nowe koncepcje metod i narzędzi organizacji dostępu do zapisanej wiedzy w środowisku cyfrowym.

**Koncepcja/Metody badawcze:** Artykuł ma charakter ogólnej teoretycznej refleksji nad ewolucją metod i narzędzi organizacji wiedzy. Na podstawie piśmiennictwa przedmiotu przeanalizowano proces kształtowania się koncepcji organizowania dostępu do zapisanej wiedzy w środowisku cyfrowym oraz przedyskutowano jego uwarunkowania technologiczne i społeczne. Zwrócono uwagę na konsekwencje rozwoju zjawiska big data (gigadanych, megadanych) dla organizacji wiedzy oraz nowe kierunki zastosowań systemów organizacji wiedzy (SOW).

**Wyniki i wnioski:** Wśród czynników technologicznych, które najsilniej oddziałują na kształtowanie nowych koncepcji organizowania wiedzy w świecie cyfrowym oraz kierunków modyfikowania tradycyjnych i tworzenia nowych modeli SOW wskazano: rozwój technologii przetwarzania języka naturalnego (NLP), upowszechnienie hipertekstu, rozwój technologii semantycznych, metod maszynowego uczenia i technologii Web 2.0, wykorzystanie technik wizualizacji informacji, dostosowanie funkcjonalności SOW do zastosowań na urządzeniach mobilnych. Wśród uwarunkowań społecznych jako czynniki odgrywające najważniejszą rolę w wyznaczaniu kierunków ewolucji metod i narzędzi organizacji wiedzy wskazano: nastawienie na obsługę użytkownika nieprofesjonalnego, dostosowanie metod budowy, struktury i funkcjonalności SOW do rosnącej różnorodności reprezentowanych za ich pomocą obiektów informacyjnych oraz pytań użytkowników kierowanych do serwisów informacyjnych.

**Oryginalność/Wartość poznawcza:** Zgodnie z wiedzą autorki, przedstawione rozważania są pierwszą w polskim piśmiennictwie kompleksową próbą prześledzenia wpływu rewolucji cyfrowej na rozwój metod i narzędzi organizacji wiedzy.

## Słowa kluczowe

Big data. Gigadane. Klasyfikacja fasetowa. Organizacja wiedzy. Rewolucja cyfrowa. System organizacji wiedzy. Środowisko cyfrowe. Tezaurus.

*Otrzymany: 8 czerwca 2019. Zrecenzowany: 22 czerwca 2019. Poprawiony: 30 czerwca 2019. Zaakceptowany: 11 lipca 2019.*

<sup>1</sup> W formie skróconej treść artykułu przedstawiona została na V Międzynarodowej Konferencji Naukowej „Nauka o informacji w okresie zmian. Rewolucja cyfrowa – dziś i jutro: infrastruktura, usługi, użytkownicy”, Warszawa, 13–14 maja 2019 r.

## 1. Wprowadzenie

Celem artykułu jest przeanalizowanie zmian zachodzących w organizacji wiedzy pod wpływem rewolucji cyfrowej oraz identyfikacja kluczowych czynników, które wyznaczają nowe koncepcje metod i narzędzi organizacji dostępu do zapisanej wiedzy w środowisku cyfrowym.

Organizacja dostępu do utrwalonej wiedzy polega na takim przekształcaniu środowiska, w którym jest ona zapisywana, gromadzona i udostępniana, aby zapewnić sprawne wyodrębnianie ze zgromadzonej kolekcji zapisów wiedzy tych jej elementów, które zaspokajają potrzeby osób poszukujących określonych treści w związku z zadaniami wykonywanymi przez te osoby. Organizacja wiedzy ma zatem za zadanie maksymalnie ułatwić dostęp do zapisów relewantnej informacji i wiedzy. Środowiskiem udostępniania utrwalonej wiedzy jest przestrzeń fizyczna lub wirtualna i należące do niej obiekty oraz wykonywane na nich działania. Jak każde środowisko, również środowisko udostępniania wiedzy jest określane przez konfigurację funkcjonujących w nim obiektów, ich wzajemne powiązania oraz cechy charakterystyczne dla otoczenia, w którym dokonują się zachodzące między nimi procesy. Obiektami środowiska udostępniania wiedzy są zapisy wiedzy (dokumenty, obiekty informacyjne), artefakty wykorzystywane do ich organizowania i przeszukiwania, które obecnie nazywa się systemami organizacji wiedzy (SOW) oraz ich elementy, tj. tworzone za ich pomocą reprezentacje zapisów wiedzy (charakterystyki wyszukiwawcze, metadane). Działaniami wykonywanymi na tych obiektach są procesy organizacji wiedzy (POW) oraz ściśle z nimi związane procesy wyszukiwania informacji.

Kluczowym elementem środowiska udostępniania wiedzy jest technologia informacyjna wykorzystywana do zapisywania, gromadzenia oraz dystrybucji wiedzy w społeczeństwie. Charakter tej technologii oraz jej rozwój determinują możliwości wyszukiwania informacji, a tym samym również możliwości realizacji POW i charakter stosowanych SOW. Konsekwencją rozwoju technologii informacyjnej jest powstawanie nowych metod organizacji wiedzy i nowych modeli SOW, dostosowanych do aktualnych możliwości wyszukiwania informacji. A zatem, technologia informacyjna i jej stałe doskonalenie mają decydujący wpływ nie tylko na stopniowe przekształcanie, ale także na całkowitą zmianę środowiska, w którym realizowane są procesy informacyjne związane ze społecznym transferem wiedzy. Taką zmianę, wywołaną postępującą rewolucją cyfrową, w dobitny sposób demonstruje w ostatnim czasie migracja niemal wszystkich rodzajów usług informacyjnych do środowiska cyfrowego. Wraz z migracją usług informacyjnych do środowiska cyfrowego migruje też ich instrumentarium, tj. narzędzia i metody realizacji tych usług, w tym metody realizacji POW i modele SOW dostosowywane do wymagań i możliwości tego środowiska.

Cyfryzacja usług informacyjnych i ich narzędzi jest procesem zachodzącym od kilkadziesiąt lat, a dzisiaj mamy do czynienia z jego już bardzo zaawansowanym, i dlatego spektakularnym, etapem. Proces ten rozpoczął się jednak ponad 60 lat temu, wraz z pierwszymi zastosowaniami cyfrowej technologii komputerowej w działalności informacyjnej, gdy powstały pierwsze komputerowe systemy informacyjne i pierwsze bazy danych. Ich tworzeniu od początku towarzyszyło opracowywanie i wdrażanie również nowych metod zapewniania dostępu do zawartych w nich zapisów wiedzy oraz nowych narzędzi wykorzystywanych w tych procesach. Rozważając zatem wpływ rewolucji cyfrowej na ewolucję metod i narzędzi organizacji wiedzy, trzeba przyrzeć się całemu przebiegowi procesu cyfryzacji usług informacyjnych, w których są one wykorzystywane oraz czynnikom, które

najsilniej oddziaływały na kształtowanie tych metod i narzędzi na kolejnych etapach ich ewolucji. W niniejszym artykule rozważania o przemianach metod i narzędzi organizacji wiedzy oraz oddziałujących na nie czynnikach, związanych z rozwojem rewolucji cyfrowej, oparte zostały na jakościowej analizie treści międzynarodowego i polskiego piśmiennictwa przedmiotu, które stanowi zarówno źródło omawianych obserwacji, jak i ich dokumentację.

## 2. Etapy ewolucji metod i narzędzi organizacji wiedzy w czasach rewolucji cyfrowej

Choć korzeni rewolucji cyfrowej doszukać się można jeszcze w XIX w., zwykle przyjmuje się, że została ona zapoczątkowana w latach 50. XX w., gdy stosowaną wcześniej technologię mechaniczną i analogową technologię elektroniczną zaczęła zastępować elektronika cyfrowa. Przejście to następowało jednak stopniowo i nierównomiernie w różnych częściach świata. O zaawansowanym stopniu cyfryzacji usług informacyjnych w krajach rozwiniętych można mówić dopiero od końca lat 70. XX w. Trzeba jednak zauważyć, że w tym pierwszym okresie powstała większość koncepcji technologicznych o kluczowym wpływie na dalszy rozwój narzędzi cyfrowych: od tranzystora, układów scalonych i podstawowych koncepcji architektury komputera, przez modele wyszukiwania informacji, metody wyszukiwania pełnotekstowego, po opracowanie protokołu przesyłania danych między sieciami komputerowymi. To wówczas krystalizował się sposób współczesnego myślenia o organizowaniu informacji i wiedzy w środowisku cyfrowym.

Wyraźne przyspieszenie procesu digitalizacji zasobów informacji i wiedzy rozpoczęło się dopiero w połowie lat 80. XX w., przyczyniając się do dynamizacji rozwoju nowych metod i narzędzi organizacji wiedzy dla usług informacyjnych prowadzonych w środowisku cyfrowym. Dzięki skonstruowaniu mikrokomputerów, a następnie szerokiemu udostępnieniu technologii sieci teleinformatycznych, w szczególności Internetu, w okresie tym nastąpiło upowszechnienie zastosowań technologii komputerowej, z czym m.in. wiąże się dostosowanie usług informacyjnych w środowisku cyfrowym do potrzeb i umiejętności szerokiej publiczności. Wreszcie od końca pierwszej dekady XXI w. można mówić o kolejnej fazie rozwoju i zastosowań technologii cyfrowej, związanej z wielkimi zbiorami danych (big data, gigadanych, megadanych) i nowym podejściem do ich przetwarzania, które otworzyło nowe możliwości wykorzystywania cyfrowo zapisanej informacji i wiedzy w coraz większej liczbie obszarów ludzkiej działalności.

