The Economics of the Topic Maps Reference Model

[XML 2002 Conference and Exposition](), December 8-13, 2002, Baltimore

Keywords: Data interchange, Design, Distributed Systems, E-business, E-Government, Enterprise applications, Enterprise Content Management, Internet, Interoperability, ISO 13250, ISO Reference Model, Knowledge Management, Metadata, Ontology, Publishing, RDF, Semantic Web, Structure, Topic Maps, WWW, XML, SGML

Steven R. Newcomb, Ph.D.
Independent Consultant
Coolheads Consulting
Allen
Texas
USA
srn@coolheads.com
http://www.coolheads.com/srn.cv.htm

Biography

> Steven R. Newcomb is an independent consultant, a co -chair of the annual Extreme Markup Languages Conferences, and a co -editor of the Topic Maps (ISO/IEC 13250:2000) and HyTime (ISO/IEC 10744:1997) standards.

Abstract

The draft ISO standard Reference Model of the Topic Maps paradigm meets technical requirements that are strikingly parallel to the same economic principles that are evidently most conducive to worldwide economic growth.

Table of Contents

Introduction

In recent decades, the thinking of the economist Friedrich A. Hayek has powerfully shaped -- and is still shaping -- the economies of many nations.

Two of Hayek's key economic ideas are:

1. An economy is most likely to flourish when its government limits itself to maintaining an impartial system of laws that allow innovative economic activities to proceed openly and freely. An economy is less likely to flourish when government attempts to dictate economic outcomes, to favor specific business models, or to favor specific persons and/or organizations. One of the hallmarks of an economic policy that promotes good economic health is that it is impossible to predict who will be the disproportionate winners, and who will be the disproportionate losers, at least over the long term.
2. It is in everyone's economic interest that all economic decision makers (i.e., everyone who trades) take the fullest possible advantage of all available and relevant knowledge. Everyone suffers when ignorance causes decisions to be made that reduce human productivity, or that miss opportunities to increase it. When knowledge is applied in the ways that best conserve or increase human productivity, everyone benefits.

The Topic Maps Reference Model proposes a simple way of thinking about knowledge that, when systematically applied, is expected

- to increase the quantity, diversity, and utility of available knowledge, and
- to allow knowledge asset owners to exploit the markets for their knowledge assets more efficiently and comprehensively.

The Topic Maps Reference Model's ability to conserve and promote diversity, even while facilitating knowledge aggregation, makes it an attactive tool for those who seek ways of applying Hayek's principles in creating a "knowledge economy".

How the Topic Maps Reference Model Works

The Topic Maps Reference Model is basically a distinction -- and a well-defined boundary -- between two things:

1. The minimum set of structural features that knowledge must always be regarded as having, if we are to have a convenient way of aggregating any kinds of independently-maintained knowledge with any other kinds of independently-maintained knowledge, in a lossless, predictable, useful, and affordable fashion.
2. All the other features that knowledge can have, including those that may need to be constrained in diverse ways within diverse contexts.

The Topic Maps Reference Model takes the position that the minimum set of structural features that must be common to all knowledge, in order to allow all kinds of knowledge to be aggregated, are a set of constraints on the structure of semantic networks. The kind of semantic network that is defined by the Reference Model is called a "topic map graph".

In a topic map graph, every node is a surrogate for exactly one subject (as in "subject of conversation"), and no two nodes are surrogates for the same subject. All nodes are connected to each other by nondirectional arcs. The Reference Model provides exactly four kinds of arcs, each of which is used in the same very specific way in each "assertion". In the Reference Model, assertions are the primary units of knowledge. Every assertion is a set of specific nodes interconnected in specific ways by specific kinds of arcs. Assertions represent relationships between subjects, and in each assertion, each related subject plays a specific role (called a "role type"), which is itself a subject. Each assertion can itself be an instance of an "assertion type", which is also a subject.

