

Author

Bernard Vatant [1] is an independent Knowledge Representation Consultant, with a background in Mathematics teaching and Astronomy popularization. He's presently working in the frame of **Mondeca Topic Navigator** [2] development and applications.

Abstract

The Topic Maps standard, as expressed in XTM 1.0 specification [3], contains no explicit guidelines for effective building of consistent and meaningful TM. This paper proposes a general methodology for Topic Map building and management in complex information environments involving a great diversity of resources, objects, concepts and actors. The methodology focuses on the distinction between "objective" levels of identified Objects and Resources, and "conceptual" level, and proposes ways for structuring, linking and managing these different levels.

The question of Topics identity and naming is addressed, as well as the problem of mixing hierarchical structures and non-hierarchical relationships, at both objective and conceptual levels, using the concept of entangled hierarchies.

Besides technical aspects, management of human expertise, definition of administration and authoring levels, are a crucial part of consistent development and updating. Topic Maps applications can tackle these issues with collaborative multi-level tools, using scopes for filtering.

An use case is presented : Mondeca Topic Navigator is the management tool for a collaborative resource data base for the Semantopic Universe [4] : people, organizations, tools, projects and events involved in the development of Topic Maps, Semantic Web and other relative knowledge technologies.

Contents

Introduction : Managing Complexity

1. Survey of Resources

2. Identification of Objects

3. Choice of Concepts : Reference Ontologies

4. Definition of Classes and Types : Entangled hierarchies

5. Building Knowledge by multi-level Semantic Associations

6. Human Management : Administration, Authoring and Updating

7. Mondeca Topic Navigator documentation and the Semantopic Universe

8. Perspectives

* Notes and Resources

Introduction : Managing Complexity

Complexity is the main issue of many information and knowledge environments, and the Web itself, chaotic representation of the real world, seems to achieve the complexity paradigm. Many definitions of complexity have been proposed, from the physicists theories of non-linear phenomena, to linguistics and social sciences descriptions. [5] From an information and knowledge management viewpoint, a definition of complexity will include at least the following characteristics:

- ✱ Abundance and heterogeneity (diversity of formats) of information resources
- ✱ Presence of many multidimensional subjects and objects, multi-connected to this information
- ✱ Variety of expertise and responsibility levels of authors, administrators, users of this information

The Topic Map approach provides tools to tackle these specific issues. Managing complexity is in fact what Topic Map technology is all about. The proposed methodology is therefore mainly addressing representation and management of information in such complex environments. Corporate R&D or Administration Intranets, Information or Business Web Portals are some among many application fields of such a methodology, not to mention the potential development of educational tools in various fields, from Social Sciences to Biology, Health and Environment Technologies, wherever classical knowledge representations are unable to manage the complexity issues.

1. Survey of Resources

Trying to build a Topic Map from scratch, as a representation of the complex real world - or at least the part of it the TM engineer is interested to represent - will lead most surely to hard and long ontological debate about the nature of inhabitants of this "real world". A more pragmatic approach is to consider the set of all available information sources - addressable resources - as a satisfying proxy, given it generally already contains many representations of, and documentation on just about every object or subject in the environment. The very abundance of these resources and the need to manage them is in fact the main concern of the TM engineer in complex and multidimensional information environments. A survey and good knowledge of what is available and pertinent, what is stable or volatile at this ground floor, should be therefore the first fundamental task for the TM engineer.

Among those resources, a distinction of status is to be made on the basis of permanence and stability of both addresses and contents. A resource considered as permanent both in address and content, e.g. a company Home Page, a specification or standard reference URL, etc. should be considered as the subject of a topic by itself. In the Topic Map terminology, such stable resources will be called "addressable subjects". Such subjects present the great quality to be identified on a non-ambiguous basis, and offer strong grounding for sharable information. Survey of stable identified resources is therefore the ground floor for Topic Map representation of a complex environment.

More volatile resources will be attached as non-permanent occurrences of various topics in the Map, and will need proper updating process, either human or system managed. This point will be addressed later on in the human management section.

2. Identification of Objects

The second step following the above survey is : What are those Resources about? First they give information about **Identified Objects**, that is objects well defined by some non-ambiguous public identifier(s) : People, Companies, Trademarks, Catalog References, Laws and Regulations, Indexed Books or Musical Works, Records, Standards, Administrative and Geographical Subdivisions, past or present, known by some individual name, reference number in some catalog, index, directory.

For the semantic clarity, any such object should be defined as a Topic distinct from any stable resource about it, as defined above at stage 1. For instance, Mondeca – the company – should be represented as a Topic distinct from the Topic representing the website www.mondeca.com, because roles played in associations by these two Topics are not the same ones, and one would want to make distinct the information to be found on the website from any other information about Mondeca.

