

NKOS Workshop 2022 (6–7 października 2022, zdalnie)

Tegoroczne warsztaty Networked Knowledge Organization Systems (NKOS), które odbyły się w dniach 6–7 października 2022 r. w ramach 20. edycji International Conference on Dublin Core and Metadata Applications, miały charakter zdalny. Współprzewodniczącymi komitetu programowego byli: Joseph Busch (Taxonomy Strategies), Douglas Tudhope (University of South Wales) oraz Marcia Zeng (Kent State University), znani badacze i praktycy od wielu lat zajmujący się systemami organizacji wiedzy (SOW). Podczas dwudniowego spotkania odbyło się sześć sesji, w tym jeden panel dyskusyjny, składający się z trzech krótkich wystąpień, oraz pięć sesji, w ramach których wygłoszono dziewięć referatów. Po zakończeniu każdej z nich odbyła się dyskusja dotycząca przedstawionych prezentacji.

Uczestników warsztatów przywitał Joseph Busch, który zaprezentował podstawowe informacje na temat sieci NKOS oraz warsztatów, podkreślając duże zainteresowanie tegorocznym spotkaniem. Pierwszym elementem programu był panel dyskusyjny na temat porównywania klasyfikacji. Wzięli w nim udział Joseph Tennis (University of Washington), Mark Butler (Voise Inc.; University of California, Berkeley) oraz Aida Slavic (Universal Decimal Classification Consortium). Prelegenci zostali dobrani jako reprezentujący różne grupy zainteresowane klasyfikacjami: teoretyków, użytkowników oraz twórców. Wstępem do dyskusji były krótkie prezentacje. W swoim wystąpieniu Tennis podkreślił wagę znajomości i zestawiania różnych klasyfikacji dla poszerzenia wiedzy na ich temat. Porównanie takie może, zdaniem Tennisa, mieć charakter diachroniczny w przypadku analizy jednej klasyfikacji, synchroniczny, jeżeli dotyczy kilku narzędzi lub może dotyczyć zmiany przeznaczenia klasyfikacji. Te trzy rodzaje analizy nie wykluczają się wzajemnie i mogą być prowadzone równolegle. W następnej prezentacji Butler zastanawiał się nad możliwością identyfikacji zbioru uniwersalnych cech, które umożliwiają porównanie klas, ich układu oraz sposobu stosowania klasyfikacji z perspektywy osoby wykorzystującej schematy do automatycznego przydzielania metadanych. Butler podkreślił, że jeden ze sposobów rozwiązania tego problemu może pochodzić z obszaru sztucznej inteligencji, szczególnie przetwarzania języka naturalnego, gdzie tworzone są schematy opisujące atrybuty modeli predykcyjnych (ang. *model cards*) i zbiorów danych (ang. *data cards*). Ich celem jest pomoc w zrozumieniu zalet i wad danego rozwiązania. Uzupełnieniem tych przemyśleń była ostatnia prezentacja – Slavic, ukazująca perspektywę twórcy dużej klasyfikacji uniwersalnej na podstawie zmian w symbolach pomocniczych Uniwersalnej Klasyfikacji Dziesiętnej.

Różnorodne zastosowania tych symboli pociągają za sobą konieczność ostrożnej modyfikacji tablicy ze względu na praktyczne konsekwencje dla użytkowników stosujących schemat do opisywania swoich zbiorów.

