Authority framework in 1.6 and CILEA's customization for the Hong Kong University

Andrea Bollini (CILEA)
Allen Lam (The University of Hong Kong)
Susanna Mornati (CILEA)
David T Palmer (The University of Hong Kong)

Universities and researcher centers are rethinking their communication strategies, highlighting the quality of their research output and the profiles of their best researchers. Listing publications from an Expert Finder system may represent a solution. But providing an Expert Finder system within an IR is a more innovative approach. This idea was developed by the University of Hong Kong Libraries and applied to their IR, HKU Scholars Hub at http://hub.hku.hk/, powered by DSpace.

This presentation shows how the HKU requirements were implemented by CILEA in the context of the ResearcherPage@HKU project. Using the new authority control framework by Larry Stone, introduced in DSpace 1.6.0, an Expert Finder system can be nicely integrated with DSpace but kept technically separated. Its components can evolve separately and are easier and cheaper to maintain. The author (Bollini) has contributed in porting the authority control framework, originally implemented for the XMLUI, to the JSPUI, and extending its architecture to support browse and search variants.

The main objective of the framework is to provide admissible metadata values from value lists that can be maintained independently from DSpace. These values can be represented by unique identifiers (IDs, URIs, etc.) that make them independent from their representation (e.g. a translation in a different language). Before 1.6 release, DSpace had a limited capacity to manage value lists. The authority control framework overcomes these limitations and allows integration with dynamic lists taken from external systems through webservices, database queries, and so on. The lists are available also from the admin UI. AJAX technologies can be applied: autocomplete, and partial refresh. These can facilitate users' experience. Long lists are managed.

The ResearcherPage@HKU project consists of a new entity, the ResearcherPage (RP), where all data about a researcher are stored. The RP becomes an authority list for the dc.contributor metadata in the IR. It is partly populated by direct input and partly from machine loads of data extracted from external databases. The researcher can also decide about the public visibility of metadata. Each individual has one established Roman script name, with a possible Hanzi script (Chinese 中文) name also existing in parallel synonymy. Many other variant names in any UTF-8 supported script may then be associated with these established headings, allowing retrieval of the established heading(s) after searching with any of the variants. RPs are strongly integrated in DSpace, but they are an independent software system. Their persistence is linked to specific database tables. Access to data uses a separate connection, with Hibernate as a framework of abstraction from the DBMS. Search is implemented with Hibernate Search, input forms and presentation layer with Spring MVC. Spring Framework is also used to configure other aspects of the system, e.g. transactionality.

This presentation describes in detail all the rich functionalities developed for the RPs and their technical design and implementation, e.g. submission, search, matching for deduplication, confidence level, authority key for unique matching and persistence.

Conclusion. The authority control framework allows implementing relevant "researcher-centric" functionalities without changing the core of DSpace. This solution is more portable to new DSpace releases and can evolve separately. Moreover, populating

metadata values from authority lists improves semantic interoperability between repositories (e.g. homonym disambiguation) where metadata are exposed in standard formats (e.g. FOAF) and richer ones (e.g. MODS, RDF).