

## W STRONĘ UCZELNIANEJ BAZY WIEDZY NOWEJ GENERACJI

**Streszczenie:** Obecnie niezwykle istotne jest, aby zasoby bibliograficzne muzeów oraz bibliotek były ogólnodostępne. Artykuł podpowiada, w jaki sposób stworzyć taką bazę wiedzy, która łączy w sobie narzędzie prezentacji zasobów i opisów bibliograficznych, system raportujący dokonania poszczególnych jednostek naukowych oraz odkrywa nieoczywiste paralele pomiędzy poszczególnymi zasobami.

**Słowa kluczowe:** baza wiedzy, *open source*, VIVO

### Wprowadzenie

W dzisiejszych czasach dostęp do informacji i wiedzy powinien być prosty i powszechny. Dlatego tak ważne jest, aby zasoby bibliograficzne muzeów oraz bibliotek (publicznych lub akademickich) były ogólnodostępne. W tym celu proponujemy bazę wiedzy nowej generacji, która zaspokoi powyższe potrzeby.

Baza ta łączy w sobie narzędzie prezentacji zasobów i opisów bibliograficznych, system raportujący dokonania poszczególnych jednostek naukowych oraz odkrywa nieoczywiste związki pomiędzy poszczególnymi zasobami.

### 1. Najważniejsze wymagania uczelnianej bazy wiedzy

Jednym z najważniejszych wymagań stawianych uczelnianej (i nie tylko) bazie wiedzy jest dopasowanie pod konkretne potrzeby. Dotyczy to zarówno wymagań wynikających z polskiej specyfiki i przepisów, jak i potrzeb konkretnej jednostki naukowej, np. punktacja ministerialna czy procentowy udział autora w tworzeniu publikacji. Ponadto taka baza wiedzy powinna charakteryzować się elastycznym schematem danych i prostotą konfiguracji, jak również wygodą wprowadzania danych. W tym celu pomocne mogą okazać się kompleksowe formularze (kreatory) oraz import/eksport masowy danych z zewnętrznych źródeł, np. Scopus<sup>1</sup> czy Polskiej Bibliografii Naukowej PBN<sup>2</sup>. Oczywiście nowoczesny sys-

<sup>1</sup> <http://www.scopus.com> [12.10.2015].

<sup>2</sup> <http://pbn.nauka.gov.pl> [12.10.2015].

tem przechowywania i prezentacji treści nie może obyć się bez zaawansowanego wyszukiwania – pełnotekstowego oraz z wykorzystaniem operatorów logicznych, takich jak „i”, „lub”, „nie”. Dużą uwagę należy poświęcić jakości przechowywanych i wprowadzanych danych. Dlatego taki system powinien dysponować modulem kontroli jakości danych zarówno na etapie wprowadzania danych do systemu, jak i okresowo w celu podniesienia jakości danych. Przez kontrolę jakości danych rozumie się raportowanie o nieścisłościach w opisach bibliograficznych, np. brak numeru lub błędny numer ISBN, wykrywanie duplikatów danych, tzn. wielokrotnie występująca ta sama publikacja, konferencja lub osoba, lecz pod wieloma nazwami. Do głównych przyczyn występowania duplikatów zalicza się literówki, skróty czy też zmianę nazwiska przez autora.

## 2. VIVO – za, a nawet przeciw

Jednym z popularnych, szczególnie w Ameryce, systemów baz wiedzy jest VIVO<sup>3</sup>, które spełnia wiele z powyższych wymagań. Do głównych zalet tego systemu należy zaliczyć dostępność (*open source*), prostotę instalacji i podstawowej obsługi. Ponadto system ten oparty jest na ontologii. Według A.M. Gadowskiego (Meta-teoria TOGA<sup>4</sup>) ontologia w uogólnionym sensie systemowym to zestaw terminów opisujących „to, co jest” lub „może być” w wybranej dziedzinie percepcji i działania agenta inteligentnego lub robota dla osiągnięcia zadanego celu. Taka ontologia może dotyczyć np. fragmentu rzeczywistości czy jakiejś teorii, mniej lub bardziej dokładnie określonych w pamięci agenta. Aby zapewnić jednoznaczność przekazu informacji/wiedzy dotyczącej określonej rzeczywistości/dziedziny, na zadanym poziomie ogólności, wykorzystuje się kategoryzację oraz hierarchizację<sup>5</sup>.

