Semantic Web Development

Technical Proposal

February 4, 2000

BAA 00-07
Program title:
    DAML
Proposal title:
    Semantic Web Development
Prime offeror:
    Massachusetts Institute of Technology Laboratory for Computer Science and
    World Wide Web Consortium
Principal Investigator:
    Tim Berners-Lee, MIT W3C<timbl@w3.org>
Co-Principal Investigators:
    David R. Karger <Karger@mit.edu>, Lynn Andrea Stein <las@lcs.mit.edu>, Ralph
    R. Swick <swick@w3.org>, Daniel J. Weitzner <djweitzner@w3.org>
Sub contractors:
    Two to three individuals acting as MIT staff.
Technical Contact:
    Ralph R. Swick, <swick@w3.org>
Administrative Contact:
    Paul Powell, <ppowell@mit.edu>
Type of business:
    Educational
Duration of effort:
    4.25 years

A. Executive Summary

When the World Wide Web was proposed a decade ago it was envisioned not only as a medium for human communication but also one of machine communication. The second half of that hope is as yet unrealized, with the frustrating result that vast amounts of data available to the human enquirer cannot practically be analysed and combined by machine. At the outset of the Web, the field of hypertext was one which had shown much initial promise but little wide scale deployment. Its conversion to a global scalable system was to change that. In the meantime, much work in knowledge representation (KR) -- in ontology, interchange languages, and agent infrastructure -- has demonstrated the viability of KR as a basis for agent interaction while simultaneously highlighting the importance of support for heterogeneity and decentralization. At present, the field of knowledge representation has shown much initial promise but little truly widespread deployment.

The Semantic Web concept is to do for data what HTML did for textual information systems: to provide sufficient flexibility to be able to represent all databases, and logic rules to link them together to great added value. The first steps in this direction were taken by the World-Wide Web Consortium (W3C) in defining Resource Description Framework (RDF) [Lassila et al. 1999], a simple language for expressing relationships in triples where any of the triple can be a first class web object. This basis has the decentralized property necessary for growth. The proposed project is to utilize and demonstrate the great power of adding, on top of the RDF model (modified as necessary) the power of KR systems. We refer to this augmented language as the Semantic Web Logic Language, or SWeLL.

We propose to build on the DARPA Agent Markup Language (DAML) infrastructure to provide precisely such an interchange between two or more rather different kinds of applications. The first of these involves structured information manipulations required to maintain the ongoing activities of an organization such as the W3C, including access control, collaborative development, and meeting management. In the second, we will address the informal and often heuristic processes involved in document management in a personalized information environment. Integrated into both environments will be tools to enable authors to control terms under which personal or sensitive information is used by others, a critical feature to encourage sharing of semantic content.

Optionally, we will also explore applications to spoken language interfaced discourse

systems, automation and automated application construction, and intentional naming of networked resources (by function rather than by a fixed naming scheme). By using the uniform structure of the Semantic Web in each of these applications, we will demonstrate the ability of this technology to build bridges between heterogeneous components and to provide next-generation information interchange.

The Internet being a medium whose level of inherent security is very low, it is expected that digital signature technology will be essential in verifying the steps involved in most discussions. Indeed, the fundamental rules controlling input to a system will not simply be logic, but will combine semantics with the security of digital signature. The project's access control system will employ digital signature. We expect to incorporate digital signature in a way consistent with industry work on digital signature for Extensible Markup Language (XML) [Eastlake].

The project will involve the creation of interoperating systems to prototype the Semantic Web ideas. These will necessarily include simple tools for authoring, browsing, and manipulating the language underlying these applications, as well as the systems themselves.

The work will be deployed along three axes: by adoption by W3C staff internally and by partners such as LCS Oxygen and other DAML participants; by dissemination through and codevelopment with the Open Source community, and when appropriate, by facilitation of consensus around interoperable standards for the Semantic Web using the W3C processs.

B. Innovative Claims and Program Vision

This project is a key move in the transition to a *Semantic Web* software and information environment, in which complex applications can interoperate by exchanging information with a basis in high level logic and well-defined meaning. The project is constructed from software modules which, while performing quite dissimilar functions, exchange information using the languages of the Semantic Web, including the DARPA Agent Markup Language (DAML). The project will form a practical testbed of these ideas. If successful, it will also provide a model to lead both military and commercial development in the future, generating a critical mass of existing systems on the Semantic Web to trigger wide growth of the technology.

We are particularly concerned with information management and information flow on the Semantic Web. Two contrasting subsystems are proposed as the essential basis of our development, with a number of optional projects. One system is an organizational management system which implements the social and managerial processes -- the "social machines" of an organization, the World Wide Wide Consortium (W3C). This system uses well-defined logical algorithms to implement such things as resource management (meeting planning, calendars, etc.), secure web-site access control, and tools to enable individuals to control use of personal information contributed to the Semantic Web. The W3C's "Live Early Adoption and Demonstration" policy of using its own technology makes it a particularly appropriate testbed.

The other central system is a personal and group information system based on our previous *Haystack* work. Haystack [Adar et al.] is a personal information environment: It stores a particular individual's documents and other on-line information, together with automatically generated annotations concerning the context in which the information is encountered, providing content- and context-based retrieval. Haystack also uses heuristic algorithms to analyze personal activity to find patterns and relevant resources within a user's information environment, making it an interesting contrast to the more deliberative W3C systems. While the two systems differ on a scale from secure enforcement of rigid policy to enhanced guesswork about relevant material, they share an internal model for the storage of information which maps well onto the XML-based Resource Description Framework (RDF) developed by W3C. This makes these two systems excellent candidates for integration using Semantic Web techniques. Using RDF to interchange raw data and higher level logic languages to be developed within the DAML community for the exchange of inference rules, the crucial step is to gain added

value from interoperability at the semantic level.