Odwołując się do tych ogólnych obserwacji, na potrzeby niniejszego artykułu zastosowano uproszczoną periodyzację rewolucji cyfrowej, wskazując trzy jej okresy o cechach istotnych dla analizy zmian zachodzących w podejściu do projektowania metod i narzędzi organizacji wiedzy: wczesny okres rewolucji cyfrowej, przypadający na pierwsze trzy dekady drugiej połowy XX w., okres przyspieszenia cyfryzacji i związanego z nim upowszechnienia usług informacyjnych w środowisku cyfrowym, przypadający na kolejne ćwierćwiecze oraz okres współczesnej rewolucji cyfrowej, związany z dynamicznym rozwojem wielkich zbiorów danych i technologii do ich przetwarzania, który obserwujemy w ostatniej dekadzie.

Trzeba zaznaczyć, że, podobnie jak w przypadku periodyzacji rewolucji cyfrowej, również etapy ewolucji metod i narzędzi organizacji wiedzy, które stanowią konsekwencję rozwoju rewolucji cyfrowej wskazać można tylko w przybliżeniu, obserwując nasilanie się pewnych

zjawisk i pojawianie nowych rozwiązań implikujących istotne, jakościowe zmiany w podejściu do organizowania dostępu do utrwalonej wiedzy. Stosując taki punkt widzenia, można uznać, że pierwszy etap ewolucji metod i narzędzi organizacji wiedzy w dobie rewolucji cyfrowej przypada na jej wczesny okres, tj. okres od początku tej rewolucji i najwcześniejszych zastosowań technologii komputerowej do wyszukiwania informacji w latach 50. XX w. do wyraźnego przyspieszenia cyfryzacji zasobów i usług oraz upowszechniania dostępu do komputerów w latach 80. XX w. Trzeba jednak pamiętać, że przez te blisko trzy dekady równolegle stosowane i rozwijane były metody i narzędzia organizacji wiedzy ukierunkowane na wykorzystywanie w hybrydowym środowisku udostępniania wiedzy, na które składało się tradycyjne środowisko analogowe oraz stopniowo wypierające je środowiska technologii mechanicznych, elektroniki analogowej i elektroniki cyfrowej. Istotne jest również to, że technologie mechaniczne i analogowe stanowiły swoiste preludium dla rozwoju technologii cyfrowej, opierając się na analogicznej koncepcji kodowania informacji za pomocą prostych znaków i algorytmicznego (dokładnego lub heurystycznego) przetwarzania informacji.

Drugi etap ewolucji metod i narzędzi organizacji wiedzy w epoce rewolucji cyfrowej otwiera wspomniane przyspieszenie cyfryzacji zasobów i usług informacyjnych, które w krajach rozwiniętych wyraźnie uwidoczniło się w drugiej połowie lat 80. XX w., i które od początku lat 90. XX w. zostało spotęgowane przez upowszechnienie dostępu do Internetu i rozwój usług sieciowych. Z tym okresem wiążą się przede wszystkim zmiany ilościowe dotyczące cyfryzacji zapisanej informacji i wiedzy, usług zapewniających globalny i coraz powszechniejszy dostęp do tych cyfrowych kolekcji oraz szybkie poszerzanie się i coraz większe różnicowanie środowiska użytkowników tych zasobów i usług.

Trzeci etap ewolucji metod i narzędzi organizacji wiedzy wyznaczają przemiany dokonujące się we współczesnym okresie rewolucji cyfrowej, związane z powstaniem wielkich zbiorów danych oraz nowych technik ich przetwarzania, w dużym stopniu opartych na zaawansowanych metodach sztucznej inteligencji. Dziś trudno jest jeszcze jednoznacznie ocenić konsekwencje tych przemian dla organizacji wiedzy, jednak obserwując wpływ technologii big data na wiele dziedzin życia i badań naukowych, można spodziewać się, że także dla nauki o informacji oraz dla organizacji wiedzy będą one znaczące (por. Sośńska-Kalata, 2018).

Rewolucja cyfrowa oznacza nie tylko zmiany technologiczne związane z wynalezieniem i upowszechnianiem zastosowań cyfrowych narzędzi zapisu, przetwarzania i transferu informacji, ale także, a może przede wszystkim, powodowane przez nie zmiany zachowania społeczeństwa. Kolejne etapy rewolucji cyfrowej wyznaczają zatem przemiany zachodzące zarówno w wymiarze technologicznym, jak i społecznym, a na następującą wraz z tymi zmianami ewolucję metod i narzędzi organizacji wiedzy również wpływały ściśle skorelowane ze sobą czynniki związane z obydwojema tymi wymiarami. W dalszej części artykułu, podejmując próbę prześledzenia ewolucji metod i narzędzi organizacji wiedzy, która dokonuje się pod wpływem rewolucji cyfrowej, rozważaniom zostały poddane kluczowe wydarzenia i osiągnięcia, związane z opracowaniem nowych rozwiązań lub wyznaczeniem nowych podejść do problemu organizowania dostępu do zapisanej wiedzy. Wydarzenia, osiągnięcia i koncepcje sygnalizujące istotne zmiany w podejściu do zagadnień organizacji wiedzy w rozważaniach tych osadzone są w kontekście ogólnej periodyzacji zmian technologicznych i społecznych sygnalizujących postępującą rewolucję cyfrową.

### 3. Dziedzictwo organizacji wiedzy u progu rewolucji cyfrowej

W okres rewolucji cyfrowej organizacja wiedzy weszła z koncepcjami zapewniania dostępu do wiedzy zapisanej w książkach i dokumentach, wypracowanymi w tradycji przede wszystkim metodyki bibliograficznej i dokumentacyjnej, bardzo już solidnej i stopniowo poddawanej standaryzacji warunkującej współpracę na szerszą skalę. Swoją metodyką rejestrowania zawartości gromadzonych kolekcji dysponowały też inne dziedziny zajmujące się przechowywaniem i udostępnianiem dziedzictwa kulturowego (archiwa, muzea), a także różne dziedziny działalności praktycznej (biznesowej, administracji publicznej, ochrony zdrowia), w których rosła świadomość użyteczności rejestrowania, gromadzenia i wykorzystywania rozmaitych informacji dla lepszej, skuteczniejszej, bardziej efektywnej realizacji ich zadań.

Tradycyjna metodyka organizowania dostępu do wiedzy była dostosowana do możliwości technologii opartej na papierze, inwentarzach i kartotekach oraz porządkowaniu dokumentów i ich opisów w przestrzeni fizycznej. Metody organizacji zapisów wiedzy w ich kolekcjach musiały uwzględniać ich fizyczność, która wymuszała np. ustalanie jednego, możliwie najlepszego miejsca poszczególnych zapisów w strukturze kolekcji. Najlepsze miejsce oznacza oczywiście takie, w którym najłatwiej (najszybciej, najtrafniej) można odnaleźć dany zapis, gdy jest on potrzebny. Na świat i wiedzę o świecie patrzemy jednak z wielu punktów widzenia, w różnych kontekstach, odpowiednio do różnych potrzeb i celów. Ze względu na te różne konteksty, takich najlepszych miejsc dla zapisów poszczególnych fragmentów wiedzy zwykle jest wiele. Metoda organizowania pośredniego dostępu do zapisów wiedzy poprzez tzw. informację pochodną, a więc ich opisy (metadane), zapewniła zwiększenie możliwości dostępu do zawartości kolekcji według różnych kryteriów, nadal jednak fizyczność zapisów metadanych wymuszała ograniczanie liczby tych kryteriów. Tymczasem ilościowy rozwój zapisów wiedzy w różnych formach oraz bogactwo i rosnąca szczegółowość zawartych w nich treści wymagały zapewniania coraz bardziej szczegółowych i wieloaspektowych struktur organizacyjnych, umożliwiających szybkie i trafne wyszukiwanie. Z kolei rozwój usług bibliotek, agend bibliograficznych i dokumentacyjnych coraz silniej uświadamiał potrzebę współpracy zarówno w skali krajowej, jak i międzynarodowej, która wymagała ujednoczenia metod i narzędzi pracy, w tym ówczesnych SOW. A zatem już w drugiej połowie XIX w. zaczęła kształtować się koncepcja wspomnianej wcześniej standaryzacji tych narzędzi, widoczna np. w upowszechnieniu w amerykańskim bibliotekarstwie Klasyfikacji Dziesiątej Deweya (KDD) i zasad opracowania przedmiotowego Charlesa A. Cuttera (*Rules for a Dictionary Catalogue*, 1876) oraz w koncepcji Uniwersalnej Klasyfikacji Dziesiątej (UKD), jako międzynarodowego SOW przeznaczonego dla usług bibliograficznych i dokumentacyjnych.

W dziedzinach, w których wielkość gromadzonych zbiorów danych przekroczyła możliwości racjonalnie szybkiego ich przetworzenia już w XIX w. opracowano pierwsze metody mechanizacji dostępu do zawartej w nich wiedzy. Często przywoływanym przykładem są maszyny tabulacyjne Hermana Holleritha (1860–1929) użyte w końcu lat 80. XIX w. do przetwarzania danych Naczelnego Lekarza Stanów Zjednoczonych o umieralności ludności w tym kraju, a następnie do przetworzenia danych Urzędu Statystycznego Stanów Zjednoczonych z powszechnego spisu ludności z 1890 r. (zob. np. Gawrysiak, 2008, 81–85). Maszyny Holleritha, wykorzystujące karty perforowane, szybko znalazły zastosowanie



w przetwarzaniu danych ze spisów ludności kolejnych państw (Austrii, Kanady, Włoch, Rosji). Wkrótce maszyny przeznaczone do tabelaryzacji danych zostały rozbudowane o funkcje pozwalające przeprowadzać operacje sumowania na zestawionych liczbach oraz ich podział na podzbiory odpowiadające różnym typom danych, dając początek maszynom kalkulacyjnym i księgowym. Produkująca te maszyny firma Tabulating Machine Company, założona w 1896 r. przez Holleritha, w 1924 r. przekształcona została w International Business Machines (IBM), a w okresie międzywojennym biurowe maszyny księgowe wykorzystujące karty perforowane stały się standardowym wyposażeniem działów księgowości dużych korporacji. Maszyny Holleritha były też inspiracją dla Paula Otleta (1868–1944), który w opublikowanym w 1934 r. *Traktacie o dokumentacji*, odwołując się do ich możliwości, naszkicował własną koncepcję mechanicznego urządzenia ułatwiającego dostęp do wiedzy zapisanej w dokumentach (Otle, 1934, 387–391).

