

# I. ROZPRAWY, BADANIA, MATERIAŁY

## TAGOWANIE KONTROLOWANE – OKSYMORON CZY PRZYSZŁOŚĆ JĘZYKÓW INFORMACYJNO- WYSZUKIWAWCZYCH?

Anna Górska  
Biblioteka Europejskiego  
Trybunału Sprawiedliwości  
w Luksemburgu

*Folksonomie, tagowanie kontrolowane, unitermy, tezaury fasetowe, FaceTag*

### Tagi jako narzędzie systemów bibliotecznych

O porównywaniu zasobów sieciowych z bibliotecznymi oraz metod ich przeszukiwania i katalogowania mówi się od momentu digitalizacji zbiorów bibliotecznych. Barbara Tillet<sup>1</sup> w artykule pt. *Library of Congress controlled vocabularies, the Virtual International Authority File, and their Application to the Semantic Web* pisze, powołując się na Tima Bernersa-Lee<sup>2</sup>:

„Biblioteki oraz twórcy Semantic Web mają wspólny cel w określaniu pojęć i bytów oraz ujednocnieniu ich różnych form”<sup>3</sup>.

Pisząc o Web 2.0, wykazuje także analogie między systemem tagów i językami wyszukiwawczymi używanymi przez biblioteki:

„Web 2.0 jest często wykorzystywany do opisu interaktywnych systemów w Internecie – aplikacje, które umożliwiają użytkownikom przystosowanie do ich potrzeb znalezionych obiektów poprzez dodawanie własnych komentarzy, tworzenie własnych „folksonomii” lub tagów – wszystkiego, co nazwalibyśmy

<sup>1</sup> B. Tillet, C. A. Harper: *Library of Congress controlled vocabularies and their Application to the Semantic Web*. „Cataloging and Classification Quarterly” 2007, vol. 43, no. 3/4.

<sup>2</sup> T. Berners-Lee, J. Hendler, O. Lasilla: *The Semantic Web*. „Scientific American” 2001, vol. 284, no. 5, p. 34-43.

<sup>3</sup> Libraries and the developers of the Semantic Web share goals for naming concepts, naming entities, and bringing different forms of these names together (wszystkie tłumaczenia pochodzą od autorki.)

hasłami przedmiotowymi lub punktami dostępu – w celu łatwiejszego znalezienia i większej użyteczności w obrębie własnej przestrzeni<sup>4</sup>.

Opisywanie zawartości zasobów Internetu jest zatem porównywalne do opisywania zawartości zbiorów bibliotecznych, archiwalnych lub muzealnych.

## Co to jest tagowanie?

Wszelkie definicje tagowania sprowadzają się do stwierdzenia, że tagowanie, polega na nadawaniu słów kluczowych obiektom dostępnym w zasobach sieciowych. Cechą charakterystyczną tego procesu jest udział użytkownika, bo to on taguje zbiory i jednocześnie sam z nich korzysta. Wykonuje zatem pracę katalogera, który wydaje się już być niepotrzebnym ogniwem łańcucha tworzenia metadanych.

Tagi mogą określać wszelkie relacje między użytkownikiem a tagowanym obiektem, nie zawsze więc określają temat czy formę dokumentu, mogą określać jego przeznaczenie (print), autoodniesienie (mój/moje, nowe), subiektywną opinię (słabe; kiepskie zdjęcie) etc.<sup>5</sup>

Tagowanie jest procesem, który prowadzi do stworzenia systemu informacyjno-wyszukiwawczego określanego jako folksonomia<sup>6,7</sup>. Vander Wal rozróżnia dwa rodzaje folksonomii – ograniczoną (*narrow folksonomy*) oraz ogólną (*broad folksonomy*)<sup>8</sup>.

Folksonomie ograniczone służą głównie tworzącemu je użytkownikowi. Dostęp do otagowanych dokumentów jest ograniczony do jednej osoby lub niewielkiego grona znajomych, którzy informacje na temat dokumentów oraz sposobów dotarcia do nich wymieniają poza środowiskiem folksonomii. Folksonomie ogólne charakteryzują się szeroką dostępnością i wyszukiwalnością

---

<sup>4</sup> Web 2.0 is often used to describe user interactive systems on the Internet – applications that enable users to customize things they find, adding their own comments, their own „folksonomies”, or tags – that we’d call subject headings or access points – to make material they wish to use more easily findable and usable in their own space.

<sup>5</sup> M. Roszkowski: *Tagowanie, folksonomie i systemy organizacji wiedzy*. 2009 [online]. [dostęp: 20.08.2012]. Dostępny w World Wide Web: <[http://www.slideshare.net/rosz\\_mar/tagowanie-folksonomie-i-systemy-organizacji-wiedzy](http://www.slideshare.net/rosz_mar/tagowanie-folksonomie-i-systemy-organizacji-wiedzy)>. Przykłady własne.

<sup>6</sup> S. Hayman, N. Lothian: *Taxonomy directed folksonomy: Integrating user tagging and controlled vocabularies for Australian education networks (2007) IFLA 73* [online]. [dostęp 26.08.2012]. Dostępny w World Wide Web: <[http://www.ifla.org/IV/ifla73/papers/157-Hayman\\_Lothian-en.pdf](http://www.ifla.org/IV/ifla73/papers/157-Hayman_Lothian-en.pdf)>.

