

## Wybrane problemy porządkowania sieci semantycznych na przykładzie melioracji rekordów bibliograficznych

### 1. Wstęp

Systemy informatyczne są przydatnymi narzędziami do usprawniania działań i procesów, szczególnie dla zadań rutynowych, dobrze ustrukturyzowanych, sformalizowanych i algorytmizowalnych. Komputery po prostu wykonują każdą czynność – zgodnie z zadanym algorytmem – szybciej niż zrobiliby to ludzie. Głównymi **cechami**<sup>3</sup> **algorytmów** są: niezależność od użytych środków i symboli (danych zakodowanych i zdigitalizowanych), proste kroki na każdym etapie, które nie wymagają „mądrych decyzji, subtelnych sądów lub intuicji”<sup>4</sup>, niezawodność. Automatyczne wykonanie instrukcji krok po kroku nie wymaga konieczności „rozumienia” sensu i celu wykonywanych czynności. Poprawność ich działania może być sprawdzana tylko na poziomie składni (syntaktyki) – stąd komputery mogą być opisane jako **silniki syntaktyczne** lub co najwyżej **inferencyjne**<sup>5</sup>. Kłasyk teorii informacji C.E. Shannon podkreśla, że „semantyczne aspekty komunikacji są nieistotne dla problemu inżynierskiego”<sup>6</sup>.

Opis semantyczny danych jest potrzebny np. dla poprawy jakości komunikacji elektronicznej i w tym celu opatrywane są one dodatkowymi atrybutami czyli **metadanymi**, aby poprawić jakość działania, np. wyszukiwarek internetowych. Jednak przetwarzanie metadanych z inżynierskiego punktu widzenia odbywa się nadal na poziomie syntaktyki. Od końca XX w. prowadzone są prace

---

<sup>1</sup> Szkoła Główna Handlowa w Warszawie, Instytut Informatyki i Gospodarki Cyfrowej.

<sup>2</sup> Uniwersytet Łódzki, Katedra Informatyki.

<sup>3</sup> D. Dennett, *Dźwignie wyobraźni i inne narzędzia do myślenia*, Copernicus Center Press, Kraków 2015, s. 195–196.

<sup>4</sup> Ibidem, s. 196.

<sup>5</sup> Ibidem, s. 258–260.

<sup>6</sup> C.E. Shannon, *A Mathematical Theory of Communication*, „The Bell System Technical Journal” 1948, vol. 27, July, October, s. 379–423, 623–656.

zmierzające do stworzenia semantycznego Internetu<sup>7</sup> (ang. *Semantic Web*) lub Web 3.0 i ustalenia standardów opisywania treści w Internecie, które umożliwiłyby automatom (np. autonomicznym agentom, robotom wyszukiwarek) przetwarzanie danych w sposób odpowiedni do ich znaczenia<sup>8</sup>. Pomocne do tego jest oznaczanie (tagowanie) obiektów dla łączenia danych (słów) w sieć semantyczną (znaczeniową). Oznaczanie semantyczne nie zawsze może być adekwatne i przydatne dla potencjalnych użytkowników, stąd wyniki wyszukiwań nadal często nie spełniają ich oczekiwań. Bowiem **znaczenie (sens) ma naturę (inter)subiektywną, kontekstową i czasową**.

Przykładem tworzenia sieci semantycznych jest katalogowanie zasobów bibliotecznych. Rekordy bibliograficzne (tworzone i meliorowane), dotyczące konkretnej pozycji, są wiązane relacjami semantycznymi (w sieć) zgodnie z określonymi regułami i procedurami. Autorzy treści w Internecie rzadko podlegają takim rygorom.

Ponadto języki (zwłaszcza naturalne) ewoluują w czasie: oddolnie, w sposób spontaniczny i niezamierzony bądź odgórnie, według projektów polityczno-ideowych. Zdaniem językoznawcy, prof. J. Bralczyka, słowa nie są logiczne, precyzyjne i spójne, a ich uporządkowanie naukowe jest trudne<sup>9</sup>. Ewolucja języka naturalnego w czasie przejawia się jako **zmiana semantyczna** (ang. *semantic shift/change*) używanych słów.

