

Biblioteka cyfrowa jako repozytorium OAI-PMH

Protokół OAI-PMH (wersja 2.0)

Protokół OAI-PMH jest protokołem służącym do selektywnego pobierania metadanych opisujących obiekty cyfrowe. Umożliwia on automatyczną wymianę informacji pomiędzy systemami gromadzącymi takie metadane, a systemami chcącymi te dane uzyskać. W opisie protokołu zdefiniowane są następujące pojęcia:

- **Harvester** - klient protokołu OAI-PMH, który pobiera metadane na temat obiektów cyfrowych.
- **Repozytorium** - system udostępniający metadane na temat obiektów cyfrowych.
- **Element** - pojedynczy obiekt opisywany przez repozytorium, posiadający unikalny identyfikator.
- **Rekord** - metadane elementu zapisane w konkretnym schemacie.
- **Zbiór** - sposób grupowania elementów zebranych w danym repozytorium. Zbiór może zawierać podzbiory, a każdy element może należeć do kilku podzbiorów.

Kryteriami selektywnego pobierania zdefiniowane w protokole OAI-PMH to: przynależność elementów do zbiorów oraz data modyfikacji rekordów opisujących elementy. Komunikacja przy pomocy protokołu OAI-PMH oparta jest na schemacie żądanie-odpowiedź i bazuje na protokole HTTP. Odpowiedzi zapisane są w formacie XML. W protokole OAI-PMH zdefiniowano następujące typy żądań:

- **Identify** - pobiera informacje na temat repozytorium. Są to informacje takie jak nazwa, data modyfikacji najstarszego rekordu, lista "przyjaznych" repozytoriów, etc.
- **ListMetadataFormats** - pobiera informacje o schematach metadanych dostępnych w danym repozytorium. Protokół OAI-PMH wymaga, aby każde repozytorium wspierało conajmniej schemat Dublin Core.
- **ListSets** - pobiera informacje o zbiorach zdefiniowanych w repozytorium. Protokół OAI-PMH w żaden sposób nie narzuca zasad definiowania zbiorów takich jak np. zakresy tematyczne czy formaty obiektów cyfrowych.
- **GetRecord** - pobiera pojedynczy rekord z repozytorium.
- **ListIdentifiers/ListRecords** - pobiera listę identyfikatorów/rekordów spełniających zadane kryteria. Kryteriami te mogą dotyczyć daty modyfikacji rekordu oraz przynależności do konkretnego zbioru.

Protokół zawiera również dodatkowe elementy, takie jak możliwość kompresji przesyłanych metadanych, obsługę usuniętych elementów, podział przesyłanych danych na pakiety, etc. Pełny opis protokołu można znaleźć tutaj.

Zewnętrzne usługi i promocja repozytorium

Obecnie dostępne jest wiele zewnętrznych usług związanych z repozytoriami OAI-PMH. Usługi te pozwalają m.in. sprawdzić zgodność danego repozytorium ze specyfikacją protokołu, gromadzą statystyki repozytorium czy pozwalają na przeszukiwanie metadanych zebranych w repozytorium. Posiadając repozytorium zgodne z protokołem OAI-PMH warto zarejestrować je w najpopularniejszych tego typu usługach. Usługi te opisano poniżej, w zalecanej kolejności wykorzystania ich.

Eksplorator repozytoriów OAI

Usługa ta dostępna jest pod adresem <http://re.cs.uct.ac.za/>. Umożliwia ona:

- Testowanie zgodności repozytorium ze specyfikacją protokołu OAI-PMH - możliwe po podaniu bazowego adresu repozytorium i wybraniu opcji "Test the specified/selected baseURL". Repozytoria, które przejdą test mogą być opcjonalnie zapisane na liście predefiniowanych repozytoriów dostępnej na głównej stronie usługi "OAI Repository Explorer".
- Przeglądanie zawartości repozytorium - możliwe po wprowadzeniu bazowego adresu repozytorium i wybraniu typu żądania, które ma być przesłane. Możliwe jest przesłanie wszystkich typów żądań zdefiniowanych w ramach protokołu OAI-PMH.

Zalecamy wykonanie testów zgodności swojego repozytorium z protokołem OAI-PMH i dodanie go do listy predefiniowanych repozytoriów.

Oficjalny rejestr repozytoriów OAI

Rejestr ten dostępny jest pod adresem <http://www.openarchives.org/Register/BrowseSites>. Umożliwia on przejście listy zarejestrowanych oficjalnie repozytoriów oraz dodanie swojego repozytorium do tej listy. Przed dodaniem repozytorium do rejestru, repozytorium to jest dokładnie testowane pod kątem zgodności z protokołem. Test zgodności jest bardziej szczegółowy niż test dostępny w usłudze "Eksplorator repozytoriów OAI". Rejestracja repozytorium możliwa jest pod adresem <http://www.openarchives.org/data/registerasprovider.html>. Powiadomienia o kolejnych etapach procedury wpisania repozytorium do rejestru przesyłane są na adres e-mail administratora podawany przez repozytorium w odpowiedzi na żądanie Identify.