In the past, such interworking would have only operated at the level of raw data. While the systems would have been able to share, for example, the fact that a person was a member of a group, they would not have been able exchange the reason why. To ask why is to ask for a justification -- a series of steps by which the fact was derived. The Semantic Web Logic Language (SWeLL) -- part of the DAML family -- will allow this justification to be passed between systems.
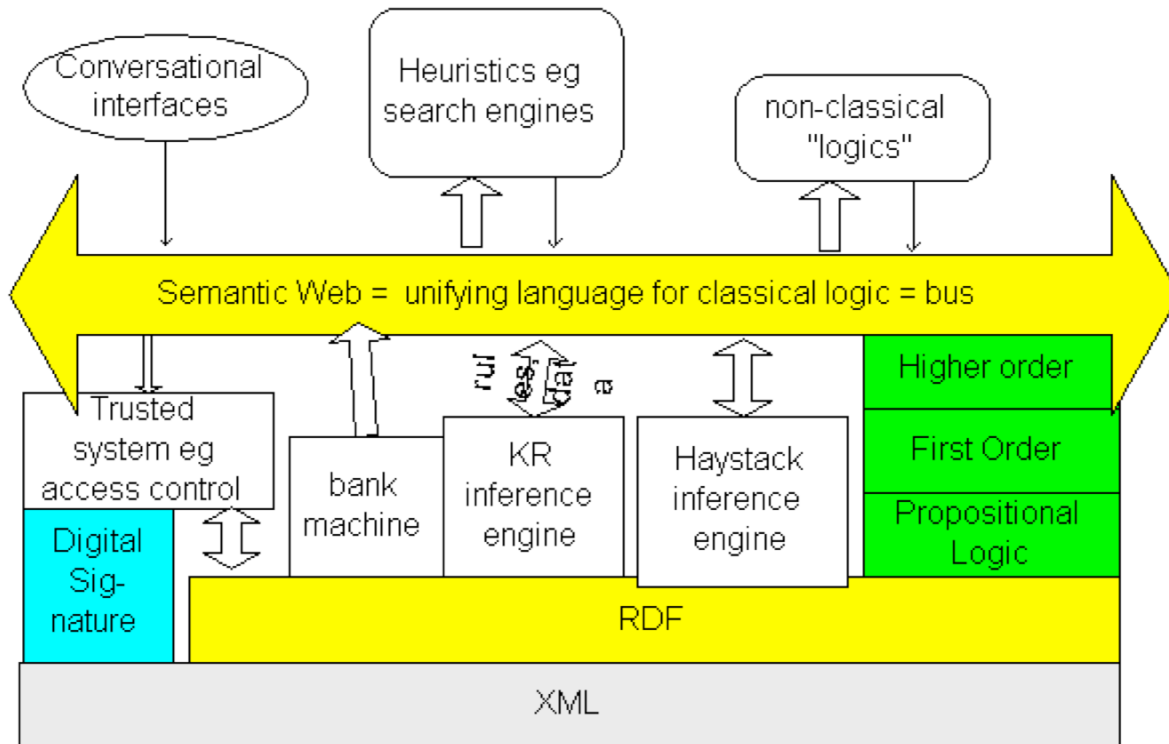
The World Wide Web currently links a heterogeneous distributed decentralized set of systems. Some of these systems use relatively simple and straightforward manipulation of well-characterized data, such as an access control system. Others, such as search engines, use wildly heuristic manipulations to reach less clearly justified but often extremely useful conclusions. In order to achieve its potential, the Semantic Web must provide a common interchange language bridging these diverse systems. Like HTML, the Semantic Web language should be basic enough that it does not impose an undue burden on the simplest web software systems, but powerful enough to allow more sophisticated components to use it to advantage as well.

The Semantic Web, then, will serve as an interchange "bus" for on-line data. In effect, it will allow any web software to read and manipulate data published by any other web

software. SWeLL will enable this interchange. The pairing of simple, predictable, reliable systems with complex, unpredictable, heuristic systems is one of the novel possibilities opened by the Semantic Web. This combination of systems is new ground which this project will specifically explore.

In importing structured information into a heuristic system such as Haystack, we are able to enhance the operation of the heuristic system. The Haystack system currently looks at the contents of documents, and a user's interactions with them, in order to determine things such as subjective "quality" and relevance". Imagine now that all information available to the user from the W3C administrative system is imported into Haystack. The Haystack inference engines will be able to flow semantic links through the social structure and be able to include concepts of W3C documents status. The information space is much richer. When before two documents might have seemed syntactically related, the versioning system might reveal that one was in fact an earlier version of the other. The structured information imported from W3C is also enhanced by being integrated into the semantic network of a haystack, making it addressable through the heuristic retrieval machinery built into Haystack.

In the other direction there are at least two interesting ways in which heuristic systems can export information -- through the semantic web -- to hard logical systems. In one case, a heuristic is used to find a solid argument, which can then be directly accepted by the logic system (e.g., using the proof verifier). In the other, a heuristic system (such as a search engine) cannot logically justify its results, but has achieved such a level of success that those results are authorized for input into an otherwise strictly logical system as hard data. In this case, the accepting system is simply taking the heuristic system's word for it, rather than analyzing the justification further. Digital signature here plays a crucial role in ensuring the validity of such assertions. Personal information management schemas will create a trusted environment in which authors of semantic content will be willing to open their metadata to other's haystacks.

C. Software Components Proposed

The research goals set out in Section B require the design of software components to enable users and their machines to interact with metadata on the Semantic Web. These new Semantic Web tools will be designed with an eye toward general application, but targeted initially toward the two testbed environments of this project: Semantic Web-based W3C organizational tools and Haystack interfaces with the Semantic Web.

W3C Organizational Tools

The W3C operation is distributed globally with over 380 Member organizations. With over 50 active working groups including participants from hundreds of member representatives, numerous documents in being edited by many collaborators, active online publishing of these documents are various stages of maturity, the W3C makes an ideal testbed for development of decentralized Semantic Web tools. As the Web community at-large depends heavily on the technical information presented at the W3C site, these tools will also be tested by the pressures of the public Web.

We propose to build a suite of Semantic Web-powered tools to assist in the management of and participation in the W3C's work. The basic functions of the tools will include:

- Semantic content authoring tools
- SWeLL-based proof generators
- SWeLL-based proof validators
- Access control: controlling access to documents on various parts of the W3C site
- Personal INformation Schema (PINS) controls: using DAML and SWeLL, users will be able to attach conditions to use and re-use of information they contribute to the Semantic Web

Our plans to build organizational management tools with Semantic Web components is illustrated by our proposed approach to access control. Employees of W3C Member organizations are granted access to portions of the W3C Web site that may differ from the access granted to non-Members. Today, W3C webmasters maintain several databases to manage access permissions for all members. Using Semantic Web tools, we wish to delegate to the Member company primary contact and to the chairs of the working groups the authority to authorize individuals to join restricted mail distribution lists, to publish Web pages at specific locations on the W3C Web site, to register for meetings and workshops, and to access materials on other parts of the Web

site. This delegation must be done by following a chain of assertions regarding the resources being access controlled and the individuals whose access is being authorized or verified. To enable this the software components required include:

- A web browser or a http proxy modified to send proofs of access (these will likely be based on the Open Source W3C Amaya or Mozilla and W3C Jigsaw platforms);
- An application for generating access proofs in simple cases for that browser;
- A web server modified to accept SWeLL access justifications;
- A generic SWeLL proof checker employed as access verifier by that server;
- A version of that checker modified to include XML Digital Signature verification.
- A PINS tool for managing working group membership information while preserving social constraints.

In addition to the access control system, we will use the Semantic Web tools as building blocks to help bring functions such as document version management, issue tracking, personal information management, and event scheduling.

