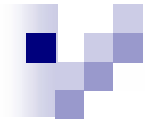


PBC 2009

zespół d/s synonimów

Tomasz Kalota, Biblioteka Uniwersytetu Wrocławskiego



Skład zespołu

- Danuta Murzynowska, Biblioteka Narodowa
- Marta Dmitruk, Biblioteka Narodowa
- Joanna Potęga, Biblioteka Narodowa
- Monika Jędralska, Biblioteka Wyższej Szkoły Humanitas w Sosnowcu
- Skarbimir Kwiatkowski, Biblioteka Uniwersytecka w Toruniu
- Wojciech Kowalewski, Wojewódzka Biblioteka Publiczna w Krakowie
- Tomasz Kalota, Biblioteka Uniwersytetu Wrocławskiego
- Adam Dudczak, PCSS



Plan prezentacji

- Jak działa mechanizm synonimów w dLibrze.
- Jak działa import słowników synonimów w dLibrze.
- Jak obecnie wykorzystywany jest mechanizm synonimów w bibliotekach cyfrowych.
- Omówienie przypadków wykorzystania wraz z przykładami.
- Wnioski i propozycje usprawnień.
- Propozycja stworzenia mechanizmu synchronizacji synonimów pomiędzy bibliotekami cyfrowymi.
- Dyskusja



Jak działa mechanizm synonimów w dLibrze

Idea tzw. synonimów ma na celu zwiększenie możliwości mechanizmu wyszukiwania zasobów w bibliotekach cyfrowych opartych na systemie dLibra. System dLibra w sposób automatyczny rozszerza zapytanie użytkownika o słowa/frazy przechowywane w słowniku synonimów. Dzięki takiemu postępowaniu użytkownik ma możliwość znalezienia większej liczby potencjalnie interesujących publikacji. Mechanizm ten został stworzony również aby zapytanie użytkownika mogło być niezależne od stosowanych w konkretnej bibliotece konwencji opisu obiektów cyfrowych (np. wykorzystanie różnych terminów na określenie tego samego pojęcia).



Jak działa mechanizm synonimów w dLibrze

Rozszerzanie zapytań odbywa się automatycznie w przypadku gdy zapytanie użytkownika zostanie dopasowane do dokładnie jednej grupy synonimów. W takim przypadku dLibra uzupełnia wpisane zapytanie o wszystkie słowa/frazy zawarte w wybranej grupie synonimów. Rozważmy następujący przykład, użytkownik wpisuje słowo **Kopernik** w słowniku synonimów znajduje się grupa **Copernicus** zawierająca dwie wartości **Kopernik i Copernicus**. Zapytanie które trafi do indeksu będzie miało następującą postać: **Kopernik OR Copernicus**.



Jak działa mechanizm synonimów w dLibrze

Mechanizm obsługujący synonimy na obecną chwilę nie pozwala na tworzenie synonimów z wyrażen składających się z kilku słów, synonimem może być tylko pojedynczy wyraz. Znajomość tej zasady jest bardzo ważna dla redaktorów ponieważ nie uwzględnienie jej podczas konstruowania synonimów będzie skutkowało niewłaściwym działaniem mechanizmu wyszukiwawczego bazującego na synonimach. Właściwość tę najlepiej ilustruje poniższy przykład:

Dla wartości ***Gebauer, Johann*** synonimem nie może być ***Gebawer, Johannes*** ponieważ synonim nie może składać się z dwóch słów. W takim przypadku należy stworzyć dwa synonimy, osobno ***Gebawer*** i osobno ***Johannes***.



Jak działa import słowników synonimów w dLibrze

Aplikacja redaktora systemu dLibra umożliwia zaimportowanie słownika synonimów z formatu MARC. Mechanizm ten jest dostępny w wersji 4.0, został on szczegółowo opisany w dokumentacji dotyczącej rozszerzeń aplikacji redaktora/administratora systemu dLibra.



Wykorzystanie mechanizmu synonimów w BC.

- Nazwy osobowe
- Nazwy własne
- Nazwy pospolite
- Format zapisu dat
- Zapis nazw języków
- Inne



Importowanie synonimów z języków formalnych. Problem aktualność słownictwa

Formy JHP BN:

- **Domy dziecka** (hasło wzorcowe)
- Domy sierot (forma odrzucona)
- Sierocińce (forma odrzucona)

W hasłach odrzuconych pojawiają się często formy przestarzałe lub rzadko używane. Mogą nie odzwierciedlać potrzeb użytkowników.

Propozycje dodatkowych form:

- Dom dziecka
- Sierociniec
- Bidul

Najprostszym sposobem dodawania synonimów jest tworzenie liczby pojedynczej lub mnogiej od hasła wzorcowego.