<sup>7</sup> Termin, który utworzył Thomas Vander Wal w 2004 r. podczas dyskusji nad tematowaniem zasobów serwisów Furl, Flickr i Del.icio.us, zob. [www.vanderwal.net/folksonomy.html](http://www.vanderwal.net/folksonomy.html), [online]. [dostęp 15.02.2012]. Wcześniej, w 1996 r., zjawisko klasyfikacji i kategoryzacji zjawisk świata zewnętrznego przez osoby niezajmujące się tym zawodowo zostało nazwane etnoklasyfikacją (*ethnoclasification*) za sprawą Susan Leigh Star (za Spiteri). Termin *folksonomia* razi językoznawców swoją niespójnością, tak samo jak zaproponowana przez Justynę Hofmoki *kumplonomia*.

<sup>8</sup> T. Vander Wal: *Explaining and Showing Broad and Narrow Folksonomy*. [online]. [dostęp 12.05.2012]. Dostępny w World Wide Web: <[http://www.personalinfocloud.com/2005/02/explaining\\_and\\_.html](http://www.personalinfocloud.com/2005/02/explaining_and_.html)>, O.Le Deuff: *Folksonomies. Les usagers indexent le web*. „Bulletin des Bibliothèques de France” 2006, vol. 51, no 4, [online]. [dostęp 12.05.2012]. Dostępny w World Wide Web: <<http://bbf.enssib.fr/consulter/bbf-2006-04-0066-002>>.

dokumentów, także dla osób spoza grupy. Jeśli przy okazji korzystania z dokumentów utworzy się grupa, ma ona charakter wtórny, powstaje dzięki korzystaniu ze stworzonej folksonomii, realizując społeczną funkcję tagowania.

W przypadku folksonomii ograniczonej kontrola słownictwa nie jest tak istotna jak w przypadku folksonomii ogólnej. Rola społeczna folksonomii ograniczonej kreuje nową jakość systemu informacyjno-wyszukiwawczego, do którego klucz posiadają wyłącznie uczestnicy danej społeczności. Dlatego do tworzenia słownictwa oraz posługiwania się nim w procesie wyszukiwania nie są potrzebne uniwersalne reguły. Wszelkie informacje dotyczące procesu wyszukiwania informacji oraz dodawania nowych tagów są wymieniane poza systemem. Jeszcze inną jakość ma folksonomia stworzona na użytek wyłącznie jednego użytkownika, przez niego samego. Może ona stanowić podtyp wąskiej folksonomii. Tagi z autoodniesieniem mają tu rację bytu, na przykład opatrując grupę zdjęć tagiem „Festiwal 2008” tylko osoba tagująca wie, że chodzi o festiwal organowy w Warszawie, na który co roku chodzi, a nie np. o festiwal rockowy w Gdańsku. Tagi w tego typu folksonomii nie potrzebują rozbudowanej hierarchii ani semantyki, ponieważ użytkownik ma w głowie wszystkie powiązania i znaczenia (czyli to, co zazwyczaj zawiera artykuł hasłowy), i z łatwością porusza się po swoim systemie, korzystając z tagów, które sam nadał. Wykorzystanie tej techniki porządkowania danych okazało się bardzo skuteczne w przypadku *personal information management* (PIM)<sup>9</sup>. W takim przypadku powstaje zamknięty obieg informacji, który upodabnia tak stworzony system informacyjno-wyszukiwawczy do niewielkich zbiorów katalogowanych z użyciem klasyfikacji lub słów kluczowych.

## Zalety i wady tagowania

„Folksonomia reprezentuje jednocześnie niektóre najlepsze i najgorsze aspekty organizacji wiedzy”<sup>10</sup>, a ponieważ jakość systemu wyszukiwawczego opiera się na sile semantycznej języka użytego do tworzenia i wyszukiwania metadanych oraz na jakości jego wizualizacji<sup>11</sup>, rozpatrzmy zalety i wady tagowania, biorąc pod uwagę te dwa aspekty.

## Słownictwo

Słownictwo użyte w procesie tagowania jest często niespecjalistyczne, ogólne, kolokwialne i bywa nacechowane emocjonalnie. Jest to główna cecha róż-

---

<sup>9</sup>A. Mathes: *Folksonomies – cooperative classification and communication through shared metadata*. [online]. [dostęp: 20.08.2012]. Dostępny w World Wide Web: <<http://adam-mathes.com/academic/computer-mediated-communication/folksonomies.pdf>>; T. Vander Wal: *Explaining and showing broad and narrow folksonomies*. [online]. [dostęp: 20.08.2012]. Dostępny w World Wide Web: <[www.vanderwal.net/random/entrysel.php?blog=1635](http://www.vanderwal.net/random/entrysel.php?blog=1635)>.

<sup>10</sup>A folksonomy represents simultaneously some of the best and worst in the organisation of information. Ibidem.