Zarówno postępująca standaryzacja tradycyjnych SOW, jak i pomysły dotyczące mechanizacji procesów wyszukiwania informacji oraz zastosowanie w UKD modelu SOW umożliwiającego tworzenie wieloaspektowej reprezentacji dokumentów, stanowiły wyraźne i ważne działania w kierunku dostosowania metod i narzędzi organizacji wiedzy do nowej epoki, w której rosnąca liczba zapisów informacji i wiedzy wymagała nowego podejścia. Kolejnym ważnym krokiem w tym kierunku była koncepcja klasyfikacji fasetowej opublikowana w 1933 r. przez matematyka i bibliotekarza – Shiyali Ramamritę Ranganathana (1892–1972). Dopiero jednak pojawienie się technologii komputerowej i przeniesienie usług informacyjnych do środowiska cyfrowego umożliwiły pełne wykorzystanie tych koncepcji.

#### 4. Organizacja wiedzy na wczesnym etapie rewolucji cyfrowej

Początek rewolucji cyfrowej wiąże się zwykle z wynalezieniem tranzystora w 1947 r., a następnie uruchomieniem produkcji układów scalonych. Wydarzenia te przypadają na ten sam czas, który uznaje się za początek kształtowania się współczesnej nauki o informacji. Jednym z wydarzeń, które miały największy wpływ na formowanie koncepcji tej dyscypliny była zorganizowana w 1958 r. w Waszyngtonie Międzynarodowa Konferencja Informacji Naukowej (International Conference on Scientific Information, ICSI-1958), podczas której zdefiniowano jej agendę badawczą i przedstawiono wyniki aktualnych badań w najbardziej newralgicznych zakresach (zob. np. Saracevic, 2010; Sosińska-Kalata, 2017). Wśród siedmiu, uznanych za najważniejsze, obszarów problemowych, w których zorganizowano prezentację referatów aż trzy dotyczyły organizacji dostępu do utrwalonej informacji i wiedzy w nowych warunkach technologicznych: *Organization of information for storage and search. Comparative characteristics of existing systems* (Organizacja informacji w celu przechowywania i przeszukiwania. Komparatywna charakterystyka systemów istniejących); *Organization of information for storage and retrospective search. Intellectual problems and equipment considerations in the design of new systems* (Organizacja informacji w celu przechowywania i retrospektywnego przeszukiwania. Intelektualne i narzędziowe aspekty projektowania nowych systemów); *Organization of information for storage and retrospective search. Possibility for a general theory* (Organizacja informacji w celu przechowywania i retrospektywnego przeszukiwania. Możliwość ogólnej teorii) (NAS, 1959, vol. I, xix-xxiv; Saracevic, 2010, 2572).

Warto przypomnieć, że na ICSI-1958 swoje badania omawiali naukowcy, których nowatorskie pomysły miały ogromny wpływ na późniejszy rozwój nauki o informacji, a nazwiska tych badaczy na trwałe weszły do historii tej nauki. Byli to m.in.: Eugene Garfield, (*Unified Index to Science*), Mortimer Taube (*Conventional and Inverted Grouping of Codes for Chemical Data; The Comac: An Efficient Punched Card Collating System for the Storage and Retrieval of Information*), Cyril Cleverdon (*The Evaluation of Systems Used in Information Retrieval*), Brian C. Vickery (*Subject Analysis for Information Retrieval; The Structure of Information Retrieval*), Douglas J. Foskett (*The Construction of a Faceted Classification for a Special Subject*), Jean-Claude Gardin (*On the Coding of Geometrical Shapes and Other Representations, with Reference to Archeological Documents*), Calvin C. Mooers (*A Mathematical Theory of Language Symbols in Retrieval*). Peter H. Luhn był organizatorem sesji poświęconej intelektualnym i narzędziowym aspektom organizacji informacji dla celów przechowywania i wyszukiwania.

Przegląd tematyki referatów opublikowanych w dwóch tomach materiałów konferencyjnych ICSI-1958 jednoznacznie wskazuje, iż na początku rewolucji cyfrowej głównym problemem badawczym w zakresie organizacji wiedzy było opracowanie metod i narzędzi umożliwiających algorytmiczne wyszukiwanie dokumentów o określonej tematyce (NAS, 1959). Tak postawiony problem wymagał nowego sposobu tworzenia reprezentacji wyszukiwanych dokumentów, zapewniającego wyszukiwalność (ang. *findability*) tych dokumentów odpowiednio do potrzeb użytkowników za pomocą urządzeń mechanicznych, a następnie za pomocą algorytmów komputerowych. Wielki wpływ na sposób rozwiązywania tego problemu miały przede wszystkim teorie logiczne i modele matematyczne oraz równolegle rozwijające się badania nad automatycznym przetwarzaniem języka naturalnego. Na pierwszy plan wysunięto dopasowanie reprezentacji dokumentu (charakterystyki wyszukiwawczej, metadanych dokumentu) do kryteriów wyszukiwania i postrzeganie elementów SOW jako kluczy wyszukiwawczych identyfikujących zbiory potencjalnie relewantnych dokumentów. Zaproponowana wówczas przez Mooersa (1919–1994) metoda wyszukiwania oparta na algebrze Boole'a na długo stała się dominującym modelem wyszukiwania informacji. Opracowana przez Luhna (1896–1964) koncepcja indeksów KWIC, użyta do generowania tzw. permutacyjnych indeksów tytułowych, była jedną z pierwszych metod automatycznego przetwarzania tekstu naturalnego w celu wyszukiwania informacji. Koncepcja Garfielda (1925–2017) wykorzystania cytowań bibliograficznych jako podstawy wyszukiwania piśmiennictwa wskazała nie tylko nowatorski sposób organizowania dostępu do treści publikacji, ale otworzyła też drogę rozwojowi bibliometrii i naukometrii. Rozwinięta przez Mooersa i Taubego (1910–1965) metoda indeksowania współrzędnego za pomocą terminów derywowanych z tekstów dokumentów (słów kluczowych, unitermów, deskryptorów) wytyczyła główny kierunek rozwoju języków indeksowania – pierwszych SOW przeznaczonych do tworzenia reprezentacji dokumentów i wyszukiwania informacji w środowisku komputerowym. W słynnych eksperymentach Cranfieldzkich (1957–1967) badano wprawdzie efektywność wyszukiwania na podstawie reprezentacji dokumentów wyrażonych za pomocą różnych ówczesnych SOW, w tym haseł przedmiotowych, języków deskryptorowych, UKD i klasyfikacji fasetowych, jednak uzyskane wyniki wskazywały, że najlepsze rezultaty osiąga się stosując w indeksowaniu i wyszukiwaniu swobodne słowa kluczowe (Cleverdon, 1967). Mimo że później z wynikami tymi wielokrotnie dyskutowano, krytykując i kwestionując użyte w testach metody, rezultaty badań Cranfieldzkich miały

wielki wpływ na to, że środowisko specjalistów wyszukiwania informacji straciło zainteresowanie tradycyjnymi SOW i w ogóle kontrolą słownictwa wyszukiwawczego (ang. *search terms*, tj. wyrazów używanych do indeksowania i wyszukiwania dokumentów). Po krótkim i owocnym okresie początkowej bliskiej współpracy środowisk badaczy organizacji wiedzy i wyszukiwania informacji, od lat 70. XX w. nastął czas izolacjonizmu badań prowadzonych w tych obszarach, który trwa do dzisiaj (por. Hjørland, 2012).

Na wczesnym etapie rewolucji cyfrowej badania dotyczące rozwoju metod i narzędzi organizacji wiedzy obejmowały dwa główne nurty: związany z ukształtowaniem, rozwojem i upowszechnieniem koncepcji tezauryso-wyszukiwawczego oraz skupiony na rozwinięciu teorii klasyfikacji fasetowej i upowszechnieniu tego modelu. Rezultaty tych badań stanowią trwały wkład w oryginalny dorobek zarówno organizacji wiedzy, jak i w ogóle nauki o informacji, który znalazł szerokie wykorzystanie także poza macierzystą dyscypliną (zob. też Sosińska-Kalata, 2008; Broughton, 2015, 339–373).

Koncepcja tezauryso-wyszukiwawczego jako narzędzia kontroli słownictwa i prezentacji semantycznej struktury słownictwa naturalnego używanego w indeksowaniu i wyszukiwaniu pojawiła się już w końcu lat 50. XX w., a powszechnie przyjęty model organizacji słownictwa wyszukiwawczego za pomocą tezauryso opublikowany został w 1974 r. w postaci międzynarodowej normy ISO 2788 *Documentation – Guidelines for the Establishing and Development of Monolingual Thesauri* (Gilchrist, 2003; Aitchison, J., Dexter Clark, S., 2003; Sosińska-Kalata, 2006). Najogólniej powiedzieć można, że u podstaw tej koncepcji leżą cztery przesłanki: (1) indeksowanie współrzędne; (2) reprezentacja pojęć składających się na poszukiwane tematy za pomocą najlepiej znanych użytkownikom terminów derywowanych z tekstów publikacji; (3) kontrola słownictwa w celu zapewnienia jednolitości reprezentacji pojęć w indeksowaniu i wyszukiwaniu; oraz (4) identyfikacja terminów wyszukiwawczych reprezentujących relewantne pojęcia za pomocą sieci relacji semantycznych ukazujących ich kontekst (por. Sosińska-Kalata, 2006, 142–143).

