

ASPECT Project – WP5 Adopting Standards and Specifications for Educational Content

CNDP's Vocabularies: potential use cases and workflows

Author: Rosa Maria Gómez de Regil
Institution: CNDP
Date: July 2009

MOTBIS thesaurus

Description

MOTBIS is the French educational thesaurus used to index the educational information. MOTBIS represents the education science domain and the general French curriculum contents. The thesaurus is mainly used in the French secondary school libraries, teacher's universities and at the SCÉRÉN network¹, by teachers, pupils and librarians.

MOTBIS is managed by the CNDP. Different documentation and education specialists maintain and update the thesaurus. The copyright is shared within the SCÉRÉN network (CNDP – CRDP).

The work is lead by the orientation committee formed by 16 persons²:

- 1 project manager
- 1 controlled vocabularies expert
- 3 librarians
- 1 educational expert (IUFM)
- 10 experts from different French regions (interacadémique)
- external experts for punctual needs.

Characteristics and functionalities

- The thesaurus is available online at <http://motbis.fr>
- The thesaurus is on sale to be included in other applications.
- Motbis use a poly-hierarchical structure since 2009
- The terms are presented in alphabetical, rotated and systematic displays
- New subjects are proposed according to the national educational policy, for example :
 - o Sustainable development
 - o Handicap
 - o Computing
 - o Economy
 - o Places
 - o Dates & periods.
- Each subject is developed in a specific working group (3 or 4 persons). The groups work mainly by mail.
- Users can propose a new term using an online form.

¹ The SCÉRÉN (Services, Culture, Éditions, Ressources pour l'Éducation Nationale) network is formed by the CNDP (national centre), 32 CRDP (regional centres), 85 CDDP (departmental centres) and CLDP (local centres).

² Only the project manager works full-time.

The screenshot shows the MOTBIS 2009 web interface. The main content area displays the search results for the term "éducation (NT 2105)". It features a central box for the term with several surrounding boxes: "Terme générique" (None), "Employé pour" (None), "Termes(s) associés(s)" (enseignement [NT 2105], psychologie de l'éducation [NT 2205]), and "Termes(s) spécifique(s)" (éducation comparée, éducation extra-scolaire, éducation interculturelle). A left sidebar contains a list of terms under "Liste alphabétique".

Motbis Thesaurus Interface

Tools

MOTBIS is managed using the G3i software, developed within the SCÉRÉN network. G3i is client/server based software.

At the moment G3i can only export the thesaurus in 3 non-standardized formats:

- .txt
- .xml
- .CSV

The screenshot shows the G3i Editing Tool interface. It displays a list of terms on the left, including "accessoire vestimentaire [TAG]", "accident", "accident chimique", etc. On the right, there are sections for "Notions génériques", "Notions spécifiques", and "Notions associées". At the bottom, there are radio buttons for "Hiérarchie", "Tête de hiérarchie", and "Thématique", along with other relationship options like "Inséré seul", "Employé pour", "Générique de", etc.

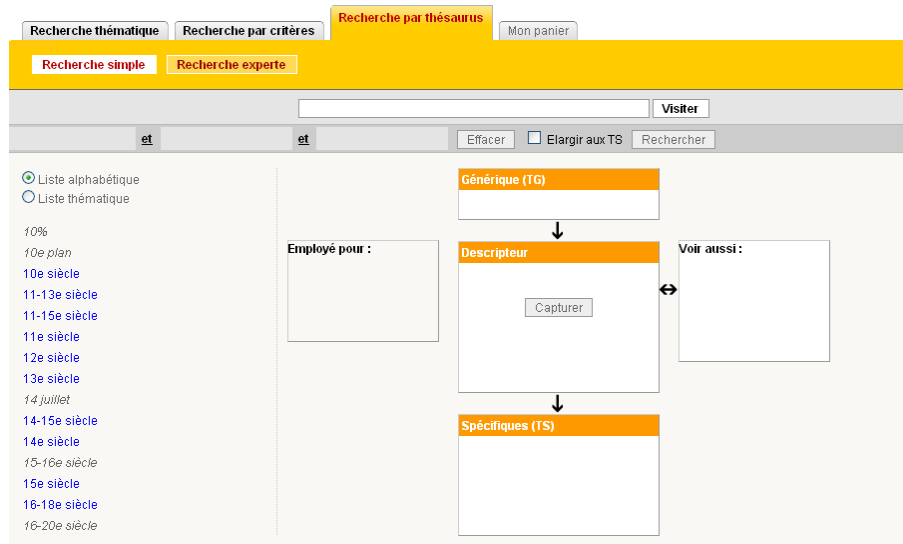
G3i Editing Tool

Terms and concepts workflow

- One subject is chosen by the working group.
- If there are some particular or very important terms to consider, these are compared to the thesaurus content.
- The working group analyses the terms and proposes to put the term in a specific place of the hierarchy.
- Each new candidate term is proposed to the orientation committee through an online questionnaire.
- The orientation committee approves or rejects (with a detailed explanation) the candidate term.
- The conclusions are summarized in a rapport including the approved terms, the rejected terms and the terms in doubt.
- An annual meeting is organised and the issues are discussed by the experts.
- A rule system is being established to select the terms that are going to be included in Motbis.