Many Objects are rather simple and unidimensional, such as referenced products, books, articles of law or regulation, and will be simply defined as "elemental topics". But a more complex and

multidimensional Object like e.g. a big company, a prolific author, a town, an university or research center, shows a variety of aspects, parts, characteristics, linked by inner relationships. That means such an Object should not, to begin with, be considered as a single topic. The best way to represent it is indeed a Topic Map of its many aspects, every one of them bound to be handled like a separate topic or association. A proper "reification" of the Topic Map will allow the represented object to be handled like a single topic at upper levels if necessary. This fractal approach [6] of the representation of Objects will avoid the creation of overloaded topics at any given level.

3. Choice of Concepts : Reference Ontologies

To represent knowledge about the above Objects and Resources, the TM engineer will need the vocabulary and ontology living at the conceptual level : types, categories, representations, ideas, relationships, whatever cannot be defined as an Object in the above defined acceptance. There is no absolute reference at that level, and construction of sharable concepts will always be the result of some agreement, inside a given community or context, on the sense of a given word or expression.

Given the potential source of ambiguity involved here, it's indeed highly recommended that Topic Map authors choose, as far as possible, subject indicators for the concepts they use in any public existing authoritative ontology, index, thesaurus, dictionary ... to ensure non-ambiguous, durable and sharable definitions. That does not mean that anyone of such resources should be an unique reference source for any universal definition of concepts. The choice of concepts should in fact keep away from those two dangerous extremes, the "chaotic" and "unique thought" ways.

In the "chaotic way", each TM author chooses its own concepts, its own glossary, its own resources, yielding very colorful but somehow non sharable Maps. This way should be used only by authors wishing to show a very personal view of the world, with the shortcoming that this Map will be quite impossible to merge with other ones in the same field. In the opposite "unique thought way", each TM author will try to reference to the same unique and universal ontology. No definite and universal consensus has ever been set for that matter, despite repetitive attempts from Aristotle's categories to more recent systems like Cyc Ontology [7] or Standard Upper Ontology [8], each of them getting its lot of enthusiastic supporters and fighting detractors. Construction of sharable ontologies will not be addressed further in this paper, but remains a crucial relative issue, for which new solutions using semantic tools are emerging. See Jack Park's paper in this same conference [9]

4. Classes and Types : Entangled hierarchies

Even if there is no universal ontology, among widely shared concepts is the hierarchical relation, appearing under various appellations : category/subcategory, set/subset, class/subclass, with their variants type/subtype, whole/part, parent/child ... These hierarchies are non ambiguous and mostly sharable when applied to sets of clearly identified and defined objects (or living creatures), and they are known there as taxonomies : Botanic and Zoology have produced the most famous and efficient examples of this method. But efficiency of a tool is always limited to a definite range. Pertinence of a taxonomic classification method is linked to some peculiar features of the application field.

Each individual object is an instance of an identified class of quite similar objects :

This wagtail in my garden is one among so many wagtails I would call the same name.

The identification is linked to the presence of characteristic properties :

The wagtail is a passerine with black hat, long wagging tail etc ...

The class/subclass relationship support inheritance of the above properties :

A "yellow wagtail" has all the properties of any wagtail, plus the needed yellow and green colors.

Hence the hierarchical classification, which has some reasonable claim to universal consistency, since all ornithologists and bird watchers will agree on it :

Living Creature > Animal > Bird > Passerine > Wagtail > Yellow Wagtail

Another consistent example will be - for those who care more about bikes than birdies :

Vehicles > Two-Wheeled > Motorcycles > Harley-Davidson > Electra Glide > 74 FLHB

In this example, it's interesting to note that "Harley-Davidson" is implicitly considered as a type of motorcycles, in a taxonomy quite similar to the ornithologist's. But now consider the following:

Companies > USA > Wisconsin > Milwaukee > Harley-Davidson

A more administrative view ... but consistent all the same ... There "Harley-Davidson" is considered as the company located in Milwaukee, Wisconsin, USA. The regional names are shorthand for "US Companies" "Wisconsin Companies" and "Milwaukee Companies". The two above hierarchies carry somehow completely independent viewpoints on the same subject. But is it the same subject when the biker asks : "Hey, where did you find that fantastic Harley-Davidson?", whereas the statistician or manager asks : "Where can I get the Harley-Davidson sales figures for the year 1999?" ... difficult question ... probably it's not really the same subject, but they are somehow aspects of the same one and could not be made distinct in some resources.

If the TM engineer is to decide anyway there should be a single Topic named "Harley-Davidson" in its Topic Map - and he has to do so if he wants to respect the so-called Topic Naming Constraint - he will have to find a way to make coexist these two views of the world. A proposed method is to let coexist the two above hierarchies, but to attach different scopes on the associations representing them in the Topic Map : a scope "bikes" for the first ones, and a scope "companies" for the second ones.