Głównym celem tego artykułu jest ukazanie **wybranych problemów semantycznych** (w kontekście systemów informacyjnych), **zmian semantycznych w języku oraz roli czynnika ludzkiego przy melioracji (porządkowaniu, aktualizowaniu) rekordów bibliograficznych**. Metodologia zastosowana w pracy opiera się na podejściu społeczno-poznawczym (interdyscyplinarnym, kognitywistycznym) – obejmującym analizę systemową, filozofię (epistemologię), logikę, lingwistykę oraz na obserwacji uczestniczącej w działaniu bibliotek.

---

<sup>7</sup> Internet semantyczny jest rozumiany jako ideał – superuporządkowana hiperbiblioteka. Pozostałą część Internetu porównać można do wirtualnego Hyde Parku (lub wirtualnego pchlego targu), w którym – na zasadach demokratycznych – każdy może mówić (sprzedawać) i czytać (kupować), nie domagając się „certyfikatów jakości” i bez cenzury.

<sup>8</sup> J. Papińska-Kacperek, *Usługi cyfrowe. Perspektywy wdrożenia i akceptacji cyfrowych usług administracji publicznej w Polsce*, Wydawnictwo UŁ, Łódź 2013, s. 45.

<sup>9</sup> J. Bralczyk, *1000 słów*, Prószyński Media & Agora, Warszawa 2017.

## 2. Digitalizacja a semantyka

W procesie digitalizacji, obiekty świata zewnętrznego są kodowane i zapisywane w formie binarnej. Pliki danych wymagają dodatkowo przetwarzania np. redukcji redundancji, integracji, standaryzacji, agregacji, interpretacji. Ponadto dla opatrzenia danych w Internecie **metadanymi**<sup>10</sup>, tworzone są – wzorowane na standardach formatów stosowanych w bibliotekach – opisy semantyczne (w bibliotekach są to np. hasła przedmiotowe, a tu – etykiety, znaczniki, tagi semantyczne). Większość tych działań wymaga zrozumienia, wyboru i decyzji kompetentnych ludzi.

Od czasów Web 2.0 powstają ogromne niekontrolowane zasoby treści cyfrowych. Ważnym elementem wizji sieci semantycznej Web 3.0 Bernersa-Lee<sup>11</sup> jest zaufanie do rzetelności poszukiwanych treści. Nadal nie powstał globalny standard semantycznego oznaczania obiektów – **pytanie: Czy jest to w ogóle możliwe?** Nie wiadomo bowiem na podstawie czyjej i jakiej wiedzy<sup>12</sup> tworzyć semantyczne opisy obiektów. Jak wyznaczać (wybierać?) ekspertów? Kto miałby za to odpowiadać, a kto finansować czynności z tym związane? Wiedza zmienia się np. w wyniku nowych odkryć, zatem relacje w budowanych sieciach semantycznych również są dynamiczne. Należy zaznaczyć, że tzw. wiedza obiegowa (ang. *common knowledge*) nie zawsze jest tożsama z wiedzą naukową<sup>13</sup>. Społecznościowe tagowanie (ang. *folksonomy*) jest stosowane na portalach takich, jak Twitter, Flickr czy Printerest. Koncepcję społecznościowego semantycznego tagowania opisano w wielu pracach badawczych, np. Semidrops<sup>14</sup> lub

---

<sup>10</sup> Np. zapisanych w RDF lub OWL (ang. *Resource Description Framework, Web Ontology Language*), opartych na XML-standardach formatów (Dublin Core, UNIMARC, MARC21, EAD, ONIX, CDWA, METS i in.) stworzonych przy współpracy informatyków z bibliotekarzami – bibliografie i katalogi biblioteczne są też przykładami zbiorów metadanych. Pojęcie metadane jest czasami używane wieloznacznie: jako „dane o danych” – dostarczają opisy takie, jak bibliografie itp., jako „dane ponad danymi” – określają, jak dane mają być interpretowane – opisują miejsca zapisu danych – tak jak atrybuty w bazach danych, lub legendy na mapach i in. Metadane też są danymi i potrzebowałyby opisujących je kolejnych metadanych i w ten sposób *ad infinitum*, dopóki nie zastosujemy wyjaśnienia spoza systemu formalnego (stąd ograniczenia systemów informatycznych jako silników syntaktycznych).

<sup>11</sup> D. Artz, Y. Gil, *A Survey of Trust in Computer Science and the Semantic Web*, „Web Semantics: Science, Services and Agents on the WWW” 2007, 5(2), s. 58–71.

<sup>12</sup> Wiedza – rozumiana jako obraz rzeczywistości – nie zawsze jest obiektywna – co najwyżej intersubiektywna.