Arc Types

AT CR AC Cx
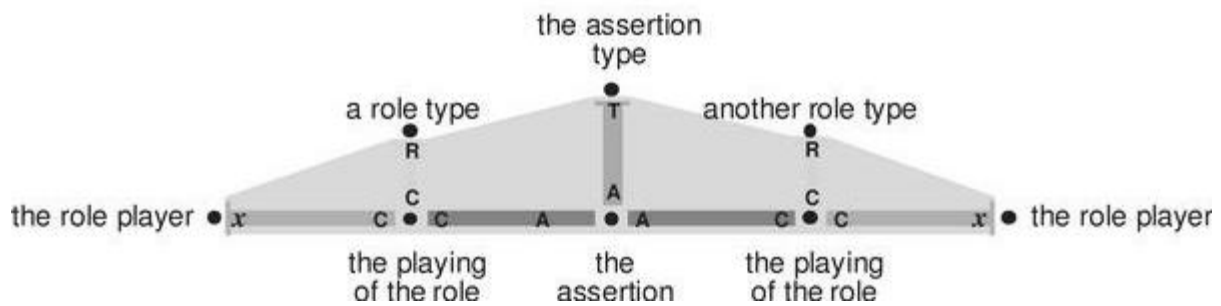
The structure of an assertion that has two role types.
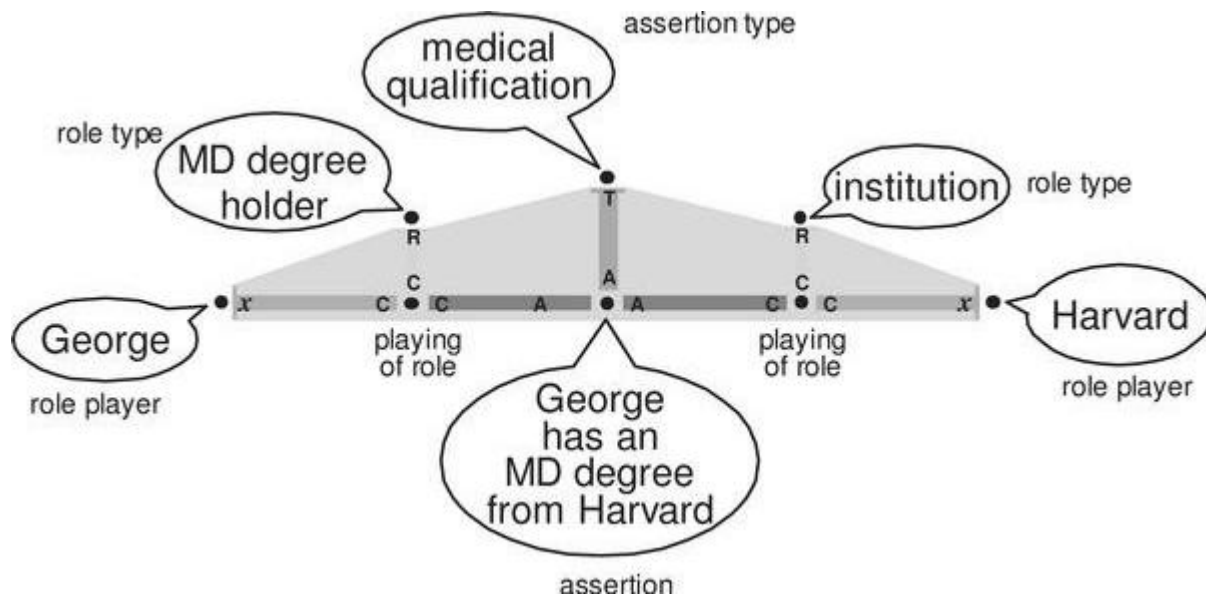
An instance of an assertion. The subject of this assertion is the notion that George has an MD degree from Harvard.

The abstract structure of assertions is rich and complex enough:

- to allow all kinds of assertions to describe themselves comprehensively and compellingly,
- to allow every substantive aspect of every assertion to be a role -player in any number of other assertions, and
- to allow any knowledge to be aggregated with any other knowledge, without changing or compromising the integrity of any of the knowledge that comprises the aggregated whole, and without introducing redundancy of any kind.

However, the abstract structure of assertions is fixed; its richness never varies. The rigidity of the structure is such that, for example, it is not possible for an assertion to have a role player in the absence of an explicit role type (role) for it to play.

Assertions themselves, however, are unlimitedly diverse in their semantics, and in the functions they can serve within the context of diverse Topic Maps Applications. Indeed, just as the Topic Maps Reference Model imposes no constraints on the subjects that nodes can represent, it imposes no constraints on subjects that are the relationships represented by assertions, nor on relationship types (assertion types) of which they are instances, nor on the roles (role types) that subjects can play in such relationships.

In order to protect the diversity of Applications, the Reference Model imposes no constraints whatsoever with respect to assertion types. This policy has consequences that some might consider extreme. Consider, for a moment, the fact that the abstract structure of assertions forbids any single assertion from having two role players that play the same role. This constraint is necessary in order to make it possible for the Reference Model to guarantee predictable, deterministic knowledge aggregation without integrity losses. However, in many circumstances, it is necessary for a set of subjects to play a single role in a single assertion. If each member of the set is also represented as a node, there should be some representation of their membership in the set, and then the set (which is a distinct subject in its own right) can play the role. The membership of a set member in a set is a relationship that can easily be represented by an assertion, and such an assertion should be an instance of an assertion type used to represent the membership of a member in a set. But the Reference Model does not require support for such an assertion type; it doesn't require support for *any* assertion types, leaving all such decisions for the designers of Topic Maps Applications.