<sup>11</sup>Por. M. Halvey, M. T. Keane: *An Assessment of Tags Presentation Techniques*. [online]. [dostęp: 12.05.2012]. Dostępny w World Wide Web: <<http://www2007.org/posters/poster988.pdf>>.

niąca tagowanie od języków systemów informacyjno-wyszukiwawczych o kontrolowanym słownictwie. Tagi sformułowane jako terminy niespecjalistyczne ułatwiają dostęp do dokumentu dla większego grona użytkowników niż specjaliści w danej dziedzinie. Tę samą zaletę niesie ze sobą użycie słów powszechnie używanych lub kolokwialnych, na przykład w kontrolowanym języku haseł przedmiotowych KABA terminem przyjętym jest termin „toksykomanii”, podczas gdy przeciętny użytkownik poszukujący dokumentów na ten temat użyłby terminu „narkomanii”<sup>12</sup> (w serwisie wykop.pl nie znajdziemy tagu „toksykomanii”, za to kilkaset artykułów otagowanych „narkomanii”). Oczywiście można by argumentować, że JHP KABA jest językiem stworzonym na potrzeby bibliotek naukowych, ale należy pamiętać, że z katalogów tych bibliotek korzystają nie tylko naukowcy.

Nadawanie zbyt ogólnych haseł jest problematyczne w sytuacji, gdy system wyszukiwawczy nie przyjmuje operatorów boole'owskich ani nie buduje haseł złożonych. Wtedy hasło ogólne nie przyczynia się do ułatwienia procesu wyszukiwania obiektu nawet dla samego autora takiego hasła. Użytkownicy serwisu Library Thing, tagując książkę pt. *Angels and Demons* Dana Browna, użyli tagów „science” lub „art.”, które nie tylko z punktu widzenia katalogowania przedmiotowego są niepoprawne. Także użytkownikowi niebędącemu zawodowym katalogerem nic takie hasło nie powie o książce. Identyczne tagi otrzymały książki Richarda Dawkinsa, Carla Sagana, czy *The Magic School Bus*, popularnonaukowa seria dla dzieci. W tej sytuacji kryterium spójności tematowania, rozumiane jako nadawanie tych samych haseł dokumentom o pokrewnej zawartości, nie może być spełnione.

Przyczyną wielu wad folksonomii jest słabość semantyczna tagów:

- polisemia i synonimia. W pierwszym przypadku – ten sam termin może być przypisany do kilku różnych obiektów. W drugim – mamy do czynienia z tym samym znaczeniem różnych słów. Noruzi<sup>13</sup> uważa, że synonimia jest większą wadą niż polisemia, ponieważ użytkownik nie ma pewności, że system zwrócił mu w odpowiedzi wszystkie obiekty danego zbioru obejmujące ten sam temat. Irański badacz ukazuje ten problem, posługując się przykładem z serwisu De.li.cio.us, gdzie dla określenia Nowego Jorku użyto następujących terminów: NewYorkCity, New\_York\_City, New-York-City, New.York.City, New-York, NewYork, New.York, NYC, NY. Zdarza się, że terminy, które wizualnie stanowią podkategorie, w kontrolowanym słownictwie byłyby terminami odrzuconymi, na przykład w Library Thing, po wyborze z chmury tagów terminu „ebook”, otwiera się następna chmura tagów, w której znajdujemy na przykład termin „electronic book”, „E-books”. Tag „ebook” został użyty 174 819 razy, tag „electronic book” – 339 razy, a tag „E-books” 2001 razy, przy czym dokumenty tak zatematowane otrzymały takie hasło ze względu na formę, a nie na treść<sup>14</sup>.

<sup>12</sup> Termin „narkomanii” jest terminem odrzuconym do „toksykomanii”. Użycie terminu narkomanii w procesie wyszukiwania generuje fałszywe odpowiedzi, ponieważ w przypadku haseł przedmiotowych rozwiniętych nie ma tropów wyszukiwawczych (rozwiniętych) z odrzuconym terminem narkomanii.

<sup>13</sup> A. Noruzi: *Folksonomies: (Un)controlled Vocabulary?* „Knowledge Organization” 2006, vol. 33, no. 4.

<sup>14</sup> Library Thing. [online]. [dostęp: 21.04.2012]. Dostępny w World Wide Web: <<http://WWW.librarything.com/>>.

Wyszukanie książki na temat książek elektronicznych jest prawie niemożliwe – z technik wyszukiwania użytkownikowi pozostaje przeglądanie (*browsing*). Niewielu użytkownikom przychodzi do głowy, aby dokumentowi nadawać wszystkie lub kilka tagów synonimicznych<sup>15</sup>,

- liczba gramatyczna<sup>16</sup>. Brak zasad powoduje dublowanie terminów lub odwrotnie – gdy istnieje tag tylko w liczbie pojedynczej, osoba używająca podczas wyszukiwania liczby mnogiej może nie znaleźć szukanego obiektu. Jeśli na Library Thing chcemy wyszukać książkę pt. *Stravaganza: City of Stars* za pomocą hasła przedmiotowego „horse”, nie znajdziemy tej pozycji, będzie ona dostępna tylko wtedy, gdy posłużymy się hasłem „horses”,

- terminy złożone. Inny zbiór możemy otrzymać, gdy użyjemy terminu „historia filozofii” i inny w przypadku wyszukiwania z użyciem terminów „filozofia”, „historia”. Niektóre serwisy (Library Thing) używają zestawów tagów odsyłających do obiektów na dany temat. Powiązane terminy do złudzenia przypominają rozwinięte hasła przedmiotowe lub deskryptory złożone w tezaurusach fasetowych:

- Archaeology – Methodology,
- Excavations (Archaeology) – North America,
- United States – History – 1933-1945.

Każdy element takiego hasła prowadzi do innego zestawu dokumentów, wszystkie jednak zawierają przynajmniej jeden z terminów będących składnikiem takiej grupy tagów. Jednak pomimo podobieństwa prezentacji słownictwa do tezaurusów lub JHP mechanizmy wyszukiwawcze są tu zupełnie odmienne.

