

Babel And Topic Maps



Illustration 1 "The Tower of Babel" Gustave Dore, 1866

Patrick Durusau, Patrick@Durusau.net

Introduction

The “how” of topic maps has been explained in at least one book¹, numerous presentations², and is the principal subject of several websites.³ Most of those explanations presume a particular “flavor” of topic maps and all fail to explain the “why” of topic maps.

The “why” of topic maps involves more than “Indexes are hard to combine automatically,” or “It is easier to arrive at a particular location if you have a good map,” or even “If I had a better glossary or thesaurus, I might understand what you are saying.” These are good issues, but solutions to them alone are hardly enough to justify the excitement seen in topic map discussions.

Diversity, in all its enriching and frustrating glory, is the basic “why” of topic maps. Historical perspective on diversity can help us make rational judgments about whether topic maps fit a particular situation. Or what sort of topic map works best in a given situation. Or claims made about the “semantic web,” “ontologies,” “reification,” “topic maps,” and other terms in the “fog of marketing.”⁴

Quite a bit of historical perspective, it turns out. We begin the story in Babel, a city in Mesopotamia, some 6,000 years ago.

1 XML Topic Maps, Parks and Hunting, eds., Addison-Wesley, 2003

2 See, www.coolheads.com (under Presentations), www.ontopia.com (under topic mapping)

3 See, Further Reading for a listing of some current web based resources on topic maps.

4 As in the more common “fog of war,” truth is the first casualty.

The Tower of Babel

The exact location and nature of the Tower of Babel has been lost in the depths of history. The story of its builders was carried forward in several traditions, the account in Genesis 11:1-9 being among the better known.⁵ One popular translation reads in part as follows:

Then they said, "Come, let us build ourselves a city, and a tower with its top in the heavens, and let us make a name for ourselves; otherwise we shall be scattered abroad upon the face of the whole earth."

The Lord came down to see the city and the tower, which mortals had built.

And the Lord said, "Look, they are only one people, and they have only one language; and this is only the beginning of what they will do; nothing that they propose to do will now be impossible for them.

Come, let us go down and confuse their language there, so they will not understand one another's speech.

So the Lord scattered them abroad from there over the face of all the earth, and they left off building the city.⁶

The most salient part of the story being: "... so they will not understand one another's speech." Whether or not one accepts this account as the origin of that diversity, or the post-Flood settlement account of Noah's descendants,⁷ it is clear that linguistic diversity is seen as a barrier to understanding.

Linguistic diversity, however, is not simply a matter of having different native languages. Specialized terminology, corporate norms, regional dialects, differing views of data or data models, within a single language, company, or organization can lead to an understanding gap as large as between a native Bantu speaker and a Irish Boston cabby.

5 S.N. Kramer suggests the Sumerian epic, "Enmerkar and the Lord of Aratta" as a parallel example. The "Babel of Tongues": A Sumerian Version. *Journal of the American Oriental Society* 88: 109, 111 1968

6 New Revised Standard Version

7 Genesis 10

Overcoming Diversity?

Greater understanding, despite linguistic diversity, has long been sought to achieve common goals (or avoid disasters). One of the more commonly suggested routes to greater understanding is having a single common language or even a uniform ontology or world view. Contenders for that role include, Esperanto⁸ and Loglan,⁹ while there is no shortage of “universal” ontology projects such as the Cyc Project,¹⁰ and of course, the Semantic Web activity at the W3C.¹¹

The irony of such universal language or ontology efforts should not be lost on any student of history. Consider that there is now an acknowledged diversity of languages and ontologies. What is the likely impact of adding one more, except to incur the attendant cost of learning the latest “in” language/ontology, and adding to the burden of understanding of those who have not “taken the cure” as it were?

