

Barbara Sosińska-Kalata  
Uniwersytet Warszawski  
b.sosinska@uw.edu.pl

## NOWE NARZĘDZIA ORGANIZACJI WIEDZY A JAKOŚĆ USŁUG INFORMACYJNYCH

### NEW TOOLS OF KNOWLEDGE ORGANIZATION AND QUALITY OF INFORMATION SERVICES

**ABSTRACT:** Providing effective access to the content of information resources published in the Internet invariably is one of the basic tasks for information professionals. Nowadays two conceptions of improving the quality of this access are developed. The first one is the paradigm of high-quality smart metadata connected with the project of the Semantic Web. The second one is the idea of Social Web and tools of collaborative generating of digital contents as well as descriptive metadata, by non-professionals. There is still slow progress in implementation of the Semantic Web tools, hence there are hopes that prosument attitude of the Internet users and dynamic development of the tools for social tagging of digital contents in shorter perspective may provide more effective organisation and selection of networked resources which will better fit to diverse needs of their users. In the paper there are discussed fundamental features of these tools, focusing on credibility of social generated metadata and ensured by them quality of access to digital contents in various kinds of information searching contexts.

#### WPROWADZENIE

Zapewnienie łatwego i efektywnego dostępu do treści zasobów informacyjnych opublikowanych w Internecie jest bez wątpienia jednym z najważniejszych obecnie zadań, które stoją przed specjalistami informacji. Internet stał się głównym medium komunikacji zarówno specjalistów jak i szerokiej publiczności. Globalna sieć staje się coraz bardziej otwarta i społeczna, a pod względem zasobów informacyjnych - coraz bardziej zróżnicowana. Sytuacja ta stawia nowe wyzwania związane organizacją dostępu do tych zasobów. Mimo iż od wielu lat prowadzone są badania na tym polu, a w rozwijającym się już blisko dwie dekady ogólnie dostępnym Internecie wdrożono wiele różnych narzędzi przeszukiwania i przeglądania informacji, kwestia wypracowania wystarczająco efektywnej metody indeksowania i wyszukiwania nadal pozostaje otwarta.

Jakkolwiek dyskusyjne jest czy w odniesieniu do przeszukiwania zasobów Internetu ma sens odwoływanie się do klasycznych pomiarów efektywności wyszukiwania informacji, to jednak optymalizacja narzędzi dostępu do tych zasobów nadal wiąże się z równoczesnym spełnieniem właśnie klasycznych jej kryteriów: optymalnej kompletności i optymalnej dokładności wyszukiwania. Od niech niezmiennie zależy to, co jest kluczowe dla jakości usług informacyjnych, tj. rzeczywista ich użyteczność dla celów, w jakich użytkownicy z tych usług korzystają.

#### OGRANICZENIA TRADYCYJNYCH NARZĘDZI

Z tymi dwoma kryteriami oceny rezultatów wyszukiwania wiążą się główne oczekiwania wobec narzędzi dostępu do cyfrowych treści Internetu. Z jednej strony powinny one bowiem zapewnić dostęp do możliwie największych zasobów opublikowanych w Internecie treści, a z drugiej – powinny zapewnić taki wgląd w te treści, który umożliwi efektywną selekcję z ogromnych zasobów optymalnej wielkości podzbiorów treści relewantnych do potrzeb użytkownika. Postulatów tych w wystarczającym zakresie nie spełniają dwa powszechnie dziś stosowane narzędzia organizacji dostępu do zasobów sieci, czyli wyszukiwarki wykorzystujące automatyczne indeksowanie pełnotekstowe, oparte na prostej analizie syntaktycznej i statystycznej tekstów oraz systemy kategoryzacji czy taksonomii tematycznych zasobów sieciowych, wykorzystujące manualne klasyfikowanie internetowych treści przez odpowiednio przygotowanych klasyfikatorów (zob. Sosińska-Kalata, 2002).

#### Wyszukiwarki i indeksowanie automatyczne

Wyszukiwarki internetowe, z których największe obejmują swoim działaniem mniej niż jedną trzecią zasobów opublikowanych w Internecie, są dziś uważane za narzędzia najlepiej spełniające oczekiwania użytkowników. Zwłaszcza wśród tzw. *digital natives*, generacji epoki Internetu (zob. Prensky, 2001), powszechne jest przekonanie, że jeśli nawet wyszukiwarki takie jak Google po wpisaniu jednego czy dwóch słów kluczowych nie zaspokoją wszystkich ich potrzeb informacyjnych, to i tak w znacznym stopniu satysfakcjonujący skłonni są oni uznać uzyskany w ten sposób zestaw informacji. Internet jest narzędziem codziennego użytku, więc przeszukiwanie go musi być łatwe, generalnie systemy informacyjne w epoce informacji muszą być narzędziami łatwymi w użyciu. Ta postawa nowego pokolenia użytkowników usług informacyjnych ma i na pewno będzie mieć istotny wpływ na projektowanie narzędzi dostępu do zasobów informacyjnych. Oczekiwanie łatwości ich użycia nie jest jednak równoznaczne z rezygnacją z jakości efektów tego użycia. Status głównego narzędzia komunikacji powoduje, że od narzędzi organizacji dostępu do wiedzy zawartej w zasobach Internetu w coraz większym

stopniu oczekuje się też zapewnienia selekcji informacji o wysokim stopniu wiarygodności, często wąsko specjalistycznej, nie zawsze łatwo werbalizowalnej. Parafrazując jedno z praw Ranganathana, można byłoby stwierdzić, że każda informacja ma (może mieć) swojego użytkownika, a zatem każda powinna być odnajdywana. Rozszerzanie zasięgu działania wyszukiwarek internetowych sukcesywnie dokonywane jest zarówno przez rozwój funkcjonalności systemów ogólnych, jak i tworzenie wyszukiwarek specjalistycznych, działających w tzw. głębokiej czy niewidocznej sieci. Jednak poza problemem objęcia zasięgiem działania wyszukiwarek jak największych zasobów opublikowanych w Internecie treści, najistotniejsze ograniczenia tych narzędzi są konsekwencją stosowanych w nich metod indeksowania automatycznego. W istocie postęp w dziedzinie metod indeksowania automatycznego jest skromny, a to co postrzegamy jako efekt poprawy ich jakości wynika jedynie ze wzrostu mocy obliczeniowej komputerów i masowości przetwarzania danych tekstowych. Do automatycznej identyfikacji treści zasobów informacyjnych stosuje się metody de facto niewiele różniące się od tych, które stosowane były 30 lat temu. Nie zapewniają one wglądu w treści cyfrowe na poziomie semantycznym, a w konsekwencji wyniki opartego na nich wyszukiwania zależne są od formy (formatu, języka, etc.) prezentacji tych treści.