Haystack

The Haystack system supports the storage, organization, retrieval, and (eventually) sharing of an individual's documents and other knowledge. Haystack was designed with a rich data representation that is quite similar to the RDF data model. Haystack already contains numerous tools that gather information about the user into this data model, as well as tools for modifying, searching and navigating this data model. Much of the data within Haystack is generated by automated services that "reason" about extant data in order to generate new data.

We will augment Haystack to import data from and export data to the Semantic Web. As numerous applications begin to make their information available in the Semantic Web format, it becomes easy for Haystack to import this information to augment its user's information space. And because of the similarity of Haystack's model to RDF, it is also easy for Haystack to export its own information for access by other Semantic Web enabled systems. This requires two basic tools:

- A Semantic Web parser for Haystack
- A tool that exports Haystack data to the Semantic Web, providing URIs for internal Haystack objects

Haystack is designed as a *personal* information repository; we therefore expect and hope that over time each Haystack will develop a specialized vocabulary of metadata

tuned to its own user. This makes the exchange of information challenging, both between multiple Haystacks and between Haystack and other tools that do not have knowledge of this specialized metadata. The Semantic Web will provide the mechanism for this information exchange. We propose the following components.

- A component that exports the "meaning" of a user's specialized metadata for use by other systems
- A component that imports Semantic Web assertions and uses them to convert external data to conform to the Haystack user's specialized metadata.
- A component that exports proofs explaining *why* a given piece of information is present in a Haystack, so that other visiting Haystacks can decide whether they too would like to believe that piece of information.

While the Semantic Web will support the exchange of information from different vocabularies, human assistance will be required to define those vocabularies and the connections between them. Users will need tools to define their own metadata structures for others. And they will need to examine others' metadata structures in order to relate them to their own. We propose the following components.

- An authoring tool that will let users describe their metadata and the rules their Haystack uses to generate it.
- A browsing tool that lets users examine the Semantic Web information at another site and guide their Haystack in relating that information to their own metadata.

General Shared Components

Haystack will also exploit the general purpose components being designed for the Semantic Web. We will specialize the following components for use in Haystack.

- The proof constructor will be used to explain how certain Haystack information was generated. This will require augmenting the proof system with fuzzy notions of relevance.
- The proof verifier will be used (essentially unchanged) to decide whether to make use of proofs generated by other systems
- The access control system described previously will be used to limit the sharing of Haystack's information to authorized individuals.

Personal Information Management Preferences

In order to encourage users to share semantic content, we will provide a Personal

Information management Schema (PINS) tool to enable users to attach usage restrictions to information, especially personalized information, made part of the Semantic Web. The tools that enable users to interact with Semantic Web content will be aware of these preferences and help users' applications to follow the permissions attached to metadata by its original creator.

Full Oxygen Integration Option

As an option, we propose to explore strategies for integrating the Semantic Web tools developed in the early phase of this project with the LCS Oxygen Project. While Haystack, an Oxygen component, is already a leading part of this DAML initiative, we propose to create links between this and other Oxygen components including tools to take advantage of advanced work on speech, automation, and networking.

D. Technical description

The power and explosive growth of the World Wide Web is based on two very simple concepts; a way of creating names for anything that is accessible on the Internet and the one-way *navigational* link from a document to such a nameable thing. These navigational links direct the readers' attention to other Internet resources, but they do not capture the author's reason for having created the link. To capture such a relationship, today's web authors use natural language, making such information all but inaccessible to automated processing.

Over time, a community builds up a shared vocabulary for the concepts with which it works. Separate communities that have similar experiences will have common concepts but different vocabularies. In the Semantic Web, we will use the naming power of the World Wide Web to attach names to the concepts as used by each community and we will have a new kind of link -- a link with meaning -- that describes the relationships between the names used by one community and the names used by another. Thus, the Semantic Web will contain bridges between these islands of local (community) vocabulary and will enable extended information access across community boundaries.

An important feature provided by the Semantic Web is the ability to respond to the query "What is the basis for that statement?"; to help a user to answer the question "Do I believe that?", or alternatively, "How much risk is there to my achieving my objectives if I act on the basis of that statement?". The language of the Semantic Web must thus be able to represent information destined for automated processing. Resource Description Framework (RDF) is a existing industry language which will form the basis for this language, which we refer to as SWeLL, the Semantic Web Logic Language. RDF is a flexible and general language for the expression of assertions and allows the exchange of raw facts between heterogeneous systems (see diagram).

Example: Simple Search Made Sensible

We provide an example illustrating how the Semantic Web permits bridges between the vocabularies developed by independent communities.

Today, we might ask "What is bmw7's power?" and get no answer back because BMW's web page gives the answer in German terminology. With the Semantic Web, we might retrieve an answer based on heuristically driven classical logic:

- What is bmw7's power?

1. search(bmw7, power, x) yields a nil result
2. search(bmw7, subclass, x) yields x=beamer, car, thing
3. search(beamer, power, x) yields a nil result
4. search(power, equivalent, x) yields x=Macht
5. search(bmw7, Macht, x) yields x=160kw
6. deduce(bmw7, power, 160kw)

- The bmw7's power is 160kw.

And, of course, the Semantic Web allows us to investigate further:

- Why?
  Because power is the same as Macht and the catalog from BMW says the Macht of the bmw7 is 160kw.

In general, individual queries to the Semantic Web will be very specialized circumscribed requests made by special-purpose applications. Want-to-buy requests, invoices, web searches, and even check-writing are each examples of such transactions. These Semantic Web applications do not rely on the full power of SWeLL to make their requests. And, when processing these requests, the fulfilling applications will likely use heuristics or restricted inference rules in order to answer the requests. Inference engines typically have a query language specially designed to constrain the question so that the answer can be produced in a tractable fashion. This is how most systems, including the components of this project, work. However, interchange between components can be embedded in a sufficiently general language to ensure that virtually any components have the potential to interoperate over the Semantic Web.

Logic of Authority

In practice -- in a distributed decentralized world -- some sources will be more reliable (on some topics) than others and different sources may even assert mutually contradictory information. RDF is based on XML for which digital signature standards are expected to emerge early in the life of this project. XML Digital Signature will be used to provide an authentication component in this project. SWeLL processors will generally be aware of the source and authority for each piece of information. Information will be processed differently as a function of trust in its source.

It is not enough for facts to be shared between different data systems, or even for signed data to be shared. Several logic-based languages have been used with some success as knowledge interchange languages (e.g. SHOE, KIF, etc.). In these systems

RDF-like facts are augmented with rules that combine logical information with directives for effecting inference. Such directives are not always exchangeable between different types of KR systems, but the logical relationships must be expressible in SWeLL. For this, SWeLL extends RDF by including negation and explicit quantification.