And leaving aside the diversity of languages/ontologies within government agencies, corporations and other organizations, the history of computer science leaves little hope for a uniform language of any sort originating from its quarter. Consider that there are some 2,500 computer languages,¹² all with their advocates and adherents. And that does not include markup languages, whose supporters have been quite prolific (or promiscuous) in formulating new and often incompatible languages.¹³

But government agencies, corporations and other organizations touch every part of daily life. Can their diversity in languages and ontologies simply be ignored? Not to mention the diversity found among the oft neglected users, voters, consumers and people in general. Is there any hope of a solution that embraces diversity?

8 Esperanto League for North America, <http://www.esperanto-usa.org/>

9 Loglan: <http://www.loglan.org/>

10 Cyc Project, <http://www.cyc.com>

11 Semantic Web, <http://www.w3.org/2001/sw/>

12 The Language List, <http://people.ku.edu/~nkinners/LangList/Extras/langlist.htm>

13 Robin Cover is the premier chronicler of markup languages. <Http://www.oasis-open.org/cover>

Embracing Diversity

The “why” of topic maps is diversity— it is able to embrace a diversity of languages and ontologies and, at the same time, to provide a basis for understanding what is being said. And that diversity includes even the languages in which topic maps are written. It would make little sense to decry efforts to create a particular universal language only to replace it with another.

Without reference to any particular topic map language or data model, the following is a brief sketch of how topic maps enable the embracing of diversity in languages and ontologies.

What is the most common phrase you hear when either explaining or having something new explained to you? “See what I mean?” The speaker has pointed out a number of bits of information in hopes that when you see whatever is being explained again, you will recognize it.

Of course, even the smallest speaker can't reside inside a computer, so the topic maps paradigm says that for whatever you want to talk about, you need to record the information necessary for someone else to recognize the subject you are describing. Doesn't matter whether the subject is animal, vegetable, mineral, an abstract concept, historical or legendary, a relationship or an example of any of the foregoing. In short, whatever you want to talk about and however you want to describe it is legitimate in the topic maps paradigm.¹⁴

But that by itself does not embrace diversity. People are free now to use whatever methods they like, so where's the embrace?

Good question! The topic maps paradigm embraces diversity and enables understanding by enabling the descriptions of subjects to be “merged” with other descriptions of the same subjects. It is this subject based merging that gives topic maps the subject-centric orientation that is the essence of the paradigm.

¹⁴ All particular implementations of topic maps will have some restrictions on what can or cannot be said and how. Users must evaluate those limits for themselves in choosing topic map solutions.

Subject Based Merging

A common form of subject based merging occurs when two different descriptions of a subject are the same. Literally the same. Two different users say “tomato” and neither one says “tomate” so the representations for that subject “merge.”¹⁵ This is more than most information systems can do now, but is not dissimilar to the quest for a universal language or ontology mentioned earlier.

What if there are two users, one who says “tomato” and the other one who says “tomate?” The information recorded from both of those users should be viewed together, but how can that ever happen? The answer to that question depends upon how much information the respective users have supplied or that is supplied by someone else.

Assume the two tomato/tomate lovers enter the following data respectively:¹⁶

Tomato Lover:

Name: tomato

Genus: *Lycopersicon esculentum*

Forecast: <http://www.ers.usda.gov/briefing/tomatoes/index.htm>

Origin: South America

Tomate Lover:

Name: tomate

Related Plant: nightshade

NameFrom: Nahuatl word “tomatl”

Events: <http://www.spain-info.com/Culture/tomatofight.htm>

¹⁵ The term “merge” actually means “are seen as one entry.” The topic maps paradigm does not compel any particular processing of the representatives of subjects.

¹⁶ I have shamelessly stolen the idea for this example from Sam Hunting's chapter “How to Start Topic Mapping Right Away with the XTM Specification” in XML Topic Maps, Park and Hunting, eds. Addison-Wesley, 2003. ISBN 0-201-74960-2

What our two tomato/tomate lovers would like to see would be:

Name(EN): tomato

Name(ES): tomate

Events: <http://www.spain-info.com/Culture/tomatofight.htm>

Forecast: <http://www.ers.usda.gov/briefing/tomatoes/index.htm>

Genus: *Lycopersicon esculentum*

NameFrom: Nahuatl word “tomatl”

Origin: South America

Related Plant: nightshade

That is to say, all the information about the same subject, in this case a tomato/tomate, all presented together to both users, even though they entered their information in their own ways.