Dodane hasła mogą być przejawem bardziej lub mniej trwałej mody językowej (tu: książka Mariusza Maślanki Bidul)



Importowanie synonimów z języków formalnych. Problem dostępu wielojęzycznego

Jan Kanty (św. ; 1390-1473)


- Ioannes Cantius
 - Iohannes de Canthy
 - Ioannes de Kanthi
 - Ioannes de Martha
 - Ioannes Maleus
 - Ioannes Wacięga
 - Jan z Kęt
 - Jan z Malca
 - Johannes de Canthy
 - Johannes Wacięga
 - Kanty, Jan
- Użytkownicy nieznający języka polskiego bez wariantów językowych nie mają szansy na odnalezienie szukanych zbiorów.
- Czy Europeana jest w stanie rozwiązać problemy związane z różnymi wersjami językowymi na poziomie ponadregionalnym?
- Co ze specjalistami z różnych dziedzin, którzy zapytania mogą tworzyć na podstawie tekstów źródłowych?



Importowanie synonimów z języków formalnych. Problem transliteracji

Dostoevskij, Fëdor Mihajlovič (1821-1881)


- Dostoevskij
 - Dostoevski
 - Dostoevsky
 - Dostoyevsky
 - Dostoevskij
 - Dostojevski
 - **Dostojewski**
 - Dostoevsky
 - Dostoïevski
 - Dastaeŭski
 - Dostoïeffsky
- Formy odrzucone w hasłach BN zawierają hasła utworzone na podstawie norm transliteracji, które obowiązywały przed PN-ISO 9:2000. Dodaje się też często formy odrzucone z NUKAT, który pochodzą z wydań w różnych krajach i wprowadzają zagraniczne normy transliteracji.
- Czy uzasadnione jest ograniczanie synonimów do najprostszej spolszczonej formy?
- Może praca doświadczonych katalogerów przyda się na przykład przy przeszukiwaniu wielojęzycznym?



Wykorzystanie synonimów w BC.

Format zapisu dat.

- ok.1930
 - ca 1930
 - Lafon (fl. ca 1840)
 - Lafon (czynny ok. 1840)
 - Fauquet, Louise (fl. ante 1847-1852)
 - Fauquet, Louise (czynny przed 1847-1852)
- Jakie skróty przy datach?
Polskie – zrozumiałe dla polskich odbiorców
Łacińskie – międzynarodowy charakter



Wykorzystanie synonimów w BC.


Zapis nazw języków.

- spa
hiszpański
- eng
angielski



Wnioski i propozycje usprawnień

- Synonimy są wykorzystywane przez bibliotekarzy i są potrzebne w bibliotekach cyfrowych
- Synonimy mogą przysłużyć się lepszemu pozycjonowaniu treści dostępnych w bibliotekach cyfrowych
- Rozszerzenie funkcjonalności dLibry w zakresie obsługi synonimów złożonych z wielu słów jest bardzo istotne z punktu widzenia poprawnego wykorzystania synonimów w bibliotekach cyfrowych. Mechanizm tworzenia synonimów powinien być intuicyjny.
- Wykorzystanie gotowych słowników (np. Tezaurusz Dziedzictwa Kulturowego - <http://historiasztuki.uni.wroc.pl/tezaurus.html>) może obniżyć koszty opracowania zbiorów w BC oraz zwiększyć ich widoczność w sieci.
- Czy możliwa jest synchronizacja synonimów pomiędzy bibliotekami cyfrowymi i czy mechanizm synchronizacji ułatwi prace redaktorom?



Synchronizacja synonimów – koncepcja w wersji beta

- FBC pełni rolę koordynatora informatycznego oraz dystrybutora słowników.
- Biblioteki cyfrowe synchronizują się co jakiś czas (dzień, tydzień, miesiąc?) w celu synchronizacji danych.
- Synchronizacja polega na pobraniu danych dostarczonych przez inne BC oraz na przesłaniu własnych propozycji synonimów do FBC.
- Tutaj pojawia się pytanie co w sytuacji kiedy BC tworząca synonim po jakimś czasie zdecyduje, że jest on niewłaściwy i usunie z własnej bazy? Czy usunięcie też powinno być synchronizowane z centralnym słownikiem? Czy da się osobno przechowywać synonimy dostarczane przez różne BC? Pozwoliłoby to tworzyć ranking jakości oraz ilości synonimów dostarczanych przez BC
- W dostarczaniu synonimów uczestniczą BC, które zgłoszą do tej pracy redaktorów. Dostarczanie synonimów nie powinno być anonimowe.
- Co jakiś czas (miesiąc?) powinien być generowany raport zawierający informacje o nowych lub zmienionych synonimach.
- Słownik mógłby istnieć w dwóch wersjach: dystrybucyjnej i roboczej. Do wersji roboczej miałyby dostęp biblioteki które zgłosiły redaktorów do współpracy. Po zaakceptowaniu zmian przekazanych w raporcie, zawartość wersji roboczej byłaby synchronizowana z wersją dystrybucyjną.