Another possible approach would be to consider "Harley-Davidson bikes" and "Harley-Davidson company" as different Topics, link them through an Association, and reify afterwards this association to create some upper level "Harley-Davidson" Topic. This second approach seems to be less natural and to carry more technical difficulties for further sharing and merging.

The above type of situation is more rule than exception in complex knowledge fields, and maybe could be pointed out as another characteristic of complexity itself. That's a main reason for the great difficulty, not to say the sheer impossibility, to represent this complexity by some unique hierarchical tree, which is yet the pretension of still widely used Dewey's Decimal Classification [10], or in a less dogmatic approach, but basically the same one, great Web directories such as Open Directory [11].

This leads us to the notion of entangled hierarchies : Every object (in the above defined meaning) and hence every Topic representing an object in the Map, is bound to be considered as the intersection of several hierarchical trees. As far as unity of hierarchy is not required, one can make a far more accurate use of hierarchical relationships, that is restricting them to strict taxonomic classifications of objects. Not forcing hierarchical associations where they can't fit will avoid the type of mismatched classification found in the above quoted directories. In fact, the definition of hierarchies at a purely conceptual level, well grounded in academic and documentation habits, may be completely avoided. Widely used so-called "top-level" categories like "entertainment", "environment", "health" or even more specific ones like "biodiversity" or "water management" are such examples of multidimensional concepts which should not be forced in any hierarchical classification.

5. Building Knowledge by multi-level Semantic Associations

Consistent Topic Maps may be limited to the above defined elements : representation of resources, identification of objects, and definition of classes and entangled hierarchies. In such representations, that one would maybe tempted to call "objective", concepts are only used for definition of types.

But the TM engineer will want, in most cases, to build a more refined knowledge representation, and for that include meaningful or semantic relationships. These relationships will use concepts to define roles in basically non-hierarchical associations, either between objects, or between concepts, or between assertions about the former or the later. Below are listed some out of many possible examples of these semantic associations, each of them creating a level of representation.

- ※ Concepts may be used to define non-hierarchical associations between objects, such as ...
- ... "partnership" to define the relationship between two companies
- ... "illustration" for the relationship between a book and its illustrator
- ... "usage" for the relationship between a research center and an instrument aboard a satellite
- ... or to take back again the above example, an association "starring" may link "Harley-Davidson", "Peter Fonda" and "Easy Rider" Topics in some Topic Map of American Popular Culture.

Those non-hierarchical associations between objects will make available new directions of navigation, orthogonal to the hierarchical ones. In fact, they are what will make the real different flavor for end users between a classical hierarchical directory and a Topic Map directory.

✱ Concepts may be used to define and relate system functions to their objects and actors . In the environment and information system, both humans and system parts play functional roles, by which they are related to each other, and to the concepts involved in these functions. Reification of each function by a conceptual Topic, and associations linking this Topic to people and/or system parts or tools managing the function, builds a functional level of representation.

✱ Semantic associations may be used to structure a dictionary of concepts used in the Topic Map. In a dictionary, terms both define and are defined by one another in a recursive way, and in fact, dictionaries are resources ready-made to be represented as Topic Maps through “definition” and “example” associations, and are one of many promising field for TM development. This level may be useful for proper definition of technical terms in a defined context. It may be an extract from some existing dictionary, reorganized in a TM form, or a built-in proper resource. That’s what has been done for the Mondeca Topic Navigator Help, in which a “micro” Topic Map has been built to present the Topic Maps terms and concepts, and functions of the software.

✱ At an upper level, semantic associations may be used to implement viewpoints. An association being reified, that is being made a subject of a Topic, further associations can represent viewpoints on this subject : agreement, discussion, commentary, expanding, refutation of any of those etc. In our biker example, the association “Easy Rider” – “Peter Fonda” – “Harley-Davidson” , being some sort of “cultural subject” in itself, may be associated to movie critics, sociology reviews articles ... through such associations. This type of level would be very useful in the TM representation of on line Forum archives, allowing a more effective search for viewpoints than the classical “thread” organization.

6. Human Management : Administration, Authoring and Updating

The problem of occurrences selection and updating may present very different features, depending on the scope of the represented environment and system. Hence the methodology should adapt to these differences, the main one being between closed and open systems.

In a closed system like e.g. a corporate intranet, where all occurrences of Topics will be internal resources, TM administrator(s) should get some control or at least get streaming information on what is going on at the ground level of resources, such as who is responsible of which resources, on what time basis these resources are updated, what are the foreseeable changes in the system addresses structure etc. If a TM structure is chosen to represent such an intranet, the final objective should be to implement in the system capacities for the TM to take into account whatever changes occur in the resources, with minimal human intervention, in the ideal case confined to the system upper management. Updating of corporate resources at ground level should not imply extra work for the TM management team, as far as the addresses structure of Intranet resources are not changed. Choosing a TM management should even lead to more system stability, and “semantic” definition of resources.