Dyskusja, do której punktem wyjścia były przedstawione wystąpienia, dotyczyła m.in. znaczenia mapowania zmian w klasyfikacji. Prelegenci przedstawili odmienne punkty widzenia zależne od sposobu wykorzystywania klasyfikacji. Slavic podkreśliła, że jest ono istotne, jednak podstawowym problemem w bibliotekach pozostaje reorganizacja fizycznego uporządkowania dokumentów zgodnie z nową wersją schematu. W zastosowaniach wykorzystywanych przez Marka Butlera ta kwestia nie występuje, a odwzorowania mogą być wykorzystywane do zmiany sposobu grupowania elementów. Z kolei dla Tennisa mapowanie pokazuje historyczne zmiany w sposobie patrzenia na świat. Należy jednak pamiętać, że jest to tylko jedna z wielu perspektyw przyglądania się przeszłości. Inny problem poruszony w dyskusji dotyczył zastępowania tradycyjnych klasyfikacji w środowisku cyfrowym przez automatyczne klasyfikowanie tekstów. Prelegenci zgodzili się, że nie mamy w tym przypadku do czynienia z alternatywą, lecz raczej uzupełnieniem dotychczasowych możliwości grupowania tekstów.

W sesji drugiej zaplanowano wystąpienie autorstwa Patricii Harpring (Getty Research Institute), dotyczące równości i włączania marginalizowanych dotychczas zagadnień do słownictwa kontrolowanego tworzonego przez Getty Research Institute (GRI). Kwestia zachowania neutralności jest istotna w przypadku słowników Getty'ego m.in. ze względu na chęć poszerzenia zasięgu odbiorców narzędzi, którzy obecnie pochodzą głównie ze Stanów Zjednoczonych i Europy. Równość kulturowo-językowa jest wspierana na różne sposoby, np. przez przydział identyfikatorów nie tylko do pojęć, ale również do terminów w różnych językach, dzięki czemu można odwołać się indywidualnie do każdego z nich. W GRI prowadzone są liczne programy mające na celu zachowanie neutralności i objęcie nowych obszarów, nieuwzględnianych dotychczas w wystarczającym stopniu. Autorka przedstawiła szereg przykładów problemów związanych z inkluzywnością projektowanych narzędzi opisu zbiorów, wskazując na różnice w stosunku do rozwiązań przyjętych w innych zbiorach słownictwa kontrolowanego. Na przykład termin *illegal aliens* został zastąpiony w Art & Architecture Thesaurus (AAT) terminem *undocumented immigrants* i oznaczony w słowniku jako pejoratywny. Ten sam termin został zastąpiony w Library of Congress Subject Headings nazwą *Noncitizens*, która występuje w AAT jako etykieta osobnego pojęcia. Również w strukturze hierarchicznej można znaleźć terminy wyodrębnione według różnych kryteriów w celu zapewnienia równoważności podejścia do różnych kultur, w tym uwzględnienie punktu widzenia społeczności rdzennych (np. *granite (rock)* obok *pounamu* w AAT). W przywoływanym słowniku wprowadzane są także różne terminy i oznaczenia związane z płcią kulturową (ang. *gender*). Podobne przykłady znajdują się w innych zbiorach słownictwa GRI. Referat dostarczył ciekawego

wglądu w najnowsze zmiany w ważnych i mogących uchodzić w wielu obszarach za wzorcowe narzędziach GRI. Wydaje się, że istotne pytanie dotyczy zależności między modyfikacjami wprowadzanymi w etykietach językowych a codzienną praktyką osób posługujących się tymi językami.

Sesja trzecia rozpoczęła się od wystąpienia Marcii Zeng (Kent State University) i Yi Hong (MedeAnalytics, Inc.) na temat 11. wydania International Classification of Diseases (ICD – Międzynarodowej Statystycznej Klasyfikacji Chorób i Problemów Zdrowotnych). Wystąpienie zostało oparte na artykule zamieszczonym w *ISKO Encyclopedia of Knowledge Organization* (Hong & Zeng, 2022). Prelegentki przedstawiły podstawowe informacje na temat klasyfikacji zapewniającej interoperacyjność danych zdrowotnych zarówno na poziomie krajowym, jak i przede wszystkim międzynarodowym. Poruszyły również zagadnienie dodatkowych narzędzi usprawniających pracę z ICD. Podstawowe wyzwania dotyczą zakresu klasyfikacji i wprowadzanych do niego zmian, które pociągają za sobą m.in. konieczność rewizji całej struktury. Ponadto schemat musi być wynikiem uzgodnień w ramach szerokiej grupy użytkowników. Autorki przedstawiły znaczne zmiany w treści, strukturze, notacji i identyfikacji pojęć, które zostały wprowadzone w ostatnim wydaniu schematu, np. wprowadzenie usług z zakresu medycyny tradycyjnej, klas reprezentujących różne warianty SARS-CoV-2, czy też identyfikatorów URI dla wszystkich uwzględnionych jednostek. Jak podkreśliły autorki, 11. wydanie ICD jest dobrym przykładem wykorzystania najnowszych technologii cyfrowych i zasad organizacji wiedzy (OW).

