TERMINO-SEMANTIC STAKES IN THE ELABORATION PROCESS OF SC36 STANDARDS

contribution to 2 expert reports recommended during Seoul plenary session (September 2003)

Presentation to the SC36 plenary session, March 2004
And elements of thoughts for the WG1
Montreal, March 2004

Henri HUDRISIER, Mokhtar BEN HENDA
Rachid ZGHIBI, Laurent ROMARY, Jean-Michel BORDE

SEM@TICE group, AUF
In collaboration with Jean DELAHOUSSE (MONDECA)

French Speaking University Agency (FUA)

This work is a contribution to both expert reports about termino-semantic stakes decided during the SC36 plenary in Seoul, September 2003.

➔ Report on standardized terminology ISO/IEC/TC37 : Frank Farance, Henri Hudrisier and Mitsuru Ikeda

➔ Report on Topic-Maps : Frank Farance, Henri Hudrisier, Jon Mason and Lassi Nirhamo

THE TERMINO-SEMANTIC NEEDS

- Needs for standardized terminologies for communication between experts and vocabulary stabilization in a limited number of languages ;
- Needs for multilingual standardized terminology : more than two or three languages ;
- Needs through fields of interest and levels of specialization of experts or by fields of cultural and linguistic specifications ;
- Needs for concepts systems for the development of MLR;
- Needs for concepts cartographies and ontologies to master :
  1- a technical and conceptual global vision ;
  2- prospective and governance of the standard development.

Needs for a standardized communication terminology between experts and needs for a vocabulary stabilization in a limited number of languages.

This need will be assured by the terminological standard ISO/IEC 19781 which mapping could be later assured with TC37 terminological standards.

Needs for a multilingual standardized terminology

- The SC36 terminological needs are not limited to a monolingual or at most bilingual communication between e-learning experts.
- TC37 has defined some methods starting from concepts and pointing in a second level at languages registries and at a third level at terms.
- A terminology is constituted elaborating at first a concepts system that allows answering multilingual terminological needs without blooming out le linkage system that ties terms to concepts, this would allow elaborating a common terminological basis that supports a large variety of languages.

The needs for conceptual expressions through fields of interest and through specialization levels of experts or through fields of cultural and linguistic specifications.

- Experts need master non only terms within a variety of national languages or specialized terminology; they need to come to a consensus about concepts systems which definitions cope with different items they are interested in.
- These items could be components, processes, individuals, institutions, etc.
- The formalization of concepts systems that constitute the first step of elaborating a TC37 terminology, facilitate the consensual agreement about these items and notions.
- Experts groups, thus naturally develop sub-sets of the conceptual system corresponding with for instance such or such WG.
- The importation of pre-existent terminology or structured thesaurus would certainly enrich this first approach.

Needs for concepts systems for MLR development

- The development of MLR within WG4 requires a mastering of metadata in terms of concepts and concepts domains.
- The TC37/SC3 standardized terminologies offer large facilities since they generate at first level concepts systems.

Needs for concepts cartographies and ontologies to master :

1- *a technical and conceptual global vision*

- Beyond MLR development, the SC36 enlarged experts community needs a synoptic vision of the different items which characterize ICT for education and training as such as the professional and usage processes associated to them.
- It also needs lexicons not only built in a limited or extended set of languages, but requires also concept systems articulated between them and describing a large variety of items on which consensus should be already taken as well as interoperability elements which will generate the different assets of the standard.
- In other words, SC36 standard will be constructed either by conceptual analysis or by the confrontation of technical and logical Sector-based specifications.

2- prospective and governance of standards development

- Since the termino-semantics method is applied, it would conceptually help master global and detailed technical domains (components and processes), professions and usages that concern us.
- In addition, the global vision of the conceptual system organized in cartographies and ontologies facilitate establishment of a development program at medium and long terms (prospective vision) and by the mean time, contribute to a good standards governance.

COMPLEMENTARITY BETWEEN TERMINOLOGY (PRODUCED FROM CONCEPTS SYSTEMS-TC37), CONCEPTS CARTOGRAPHIES (TOPIC MAPS) AND ONTOLOGIES

- TC37 terminologies allow fit out a relatively Nearsighted vision of the conceptual universe; however, the mechanisms of a terminological data base allow instantiate terms and definitions in different languages .
- Ontologies are built essentially on basis of standardized and interoperable terminologies. They have the potential to manage complex and rich relations between contents introducing semantics into the relations between objects.
- Cartographies of concepts (Topic Maps) allow representation of ill-assorted subjects, their relations, their organization to federate and have access to distributed contents within networks. Topic Maps allow cohabitation of different Topic Maps on the same contents in order to reflect needs and the different pints of views.