#### Katalogi i manualne klasyfikowanie

Z kolei wobec braku metod automatycznej klasyfikacji treści zasobów informacyjnych, które można byłoby wdrożyć na szerszą skalę, stosowane w praktyce manualne przygotowywanie metadanych do katalogów zasobów sieciowych opartych na ustrukturyzowanym słownictwie kontrolowanym, schematach klasyfikacyjnych czy kategoryzacjach tematycznych, implikują wysokie koszty, nie zawsze są też w stanie – szczególnie w przypadku dużych serwisów ogólnych - zapewnić oczekiwanego tempa aktualizacji gromadzonych informacji. Takie katalogi zawierają przygotowane przez profesjonalistów wysokiej jakości metadane i rejestrują zweryfikowane, wartościowe zasoby sieciowe, ale zwykle obejmują względnie małą część zasobów Internetu. Jest to zatem podejście dość ekskluzywne, które jednak mimo dość często spotykanej opinii o małej przydatności dla użytkowników (por. Kłopotek, 2002, s. 17), nadal jest szeroko stosowane. Z pewnością najlepiej sprawdza się ono w serwisach specjalistycznej informacji naukowej, takich jak tzw. *subject gateways*, gdzie często do organizacji zasobów wykorzystywane są bardziej lub mniej zmodyfikowane tradycyjne języki informacyjno-wyszukiwawcze (por. Roszkowski, 2009). Podejście to wykorzystują jednak także wielkie serwisy informacyjne, takie jak np. Yahoo!, nadal zatrudniający ponad 150 klasyfikatorów do redagowania uniwersalnego katalogu zasobów Internetu. Trzeba też pamiętać, że na tym podejściu do organizacji treści sieciowych opiera się też

powszechne porządkowanie informacji w skali mikro, czyli za pomocą tzw. taksonomii (zob. Tomaszczyk, 2007) w różnego typu serwisach hipertekstowych: organizacyjnych, społecznościowych, e-commerce, etc.

## NOWE NARZĘDZIA

Krytycznym problemem jakości dostępu do treści cyfrowych w wielkich zasobach Internetu jest pozyskiwanie metadanych, zwłaszcza metadanych deskrypcyjnych. Z tym zagadnieniem wiążą się poszukiwania nowych systemów organizacji wiedzy, które ułożyły się w dwa w pewnym sensie rywalizujące nurty badań (zob. np. Morville, 2005, s. 119-154). Nurty te można jednak postrzegać także jako w pewien sposób komplementarne i tworzące coś, co można byłoby nazwać semantyczno-społecznym podejściem do organizacji zasobów sieci.

### Semantic Web, silne metadane i sztuczna inteligencja

Pierwszy obejmuje badania skupione wokół wykorzystania technik sztucznej inteligencji w identyfikacji i przetwarzaniu treści cyfrowych, w tym konceptu ontologii informatycznej i projektu Semantic Web. Kwintesencją tych prac jest wyposażenie jednoznacznie identyfikowalnych cyfrowych treści w wysokiej jakości metadane reprezentujące ich warstwę pojęciową, powiązania strukturalne i atrybuty determinujące ich użycie oraz stosowanie na wielką skalę standaryzacji metadanych, która warunkuje współdzielenie zasobów informacyjnych przez różne systemy. W tym klasycznym inżynierskim podejściu wysoka jakość tych semantycznie nasyconych, silnych metadanych (ang. *smart metadata*) jest kwestią kluczową, a jakość usług informacyjnych, w których takie narzędzia byłyby wykorzystane jest pochodną jakości tych metadanych. Precyzyjny opis treści cyfrowych, ich weryfikacja i możliwość semantycznego przetwarzania pozwolić powinny na dokładniejsze przeszukiwanie zasobów sieci i efektywniejsze ich filtrowanie odpowiednio do potrzeb użytkownika. W skali mikro, czyli zastosowań tego rodzaju narzędzi w korporacyjnych intranetach, teza ta znajduje praktyczne potwierdzenie. Jednak choć Konsorcjum 3W, patronujące projektowi Semantic Web, opracowało pełną architekturę niezbędnych standardów metadanych, stanowiącą fundament tego projektu i wdrożyło większość z nich, to ich stosowanie w praktyce na szeroką skalę w Internecie jest zapewne jeszcze sprawą dość odległej przyszłości. Z konceptualnego punktu widzenia projekt Semantic Web jest wielce obiecujący, w praktyce jednak główny problem tkwi znowu w pozyskaniu metadanych wysokiej jakości koniecznych do automatycznego przetwarzania semantycznego.

## Social Web, „mądrość tłumu” i narzędzia społecznego klasyfikowania (tagowania)

Drugi kierunek wiąże się z rozwijaniem narzędzi, które ogólnie określa się mianem narzędzi Internetu społecznościowego (*Social Web*) czy też Web 2.0. Ich idea zasadza się na wykorzystaniu do tworzenia cyfrowych treści, w tym metadanych służących organizacji zasobów informacyjnych, potencjału tkwiącego w – nawiązując do określenia Alвина Tofflera - prosumenckich postawach użytkowników Internetu. U podstaw tego podejścia leży zatem przekonanie, że współczesny użytkownik informacji nie jest biernym jej odbiorcą, lecz aktywnym uczestnikiem rozmaitych procesów informacyjnych, związanych zarówno z publikowaniem własnych treści, jak i weryfikacją treści opublikowanych przez innych, ich oceną, streszczaniem i komentowaniem, a także indeksowaniem i organizowaniem zasobów cyfrowych – dla siebie i dla innych, którzy dzielają podobne zainteresowania. Z przekonaniem tym łączy się druga istotna dla tego podejścia teza, tj. teza o „mądrości tłumu”. Zgodnie z nią uśrednione, niezależne lecz skoordynowane decyzje dużej liczby osób, które nie muszą być profesjonalistami są zwykle trafniejsze niż decyzje pojedynczych ekspertów (Surowiecki, 2005, s. XIII, XV, XVII). Według Jamesa Surowieckiego dowodem tej „mądrości tłumu” jest m.in. sukces wyszukiwarki Google, której sprawność w selekcjonowaniu relewantnych stron opiera się na badaniu prowadzących do nich hiperłączy, co interpretować można jako badanie opinii użytkowników tworzących te hiperłącza. „Mądrość tłumu” warunkuje też jakość usług informacyjnych, w których wykorzystuje się społecznościowe narzędzia organizacji wiedzy.

### SPOŁECZNE KLASYFIKOWANIE, TAGI, FOLKSONOMIE

Jednym z pierwszych projektów opartych na takich założeniach był zarządzany przez firmę Netscape opensoursovyy projekt DMOZ - największy obecnie katalog internetowy *Open Directory*, tworzony od 1998 r. przez ponad 80 tysięcy wolontariuszy z całego świata. Dzisiaj rejestruje on ponad 4.5 mln stron internetowych i zawiera blisko 600 tysięcy kategorii tematycznych. Z metadanych zgromadzonych w tym katalogu korzystają setki wyszukiwarek i katalogów wielkich portali, m.in. Google, Netscape Search, AOL Search, Lycos, HotBot, DirectHit.

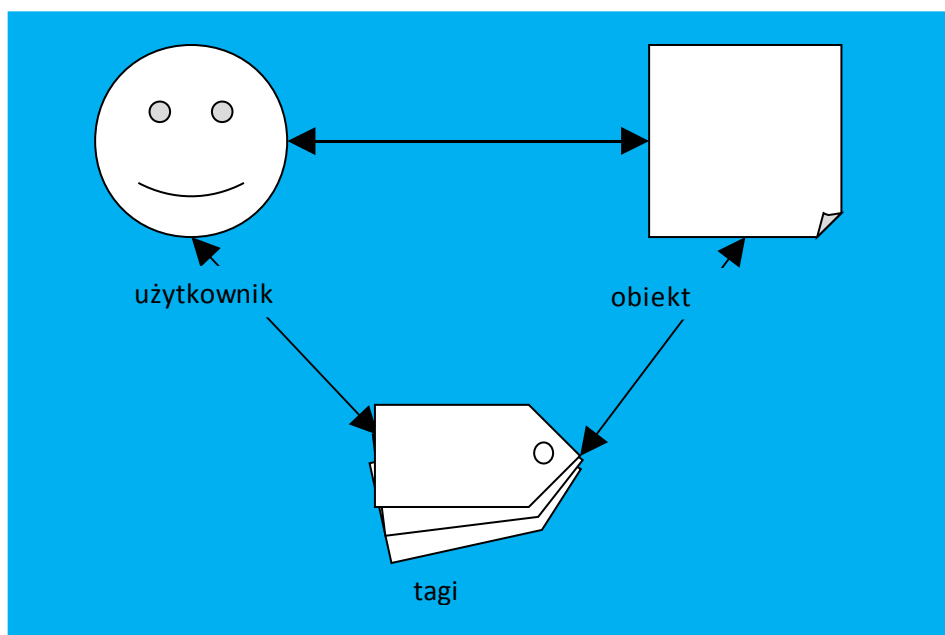
Tworzenie *Open Directory* oparte na pracy wolontariuszy, podporządkowanej pewnym redakcyjnym zasadom, przypomina sposób realizacji obecnie najpopularniejszego społecznościowego narzędzia, jakim jest Wikipedia. Notabene poza tworzeniem treści encyklopedycznych jej redaktorzy również te treści klasyfikują, przyporządkowując artykuły do istniejących już lub zaproponowanych przez siebie kategorii. Zasada „neutralnego punktu widzenia”, który negocjowany jest między różnymi użytkownikami, stanowi tu podstawę