But to achieve its full potential, the Semantic Web must do more. The Semantic Web is composed of heterogeneous systems. No real system can answer every possible question about its data (i.e. none can compute the deductive closure of its information). Different systems make different compromises, affording different sets of inferences. For this reason, one application will often be able to reach a particular conclusion when another does not.

In the general case, one piece of software will need to be able to explain or justify to another why its assertions ought to be accepted. This allows an untrusted system to effectively communicate information, by compellingly justifying that information. The steps of such a justification amount to a logical proof of the assertion, and the final key element of the Semantic Web is a proof verification system that allows justifications and inferences to be exchanged along with raw data. Thus, a heuristic of one system may posit a relationship (such as that a person may access a web page) and succeed in finding a logical justification for it. This can be conveyed to a very rigid security system that could never have derived this relationship or its justification on its own, but which can verify that the justification is sound.

The Semantic Web works by allowing each individual system to maintain its own language and characteristics. The access control system works with finite sets of people, groups, and documents, and allows constrained expressions for access control. Haystack uses a certain set of user-programmed rules rather than a general inference engine. Prolog, SHOE, Algernon and KIF systems in turn make different decisions regarding their inference languages, and so have correspondingly different characteristics.

Organizational Modeling: A Structured Application

The W3C uses a mixture of database and web-based tools to provide administrative support for day-to-day operations. Specific challenges include balancing rapid response and autonomy of subgroups with general accountability. This challenge is heightened by having personnel distributed geographically around the world. Concepts modeled by the systems include people, W3C Member organizations, individuals, working

groups and other groups of various kinds, mailing lists, and documents. All documents involved in W3C publicly visible or internal work are on the Web. The web servers allow access to an individual web resource to be finely controlled without changing the URI of that resource.

Currently, W3C is at a similar level of development to many organizations in that applications are built as custom web access to databases. However, data exists in a variety of formats and storage media from flat files to relational databases.

There is a great deal of overlap among the W3C data systems, but this shared conceptual space more often translates into redundancy than in useful sharing of subsystems. For example, the concept of group membership exists independently in web access control, W3C membership management, mailing lists administration, and meeting registration. In an ideal world group membership would be delegated using a flexible trust system which would map the practical nature of real-life trust relationships.

Our proposed system uses Semantic Web techniques to expose this organizational data consistently. For example,

1. The W3C membership table is exposed at the W3C web site in RDF;
2. The list of those authorized access though a member company may be delegated to a remote web site;
3. The membership of a working group may be delegated to the chair;
4. The access control of files is stored as metadata close to the file, and determined by the file owner;
5. One or more annotation servers allow a third party to store information about items for which they have no write access;
6. and so on

In this context, we expect to demonstrate the operation of systems which cross domains such as one or more of the following.

1. Generic query engines will perform queries crossing domains;
2. Validation may be performed of organizational invariants across domains;
3. New application building may be done using generic tools;
4. An access proof construction client will be able to use heuristics to find a justification for access, and the access control system will be able to validate its correctness;

5. The heuristic models of haystack will be able to interoperate with the organizational management systems.

The expectation is that during the life of the project, code will be available for the validation of XML digital signatures. This will allow these systems to check information strictly with regard to its source, and particularly a validated signature.

Example Scenario:401: Oh Yeah? Prove you have access!

The Semantic Web will allow us to model social processes, making facts and rules available for sharing between heterogeneous implementations of those social processes. For example, consider access to the [W3C web site](#).

A portion of the W3C web site is confidential to the W3C Membership. Traditional web server facilities allow us to create a w3c-member group, issue accounts that belong to that group, and limit access to certain paths within the web site to that group.

But this is a poor model of the actual social process, and the W3C spends a lot of administrative resources bridging the gap: the right to access the member confidential portions of the W3C web site is actually granted to any W3C Member representative, and each W3C Member organization has a distinguished Advisory Committee Representative (AC Rep) who has the right to appoint other representatives from his/her organization. We have extended the traditional web server facilities to model groups within groups, but we have not automated the various policies that AC Reps use to appoint representatives; for example, the ACME AC Rep may say "do not grant anyone the right to represent ACME to W3C without my explicit approval" while the Zebco AC Rep says "everyone who can receive mail in the Zebco.com domain may act as a Zebco representative to W3C." The W3C system administration team has to remember and manually implement these policies. Even this is a simplification of the overall confidentiality policy; W3C has other classes of individuals such as invited experts, liaisons, and other exceptions.

In the Semantic Web we can represent the whole range of policies as logical assertions using classical logic with quoting and some axioms about digital signatures. We replace the fixed structure of groups and accounts in the web server with a component that verifies assertions of the form

Kwebmaster assures "hasAccess(Krequestor, pageID)"

where

K assures "S"

means that the statement S was signed with K or can be deduced from statements signed by K; i.e. we have the following axioms:

```
(S=>T) => ((K assures "S") => (K assures "T"))
(K assures "S") /\ (K assures "T") => (K assures "S /\ T")
```

Next, the W3C Director delegates appointment of member representatives to AC Reps by publishing a signed assertion:

```
Kdirector assures
"\forall org, Kacrep, Krep (ACRep(Kacrep, org)
/\ (Kacrep assures "memberRep(Krep, org)")
=> memberRep(Krep, org)"
```

And in the course of ACME joining W3C, the W3C Director certifies the key of Wiley, their AC Rep:

```
Kdirector assures "ACRep(Kwiley, ACME)"
```

Wiley appoints John Doe to represent ACME to W3C by publishing a signed assertion:

```
Kwiley assures "memberRep(KjohnDoe, ACME)"
```

The W3C Webmaster records his/her trust in the W3C Director on matters of member representation ala:

```
Kwebmaster assures "\forall org, rep ((Kdirector assures "memberRep(rep, org)") =>
memberRep(rep, org))"
```

Finally, the secretariat of a member confidential meeting arranges that Kwebmaster sign a statement that the meeting record is member confidential:

```
\forall K, org (memberRep(K, org) => hasAccess(K, meetingRecord))
```

When John Doe requests the meeting record, he will get an HTTP authentication challenge. To meet this challenge, he must prove to the web server that

```
Kwebmaster assures hasAccess(KjohnDoe, meetingRecord)
```

This follows from the above: the webmaster grants access to that page to any member representative, and trusts the W3C Director on matters of membership representation; the W3C Director, in turn, delegates that decision to AC Reps; Wiley has appointed John Doe a representative of ACME, and so he is granted access.

The proof-checking machinery required in the web server to make use of the statements above is similar to previous work done in TAOS [Wobber et al. 1993] with BAN logic [Burrows et al. 1990]. In the Semantic Web project we propose to show that proof checking on access can be scalable to the entire Web.

Note, too, that in principle the above statements can be recorded anywhere; they needn't be stored inside the trusted computing base of the W3C web server, and the W3C administrative staff needn't be aware that Wiley has appointed John Doe a representative of ACME. This obliges John Doe to assemble all these facts and arrange them into a proof when submitting his request for access to the page.