We will treat successively :

1. SC37/SC3 standardized terminologies ;
2. Ontologies and Topic Maps

PRESENTATION OF A STANDARDIZED METHODOLOGY FOR THE SC36

Reviewed and augmented Document

Working Group 1
Kiev
25-26 May 2003

Henri HUDRISIER

in collaboration with Annie MARCHEX, Odile ARTUR, Nadine LUCAS, Laurent ROMARY and Rachid ZGHIBI

NORMATIVE BASES ON WHICH TO LEAN ON AND WHY

1. Standard ISO 704 : Principles and methods (in terminology)
2. Standard ISO 12620 : Data categories
3. Standard ISO 16642 TMF :Terminology Markup Framework

STANDARD ISO 704 : Principles and methods (in terminology)

Is used as reference for elaborate the terminology of SC36 and to structure a network of concepts in each under-field already given (8 categories Kiev 2002) to define the types of relations (hierarchical, partitive, and relational) existing between these concepts (cf in ISO704: introduction: principal activities).

The main activities include, but are not limited to the following:

- **identifying concepts and concept relations;**
- **analyzing and modeling concept systems on the basis of identified concepts and concept relations;**
- **establishing representations of concept systems through concept diagrams;**
- **defining concepts;**
- **attributing designations (predominantly terms) to each concept in one or more languages;**
- **recording and presenting terminological data, principally in print and electronic media (terminography).**

Objects, concepts, designations and definitions are fundamental to terminology and therefore form the basis of this International Standard. Objects are perceived or conceived and abstracted into concepts which, in special language, are represented by designations and described in definitions. A set of designations belonging to one special language constitutes the terminology of a specific subject field.

STANDARD ISO 12620 "Data categories"

Is used as reference to determine the data elements which will have to be established for each concept retained for work terminographic in coherence with TMF (Terminological Markup Framework) and what is useful for our terminology of the SC36.

STANDARD ISO 16642 TMF :
TERMINOLOGY MARKUP FRAMEWORK

is used as reference to determine the structure which can be retain to record work terminographic for the SC36. This standard is based both on ISO 12620 for determinate terminogical data categories and on ISO 11179 for determinate a common structure of metadata to describe terminogical data categories.

ISO 704 : PRINCIPLES AND METHODS

DEFINITION OF CONCEPT SYSTEMS

The *terminology* of a field shall not be an arbitrary collection of *terms.* The *terminology* of a *subject field* is the collection of *designations* attributed to *concepts* making up the knowledge structure of the field. The *concepts* shall constitute a coherent *concept system* based on the relations established between *concepts*.

The unique position of each *concept* within a system is determined by the *intension*, i.e. the unique set of *characteristics* constituting the *concept*, and the *extension*.

ISO 704 : PRINCIPLES AND METHODS

A concept system serves to:

- ➔ model concept structures based on specialized knowledge of a field;
- ➔ clarify the relations between *concepts*;
- ➔ form the basis for a uniform and standardized *terminology*;
- ➔ facilitate the comparative analysis of *concepts* and *designations* across languages;
- ➔ facilitate the writing of *definitions*.

ISO 704 : PRINCIPLES AND METHODS

1- TYPES OF CONCEPT SYSTEMS

➔generic concept system: a system in which all the *concepts* in a vertical series relate to each other as *generic* and *specific concepts* ;

➔ partitive concept system: a system in which all the *concepts* in a vertical series relate to each other as a whole and its parts ;

➔ associative concept system: a system in which all the *concepts* relate to each other by association. The type of *associative relation* between any two *concepts* may vary within a system.

ISO 704 : PRINCIPLES AND METHODS

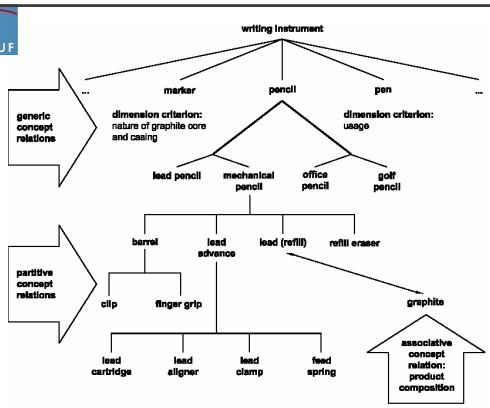
2- DEVELOPING CONCEPT SYSTEMS

➔ selecting the *concept field*, the preliminary *designations* and *concepts* to be treated by taking into account the *subject field*, the user group and its needs ;

➔ analyzing the *intension* and *extension* of each *concept*;

➔ determining the relation and position of these *concepts* within the *concept system* ;