Creation workflow

- When the new terms and modifications are approved, they are captured in G3i software by one identified person, usually the project manager.
 - o The CNDP plans to open the capturing procedure to other members of the working groups and to have a workflow system included in the software (G3i or another software)
- Periodically, the thesaurus terms are extracted in a home-xml format to be included in the BCDI library software, also developed by one of the members of the SCÉRÉN network (CRDP de Poitou-Charentes)
- A new validation is done after this action
- And then, the thesaurus can be updated in the website and in BCDI software



The screenshot shows the Motbis interface integrated to BCDI library software. The interface features a navigation bar with tabs: 'Recherche thématique', 'Recherche par critères', 'Recherche par thésaurus' (highlighted), and 'Mon panier'. Below this, there are buttons for 'Recherche simple' and 'Recherche experte'. A search bar is present with a 'Visiter' button. The main content area is divided into sections: 'Liste alphabétique' and 'Liste thématique' (selected), a list of terms (e.g., '10e plan', '10e siècle', '11-13e siècle', etc.), and a central workflow diagram. The diagram shows a flow from 'Générique (TG)' to 'Descripteur' (with a 'Capturer' button) to 'Spécifiques (TS)'. There are also 'Employé pour' and 'Voir aussi' fields.

Motbis integrated to BCDI library software

OTAREN system

Description

OTAREN (Outil Thématique d'Aide à la Recherche pour l'Éducation Nationale) is a conceptual tool that proposes a mapping between different controlled vocabularies:

Thesauri

- Motbis (SCÉRÉN network thesaurus) - <http://www.motbis.fr/>
- ONU thesaurus - <http://lib-thesaurus.un.org/LIB/DHLUNBISThesaurus.nsf>
- UNESCO thesaurus - <http://databases.unesco.org/thesfr/contact.html>
- AGROVOC - <http://www4.fao.org/agrovoc/>
- EUROVOC - <http://europa.eu/eurovoc/>
- GEMET - <http://www.lexicool.com/gemet-multilingual-environment-thesaurus.asp?IL=1>
- TEE (European Educational thesaurus) - <http://bdd.inrp.fr:8080/Tee/TeeWelcome.html>
- ETB thesaurus - <http://etb.eun.org/eun.org2/eun/en/etb/content.cfm?lang=en&ov=7208>
- INIST - <http://www.termsciences.fr/>

Subject headings

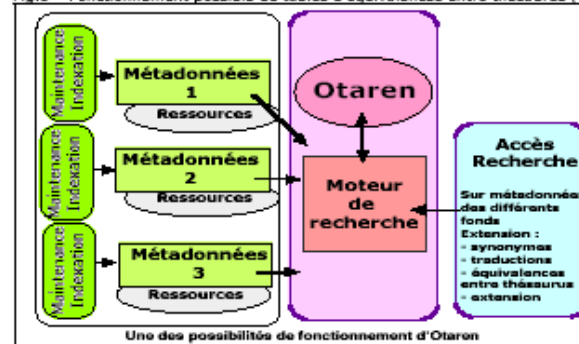
- RAMEAU (French National Library subject headings list)
- LCSH

Authority lists

- CNDP authority list
- BNF authority list - <http://catalogue.bnf.fr/noticesautorites.html>
- LCNA - <http://authorities.loc.gov/>
- UQAM - http://www.manitou.uqam.ca/manitou.dli?INVITE+source+2_uqam_0+0+uqam

The objective of this mapping is to apply search functionalities through different metadata records avoiding the database reindex process.

Fig.3 – Fonctionnement possible de tables d'équivalences entre thésaurus (Otaren)



Context³

OTAREN (Outil Thématique d'Aide à la Recherche pour l'Éducation Nationale) is the experimental prototype of a family of search and retrieval languages likely to be provided to different categories of users by the CNDP in the near future: it has been designed according to a general context and, moreover, has been developed within the framework of this special themed issue, whose main points are:

- the diversification of the ways to access information on the resources which are accessible using PCs and its consequences in terms of "locating relevant information" ;

³ The text presented in this section was written by François Feyler.