In practice the web server will have a limited ability to construct proofs for common cases of authorization, in addition to its general ability to check these proofs. But member organizations can implement arbitrarily complex policies without burdening the W3C administrative team by publishing their own signed assertions and implementing their own agent for generating authorization proofs. Computation of these assertions may be by heuristic, rather than by rigorous logical calculation, provided that the exported proof still meets the standards of SWeLL and the access control system's proof verifier.

Information Management: An Unstructured Application

Where the day to day management of the W3C is a well-structured and largely deterministic process, many web applications are more heuristic and less sharply defined. Inter-document relationships and relevance, as manipulated for example by web search engines and recommender systems, are exemplary of this less formally structured kind of relationship.

Haystack is a system that we have built for information management. In its current incarnation, we have focused on the management of individual corpora, i.e., the information with which a single user interacts. Modern information management tools -- such as information retrieval and data mining algorithms -- have been effective at facilitating interaction with global corpora such as the World Wide Web. This one-size-fits-all approach to information management has been roughly analogous to a public library: the corpus is large, broad, and impersonal, and the retrieval tools are relatively insensitive to individual and contextual differences among their users. Haystack's focus on individual corpora and personalized information management is more like one's own bookshelf. Your own haystack collects the information with which *you* interact,

captures the *contexts* within which you use it, and *learns* how you organize that information. So, for example, you can ask your haystack for "the paper I just read that was on the same topic as the email I forwarded to Pat last Thursday." This query mixes content ("topic") and context ("email I forwarded..."); it relies on interconnections ("same topic") and interactions ("paper I just read") that are meaningful only to its particular user.

Haystack already allows browsing and annotation of the Haystack internal data model. Haystack's internal data model is a lot like RDF and we will be providing translators between these two models for import/export of Haystack data. We will adapt the Haystack UI to browse and annotate Semantic Web information. This includes some rudimentary graph browsing as well as an attempt at semi-sensible "structure" browsing.

At the same time, the power of individual haystacks would be significantly enhanced by connecting them; this is a direction we will undertake using the Semantic Web. Haystacks represent the information environments of their users. It would be of great benefit if individual haystacks could be used as proxies for collaboration. After all, if you can't find what you're looking for on your own bookshelf, you are more likely to turn to your neighbor -- either for access to his bookshelf or to ask for advice -- than to head to the library. Under the current proposal, we intend to address exactly this issue. One's own haystack contains a substantial collection of one's own intelligence. It represents the areas of one's expertise. And, having learned one's perspective on that information, it embodies one's own biases and insights. By allowing one haystack to query another, we facilitate information sharing, expert identification, and community building. This will be accomplished by adding DAML facilities to Haystack, allowing one haystack to communicate with another using the power of the Semantic Web.

Information management is the heart of the W3C operation. During the course of developing a specification a large corpus of contextual information is created. This corpus includes discussion threads and rationale for each portion of the specification and detailed issues with their resolution. The W3C process requires consensus but provides for recording of dissenting opinions as a means to allow consensus to be reached. When a draft specification is later presented for advancement to a higher status, the issues list and minority statements should become part of the review process. The information management system should facilitate discovery of common threads of discourse. This is especially challenging when those threads are separated in time and

when the terminology used by the discussants changes and evolves. In the Semantic Web we intend to investigate the use of Haystack's heuristics for finding related information and develop an augmented set of authoring tools for annotating discussion threads and the evolution of terminology.

This of course introduces significant privacy and information control issues, not unlike those of the W3C's access control. We intend to adapt the W3C ACL system (to be developed under this project) for use in inter-haystack collaboration and information export. There may be substantial differences between the specific rules of access control for the W3C's activities and those of information sharing among haystacks; but this is precisely the kind of flexibility that the Semantic Web is designed to support. Haystack may accept different justifications from W3C ACL, but the underlying proof language and much of the verification infrastructure will be shared between these systems. Accordingly, we expect to adapt the Haystack UI to support manual proof verification, i.e., to browse the proof and accept/reject statements with human assistance.

As Haystack deals typically with subjective and vaguely-defined terms such as relevance and information quality, the algorithms it uses are generally heuristics controlled and adjusted by the user. This sets it apart from the mechanical, logical operations of the W3C operational machinery. The systems are in that way quite distinct. The real power of the Semantic Web will be evidenced not by inter-haystack sharing, but by the ways in which this same DAML integration allows information sharing between a haystack and a radically different information producer or consumer.

Consider, for example, the export of Haystack data to the access control system. Suppose (for some reason) a person wants to make access to their research papers available to interested academics in the field, but not to the general public. A haystack can be used to heuristically generate a collection of names of authors of papers whose contents match the specified field, together with their email addresses. This information can be exported in one of two ways. First, the haystack can export a list of the email addresses of all qualifying individuals as an RDF file on the Semantic Web. She or he then introduces into the formula for access to the documents a term which allows access by people whose email addresses are in the list. The secure system has been allowed to trust -- for a specific purpose only -- the output of a rough and ready heuristic. More extremely, the Haystack can produce, on demand, a SWeLL justification as to why a particular email address should be included. This justification includes reference to relationships, such as document relevance, that are meaningful only within Haystack.

But it also includes reference to other relationships, such as the Dublin Core's creator metadata, that may be directly manipulable by the consuming system.

Vision

There is, of course, no reason to limit communication over the Semantic Web to these two applications. Indeed, in options to this proposal, we intend to develop a variety of other DAML applications that will similarly benefit from exploitation of the universality of communication over the Semantic Web. We would then expect to create, for example, a spoken language interface to the information contained in one's Haystack; an intentional naming scheme that allows one to locate W3C working groups with relevant interests; or construct a complete vacation planner on the fly.

This project is a test on a finite scale of a system designed to scale in an unbounded fashion. Aspects of the Semantic Web's features supporting smooth evolution of technology will only be tested thoroughly when similar but independent systems are independently designed and later merged by the creation of linking rules. For example, if another organization were to model its processes independently and later connect the shared concepts (such as "meeting") to W3C's so as to directly enable shared calendars, then we would have a success.

Options

The Oxygen project, currently underway at MIT, brings numerous LCS and AI lab researchers together to build a ubiquitous computational infrastructure involving both handheld and embedded computational units. The projects fitting into Oxygen include speech technology (as a user interface), Haystack (for knowledge representation) and the Intentional Naming System (for naming and resource discovery). All of these systems could benefit by sharing a common representational infrastructure using the Semantic Web. Haystack has already been discussed; here we discuss the other potential integrations.