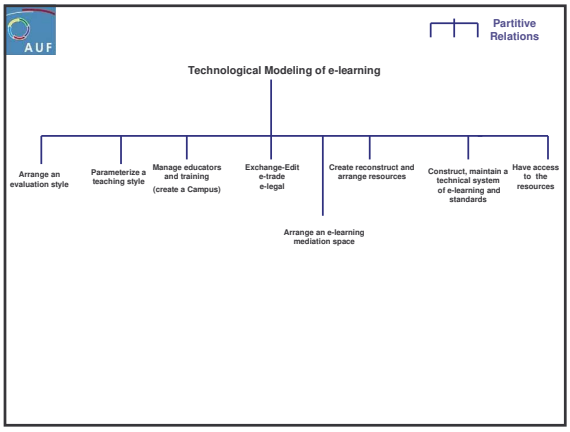
➔ formulating and evaluating *definitions* for the *concepts* based on the concept relations ;

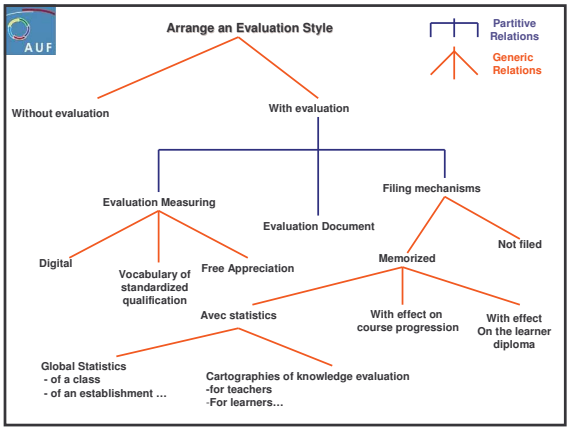


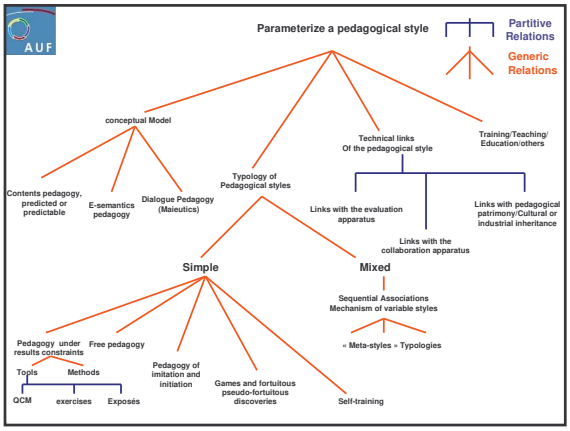
ACKNOWLEDGMENT

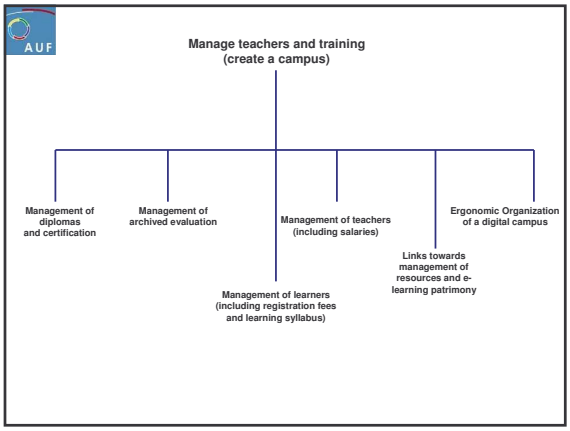
The following initiative to construct a concept systems for SC36 e-learning is very elementary.

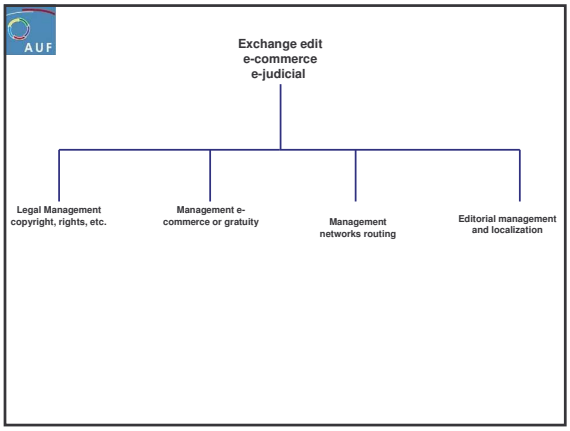
The following graphs are somehow provocative and are intended to arouse critics and accelerate a consensual elaboration of a real concept network of the domain

