

Report

Work in progress on the new Soggettario

Anna Lucarelli

(Central National Library of Florence, Department of the National Italian Bibliography,
head of the project on the Renewal of the Subject Heading List for Italian Library Catalogues)
<anna.lucarelli@bncf.firenze.sbn.it>

update of an Italian report presented to a roundtable of the Semantic Web working group
and published in "LIUC papers", n. 183, December 2005
translated by Claudio Gnoli (ISKO Italy)

Work on a prototype of the new Soggettario (the main Italian subject heading list) has begun in October 2004 at the Central National Library of Florence (BNCF). BNCF is involving in the renewal of the most used subject indexing tool of Italian libraries. The project had already produced a Feasibility Study, representing a reference for the works which have started recently. An Italian abstract of the Feasibility Study, and more documentation, can be found in the BCNF website at the Web address <<http://www.bncf.firenze.sbn.it/progetti/>>.

Together with the project team and the BNCF staff, some young external consultants are working, who are focusing on the terminological component of the system, namely the prototype of the Thesaurus. At the same time, documents are produced which define procedures, establish criteria, and give guidelines.

The list of items by which the prototype started consists of:

- the terms included in the updates provided by the Italian National Bibliography (BNI) during the period 1986-1998 (already published in the form of lists);
- the items introduced from 1999 to 2005
- other items introduced in the past years but never recorded;
- some terms from the Soggettario (1956) or from BNI updates (1956-1985), being especially outdated and needing revision.

In building the semantic networks and the hierarchies, and in making the terminological control, of

course, more terms have to be included, which come from the Soggettario, the BNI and other authoritative sources both catalographic (various indexing tools) and lexicographic (general and special directories). DDC numbers are also related to the terms in the Thesaurus.

The prototype, to be completed in April 2006, will include a sample of terms for each disciplinary area. About 5000 terms will have a complete structure, but 6000 more will be included in the Thesaurus in order to fill the semantic networks, and marked with a different working status.

For the prototype the AgroVoc software is used, which has been provided by FAO and adapted by the BNCF computing staff in order to match the specific requirements of the project. Thanks to the potential of this software, we intend to test in future on the multilingual side of terminology. Such work is likely to begin by testing links to the corresponding forms used by the Library of Congress. We are currently beginning to focus on this, and we wish that external parties be involved which are concerned with multilingual terminology in more or less specialist contexts. We will follow the road of conventions with Italian universities, which could cooperate to this development through their students and graded students. We are also looking at the developments in the work of the British BSI working group on standards for thesauri convened by Stella Dextre Clarke.

Testing on the syntactic component of the language is in progress, carried out by a team working on the basis of some terms of the Thesaurus which have been identified for possessing some special features. This will maybe lead to some changes and simplifica-

tions in the original structures coming from the GRIS model <<http://www.aib.it/aib/commiss/gris/guida.htm>> and the Feasibility Study.

Within June 2006, two basic documents will be published:

- a paper volume containing the Rules, concerning both terminology and syntax, explanations of the Thesaurus, criteria, choices made about the categories and the width of the semantic network, etc.;
- a CD containing the Thesaurus.

One side of great interest will be the e-learning process of the indexing language of the new Soggettario. This has to be undertaken only after the prototype stage is done. At that time, once the text of the Rules will have been published, it will be possible for the system to spread, to be learned, and to be tested by various indexers and partners.