Kolejna prezentacja w sesji trzeciej, autorstwa Marcina Trzmielewskiego (Université Paul-Valéry) i Claudio Gnolego (Università di Pavia), dotyczyła reprezentacji obszaru dotyczącego alergii w Integrative Levels Classification (ILC). W związku z brakiem SOW dotyczącego tej tematyki, Trzmielewski we współpracy ze szpitalem uniwersyteckim w Montpellier stworzył ontologię dotyczącą uczuleń, której podstawą była analiza potrzeb alergologów. Prelegenci podjęli próbę włączenia indukcyjnie stworzonego układu faset do uniwersalnej klasyfikacji o charakterze dedukcyjnym. Zdaniem autorów przeprowadzona analiza pokazała, że podejście indukcyjne może być przydatne podczas tworzenia uniwersalnych klasyfikacji fasetowych takich jak ILC, a schemat może zostać wzbogacony na jego podstawie o dodatkowe fasety.

Podczas dyskusji podsumowującej sesję trzecią poruszono temat szybkości wdrożenia nowej wersji ICD, w której wprowadzono liczne modyfikacje w stosunku do poprzedniej edycji. Hong podkreśliła, że obecnie nadal spotyka się w Stanach Zjednoczonych instytucje, które stosują 9. wydanie schematu, więc proces ten potrwa zapewne jeszcze wiele lat. Pokazuje to konieczność zachowania ostrożności podczas wprowadzania zmian do klasyfikacji profesjonalnych używanych na co dzień w praktyce zawodowej. Jedno z pytań dotyczyło także SOW wykorzystanych podczas projektowania ontologii z zakresu alergii. W związku z tym, że jest to

dopiero początek projektu postanowiono oprzeć się na razie wyłącznie na wiedzy alergologów, jednak przewidziano również wykorzystanie innych źródeł, takich jak inne klasyfikacje, np. ICD. Opisana pokrótce dyskusja zakończyła pierwszy dzień warsztatów.

Pierwszy referat w sesji czwartej, autorstwa Eileen Chen, Julii Bullard i Deana Giustini (University of British Columbia), dotyczył porównania wyników automatycznego i manualnego indeksowania w National Library of Medicine (NLM) w Stanach Zjednoczonych. Podstawowym narzędziem stosowanym do automatycznego indeksowania zbiorów w tej instytucji jest Medical Text Indexer (MTI). Próba badawcza obejmowała 20 artykułów (10 z wysokim Journal Impact Factor i 10 z niskim), a do automatycznego indeksowania użyto publicznie dostępnej wersji MTI, nieużywanej jednak w bibliotece od 2021 r. Pomiedzy opisami generowanymi automatycznie i stworzonymi manualnie występują znaczące różnice dotyczące liczby przyporządkowanych haseł oraz stopnia określania relewancji terminów. Zauważono także m.in., że algorytmy MTI mają tendencję do uwzględniania wyżej na liście proponowanych terminów indeksowych wyrażen związanych z mężczyznami. Uczestnicy spotkania zgodzili się, że potrzebne są dalsze pogłębione badania tego zagadnienia, które jest istotne nie tylko w kontekście indeksowania literatury biomedycznej. W dyskusji poruszono również problem jakości indeksowania manualnego, z którym porównuje się osiągnięcia systemów automatycznych. Autorzy nie współpracowali podczas badań z NLM, co przełożyło się na brak możliwości odpowiedzi na część pytań, które pojawiły się w dyskusji. W przypadku tego rodzaju projektów badawczych współpraca taka wydaje się niezbędna.