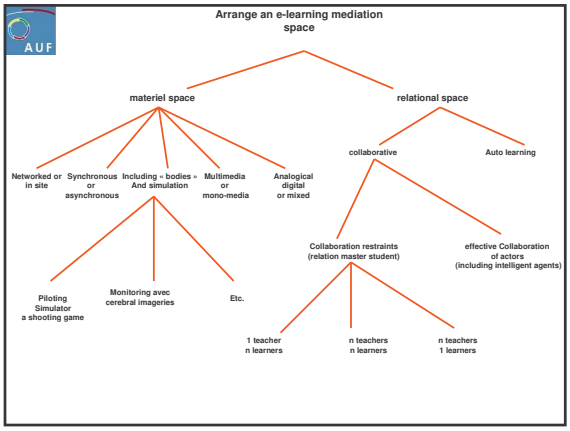

AUF
Partitive Relations
Technological Modeling of e-learning
Arrange an evaluation style
Parameterize a teaching style
Manage educators and training (create a Campus)
Exchange-Edit e-trade e-legal
Create reconstruct and arrange resources
Construct, maintain a technical system of e-learning and standards
Have access to the resources
Arrange an e-learning mediation space


AUF
Arrange an Evaluation Style
Partitive Relations
Generic Relations
Without evaluation
With evaluation
Evaluation Measuring
Evaluation Document
Filing mechanisms
Digital
Vocabulary of standardized qualification
Free Appreciation
Memorized
Not filed
Avec statistics
With effect on course progression
With effect On the learner diploma
Global Statistics
- of a class
- of an establishment …
Cartographies of knowledge evaluation
-for teachers
-For learners…


AUF
Parameterize a pedagogical style
Partitive Relations
Generic Relations
conceptual Model
Technical links Of the pedagogical style
Training/Teaching/ Education/others
Contents pedagogy, predicted or predictable
E-semantics pedagogy
Dialogue Pedagogy (Maieutics)
Typology of Pedagogical styles
Links with the evaluation apparatus
Links with pedagogical patrimony/Cultural or industrial inheritance
Links with the collaboration apparatus
Simple
Mixed
Sequential Associations Mechanism of variable styles
« Meta-styles » Typologies
Pedagogy under results constraints
Free pedagogy
Tools
Methods
QCM
exercises
Exposés
Pedagogy of imitation and initiation
Games and fortuitous pseudo-fortuitous discoveries
Self-training


AUF
Manage teachers and training
(create a campus)
Management of diplomas and certification
Management of archived evaluation
Management of learners (including registration fees and learning syllabus)
Management of teachers (including salaries)
Links towards management of resources and e-learning patrimony
Ergonomic Organization of a digital campus


AUF
Exchange edit
e-commerce
e-judicial
Legal Management copyright, rights, etc.
Management e-commerce or gratuity
Management networks routing
Editorial management and localization


AUF
Arrange an e-learning mediation space
materiel space
relational space
Networked or in site
Synchronous or asynchronous
Including « bodies » And simulation
Multimedia or mono-media
Analogical digital or mixed
collaborative
Auto learning
Piloting Simulator a shooting game
Monitoring avec cerebral imageries
Etc.
Collaboration restraints (relation master student)
effective Collaboration of actors (including intelligent agents)
1 teacher n learners
n teachers n learners
n teachers 1 learners

ISO 12620 STANDARD "DATA CATEGORIES"

TYPOLOGY OF DATA CATEGORIES

The data category specifications are divided into three major groups:

- data categories for terms and term-related information ;
- descriptive data ;
- administrative data.

The groups are further subdivided into ten sub-groups.

STANDARD ISO 12620 "DATA CATEGORIES"

GROUP 1 : DATA CATEGORIES FOR TERMS AND TERM-RELATED INFORMATION

Subgroup 1 consists of the data category term and contains a term or other information treated as if it were a term (e.g., phraseological units and standard text).

Subgroup 2 specifies data categories for term-related information.

Subgroup 3 specifies data categories for information relating to equivalence between or among terms assigned to the same or very similar concepts.

GROUP 2 : DESCRIPTIVE DATA CATEGORIES

Subgroup 4 specifies data categories for the classification of concepts into subject fields and subfields, along with other classification-related information.

Subgroup 5 specifies data categories for concept-related description, i.e., different kinds of definitions, explanations and contextual material provided to define or otherwise determine the subject field and concept to which a term is assigned.

Subgroup 6 specifies data categories for indicating relations between pairs of concepts.
Subgroup 7 specifies data categories used to express the position of concepts within concept systems.

Subgroup 8 specifies the data category *note*. This category stands alone because it can be associated with any one of the other categories and therefore cannot be subordinated to any other specific subgroup.

GROUP 3 : ADMINISTRATIVE DATA CATEGORIES

Subgroup 9 specifies data categories for documentary languages and thesauri ;

Subgroup 10 specifies data categories for all other strictly administrative information.