In an open system such as a web portal, updating of occurrences is a more tricky work, and will need more human investment. The success of the Open Directory Project [11], reaching the first rank among Web Directories, both in quantity and quality of content by involving in barely two years more than 30,000 contributors to manage more than 300,000 categories, yields fairly good evidence that as far as surveying the Web for pertinent and updated resources in a given field, “Humans do it better”. That comment on the success of ODP does not change the above remarks on the incapacity of its very hierarchical structure to deal with complexity. But the generalization of RDF [12] or other kind of semantic metadata [13] may change the problem in the future. The vision of future Web portals will certainly include large Topic Map structures linked to powerful semantic query engines.

7. Mondeca Topic Navigator Documentation and the Semantopic Universe

The above methodology has been evaluated and refined along with development of some Mondeca Topic Navigator applications. Two fairly different situations have been addressed :

*** Mondeca Topic Navigator Documentation**

This is an example of somehow simple and closed environment. The objective level is there limited to resources created on purpose, in the form of plain HTML pages, each describing either a general concept in Topic Maps terminology, or a particular concept used by the tool, or software function, either an example illustrating one of the above. Each resource is represented one-to-one by a Topic in the Topic Map.

Three levels of associations have been created :

“Definition” associations link concepts to concepts, like in a dictionary.

“Management” associations link functions to the concepts they manage.

“Illustration” associations link examples to the concepts they illustrate

To link this documentation to a more complex use case, examples have been chosen in the following Semantopic Universe

*** The Semantopic Universe**

This is a very open environment : the goal is to build a representation and communication tool for “what is going on” in the loosely defined community working towards the “new web paradigm”. This has been summed up in the title, “universe” meaning this environment is open and hopefully in expansion, and the term “semantopic” forged out a coalescence of “semantic” and “topic”.

At the objective level, we have people, organizations, tools, projects and events. In fact, the two later types are considered as associations of the three former, since a project or event is basically a meeting of people, organizations and tools. Apart from those and the “natural” hierarchical relationships, semantic associations have been created such as : “partnership” between people and/or organizations, “usage” or “development” between tools and users and/or developers.

“Conceptual subjects” are whatever the above “objects” deal with, search on, discourse about ... They will be used to eventually filter the TM for end users, by using them as scopes attached to the above objects and associations.

The Mondeca software allows a collaborative management of this Topic Map, which means every interested actor can manage on-line, as an editor, the part of it concerning the topics and associations he/she is involved or has expertise in. Remember at this level : “Humans do it better” ...

Apart of its sheer usefulness as a tool for the involved community, it will be an opportunity to evaluate the capacity of the TM technology to be widely understood and used in collaborative projects.

8. Perspectives

Developed first as a tool for knowledge representation and resources indexing, the Topic Map technology seems bound to become a more pervasive tool, pertinent wherever one needs to tackle complexity. Given the technology is still young, it has to be confronted to a variety of use cases in order to be refined and validated. The present applications development has focused mainly on resources indexing and somehow static knowledge representation. The capacity of the paradigm to get to a really central role in open complex systems will certainly be linked to the development of interactive tools allowing effective human management in dynamic collaborative environments, including the upper level of discourse associations such as agreement, disagreement, example, development, perspective, generalization ... That will be the next important challenge to consider, at a time where building collective intelligence, corporate or community knowledge, are considered to be among the main economic and social issues for the years to come. Building such tools will need there again specific implementations, but the strong Topic Map conceptual core seems in good position to play a central paradigm role in those predictable developments.

Knowledge Technologies 2001 ~ March 4 - 7 ~ Austin, Tx
Managing complex environments with Topic Maps

Notes :

- [1] Bernard Vatant home page : <http://www.universimmedia.com>
- [2] Mondeca Web Site : <http://www.mondeca.com>
- [3] XTM specification : <http://www.topicmaps.org/xtm/>
- [4] The Semantopic Universe : <http://www.universimmedia.com/semantopic.htm>
- [5] Resource links on complexity : <http://www.brint.com/Systems.htm>

- [6] Resource links on fractals : <http://forum.swarthmore.edu/library/topics/fractals/>
- [7] Cycorp Ontology : <http://www.cyc.com>
- [8] Standard Upper Ontology, an IEEE project : <http://suo.ieee.org>
- [9] Bringing Knowledge Technologies to the Classroom : <http://www.thinkalong.com/JP/KT2001.pdf>
- [10] Dewey's Decimal Classification : http://www.oclc.org/dewey/about/about_the_ddc.htm

- [11] Open Directory Project : <http://dmoz.org>
- [12] Resource Description Framework : <http://www.w3.org/TR/REC-rdf-syntax>
- [13] See for instance the Dublin Core Initiative : <http://purl.org/dc/>