For more than a decade, the Spoken Language Systems Group has been conducting research in human language technologies and their utilization in conversational interfaces -- interfaces that can understand the verbal query in order to help users create, access, and manage information. Their technologies, which have been ported to eight languages and more than a dozen applications, are based on a semantic representation of the input query, in the form of a hierarchical frame consisting of key-value pairs. The semantic frame is the central point for database access, discourse

and dialogue modelling, as well as natural language generation. Their involvement in the Semantic Web effort will provide a much needed perspective based on human language. Furthermore, many of the tools developed by the speech group work by extracting information from other systems: for example, the Jupiter system responds to weather queries by finding appropriate information on the web---currently in textual form. Representing this information semantically would improve Jupiter's ability to answer queries.

The intentional naming system (INS) [Adjie-Winoto et al. 1999] currently under development at LCS is intended for naming and resource discovery in future mobile networks of devices and services. The idea in intentional naming is to name entities based on descriptive attributes, rather than simply on network location (which is what the DNS does). This allows applications to discover resources based on *what* they are looking for, rather than *where* they are located in the network. This is natural for most applications, since they often can describe the former but do not know about the latter.

INS benefits greatly from the Semantic Web and its associated markup language, since it needs an expressive naming syntax to describe entities (e.g., the closest and least loaded printer near where you're standing right now). It also needs to be able to standardize similar services that might be advertised by their providers with different names, as in one person calling the device a "printer" and the other calling it a "paper output device." The Semantic Web makes this possible.

Because INS incorporates techniques in its resolution architecture to handle service and user mobility as well as group communication, we believe that using the Semantic Web in this system will allow us to explore and demonstrate the benefits of the semantic web in mobile networks of devices and services.

E. Evaluation

The primary means of evaluation to be used in this project are those centered around feasibility demonstration by actual deployment of the technology. Although the earlier experimental phases of this project will require more limited deployment than the later, more finished stages, we anticipate employing our own work to the extent possible as a primary means of evaluation.

In this project, we will integrate the results of our work at the earliest possible stage into the daily work of the World Wide Web Consortium. The philosophy of "Live Early Adoption and Demonstration" (LEAD) has been a operational tool at the Consortium for the last few years. It is an approach that we have found particularly effective in those areas (such as web-based collaboration) where the weakest link has been between the development of technology and its actual deployment. For example, the Consortium's practical use of its Amaya (web authoring and browsing) client and Jigsaw server has shed considerable light on interactive editing in HTTP 1.1.

Because many of the prototypical tools that we are building will be organizational process management tools, we anticipate using these tools in the running of the World Wide Web Consortium. This will put pressure on these tools to be useful and functional as well as robust; by living with them, we will also quickly learn where they must be improved. As these tools become more sophisticated, we will increasingly deploy them in critical-path Consortium operations.

The Semantic Web is primarily a tool for interoperability. This is why we are proposing to build a suite of applications, rather than a single monolithic application. The goal of interoperability between heterogeneous components that we build is one that will test the extent to which the Semantic Web is achieving its promise. The more diverse the systems interoperating, the greater the merit of the Semantic Web. Initially, we expect the process management tools and information management tools to be somewhat separate. As the Semantic Web's interoperability increases, we expect to see further synergies arise between these diverse systems. This will be a reflection of the extent to which the Semantic Web is achieving its goals. The broader set of option components will increase the rigor of this test.

Ultimately, the test of the Semantic Web--and the applications that we build for it--will be the extent to which applications built by others interoperate with ours. The next test is transmission of information between our applications and those built by

other DAML contractors. If the Semantic Web lives up to its potential, even this phase will ultimately be surpassed by one in which non-participant organizations construct applications to exploit this infrastructure. Similarly, we will judge our tools by the extent to which they are ultimately useful to other groups.

An interesting result of the project will be to find how much logical information (from inference rules and other metadata) can in practice be exchanged between systems. It may be that in certain applications much more significant information is in the logic than in any raw data, in which case the Semantic Web will enable a dramatic enhancement of interoperability. This project is designed to get as much advanced experience early on, before very large scale deployment. The diversity of systems successfully interfaced to the Semantic Web will be an important indicator of success. The deployment of the system into the running working processes of the W3C will sometimes mean that practical concerns may dictate what systems can be changed and in which ways at the time of deployment. Therefore the research aspect of the project will have to be adaptable to these practical concerns. However, it will have the advantage of running tests on live data rather than imagined or simulated scenarios, including new real life application needs which arise unforeseen during development. It will also increase the applicability of experience and even running code to second generation adopters of the technology.

H. Proprietary Claims

There are no proprietary claims by an external party on the ideas or products of this project. MIT and the researchers in this proposal have full control over all intellectual and material rights.

I. Statement of Work

This project will contribute to the development of a vibrant, ubiquitous Semantic Web by building critical Semantic Web infrastructure and demonstrating how that infrastructure can be used by working, user-oriented applications.

SWeLL Infrastructure

Basic SWeLL infrastructure components, the building blocks of all Semantic Web applications to be produced by this project, will be built first, and then refined throughout the life of the activity. The basic components are:

- SWeLL parser
- Proof generator
- Proof checker
- Authoring tool
- Schema editor
- Query language

Once the parser, proof generator, and proof checker are built, they will be used together to establish the basic functionality of logical expression in the SWeLL environment. As other components are built, they will be deployed in various combinations to develop SWeLL functionality for both the W3C web site and Haystack.

Haystack export as RDF

A Haystack HTTP server makes Haystack data available as RDF.

Foundation verification

This involves the refinement (if necessary) of SWeLL, the definition of a proof transfer language, the construction of a parser and of a proof validator for that language, and the publication of test data.

Import of W3C corpus into Haystack

The collection and state of W3C documents is made available to Haystack users.

Access validating HTTP server

With the definition of an access control schema and an extension to HTTP, we modify an existing HTTP server to use the proof engine and accept SWeLL requests for authentication. This will be tested with an HTTP client modified to send (hand-written,

not automatically generated) access justifications.

Exposure of W3C administrative systems

At least two W3C administrative systems should be available on the Semantic Web at this stage. This will involve creation of schemas for the systems involved and the generation of custom user agents or (in most cases) the modification of legacy systems to generate SWeLL.

Import/Export of Haystack in SWeLL

The export of Haystack data to the Semantic Web requires the generation of schemas and the conversion of internal formats to RDF. At around the same time we hope to demonstrate the import on demand of arbitrary data from the Semantic Web into a user's haystack. This adds Haystack's user interface to the Semantic Web toolset.

Cross-domain query demonstration

Choosing an appropriate query language, we make a query across previously disjoint domains as a test rather than live system.

Socially meaningful cross-domain operation

A cross-domain query that provides a socially useful function, modelling the real social needs of the organization, for example PINS and ACLs.

Community evangelism and deployment (years 3 and 4)

The pace of uptake of technology in the commercial community at large is difficult to predict in the current enthusiastic environment. As sufficient interest occurs that this is possible, we will coordinate its transition to the W3C standards process.

