MetaTech Consulting, Inc.

White Paper

Analysis and Synthesis of Rule Markup Language (RuleML), an Emerging Technology

Jim Thomas

April 25, 2003

Table of Contents

Analysis and Synthesis of the Rule Markup Language (RuleML)

The quantity of the content on the World Wide Web (herein referred to as the Web) is vast and continues to increase at an ever increasing rate. Without foreknowledge of where some specific document or data exists, the difficulty of finding it is increasing proportionally to the volume. Increasingly sophisticated search engines aid in efforts to locate data though they are only marginally effective in resolving the relevance of the potential web documents. Additionally, search engines at large only span a portion of the entire web. Industry has continued to grapple with technologies and techniques that will better serve the

Overview of RuleML

Organizations realize efficiencies through the employment of a variety of rules. Rules may specify who is to do what, or how something is to be done, or when (e.g. under what conditions) an action is to occur. Policy documents and operations manuals document many of an organization's rules such that they are readily identifiable. Less tangible are an organization's tacit rules – those that are not captured in formal or official documents. It is the experience of this author that it is these tacit rules that allow organizations to succeed in a dynamic, fluid operating environment.

Collectively, explicit and tacit rules constitute an organizational or industry *Rule Knowledge*. In a Business-to-Business (B2B) or Business-to-Customer (B2C) scenario a common interpretation of rules, e.g. shared Rule Knowledge, is crucial though exceedingly difficult to establish. RuleML strives to provide standardized structure and syntax to represent Rule Knowledge (Grosof, 2001). Solutions implementing RuleML will benefit from enhanced interoperability of data and processes and will realize improved efficiency as well as reduced operating costs.

Technology Basis

RuleML is not a standalone technology. Rather, it is a technology that, together with other technologies, will contribute to a collective solution – more meaningful and useful data, information, and knowledge shared between interested parties (e.g. people, process, and systems). The following subparagraphs provide a brief discussion of some of the technologies related to RuleML with a description of the nature of the interrelationship.

Symantec Web.

The World Wide Web Consortium (W3C) is an oversight and guidance body that helps to steer the evolution of the Web. They have recognized an extension of the Web, referred to as the Symantec Web, in which information is given well-defined meaning. As described by Berners-Lee, Hendler, and Lassila (2001), the Symantec Web will bring structure to the meaningful content on the Web, and thus create an environment where sophisticated tasks can be performed by intelligent agents roaming the web on behalf of users. It is not reasonable to expect that every web page document on the Web will be given full meaning. Rather, only those documents related to discrete industries or organizations are of concern here. In many instances it is necessary to include the rules associated with data for it to be considered fully well-defined. RuleML has been identified as a markup language for documenting and communicating those rules within the Symantec Web.

Ontologies.

Also central to the Symantec Web are *Ontological* specification. An ontology provides context, generally in some natural language, for data, information, and knowledge (Gruber, 1993). Just as individual pieces of information require context to have well-defined meaning, so too do rules. For example, a piece of information such as available stocks of *culture media* – the

substance used in laboratories to grow bacteria – has different relevance if your perspective is

from the Center for Disease Control than if you are a member of an international body charged

with monitoring of the proliferation of duel-use components associated with weapons of mass

destruction development. Individual industries are working to model data, information, and

knowledge within their discrete context. Data that exists in more than one context – more than

one ontology - may have different rules for each of the ontologies. Therefore, RuleML is

dependent on successful development of ontologies within the Symantec Web.

eXtensible Markup Language.

The eXtensible Markup Language, known commonly as *XML*, is Web standard for text

publishing. It provides the flexibility for the author to annotate (e.g. markup) the document with

user defined tags. As web-based documents are fundamentally text documents, this technology

has become core to the Web of today and is an enabling technology of the Symantec Web.

Being easily extensible, it has provided the foundation for a remarkable number of other markup

languages to include RuleML. XML provides the markup language while RuleML provides tags

and unique syntax for rules implementation (Boley, 2001).

Rule classifications.

The science of rules has a pedigree shared with mathematics and philosophy. While a

discussion of its lineage is beyond the scope of this paper, it is worthwhile to mention that rules,

theorems, and proofs share a tight association. Businesses of today employ rules from each of

these disciplines to accomplish their objectives. Analysts capturing explicit and tacit business

rules categorize as follows (Wagner, 2002):

1. Integrity rules that specify that something must be true for something else to occur. This is typified by the insistence of a back to only approve a withdrawal from an account if there are sufficient funds available.

2. Deductive rules that involve inference or mathematical calculations. Again, using the banking analogy, a deductive rule is applied to calculate the interest due on a loan.

3. Reactive rules that identify an action that is to be performed in response to an event. This is generally referred to as an *Event-Condition-Action* rule. A readily available banking example of this type of rule is captured in the scenario when a check is submitted for payment against an account with insufficient funds. In this scenario, the event is the submission of the check for payment. The condition is the insufficient funds status. The action may be the awarding of a penalty fee against the account and an issuance of a letter to the account holder.

4. Deontic assignment rules that specify powers, rights, and duties of system actors. One need not overextend the banking analogy to draw an example for this rule type. A banking institution may limit the right to approve unsecured loans to senior loan managers while allowing secured loans to be approved by all loan managers.

Though RuleML is expected to be able to represent all of the aforementioned rule types, the current implementation engines support only reactive rules.