Tematem kolejnego wystąpienia, którego autorką była Ziyoung Park (Hansung University), było tworzenie listy opisów koreańskich SOW. Opisy te zostały włączone do bazy BARTOC. Dodatkowo autorka stworzyła strony wiki dla systemów, które powiązała ze wspomnianą bazą. Obecnie BARTOC zawiera wpisy dotyczące około 200 koreańskich SOW. Tego rodzaju przedsięwzięcia są ciekawe ze względu na stosunkowo niewielką wiedzę na temat OW w Korei Południowej w międzynarodowym środowisku informatologicznym¹. Planowane są dalsze prace nad wprowadzaniem nowych opisów i poprawianiem stron wiki.

Pierwsze wystąpienie w kolejnej sesji wygłosił L. P. Coladangelo (Kent State University). Dotyczyło ono wizualnej reprezentacji pojęć w postaci np. grafik bądź nagrań wideo. W modelu Library Reference Model (LRM) nomen, czyli etykieta reprezentująca pojęcie, jest niezależną jednostką, do której przypisany jest zestaw atrybutów. Autor przedstawił na przykładach te cechy nomen, które mogą odnosić się do jego wizualnej wersji, w tym formę oraz aspekty techniczne. Dyskusja, która odbyła się po referacie, dotyczyła m.in. kontekstów, w których

¹ Jednym z systemów, który jest lepiej opisany w literaturze światowej, jest Koreańska Klasyfikacja Dziesiętna (zob. np. Oh, 2020).

wizualne nomen mogą zastąpić środki językowe. Prelegent podkreślił, że widzi miejsce dla takich rozwiązań przede wszystkim w środowisku międzynarodowym, gdzie duże znaczenie mogą mieć piktogramy, jednak w większości przypadków ten rodzaj reprezentacji pojęć może być tylko uzupełnieniem etykiet językowych. Wystąpienie Coladangelo było ciekawe i inspirujące do przemyśleń na temat znaczenia i charakteru etykiet stosowanych w SOW.

Autorkami drugiego wystąpienia w tej sesji były Gema Bueno-de-la-Fuente (University of Zaragoza), Carmen Agustín-Lacruz (University of Zaragoza), Mariàngela S. L. Fujita, (Universidade Estadual Paulista) i Ana Lúcia Terra (Universidade de Coimbra). Jego tematem były zasady i wytyczne dotyczące OW w repozytoriach instytucjonalnych. Przedstawione analizy miały charakter eksploracyjny i dotyczyły w sumie dziewięciu repozytoriów skategoryzowanych jako lokalne, narodowe bądź międzynarodowe. W referacie pokazano problemy z gromadzeniem danych na ich temat, a także zróżnicowane uwzględnienie zagadnień związanych z OW w wybranych kategoriach repozytoriów, tj. najsłabsze w zbiorach lokalnych, a najlepsze w zbiorach międzynarodowych. W związku z tym, że jedną z podstawowych funkcji repozytoriów jest zapewnienie dostępu do zbiorów, słabość w tym zakresie wydaje się sporym problemem. Szczególnie istotne są braki na poziomie podstawowym, jakim są repozytoria lokalne.

