

04. Wyszukiwanie i przeglądanie zasobów w bibliotece cyfrowej

System dLibra oferuje swoim użytkownikom wiele sposobów dotarcia do interesującego obiektu cyfrowego. Podstawowym sposobem jest wyszukiwanie. Ponadto możliwe jest przeglądanie zawartości biblioteki cyfrowej poprzez kolekcje czy indeksy słów. Dodatkowo każdy obiekt cyfrowy w systemie dLibra posiada swój własny, unikalny link dostępu oraz [identyfikator OAI](#).

Wyszukiwanie lokalne

Wyszukiwanie lokalne obiektów cyfrowych w systemie dLibra jest podstawowym mechanizmem wykorzystywanym przez użytkowników stron internetowych. Jest to najprostszy i najłatwiejszy w wykorzystaniu mechanizm odnajdywania interesujących pozycji. Mechanizm ten wyszukuje w zgromadzonych lokalnie zasobach obiekty i prezentuje listę wyników wyszukiwania poczynając od elementów pasujących najbardziej do zapytania. Informacjami przeszukiwanymi są metadane oraz ewentualny tekst publikacji jeśli jest dostępny. Mechanizm wyszukiwania w systemie dLibra oparty jest na silniku wyszukiwawczym [Lucene](#). Wszelkie właściwości mechanizmu wyszukiwania (np. konstrukcja zapytań wyszukiwawczych, czynniki wpływające na kolejność wyników wyszukiwania) bazują na silniku [Lucene](#) i w jego dokumentacji można odnaleźć więcej informacji na temat zasad jego działania.

Wykorzystanie słownika synonimów

Wyszukiwanie lokalne oferuje również mechanizm rozszerzania zapytań o słowa/frazy. Budowa tego słownika została opisana [tutaj](#).

Automatyczne rozszerzanie zapytań

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*.

W wyszukiwaniu prostym nie ma możliwości aby wyłączyć mechanizm rozszerzania zapytań jest to możliwe tylko w wyszukiwaniu zaawansowanym. Odnosząc to stwierdzenie do powyższego przykładu w wyszukiwaniu prostym nie mamy możliwości aby zapytać się tylko o publikacje opisane (lub zawierające) tylko słowem *Kopernik*. Zapytanie to będzie zawsze rozszerzane.

Należy również nadmienić, że rozszerzanie zapytań o wartości ze słownika jest dostępne tylko dla zapytań składających się tylko z jednego słowa lub fraz. Przykłady zapytań dla których podjęta zostanie próba dopasowania to: *Kopernik*, *"bitwa pod grunwaldem"*, *"Bolesław Prus"*. W przypadku zapytań wielowrazowych dLibra nie rozszerza ich treści, przykłady zapytań, które nie zostaną rozszerzone to: *bitwa pod grunwaldem*, *Bolesław Prus*.

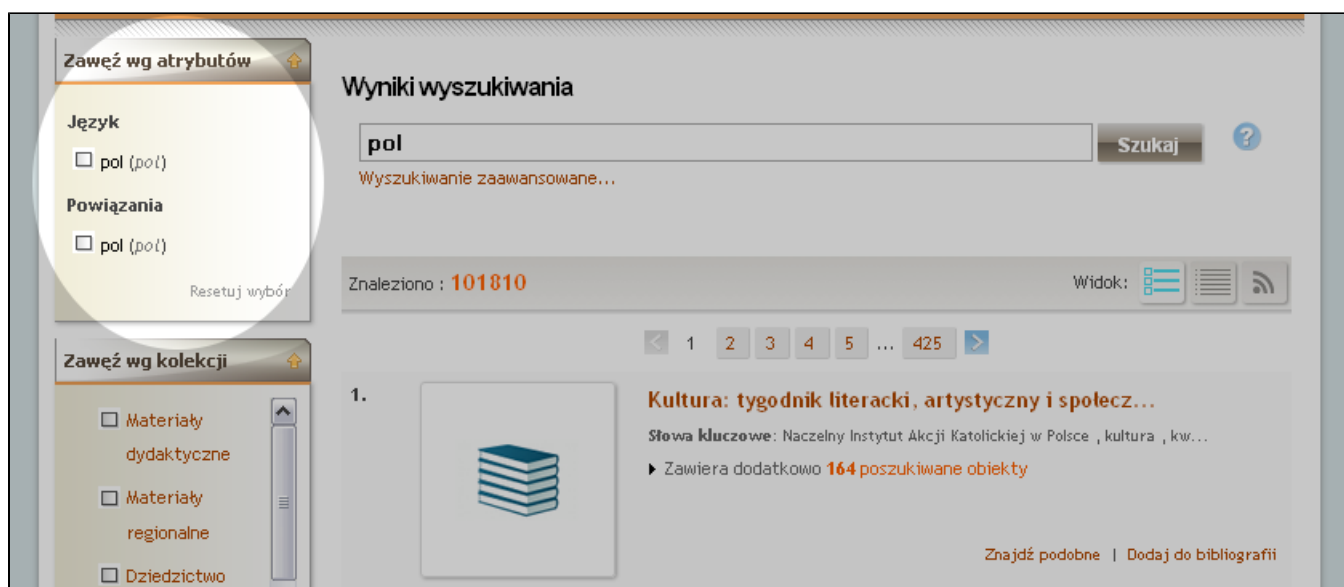
Zawężanie zapytań wg wartości z słownika synonimów