Poszukiwanie ogólnej teorii organizacji zasobów informacyjnych dla potrzeb ich przechowywania i przeszukiwania wiąże się z rozwojem i upowszechnieniem koncepcji analizy i klasyfikacji fasetowej, jako opartej na przejrzystych zasadach logicznych metody wieloaspektowej strukturyzacji zbiorów pojęć reprezentowanych w indeksowaniu i wyszukiwaniu. Kluczowe znaczenie miało przyjęcie tej koncepcji jako podstawowego modelu nowoczesnych klasyfikacji tworzonych na potrzeby organizowania zasobów specjalistycznych baz danych bibliograficznych przez brytyjską Classification Research Group (CRG, 1952–1968). W 1957 r. na międzynarodowej konferencji w Dorking, zorganizowanej przez ASLIB (Association of Special Libraries and Information Bureaux) pod patronatem FID (Fédération Internationale de Documentation) i poświęconej klasyfikacji i wyszukiwaniu informacji, grupa badaczy tworzących CRG przedstawiła manifest, w którym klasyfikacja fasetowa zadeklarowana została jako podstawa organizacji zasobów wszelkich systemów wyszukiwania informacji (McIlwaine & Broughton, 2000; Satija, 2017). W drugiej połowie XX w. brytyjska CRG stała się najważniejszym zespołem badawczym, zajmującym się zarówno rozwojem teorii klasyfikacji fasetowej, jak i tworzeniem fasetowych systemów klasyfikacyjnych dla różnych dziedzin i branż. Celem badań CRG było wypracowanie nowego modelu ogólnej klasyfikacji, w którego koncepcji wykorzystana została teoria klasyfikacji fasetowej Ranganathana oraz teoria poziomów integracyjnych (ang. *theory of integrative levels*), która miała zastąpić szybko dezaktualizujący się tradycyjny podział



dyscyplinarny, stosowany dotąd jako podstawa organizacji wiedzy za pomocą systemów klasyfikacyjnych. Do końca działania CRG członkom tej grupy badawczej nie udało się zbudować nowej klasyfikacji ogólnej opartej na tym modelu, ale wypracowane przez nich założenia teoretyczne oraz ich implementacja w kilkunastu klasyfikacjach specjalistycznych, a także w semantycznej organizacji deskryptorów w systemie *Thesourofacet* i później w wielu innych tzw. tezaurusach fasetowych, miały wielki wpływ na ukierunkowanie współczesnych metod organizowania zasobów utrwalonej wiedzy dla potrzeb wyszukiwania informacji w środowisku cyfrowym.

W tym pierwszym okresie rewolucji cyfrowej, świadomość konieczności sformalizowania reprezentacji pojęć i relacji semantycznych wyrażanych w języku naturalnym dla potrzeb komputerowego przeszukiwania treści zawartych w dokumentach skłoniła niektórych badaczy do podejmowania również prób opracowania narzędzi zapewniających odpowiednio zaawansowany poziom takiej formalizacji. Warto przypomnieć koncepcję kodów semantycznych tworzonych w latach 50., 60. i 70. XX w. (*Semantic Code ASM-WRU* J.W. Perrego i A. Kenta, Język Rx-kodów E.F. Skorochodki, projekt języka semantycznego dla polskiego systemu SINTO opracowany przez Olgerda A. Wojtasiewicza i Bożennę Bojar). Kody semantyczne tworzone na wczesnym etapie rewolucji cyfrowej, które za pomocą symbolicznej notacji i reguł składni logicznej miały zapewnić czytelną maszynowo i wysoce szczegółową reprezentację pojęciowej zawartości tekstów, można uznać za wczesne formy współczesnych ontologii. Zastosowanie tych narzędzi nie wyszło jednak poza fazę projektów i eksperymentów (Perry & Kent, 1958; Skorochodko, 1971; Czerny, 1978; Wojtasiewicz & Bojar, 1977/1978).

## 5. Organizacja wiedzy w okresie przyspieszenia cyfryzacji usług i zasobów informacyjnych

Przyspieszenie przemian związanych z rewolucją cyfrową, które rozpoczęło się w końcu lat 70. XX w. wiąże się z jednej strony z coraz szerszym zastosowaniem wielkich komputerów w instytucjach różnych sfer życia społecznego (nauce, biznesie, administracji), a z drugiej – z zapoczątkowaną wynalazkiem mikroprocesora rewolucją mikrokomputerową, czyli upowszechnieniem indywidualnego korzystania z komputerów w pracy zawodowej, a następnie w życiu codziennym. Generalnie zjawiska te nie spowodowały w tamtym okresie gwałtownych zmian w dotychczasowym podejściu do metod i narzędzi organizacji wiedzy, jednak zwracały coraz mocniej uwagę na konieczność ich zrewidowania w świetle zmian zachodzących w środowisku użytkowników serwisów informacyjnych.

Wraz z rozwojem bibliograficznych i pełnotekstowych baz danych przede wszystkim rosła liczba specjalistycznych języków deskryptorowych i ich tezaurusów. Postępy w przetwarzaniu języka naturalnego wykorzystywano do doskonalenia technik automatycznego indeksowania i wyszukiwania pełnotekstowego. W bazach dziedzinowych zastosowanie znalazły też klasyfikacje specjalistyczne, raczej jednak w przeglądaniu niż w wyszukiwaniu informacji.

Szybko rosnąca liczba różnych SOW stosowanych lokalnie w różnych bazach utrudniała współpracę i dostęp do poszukiwanych informacji, zgromadzonych w różnych zbiorach. Konieczne stało się zatem wypracowanie nie tylko zasad standaryzacji budowy SOW, ale

także zapewnienie ich semantycznej kompatybilności i przekładalności, a więc pewnego typu ich interoperacyjności. W rezultacie podjęto poszukiwania dodatkowych metod i narzędzi organizacji wiedzy przeznaczonych do operowania na poziomie integracji dostępu do różnych zbiorów informacji o piśmiennictwie, zakodowanych w różnych SOW (np. fasetowa klasyfikacja Broad System of Ordering, powstała z inicjatywy UNESCO) i narzędzi umożliwiających przekłady między różnymi językami indeksowania i klasyfikowania. W 1985 r. została opublikowana norma ISO 5964 *Guidelines for the Establishment and Development of Multilingual Thesauri*, regulująca zasady organizowania wielojęzycznego dostępu do zasobów zaindeksowanych za pomocą kontrolowanego słownictwa deskryptorowego.

Zapoczątkowana jeszcze w latach 60. i 70. XX w. konwersja katalogów bibliotecznych z formatu analogowego na format elektroniczny w latach 80. na szeroką skalę wprowadziła do środowiska cyfrowego również tradycyjne SOW stosowane w bibliotekarstwie: języki haseł przedmiotowych i biblioteczne systemy klasyfikacyjne, przede wszystkim ogólne klasyfikacje takie jak KDD, UKD i Klasyfikacja Biblioteki Kongresu (KBK). Przeniesienie tych narzędzi do nowego środowiska technicznego początkowo nie wpłynęło jednak w istotny sposób na realizowany przez nie model organizowania dostępu do utrwalonej wiedzy. Rewizję tego modelu wymusiło dopiero pojawienie się na początku lat 90. XX w. kolejnego akceleratora przemian, tj. teleinformatycznej sieci Internet i systemu WWW, a wraz z nimi nowej generacji katalogów OPAC (Online Public Access Catalog), dobitnie ujawniając niedostosowanie tradycyjnych SOW do możliwości wyszukiwania w środowisku cyfrowym i możliwości samodzielnego korzystania z nich przez tzw. użytkowników nieprofesjonalnych.

Od końca XX w. główny nurt badań organizacji wiedzy był więc skupiony przede wszystkim na dostosowywaniu tradycyjnych SOW (głównie języków haseł przedmiotowych, systemów klasyfikacyjnych i języków deskryptorowych) do funkcjonowania w sieciowym środowisku cyfrowym. Słowniki tych SOW oraz instrukcje indeksowania lub klasyfikowania za ich pomocą przenoszono do formatu elektronicznego, najpierw w postaci słowników i tablic elektronicznych wydawanych na płytach CD-ROM i dostępnych tylko lokalnie, a następnie w postaci serwisów online aktualizowanych w trybie bieżącym i udostępnianych globalnie. W taki sposób tradycyjne SOW stosowane w praktyce bibliotecznej i bibliograficznej przekształcone zostały na modele bibliograficzne i systemy metadanych do opisu, wyszukiwania i zarządzania rekordami bibliograficznymi.

Samo przekształcenie tradycyjnych SOW w czytelne maszynowo systemy metadanych nie mogło jednak zmienić specyfiki tych narzędzi, tj. tego, że zostały one zbudowane jako specjalistyczne narzędzia pracy profesjonalistów informacji i były konstruowane przede wszystkim w celu maksymalizacji efektywności wyszukiwania informacji, zwykle kosztem wysokiego stopnia złożoności ich struktury i wysokiego poziomu trudności stosowania ich w praktyce. Cyfryzacja słowników i reguł indeksowania za pomocą tych narzędzi również zasadniczo służyła tylko profesjonalistom – katalogerom. Użytkownicy nieprofesjonalni, mimo iż te narzędzia niekiedy udostępniano publicznie, w praktyce rzadko po nie sięgali, zarówno do przeszukiwania bibliotecznych katalogów online, jak i do przeszukiwania bibliograficznych i pełnotekstowych baz danych. Taką sytuację obserwujemy także obecnie.