zapewniania jakości zarówno treści, jak i dostępu do nich przez system kategorii tematycznych (por. Weinberger, 2007).

#### Modelowe cechy społecznego klasyfikowania

Podjęcie to różni się jednak nieco od koncepcji całkowicie swobodnie tworzonych klasyfikacji społecznościowych, czy też rozproszonych systemów klasyfikacyjnych budowanych przez użytkowników serwisów Web 2.0, które określa się mianem folksonomii (Guy i Tonkin, 2006). Za główną ich cechę uznaje się bowiem właśnie to całkowicie swobodne indeksowanie obiektów (treści) cyfrowych za pomocą niezależnie wybieranych wyrażań - tzw. tagów (znaczników, etykiet, kategorii) oraz społeczny kontekst tego wspólnego indeksowania. Działające w ten sposób tzw. systemy tagowania (ang. *tagging systems*) rejestrują równocześnie trzy rodzaje powiązań (por. Vander Wal, 2005):

- poszczególnych tagów z określonymi obiektami (treściami) cyfrowymi,
- poszczególnych użytkowników z opisywanymi przez nich obiektami cyfrowymi
- poszczególnych użytkowników z używanymi przez nich tagami (rys. 1).



**Rysunek 1: Podstawowy model tagowania (na podstawie: Smith, 2008, s. 4)**

W takim indeksowaniu wybierane przez poszczególnych użytkowników tagi reprezentują ich indywidualny punkt widzenia, wzbogacając wielowymiarowość opisu określonych treści. System tagowania umożliwia dzięki temu dostosowanie organizacji treści opisywanych przez

zbiorowość różnych użytkowników do różnych indywidualnych potrzeb, wiedzy językowej i dziedzinowej oraz perspektyw poznawczych.

Pomysł wykorzystania do organizacji dostępu do zbiorów informacji słów kluczowych swobodnie wybieranych przez nie-ekspertów nie jest oczywiście nowy. Od dawna wiele czasopism naukowych wykorzystuje słowa kluczowe wskazane przez autorów jako indeksowanie uzupełniające, które zasila również bibliograficzne bazy danych. Opisywanie stron HTML publikowanych w systemie WWW przez twórców je użytkowników za pomocą metadanych z tagiem „Keywords” było jedną z najwcześniejszych cech tego systemu, która miała wyszukiwarkom zapewnić trafne selekcjonowanie dokumentów. Trzeba jednak pamiętać, że w obu wspomnianych przypadkach nie chodzi jednak o opis treści dokumentów przez użytkowników lecz przez autorów – w pierwszym na ogół uznawany za solidny i użyteczny, w drugim – jak pokazała praktyka – na tyle niewiarygodny, że większość współczesnych wyszukiwarek z tych metadanych w ogóle nie korzysta.

#### Najwcześniejsze systemy społecznego tagowania

Zdaniem Thomasa Vander Wala, najwcześniej użytym systemem tagowania był Lotus Magellan – oprogramowanie z 1988 r., które umożliwiało użytkownikom zapisywanie komentarzy czy adnotacji do dokumentów i innych obiektów cyfrowych przechowywanych na twardym dysku, a następnie odnajdywanie materiałów związanych z określonym tematem czy zadaniem przy użyciu technik wyszukiwania pełnotekstowego przeprowadzanego na tych adnotacjach. Za wczesny system tagowania uznać można również narzędzie zastosowane w połowie lat 90. przez jeden z pierwszych komercyjnych internetowych serwisów komunikacyjnych – amerykański CampuServ. Umożliwiało ono użytkownikom forów dyskusyjnych CampuServ opisywanie za pomocą słów kluczowych treści wątków i postów (Vander Wal, 2007).

#### Gwałtowny rozwój systemów tagowania w ostatnich pięciu latach

Prawdziwa jednak eksplozja społecznego indeksowania treści opublikowanych w Internecie nastąpiła wraz z pojawieniem się systemów tagowania w pierwszych serwisach społecznościowych. W 2003 r. serwis Delicious zaoferował użytkownikom narzędzia opisywania zakładki do stron internetowych za pomocą niekontrolowanych słów kluczowych. W 2004 r. podobny system udostępnił serwis Flickr, na którym użytkownicy swobodnie dobrane etykietami mogą opisywać cyfrowe fotografie. W kolejnych latach narzędzie to udostępnione zostało użytkownikom wielu działających już wcześniej serwisów społecznościowych (np. Technorati, You Tube), powstawały też liczne następne, umożliwiając użytkownikom

organizowanie rozmaitych rodzajów zasobów na własny użytek lub też na użytek społeczności, z którą się identyfikują (np. Connotea i CiteULike – serwisy porządkowania zakładki do tekstów naukowych, SlideShare – serwis porządkowania slajdów, Ma.gnolia – serwis porządkowania zakładki, Library Thing – serwis wspólnego katalogowania książek, Last.fm i MusicBrainz – serwisy porządkowania plików muzycznych, Buzzillions – serwis gromadzący opinie o produktach dostępnych w sklepach internetowych).

Popularność tagowania w serwisach społecznościowych spowodowała zainteresowanie tym narzędziem także ze strony e-handlu (np. Amazon, Etsy, sklepy spożywcze jak polska Alma24), który odkrył w nim możliwość ułatwienia klientom dostępu do produktów, zachęty do zakupów, a firmom równocześnie monitorowanie opinii i upodobań. Systemy tagowania udostępnia wiele serwisów dziennikarskich, zarówno zawodowych (np. tvn24.pl), jak i tzw. obywatelskich (np. Wiadomości24.pl). Narzędzia do porządkowania treści prywatnych swobodnie wybranymi słowami kluczowymi oferuje Gmail (poczta elektroniczna serwisu Google) i iTunes (agenda firmy Apple sprzedająca pliki muzyczne online). Systemy tagowania wprowadza do swych serwisów internetowych także coraz więcej organizacji i instytucji działających w domenie publicznej, w tym coraz więcej różnego typu bibliotek. Niektóre biblioteki jako uzupełnienie albo zamiast swojego katalogu wykorzystują system tagowania serwisu Library Thing (ciekawym przykładem jest Delany Public Library). Inne wykorzystują zakładki z serwisu Delicious zamiast stosowanych dotąd własnych kolekcji linków do zasobów sieci interesujących ich użytkowników (np. Tunder Bay Public Library, Nashville Public Library, College of New Jersey Library). Narzędzia do tagowania kolekcji zarejestrowanych w katalogach bibliotecznych zaoferowało wielu producentów zintegrowanych systemów bibliotecznych w postaci tzw. *social OPAC*, m.in. obecne na polskim rynku firmy Ex Libris (w 2007 r. wyposażała w funkcję tagowania system Aleph Primo), i VTLS (w lipcu 2009 r. zaanonsowała produkt Chamo - „społeczny OPAC” z systemem tagowania; zob. VTLS..., 2009).