W ostatniej sesji, która miała bardziej praktyczny charakter, wygłoszono dwa referaty. Pierwszy z nich, autorstwa Julaine Sashanie Clunis (University of Wisconsin-Milwaukee), dotyczył wykorzystania klinicznych SOW w analizie badań dotyczących COVID-19. Prelegentka podkreśliła, że w przypadku SOW z zakresu medycyny niezwykle ważne są powiązania między różnymi schematami umożliwiające zapewnienie interoperacyjności na poziomie semantycznym. W praktyce poziom wzajemnego odwzorowania systemów jest zróżnicowany. Niezbędne są zatem łatwe w stosowaniu metody ich mapowania. Autorka sprawdziła, w jaki sposób ten proces może zostać wsparty przez zastosowanie narzędzia KNIME – Konstanz Information Miner oraz w jaki sposób można wykorzystać uzyskane wyniki do adnotowania dokumentacji eksperymentów medycznych. Clunis przedstawiła sposób wykorzystania oprogramowania do mapowania wybranych ontologii, takich jak COVID-19 Vocabulary Ontology, COVID-19 Ontology czy też Coronavirus Infectious Disease Ontology. Wyniki mapowania, podobnie jak adnotowania, są bardzo obiecujące z punktu widzenia ich jakości oraz możliwości obniżenia kosztów działalności instytucji medycznych. Występują też jednak pewne problemy, dotyczące np. jakości zastosowanych algorytmów. Przedstawione wystąpienie było dobrym uzupełnieniem wcześniejszej prezentacji autorstwa Zeng i Hong, pokazującym praktyczne zastosowanie systemów medycznych. Warto obserwować postępy w zakresie badań nad medycznymi SOW, gdyż często są to analizy nowatorskie, które można wykorzystać w kontekście innych domen. Ostatni element został zresztą podkreślony przez autorkę wystąpienia.

Ostatnie wystąpienie zaprezentowane przez L. P. Coladangelo (Kent State University) również dotyczyło problematyki zdrowotnej, jednak nieco innego jej aspektu. Przedstawił on prace nad Linked Fitness Training (LiFT) Ontology, systemem z zakresu treningu siłowego. Główną przyczyną rozpoczęcia prac nad tego rodzaju ontologią był brak rozwiązania, które pozwoliłoby na powiązanie zasobów dotyczących tej aktywności. Prace były oparte na analizie domen i metodzie fasetowej z wykorzystaniem metodologii Ontology Development 101. Autor przedstawił w swoim referacie podstawowe klasy i własności. W celu ulepszenia ontologii planowane jest zastosowanie metody delfickiej, studium przypadku oraz person.

Spotkanie zakończył krótkim podsumowaniem Joseph Busch, dziękując prelegentom za przygotowanie wystąpień. Główną osią tematyczną tegorocznych warsztatów NKOS były nauki o zdrowiu i medycyna, jednak podczas spotkania pojawiły się także inne wątki, np. kwestia inkluzywności SOW, dostępu do narzędzi porządkowania zbiorów i zjawisk czy też projektowania ontologii. Warto śledzić wydarzenia NKOS, ponieważ zapewniają wgląd w najnowsze, często eksploracyjne, badania dotyczące szeroko rozumianej OW. Prezentacje oraz abstrakty zostały zamieszczone na stronie grupy NKOS (<https://nkos.dublincore.org/2022NKOSworkshop/NKOS2022.html>). Planowana jest także publikacja artykułów w specjalnym numerze czasopisma *The Electronic Library*.

Bibliografia

- Hong, Y., & Zeng, M. (2022). International Classification of Diseases (ICD) [online]. In: B. Hjørland & C. Gnoli (eds.), *ISKO Encyclopedia of Knowledge Organization*, [17.10.2022], <https://www.isko.org/cyclo/icd>
- Oh, D.-G. (2020). Korean Decimal Classification (KDC) [online]. In: B. Hjørland & C. Gnoli (eds.), *ISKO Encyclopedia of Knowledge Organization*, [17.10.2022], <https://www.isko.org/cyclo/kdc>

Bartłomiej Włodarczyk

ORCID 0000-0001-9229-4656

Wydział Dziennikarstwa, Informacji i Bibliologii
Uniwersytet Warszawski

Nadesłano: 17 października 2022.