Publiczne udostępnienie Internetu nie tylko gwałtownie poszerzyło dostępność usług informacyjnych, ale przyspieszyło też rozwój metod i technik sztucznej inteligencji, czyli tzw. technologii inteligentnych. Na lata 90. XX w. przypada pojawienie się nowych metod

i narzędzi przetwarzania informacji umożliwiających dostęp do zapisanej wiedzy. Jako najważniejsze z nich wskazać należy: metody maszynowego uczenia się; metody eksploracji danych i tekstów; doskonalone metody i narzędzia przetwarzania języka naturalnego, które umożliwiły dalszy rozwój automatycznego indeksowania i klasyfikowania; zaawansowane SOW takie, jak leksykalne bazy danych, sieci semantyczne, mapy tematyczne i ontologie; narzędzia komputerowej wizualizacji informacji. Przegląd artykułów opublikowanych na łamach czasopisma *Knowledge Organization* w ostatnich dwóch dekadach wyraźnie pokazuje, że wszystkie te nowe metody i narzędzia spotkały się z wielkim zainteresowaniem badaczy organizacji wiedzy. Częściej jednak opisywali oni te innowacje, opracowane poza organizacją wiedzy, rozważając ich przydatność do celów organizacji wiedzy, niż integrowali je z własnymi metodami organizowania dostępu do utrwalonej wiedzy (Hjørland, 2012; Ibekwe-SanJuan & Bowker, 2017).

Z szybkim rozwojem usług internetowych w końcu lat 90. XX w. jest związane pojawienie się architektury informacji, która przez jednych jest traktowana jako subdyscyplina nauki o informacji, przez innych jako odrębna interdyscyplinarna specjalność zajmująca się projektowaniem różnych typów serwisów informacyjnych. W dziedzinie tej znalazły zastosowanie modele organizacji wiedzy wypracowane w okresie wcześniejszym, przede wszystkim tezaury jako narzędzie kontroli słownictwa wyszukiwawczego oraz metoda analizy fasetowej stosowana w taksonomiach fasetowych, narzędziach fasetowego przeglądania i przeszukiwania zasobów informacyjnych projektowanych serwisów. Choć jest to odrębne zagadnienie, które w niniejszym artykule nie będzie rozwijane, można zaryzykować tezę, że to właśnie architektura informacji w największym stopniu przyczyniła się do upowszechnienia tych metod i narzędzi organizacji wiedzy w szerokim, pozabibliotecznym środowisku współczesnych usług informacyjnych.

Na przeformułowanie koncepcji SOW i użytkowania narzędzi organizacji wiedzy w środowisku cyfrowym największy wpływ miał jednak rozwój bibliotek cyfrowych. Rzadko pamięta się o tym, że najszerzej obecnie wykorzystywana na świecie wyszukiwarka internetowa Google powstała w 1998 r. w ramach badań dotyczących właśnie technologii dla bibliotek cyfrowych, tj. w ramach Stanford Digital Library Technologies Project (SDLTP), projektu badawczego realizowanego na Uniwersytecie Stanforda w latach 1995–2004 (Stanford, n.d.). Jednak narzędzia takie jak Google, podobnie jak metody automatycznego indeksowania i wyszukiwania, czy też ontologie, trafiły do organizacji wiedzy z zewnątrz, jako wynik badań prowadzonych przez informatyków zajmujących się algorytmami wyszukiwania informacji. Wraz z rozwojem Internetu i systemu WWW w ramach organizacji wiedzy zaczęły się natomiast rozwijać koncepcje sieciowych systemów organizacji wiedzy (SSOW, Networked Knowledge Organization Systems, NKOS), których zadaniem jest wspieranie procesu automatycznego indeksowania i wyszukiwania informacji publikowanych w środowisku sieciowym przez porządkowanie i kontrolę systemu leksykalnego wykorzystywanego w tych procesach.

Prace nad SSOW zostały zainicjowane podczas warsztatów zorganizowanych przez ACM Digital Libraries w latach 1997 (*Thesauri and metadata*) i 1998 (*Application of terminology and classification tools for digital collection development and network-based search*). Generalnie koncepcja SSOW jest związana z przeniesieniem teorii i metodyki tradycyjnych systemów porządkowania pojęć i terminów do środowiska sieciowego i wykorzystaniem technologii sieciowych do optymalizacji procesów indeksowania i wyszukiwania informacji

za ich pomocą. W końcu lat 90. XX w. rozpoczęto też prace nad formatami danych do reprezentacji semantycznej struktury słownictwa SSOW, wśród których najbardziej znany jest format Simple Knowledge Organization System (SKOS) – opublikowany w 2008 r. i rekomendowany przez W3C (WWW Consortium) jako podstawowy standard opisu systemu leksykalnego SOW w środowisku sieciowym (zob. np. Roszkowski, 2009). SKOS jest jednym ze standardów Sieci Semantycznej (ang. *Semantic Web*), opartym na standardzie RDF (Resource Description Framework) i służy przede wszystkim do publikowania słowników SOW w środowisku sieciowym, w modelu danych powiązanych (ang. *linked data*).

SSOW są obecnie szeroko stosowane w usługach informacyjnych w WWW, zarówno w serwisach informacji o piśmiennictwie (bazach bibliograficznych dostępnych online, w bibliotekach cyfrowych i innych rodzajach pełnotekstowych systemów informacyjnych), jak i jako narzędzia online katalogerów w bibliotekach, a także w komercyjnych serwisach internetowych, gdzie wykorzystywane są w szczególności przez architektów informacji w organizowaniu informacji zawartej w dużych serwisach internetowych. SSOW obejmują różne rodzaje schematów metadanych służących kodowaniu semantycznej organizacji słownictwa wyszukiwawczego – zarówno takich, które są rezultatem przekształcenia tradycyjnych SOW (tezaurusów, kartotek haseł wzorcowych, systemów klasyfikacyjnych) i innych wykazów słownictwa (glosariusze, bazy leksykograficzne), jak i takich, które są narzędziami *digital-born*, powstałymi w celu organizacji dostępu do wiedzy utrwalonej w formacie cyfrowym i opublikowanej w środowisku sieciowym (pierścienie synonimów, taksonomie, sieci semantyczne, mapy tematyczne, ontologie).

Pojawienie się koncepcji Web 2.0 w połowie pierwszej dekady XXI w. dało kolejny impuls do rewizji koncepcji zapewniania dostępu do cyfrowych zapisów wiedzy w Internecie. Technologia Web 2.0 zmieniła model komunikacji w serwisach internetowych, umożliwiając użytkownikom partycypację zarówno w publikowaniu treści w serwisach internetowych, jak i w zarządzaniu tymi treściami. Wraz z rozwojem mediów społecznościowych zaczęła kształtować się społeczna aktywność m.in. w zakresie indeksowania (tagowania, katalogowania) zasobów cyfrowych. W ten sposób powstał nowy model SSOW: folksonomia. System ten składa się ze zbiorów wyrażen swobodnie wprowadzanych przez użytkowników jako metadane wyszukanych lub opublikowanych przez nich dokumentów lub obiektów cyfrowych. Brak kontroli tego słownictwa, częste występowanie w nim wyrażen o niskiej wartości informacyjnej lub błędnie zapisanych początkowo budziły sceptycyzm środowiska organizacji wiedzy. Dość szybko okazało się też, że społeczne indeksowanie nie jest zjawiskiem powszechnym i znajduje zastosowanie tylko w niektórych obszarach (por. Sosińska-Kalata, 2011, 98–107). Stopniowo jednak, zwłaszcza wykorzystywanie folksonomii przez użytkowników naukowych serwisów społecznościowych, a także przez użytkowników katalogów online niektórych bibliotek (najczęściej publicznych), zwróciło uwagę badaczy na wartości związane z treściami tworzonymi przez użytkowników (ang. *user generated content*). Należy do nich przede wszystkim aktualność terminologii stosowanej przez użytkowników, która kontrastuje z szybkim starzeniem się terminologii wykorzystywanej w leksyce SOW z kontrolowanym słownictwem. Sukcesy wielu projektów opartych na crowdsourcingu, w tym Wikipedii i licznych projektów m.in. z zakresu humanistyki cyfrowej, przemawiają za włączeniem folksonomii do poszukiwań nowych modeli SSOW dla rozwijającego się sieciowego środowiska cyfrowego. Systemy tagowania po raz kolejny wykazały też, że złożona składania tradycyjnych SOW stanowi nie tylko barierę

strukturalną, utrudniającą wieloaspektowe wyszukiwanie automatyczne, ale również barierę funkcjonalną, ograniczającą praktyczną użyteczność tych systemów dla użytkowników nieprofesjonalnych, którzy stanowią główną grupę docelową SOW stosowanych w usługach informacyjnych w środowisku cyfrowym.

Od dawna znany problem starzenia się struktur wiedzy odwzorowanych w relacyjnej strukturze słownictwa SOW badacze organizacji wiedzy próbują rozwiązać również w dobie coraz szybciej postępującej rewolucji cyfrowej, pokładając duże nadzieje w potencjale technologii cyfrowej. W 2007 r., na VIII Konferencji Hiszpańskiego Oddziału ISKO-ESP, która odbyła się w mieście León, sformułowano zestaw postulatów rozwoju nowych SSOW, które zachowają odporność na ten proces. Zalecenia te zostały opublikowane pod nazwą *Manifestu z León* (León Manifesto, 2007):

- (1) Nasilająca się interdyscyplinarność wiedzy wymaga zasadniczo nowych SOW, opartych na gruntownej rewizji zasad tworzenia tradycyjnych klasyfikacyjnych SOW ukierunkowanych na podział dyscyplinarny.
- (2) Innowacyjne podejście jest zarówno pożądane jak i możliwe dzięki nowej technologii sieciowej, a więc powinno być implementowane w projektowaniu nowych SOW.
- (3) Zamiast dyscyplin naukowych podstawowymi jednostkami organizacji nowych SOW powinny być klasy zjawisk (elementów) świata, tak jak są one reprezentowane w wiedzy człowieka o poznawanej rzeczywistości.
- (4) Nowe SOW powinny umożliwiać użytkownikom przechodzenie od jednej do innej perspektywy lub punktu widzenia, odzwierciedlając wielowymiarowy charakter ludzkiego myślenia; w szczególności powinny umożliwiać niezależne wyszukiwanie poszczególnych elementów i zjawisk świata, poszczególnych teorii dotyczących tych zjawisk i zachodzących między nimi związków oraz poszczególnych metod ich badania.
- (5) Związki między zjawiskami, między zjawiskami i teoriami dotyczącymi ich poznania oraz między zjawiskami i metodami ich badania powinny być wyrażane technikami analityczno-syntetycznymi wypracowanymi w klasyfikacjach fasetowych<sup>2</sup>.