- zerową wartość semantyczną niosą ze sobą tagi subiektywne i zbyt ogólne, jak na przykład „current affairs”, „current events”, „wishlist”, „new owned”<sup>17</sup>.

Wynikająca z definicji powszechność tagowania rozumiana jako dostępna każdemu użytkownikowi możliwość katalogowania zbiorów niesie ze sobą niebezpieczeństwo błędnego etykietowania, które może innych użytkowników wprowadzić w błąd. Według sformułowania Oliviera Le Deuff’a powstaje „semantyczna wieża Babel”, co prowadzi do uzyskania niesatysfakcjonujących rezultatów wyszukiwania i straty czasu.

W 2009 r. na Uniwersytecie Warszawskim pracownicy Biblioteki Narodowej Paweł Rygiel i Bartosz Włodarczyk przeprowadzili eksperyment polegający na otagowaniu dokumentów ikonograficznych przez trzy grupy: studentów, specjali-

---

<sup>15</sup> W serwisie Library Thing tagi „ebook” oraz „electronic book” otrzymała książka Jamesa Boyle’a *Public Domain: Enclosing the Commons of the Mind*. Tagi „ebooks” i „E-books” otrzymał podręcznik *E-books handboek 2010* Michiela Hatzenboera. Nie ośmielam się wysunąć przypuszczenia, że tagi w tych przypadkach nadał profesjonalista. Brak natomiast w tym serwisie książki, która otrzymałaby wszystkie trzy tagi.

<sup>16</sup> L. Spiteri: *Structure and form of folksonomy tags: the road to the public library catalogue*. [online]. [dostęp: 15.05.2012]. Dostępny w World Wide Web: <[www.ala.org/lita/ital/sites/ala.org.lita.ital/files/content/26/3/spiteri.pdf](http://www.ala.org/lita/ital/sites/ala.org.lita.ital/files/content/26/3/spiteri.pdf)>.

<sup>17</sup> Przykłady za M. Santoro: *Questa sera si cataloga per oggetto*. Tego typu tagi mają sens w przypadku *narrow folksonomy*. Rozważania tego artykułu obejmują jednak wyłącznie *broad folksonomy* jako system, z którego korzysta szerokie grono użytkowników niezwiązanych ze sobą relacjami spoza systemu informacyjno-wyszukiwawczego.

stwów dziedzinowych oraz bibliotekarzy zajmujących się katalogowaniem<sup>18</sup>. Podstawowe pytania, na które szukano odpowiedzi w badaniu, były następujące: jaką wartość informacyjną ma społeczne tagowanie, czy jest ono konkurencją dla tradycyjnych metod indeksowania i czy można zjawisko tagowania postrzegać w kategoriach rozwoju języków informacyjno-wyszukiwawczych. Okazało się, że wiarygodną informację gwarantuje wyłącznie specjalista, chociaż amator także może stworzyć wiarygodny opis, dodając takie elementy, o których nie miał wiedzy specjalista, lub tagi o wyższym stopniu szczegółowości. Niestety tagi nadane przez amatorów na ogół jednak wprowadzają w błąd lub są zbyt ogólne. Na przykład, wśród 27 dokumentów ikonograficznych był barwny drzeworyt japoński z 19 w. Na cztery szczegółowe hasła specjalistyczne przypadło jedno poprawne hasło amatorskie. Na drugim biegunie uplasowały się tagi jak np.: malarstwo azjatyckie, grafika chińska, akwarela. Autorzy badania zaobserwowali, że twórcami projektu „Steve-museum”, że analiza tagów ukazuje zachowania informacyjne użytkowników, co daje materiał do pracy nad doskonaleniem istniejących słowników języków informacyjno-wyszukiwawczych.

W 2006 r. M. Guy i E. Tonkin<sup>19</sup> przeprowadzili badanie na próbcie tagów z serwisów Flickr i de.li.cio.us. W obu procent błędnych tagów plasował się między 30-40% słownictwa. Błędy polegały głównie na tworzeniu niepoprawnych gramatycznie form, niepoprawnej ortografii lub tworzeniu kilkunastu złożonych.

Używanie słownictwa nacechowanego emocjonalnie jest dyskusyjne. W systemach wyszukiwawczych instytucji naukowych lub serwisów o naukowych ambicjach używanie tak sformułowanych tagów jest wadą, pożądane natomiast może być w katalogach stworzonych dla celów usługowo-handlowych<sup>20</sup>.

## Organizacja pracy

Dużą zaletą, która w pewnych sytuacjach może stać się wadą tagowania, jest błyskawiczne tempo wprowadzania neologizmów, nowych pojęć i nazw do systemu. Nie muszą one być weryfikowane na podstawie wiarygodnych źródeł, zatem pojawiają się w środowisku wirtualnym natychmiast po pojawieniu się dokumentu, który mają opisać.

Zaletą tagowania jest także bardzo prosty proces tworzenia tagów, porządkowania zbiorów, utrzymywania systemu. Łączy się to z bardzo niskimi kosztami. Ponieważ nie ma reguł, nie trzeba szkolić katalogerów, a oprogramowanie i bazy do przechowywania dokumentów są powszechnie dostępne w Internecie.