<sup>13</sup> Teorie naukowe czasami przeczą sobie. Wybór jednej z nich jest arbitralny i subiektywny.

<sup>14</sup> D. Torres, A. Diaz, H. Skaf-Molli, P. Molli, *Semidrops: A Social Semantic Tagging Approach for Emerging Semantic Data*, 2011 1. 340–347, 10.1109/WI-IAT.2011.51.

Lexitags<sup>15</sup>, ale zwykle opierają się one na jednej aplikacji i nie ma udanego wdrożenia dla całego Internetu. Społecznościowe oznaczanie treści pełni swoją funkcję w jednym konkretnym serwisie, ułatwiając jego użytkownikom odnajdywanie obiektów lub kojarzenie podobnych. W aplikacjach tych autorzy sami, zgodnie ze swoim przekonaniem, wpisują słowa kluczowe, dla niektórych być może irracjonalnie. Może zastanawiać zasadność wyboru takich, a nie innych tagów, hasztagów czy słów kluczowych.

Interpretacja danych (np. słów użytych w tagach) wymaga **wcześniejszej** wiedzy użytkownika, **jak** rozumieć (np. dane słowo lub model danych), oraz jego doświadczenia **jak** taka informacja może być użyta<sup>16</sup>. L. Floridi pisze: „Informacja jest wciąż nieuchwytnym pojęciem”<sup>17</sup>. Niekonsekwencja w interpretacji poszczególnych pojęć przez odbiorcę zależy od stopnia zgodności kulturowej i konceptualnej (używanego języka) oraz od indywidualnej językowej i komunikacyjnej kompetencji użytkowników<sup>18</sup>. R. Koseleck<sup>19</sup> twierdzi, że pojęcie jest czymś więcej niż słowo, z którym jest związane, a znaczenie słowa wskazujące na oznaczone pojęcie, zależne jest od kontekstu (sytuacji społecznej). Wiedza, czyli „znajomość” (niekoniecznie **znaczenia**) słów może być różnie wyrażana, np.:

- Tak, znam to słowo ze słyszenia, ale nie wiem, co to znaczy.
- Tak, znam definicję słownikową (nawet dla różnych dziedzin).
- Tak, znam przykłady dla klasy, nazwanej tym słowem (teoria prototypu)<sup>20</sup>.
- Tak, wiem kiedy, gdzie i jak używać tego słowa.
- Tak, znam kolokacje (z którymi innymi słowami można go użyć).

Tylko to ostatnie stwierdzenie daje się w pewnym stopniu sformalizować i wówczas przedstawienie relacji słów jako sieci nie nastrocza trudności.

Twórcy kodu semantycznego mogą korzystać z różnych zaufanych źródeł, np. z wiedzy naukowej (po rozwiązaniu dylematu, kto określa naukowość nauki), lub ze stworzonych społecznościowo skojarzeń pojęć. W koncepcji wyszukiwarek Web 3.0 dopuszcza się tworzenie kolokwialnych zapytań, zatem wiedza

---

<sup>15</sup> C. Veres, *LexiTags: An Interlingua for the Social Semantic Web*, <https://www.semantic-scholar.org>, 2011.

<sup>16</sup> A.Z. Wassilew, *Cena a wartość w gospodarce cyfrowej – wybrane problemy*, „Ekonomiczne Problemy Usług” 2016, 122, s. 165–176.

<sup>17</sup> L. Floridi, *Open Problems in the Philosophy of Information*, „Metaphilosophy” 2004, 35(4), s. 554–582.

<sup>18</sup> R.T. Bell, *Sociolinguistics: Goals, Approaches and Problems*, Londyn: Batsford 1976, s. 252.

<sup>19</sup> R. Koselleck, *Semantyka historyczna*, Wydawnictwo Poznańskie, Poznań 2001, s. 42.

<sup>20</sup> *Cognition and Categorization*, E. Rosch, B.B. Lloyd (red.), Lawrence Erlbaum Associates, Hillsdale 1978.

zgrupowana w chmurach tagów portali społecznościowych jest użyteczna, ale wymaga weryfikacji. Tu rodzi się kolejne pytanie: Kto miałby to robić? Decyzje, z jakiej wiedzy korzystać, podejmowane w trakcie tworzenia znaczników semantycznych, mogą być czynnikiem ograniczającym czy cenzurującym – tak jak ostrzegął S. Lem iż „możliwe, że nie rozstrzygniemy (...) zagadki świadomości, lecz zdołamy za to imitować inteligencję tak doskonale, że otoczeni legionami (...) imitacji, coraz dokładniej będziemy wchodzić pod ich (...) kuratelę”<sup>21</sup>. Stworzenie uniwersalnego standardu opisu semantycznego jest bardzo trudne (nawet niemożliwe) i z powodu ewolucji języków, w tym zmian znaczenia wyrazów.