Manifest z León zasadniczo jest kontynuacją i rozwinięciem koncepcji budowy ogólnego, uniwersalnego SOW, sformułowanej 50 lat wcześniej w Dorking. Zgodnie z tymi postulatami od 2004 r. międzynarodowa grupa badawcza pod kierunkiem Claudio Gnoli realizuje projekt Integrative Levels Classification – nowego klasyfikacyjnego SSOW opartego na teorii poziomów integracyjnych i zasadach klasyfikacji fasetowej (Gnoli & Poli, 2004; Gnoli et al., 2011).

## 6. Organizacja wiedzy na współczesnym etapie rewolucji cyfrowej

Jak wspomniano wcześniej, w ciągu ostatniego dziesięciolecia najważniejszym aspektem rewolucji cyfrowej stało się zjawisko big data (gigadanych, megadanych), czyli rozwój wielkich zbiorów danych i narzędzi do ich automatycznej analizy. We współczesnym piśmiennictwie, zarówno naukowym, jak i publicystycznym, politycznym, biznesowym czy ekonomicznym, popularność hasła „big data” jest już większa niż popularność do

<sup>2</sup> Przekład własny autorki.



niedawna dominujących haseł „Web 2.0” i „media społecznościowe”. Zjawisko big data ma coraz większe konsekwencje zarówno technologiczne, jak i metodologiczne, społeczne i epistemologiczne. Przekształca model badań w coraz większej liczbie nauk i model działania w coraz większej liczbie obszarów życia społecznego, a także indywidualnego.

Ostatnie postępy w technikach przetwarzania danych doprowadziły do powstania bardzo już sprawnych algorytmów uczenia maszynowego, które są w stanie wykorzystać ogromne ilości danych, pozostawianych przez codzienne korzystanie przez ludzi z urządzeń cyfrowych, do tworzenia modeli predykcyjnych. Lista aplikacji, w których obecnie wykorzystywane są algorytmy big data stale rośnie. Jako przykłady można wymienić: wykorzystywane przez wyszukiwarki systemy podpowiedzi i korekty terminów stosowanych do wyszukiwania; systemy rekomendacyjne, takie jak stosowane np. przez Amazona, Netflixa i wiele systemów e-commerce, które sugerują, co poszczególni użytkownicy/klienci mogą chcieć czytać, oglądać czy kupić w oparciu o ich poprzednie kliknięcia i zakupy; systemy kierowania reklam dostosowanych do profili odbiorców; narzędzia analiz giełdowych do wykrywania trendów; narzędzia analizy sentymentów, eksploracji opinii i wizualizacji informacji; systemy diagnostyki medycznej i technicznej. Coraz lepsza jakość automatycznego tłumaczenia za pomocą Google Translator jest dobrą ilustracją jakościowego skoku, który metody przetwarzania wielkich zbiorów zapewniły komputerowemu przetwarzaniu języka naturalnego. Niektórzy badacze uważają, że algorytmy big data, chociaż nie są niezawodne, osiągnęły poziom nie tylko akceptowalny dla ludzi, ale nawet konkurencyjny wobec usług ludzi-specjalistów (Ibekwe-SanJuan & Bowker, 2017, 188). Ponadto algorytmy te działają nieinwazyjnie, w sposób niewidoczny dla użytkownika, gromadząc dane i przetwarzając je tak, aby dostarczyć użytkownikowi wartościowych sugestii i zaleceń bez wymagania od niego jakichkolwiek dodatkowych działań.

Rozwój algorytmów big data i ich zastosowań każe zastanowić się nad znaczeniem SOW tworzonych przez ludzi, ich adekwatnością wobec potrzeb organizacji dostępu do utraconej wiedzy w globalnym środowisku cyfrowym oraz ich zdolnością do nadążania za stale rosnącymi rozmiarami dostępnych danych na określone tematy i z zakresu poszczególnych domen. W ostatnich latach kilku badaczy organizacji wiedzy podjęło próbę oceny użyteczności SOW w tym nowym środowisku udostępniania wiedzy i szans ich dalszego rozwoju.

Kwestię tę w 2012 r. poruszył m.in. Birger Hjørland w artykule o prowokacyjnym tytule *Is Classification Necessary after Google?* (Hjørland, 2012). Szeroko interpretując pojęcie klasyfikacji i obejmując nim wszystkie rodzaje SOW służących do wskazywania miejsca poszczególnych pojęć w konceptualnej strukturze wiedzy, na pytanie postawione w tytule artykułu Hjørland odpowiedział twierdząco, przekonując, że w epoce zalewu informacji i wielkich zasobów danych zapisanych cyfrowo klasyfikacja jest potrzebna bardziej niż kiedykolwiek wcześniej. Równocześnie uznał, że stosowane dotychczas SOW nie są w stanie sprostać potrzebom wyszukiwania informacji we współczesnym środowisku udostępniania wiedzy: globalnym, cyfrowym, sieciowym i wypełnionym przez technologie inteligentne. Wśród wskazanych przez Hjørlanda wymagań, które powinny spełnić SOW dostosowane do współczesnego środowiska udostępniania wiedzy warto wymienić przede wszystkim: odejście od silnej standaryzacji budowy SOW i zastąpienie arbitralnej projekcji struktury wiedzy podejściem elastycznym, uwzględniającym różne interpretacje i epistemologie; dostosowanie kryteriów klasyfikacji pojęć do kryteriów relewancji wyszukiwania informacji; dostosowanie zasad indeksowania do problemów wyszukiwawczych użytkowników;

zapewnienie wysokiej jakości indeksowania opartego na wiedzy dziedzinowej; wykorzystanie w tworzeniu i aktualizacji SOW możliwości technologii inteligentnych, maszynowego uczenia się, eksploracji danych i tekstu.

Z kolei Dagobert Soergel kilka lat później sformułował dość niejednoznaczną ocenę użyteczności SOW w erze cyfrowej (Soergel, 2015). Z jednej strony potwierdził wszechobecność wiedzy we wszystkich ludzkich działaniach, co powinno zapewnić niezbedność SOW w każdej dziedzinie i zastosowaniach wymagających intensywnego korzystania z wiedzy. Z drugiej strony przyznał, że wiele osiągnięć składających się na współczesny postęp w zakresie automatycznych technik pozyskiwania, reprezentacji i rozpowszechniania wiedzy jest rezultatem prac specjalistów innych dyscyplin. Ontologie, systemy z bazami wiedzy i systemy eksperckie, platformy wyszukiwania informacji i technologie Sieci Semantycznej zostały opracowane poza społecznością organizacji wiedzy. Soergel wezwał zatem do nawiązania współpracy między w dużej mierze oddzielnymi środowiskiem badaczy organizacji wiedzy i środowiskami badaczy ontologii, modelowania danych i technologii semantycznych, przekonując, że taka współpraca jest niezbedna do rozwiązania wielu problemów wymagających lepszych metod i narzędzi. Trzeba jednak zauważyć, że do takiej współpracy od dawna bezskutecznie wzywa wielu badaczy nauki o informacji (zob. Sosińska-Kalata, 2017, 44–45).

W 2016 r. w specjalnym numerze *Knowledge Organization* (KO) została przedstawiona debata na temat znaczenia tezaury dla wyszukiwania informacji online i innych aplikacji intensywnie wykorzystujących wiedzę. Warto przypomnieć, że niedawno gruntownej rewizji poddane zostały międzynarodowe standardy budowy tezaury (ISO 25964 – 1: 2011 *Information and documentation – – Thesauri and interoperability with other vocabularies. Part 1: Thesauri for information retrieval*; ISO 25964–2: 2013 *Information and documentation – – Thesauri and interoperability with other vocabularies. Part 2: Interoperability with other vocabularies*). Wśród sześciu artykułów opublikowanych we wspomnianym numerze specjalnym KO autor tylko jednego, Birger Hjørland, wyraził przekonanie, że użyteczność dotychczasowej koncepcji tezaury we współczesnym środowisku wyszukiwania informacji wyczerpała się, jakkolwiek zaznaczył, że nadal pozostaje miejsce dla „more flexible semantic tools based on proper studies of domains” (Hjørland, 2016, 13). Autorzy pozostałych artykułów dowodzili natomiast, że potrzeba stosowania w wyszukiwaniu takich narzędzi jak tezaury jest nadal aktualna, prezentując jednak argumenty dotyczące przede wszystkim braku dobrych narzędzi semantycznych wspierających wyszukiwanie informacji specjalistycznej, w szczególności w takich środowiskach jak korporacyjne intranety, systemy zarządzania dokumentami, biblioteki obrazów, katalogi muzeów, archiwa zapisów dźwiękowych, pełnotekstowe bazy agencji prasowych, wielojęzyczne systemy zarządzania zasobami cyfrowymi (White, 2016; Kempf & Neubert, 2016; Tudhope & Binding, 2016; MacFarlane, 2016; Garcia-Marco, 2016). W kilku istotnych kwestiach istnieje jednak wyraźna zgoda. Przede wszystkim podkreśla się zbędność tezaurusów ogólnych, których rolę skutecznie przejęła wyszukiwarka Google. W sytuacji, gdy podstawą wyszukiwania są metadane deriwowane z tekstów dokumentów, większość badaczy za zbędną uznaje też normatywną kontrolę słownictwa wyszukiwawczego. Uzasadnienie traci także przestrzeganie silnej standaryzacji budowy tezaurusów według dotychczasowych zasad: z jednej strony do wyszukiwania w wąskich, specjalistycznych domenach przydatna jest dokładniejsza niż przyjęta w normach specyfikacja relacji semantycznych, a z drugiej – w praktyce coraz rzadziej

przestrzega się zalecanych standardów. Wreszcie, zgodnie powtarzanym postulatem jest konieczność stosowania w tworzeniu i aktualizacji tezaurusów nowoczesnych technologii inteligentnych, w szczególności algorytmów statystycznych, technik NLP, eksploracji tekstów, automatycznego klasyfikowania, maszynowego uczenia się i wizualizacji informacji.