<sup>18</sup> Prezentacja pt. *Opracowanie rzeczowe zbiorów ikonograficznych – hasła przedmiotowe, słowa kluczowe, tagi? Czy folksonomia może wspomóc rozwój jhp?* została przedstawiona na XI Ogólnopolskich Warsztatach *Języki informacyjno-wyszukiwawcze – teraźniejszość i przyszłość*. Warszawa, 3-5 czerwca 2009 r.

<sup>19</sup> M. Guy, E. Tonkin: *Folksonomies: tyding up tags?* „D-Lib Magazine” 2006, vol. 12, no. 1.

<sup>20</sup> O ile nie zachodzi zjawisko tzw. astroturfingu – nieuczciwego marketingu, procederu, w którym koncerny opierają promocję swoich produktów na udawanej kampanii społecznej. Niektórzy badacze – np. O. Le Deuff – widzą w tym wadę i zagrożenie dla społecznych funkcji tagowania. O. Le Deuff: *Folksonomies...*, op. cit.



## Funkcje społeczne

Pożyteczną cechą wykazywaną przez badaczy<sup>21</sup> jest tworzenie grup, społeczności dzielących zainteresowania. Serwisy używające tagów umożliwiają sprawdzenie, który z użytkowników posłużył się takimi samymi terminami oraz jakie obiekty oznaczył w ramach swojej dziedziny zainteresowań.

Piotr Tańkowski, analizując zjawisko folksonomii przez pryzmat „mądrości tłumów”, która przede wszystkim manifestuje się w Internecie, zwraca jednak uwagę, że mądrość ta jest w wysokim stopniu niedoskonała<sup>22</sup>.

Jako bardzo istotną wadę folksonomii wykazuje się niemożność wykorzystania przez użytkownika zdolności poznawczych w procesie wyszukiwania informacji. Oznacza to, że wyszukując za pomocą tagów użytkownik nie ma możliwości poprawienia wyników przyszłych wyszukiwań przez zdobycie doświadczenia i poznania zasad działania systemu, przez uczenie się – ponieważ słownictwo folksonomii nie jest kontrolowane i nie rządzi się żadnymi regułami, nie można się tych reguł nauczyć, czy to przez zapoznanie się z nimi, czy też przez własne spostrzeżenia dotyczące takich zasad<sup>23</sup>.

## Wizualizacja

Kolejnym czynnikiem decydującym o jakości pozyskiwania informacji z katalogów jest forma przedstawienia słownictwa użytkownikowi. W przypadku folksonomii są to przeważnie chmury tagów. Terminy są w nich ustawione w porządku alfabetycznym z graficznym wyróżnieniem najczęściej używanych.

Zalety takiej prezentacji słownictwa pokazują autorzy artykułu pt. *Getting our head in the clouds: Toward evaluation studies of tagclouds*<sup>24</sup>. Wizualizacja w formie chmury tagów uwydatnia również wadę systemu polegającą na braku terminu szczegółowego na pierwszym poziomie prezentacji słownictwa (zwraca na to uwagę A. Noruzi)<sup>25</sup>. Szczegółowość terminów zależy od uznania i stanu wiedzy użytkownika, który formułuje tag. We wspomnianym eksperymencie przeprowadzonym na Uniwersytecie Warszawskim zdjęcie przedstawiające króla Wielkiej Brytanii Edwarda VIII podczas spotkania z politykami niemieckimi Hermannem Georgiem Fiedlerem oraz Ferdinandem Grafem von Zeppelin

---

<sup>21</sup> A. Noruzi: *Folksonomies...*, op. cit.; E. Quintarelli: *Folksonomies: Power to the people*. [online]. [dostęp: 15.05.2012]. Dostępny w World Wide Web: <[www.iskoi.org/doc/folksonomies.htm](http://www.iskoi.org/doc/folksonomies.htm)>; O. Le Deuff: *Folksonomies...*, op. cit.

<sup>22</sup> P. Tańkowski: *Etykietowanie treści w serwisie Podaj.net*. EBIB Elektroniczny Biuletyn Informacyjny Bibliotekarzy. 2010, nr 4 (113). [online]. [dostęp: 14.09.2012]. Dostępny w World Wide Web: <<http://www.nowyebib.info/2010/113/a.php?tafilowski>>; K. Stępień: *Folksonomie, czyli społecznościowe opisywanie treści*. Warszawa 2010.

<sup>23</sup> G. Macgregor, E. McCulloch: *Collaborative tagging as a knowledge organisation and resource discovery tool*. „Library Review” 2006, vol. 55, no. 5.

<sup>24</sup> A. W. Rivadeneira, D. M. Gruen, M. J. Muller, D. R. Millen: *Getting our head in the clouds: Toward evaluation studies of tagclouds*. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems 2007, pp. 995-998; podają za K. Knautz, S. Soubusta, W. G. Stock: *Tag Clusters as Information Retrieval Interfaces*. Proceedings of the 43<sup>rd</sup> Hawaii International Conference on System Sciences 2010.

<sup>25</sup> A. Noruzi: *Folksonomies...*, op. cit.

otrzymało tagi o różnorodnej szczegółowości: Ludzie; mężczyźni; Edward VIII; Fiedler, Hermann Georg; Zeppelin, Ferdinand, Graf von. W serwisie o budowie analogicznej do Library Thing do hasła szczegółowego nie trafi się inaczej, niż wybierając kolejne terminy o coraz większej szczegółowości, na przykład przeglądając chmurę tagów w serwisie Library Thing nie znajdziemy terminu „Bhagavad Gita”, termin ten znajduje się jako tag w kategorii „Yoga”.