### 3. Zmiany semantyczne wyrazów

Obiekty w świecie rzeczywistym są zwykle opisywane w języku naturalnym i tylko w nim można stwierdzić, czy dane zdanie ma sens i czy jest prawdziwe<sup>22</sup>. Wymagana jest dodatkowa wiedza (założenia) o warunkach, w których coś można potwierdzić lub zanegować.

Ciągła ewolucja języka, jako dynamiczny proces<sup>23</sup>, jest zależna od jego użytkowników. Stosowane słowa zmieniają pierwotny sens, pojawiają się nowe zjawiska i koncepcje, a terminy odnoszące się do starych kontekstów często zyskują dodatkowe znaczenia. Nowe słowa są zapożyczane z innych języków, powstają też dialekty i profesjonalne żargony<sup>24</sup>. Wiele zapożyczeń uzyskuje często konotacje, których rodzime słowa nie posiadają. Dynamika zmian języka jest tak duża, że np. zaniechano aktualizacji trzeciej papierowej edycji *Oxford English Dictionary*, bowiem przygotowanie do druku nie nadążało za bieżącymi zmianami w języku angielskim<sup>25</sup>.

Dla każdego języka, prowadzone są niekończące się studia nad ciągłą zmianą semantyczną wyrazów. W bogatej literaturze na ten temat<sup>26</sup> znajdujemy

<sup>21</sup> S. Lem *Okamgnienie*, Wydawnictwo Literackie, Kraków, 2000, s. 103.

<sup>22</sup> A.Z. Wassilew, J. Papińska-Kacperek, *Wybrane problemy semantyczne we współczesnych systemach informatycznych*, „Roczniki KAE” 2017, 44(2), s. 23–33.

<sup>23</sup> A. Blank, *Why do New Meanings Occur? A Cognitive Typology of the Motivations for Lexical Semantic Change*, „Historical Semantics and Cognition” 1999, (13).

<sup>24</sup> E. Sapir, *Język. Wprowadzenie do badań nad mową*, Universitas, Kraków 2010.

<sup>25</sup> J. Gleick, *Informacja. Bit, wszechświat, rewolucja*, Znak, Kraków 2012.

<sup>26</sup> A. Blank, op. cit.; L. Bloomfield, *Language*, Compton Printing Ltd, London & Aylesbury 1935; J. Grzega, M. Schoener, *English and General Historical Lexicology*, Katholische Universität Eichstätt-Ingolstadt, 2007; J.M. Rodd, R. Berriman, M. Landau, T. Lee, C. Ho,

przykłady **zmian semantycznych**, związanych m.in. z przesunięciami w taksonomii<sup>27</sup>, oceną<sup>28</sup>, skojarzeniami<sup>29</sup>, interpretacją<sup>30</sup> itd.

Wydaje się więc, że słowa mogą być używane zgodnie z kontekstem o różnym znaczeniu. Zatem można traktować je jak dane, a najistotniejsze jest znalezienie najważniejszych możliwych relacji łączących je w sieć semantyczną.

#### 4. Melioracja rekordów bibliograficznych

Dla porządkowania Internetu można skorzystać z doświadczeń melioracji rekordów bibliograficznych w bibliotekach. Rekordy bibliograficzne są szczególnym przypadkiem sieci semantycznej (znaczeniowej) – w przypadku konkretnej książki relacje semantyczne łączą autora, tytuł, hasła przedmiotowe, wydawnictwo itp. Do prezentacji i wymiany skatalogowanych i powiązanych rekordów służą różne standardy formatów cyfrowych. Wprowadzono je dla formatowania struktury danych, łatwiejszej do przetwarzania przez komputery. Zawierają pola ze szczegółowymi informacjami o zapisanych danych bibliograficznych lub rekordach referencyjnych – głównie o strukturze drzewa. Dla danego hasła mogą być użyte pojęcia szersze, węższe bądź kojarzeniowe.