- the contribution of documentary searches both in free text (summary words, keywords, etc.) and controlled language (thesaurus descriptors) ;
- the issue of distinguishing between indexing languages and search interfaces.

The possibility of carrying out “full text” searches on a set of fields (title, summary and indexing words, both controlled and otherwise...) on a PC which arose in the early 1980s has highlighted three phenomena.

- Searches which are only carried out on a controlled indexing area by the indexing language still produce “false negatives”, even if these are very effective in terms of countering “false positives”. Good prior knowledge of the installed documentary language as well as the enhanced search procedures (using up-posting on the thesaurus) enables users to get good results without reaching completeness in terms of relevance.
- “Full text” searches often enable you to unearth totally relevant documentary references which could not be found using the installed documentary language. On the other hand, “full text” searches can also produce false positives (searches on ambiguous strings) at the same time as sometimes very significant false negatives.
- The chances of improving the search results are increased if the user knows how to use top-down posting (in the case of a theme-based search with a thesaurus) and truncation together, while remaining aware of any potential ambiguity in certain strings. However, with the majority of tools in use today, this type of search essentially corresponds to an “expert” documentary search; a target which has yet to be achieved in a training system in documentary searches but, by all accounts, is not the case with the majority of students, at the beginning of their secondary education at least.

Application

OTAREN has been implemented into Pertimm's semantic search engine technology to experiment the semantic filtering possibilities in a full-text and metadata document corpus.



The screenshot shows the 'Outil de Recherche Sémantique' interface. The search bar contains 'prehistory' and the search button is labeled 'Rechercher'. Below the search bar, there is a checkbox for 'Recherche en mode concept' and a link to 'Réinitialiser votre recherche'. The search results are displayed in a list, with the first result titled 'Thématique : Education et image(s) 1 - Dans les enseignements et les activités éducatives' from 'Educasources'. The second result is 'Boutiques de musées - Les jeux de la Préhistoire' also from 'Educasources'. On the right side, there is a 'Thésaurus' sidebar with sections for 'TERMES GÉNÉRIQUES', 'TERMES SPÉCIFIQUES', and 'TERMES ÉQUIVALENTS ET ASSOCIÉS'. The 'TERMES ÉQUIVALENTS ET ASSOCIÉS' section shows 'préhistoire' and 'Andy Warhol' as associated terms. Below the thesaurus is a section for 'Expressions associées'.

OTAREN integrated to Pertimm, a semantic search engine

Other controlled vocabularies

Nomenclatures

Description

The CNDP has defined a set of nomenclatures to describe educational resources according to the French curriculum.

The nomenclatures are combined with Motbis thesaurus to improve the indexation, cataloguing and retrieval of educational resources.

The nomenclatures are managed and maintained by the CNDP. Different documentation and education specialists from the SCÉRÉN network are also involved in these activities.

Nomenclatures types

- a) Specifically designed to describe educational resources
 - Subject fields
 - Educational levels
 - Intended users
- b) General authority lists
 - Document format
 - Contributor roles
 - Periods
 - Languages
 - Countries
 - Document type
- c) Specifically designed to the SCÉRÉN network productions
 - Publishers list
 - National collections

Characteristics and functionalities

- The nomenclatures are available online only in the intranet.
- The design departure point is the French curriculum.
- The nomenclatures are used in the educational metadata system Numéribase and in the SCÉRÉN productions' database BNPR.

Creation workflow

- The design and maintenance work is lead by a project manager at CNDP.
- The project manager works with the subject field experts of the CNDP's Publishing Department to define de nomenclatures' entries.
- The nomenclatures are submitted to the Ministry of Education for a global validation.
- The updating workflow is being studied.

Tools

- The nomenclatures are managed in a relational database and an excel file.
- Mapping systems between different versions of the nomenclatures exist.

LOM Vocabularies

Description

The CNDP use the IEEE LOM vocabularies within the educational metadata system Numéribase.

These vocabularies are managed and maintained by the CNDP according to the evolutions of the LOMFR metadata application profile.

LOMFR vocabularies

- General.Language
- General.DocumentType
- Technical.Format
- Technical.Requirement.OrComposite.Type
- Educational.LearningResourceType
- Educational.Activity
- Rights.Cost
- Relations.Type

Tools

- The vocabularies are managed in a relational database.
- The vocabularies are included in the metadata record xml export.

Issues and perspectives

- SKOS file import for Motbis and OTAREN
- Semantic interoperability
- Semantic Web vocabularies management
- Integration of the new standards as ISO 25964 (thesauri and interoperability with other vocabularies).