W moim przekonaniu, koncepcja tezauryusa zachowuje użyteczność we współczesnym środowisku cyfrowym przede wszystkim jako metoda wspierania wyszukiwania informacji dzięki dostarczaniu ustrukturyzowanych semantycznie zbiorów specjalistycznego słownictwa wyszukiwawczego i ta jego rola ma największe znaczenie. Wysiłki normalizacyjne środowiska organizacji wiedzy w ostatnich kilkunastu latach skupione były na rozwiązaniu problemu interoperacyjności tezaurusów i innych SOW, gdy tymczasem problemem obecnie najważniejszym jest raczej zapewnienie wyszukiwalności relewantnej informacji, zwłaszcza tam, gdzie niezbędne jest precyzyjne spełnianie kryteriów relewancji.

Warto też przytoczyć optymistyczną opinię Alego Shiri, który próbował wykazać możliwości wykorzystania SOW w analizie big data. Stwierdził on, że SSOW realizowane w modelu linked data mogą stanowić podstawę dla automatycznej klasyfikacji i wizualizacji semantycznej struktury dużych zbiorów danych (Shiri, 2014). Lista wskazanych przez Shiri możliwych zastosowań SSOW w analizie big data obejmuje: przydzielanie metadanych, automatyczną i pół-automatyczną analizę tekstu, kategoryzację treści, tworzenie widoków fasetowych, tworzenie metadanych, mapowanie (odwzorowywanie) wiedzy, wyznaczanie powiązań między wielkimi zbiorami danych, wyznaczanie powiązań między zbiorami słownictwa a tekstami nieustrukturyzowanymi, systemy rekomendacji, semantyczne wyszukiwarki, automatyczne uzupełnianie terminów, automatyczną eliminację wieloznaczności, interaktywną/automatyczną rozbudowę terminów, wizualizację powiązań i metadanych. Ten bogaty wykaz ma jednak na razie charakter tylko postulatywny, jakkolwiek coraz więcej jest badań, w których podejmuje się próby wykorzystywania SOW w tego rodzaju zastosowaniach (zob. np. Café & Rocha Souza, 2017; Claverly & Muir, 2018).

## 7. Wnioski

Cyfrowa transformacja środowiska udostępniania utrwalonej wiedzy w fundamentalny sposób wpływa na ewolucję organizacji wiedzy, zarówno w jej praktycznym, jak i teoretycznym wymiarze. Omówiona w artykule analiza zakresu i charakteru tego wpływu pozwala sformułować kilka ogólniejszych wniosków, które prezentuje poniższa lista:

- (1) Przede wszystkim stwierdzić należy, że specyfika cyfrowego środowiska, jego możliwości i ograniczenia technologiczne, od początku tworzenia specjalistycznych narzędzi organizacji dostępu do zapisów wiedzy za pomocą komputerów bezpośrednio decydowała i decyduje o kształtowaniu koncepcji tworzenia nowych SOW i o sposobach przekształcania tradycyjnych SOW.
- (2) Można też zauważyć, że na kształtowanie metod i narzędzi organizacji wiedzy wykorzystywanych w czasach postępującej rewolucji cyfrowej stopniowo rosnący wpływ miało dostosowanie tych narzędzi do kompetencji wyszukiwawczych nieprofesjonalnych użytkowników końcowych (ang. *end users*); obecnie antycypowane kompetencje i preferencje użytkowników końcowych odgrywają rolę decydującą i coraz większą wagę przywiązuje się do poznania kompetencji różnych środowisk

zawodowych, w których pracy nowoczesne SOW mają ułatwiać wyszukiwanie relewantnej informacji.

- (3) Do końca XX w., mimo cyfryzacji słowników i instrukcji indeksowania, tradycyjne SOW zachowały charakter narzędzi profesjonalnych, ukierunkowanych na maksymalizację efektywności wyszukiwania kosztem wysokiego stopnia złożoności ich struktury i wysokiego poziomu trudności stosowania w praktyce.
- (4) Dwie koncepcje organizacji wiedzy ukształtowane i rozwinięte na wczesnym etapie rewolucji cyfrowej stanowią najbardziej trwałe dorobek tej dyscypliny, który w sensie metodologicznym zachowuje użyteczność do dziś:
  - a. koncepcja tezaursu informacyjno-wyszukiwawczego jako metoda wspierania wyszukiwania informacji dzięki dostarczaniu ustrukturyzowanych semantycznie zbiorów specjalistycznego słownictwa wyszukiwawczego,
  - b. koncepcja analizy fasetowej i klasyfikacji fasetowej jako model wieloaspektowej reprezentacji i strukturyzacji systemów pojęć i terminów wyszukiwawczych.
- (5) W dotychczasowym przebiegu rewolucji cyfrowej Internet był najsilniejszym akceleratorem przemian koncepcji, metod i modeli organizacji wiedzy; przyspieszył zarówno rozwój technologii, które przekształcają sposób organizowania dostępu do utrwalonej wiedzy, jak i spowodował tzw. społeczny zwrot w podejściu do organizowania tego dostępu, wysuwając na pierwszy plan potrzeby, oczekiwania i możliwości użytkowników nieprofesjonalnych.
- (6) Wśród czynników technologicznych, które najsilniej oddziałują obecnie na kształtowanie nowych koncepcji organizowania wiedzy w świecie cyfrowym oraz kierunków modyfikowania tradycyjnych i tworzenia nowych modeli SOW, wskazać należy:
  - a. rozwój technologii NLP,
  - b. rozwój technologii semantycznych,
  - c. rozwój metod maszynowego uczenia,
  - d. rozwój technologii Web 2.0 i metod crowdsourcingu,
  - e. wykorzystanie technik wizualizacji informacji,
  - f. dostosowanie funkcjonalności SOW do zastosowań na urządzeniach mobilnych,
  - g. upowszechnienie cyfrowych zapisów informacji i wiedzy poprzez coraz szerszej zakrojoną digitalizację rezultatów badań naukowych i artefaktów szeroko rozumianej kultury oraz upowszechnienie cyfrowego utrwalania i publikacji rezultatów różnego rodzaju ludzkiej działalności,
  - h. rosnącą różnorodność zapisów informacji i wiedzy stanowiących przedmiot procesów organizacji wiedzy,
  - i. rozwój metod gromadzenia i analizy wielkich i różnorodnych zbiorów danych (big data).
- (7) Wśród uwarunkowań społecznych jako czynniki odgrywające najważniejszą rolę w wyznaczaniu kierunków ewolucji metod i narzędzi organizacji wiedzy za najważniejsze uznać należy:
  - a. nastawienie na obsługę użytkownika nieprofesjonalnego,
  - b. rosnącą różnorodność pytań użytkowników, które kierowane są do serwisów informacyjnych,
  - c. nastawienie na szczegółowe poznanie specyfiki potrzeb informacyjnych środowisk zawodowych, specjalistycznych i społecznych, które mają być użytkownikami

projektowanych SOW, ich epistemologii i kompetencji oraz dostosowanie do nich budowy i funkcjonalności narzędzi,

- d. wykorzystanie potencjału crowdsourcingu w tworzeniu, aktualizacji i doskonaleniu SOW.

Z dużą dozą pewności można założyć, że cyfryzacja zasobów utrwalonej informacji i wiedzy oraz usług informacyjnych w kolejnej dekadzie będzie nadal determinować kierunki rozwoju metod i narzędzi organizacji wiedzy. Będą one musiały sprostać dwóm nasilającym się tendencjom: wymaganiom użytecznego dostępu do coraz większych zbiorów oraz potrzebom, oczekiwaniom i umiejętnościom użytkowników, którzy z jednej strony wymagają narzędzi prostych i intuicyjnych w obsłudze, a z drugiej informacji i wiedzy o coraz bardziej zróżnicowanej postaci i coraz lepiej dostosowanych do zadań, w których mają być one wykorzystywane. Potrzeba dobrych SOW w takim kontekście wydaje się bezdyskusyjna i sądzić można, że zakres stosowania tego typu narzędzi będzie coraz szerszy, wykraczając poza tradycyjny obszar wspierania procesów wyszukiwania zapisów informacji i wiedzy oraz przeglądania i zarządzania ich kolekcjami. Jednak trzeba pamiętać, że zapewnienie funkcjonalności nowoczesnych SOW w rozwijającym się cyfrowym środowisku udostępniania wiedzy wymaga coraz głębszych zmian w podejściu do ich projektowania i konstruowania.