Przekłada się to na elastyczny schemat danych oraz bogatą semantykę wprowadzonych danych. Oznacza to prostotę wprowadzania zmian, dodawania nowych pól do opisów bibliograficznych oraz jednoznaczność tychże opisów bibliograficznych (np. wprowadzona raz nazwa konferencji może być wielokrotnie wykorzystana i połączona z innymi wydarzeniami, osobami, publikacjami itp.).

Aczkolwiek system ten nie jest pozbawiony wad. Do najważniejszych można zaliczyć brak kompleksowych formularzy oraz kontroli jakości danych i wolne działanie dla większych zbiorów danych. Ponadto wykorzystuje wiele technologii (Java, Javascript, Freemake Template, XML, SPARQL, RDF, baza relacyjna), co skutkuje bardzo czasochłonnym i skomplikowanym wprowadzaniem zmian. W związku z tym proponujemy Państwu nową semantyczną bazę wiedzy pozbawioną tych wad, która została opisana w kolejnym punkcie.

## 3. Nowa uczelniana semantyczna baza wiedzy

Implementacja proponowanej nowej uczelnianej semantycznej bazy wiedzy została podzielona na dwa główne etapy. W pierwszym etapie planujemy wdrożyć funkcjonalności, którymi dysponuje VIVO i nie tylko. Mianowicie system ten umożliwi wprowadzanie opisów

<sup>3</sup> <http://vivoweb.org/about> [8.10.2015].

<sup>4</sup> Gadowski A.M., *Ontology and Knowledge: Meta-ontological Perspective* 1993.

<sup>5</sup> <https://p..wikipedia/wiki/Ontologia> [8.10.2015].

bibliograficznych oraz zostanie oparty na ontologii. Ponadto zostaną zdefiniowane formularze, które zdecydowanie uproszą proces wprowadzania opisów bibliograficznych, oraz zostanie zaimplementowany moduł importu/eksportu danych zgodny ze schematem opisu bibliograficznego PBN. Pozwoli to na wygodne raportowanie wyników jednostek naukowych zgodnie z rozporządzeniem Ministerstwa Nauki i Szkolnictwa Wyższego.

Natomiast w drugim etapie prac implementacyjnych planujemy zwiększyć elastyczność systemu poprzez automatycznie generowane kompleksowe formularze [kreatory wprowadzania (meta)danych] na podstawie ontologii i plików konfiguracyjnych oraz importu danych z innych źródeł, np. takich jak Scopus. Ponadto rozszerzymy funkcjonalność systemu o kontrolę jakości danych, m.in. automatyczne wykrywanie duplikatów danych, sprawdzanie autorstwa oraz moduł wykrywania i dedukcji nietrywialnych związków pomiędzy wprowadzonymi bytami (autorami, publikacjami, wydarzeniami) do bazy wiedzy na podstawie ontologii, np. związki pomiędzy autorami.

Poza tym nowa uczelniana baza wiedzy zostanie udostępniona w ramach licencji *open source* oraz stanowić będzie integralną część systemu dLibra<sup>6</sup>. dLibra to dedykowany system do budowy bibliotek cyfrowych, który ukierunkowany jest na udostępnianie dokumentów pochodzących z bibliotek (akademickich i publicznych), instytucji pozarządowych, fundacji, jednostek publicznych, firm komercyjnych, a także zbiorów prywatnych.

## Podsumowanie

Nowa uczelniana baza wiedzy to nowoczesne rozwiązanie wykorzystujące ontologię i semantyczną bazę danych. Ta baza wiedzy to nie tylko repozytorium osiągnięć naukowych pozwalające na wygodne wprowadzanie danych oraz dysponujące precyzyjną kontrolą jakości danych, ale również baza wiedzy, która pozwoli na odkrywanie nieoczywistych związków pomiędzy autorami, wydarzeniami oraz jednostkami naukowymi.

---

<sup>6</sup> <http://dingo.psnk.pl/dLibra/lang-pref.pl> [13.10.2015].