GROUPE 1 Data categories link to the term *Categories du terme et de l'information liée aux données concernant terme*		
Subgroup 1 consist of data category term and contains a term or a syntagm ***Le sous-groupe 1 est composé du terme de la catégorie de données et contient un terme ou une autre information traitée en tant que terme (par exemple : les syntagmes ou le texte normalisé).***		
A.1 Term *Term* <the name of a term>		
Subgroup 2 is used for data categories link to the term. ***Le sous-groupe 2 porte sur les catégories de données qui fournissent de l'information liée au terme.***		
A.2 Information link to the term ***Information*** *concerning the term*	A.2.1Term typology *Type of term*	A.2.1.1Entry term A.2.1.2 Synonym A.2.1.3 Quasi-synonym A.2.1.4 international scientific term A.2.1.5 common name A.2.1.6 Internationalism A.2.1.7 Full form A.2.1.8 Abbreviated form of term A.2.1.9 Variant A.2.1.10 Translated form A.2.1.11 Transcribed form A.2.1.12 Romanized form A.2.1.13 Symbol A.2.1.14 Formula A.2.1.15 Equation A.2.1.16 Logical expression A.2.1.17Materials management categories A.2.1.18 Syntagm

	A.2.2 Grammar	A.2.2.1 Part of speech A.2.2.2 Grammatical gender A.2.2.3 Grammatical number A.2.2.4 Animacy A.2.2.5 Noun class A.2.2.6 Adjective class
	A.2.3 Usage	A.2.3.1 Usage note A.2.3.2 Geographical usage A.2.3.3 Register A.2.3.4 Frequency A.2.3.5 Temporal qualifier A.2.3.6 Time restriction A.2.3.7 Proprietary restriction
	A.2.4 Term formation	A.2.4.1 Term provenance A.2.4.2 Etymology
	A.2.5 Pronunciation	
	A.2.6 Syllabication	
	A.2.7 Hyphenation	
	A.2.8 Morphology	A.2.8.1 Morphological element A.2.8.2 Term element
	A.2.9 Term status	A.2.9.1 normative authorization A.2.9.2 Language planning qualifier A.2.9.3 administrative status A.2.9.4 process status
	A.2.10 Degree of synonymy	

Le sous-groupe 3 porte sur les catégories de données concernant l'équivalence entre les termes affectés aux mêmes notions ou à des notions très proches.		
A.3 Equivalence	A.3.1 Degree of equivalence A.3.2 False friend A.3.3 Directionality A.3.4 reliability code A.3.5 Transfer comment	
GROUP 2 : Descriptives data categories *Catégories de données descriptives*		
Sub-group 4 specifies data categories for the classification of concepts into subject fields and sub-fields, along with other classification-related information *Le sous-groupe 4 porte sur les catégories de données concernant la classification des notions en domaines et en sous-domaines et sur d'autres informations touchant à la classification.*		
A.4 subject field (8 categories)	A.4.1 classification system A.4.2 classification code (number)	
Sub-group 5 specifies data categories for concept-related description , i.e. different kinds of definitions, explanations and contextual material provide to define or otherwise determine the subject field and concept to which a term is assigned.		

A.5 Concept related-description	A.5.1 Definition A.5.2 Explanation A.5.3 Context A.5.4 Example A.5.5 Non textual illustrations A.5.6 Unit (perhaps for notation) A.5.7 Range A.5.8 Characteristics	
Sub-group 6 specifies data categories for indicating relations between pairs of concepts		
A.6 Concept relations	A.6.1 generic Relation	
	A.6.2 partitive Relation	
	A.6.3 sequential Relation	A.6.3.1 temporal Relation
		A.6.3.2 spatial Relation
	A.6.4 associative Relation	
Sub-group 7 specifies data categories used to express the position of concepts within concept system		

A.7 Notional Structures	A.7.1 Notions System	
	A.7.2 Notion Position	A.7.2.1 Extended Notion A.7.2.2 Superordinated Notion A.7.2.3 Subordinate Notion A.7.2.4 Coordinate Notion A.7.2.5 Connected Notion
Sub-group 8 specifies the data category Note. This category stand alone because it can be associated with any one of other category and there fore cannot be subordinated to any other specific sub-group		
A.8 Note		
GROUP 3 : Administrative data categories		
Sub-group 9 specifies data categories for documentary languages and thesaurus		