---

M.G. Gaskell, M.H. Davis, *Learning New Meanings for Old Words: Effects of Semantic Relatedness*, „Memory & Cognition” 2012, 40(7), s. 1095–1108.

<sup>27</sup> *Zawężanie, specjalizacja* (nowe znaczenie bardziej precyzyjne, np. *słowo* pierwotnie jako *mowa, piwo – napoje*); *poszerzenie, uogólnienie* (przesunięcie w górę w taksonomii, np. *naród polski* dawniej tylko szlachta); *rozgałęzienie* (słowo staje się wieloznaczne, uzyskuje nowe znaczenie, np. *okno, mysz, wirus, sieć*, w kontekście informatycznym).

<sup>28</sup> *Hiperbola* (przesunięcie ze słabszego na silniejsze znaczenie, np. *całe wieki czekam*); *degeneracja* (*pejory, pogorszenie* – nowe znaczenie jest negatywne, np. *czerstwy, dostateczny, spekulant, śmieć*); *elewacja* (nowe znaczenie dodaje ładunek dodatni, np. *pasja* dawniej *gniew*).

<sup>29</sup> *Pospoliczenie nazw marek* (*genericization* – nazwy marek zastępują przedmiot, np. *walkman, xerox*); *metafora* (oparta na semantycznym podobieństwie rzeczy, np. *stopka* strony); *metonimia* (oparta na bliskości między pojęciami, np. *słucham Bacha*); *antonomazja* (frazja lub epitet zastępuje własne imię, np. *Elvis Presley jako Król Rocka*); *eponim* (nazwa własna używana dla idei „zwykłego” słowa, np. *bikini*); *synekdocha* (oparta na relacji całość-część np.: *Manchester* jako drużyna piłkarska); *antyfrazja* (oparta na kontrastowym aspekcie pojęć, np. *lüksowa panienka* jako „zawód”); *auto-antonymy* (przesunięcie znaczenia i koncepcji słowa na komplementarne przeciwieństwo np.: *strasznie* w sensie *dobrze*); *analogia* (podobieństwo synonimiczne, np. *łapać* jako *rozumienie*).

<sup>30</sup> *Nowomowa* (w celu ograniczenia wolności myśli, pełne zagadnienia zredukowane do prostych terminów o uproszczonym znaczeniu, np. *bojownik o wolność*); *dyskurs* (implikuje różne konotacje, np.: *wojna z terroryzmem*), w konsekwencji powoduje niejednoznaczność używanych słów, np. *sprawiedliwość dziejowa*.

Wypełnienie pól rekordów bibliograficznych jest wykonywane głównie przez bibliotekarzy i jest regulowane przez normy spoza systemu informatycznego. Obiekty (pojęcia) rzeczywistości są zakodowane w formie wygodnej zarówno dla bibliotekarzy, jak i dla użytkowników bibliotek.

Do wypełniania pól rekordów wzorcowych wykorzystywane są wszelkie dostępne źródła informacji: encyklopedie (także Wikipedia), słowniki, tezaury, bazy danych bibliotek narodowych, specjalistyczne publikacje, wyszukiwarki i strony internetowe, informatorzy, publikacje promocyjne, korespondencja itp. Czasami jednak pojawia się problem zaufania do źródeł informacji<sup>31</sup>.

Wiele zarejestrowanych danych wymaga aktualizacji lub poprawienia – nie zawsze jest to możliwe do wykonania automatycznie przez system informatyczny. Istnieją ustanowione zasady i procedury postępowania w wymienionych poniżej sytuacjach, które wymagają interwencji doświadczonych bibliotekarzy, aby wybrać pomiędzy proponowanymi dostępnymi opcjami<sup>32</sup>. Ważnymi czynnikami powodzenia jest erudycja, intuicja, wystarczająco dużo czasu na poszukiwania<sup>33</sup>.

Melioracje rekordów bibliograficznych wykonywane są, gdy zachodzi taka potrzeba. Wiele danych realnego świata zmienia się w czasie i nie chodzi tylko o zmianę semantyczną języka. Podstawą do katalogowania jest istniejąca publikacja – fizyczna lub cyfrowa. Głównym zadaniem jest jednoznaczne opisanie każdego rekordu. Czasami potrzebne są dodatkowe informacje, które nie zawsze są łatwo dostępne – często wymagają lat doświadczeń i szerokiej wiedzy. Najbardziej przydatne informacje dla użytkowników to autor, tytuł i temat publikacji. Hasła przedmiotowe mogą być takie same, jak słowa kluczowe publikacji, ale czasami są też opisane poprzez słowa, które nie pojawiają się w dokumencie.