Despite the fact that different topic maps may conform to diverse Topic Maps Applications, they can nonetheless be aggregated ("merged") under the Reference Model; the Reference Model provides a platform on which all knowledge can be aggregated, regardless of its diversity. The points of aggregation are subjects, and aggregation is effectively required by the "one subject per node, one node per subject" rule. Each node (each node's subject) can be a role player in any number of assertions (relationships), and the assertions can be made within any number of ontologies. (In the context of the Reference Model, an *ontology* is mainly a set of assertion types and role types, and it is the essence of the definition of a Topic Maps Application.) Assertions describe themselves by their supporting role type and assertion type subjects, so their ontological diversity need not be compromised by aggregation with assertions created within foreign TM Applications. The diversity of assertion types that can result from aggregating diverse kinds of knowledge can only increase, and can never decrease, their individual usefulness.

So what does the Reference Model mean for the economics of the creation, maintenance, aggregation, marketing, and utility of knowledge?

(In what follows, I'm only partially answering the above question, because I'm omitting the usual discussion of how topic maps can improve the findability of materials relevant to subjects, overcome the problem of multiple and/or colliding names for subjects, support diverse usage scenarios, etc. etc. I'm assuming that readers of this paper already know that material, or at least can easily find it elsewhere. Much of that discussion would be outside the scope of this paper anyway, because most such features, and even the famous "TAO of Topic Maps", is necessarily in the realm of specific Topic Map Applications, such as the Standard Application.)

Many more (and more diverse) knowledge creators and maintainers, and many more and more diverse knowledge assets. Small, independent businesses can own and maintain extremely specialized knowledge assets, and, despite their small size, they can exploit them in a wide variety of markets by licensing them to aggregators that serve each of those diverse markets. From the perspective of small knowledge asset creator/maintainers, it's tantamount to "outsourcing the marketing and sales."

Many more (and more diverse) knowledge aggregators. Knowledge aggregation can become more of a cottage industry, in addition to remaining a way for the big publishers and portal operators to add value to their offerings, and reap the rewards of being big. Multiple aggregators may add different kinds of value and address different markets. Publishers and portal owners can still add value by aggregating, but now they are in a better position to outsource knowledge aggregation tasks, and to use knowledge assets owned by others to increase the exploitability of their own. Because of the minimal disciplines imposed by the Reference Model, aggregation reliably adds whatever real value can be realized, and the whole can often be greater than the sum of the parts. Each subject remains unique, even while everything known about it is directly connected to it. True, the cost of aggregating a group of knowledge assets the first time may still be substantial, but the cost of aggregating their subsequent versions on an ongoing basis is minimized, because investments in mappings that have not been affected by changes retain their value. The knowledge of how to aggregate a set of evolving knowledge assets becomes a core

competency of a specialized business.

Many more, larger, more diverse, and more specialized aggregations of knowledge. Aggregators can license their products to other aggregators. One can easily imagine many layers of aggregation, in which smaller aggregations become aggregated into very large and/or very specialized aggregations. Very specialized aggregations may be created by aggregating a great deal of knowledge, and then deleting whatever is outside the specialty. Large aggregators may choose to use such specialized aggregations to acquaint customers with the value of their offerings.

More specialization and diversity among publishers and portal owners. Increasingly, these will be in the business of delivering consumers to aggregators, and their market approaches will diversify into various GUIs and services that will enhance productivity of the users of knowledge.

Conclusion

In general, the Topic Maps Reference Model is designed in such a way as to increase the likelihood that knowledge will be found and used when needed, that the creators and maintainers of knowledge will be rewarded for their efforts and contributions, and that the opportunity to be in the knowledge aggregation business will be very widely distributed, instead of being concentrated in relatively few organizations.

The general impact is to increase the knowledge economy's diversity.

The knowledge business becomes even more outsourcing-oriented. Indeed, since all knowledge is intimately connected with all other knowledge, and since nobody can possibly know everything, it follows that, for the sake of our economic health, we should always strive to provide for a knowledge economy that maximizes the opportunities for outsourcing relationships, that maximizes the efficiency with which contributors to the value chain are rewarded, and that minimizes the cost of aggregating arbitrary combinations of constantly changing bodies of knowledge.

If Hayek is right, then the efficiency with which economies perform can be significantly enhanced by applying the Topic Maps Reference Model in the way described above.