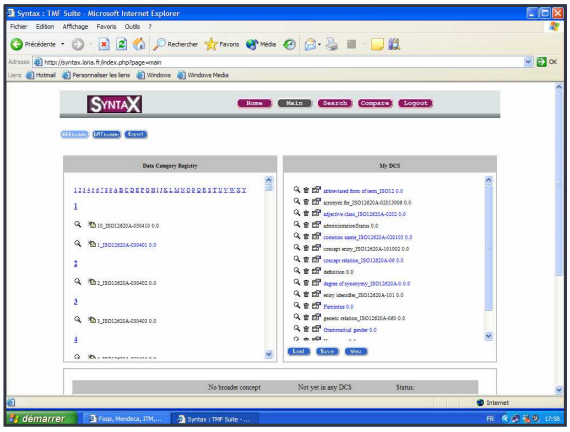
A.9 Documentary Langage	A.9.1 Thesaurus Name	
	A.9.2 Thesaurus Descriptor	A.9.2.1 The most generic Descriptor A.9.2.2 Generic Descriptor A.9.2.3 Restricted term A.9.2.4 Related Term
	A.9.2 Non-descriptor	
	A.9.3 Key-word	
	A.9.4 Index Title	
Sub-group 10 specifies data categories for administrative data categories		
A.10 Administrative Category	A.10.1 Transaction of terminological work	
	A.10.2 Function related with terminological management	A.10.2.1 Date A.10.2.2 Responsibility
	A.10.3 Administrative Subsets	A.10.3.1 Subset/client A.10.3.2 Subset/initial client A.10.3.3 Subset/project A.10.3.4 Subset/initial project A.10.3.5 Subset/product A.10.3.6 Subset/application A.10.3.7 Subset/ environment A.10.3.8 Subset/commercial A.10.3.9 Subset/security

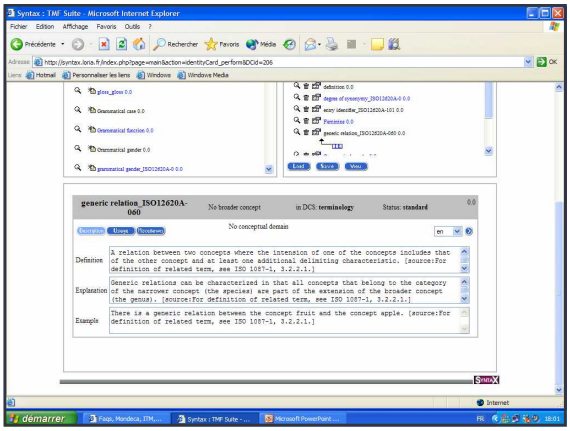
	A.10.4 Information concerning permission	A.10.4.1 Authorization Function A.10.4.2 Authorization Code A.10.4.3 Authorization password A.10.4.4 Professional Title
	A.10 5 User Suggestion	
	A.10.6 Administrative designations	A.10.6.1 consequential Term A.10.6.2 terminological sorting key A.10.6.3 Research Term
	A.10.7 Language indicator	
	A.10.8 External Text	
	A.10.9 Filing Sequence	
	A.10.10 Entry Type	
	A.10.11 Element Status	
	A.10.12 Forwarding Data base	
	A.10.13 Entry source	
	A.10.14Notion identifier	
	A.10.15lEntry identifier	
	A.10.16 File identifier	
	A.10.17 File identifier	
	A.10.18 Forwarding	A.10.18.1 Forward of type "See" A.10.18.2 Forward of type "See also" A.10.18.3 Inverted term A.10.18.4 Permuted term A.10.18.5 Homograph A.10.18.6 Antonym

	A.10.19 Source	
	A.10.20 Source Identifier	
	A10.21 Identifier	A.10.21.1 URL A.10.21.2 FPI
	A.10.22 Origin	A.10.22.1 Origin-person A.10.22.2 Origin-institution A.10.22.3 Origin-Data base

Handling of this data categories is complex.

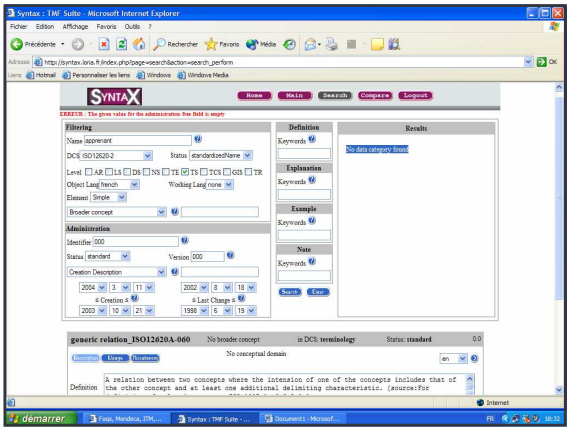
LORIA (Laboratoire Lorrain pour la Recherche en Informatique et Automatisation : France) has developed the SYNTAX software that allow managing these data categories and notably constituting « nodes of data categories » to differentiate usages in different languages.





Exemples d'aménagement des catégories de données en bouquets spécifiques adaptées selon les langues : anglais, français et arabe.