Ważną kwestią jest ustalenie imienia i nazwiska autora. Nie ma problemu, gdy jeden autor jest powiązany tylko z jednym tytułem. Jednak w bazie bibliotecznej mogą istnieć inne formy imienia, a tylko jedna z nich musi być wybrana

---

<sup>31</sup> W katalogu znajdują się dwie tłumaczki literatury dziecięcej: *Morkowiczowa Janina (1875–1960)* i *Bernsteinowa Rozalia* (pseudonim). Są dwie sprzeczne opinie: według historyka literatury dr Michała Fijałkowskiego jest to ta sama osoba, a według dr hab. Piotra Oczko UJ – nie, <http://dom-echa.blogspot.com/2016/09/kim-bya-rozalia-bernsteinowa.html>; <http://pokrewne-dusze-maud.blogspot.com/2016/08/czy-wiemy-cos-o-rozaliibernsteinowej.html>

<sup>32</sup> W rekordzie *Hertz, Zofia* znajduje się rok urodzenia 1911 – używała go przez całe życie. Według przedwojennych dokumentów jej rok urodzenia to 1910. Dylemat, który z nich wybrać do danych bibliograficznych. Pierwsza data jest bardziej rozpowszechniona w publicznym dyskursie, choć nie jest prawdziwa.

<sup>33</sup> Autor książki z 1925 r. podpisał się inicjałami *R.G.* i tak został skatalogowany. Teraz dzięki Internetowi łatwiej było zidentyfikować, że autorem jest *Roman Grus*. Choć był leśnikiem, napisał *Język polski: (jego bogactwo, piękność i siła tajemnicza)*.

jako główna. Czasami sam autor podpisuje się różnymi pseudonimami, w niektórych przypadkach tworzone są różne rekordy referencyjne dla tej samej osoby. Wątpliwość może pojawić się m.in. w związku z transkrypcją i/lub transliteracją oryginalnego imienia zapisanych głównie alfabetami niełacińskimi<sup>34</sup>, a w niektórych przypadkach przy spolszczeniu imienia<sup>35</sup>.

Kolejnym powodem do melioracji są zmiany nazwisk kobiet wychodzących za mąż<sup>36</sup> lub imiona ludzi po zmianie płci<sup>37</sup>. Związane z daną osobą dane personalne<sup>38</sup> nie są stałe przez całe życie i czasem trzeba zmodyfikować odpowiednie pola.

Na sprawdzonych i uporządkowanych danych powinno opierać się tworzenie bibliografii w artykułach naukowych. Niezależnie od stosowanych standardów cytowania (oksfordzki czy harwardzki) brak prawidłowych danych o cytowanej publikacji lub błędne ich wykorzystanie tworzy problemy np. w zbieraniu liczby cytowań czy obliczaniu indeksu Hirscha<sup>39</sup>. Kolejne osoby cytujące daną publikację często powielają błędne opisy, co skutkuje wielokrotnym użyciem np. nieprawidłowego brzmienia nazwiska lub kolejności imion<sup>40</sup>. Często w przypadku długiej listy słów tworzących dane personalne pojawia się wątpliwość, które wyrazy są imieniem, a które nazwiskiem. Różne opisy bibliograficzne

---

<sup>34</sup> Niełatwo określić czy Александър, Aleksander, Alexander, Aleksandr, Василев, Wasilew, Wasylew, Wassilew, Vasilev, Vassilev itd. odnoszą się do tej samej osoby. Ten sam problem może się pojawić nawet przy identycznej pisowni.

<sup>35</sup> W bazie pojawiła się niepotrzebnie pozycja *Hofmann, Franciszek (1845-1897)* – początkowo nie było oczywiste, że nie jest Polakiem. Ustalono, że był austriackim prawnikiem: *Hofmann, Franz (1845-1897)*. Do identyfikacji jego tożsamości pomogły tytuły jego publikacji.

<sup>36</sup> W katalogu znajduje się *Wiącek, Alicja* jako autorka kilku książek. Pierwsza wątpliwość: czy autorem wszystkich jest ta sama osoba. Publikacje dzieli duży odstęp czasu (40 lat) i tematyka. Podczas badania w źródłach okazało się, że to imię i nazwisko odnosi się do trzech osób, z których dwie zmieniły w międzyczasie stan cywilny i nazwiska. W rezultacie powstały hasła: *Wiącek, Alicja (przewodnik turystyczny)* => *Łukowska, Alicja, Wiącek, Alicja (język podreęczniki)* => *Jankowiak, Alicja, Wiącek, Alicja (historia literatury)* => *Wiącek, Alicja*.