W ciągu zaledwie czterech lat społeczne klasyfikowanie z funkcjonalności charakterystycznej dla społecznościowych serwisów organizacji linków do różnych zasobów sieci stało się własnością, która zdaje się być postrzegana jako niezbędna w niemal każdej nowej aplikacji internetowej. Nie jest jednak do końca jasne, czy jest to wynik faktycznej użyteczności tych narzędzi organizacji wiedzy, czy też tylko moda, za którą kryje się przekonanie, że wyposażenie serwisu w system tagowania i generowanie własnej folksonomii zapewni serwisowi lepszy publiczny wizerunek, sugerując otwartość na współpracę z użytkownikami.



## Rodzaje i cechy specyficzne systemów tagowania

Systemy tagowania nie są jednorodne. Można wskazać wśród nich podtypy, różniące się zasadami opisywania zasobów, ich rodzajem, przestrzenią działania. Na podstawie przytoczonych wcześniej przykładów takich systemów można zaproponować następującą ich ogólną typologię:

- systemy klasyfikowania społecznościowego zakładek do zasobów Internetu, których celem jest współorganizowanie zasobów informacyjnych, ale które nie służą publikowaniu własnych zasobów (np. Delicious, Connotea, CiteULike, polska Linkologi.pl)
- systemy współdzielenia się zasobami informacyjnymi, umożliwiające publikowanie i porządkowanie określonego rodzaju treści cyfrowych (np. Flickr, Technorati, You Tube, Last.fm, Library Thing)
- systemy tagowania do organizacji informacji osobistej (PMI – personal management information; np. Gmail i iTunes)
- systemy tagowania wykorzystywane w handlu internetowym (np. Buzzillions, Amazon, Alma.pl)
- systemy tagowania zawartości poszczególnych witryn internetowych (np. serwisy dziennikarskie TVN24.pl i Wiadomości.pl, serwisy instytucjonalne czy organizacyjne, jak IFLA.org).

Systemy tagowania różnią się też pod względem zasad dotyczących klasyfikowania, zasad porządkowania tagów, metod prezentacji tagów itp.

Kolejny obszar różnic dotyczy przewidzianego w serwisie udziału użytkowników w korzystaniu z systemu tagowania. Część systemów tagowania pozwala użytkownikom jedynie przeglądać zasoby serwisu za pomocą tagów, nie daje im jednak możliwości samodzielnego ich opisywania (np., używane na stronach mediów prasowych, albo instytucji i organizacji). Takiego systemu tagowania nie można zatem traktować jako społecznego narzędzia organizacji dostępu do treści cyfrowych. Niektóre serwisy pozwalają użytkownikom opisywać zasoby informacyjne, ale w pewnym stopniu kontrolują użycie określonych tagów – tego rodzaju nadzór nad słowami kluczowymi używanymi jako tagi prowadzi np. Amazon. W pełni społecznościowe klasyfikowanie zapewniają systemy tagowania, które nie stosują żadnej kontroli używanego słownictwa.

Typowe społecznościowe systemy tagowania charakteryzują przede wszystkim następujące cechy:

- indywidualizacja opisu: użytkownik może tworzyć pewien własny sposób przeglądania swojej własnej kolekcji obiektów cyfrowych, we własnym języku, odpowiednio do swoich zainteresowań

- indeksowanie fasetowe, czyli wieloaspektowość opisu: użytkownik może każdemu obiektowi przypisać tyle tagów, ile uzna za stosowne; każdy tag reprezentuje cechę obiektu rozpatrywaną z istotnego dla użytkownika punktu widzenia
- indywidualizacja opisu obiektów za pomocą tagów pozwala reprezentować unikalne aspekty ich recepcji, które mogą być użyteczne również dla innych użytkowników
- w serwisach społecznościowych, fasetyzacja opisu może osiągać bardzo wysoki poziom; wieloaspektowość opisu poszczególnych obiektów cyfrowych potęguje kumulacja rozmaitych tagów przypisanych danemu obiektowi przez różnych użytkowników
- demonstracja „mądrości tłumu” – społecznego konsensusu w opisie zasobów informacyjnych i poszczególnych obiektów cyfrowych
  - pomiar popularności tagów jako projekcja społecznej recepcji zbiorów. Dla zgromadzonych tagów system oblicza frekwencję użycia, niekiedy relatywizowaną do liczby użytkowników wprowadzających dany tag; wynik tego pomiaru wizualizowany jest najczęściej w sugestywnej formie tzw. chmury tagów (chmury znaczników), albo też w postaci list z numeryczną informacją o częstotliwości użycia poszczególnych kategorii; tagi uporządkowane są alfabetycznie, a stopień popularności wyrażany jest za pomocą wielkości czcionki (w chmurach) lub liczby użyć (na listach); taka chmura czy lista tagów pokazuje zatem tematy, na których skupia się uwaga społeczności korzystającej z danego serwisu, a także tendencje (język, aspekty) w sposobie ich opisywania
  - w opisie poszczególnych obiektów gromadzone są tagi, wprowadzane przez wszystkich użytkowników zainteresowanych tymi obiektami; taki opis, demonstrowany niekiedy w postaci chmury tagów, stanowi więc wynik pewnego rodzaju społecznego konsensusu dotyczącego danego obiektu: strony internetowej, książki, fotografii, produktu, etc.
- systemy tagowania są narzędziami przeglądania zasobów informacyjnych według określonych atrybutów, reprezentowanych przez poszczególne tagi; w przeglądaniu zasobów informacyjnych tagi mogą być agregowane, co pozwala stopniowo zawęzić obszar przeglądanych treści według aspektów interesujących użytkownika
- tagi mogą być współdzielone i łączone przez systemy tagowania używane przez wiele serwisów. Np. agregowane mogą być tagi serwisów Technorati, Flickr i Delicious (Smith. 2008, s. 3).

Ta ostatnia cecha, w niewielkim stopniu dotąd wykorzystywana, może stanowić o wielkim potencjale społecznych systemów tagowania dla organizacji zasobów Internetu w skali ogólnej. Rozwijające się systemy tagowania dostarczają bowiem niezwykle bogatego słownictwa opisu

różnorodnych treści, powiązanych z określonymi zasobami cyfrowymi. A zatem, dostarczają wielkie ilości deskrypcyjnych metadanych.

### Rodzaje tagów

Słownictwo używane w systemach tagowania do opisu różnego typu obiektów cyfrowych reprezentuje różne perspektywy poznawania świata, różne kultury i subkultury, różne poglądy i różne poziomy wiedzy. Stanowi więc niezmiernie interesujące źródło wiedzy nie tylko o zasobach informacyjnych Internetu, ale też o ich użytkownikach. Jest to więc niezmiernie ciekawe pole badań zachowań użytkowników i badań takich na świecie stopniowo prowadzi się coraz więcej.