Sytuacja w której dopasowanie jest jednoznaczne jest scenariuszem stosunkowo prostym. Ponieważ te same słowa/frazy mogą występować w różnych grupach synonimów, należących do słowników zdefiniowanych dla różnych atrybutów (por. [02. Słownik synonimów](#)) dochodzi czasem do sytuacji w której do zapytania użytkownika dopasowana zostanie więcej niż jedna grupa synonimów. W tej sytuacji Aplikacja Czytelnika automatycznie rozszerzy zapytanie o wszystkie wartości z wszystkich grup synonimów łącząc je logicznymi spójnikami OR(lub). Po wyszukaniu odpowiednich wyników czytelnik otrzyma możliwość zawężenia zapytania do konkretnych grup synonimów zawartych w atrybutach, których lista pojawi się po lewej stronie wyników wyszukiwania. Na liście poszczególne grupy synonimów, dla danych wyrażen z zapytania, są agregowane wg atrybutów.

Rozważmy następujący przykład:

- treść zapytania: pol (w znaczeniu języka polskiego)
- przykładowe grupy synonimów:
 - pol (atrybut: Język): pol, polski, pl
 - pol (atrybut: Powiązania): pol, polskie

Odnajdując takie grupy synonimów domyślnie zostanie wydane zapytanie *pol OR pl OR polski OR polskie*.



Jeśli w lewym panelu zostanie wybrana jakaś grupa synonimów wówczas zapytanie zostanie zawężone do atrybutu, do którego dana grupa należała. Jeśli więc użytkownik zawęzi zapytanie do grupy *pol* z atrybutu *Język*, wówczas wydane zostanie zapytanie na zasadzie *Język=(pol OR polski OR pl)*.

Zawężanie zapytań wg kolekcji

Po wykonaniu zapytania istnieje również możliwość zawężenia go wg kolekcji. Opcja ta widoczna jest, podobnie jak opcja zawężania wg atrybutów (patrz wyżej), w lewym panelu na stronie z wynikami wyszukiwania. Opcja ta jest dostępna już w wersji 4.0 Aplikacji Czytelnika systemu dLibra, jednakże w najnowszej (tj. 5.x) dodana została możliwość wyboru wielu kolekcji. Wybór kolekcji spowoduje, iż rekordy spełniające zapytanie będą wyszukiwane tylko w wybranych kolekcjach i ich podkolekcjach.

W przypadku wyszukiwania prostego drzewo kolekcji w panelu "Zawęż wg kolekcji" ogranicza się jedynie do pierwszego poziomu (tzn. podkolekcji kolekcji głównej), w przypadku wyszukiwania zaawansowanego w momencie wyboru kolekcji zawierających podkolekcje drzewo się rozwija.

Jeśli w panelu nie jest wybrana żadna kolekcja wówczas wyszukiwanie dokonuje się w kolekcji głównej lub, w przypadku wyszukiwania prostego, może to również oznaczać wyszukiwanie w kolekcji nienależącej bezpośrednio do kolekcji głównej.

Wyszukiwanie zasobów rozproszonych

Poza wyszukiwaniem lokalnym istnieje możliwość włączenia do biblioteki cyfrowej dLibra wyszukiwania zasobów rozproszonych. Wyszukiwanie rozproszonych zasobów bazuje na protokole [OAI-PMH](#) i możliwe jest przy wykorzystaniu usługi wyszukiwania zasobów rozproszonych portalu [Federacji Bibliotek Cyfrowych](#). Wyszukiwanie zasobów rozproszonych polega na tym, że lokalna biblioteka cyfrowa przekierowuje zapytanie użytkownika do serwisu [FBC](#). Następnie wyświetlana jest lista elementów znalezionych we wszystkich bibliotekach cyfrowych, które wchodzi w skład [FBC](#) (aktualna lista znajduje się [tutaj](#)). Lista ta prezentowana jest użytkownikowi na stronach [FBC](#). Serwis [Federacja Bibliotek Cyfrowych \(FBC\)](#) jest kolejnym etapem budowy sieci rozproszonych bibliotek cyfrowych i repozytoriów w Polsce. Nazwa serwisu FBC odzwierciedla jego charakter - serwis ten jest zbiorem zaawansowanych usług sieciowych opartych na zasobach cyfrowych dostępnych w polskich bibliotekach cyfrowych i repozytoriach uruchomionych w sieci [PIONIER](#). Zasoby te współtworzone są przez wiele instytucji naukowych i publicznych, takich jak wyższe uczelnie, biblioteki, archiwa, muzea czy ośrodki badawcze. Serwis [FBC](#) utrzymywany jest przez [Poznańskie Centrum Superkomputerowo-Sieciowe](#) afiliowane przy [Instytucie Chemii Bioorganicznej PAN](#).






