<sup>37</sup> Po zmianie płci przez braci Wachowskich zmieniono imiona w katalogu: *Wachowski, Larry* => *Wachowski, Lana* oraz *Wachowski, Andy* => *Wachowski, Lilly*.

<sup>38</sup> W katalogu istniała forma *Middleton, Catherine Elizabeth (1982-)*. Po ślubie i wejściu do rodziny królewskiej rekord został zmieniony na: *Katarzyna (księżna Cambridge; 1982-)*. W ten sam sposób rekord dla *Markle, Meghan (1981-)* po ślubie z księciem Harrym został zmodyfikowany na *Meghan (księżna Sussex; 1981-)*. W takich przypadkach imiona są spolszczane (jeśli to możliwe), a w nawiasie podaje się tytuł honorowy.

<sup>39</sup> Np. autorka tego artykułu wg Google Scholar co najmniej 3 razy w drukowanych publikacjach cytowana jest jako Papińska-Kacperek. Pierwsza osoba nie zauważyła, że edytor Word przekształcił nazwisko, a kolejne albo popełniły ten sam błąd, albo powielają pierwszy opis.

<sup>40</sup> Np. naukowiec z Hongkongu, dr Tien Ee Dominic Yeo poprawnie cytowany jest jako Yeo T.E.D., ale autorzy cytują go też jako Yeo, D T E, a nawet jako T e Dominic Yeo.



mogą przysparzać problemów redaktorom, wydawcom czy wreszcie bibliotekarzom, którzy w końcu spróbują odkryć, jak naprawdę nazywa się autor oraz jak powinno się zapisać jego dane w rekordach bibliograficznych.

## 5. Podsumowanie i wnioski

Systemy komputerowe, choć bardzo przydatne w rutynowych i algorytmizowalnych zadaniach, operują jednak na danych w systemach formalnych, tj. tylko na poziomie składni, i z tego powodu nie zawsze sprawdzają się, gdy zadanie dotyczy znaczenia i zrozumienia. Na chwilę obecną do rozwiązywania problemów semantycznych nadal potrzebni są ludzie.

Ewolucja języków naturalnych jest procesem ciągłym, który zachodzi podczas używania ich w komunikacji między ludźmi. Jego dynamika przejawia się adaptacją nowych słów, tworzeniem nowych terminów i pojęć czy zmianą ich sensu.

W folksonomii dla niektórych stron lub aplikacji internetowych ważną rolę w tworzeniu chmur tagów, a tym samym sieci semantycznej dla ich baz wiedzy, odgrywają użytkownicy. Jednak pojawienie się zbyt wielu tagów może tylko zwiększyć jeszcze bardziej niejednoznaczność pojęć.

Melioracja rekordów bibliograficznych w katalogach bibliotek jest szczególnym przypadkiem sposobu, w jaki dane (słowa) są łączone w ogólną sieć semantyczną. Bazy danych w bibliotekach tworzone są ręcznie lub poprzez digitalizację danych analogowych. Ponadto dla bibliotekarzy i użytkowników konieczne jest dalsze przetwarzanie rekordów, które jest wykonywane przez pracowników zdolnych do „mądrych decyzji i subtelnych osądów”. Wszystkie te zadania wymagają ludzkiego kapitału: erudycji, doświadczenia, inwencji, intuicji oraz często głębszych poszukiwań, badań i trafu.

Systemy informatyczne dla ich skuteczności potrzebowałyby dobrze ustrukturalizowanych i sformalizowanych danych, co wymagałoby standaryzacji i aktualizacji źródeł, wprowadzenia jednego wspólnego języka i pisowni oraz przyjęcie jednej normy transliteracji, co byłoby możliwe w jakimś hipotetycznym systemie totalitarnym. Sztuczna inteligencja i uczenie maszynowe mogą mieć zastosowanie, gdy będą w stanie same rozwiązać powyższe przypadki. Na tym etapie można tylko polegać na melioracji rekordów bibliograficznych przez ludzi.