Z punktu widzenia organizacji dostępu do wiedzy interesujące jest na to, w jakim zakresie wyrażane w tagach aspekty opisu treści cyfrowych odbiegają od aspektów standardowo wykorzystywanych w klasycznych systemach organizacji wiedzy. Badania dotyczące tej kwestii pokazują, że bardzo różnorodne słownictwo używane jako tagi można przyporządkować do stosunkowo niedużej liczby kategorii, z których część pokrywa się z klasycznymi aspektami opisu dokumentów (treść, forma zasobu, autorstwo, źródło), ale też wiele to aspekty w profesjonalnych opisach nie brane pod uwagę (opinia, zadanie, przedstawienie).

Tabela 1: Główne rodzaje tagów (na podstawie: Smith, 2008, s. 67, Zollers, 2007)

tagi treści	css, webdesign, politics, food, family, travel, mleko, papier
tagi zasobu	blog, video, book, podcast, gry
tagi źródła lub autorstwa	nytimes, newsweek, vanderwal
tagi opinii	cool, *****, defective_by_design, bad_music, awesome, intresting_book, ciekawostki
tagi osobistego użytku zasobów	my, mine, mystuff
tagi oznaczające zadania	do_przeczytania, todo, download
tagi performatywne (ang. <i>play &amp; performance</i> )	squaredcircle, seenlive, aka_vogon_poetry

### Zalety i wady systemów tagowania

Systemy tagowania są narzędziem modnym i popularnym wśród użytkowników. Ich niewątpliwą zaletą jest wielka łatwość zbiorowego lub indywidualnego opisywania cyfrowych

treści, które interesują użytkownika. Dla zwolenników społecznościowego Internetu duże znaczenie ma to, że w przeciwieństwie do autorytatywnie narzuconych systemów kategoryzacyjnych, takich jak klasyfikacje biblioteczne czy taksonomie organizujące zawartość witryn internetowych, społeczne klasyfikacje są demokratycznymi narzędziami tworzenia dostępu do wiedzy, bo każdemu pozwalają organizować jej cyfrowe zasoby według własnych pomysłów i preferencji, za pomocą własnego języka. Sprzyja to rozwojowi internetowych społeczności dzielących określone zainteresowania i wokół nich budujących własne repozytoria wiedzy, często tworzących specyficzne subkultury. Odrzucenie standardowych narzędzi, ujednocających i zwykle generalizujących reprezentację treści oraz zapewnienie całkowitej swobody organizowania dostępu do informacji i wiedzy sprzyja także aktywowaniu społeczności lokalnych, wspieraniu ich tradycji i tożsamości (por. Boast i in., 2007).

Timo Hanney z Nature Publishing Group, która prowadzi serwis Connotea - społecznościowy serwis zakładki do treści naukowych, zwraca uwagę na to, że systemy tagowania pozwalają opisywać piśmiennictwo naukowe za pomocą najbardziej aktualnego słownictwa używanego w komunikacji naukowej i ma to istotne znaczenie dla rozwoju nauki. Z jednej strony takie systemy umożliwiają sprawniejsze zarządzanie coraz większymi zasobami różnych typów informacji, do których dostęp jest niezbędny w pracy naukowej: nie tylko piśmiennictwem, ale też danymi badawczymi i danymi uzyskanymi z ich analiz. W Internecie są obecne wszelkie rodzaje informacji potrzebnych naukowcom, aby mogli monitorować aktualny stan wiedzy w ich dziedzinach i być efektywni w swoich pracach. Szczególnie ważny jest dostęp do informacji najnowszej, dla której często nie ma jeszcze przyjętego szerzej nazewnictwa i która nie została jeszcze jednoznacznie sklasyfikowana. Taka informacja w tradycyjnych systemach organizacji wiedzy ukryta jest pod nazwami ogólnymi i arbitralnie przyporządkowywana kategoriom w sposób niezgodny z aktualnym stanem wiedzy badaczy. Niemożliwe jest, aby organizacje odpowiadające za znormalizowany opis piśmiennictwa naukowego za pomocą słownictwa kontrolowanego czy klasyfikacji były w stanie zreklasyfikować całą wiedzę naukową odpowiednio do aktualnego stanu poznania, jednak możliwe wydaje się, że sama społeczność naukowców przez wspólne kategoryzowanie robi to przynajmniej w odniesieniu do piśmiennictwa, które jest dla nich szczególnie interesujące lub użyteczne. Jeszcze inną wartością systemów tagowania dla świata nauki może być to, że systemy te są obszarem kształtowania się i uzgadniania pomiędzy specjalistami aktualnej terminologii (Smith, 2008, s. 36).

Społeczne aspekty systemów tagowania powodują, że systemy te powinny mieć też znaczną wartość dla wspierania dzielenia się wiedzą i współpracy w środowisku organizacyjnym.

Współdzielenie kluczowych zasobów i tworzenie wspólnego słownictwa może pomagać w zmniejszaniu dystansu poznawczego między członkami organizacji czy pracownikami przedsiębiorstwa. To z kolei powinno służyć promowaniu społecznego kapitału, podstawowego czynnika we współdzieleniu wiedzy (por. Sinclair i in., 2008).

Z ekonomicznego punktu widzenia niebagatelną zaletą systemów społecznego tagowania jest to, że dostarczają one wielkie zasoby tanich metadanych deskrypcyjnych wzbogacając tym samym wgląd w semantykę cyfrowych treści.

Jednak z punktu widzenia efektywności wyszukiwania systemy tagowania mają też wiele cech dyskusyjnych. Słaby jest jeszcze stan poznania motywacji i sposobów używania tych narzędzi, więcej jest zatem dotyczących ich pytań aniżeli odpowiedzi. Pytania kluczowe o jakość społecznego klasyfikowania wiążą się z oceną klasyfikującej społeczności. W większości serwisów udostępniających użytkownikom funkcję opisywania zasobów użytkownikiem może zostać każdy, kto w serwisie się zarejestruje, a warunki rejestracji nie stawiają żadnych istotnych barier. Zasady uczestnictwa są zatem bardzo liberalne, a w konsekwencji metadane opisujące zasoby informacyjne tworzy bardzo zróżnicowana społeczność.

Zdaniem Gene'a Smitha istnieją cztery podstawowe płaszczyzny napięć związane z oceną użyteczności systemów tagowania (Smith, 2008, s. 15):

Osobisty <--> społeczny	Czy ludzie opisują tagami zasoby informacyjne dla swojego własnego użytku? Czy też motywacją do tworzenia tagów jest dzielenie się informacją z jakąś grupą, albo chęć bycia postrzegany jako osoba znająca się na pewnej problematyce, albo jeszcze jakieś inne społeczne czynniki?
Idiosynkratyczny <--> standardowy	Czy tagi powinny być całkowicie unikalne i idiosynkratyczne? Może powinny być one standaryzowane, aby mogły być używane w przeglądaniu i wyszukiwaniu?
Swoboda <--> kontrola	Czy system daje użytkownikom całkowitą swobodę? Czy też wpływa lub kontroluje wybór tagów (np. za pomocą sugestii)?
Amator <--> ekspert	Jakie kwalifikacje mają ludzie opisujący zasoby informacyjne? Czy tagi wprowadzone przez amatorów i przez ekspertów mają równą wartość? Jak odróżnić wyrażone w tagach popularne opinie i opinie ekspertów wówczas, gdy nie są one zgodne?