A few monolithic publishing organizations cannot reasonably be expected to plan the distribution of knowledge in such a way as to maximally benefit the overall economy. Such an expectation is as unreasonable -- and as disreputable -- as the expectation that the economy of a large country can be efficiently planned by its national government. Such plans will always be suboptimal, and will inhibit the development of a multitude of complex, diverse, economically efficient knowledge-transference relationships between those who have knowledge and those who can put it to use. If it's really true that we're moving toward a "knowledge economy", then we must free that economy to define itself. In a wide-open knowledge economy, everybody who has knowledge wins, everybody who uses knowledge wins, and everybody else wins, too. The Topic Maps Reference Model can be a powerful tool in the quest to open and diversify the knowledge economy.

Suggestions for Public Policy

If a government wanted to seize the opportunity afforded by the Topic Maps Reference Model to improve economic productivity, what should it do?

I suppose Hayek would routinely suggest that laws and regulations that encourage economic participation by, and do not have the effect of discriminating against, small, highly specialized knowledge creators, maintainers, and aggregators, should be established and maintained. Hayek might suggest examining all existing regulations to see whether they might tend to diminish the diversity of the knowledge industry, or to exclude small players from access to the maximum quantity and diversity of potential customers. He might suggest that government bear in mind the fact that, in the end, it is individuals who know things, and it is individuals who make decisions, and therefore the efficiency of the knowledge economy depends on the efficiency with which knowledge-carrying channels can, in effect, be established between individuals. He might point out that one size does not fit all, and the character of the optimal knowledge-carrying channel between any two individuals may even be unique to the combination of the knowledge itself and the two individuals.

Acting on Hayek's advice, a national leader like Margaret Thatcher might ask herself whether her government is occupying economic roles in the knowledge economy that would be more efficiently occupied by private interests, and, if any are found, to "privatize" them forthwith.

While there are many areas in which intellectual property reform is sorely needed in the United States and elsewhere, intellectual property reform cannot be accomplished on the basis of theory alone. Such reform will be best informed by the experiences of a knowledge economy whose diversity is actually and already burgeoning.

So, the question becomes, "What can government do to encourage the burgeoning of diversity in the knowledge economy?" I think the answer is simple: the government, which, if well-run, is among the biggest knowledge consumers in any economy, should become a smart customer. It should set its own knowledge acquisition policies, for its own self-interested reasons, in such a way as to minimize the cost of aggregating the knowledge it acquires. The Topic Maps Reference Model shows that it is not necessary for the government to dictate ontologies to its information suppliers, in order to allow knowledge aggregation to be accomplished at low cost and with high value.

It's interesting to reflect, for a moment, on the history of XML. The TopicMaps Reference Model represents the next stage of information interchange conventions. With it, we're moving beyond our old concerns about standardizing the parsing function so that we can achieve *information* interchange at predictable cost. Now we're responding more directly to our deeper concern: controlling the cost of *knowledge* interchange. It's instructive to remember how it happened that SGML became such a powerful force in the publishing world that, when the World Wide Web came along, SGML was basically the only rational choice as a basis for HTML, and, ultimately, for SGML's twin brother, XML. Briefly (and too simply), the primary reason for the success of SGML was its early adoption by the U.S. Department of Defense, which in 1987 began requiring its weapons systems documentation to be delivered in SGML form. When that famous decision was made, it was made in SGML's favor because there was no other candidate that could meet the requirements, and the economics of DOD's need for standardization in this area were acute and desperate.

Governments face a similarly acute and desperate situation today, but this time it's in the arena of intelligence gathering and integration. The ability of the Topic Maps Reference Model to support the aggregation of diverse knowledge from diverse, independent sources is exactly what's needed in order to allow government agencies and privately-owned knowledge to be aggregated efficiently and, more to the point, quickly enough to avert terrorist disasters. I discussed this problem, and proposed the Topic Maps Reference Model as a key part of the solution, in my talk at Extreme Markup Languages 2002, "Forecasting Terrorism: Meeting the Scaling Requirements" (http://www.coolheads.com/SRNPUBS/extreme2002/forecasting-terrorism.htm).

There is good reason for governments to adopt the Topic Maps Reference Model as their preferred model of knowledge aggregation, and there is no alternative that offers the Reference Model's combination of integrity preservation, predictability, scalability, and so on. It would not be the worst side-effect of the worldwide terrorism crisis if it led, ultimately, to a far more robust and diverse worldwide knowledge economy, just as the success of SGML/HTML/XML is not the worst hangover effect of the defense buildup during the Cold War.