Cataloguing documents in non Roman scripts in SUDOC
the French university libraries network

Danièle DUCLOS-FAURE
(daniele.duclos-faure@education.gouv.fr)

[Abridged translation by Annick Bernard]

According to an official survey conducted by the French Ministry of Education among 103 university libraries, 44 libraries hold 1 million books in non-Roman scripts. Among them, 20 libraries hold 73 000 titles in Arabic (mainly), Persian, Ottoman and Pashto.

Nearly a third of the non Roman titles do not appear in any catalogue, due to lack of relevant competence in the concerned language. Most of the existing records appear in card catalogues and include the original script, while 44 % of them, which have been converted and loaded in the library system with romanization only, can be found in SUDOC, the union catalogue for academic institutions and a cataloguing facility shared by 113 libraries (march 2002). With 5 millions records and 13 millions locations, SUDOC uses only Roman scripts for the time being. Bibliographic and authority records are structured in UNIMARC format [www.sudoc.abes.fr].

The interdisciplinary project *"Langues et civilisations du monde"* aims at bringing together the teaching and research units of nine academic institutions in Paris, all specialized mainly in Central and East European, Asian, and African studies. Their respective libraries and catalogues of over 1,6 million books will merge into one: BULAC (Bibliothèque universitaire des langues et des civilisations), a library open to the academic community as well as to a wider readership.

A working group of ten librarians assisted by six other groups of expert librarians all specialists of various Oriental languages and non Roman scripts was set up to discuss the feasibility of a unique catalogue: their report concluded that the BULAC should have a multi-script catalogue and that it should be a member of the higher education cataloguing network SUDOC. Consequently, it has become a priority for SUDOC to develop keying, indexing, search and display functions for non Roman scripts [www.sup.adc.education.fr/bib/].

The following conditions have to be met:
-    Use of UNICODE (UTF 16) for storing, indexing and displaying the various character sets: by now, SUDOC is in the process of implementing UNICODE [www.unicode.org];
-     Use of romanization systems reversible as far as possible: ISO 233-2 standard, widely used in France, is recommended for romanizing Arabic [www.iso.org] [www.afnor.fr];
-     Use of UNIMARC format for both bibliographic and authority records which allows repeating the fields in original and romanized scripts in the same record; [http://www.ifla.org/VI/3/p2001/guideindex.htm];
-    Use of a real authority file where all possible forms of authors and uniform titles will have to be recorded;
-    use of ID numbers of authority records to link bibliographic records to the authority file.

These conditions are necessary to arrive at accurate and comprehensive results whether the search request is done via authorized heading or its variants, or still via original or Roman script.

Regarding authority files for names and titles, some recommendations of the report are related to the format for structuring the authority data, that is to the form of authorized and variant headings:

-    Data structure should include
     o   at the record level : nationality, language(s) of the author (language of the country of origin ; language of the title);
     o   at the form level: script of the form; romanization standard;

-    form of authorized headings to be displayed in the bibliographic record: it is recommended to mention
     o   form in original script;
     o   romanized form, according to ISO 233-2 for Arabic;

- o   French form commonly used;

- variant forms: cannot be displayed in a bibliographic record, but have to be indexed so that they can be used for accessing the record
    - o   variants of the name;
    - o   names complying with American romanization standards (ALA-LC), if different;
    - o   names complying with older romanization systems, if different;
    - o   any other form of the name appearing on the document.

The above mentioned set of recommendations aims at complying with the needs of the professional community as well as a large readership whereby a multi-script catalogue should allow the development of the higher education union catalogue and the choice of BULAC's integrated library system, making sure furthermore that both interact efficiently. SUDOC will convert its data to UNICODE during 2003, and its developments to implement non Roman scripts should be completed in 2004.

[an error occurred while processing this directive]