Obok chmury tagów w serwisach znajdziemy także indeksy tagów. Indeks pozwala na przeglądanie zawartości zasobu, dzięki czemu użytkownik o niesprecyzowanym przedmiocie poszukiwań może podążać za podpowiedzią w postaci terminów z indeksu<sup>26</sup>. W przypadku słownictwa kontrolowanego terminy wyszukiwawcze stanowią deskryptory, hasła przedmiotowe, symbole klasyfikacyjne itd., w folksonomiach takie terminy to tagi będące tropem informacyjnym (*information scent*), za którym podąża użytkownik, aby znaleźć potrzebną mu informację<sup>27</sup>.

Badania dowiodły, że możliwość przeglądania istniejącego słownictwa jest częściej wykorzystywana niż formułowanie zapytania wyszukiwawczego. Przeglądanie jest szczególnie ważne w przypadku przeszukiwania zbiorów obcojęzycznych<sup>28</sup>, kiedy użytkownikowi nie zawsze jest łatwo sformułować instrukcję wyszukiwawczą.

## Folksonomie a system unitermów

W 1953 r. Mortimer Taube opracował dla NASA tzw. *Uniterm Indexing System* – system informacyjno-wyszukiwawczy posługujący się unitermami. Cechy charakterystyczne systemu unitermów były następujące:

- unitermy posiadały równy status wyszukiwawczy (każdy uniterm mógł być samodzielnym punktem dostępu),
- unitermy nie mogły być terminami złożonymi,
- stanowiły pojedyncze jednostki leksykalne,
- nie tworzyły hierarchii ani klas,
- były sformułowane w języku naturalnym,
- unitermy charakteryzujące dokument były zaczerpnięte z treści dokumentu, a nie z kontrolowanej listy terminów,
- postkoordynacja.

Do roku 1961 system ewoluował. Zmiany obejmowały rezygnację z równego statusu wyszukiwawczego, wprowadzenie hierarchii oraz wskaźników roli i modyfikatorów. Była to prosta droga do stworzenia tezaury<sup>29</sup>.

Tagi, charakteryzujące się większością cech, pierwotnego systemu Taubego, zdają się być unitermami przeniesionymi w środowisko wirtualne. Uni-

<sup>26</sup> Eksperyment Halvey'a i Keane'a wykazał, że użytkownikom najłatwiej jest wyszukiwać terminy w pionowym indeksie alfabetycznym. M. Halvey, M. T. Keane: *An Assessment...*, op. cit.

<sup>27</sup> Por. P. Pirolli, S. K. Cart: *Information Foraging*. [online]. [dostęp: 20.08.2012]. Dostępny w World Wide Web: <<http://act-r.psy.cmu.edu/papers/280/uir-1999-05-pirolli.pdf>>.

<sup>28</sup> J. Sinclair, M. Cardew-Hall: *The folksonomy tag cloud: When is it useful?* „Journal of Information Science” 2008, vol. 34, no. 1.

<sup>29</sup> L. Bielicka: *Metodyka i organizacja opracowania tezaury*. Warszawa 1988.



termy sprawiały te same problemy co tagi: brak hierarchii oraz brak kontroli słownictwa – polisemię i synonimiczność. Złożeń w systemie unitermów się nie używało, w folksonomiach stosuje się je nieskutecznie (brak reguł zapisu terminów złożonych lub niedopuszczalność ich stosowania).

Na podstawie obserwacji tendencji przekształceń systemu unitermów, można zaryzykować prognozę rozwoju folksonomii w kierunku tezaurusów.

## Propozycje

W celu wyeliminowania słabości folksonomii zaczęto mówić o kontroli słownictwa. Pisze o tym Alireza Noruzi<sup>30</sup>:

„(...) dla systemów posługujących się folksonomiami słownictwo kontrolowane jest rzeczą niezbędną dla zapewnienia spójności tagowania w bazie danych i w działalności tagujących”<sup>31</sup>.

Podobny wniosek sformułowała także w swojej książce Patricia Harpring<sup>32</sup>:

„Przy założeniu, że jednym ze wspólnych celów społeczności korzystającej z dziedzictwa kulturowego jest ulepszenie dostępu do wirtualnej wiedzy o sztuce i kulturze materialnej, niezwykle ważne jest słownictwo kontrolowane. Jest ono niezbędne w fazie indeksowania, ponieważ bez niego katalogerzy będą używać różnych terminów w odniesieniu do tej samej osoby, miejsca lub rzeczy, a w procesie wyszukiwania użytkownicy mogą posługiwać się różnymi synonimami lub hasłami związanymi. Ponieważ użytkownicy najczęściej nie są specjalistami, potrzebują wskazówek, bo mogą nie znać poprawnego terminu”.

W ostatnich latach w badaniach nad folksonomiami widać dwie tendencje. Jedna z nich wskazuje folksonomie jako system koegzystujący z istniejącymi systemami organizacji wiedzy i językami informacyjno-wyszukiwawczymi<sup>33</sup>, druga ukazuje drogi rozwoju folksonomii w kierunku tagowania kontrolowanego, które powinno zachować wszystkie zalety tagowania społecznego, ale także zaadaptować najlepsze cechy systemów kontrolowanych (Noruzi, Spiteri, włoski projekt Facetag).