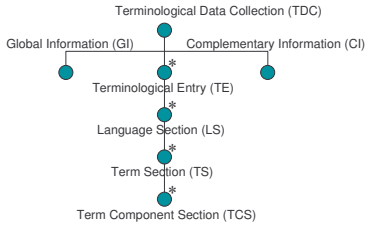
A.2.2.2 : grammatical case			
General	English	French	Arabic
Masculine	Boy	Logiciel	برنامج
Feminine	Girl	Base de données	قاعدة بيانات
Neutral	Software	-	-
Others			
A.2.2.3 : Grammatical number			
Singular	Professor	Professeur	أستاذ
Plural	Professors	professeurs	أساتذة
Duel	-	-	أستاذان
unnumbered			
Others			



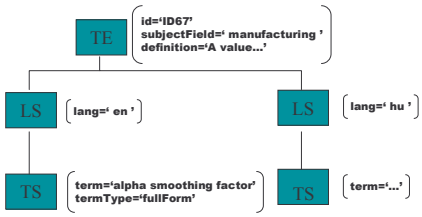
STANDARD ISO 16642 TMF
TERMINOLOGY MARKUP FRAMEWORK

- Although the terminological methodology has been standardized, through ISO 704 and ISO 12620 standards, terminological instances while respecting these standards have developed incompatible terminological basis.
- To respond to future semantic Web stakes, it became urgent to develop a permanent standard allowing this interoperability. They concur in that TMF which is based (like MLR) upon ISO 11179 standard.

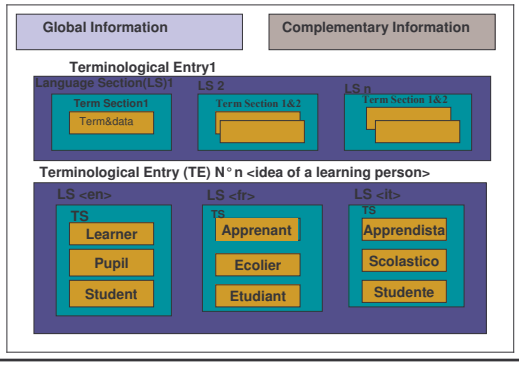
TMF : META-MODEL



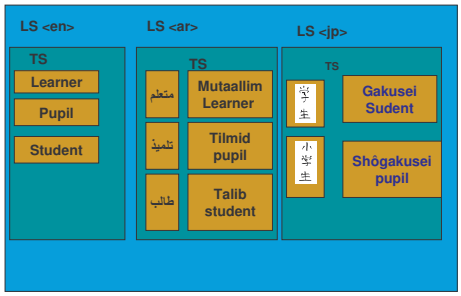
TMF : INFORMATION MODEL



Terminological Data Collection



Terminological Entry (TE) N° n <idea of a learning person>



TOPIC MAPS AND ONTOLOGIES

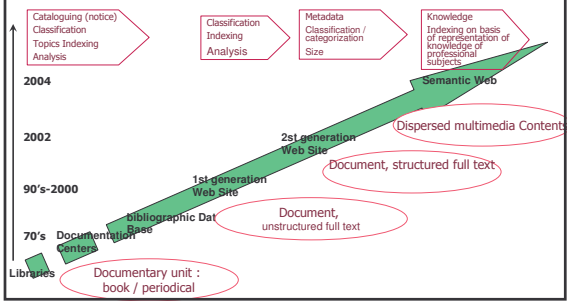
Prepared by

Henri HUDRISIER, Jean DELAHOUSSE, Rachid ZGHIBI, Mokhtar BEN HENDA, Laurent ROMARY

Ontologies and Topic Maps are complementary tools that aim at giving a more global vision than terminologies and concepts systems.

If I were asked to put order in the empire, my first concern would be to reestablish the meaning of words.
Confucius

Definitions : contents structure and architecture



Ontology

- ✓ Basic structured vocabularies, validates by all actors of a community giving a non ambiguous sense to published information and knowledge.
- ✓ develop, maintain, improve information and knowledge organization which respond to the world complexity, to the diversity of points of view, and to the rapid growth of knowledge.
- ✓ determine and use relations and logic rules between concepts, permitting an efficient use of intelligent agents.

Ontology can include :

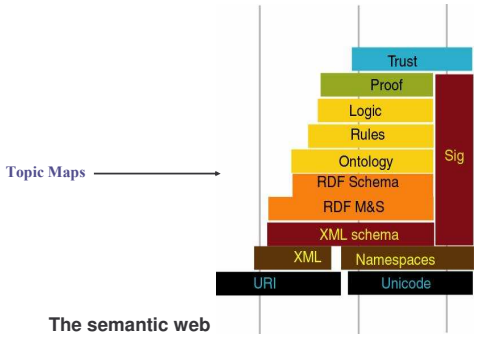
1. *Objects classes to organize* (projects, persons, products, commercial documents, contracts...),
2. *Type of attributes that can be attached to objects* (reference, description, adresses, size, geographic situation...)
3. *Types of relations between objects* (an object "person" can be related through a relation "employed-by" to an object of the type "organization").