Systemy społecznego klasyfikowania generują metadane, które można byłoby określić jako „metadane mas”. Propagując systemy społecznego klasyfikowania przyjmuje się ogólne założenie, że tymi masami ludzi opisujących zasoby informacyjne kieruje dobra wola i nastawienie na rzetelną współpracę dla wspólnego dobra. Cenny ze społecznego punktu widzenia brak kontroli wprowadzania tagów prowadzi jednak do powstawania chaosu, który w świetle zasad tradycyjnej organizacji wiedzy utrudnia dostęp do informacji. Typowe zarzuty związane z brakiem kontroli słownictwa dotyczą przeniesienia do systemów tagowania synonimii i homonimii języka naturalnego oraz braku powiązań hierarchicznych, które umożliwiłyby rozszerzenie i zawężanie pola zainteresowań. Rozproszenie informacji jest też konsekwencją specyficznej synonimii między poprawnymi i niepoprawnymi grafiami tych samych wyrażen. Nie można jednak wykluczyć ani zachowań antyspołecznych ani zwykłych nieporozumień.

W praktyce stosuje się więc pewne mechanizmy ograniczające nadmierny chaos w systemach tagowania:

- sugestie serwisu dotyczące formy gramatycznej tagów zapobiegają równoległemu używaniu tych samych wyrażen w liczbie mnogiej i w licznie pojedynczej
- sugestie dotyczące pożytku z używania tych samych tagów co inni użytkownicy pozwalają redukować synonimię leksykalną
- niektóre serwisy łączą tagi synonimiczne za pomocą pierścieni synonimów; wymaga to jednak ingerencji personelu serwisu
- bywają stosowane mechanizmy generowania struktur hierarchicznych słownictwa na podstawie analizy ich współwystępowania w opisach tych samych obiektów, np. mechanizm klasteryzacji w serwisie Flickr, algorytm konkurencji w serwisach Delicious i CiteULike
- niektóre systemy organizują tagi we wcześniej określone fasety, do których użytkownicy wpisują wybrane przez siebie wyrażenia, np. w serwisie Buzzillions, służącym ocenie produktów dostępnych w sklepach online, użytkownicy umieszczają tagi w fasetach „Pros”, „Cons” i „Best uses”, w których następnie wyświetlane są tagi opisujące produkt.

Problemy związane z zachowaniami antyspołecznymi próbuje się rozwiązać np. za pomocą algorytmów obliczania popularności tagów, w których uwzględnia się nie tylko informację o frekwencji użycia, ale też o tym, ilu użytkowników wprowadziło dany tag. Pozwala to uniknąć sytuacji takiej, gdy pewien użytkownik z wielką pasją wprowadza tagi niekoniecznie związane z zainteresowaniami innych użytkowników, ale jego aktywność powoduje, że w chmurze tagów

prezentującej najczęściej używane wyrażenia dominują właśnie te wprowadzone przez aspołecznego pasjonata. Nieuwzględnianie w pomiarze popularności tagów ich powiązania z konkretnymi użytkownikami skutkuje wyeksponowaniem w serwisie treści niekoniecznie najbardziej reprezentatywnych dla społeczności tego serwisu. Dobrym przykładem może być obecna sytuacja w serwisie Connotea, zasadniczo adresowanym do specjalistów medycyny i ochrony zdrowia i wysoko ocenianym w badaniach (por. Hammond i in. , 2005b, Xia Lin i in., 2006) Tymczasem wśród najpopularniejszych w tej chwili tagów zobaczyć można zaledwie kilka związanych z tym tematem, a bardzo wiele tagów ogólnych, dotyczących Internetu, gier, filmów, usług online, bibliotek, projektowania, biznesu – i to jeszcze w rosyjskiej wersji językowej (<http://www.connotea.org/cloud> – stan chmury w dn. 21.09.2009).

Odrębny problem stanowi wykorzystywanie serwisów społecznościowego klasyfikowania przez spamerów, do własnych celów biznesowych, reklamy, celowego rozpowszechniania określonych treści czy poglądów.

#### Wyszukiwanie informacji i użyteczność chmur tagów

Przytoczony przykład chmury tagów w serwisie Connotea prowadzi do refleksji o użyteczności tej techniki organizacji dostępu do informacji. Chmura tagów jest chyba najbardziej charakterystyczną cechą zewnętrzną systemów tagowania i wizualnie atrakcyjną formą prezentacji słownictwa wyszukiwawczego. Jednak praktyczne jej używanie do pozyskiwania informacji budzi wątpliwości. T. Vander Wal uznał ją za „cute but offering little value” [uroczą, ale zapewniającą małą wartość] (Vander Wal, 2005).

Ciekawe badanie tego problemu przedstawili James Sinclair i Michael Cardew-Hall, w warunkach eksperymentalnych obserwując sposób korzystania przez użytkowników (studentów inżynierii na Australian National University) z chmur tagów lub wyszukiwarki (Sinclair i Cardew, 2008). Wstępnie założyli, że jeśli chmura tagowa nie jest wygodnym narzędziem dostępu do informacji, to po kilku próbach użytkownicy będą z niej rezygnować na rzecz tradycyjnej wyszukiwarki. Badania pokazały jednak, że chociaż duża grupa uczestników eksperymentu rzeczywiście korzystała wyłącznie z wyszukiwarki, to jednak także znaczna ich część do poszukiwań stosowała chmurę znaczników. Wynik ten jest tym bardziej interesujący, że większość uczestników eksperymentu nie miała wcześniej styczności z systemami tagowania. W całych badaniach liczba pytań, na które odpowiedź znaleziona została za pomocą chmury tagów była większa niż liczba pytań, na które odpowiedź znaleziono za pomocą wyszukiwarki. Sinclair

i Cardew-Hall uznali zatem, że chmury znaczników mają wartość jako narzędzie umożliwiające skuteczne przeszukiwanie zbiorów danych.

Szczegółowa analiza wyników tych badań pozwoliła wyodrębnić dwie charakterystyczne sytuacje, kiedy użytkownicy korzystają z chmury tagów: pierwsza to szukanie odpowiedzi na pytania ogólne, informacje dotyczące szerokich i nieprecyzyjnie określonych tematów; druga to sytuacja odpowiedzi na pytania o przedmiot, którego nazwa użyta jest jako tag. Zadania ogólne w eksperymencie wyznaczone zostały za pomocą poleceń „Wskaż tytuł artykułu, który uważasz za interesujący” i „Wskaż tytuł artykułu, który uważasz za zabawny”. Pytania te wymagały przeglądania zasobów serwisu i w tym celu jako narzędzia użyteczne sprawdziły się chmury tagowe. Chmury znaczników wspierają zatem tzw. „przypadkowe wyszukiwanie”.

Badania pokazały też, że szukanie odpowiedzi na pytania za pomocą chmury tagów wymagało sformułowania większej liczby instrukcji, niż w przypadku korzystania z wyszukiwarki. To również pokazuje, że korzystając z chmury tagowej użytkownicy nastawiają się na jej przeglądanie raczej, niż na wyszukiwanie za pomocą jednego wybranego hasła.

Zadania szczegółowe do wykonania których najczęściej używano chmury znaczników dotyczyły tematów, których nazwy użyte były jako znaczniki. W takim przypadku użytkownicy decydowali się po prostu kliknąć na odpowiedni znacznik, zamiast np. wpisywać poszukiwaną nazwę do okna wyszukiwarki.

Z przytoczonych badań wnika, że poszukiwanie informacji przez przeglądanie chmur tagów jest działaniem intelektualnie łatwiejszym niż formułowanie precyzyjnej instrukcji dla wyszukiwarki. Sformułowanie pytania do wyszukiwarki wymaga znajomości odpowiedniej terminologii, umiejętności operowania bliskoznacznikami i hiperonimami. Przeglądanie znaczników dostarcza użytkownikom pomocy w wyborze słownictwa wyszukiwawczego dzięki wizualizacji treści zasobów za pomocą chmur znaczników. Użytkownik nie wymyśla zatem terminów wyszukiwawczych, a jedynie ocenia, które spośród prezentowanych mu wyrażen najlepiej odpowiadają identyfikacji tego, co go interesuje. Zauważono np., że uczestnicy eksperymentu, dla których język angielski był drugim językiem chętniej korzystali z chmury znaczników niż z wyszukiwarki.