## Bibliografia

- Aitchison, J., Dexter Clark, S. (2003). The Thesaurus: A Historical Viewpoint with a Look to the Future. *Cataloging & Classification Quarterly*, 37 (3/4), 5–21.
- Brughton, V. (2015). *Essential Classification*. 2nd ed. London: Facet Publishing.
- Café, L.M., Rocha Souza, R. (2017). Sentiment Analysis and Knowledge Organization: An Overview of the International Literature. *Knowledge Organization*, 44 (3), 199–214.
- Cleverdon, C. (1967). The Cranfield Tests on Index Languages Devices. *Aslib Proceedings* 19 (6), 173–194.
- Claverly, P.H., Muir, L.J. (2018). Using Knowledge Organization Systems to Automatically Detect Forward-looking Sentiment in Company Reports to Infer Social Phenomena. *Knowledge Organization*, 45 (2), 152–169.
- Czerny, A.I. (1978). *Wstęp do teorii wyszukiwania informacji*. Warszawa: OIN PAN.
- Garcia-Marco, F.-J. (2016). Enhancing the Visibility and Relevance of Thesauri in the Web: Searching for a Hub in the Linked Data Environment. *Knowledge Organization*, 43 (3), 193–202.
- Gawrysiak, P. (2008). *Rewolucja cyfrowa: rozwój cywilizacji informacyjnej*. Warszawa: Wydaw. Nauk. PWN, Mikom.
- Gilchrist, A. (2003). Thesauri, Taxonomies and Ontologies – An Etymological Note. *Journal of Documentation*, 59 (1), 7–18.
- Gnoli, C., Poli, R. (2004). Levels of Reality and Levels of Representation. *Knowledge Organization*, 31 (3), 151–160.
- Gnoli, C., Pullmann, T., Cousson, Ph., Merli, G., Szostak, R. (2011). Representing the Structural Elements of a Freely Faceted Classification. In: A. Slavic, E. Civalero (eds). *Classification and Ontology: Formal Approaches and Access to Knowledge: Proceedings of the International UDC Seminar*, The Hague, 19–20 September 2011 (193–205). Würzburg: Ergon Verlag.
- Hjørland, B. (2012). Is Classification Necessary After Google? *Journal of Documentation*, 68 (3), 299–317.
- Hjørland, B. (2016). Does the Traditional Thesaurus Have a Place in Modern Information Retrieval? *Knowledge Organization*, 43 (3), 145–159.



- Ibekwe-SanJuan, F., Bowker, G.G. (2017). Implications of Big Data for Knowledge Organization. *Knowledge Organization*, 44(3), 187–193.
- Kempf, A.O., Neubert, J. (2016). The Role of Thesauri in an Open Web: A Case Study of the STW Thesaurus for Economics. *Knowledge Organization*, 43 (3), 160–173.
- León Manifesto (2007). *The León Manifesto* [online], ISKO-IT (10.05.2019), <http://www.iskoi.org/ilc/leon.php>
- MacFarlane, A. (2016). Knowledge Organization and Its Role in Multimedia Information Retrieval. *Knowledge Organization*, 43 (3), 180–183.
- McIlwaine, I.C., Broughton, V. (2000). Guest Editorial: The Classification Research Group – Then and Now. *Knowledge Organization*, 27 (4), 195–199.
- NAS (1959). *Proceedings of the International Conference on Scientific Information, Washington, D.C., November 16–21, 1958*. Washington: National Academy of Sciences.
- Otlet, P. (1934). *Traité de documentation: le livre sur le livre: théorie et pratique*. Brussels: Editions Mundaneum, Palais Mondial
- Perry, J.W., Kent, A. (1958). *Tools for Machine Literature Searching. Semantic Code Dictionary. Equipment. Procedures*. New York: Interscience Publishers.
- Rozzkowski, M. (2009). Simple Knowledge Organization System (SKOS) – reprezentacja wiedzy w sieciowych systemach organizacji wiedzy. *Zagadnienia Informatyki Naukowej*, 1, 89–102.
- Saracevic, T. (2010). Information Science. In: M. Bates, M.N. Maack (eds.). *Encyclopedia of Library and Information Sciences*. Third Edition, Boca Raton, FL, London: CRC Press, 2570–2584.
- Satija, M.P. (2017). Colon Classification (CC). Review of Concepts in Knowledge Organization. *Knowledge Organization*, 44 (4), 291–307.
- Severino, F. (2005). *What thesaurus to define EU/ACP relations? Analysis of the term development in the thesauri of the EU and other international organizations* [online]. Research presented at the ISKO Italy-UniMIB meeting: Milan, June 24, 2005. ISKO-IT [10.05.2019], <http://www.iskoi.org/doc/development.htm>
- Shiri, A. (2014). Linked Data Meets Big Data: A Knowledge Organization Systems Perspective. *Advances in Classification Research Online*, 24(1), 16–20.
- Skorochoodko, E.F. (1971). *Semanticheskie problemy avtomatizatsii informatsionnogo poiska*. Kijev: Naukova Dumka.
- Soergel, D. (2015). Unleashing the Power of Data Through Organizations: Structure and Connections for Meaning, Learning and Discovery. *Knowledge Organization*, 42(6), 401–427.
- Sosińska-Kalata, B. (2006). Tezaurusy w zmieniającym się środowisku wyszukiwania informacji. W: B. Sosińska-Kalata et al. (red.), *Informacja w sieci. Problemy, metody, technologie*. Warszawa: Wydaw. SBP, 140–156.
- Sosińska-Kalata, B. (2008). Kształtowanie się nowego paradygmatu organizacji zasobów wiedzy w społeczeństwie sieciowym. W: D. Pietruch-Reizes, W. Babik (red.). *Wymiana informacji i rozwój profesjonalnych usług informacyjnych w edukacji, nauce i kulturze na rzecz społeczeństwa opartego na wiedzy*. Katowice: PTIN, 17–23.
- Sosińska-Kalata, B. (2011). Nowe narzędzia organizacji wiedzy a jakość usług informacyjnych. W: D. Pietruch-Reizes, W. Babik (red.). *Bezpieczna, innowacyjna i dostępna informacja. Perspektywy dla sektora usług informacyjnych w społeczeństwie wiedzy*. Katowice: PTIN, 95–109.
- Sosińska-Kalata, B. (2017). Kierunki rozwoju współczesnej informatologii. *Forum Bibliotek Medycznych*, 10 (2), 25–46.
- Sosińska-Kalata, B. (2018). Big data (dane masowe) w nauce o informacji. *Zagadnienia Informatyki Naukowej*, 56 (2), 7–35.
- Spiteri, L. (1995). The Classification Research Group and the Theory of Integrative Levels [online]. *Katharine Sharp Review*, 1, 1–8, [10.05.2019], [https://www.ideals.illinois.edu/bitstream/handle/2142/78239/spiteri\\_classification.pdf?sequence=2&isAllowed=y](https://www.ideals.illinois.edu/bitstream/handle/2142/78239/spiteri_classification.pdf?sequence=2&isAllowed=y)

- Stanford (n.d.). *Working Papers Concerning the Creation of Google* [online]. Stanford University Digital Library. Stanford Digital Library Technologies Project [7.05.2019], <http://diglib.stanford.edu:8091/diglib/pub/projectdir/google.html>
- Tudhope, D., Binding, C. (2016). Still Quite Popular After All Those Years – The Continued Relevance of the Information Retrieval Thesaurus. *Knowledge Organization*, 43 (3), 174–179.
- White, M. (2016). The Value of Taxonomies, Thesauri and Metadata in Enterprise Search. *Knowledge Organization*, 43 (3), 184–192.
- Wojtasiewicz, O.A., Bojar, B. (1977/1978). *Sprawozdanie z prac nad kodem semantycznym dla SINTO* [dok. niepubl.].

---

## Knowledge Organization in the Digital World: The Impact of the Digital Revolution on the Evolution of Knowledge Organization Methods and Tools

### Abstract

**Purpose/Thesis:** The aim of the paper is to analyze the changes taking place in knowledge organization as a result of the digital revolution and to identify key factors which promote new methods and tools for organization of access to recorded knowledge in the digital environment.

**Approach/Methods:** The article is a theoretical reflection on the evolution of methods and tools of knowledge organization. With a reference to the subject literature, the process of shaping the concepts of organizing access to recorded knowledge in the digital environment was analyzed and its technological and social conditions were discussed. Attention is drawn to the development of the big data phenomenon's implications for knowledge organization and to new directions of knowledge organization systems (KOS) applications.

**Results and conclusions:** Among the technological factors that have the highest impact on shaping new concepts of knowledge organization in the digital world, and directions of modifying traditional models of KOS and creating new ones, the following are indicated: the development of natural language processing (NLP) technologies, hypertext development, the development of semantic technologies, machine learning methods and Web 2.0 technologies, information visualization techniques, adaptation of KOS functionality to applications on mobile devices. Among the social conditions, the following are identified as the factors which play the major role in determining the directions of the evolution of knowledge organization methods and tools: the focus was on non-professional user service and adaptation methods of KOS development, structure and functionality to the growing diversity of information objects represented by them and users' questions directed to information services.

**Originality/Value:** As far as the author knows, the presented considerations are the first comprehensive attempt to trace the impact of the digital revolution on the development of methods and tools for knowledge organization in the Polish research literature.

### Keywords

Big data. Digital environment. Digital revolution. Faceted classification. Gigadata. Knowledge organization. Knowledge organization system. Thesaurus.

---

*Prof. dr hab. BARBARA SOSIŃSKA-KALATA jest kierownikiem Katedry Informatologii na Wydziale Dziennikarstwa, Informacji i Bibliologii Uniwersytetu Warszawskiego oraz redaktorem naczelną czasopisma Zagadnienia Informacji Naukowej – Studia Informacyjne i członkiem Komitetu Naukoznawstwa PAN. Specjalizuje się w problematyce nauki o informacji, w szczególności organizacji wiedzy, a także ilościowych*

*badan informacji oraz historii, teorii i metodologii nauki o informacji. Opublikowała ponad 250 prac, w tym 12 książek i ponad 160 artykułów naukowych. Do jej najnowszych publikacji należą: Nauka o informacji wśród nauk o kulturze (Kultura książki i informacji, Katowice, 2017), Książka (dokument) w środowisku informacyjnym (Encyklopedia Książki, Wrocław 2017), The impact of the works by Paul Otlet and Suzanne Briet on the development of the epistemology of documentation and information science in Poland (Fondaments épistémologiques et théoriques de la science de l'information-documentation, London, 2018), Big data (dane masowe) w nauce o informacji (Zagadnienia Informacji Naukowej, 2018).*

*Kontakt z autorką:*

*b.sosinska@uw.edu.pl*

*Katedra Informatologii*

*Wydział Dziennikarstwa, Informacji i Bibliologii*

*Uniwersytet Warszawski*

*ul. Nowy Świat 69*

*00-046 Warszawa*