Naukowcy badający zjawisko folksonomii zgodnie twierdzą, że aby tagowanie wspomogło proces wyszukiwania informacji, potrzebna jest kontrola słownictwa i opracowanie choćby ogólnych reguł dla tworzenia folksonomii. Może to polegać na uporządkowaniu słownictwa w klastry<sup>34</sup>, które jednak są

<sup>30</sup> A. Noruzi: *Folksonomies...*, op. cit.

<sup>31</sup> (...) a controlled vocabulary for a folksonomy-based system is essential to ensure tagging consistency across the database and between taggers.

<sup>32</sup> Given that a shared goal of the cultural heritage community is to improve access to virtual arts and material culture information, controlled vocabularies are essential. They are necessary at the indexing phase because without them catalogers will not consistently use the same term to refer to the same person, place or thing. In the retrieval process, various end users may use different synonyms or more generic terms to refer to a given concept. End users are often not specialists and thus need to be guided because they may not know the correct term. P. Harpring: *Introduction to Controlled Vocabularies. Terminology for Art., Architecture and Other Cultural Works*. 2010.

<sup>33</sup> G. MacGregor, E. McCulloch: *Collaborative...*, op. cit.

<sup>34</sup> A. Noruzi: *Folksonomies...*, op. cit.; K. Knautz, S. Soubusta, W. G. Stock: *Tag clusters...*, op. cit.

krytykowane za tworzenie grup terminów o niejasnym pokrewieństwie i znaczeniu. Klasteryzacja nie pozwala także ograniczać znaczenia terminów i kontynuować w ten sposób poszukiwań<sup>35</sup>. Prezentacja słownictwa powinna mieć postać teaurusu lub słownika haseł przedmiotowych<sup>36</sup>.

W Bath przeprowadzono eksperyment polegający na wsparciu procesu tagowania przez klasyfikację dziesiętną<sup>37</sup>. W rezultacie uzyskano bardziej szczegółowy poziom tematowania dokumentów.

Louise Spiteri<sup>38</sup> postuluje rozwiązanie trzech najistotniejszych problemów:

- ustalenie zasad stosowania liczby pojedynczej i mnogiej,
- ustalenie zasad dla tworzenia tagów złożonych (*compound tags*),
- zapewnienie katalogerom-użytkownikom łatwego dostępu do encyklopedycznych i słownikowych źródeł online w celu weryfikacji proponowanego słownictwa, wyeliminowania homonimii, skrótów lub skrótowców.

W celu wyeliminowania synonimii i homonimii chińscy badacze proponują zastosowanie rachunku prawdopodobieństwa i statystyk<sup>39</sup>. W uniknięciu tego problemu pomaga także zastosowanie w systemie kategorii lub faset<sup>40</sup>.

Wydaje się, że z wieloma problemami uporali się twórcy FaceTag<sup>41</sup>. FaceTag jest prototypem narzędzia służącego do społecznego tagowania z uwzględnieniem relacji semantycznych stosowanego słownictwa. Projekt stworzyło trzech włoskich architektów informacji w roku 2006. Rozwiązania dotyczące organizacji słownictwa są zainspirowane projektami Flamenco, Facetious oraz Etsy. Jednym z głównych założeń projektu jest odrzucenie opartej na algorytmach metody grupowania słownictwa w kategorie i umieszczenie go w odpowiednich hierarchiach na rzecz wykonywania tych czynności przez użytkowników. Fasety utworzono na podstawie zaleceń Classification Research Group<sup>42</sup> i pośrednio klasyfikacji Ranganathana. Są to: typ dokumentu, zagadnienia (*themes*), ludzie, cele. Wyszukiwanie można ograniczyć do wybranego języka i daty publikacji. System umożliwi użytkownikowi dwa rodzaje działalności: przeglądanie/wyszukiwanie (*browsing/searching*) i oznaczanie/edytowanie (*bookmarking/editing*). Interfejs edycyjny jest dostępny tylko po stworzeniu konta użytkownika i zalogowaniu. Twórcy FaceTaga rozwiązali wiele problemów folksonomii:

---

<sup>35</sup> E. Quintarelli, A. Resmini, L. Rosati: *Facetag: Integrating Bottom-up and Top-down Classification in Social Tagging System*. „ASIS&T Bulletin” 2007 June/July.

<sup>36</sup> A. Noruzi: *Folksonomies: Why do we need controlled vocabulary?* „Webology” 2007, vol. 4, no. 2.

<sup>37</sup> M. Lykke, A. L. Hoj, L. M. Madsen, K. Golub, D. Tudhope: *Tagging behaviour with support from controlled vocabulary*. Facets of Knowledge Organisation, A tribute to Professor Brian Vickery, ISKO UK biennial conference, 4<sup>th</sup>-5<sup>th</sup> July 2011, London. [online]. [dostęp: 24.08.2012]; Dostępny w World Wide Web: <<http://www.iskouk.org/conf2011/papers/lykke2.pdf>>.

<sup>38</sup> L. Spiteri: *Structure...*, op. cit.

<sup>39</sup> L. Zhang, X. Wu, Y. Yu: *Emergent Semantics from Folksonomies: A Quantitative Study*. „Journal on Data Semantics” 2006, no. 6.

<sup>40</sup> E. Quintarelli: *Facetag...*, op. cit.