J. Management Plan

The personnel in this proposal are all university researchers.

L. Technology Sharing

Technology developed under this proposal will be propagated through the very same mechanisms that have enabled the rapid spread of today's core World Wide Web standards. Sharing DAML-based technology will occur along three dimensions, with different user communities becoming involved as the technology increasingly matures.

The first axis of technology sharing is with explicit supporting partners such as World Wide Web Consortium staff and members, the LCS Oxygen environment, and DAML collaborators. The second axis involves sharing with public metadata communities, W3C Metadata standards activities, and related open source and academic communities. Finally, the third axis of technology sharing is to move mature DAML infrastructure to the W3C standards track and, as appropriate, other standards bodies such as the Internet Engineering Task Force.

Axis one of technology sharing will occur within the confines the two chosen DAML testbeds, Haystack and the W3C community. With 60 staff in three sites around the world and over 380 member organizations internationally, the W3C community will provide a substantial testbed for initial application of DAML-based software and services developed by this project. For example, as soon as the DAML-based access control tools are stable, they will be deployed on the W3C web site for use in mediating W3C member access to W3C documents and services. Later, as authoring tools become available, they will be made available to members for use in creating W3C documents and associated metadata. Other tools such as proof checkers and heuristic document management systems will also be deployed to address W3C workflow and collaboration issues. Drawing an analogy between this phase of Semantic Web technology sharing and the original development of the World Wide Web, one may see basic DAML tools serving the same role as the original CERN phonebook did. This first phase of technology sharing will provide the DAML team with vital input on the soundness of our design and begin the process of building broader support for the DAML approach.

Through the life of the Semantic Web activity, the links with LCS's Oxygen project will also provide key technology sharing and evangelism opportunities. In particular, the corporate partners in Oxygen will be a valuable sounding board for assessing the commercial viability of DAML and an important venue for technology transfer from the laboratory to the market. Oxygen research will shed light on semantic web interface

with novel operating system internals and distributed applications.

The second axis of DAML deployment will bring tools and techniques developed with DAML to the attention of well-known public metadata communities such as the Dublin Core bibliographic schema and the OpenGIS framework for representing geographic information. An important avenue of technology diffusion will be the W3C's RDF Interest Group, a global community of metadata developers who work in this group to coordinate their efforts and provide W3C feedback on its metadata standardization efforts. While we do not expect these communities to necessarily accept the entire DAML framework they will provide vital corpora of metadata on which to test DAML. At this stage we also expect increased interaction with the communities developing the XML Signature Specification for attaching digital signatures to XML, as well as developers working on the Platform for Privacy Preferences (P3P), which will have significant overlap with IMP. Finally, in this phase we would expect to use Haystack tools to interact with these public metadata stores.

In the final axis of technology sharing, we will concentrate on bringing critical DAML infrastructure components into the W3C standards development process. This is of course not a unilateral action by W3C staff, and happens only when community consensus is that work on standards would be worthwhile. Deployment along the other two axes will hopefully bring this about. The W3C Process document describes the creation of new standardization activities. As the recognized source of Web standards, W3C is in a unique position to move concepts developed in this project into the public Web. As has been noted, W3C has already completed a first round of Web metadata standards through our RDF work. We expect that DAML will provide valuable guidance to the Web development community as a whole on evolving the Semantic Web. W3C's close liaisons with other standards bodies, especially the IETF and the WAP Forum, will also provide valuable opportunities for technology sharing.

M. Facilities

All work carried out by MIT under this proposal will be conducted at the MIT Laboratory for Computer Science, the major site for computer science research at MIT. The Laboratory provides a basic computing environment for research, including common services such as Internet access and printing. LCS provides administrative support for contract research, which will be utilized by the proposed project. LCS co-ordinates the physical plant provided to the research groups. Equipment and facilities required to support specific research programs, such as computational workstations, specialized network facilities, and dedicated servers and prototyping machines, is acquired on a per-program basis. This proposal includes an equipment budget to acquire such facilities for this project. The required facilities fall into (how many) categories: explanation of computers and servers here. MIT is unable to supply these facilities from institutional resources.

The MIT Laboratory for Computer Science is the United States host for the World Wide Web Consortium The W3C team and research groups are equipped with recent computing systems, and sufficient bandwidth for efficient work with other national and international collaborators, including video conferencing. The Consortium has a collaborative web server allowing delegated authority to work in creative shared spaces, hypertext email archives, an Event Tracking Agent for issue tracking, and automated facilities for meeting organiztion. We have three conventional telephone conference bridges and Internet video conferencing facilities. The Lab generally can offer wireless LAN (802.11+DHCP) access for meeting guests.

N. Government Furnished Property

No Government furnished property is required for this research.

O. Experience

This projects involves a new collaboration between two organizations: The World Wide Web Consortium and the MIT Lab for Computer Science.

The World Wide Web Consortium W3C was established in 1994 in response to a growing demand for a coordination body for the rapid development of Web standards. It was founded by the inventor of the Web, Tim Berners-Lee (PI on this project), and is hosted by MIT LCS in the USA, INRIA in France, and Keio University in Japan, with offices in several other countries. The W3C has over the last 6 years overseen processes of issue raising, design, consensus building and testing resulting in about 20 Recommendations including HTML, XML, XHTML, XML Transformations (XSL), Cascading Style Sheets, and Portable Network Graphics.

The need for common standards for interchange of metadata (information about Web resources) drove the Consortium to create, in the Resource Description Framework (RDF), a general knowledge representation language of limited power and great extensibility. Ralph Swick (co-PI on this proposal) was and is the W3C metadata lead and co-author of the RDF Model and Syntax Recommendation.

The W3C Technology and Society Domain, lead by Daniel Weitzner (co-PI on this proposal) has lead the Web community in developing technology to assist in solving social problems such as privacy and authentication. The PINS information management tools proposed as part of this project will draw heavily on the W3C's work on the Platform for Privacy Preferences (P3P) and work on digital signatures for XML.

The MIT Laboratory for Computer Science (LCS) is an interdepartmental laboratory, bringing together faculty, researchers, and students in a broad program of study, research and experimentation. The Laboratory's principal goal is to pursue innovations in information technology that will yield long-term improvements in the ways that people live and work. The hallmark of its research is a balanced consideration of technological capability and human utility.

Currently, LCS is focusing its research on the architectures of tomorrow's information infrastructures. In the interest of making computers more efficient and easier to use, LCS researchers are putting great effort into human-machine communication via speech understanding; designing new computers, operating systems, and communications architectures for a networked world; automating information gathering and

organization; and the theoretical underpinnings of CS.