There are a number of techniques presently and more being developed for such merging, but consider the simplest scenario. Assume a third user, one who knows that “tomato” and “tomate” are both terms for the same subject. The third user enters that information into the topic map system and as a result, anyone searching under either term will obtain all the information about that subject.

If that sounds too easy, consider that the relational database world still lacks a developed notion of subjects, which is what the merging operation turns upon. That is to say that it is necessary to think in terms of subjects and what is known about a subject in order to support any useful form of merging. Particularly if a uniform language or ontology is not going to be required for it to work. In retrospect that does seem obvious but a debt of gratitude is owed to the individuals who discovered and then persisted in explaining their insight.¹⁷

¹⁷ Steven R. Newcomb and Michel Biezunski were the originators of the topic maps paradigm and continue to be the leading lights in its development.

Summary

From a historical perspective, it should be concluded that one world language solutions are doomed to simply add to the diversity of languages that gave rise to the term “babble.” The Jewish Study Bible concluded that the building of the Tower of Babel was an “act of Promethean hubris.”¹⁸ Certainly the same could be said for any language, ontology or data model that purports to be the universal answer to linguistic, cultural, and organizational diversity.

Users should ask the following questions of any information solution they encounter:

- 1) Does it have a subject identity based means of merging information about the same subject?
- 2) Does it enable users to specify how they wish to identify their subjects?
- 3) Does it enable users to specify rules for when two instances of the same subject should be merged?
- 4) Does it force users to learn and use a new or awkward language to describe their subjects?
- 5) Is it better-adapted to the users' world views than to that of the vendor who is selling it?

Unlike the goals of more ambitious projects, the goal of the topic maps paradigm is almost prosaic.¹⁹ It is to allow anyone to talk about anything and have the opportunity to discover information from others, talking about the same subject, without regard to language, culture or other diversity.

The topic maps paradigm does not insure that goal will be achieved. It just makes it more likely than approaches which require universal homogenization of language or culture.

18 The Jewish Study Bible, Berlin and Bretter, eds., Oxford University Press, 2004, page 29.

19 Prosaic, however, does not mean it lacks in conceptual complexity. Interested readers can find the latest draft of the Topic Maps Reference Model at: <http://www.isotopicmaps.org/tmrm>. Look for the latest revision as it is still in the ISO process.

Topic Map Resources

The following is a very brief listing of resources and consultants for those interested in exploring topic maps in further detail.

Coolheads Consulting: Steven R. Newcomb, Victoria T. Newcomb and Michel Biezunski. <http://www.coolheads.com>

Cover Pages: (XML) Topic Maps <http://xml.coverpages.org/topicMaps.html>

Patrick Durusau: <http://www.durusau.net>

Networked Planet: Kal Ahmed and Graham Moore.
<http://www.networkedplanet.com>

Mondeca: Jean Delahousse, Bernard Vatan. <http://www.modeca.com>

Ontopia: Steve Pepper, Lars Marius Garshol. <http://www.ontopia.com>

SC34/WG3: <http://www.isotopicmaps.org> (Website for the ISO working group responsible for the topic maps standard.)

XML Topic Maps: Creating and Using Topic Maps for the Web, Jack Park and Sam Hunting, eds., Pearson Education Inc., 2003. (Covers XML topic map syntax only.)

Biography

Patrick Durusau is an independent consultant, advising clients on the design, implementation and integration of complex information systems.

He was the Director of Research and Development for the Society of Biblical Literature. (2000-2005)

He is the co-editor (with Steven R. Newcomb) of the Topic Maps Reference Model and Chair of the Published Subjects TC at OASIS.

Patrick is the Chair of the US National Body to JTC1/SC34, a member of the board of directors of the Text Encoding Initiative Consortium, technical lead for the OSIS Project, a joint markup project of the Society of Biblical Literature, the American Bible Society and the United Bible Societies.