Zalecamy wykonanie testów zgodności swojego repozytorium z protokołem OAI-PMH i zarejestrowanie go w rejestrze repozytoriów. Przed zarejestrowaniem swojego repozytorium sugerujemy sprawdzić, czy repozytorium to nie znajduje się już w rejestrze. Jest to możliwe, jeżeli figuruje ono na liście "friends" kogoś z zarejestrowanych repozytoriów.

Eksperymentalny rejestr repozytoriów OAI w UIUC

Rejestr ten dostępny jest pod adresem <http://gita.grainger.uiuc.edu/registry/searchform.asp>. Umożliwia on przejrzanie i przeszukanie listy zarejestrowanych repozytoriów oraz przejrzanie szczegółowych informacji na temat poszczególnych repozytoriów, a także zestawień i statystyk podsumowujących informacje na temat wszystkich zarejestrowanych repozytoriów. Aby dodać swoje repozytorium do rejestru należy skontaktować się z [Tomem Habingiem](#).

Zalecamy zarejestrowanie swojego repozytorium w rejestrze UIUC. Przed zarejestrowaniem repozytorium sugerujemy sprawdzić, czy repozytorium to nie znajduje się już w rejestrze. Jest to możliwe, jeżeli figuruje ono na liście "friends" kogoś z zarejestrowanych repozytoriów lub też jest już w oficjalnym rejestrze repozytoriów OAI.

OAlster

OAlster dostępny pod adresem <http://oalster.umd.umich.edu/o/oalster/to> wyszukiwarka obiektów cyfrowych, których opisy udostępniane są przy pomocy protokołu OAI-PMH. Umożliwia ona przeszukiwanie metadanych pobieranych cyklicznie ze zgłoszonych do projektu repozytoriów.

Aby zarejestrować repozytorium w wyszukiwarce OAlster należy wykonać następujące czynności:

1. Zarejestrować swoje repozytorium w oficjalnym rejestrze OAI.
2. Zarejestrować swoje repozytorium w rejestrze UIUC.
3. Skontaktować się z [obsługą serwisu OAlster](#) w celu dodania repozytorium.

Google Sitemaps

Google Sitemaps to testowa usługa firmy Google mająca m.in. na celu poprawę stopnia zaindeksowania repozytoriów cyfrowych przez wyszukiwarkę Google. W tym celu, w ramach usługi SiteMap, można zgłosić do Google swoje repozytorium OAI-PMH. Po zgłoszeniu repozytorium Google, poza standardowymi metodami indeksowania repozytorium (czyli przy użyciu robotów WWW), będzie również wykorzystywało protokół OAI-PMH do pobierania informacji o dostępnej treści. Dodatkowo Google udostępnia statystyki związane z indeksowaniem zgłoszonych repozytoriów oraz informacje na temat zapytań zadawanych przez użytkowników Google, a dotyczących tych właśnie repozytoriów.

Usługa dostępna jest pod adresem <https://www.google.com/webmasters/sitemaps>.

Biblioteka cyfrowa oparta na systemie dLibra jako repozytorium OAI-PMH

Od wersji 2.0.1 systemu dLibra w aplikacji czytelnika dostępny jest interfejs OAI-PMH. Interfejs ten umieszczony jest pod adresem <http://adres.biblioteki.cyfrowej.pl/dlibra/oai-pmh-repository.xml>, czyli na przykład <http://www.wbc.poznan.pl/dlibra/oai-pmh-repository.xml>. Taki właśnie adres należy podawać we wszystkich opisanych powyżej systemach na pytanie o adres bazowy repozytorium OAI-PMH.

Działanie interfejsu można sprawdzić m.in. poprzez przesłanie żądania Identify. Można to wykonać dodając do podstawowego adresu repozytorium końcówkę `?verb=Identify`, czyli na przykład <http://www.wbc.poznan.pl/dlibra/oai-pmh-repository.xml?verb=Identify>.

Oprogramowanie dLibra wspiera większość z opcjonalnych rozszerzeń protokołu OAI-PMH takich jak zbiory czy obsługa usuniętych elementów. Protokół OAI-PMH jest w oprogramowaniu dLibra wykorzystywany do realizacji mechanizmu wyszukiwania rozproszonego. Mechanizm ten dostępny jest od wersji 2.2 oprogramowania.