Metody przeglądania zasobów biblioteki cyfrowej

Poza wyszukiwaniem obiektów cyfrowych istnieje jeszcze kilka innych sposobów odnajdywania i przeglądania obiektów cyfrowych. Jest to przede wszystkim zestaw kolekcji, który w zorganizowany sposób dzieli obiekty cyfrowe na mniejsze zbiory. Naturalnym podziałem wydaje się być podział tematyczny i tak też wykorzystuje kolekcje większość bibliotek cyfrowych opartych na oprogramowaniu dLibra. Dzięki hierarchicznej organizacji kolekcji podział może rozpoczynać się od bardzo ogólnych zbiorów (np. *Materiały dydaktyczne*), które podzielone są na bardziej zawężone podzbiory (np. *Matematyka*, *Fizyka*), itd.

Innym sposobem przeglądania obiektów w bibliotece cyfrowej to indeksy utworzone z wszystkich wartości wybranego atrybutu metadanych. To jakie atrybuty będą wykorzystywane do generowania indeksów dostosowywane jest do konkretnych potrzeb instytucji wykorzystującej bibliotekę cyfrową. Przykładowo mogą to być indeksy utworzone na bazie wartości z atrybutów *Autor*, *Tytuł* czy *Temat* i *słowa kluczowe*.

Obiekty złożone, będące wynikiem zastosowania publikacji grupowej, są przedstawiane w systemie dLibra jako pewna hierarchiczna struktura, po której użytkownik może się swobodnie poruszać. Dzięki temu użytkownik ma dostęp do dowolnej części całego obiektu cyfrowego. Przykładowo jeśli jest to struktura gazety to może ona mieć postać jak przedstawiono na rysunku poniżej.

Struktura publikacji:

- ▼  Dziennik
 - ▶  1938
 - ▼  1939
 - ▶  styczeń
 - ▶  luty
 - ▶  marzec
 - ▶  kwiecień
 - ▶  maj
 - ▶  czerwiec
 - ▶  lipiec
 - ▶  sierpień
 - ▼  wrzesień
 - >   Dziennik nr200
 - >   Dziennik nr201
 - >   Dziennik nr202
 - ▶  październik
 - ▶  listopad
 - ▶  grudzień

Przykładowa struktura publikacji grupowej






















Poza wymienionymi wcześniej metodami przeglądania zasobów biblioteki cyfrowej do dyspozycji użytkownika są różnorodne kanały [RSS](#). Między innymi są to kanały z najnowszymi publikacjami, które pojawiły się w bibliotece, plany wprowadzania publikacji czy wiadomości ze strony internetowej biblioteki. Ponadto jako kanał RSS można subskrybować zapytanie wyszukiwawcze użytkownika. W tego typu kanale RSS pojawiają się te nowe publikacje wprowadzone do systemu dLibra, które spełniają kryteria zapytania wyszukiwawczego tego kanału.

Stały odnośnik i identyfikator obiektu cyfrowego

Każda publikacja w systemie biblioteki cyfrowej dLibra posiada unikalny odnośnik do strony internetowej pod którą zawsze znajdują się informacje na jej temat. Strona ta zbudowana jest z adresu strony internetowej biblioteki cyfrowej po którym następuje ciąg znaków `/publication/` oraz identyfikator publikacji. Przykładowo może to być następujący odnośnik: <http://www.wbc.poznan.pl/publication/34>, który składa się z adresu strony internetowej <http://www.wbc.poznan.pl>, ciągu znaków `/publication/` oraz identyfikatora publikacji w tej konkretnej bibliotece cyfrowej 34.

Ponadto każda publikacja w systemie dLibra posiada automatycznie wygenerowany globalny i unikalny identyfikator który jest oparty na otwartym standardzie [OAI Identifier](#). Każdy taki identyfikator składa się z przedrostka oai, domeny repozytorium w którym się dany obiekt znajduje oraz unikalnego identyfikatora w ramach tego repozytorium. Wszystkie te elementy powinny być oddzielone znakiem dwukropka :. I tak na przykład publikacja z poprzedniego przykładu miałaby następujący identyfikator: oai:www.wbc.poznan.pl:34. Szczegóły można znaleźć w samym standardzie [OAI Identifier](#).

Struktura publikacji:

- ▼  Dziennik
 - ▶  1938
 - ▼  1939
 - ▶  styczeń
 - ▶  luty
 - ▶  marzec
 - ▶  kwiecień
 - ▶  maj
 - ▶  czerwiec
 - ▶  lipiec
 - ▶  sierpień
 - ▼  wrzesień
 - >   Dziennik nr200
 - >   Dziennik nr201
 - >   Dziennik nr202
 - ▶  październik
 - ▶  listopad
 - ▶  grudzień