Challenges

No technology evolved in the absence of challenges. With new technologies, the challenges that readily come to mind deal with the immaturity of the dependent technologies and the inability of the stakeholders to generate sufficient interest to obtain funds necessary to further their cause. There are, however, other types of challenges that place RuleML at risk. They are

a) knowledge capture, and b) decentralized stakeholders. A discussion of these challenges

follows.

Knowledge capture.

As mentioned earlier, business within industries and organizations alike is accomplished

by following explicit and tacit rules. Where rules are well documented, they may be readily

translated to RuleML. The tacit rules, those that may exist wholly within the minds of the

personnel (either consciously or unconsciously) are every bit as important though dramatically

more difficult to chronicle. Even when those possessing the rules are willing to impart such

knowledge, it requires painstaking effort by the business analyst to discover the rules. It is far

more difficulty to capture those rules when the knowledge source is unwilling to share. Many of

the *grey beards* within organizations view their knowledge as power and are unwilling to share

it, possibly feeling threatened that it may be used to make their positions obsolete.

Decentralized stakeholders.

The wider the scope of standards, the greater the value they bring. The W3C and other

bodies strive to identify standards with international application, thereby providing the greatest

utility and value. The greater the scope of a standard, the greater the difficulty in gaining

necessary consensus in ratifying it. This same trend holds equally true for definition of

ontologies and rules. Industries are driven to collaborate on defining functional ontologies and

rule sets to increase their operating efficiency and ability to effectively communicate with their

peers and partners. As industries span national borders, each representing unique interests, the

decentralized nature of the effort is accentuated through protracted negotiations on the subject.

Issues

The issues or implications of RuleML touched on in the preceding discussions were couched in only the narrow context of organizations or industries. The implications, however, have the potential to extend to the entire web-enable community – all who use the internet. A discussion of the social, cultural, and political implications of RuleML follows.

Social.

From the inception of the internet, particularly from the realization of the internet, the quantity of data and information available through the technologies has continued to balloon. While some may be naïve enough to think their queries are spanning the breadth of the entire internet, even the most robust search engines access more than a significant minority of the total available content when prosecuting a query. Even if the tools were fully capable of exploiting the full content of the Web, the users would be unable to discern the data contained in their query results that are the best for their needs. The Symantec Web, with RuleML as an enabler, has the potential to rescue the users from a sea of data and allow them to function on islands of knowledge – the reduced subsets of data and information formed into actionable, relevant knowledge. Technologies hope to elevate the user community further up the information supply chain to the point that the Web provides *wisdom* beyond knowledge. The Symantec Web is a necessary step along that journey.

Cultural.

Discussions of culture frequently entail a treatment of trust. The Web influences each individual user differently, in part, to their perception of trust in its content. As RuleML and the Symantec Web mature, the quality of the data, information, and knowledge it makes available should improve in quality.

Political.

As with all other Web technologies under the governance of the W3C, the evolution of

RuleML is being steered by an international consortium, not a single business or even industry.

Participation in activities related to the technology is open to any individual or organization with

a vested interest. Through defined processed sponsored by the W3C, such as the *Request for*

Comments process that allows any individual to submit comments and comment on submissions

from others, unbiased evolution of the standards and technology is supported.

Life Cycle Implications

As an open-systems technology, RuleML belongs to the internet community rather than

any one vendor or technology integrator. Though open-systems technology and standards

generally provide greater value to the industry participants, they require a high level rigor

throughout their life cycle to ensure continued community support. This rigor manifests itself in

costs that must be shared by the interested community partners. RuleML is today an immature

technology. However, with its inclusion as an enabling technology for the Symantec Web,

substantial public and private interests have begun bringing to bear their skills, resources, and

clout to accelerate its evolution. Industry has committed continuing investment of significant

capital to ensure this technology continues to mature in synchronization with the other related

technologies.

Conclusions

RuleML is an emerging technology with the potential of significantly influencing the way the internet community accesses and applies the vast quantities data, information, and knowledge contained in *cyber* space. Through continued investment by industry, the XML-based technology can represent the rules of business to aid in realizing the Symantec Web. The variety of tools employing the technology and standards of RuleML will continue to grow particularly in the industries seeking to maximize B2B and B2C capabilities.

In this paper, the author has explored a broad range of factors related to RuleML. The emerging technology was discussed in context in which it is evolving. Peer reviewed literature, together with technical papers authored by those instrumental in defining the technology, were surveyed to support the analysis and a select few were cited to substantiate assertions.

References

Berners-Lee, T., Hendler, J. & Lassila, O. (2001), The semantic web. *Scientific American,* 284 (5), p34.

Boley, H. (2001), The rule markup language: RDF-XML data model, XML schema hierarchy, and XSL transformations. Retrieved April 23, 2003, from http://www.dfki.uni-kl.de/~boley/ruleml-mht.pdf

Grosof, B. N. (2001), Standardizing XML rules: Preliminary outline of invited talk. Retrieved April 23, 2003, from http://ebusiness.mit.edu/bgrosof/paps/ebiz+intell-web-ijcai01-wksh-paper.pdf

Gruber, T. (1993), What is an ontology? Retrieved April 23, 2003, from http://www-ksl.stamford.edu/kst/what-is-an-ontology.html

Wagner, G. (2002), How to design a general rule markup language. Retrieved April 23, 2003, from http://tmitwww.tm.tue.nl/staff/gwagner/myruleml/GRML.pdf