EBIB Biuletyn EBIB, nr 7 (134)/2012,
Polska oferta edukacyjna dla bibliotekarzy — stagnacja czy kreacja?
Dział sprawozdania

Tomasz Parkoła
Poznańskie Centrum Superkomputerowo-Sieciowe

Konferencja Open Repositories 2012

Słowa kluczowe: *otwarte repozytoria, konferencje naukowe*

Fot. 1. *Banner konferencji Open Repositories 2012 „Open Services for Open Content".*
Źródło: *OR2012* [on-line]. [Dostęp 24.09.2012]. Dostępny w World Wide Web: http://or2012.ed.ac.uk/about/.

Konferencja *Open Repositories 2012* (*OR2012*) wraz z towarzyszącymi warsztatami odbyła się w Edynburgu, w Szkocji, w dniach 9–13 lipca 2012 r. Bogaty program konferencji oraz interesujące warsztaty, a także ogromna liczba uczestników potwierdzają rangę cyklu konferencji *Open Repositories*.

Fot. 2. *OR2012. The 7th International Conference on Open Repositories*
Źródło: PARKOŁA, T. Konferencja Open Repositories 2012. W: *Zespół Bibliotek Cyfrowych PCSS. Dział Usług Sieciowych — Poznańskie Centrum Superkomputerowo-Sieciowe* [on-line]. 19 lipca 2012. [Dostęp 24.09.2012]. Dostępny w World Wide Web: http://dl.psnc.pl/2012/07/19/konferencja-open-repositories-2012/.

Konferencja składała się z wielu sesji. Szczególnie interesujące z punktu widzenia budowy bibliotek cyfrowych były te związane z przetwarzaniem dużych ilości danych (tzw. *data mining*), a także te dotyczące długoterminowego przechowywania danych.

Pierwsze z wymienionych dotykały różnych tematów, m.in. przeszukiwania ogromnych ilości danych, semantycznego wyszukiwania, agregacji metadanych i danych, ekstrakcji informacji z dokumentów tekstowych oraz przepływów pracy (ang. *workflows*) związanych z danymi tekstowymi. Przedstawiono różne systemy opracowane z myślą o przetwarzaniu dużych ilości danych, w tym:

- Evidence Finder (http://labs.ukpmc.ac.uk1), który pozwala na przeszukiwanie 2 mln dokumentów i 71 mln zdań.
- MEDIE (http://www.nactem.ac.uk/medie/), który pozwala na semantyczne wyszukiwanie informacji biomedycznych (bazuje na MEDLINE, http://www.nlm.nih.gov/pubs/factsheets/medline.html).
- Argo (www.nactem.ac.uk/Argo), który pozwala na tworzenie przepływów prac przetwarzających dane tekstowe.
- HIVE oraz rozszerzenie HIVE-ES (http://www.nescent.org/sites/hive/) ułatwiające tworzenie metadanych i słowników wartości.
- CORE (http://core-project.kmi.open.ac.uk/), umożliwiający wyszukiwanie danych i metadanych repozytoriów dokumentów, zawiera mechanizmy pozwalające na znajdowanie treści dokumentu na podstawie pobranych metadanych.

Przy rozwoju omawianych systemów wykorzystywano różne narzędzia do przetwarzania tekstu, np.:

- TextCat (http://odur.let.rug.nl/vannoord/TextCat/),
- U-Compare (http://u-compare.org/),
- OSCAR4 (https://bitbucket.org/wwmm/oscar4/wiki/Home),
- ANTRL (http://www.antlr.org/),
- MAUI (http://code.google.com/p/maui-indexer/),
- KEA (http://www.nzdl.org/Kea/),
- Sesame (http://www.openrdf.org/index.jsp),
- H2 (http://www.h2database.com/).

Warsztaty powiązane z długoterminowym przechowywaniem danych źródłowych dotyczyły przede wszystkim oprogramowania Trident (http://tridentworkflow.codeplex.com/) i możliwości jego konfiguracji oraz wykorzystania. Podczas warsztatów zaprezentowano również najważniejsze kwestie związane z długoterminowym przechowywaniem danych źródłowych, w tym zasady identyfikowania plików, które powinny podlegać migracji lub normalizacji, oraz narzędzia, które można wykorzystać do budowania procesu przechowywania danych. Omówione narzędzia to:

- Kepler (https://kepler-project.org/),
- Taverna (http://www.taverna.org.uk/),
- Ptolemy II (http://ptolemy.eecs.berkeley.edu/ptolemyII/),
- Triana (http://www.trianacode.org/).

Konferencja trwała trzy dni, podczas których zaprezentowano wiele ciekawych referatów związanych z rozwojem szeroko rozumianych repozytoriów cyfrowych, m.in.: *Build to scale*

[1] Wszystkie odesłania do stron internetowych przedstawiają wersję aktualną w dniu 24.09.2012 r.

— referat omawiający budowanie systemu wyszukiwania dla 250 mln rekordów, opartego na Apache Solr i dostarczającego wyniki wyszukiwania w ciągu co najwyżej dwóch sekund.

- *Inter-repository Linking of Research Objects with Webtracks* — referat omawiający propozycję protokołu InteRCom, który pozwala na wymianę semantycznych informacji między repozytoriami.
- *ResourceSync: Web-based Resource Synchronization* — referat przedstawiający protokół synchronizacji danych i metadanych, bazujący na doświadczeniach protokołu OAI-PMH oraz OAI-ORE.
- *Griffith's Research Data Evolution Journey: Enabling data capture, management, aggregation, discovery and reuse* — referat opisujący infrastrukturę w ramach uniwersytetu Griffith, w tym narzędzia semantyczne VIVO (http://sourceforge.net/apps/mediawiki/vivo/) oraz VITRO (http://vitro.mannlib.cornell.edu/).
- *Multivio, a flexible solution for in-browser access to digital content* — referat przedstawiający uniwersalną przeglądarkę dokumentów PDF, GIF, JPEG czy PNG.
- *ORCID update and why you should use ORCIDs in your repository* — referat omawiający aktualny stan i plany rozwojowe systemu identyfikowania naukowców ORCID (http://about.orcid.org/).
- *Digital Preservation Network, Saving the Scholarly Record Together* — referat omawiający inicjatywę powstałą w Stanach Zjednoczonych, dotyczącą budowania heterogenicznego systemu długoterminowego przechowywania (http://d-p-n.org/).

W ramach konferencji przedstawiciel Poznańskiego Centrum Superkomputerowo-Sieciowego zaprezentował referat pt. *dArceo services: advancing long-term preservation*, omawiający usługi długoterminowego przechowywania danych źródłowych dla polskich instytucji nauki i kultury, ze szczególnym uwzględnieniem materiałów tekstowych, graficznych i audiowizualnych. Zachęcamy do odwiedzenia strony konferencji *OR2012* (http://or2012.ed.ac.uk/), gdzie znajdują się prezentacje autorów oraz program konferencji.

Parkoła, T. Konferencja Open Repositories 2012. W: *Biuletyn EBIB* [online] 2012, nr 7 (134), *Polska oferta edukacyjna dla bibliotekarzy - stagnacja czy kreacja?* [Dostep: 20.10.2012] Dostępny w World Wide Web: http://www.nowyebib.info/images/stories/numery/134/134_parkola.pdf. ISSN 1507-7187.