

David Goldstein, PhD
Wise Web softWare
david_goldstein@freddiemac.com

Introduction

This paper discusses some of the factors that should be considered in fostering adoption of rule-based technology by business to expand the "semantic web". The authors have helped develop numerous applications for large military and commercial customers, ranging from one million to six-hundred million dollars in budget. In this paper we examine the state-of-the-art, noting factors that have been the most influential in widespread adoption of knowledge-based technology. We then present specific issues from the perspective of business and improvements in current technologies needed for widespread adoption by business. We conclude our technical portion with how these suggestions should be tailored for public representation and interchange of rules. Having done significant academic and military research, we try to present the needs of business in a framework that can readily be digested by the academic community.

State of the Art

Literally hundreds of government initiatives and commercial products have contributed to the state-of-the-art. DARPA, W3C, Esprit, and other organizations have attempted to create tools and standards for fostering information and knowledge sharing. With the explosion of interest in networks in the mid 1990's, the concept of a "semantic web" gained increasing popularity.

Use of rule based systems, ontologies and knowledge sharing in corporations has trailed these efforts dramatically. Without funds to explore technologies for invention's sake, little has changed in the technologies that have been adopted - other than buzzwords - for almost two decades. Decision tables, decision trees, and production rules may seem novel technologies, but they are all orthogonal representations of mechanisms that gained widespread popularity based upon Forgey's 1978 paper on RETE and N.A.S.A.'s/U.S.A.F.'s joint development of the C Language Integrated Production System (CLIPS).

CLIPS has had the largest effects on intelligent software development business of any effort: it provided "nearly free" tools that enabled business to examine the area without significant investment. JESS, ART-IM, Eclipse and numerous other reasoning tools can be directly traced to CLIPS. Virtually every business that has adopted commercial tools - even adopting tools those that claim no lineage from CLIPS - have prototyped systems using CLIPS or one of its derivatives.

This heritage is important to realize because business is loath to build infrastructure. The authors have seen companies abandon business rules completely because of a lack of infrastructure. We have seen companies give up rule-based systems for simple adoption of lower-level tools and we have also seen companies cancel large projects, with hundreds of people, after sinking tens of millions of dollars into rule repositories. Drastically increasing the adoption of rule-based systems, embracing knowledge-sharing, and offering shared ontologies will only receive a warm welcome from business once nearly seamless, business-oriented tools and interfaces are available.

Business Scenario

A good example of current commercial needs can be obtained from examining some commonalities of Wise Web software's current and prospective clients. Examining the financial arena, a common process that must be supported can roughly be described as:

- The business organization defines requirements,
- Business engineers specify these requirements in detail using non-knowledge-based tools,
- "Magic occurs" to implement these requirements into a rule language,
- "Magic occurs" to have rule engines to interoperate with the organizations infrastructure,
- Business engineers work with testers to validate the knowledge in a system,

- "Magic occurs" to get the knowledge in the system to accommodate corporate security needs, and
- The new version of the software system is deployed in a production environment.

Each instance of "magic" in the above description involves specialized software development that business considers high-risk. Significant vendor training, vendor tools, and current repository technology has yet to substantially reduce the risk on any project observed to date. In fact, the author has never seen knowledge reused in a commercial application – other than through cut and paste - despite the use of commercial repository tools, internal ontology and adoption of standards.

The obvious issues are making repository tools, ontologies, standards, etc., be improved to foster knowledge reuse. Ontologies must capture nuances in domains and bootstrap development to make using rules technologies by domain experts worthwhile – most domain experts do not understand proper use of parentheses, let alone predicate calculus.

Knowledge sharing tools must also import commercial standards to bootstrap ontologies; the common approach of importing and exporting XML is not enough. Similarly, knowledge-based tools should be written to better interoperate with other infrastructure components. Configuration of web, libraries, configuration management, build software, etc. consume millions of dollars in large, real-world projects. Current rule's tools behave in an unacceptably different manner than more traditional software with respect to these factors.

Advances Needed

Three areas seem to need significant attention for improving the reuse and sharing of knowledge. The first two are more obvious so we just touch on them: better repositories (and repository infrastructure) and ontology merging. Better repositories are crucial and more significant adoption of web-based differencing presumes that the most difficult, costly and highest risk components are solved. Repository problems can cost millions to solve.

Sharing and combining ontologies offers a great cost-savings, and so would provide great cost-benefit to business. Enabling normalizing one businesses ontology with another would foster customer-supplier relationships. Similarly, it would similarly foster building new applications and porting applications from one domain into another. Many industries realize the problems with, but still rely on, legacy "stove-pipe" systems.

Natural Language Understanding (NLU) offers obvious advantages for extracting knowledge from domain documents, web resources, etc. However, the less obvious (and perhaps more immediately useful) application of NLU is in generating domain and application documents from knowledge. Because many applications, including in both the military and financial domains, use non-standard, imprecise vocabulary and grammars, more immediate impact could be realized by creating contracts from business logic than trying to extract business logic from domain documents. Thus, probabilistic grammars, probabilistic dictionaries, and domain-specific "translators" could readily provide great value to business. Such translators could address either business problems (e.g., writing contracts) or system artifacts (e.g., load management) and be of great value.

Accommodating Public Representation and Interchange of Knowledge

Fostering the exchange and reuse of knowledge across applications, enterprises and domains can be considered at two levels: enabling reasoning and fostering commercial usage. Production rules, decision tables, and decision trees are orthogonal, so the types of representations examined in the workshop can be represented for purposes of reasoning using the same mechanisms (whether using JSR94 or another representation such as Ontolingua, KIF, etc.) However, the practical sharing of knowledge must consider numerous other metadata aspects for reuse:

- probabilities leading to a decision being made,
- sensitivity of knowledge for competitiveness,
- audit trail of information use within an organization for contract compliance,

- traceability of corporate responsibility,
- intellectual property ownership,
- property reimbursement for use,
- audit trail for evidence validity – including requirements, test cases, etc.,
- identification and versioning information, and
- approval status within an application.

We note that some of these would be best to incorporate directly in the representation because they could dramatically affect reasoning (e.g., probabilities leading to a decision) while others would be best represented as metadata.

Some of these pieces of data should be embedded in sophisticated licensing terms to ensure permitted, proper use. For example, our Financial Client 1 (FC1) offer services to their Financial Clients 2, 3 and 4. Financial Client 2 (FC2) offers some of the same services as our client Financial Client 1, so FC1 only wants a specific system in FC2 to be aware of the knowledge. Licensing structure contained within the knowledge must ensure that only a particular application at FC2 examines the knowledge. Similarly, the system and application must be properly identifiable, must handshake any compensation, internally note its use of the knowledge, validate the appropriateness of the knowledge, etc. Hence, it is the management of the exchanged knowledge – even when contained entirely within the consuming organization - that we feel requires more advancement from the state-of-the-art than the actual reasoning enabled.

Conclusion

The web offers great possibilities for enabling knowledge reuse to improve the cost-effectiveness of intelligent applications. Increasing the adoption of knowledge-based technology in industry requires that more effective tools, especially in terms of repositories, infrastructure, and knowledge management. We also discuss how such infrastructure must manage the security, use and intellectual concerns of knowledge to be exchanged, even after internalized within an organization.