## Bibliografia

- Artz D., Gil Y., *A Survey of Trust in Computer Science and the Semantic Web*, „Web Semantics: Science, Services and Agents on the WWW” 2007, 5(2), s. 58–71.
- Bell R.T., *Sociolinguistics: Goals, Approaches, and Problems*, Batsford, Londyn 1976, s. 252.
- Blank A., *Why do New Meanings Occur? A Cognitive Typology of the Motivations for Lexical Semantic Change*, „Historical Semantics and Cognition” 1999, (13).
- Bloomfield L., *Language*, Compton Printing Ltd., London & Aylesbury 1935.
- Bralczyk J., *1000 słów*, Prószyński Media & Agora, Warszawa 2017.
- Cognition and Categorization*, E. Rosch, B.B. Lloyd (red.), Lawrence Erlbaum Associates, Hillsdale 1978.
- Dennett D., *Dźwignie wyobraźni i inne narzędzia do myślenia*, Copernicus Center Press, Kraków 2015.
- Floridi L., *Open Problems in the Philosophy of Information*, „Metaphilosophy” 2004, 35(4), s. 554–582.
- Gleick J., *Informacja. Bit, wszechświat, rewolucja*, Znak, Kraków 2012.
- Grzega J., Schoener M., *English and General Historical Lexicology*, Katholische Universität of Eichstätt-Ingolstadt, 2007.
- Koselleck R., *Semantyka historyczna*, Wydawnictwo Poznańskie, Poznań 2001.
- Lem S., *Okamgnienie*, Wydawnictwo Literackie, Kraków 2000.
- Papińska-Kacperek J., *Usługi cyfrowe. Perspektywy wdrożenia i akceptacji cyfrowych usług administracji publicznej w Polsce*, Wydawnictwo UŁ, Łódź 2013.
- Rodd J.M., Berriman R., Landau M., Lee T., Ho C., Gaskell M.G., Davis M.H., *Learning New Meanings for Old Words: Effects of Semantic Relatedness*, „Memory & Cognition” 2012, 40(7), s. 1095–1108.
- Sapir E., *Język. Wprowadzenie do badań nad mową*, Universitas, Kraków 2010.
- Shannon C.E., *A Mathematical Theory of Communication*, „The Bell System Technical Journal” 1948, vol. 27, July, October, s. 379–423, 623–656.
- Torres D., Diaz A., Skaf-Molli H., Molli P., *Semdrops: A Social Semantic Tagging Approach for Emerging Semantic Data*, IEEE/WIC/ACM International Conference on Web Intelligence 2011, 1, s. 340–347, 10.1109/WI-IAT.2011.51.
- Veres C., *LexiTags: an Interlingua for the Social Semantic Web*, <https://www.semanticscholar.org>, 2011.
- Wassilew A.Z., *Cena a wartość w gospodarce cyfrowej – wybrane problemy*, „Ekonomiczne Problemy Usług” 2016, nr 122, s. 165–176.
- Wassilew A.Z., Papińska-Kacperek J., *Wybrane problemy semantyczne we współczesnych systemach informatycznych*, „Roczniki KAE” 2017, 44(2), s. 23–33.

## Źródła sieciowe

<http://dom-echa.blogspot.com/2016/09/kim-bya-rozalia-bernsteinowa.html> (dostęp: 20.07.2018).

<http://pokrewne-dusze-maud.blogspot.com/2016/08/czy-wiemy-cos-o-rozaliibernsteinowej.html> (dostęp: 20.07.2018).

\* \* \*

## Selected problems of organizing semantic networks on the example of the amelioration of bibliographical records

### Abstract

Information systems operate on digital data in formal systems. Like a syntax machine, they are able to analyse the “correctness” only on the syntactic level. The analysis of the semantics or meaning requires the intervention of human factors. The meaning of the words (or constructed sentences) is strictly related to the used natural language. Throughout the history of the dynamic world, the languages have evolved and the words have changed their meaning – it is called the semantic change (shift). The tagging in the semantic Internet is based on a library’s experience. A book’s data (the author, the title, the keywords and so on) linked by semantic relationships in the bibliographical records are a particular case of a semantic network. The bibliographical records require an amelioration (better ordering, refining) by librarians. This is illustrated in a few cases taken from practice. The main goal of this article is to consider the general semantic problems, semantic changes (shifts) in the languages, and the role of human factors for the amelioration of the bibliographical records.

*Translated by Aleksander Z. Wassilew*

**Keywords:** language evolution, semantic changes/shifts, bibliographical records, amelioration of data