Wielu uczestników eksperymentu Sinclaira i Cardew-Halla zwracało uwagę na to, że chmura tagów nie umożliwia wystarczającego zawężenia tematu poszukiwań. Jeśli nie ma znacznika nazywającego wprost poszukiwany temat, a jest to temat dość szczegółowy, przeglądanie zawartości chmury zajmuje więcej czasu niż skorzystanie z wyszukiwarki.

Podobną ocenę rezultatów wyszukiwania za pomocą tagów uzyskał P. Jason Morrison porównując efektywność dostępu do informacji za pomocą kilku folksonomii, wyszukiwarek



i katalogów (Delicious, Furl i Reddit; Google, Microsoft Live i Altavista; Yahoo! Directory i Open Directory Project). Badanie to wykazało, że folksonomie osiągają efektywność wyszukiwania porównywalną z efektywnością katalogów, jakkolwiek na ogół niższą niż wyszukiwarki. Folksonomie cechowała niższa dokładność wyszukiwania niż katalogi i wyszukiwarki, chociaż w przypadku Delicious wyniki były lepsze niż w przypadku Open Directory i równie dobre jak w przypadku wyszukiwarki MS Live. Użytkownicy o małym doświadczeniu w wyszukiwaniu osiągnęli lepsze wyniki używając folksonomii niż wtedy, gdy korzystali z katalogów, ale równie sprawnie wyszukiwali za pomocą wyszukiwarek. Folksonomie wykazywały najmniejszą sprawność w wyszukiwaniu konkretnej witryny. Badania te wykazały też, że wyniki wyszukiwania za pomocą folksonomii, katalogów i wyszukiwarek uzupełniają się, co dowodzi, że zapewniają dostęp do nieco innych zasobów informacyjnych (Morrison, 2007).

Trzeba zwrócić uwagę, że mankamentem chmur tagów jest to, iż nie zapewniają one bezpośredniego dostępu do wszystkich obiektów cyfrowych opisanych w serwisie. Zwykle punktem wyjścia w przeglądaniu chmury jest zestaw tagów w serwisie najpopularniejszych – ich liczba bywa różna, ale na ogół nie przekracza 100, podczas gdy liczba tagów używanych w dużych serwisach może sięgać dziesiątek, a nawet setek tysięcy. Dostęp do tagów mniej popularnych możliwy jest poprzez opisy obiektów (a więc drogą wyszukiwania asocjacyjnego, za pomocą wyrażen użytych w opisie obiektu uznanego za relewantny), a czasami za pomocą dodatkowych narzędzi, takich jak zestawienia tagów związanych z danym, generowane za pomocą algorytmów konkurencji. Trudno jednak ocenić w jakim stopniu narzędzia takie zapewniają korzystanie z mniej popularnych tagów. W badaniach Sinclaira i Cardew-Halla zestaw najbardziej popularnych tagów zawierał 70 wyrażen, co oznaczało, że ponad 50% zasobów ich eksperymentalnego serwisu nie było dostępne poprzez chmurę tagową wyświetlaną na pierwszej stronie. Generalnie można zatem stwierdzić, że znaczna część wartościowych zasobów opisanych mniej popularnymi tagami jest trudna do znalezienia, niejako ukryta za zasobami zaindeksowanymi wyrażeniami najczęściej używanymi. Jest przy tym oczywiste, że wyeksponowana w chmurach tagów popularność pewnych wyrażen nie jest równoznaczna z relewancją opisanych nimi zasobów.

## ZAKOŃCZENIE

Współczesny Internet i narzędzia organizacji dostępu do opublikowanej w nim wiedzy stają się coraz bardziej społeczne. Tendencja ta z pewnością będzie się pogłębiać, bo wraz z dorastaniem kolejnych roczników *digital natives* coraz silniej zmienia się społeczne nastawienie do Internetu i jego rola w komunikacji społecznej. Gwałtowny rozwój systemów zbiorowego tagowania postrzegany bywa jako przejaw zjawiska, które Walter J. Ong określił jako tzw. wtórną

oralność uznając, że kształtujący się dzięki mediom elektronicznym nowy sposób komunikacji więcej ma wspólnego z dawnymi kulturami oralnymi niż z językiem literackim obowiązującym w tradycyjnej komunikacji piśmienniczej (Ong, 1992). Słownictwo gromadzone przez systemy tagowania, deklarowany sprzeciw wobec znormalizowanego nazewnictwa stosowanego w systemach organizacji wiedzy ze słownictwem kontrolowanym taką właśnie wtórną oralność zdaje się potwierdzać. A zatem ta wtórna oralność jako atrybut coraz liczniejszych rzesz użytkowników Internetu musi być brana pod uwagę w projektowaniu metod i narzędzi organizacji cyfrowych treści opublikowanych w globalnej sieci. Stąd też mimo ciągle jeszcze zdystansowanej oceny tych narzędzi przez wielu profesjonalistów informacji, w literaturze przeważa pogląd, że systemy społecznego tagowania są narzędziami ważnymi i użytecznymi (Hammond i in., 2005a i 2005b).

Doceniając znaczenie społecznego klasyfikowania i techniki zbiorowego tagowania trzeba jednak postawić pytanie, czy istotnie słuszna jest głośna teza Claya Shirky'ego, że klasyczne systemy klasyfikowania, kategoryzacje, taksonomie i ontologie są z gruntu błędnym podejściem do organizacji wiedzy w Internecie (Shirky, 2005). Czy rzeczywiście tradycyjne narzędzia i metody organizacji wiedzy są strategiami przeszłości, które nie przystają do specyfiki (wszystkich? niektórych?) sieciowych zasobów informacyjnych i zachowań ich użytkowników? Czy zatem projekt Semantycznego Weba, który uznać można za pewnego typu kontynuację tradycyjnego podejścia jest nietrafioną utopią?

Prowadzone od kilku lat badania nad systemami społecznego klasyfikowania pokazują, że oddanie opracowania metadanych o treści zasobów informacyjnych w ręce użytkowników – jest jednak źródłem pewnych problemów. Z jednej strony niepewna jest wystarczalność wiedzy indeksujących, zwłaszcza gdy mamy na myśli klasyfikowanie nie tylko dla własnych potrzeb, ale też dla potrzeb pewnej społeczności. Dyskusyjne jest też to, kto jest bardziej kompetentnym klasyfikatorem: użytkownicy czy profesjonalni klasyfikatory. Użytkownicy, często też autorzy określonych treści, w niektórych sytuacjach świetnie znają specjalistyczny język (jeśli są to faktycznie specjaliści dziedziny), ale w innych mogą okazać się ignorantami (gdy są nowicjuszami w danej dziedzinie). Z kolei profesjonalni klasyfikatory lepiej znają zawartość całego repozytorium (np. biblioteki cyfrowej, archiwum elektronicznego, systemu *subject gateway*) i lepiej orientują się w potrzebach potencjalnych użytkowników oraz spójności użycia wyrażeń stosowanych do indeksowania.