<sup>41</sup> Włosi mają bardzo dobre doświadczenia w opracowywaniu sprawnych narzędzi wyszukiwawczych, por. *Nuovo soggettario*.

<sup>42</sup> E. Tonkin [et al.]: *Collaborative and Social Tagging Networks* „Ariadne – Web Magazine for Information Professionals” 2008 [online]. [dostęp: 20.08.2012]. Dostępny w World Wide Web: <[www.ariadne.ac.uk/issue54/tonkin-et-al](http://www.ariadne.ac.uk/issue54/tonkin-et-al)>.

- możliwość jednoczesnego wpisywania terminu w okno wyszukiwawcze interfejsu oraz wybór terminu z fasety zawężają wyszukiwanie analogicznie do systemów stosujących operatory boole'owskie,

- mechanizm podpowiadający wyszukiwane słowo prowadzi użytkownika bezpośrednio do indeksu terminów przyjętych, co eliminuje zerowy rezultat wyszukiwania,

- zwiększenie siły semantycznej tagów poprzez umieszczenie ich w hierarchii,

- system akceptuje tagi złożone,

- wykluczenie polisemii, synonimii oraz problemów z liczbą gramatyczną poprzez wprowadzenie zasad tworzenia nowego słownictwa za pomocą szczegółowego formularza, który wypełnia użytkownik, chcąc wprowadzić nowy tag.

Badania nad użytecznością tagów oraz propozycje kontroli terminów proponowanych przez użytkowników prowadzą badaczy w kierunku społecznego tworzenia tezaurusów fasetowych. Główna różnica między nimi a tradycyjnymi tezaurusami implementowanymi do systemów informacyjno-wyszukiwawczych jest taka, że terminy – deskryptory tworzą użytkownicy, a nie bibliotekarze czy specjaliści od języków informacyjnych. Bardzo szczegółowo zaprojektowane formularze do tworzenia nowych deskryptorów tezaurusa prowadzą użytkownika krok po kroku do dobrania poprawnej formy hasła i umieszczenia go w gnieździe semantycznym w odpowiedniej kategorii i w odpowiednich relacjach wobec innych haseł. Przypomina to nieco formularze dla katalogerów do tworzenia słownictwa w tradycyjnych tezaurusach.

Zaproponowane terminy są weryfikowane przez bibliotekarzy specjalizujących się w językach informacyjnych i dopiero skontrolowane i ewentualnie poprawione trafiają do systemu.

Analogiczny system kontroli słownictwa wprowadzono w LibraryThing for Libraries. Spośród grupy synonimów administratorzy wybierają najczęściej używane i wprowadzają je do systemu jako termin przyjęty wraz z terminami powiązаныmi ustawionymi w odpowiedniej hierarchii. Usuwane są terminy nieangielskie, a na ich miejsce wprowadzane anglojęzyczne. Oczywiście eliminuje się tagi bez znaczenia – subiektywne, z autoodniesieniem, zbyt ogólne<sup>43</sup>.

## Wnioski końcowe

Widać, że obserwowane rozwiązania nie różnią się zbytnio od przyjętych w dużych katalogach. Tendencją, jaką można zaobserwować w próbach „okiełznania” folksonomii, jest tworzenie systemów o cechach tezaurusa fasetowego. Do tego prowadzi porządkowanie słownictwa w fasety, tworzenie klas oraz hierarchii wewnątrz klas. Propozycje tworzenia formularzy ułatwiających dobór i wprowadzanie nowych terminów do systemu są rozwiązaniem identycznym jak procedury tworzenia słownictwa w tezaurusach (por. *Nuovo soggettario* i FaceTag). Ważną zmianą w procesie tworzenia takich systemów jest zastąpienie katalogera użytkownikiem. Łańcuch kataloger – metadane – użytkownik staje się łańcuchem użytkownik – metadane – użytkownik.

<sup>43</sup> Informacja otrzymana od Tima Spaldinga z LibraryThing – korespondencja 2.05.2012.

Zaproponowane systemy mogą się sprawdzić w dużych niespecjalistycznych bibliotekach, których użytkownicy wywodzą się z różnych środowisk i mają różne potrzeby informacyjne. Im biblioteka ma bardziej zróżnicowany zbiór, tym bardziej rośnie wśród użytkowników potrzeba przeglądania, którego jakość jest proporcjonalna do stopnia rozbudowania systemu wyszukiwania rzeczowego.

Można zadać sobie pytanie, czy to jeszcze będzie społeczne tagowanie, czy też całkiem nowy system katalogowania. Niezależnie od nazwy powstała nowa jakość, a od rozwoju technologii IT oraz „mądrości tłumów” i mądrości bibliotekarzy zależy, w jakim kierunku te systemy się rozwiną i w jakim stopniu przyczynią się do wzrostu biblioteczno-informacyjnej wartości naddanej.

## Summary

Cataloguing the Internet resources is being compared to cataloguing library resources. Social tagging is one of the methods of systematizing the Internet contents. The researchers have been deliberating for several years whether this method is useful for libraries, how it's usefulness can be increased, and how to eliminate it's shortcomings, to make it an equal tool in library information-retrieval systems. The author presents systematic review of current research on this topic, and proposals of folksonomy transformation into controlled tagging systems. She also offers a thesis concerning a tendency of folksonomy development into facet thesaurus, comparing tags to uniterms' system.