Works on ontologies respond to many need :

• to dispose of basic structured and validated vocabularies used by all the actors of a given community, giving a non ambiguous sense to the published information and knowledge contents ;

• determine and use relations and logic rules between concepts, permitting an efficient use of intelligent agents;

• develop, maintain, improve information and knowledge organization which respond to the world complexity, to the diversity of points of view, and to the rapid growth of knowledge.



The semantic web
W3C

What are Topic Maps ?

Topic Maps concept is the result of a one decade patient work thinking about the problematic of documentation indexes, of information organization and knowledge representation. Beyond the concept, realized works within the frame of ISO, and of TopicMaps.Org working group conducted to an international standard : ISO/IEC 13250. This standard contains a non official description of the object type that constitute a Topic Map, as well as an XML representation and exchange syntax (XTM 1.0). Current works within ISO context are related to a standardized application model, an abstract reference model, a query language (TMQL) and a constraint (TMCL) : a language which is not intended to be free, but controlled by ontology.

Topic Maps basics

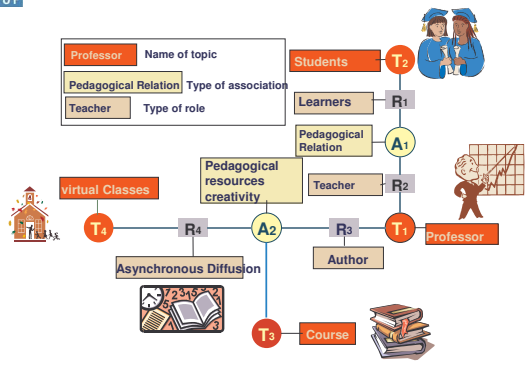
A Topic Map, formatted as an XML file or treated by a specialized software, has as objective to :

1. Afford an organized set of Topics (formal representation of well identified subjects),
2. Afford a set of link to documentary resources indexed by these Topics,
3. Afford semantic relations between these topics, expressed as Associations where each Topic plays a specific role.

A Topic Map is then an infrastructure permitting the description and organization of information and the access to them, including different mechanisms allowing the enrichment of this index of knowledge :

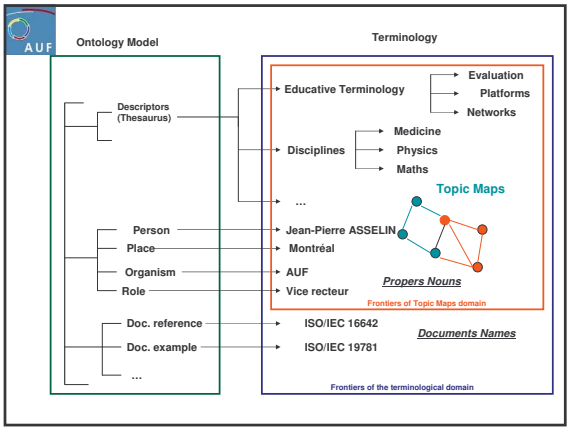
- Management of many names for each Topic
- Management of hits (occurrences) relating Topics and external indexed documents
- Management of semantic and hierarchical (or not) relations between Topics via Associations
- Simple or multiple classification of Topics by other Topics
- Filtering mechanism (Scope) of topics characteristics (names, associations, occurrences) allowing to filter and personalize indexes accordingly to users
- Universal Identification mechanism of Subjects (PSI or Published Subject Indicators) allowing the unique identification of a Subject represented within different Topic Maps
- Capacity of exchanging and merging Topic Maps.

An example of Topics modeling



Three standards gather today the Semantic Web Community :

- Terminologies (ISO/TC37),
- Ontologies (W3C/ OWL) ;
- Topic Maps (ISO/IEC/SC34: FCD 13250)



CONCLUSION

Identifying termino-semantic needs of SC36 is a first step. It is also necessary to get organized in order to profit as much as possible for the e-learning standards development. The intervention of Laurent Romary as a representative liaison of TC37 during the Paris plenary gives a quite complete panorama of the interest that termino-semantic standards have for e-learning. They could be classified into three domains corresponding to 3 types of preoccupations :

- The development of terminologies as properly stated ;
- The macroscopic organization of concepts systems in standardized cartographies (Topic Maps) or in developed ontologies within a standardized frame compatible with semantic Web (OWL) ;
- The termino-semantic assistance at different projects poles of SC36 :
  1. The conceptualization of metadata (WG4) ;
  2. The global vision of the domain (marketing Group and SC36 Governance).