Odrębną kwestią jest to, w jakich zastosowaniach sprawdzają się systemy społecznego klasyfikowania. Czy faktycznie użytkownicy tak ochoczo kategoryzują zakładki do interesujących ich stron internetowych, fotografii, blogów, artykułów naukowych, które wykorzystują w swej

pracy, równie entuzjastycznie podejść do opisywania np. zbiorów bibliotecznych? Pobieźny ogląd systemów tagowania wdrożonych rok lub dwa lata temu w katalogach trzech wielkich bibliotek akademickich (Royal and University Library w Kopenhadze, Oxford University Library i New York University Library) wskazuje dość wstrzemięźliwą postawę ich użytkowników. W tej sytuacji racjonalniejsze jest traktowanie klasyfikacji społecznościowych jako narzędzia uzupełniającego tradycyjne instrumenty i metody opracowania zasobów, a nie narzędzia zastępującego je. Wbrew tezie Claya Shirky'ego z takim podejściem będziemy mieć do czynienia zapewne także w e-handlu i w serwisach organizacji czy instytucji.

Zbiorowo generowane społeczne metadane tworzą niewątpliwie cenne zbiory słownictwa wyszukiwawczego przede wszystkim ze względu na jego aktualność i rzeczywistą aktywność w komunikacji. Równocześnie jednak są to zbiory chaotyczne. W systemach tagowania stopniowo wprowadza się już narzędzia wyodrębniania w tym chaosie pewnego porządku za pomocą analizy konkurencji i liczby współwystąpień tagów w opisach poszczególnych obiektów. Niektóre serwisy decydują się na wprowadzenie pewnych elementów manualnej kontroli tagów, np. w formie wyodrębniania pierścieni synonimów. Być może w przyszłości, dzięki badaniom takim jak związane z projektem Semantic Web oraz technikom data mining i text mining dostępne będą narzędzia jeszcze silniejsze, które bez ingerencji człowieka pozwolą to bogate słownictwo pełniej i skuteczniej wykorzystywać w organizacji wysokiej jakości dostępu do wielkich cyfrowych zasobów wiedzy.

## BIBLIOGRAFIA

- Boast, R., Bravo, M., Srinivasan, R. (2007). Return to Babel: Emergent diversity, digital resources, and local knowledge. *The Information Society* no. 23 pp. 395-403.
- Guy, M., Tonkin, E. (2006). Tidying-up tags? [online] *D-Lib Magazine* vol. 12 no 1 [14.09.2009]. Dostępny w WWW: <http://www.dlib.org/dlib/january06/guy/01guy.html>
- Golder S.A., Huberman, B.A. (2006). Usage patterns of collaborative tagging systems. *Journal of Information Science* vol. 32 no 4 pp. 198-208.
- Hammond, T., Hannay T., Lund, B., Scott, J. (2005a). Social bookmarking tools (I): a general review. *D-Lib Magazine* vol . 11 no. 4. Dostępny w WWW: <http://www.dlib.org/dlib/april05/hammond/04hammond.html>
- Hammond, T., Hannay T., Lund, B., Scott, J. (2005b). Social bookmarking tools (I): a case study – Connotea [online]. *D-Lib Magazine* vol 11 no 4 [dostęp: 14.09.2009]. Dostępny w WWW: <http://www.dlib.org/dlib/april05/lund/04lund.html>
- Hendry D.G., Jenkins, J.R., McCarthy, J.F. (2006). Collaborative bibliography. *Information Processing and Management* vol. 42 no 3 pp. 805-825.
- Kłopotek, A. (2002). *Inteligentne wyszukiwarki internetowe*. Warszawa: Akademicka Oficyna Wydawnicza Exit
- Morrison, P. J. (2008). Tagging and searching: search retrieval effectiveness of folksonomies on the World Wide Web. *Information Processing and Management* vol. 44 pp. 1562-1579.

- Morville, P. (2005). *Ambient findability*. Sebastopol, CA: O'Reilly Media
- Ong, W.J. (1992). Oralność i piśmienność. Słowo poddane technologii. Lublin: Wydaw. KUL
- Prensky, M. (2001). Digital natives, digital immigrants. Part 1. *On the Horizon* vol. 9 no 5 p. 1, 3-6.
- Rethlefsen, M. L. (2007). Tags Help Make Libraries Del.icio.us. Social bookmarking and tagging boost participation ( online). *Library Journal* 9/15/2007. Dostępny w WWW: <http://www.libraryjournal.com/article/CA6476403.html>
- Roszkowski, M. (2009). *Język informacyjno-wyszukiwawczy jako narzędzie organizacji informacji w dziedzinowych systemach hipertekstowych*. Rozprawa doktorska napisana pod kierunkiem dr hab. W. Babika. Katowice, Uniwersytet Śląski, Wydział Filologiczny 2009 [maszynopis]
- Shirky, C. (2005). *Ontology is overrated: categories, links, and tags* [online]. Clay Shirky's writings about the Internet [dostęp: 15.09.2009]. Dostępny w WWW: [http://www.shirky.com.writings/ontology\\_overrated.html](http://www.shirky.com.writings/ontology_overrated.html)
- Sinclair, J., Cardew-Hall, M. (2008). The folksonomy tag Cloud: when is it useful? *Journal of Information Science* vol. 34 no 1 pp.15-29
- Smith, G. (2008). *Tagging. People-powered metadata for the Social Web*. Berkeley: New Riders.
- Sosińska-Kalata, B. (2002). Struktury klasyfikacyjne w organizacji zasobów informacyjnych Internetu. W: *Multimedialne i sieciowe systemy informacyjne. Materiały konferencyjne pod red. Cz. Daniłowicza*. Wrocław: Oficyna Wydaw. Politechniki Wrocławskiej s. 359-368; dostępny w WWW: <http://www.zsi.pwr.wroc.pl/zsi/missi2002/pdf/s403.pdf>
- Surowiecki, J. (2005). *The wisdom of crowds*. New York: Anchor Books
- Tomaszczyk, J. (2007). Taksonomia jako narzędzie organizacji informacji. *Zagadnienia Informacji Naukowej* nr 1 (89) s. 40-49
- Vander Wal T. (2005). *Folksonomy definition and Wikipedia*[online]. Vanderwal.net [dostęp: 15.09.2009]. Dostępny w WWW: [www.vanderwal.net/random/entrysel.php?blog=1750](http://www.vanderwal.net/random/entrysel.php?blog=1750)
- Vander Wal, T. (2007). A Stale State of Tagging? [online]. Vanderwal.net [dostęp: 15.09.2009]. Dostępny w WWW: <http://vanderwal.net/random/entrysel.php?blog=1945>
- VTLS Introduces Chamo, a New Social OPAC with Drupal Support [online] (2009). VTLS Pressreleases [dostęp: 18.09.2009]. Dostępny w WWW: <http://www.vtls.com/pressrelease/VTLS-Introduces-Chamo,-a-New-Social-OPAC-with-Drupal-Support-62>
- Weinberger, D. (2007). *Everything is miscellaneous. The power of the new digital disorder*. New York: Henry Holt and Co.
- Zollers, A. (2007). *Emerging motivations for tagging: expression, performance, and activism* [online]. WWW2007 [dostęp: 15.09.2009]. Dostępny w WWW: [http://www2007.org/workshops/paper\\_55.pdf](http://www2007.org/workshops/paper_55.pdf)
- Xia Lin, Joan E. Beaudoin, Yen Bui, Kaushal Desai (2006). Exploring Characteristics of Social Classification [online]. In: *Advances in Classification Research, Vol. 17: Proceedings of the 17th ASIS&T SIG/CR Classification Research Workshop (Austin, TX, November 4, 2006)*, ed. Jonathan Furner and Joseph T. Tennis [dostęp: 10 września 2007]. Dostępny w WWW: <http://dlist.sir.arizona.edu/1790/01/lin.pdf>

Warszawa, 21.09.2009 r.