One central technology in the current LCS agenda is the Haystack project led by co-PIs David Karger and Lynn Andrea Stein. Haystack combines the function of a traditional text-document retrieval system with the gathering, representation, presentation and management of arbitrary metadata, forming a personal information management tool and a part of the semantic information space for LCS's next-generation computer systems.   Haystack builds on Karger's previous work in information retrieval (including Scatter/Gather) and Stein's research in knowledge representation (especially nonmonotonic logics), software agents (including the Sodabot agent environment in the MIT AI Lab's DARPA-funded Intelligent Room), and web manipulation (e.g., the Squeal/ParaSite structured web interface).

Other LCS members whose work will inform this project include Steve Garland of the LCS Software Devices and Systems group, principal developer of LP, a proof assistant for multisorted first-order logic, and co-designer of the Larch and IO Automata specification languages and analysis tools; Daniel Jackson, co-leader of the LCS Software Design group and author of numerous langauges and tools for describing the structural properties of software (including Alloy/Alcoa, with properties similar to SWeLL); Nancy Lynch, leader of the LCS Theory of Distributed Systems group, author of *Atomic Transactions* and *Distributed Algorithms* and co-designer of IO Automata; and Ron Rivest, co-leader of the LCS Cryptography and Information Security Group, co-inventor of the RSA public key cryptosystem, the MD5 digital fingerprinting algorithm, and the SDSI public key infrastructure.

Optional Semantic Web applications and features draw on additional expertise within LCS. The Spoken Language Systems Group (including Stephanie Seneff) has been conducting research in human language technologies and their utilization in conversational interfaces. Their technologies, which have been ported to eight languages and more than a dozen applications, are based on a semantic representation of the input query. Their involvement in the Semantic Web effort will provide a much needed perspective based on human language.

The LCS Computer-Aided Automation Group, directed by Srini Devadas, is dedicated to exploring and implementing computer-based solutions to problems in the design and control of electronic devices in the home, office, and factory. The group is engaged in research to further automated control of these electronic devices, and to leverage advances in networking technology to allow intelligent products to interact

with one another, further enhancing their usefulness.

Hari Balakrishnan, of the Software Devices and Systems group, works on wireless device networks and intentional naming schemes. His PhD work includes the Snoop Protocol, designed to significantly improve TCP performance over error-prone wireless links, and studies of the effects of bandwidth and latency asymmetry and the effects of small window sizes on TCP performance.

Q. Qualifications

Tim Berners-Lee

A graduate of Oxford University (BA Physics first class, 1976), Tim now holds the 3Com Founders chair at the Laboratory for Computer Science (LCS) at the Massachusetts Institute of Technology (MIT). He directs the World Wide Web Consortium, an open forum of companies and organizations with the mission to lead the Web to its full potential.

With a background of system design in real-time communications and text processing software development, in 1989 he invented the World Wide Web, an internet-based hypermedia initiative for global information sharing, while working at CERN, the European Particle Physics Laboratory. He wrote the first web client (browser-editor) and server, and the original HTML, HTTP and URI specifications, in 1990. Awards include ACM Software Systems Award (1995), ACM Kobayashi award (1996), the IEEE Computer Society Wallace McDowell Award (1996), Computers and Communication (C&C) award (1996); OBE (1997), Charles Babbage award (1998), a MacArthur Fellowship and The Eduard Rhein technology award (1998). Honorary degrees from Parsons School of Design, New York (D.F.A., 1996), Southampton University (D.Sc., 1996), and Essex University (D.U., 1998), Southern Cross University (1998). Distinguished Fellow of the British Computer Society, Honorary Fellow of the Institution of Electrical Engineers. Publications include:

T. Berners-Lee "*Universal Resource Identifiers in WWW*", RFC1630, 1994/6

Berners-Lee, T.J., et al, "World-Wide Web: Information Universe", *Electronic Publishing: Research, Applications and Policy*, 1992/4.

Berners-Lee, T.J., et al, "The World Wide Web," *Communications of the ACM*, 1994/8.

T. Berners-Lee, L. Masinter, M. McCahill, "*Universal Resource Locators (URL)*", RFC1738, 1994/12

T. Berners-Lee, D. Connolly "*Hypertext markup Language - 2.0*", RFC1866, 1996/5

R. Fielding, J. Gettys, J. Mogul, H. Frystyk, L. Masinter, P. Leach, T. Berners-Lee "*Hypertext Transfer Protocol HTTP 1.1*", RFC 2616, 1999/6

Tim Berners-Lee, Dan Connolly, Ralph R. Swick "*Web Architecture: Describing and*

Exchanging Data", W3C Note, 1999 / 6 / 7

Berners-Lee, T., *Weaving the Web*, Harper SanFrancisco, 1999

Dan Connolly

Dan Connolly (B.S. in Computer Science, University of Texas at Austin, 1990) is the leader of the XML Activity in the World Wide Web Consortium. He began contributing to the World Wide Web project, and in particular, the HTML specification, while developing hypertext production and delivery software in 1992.

Mr. Connolly presented a draft of HTML 2.0 at the first Web Conference in 1994 in Geneva, and served as editor until it became a Proposed Standard RFC in November 1995.

Mr. Connolly was the chair of the W3C Working Group that produced HTML 3.2 and HTML 4.0, and collaborated with others to form the W3C XML Working Group and produce the W3C XML 1.0 Recommendation.

Prior to joining the World Wide Web Consortium, Mr. Connolly was a software engineer at HaL Software Systems were he participated in the design and implementation of document database tools including building proxy servers to connect those tools to the World Wide Web.

Mr. Connolly's research interest is investigating the value of formal descriptions of chaotic systems like the Web, especially in the consensus-building process.

Connolly, ed, XML: Principles, Tools and Techniques, World Wide Web Journal (O'Reilly and Associates, 1997)

"An Evaluation of the World Wide Web as a Platform for Electronic Commerce", in Kalakota, Whinston, Readings in Electronic Commerce (Addison-Wesley, 1996)

Berners-Lee, Connolly, eds., Hypertext Markup Language 2.0, Internet Engineering Task Force RFC1866 (1995), http://www.ietf.org/rfc/rfc1866.txt

David R. Karger

David R. Karger (A.B. *summa cum laude* in Computer Science, 1989, Harvard College; Certificate of Advanced Study in Mathematics 1990, Churchill College, Cambridge University; Ph.D., 1994, in Computer Science, Stanford University) is an Associate Professor of Computer Science and a member of the Laboratory for Computer Science. He was the recipient of a David and Lucille Packard Foundation Fellowship (1997--2002), an Alfred P. Sloane Foundation Fellowship (1996--1998), the 1997 Mathematical Programming Society Tucker Prize, the 1995 ACM Doctoral Dissertation award, and a 1995 NSF CAREER grant.

Professor Karger's research interests include algorithms and information retrieval. His work in algorithms has focused on applications of randomization to graph optimization problems and led to significant progress on several core problems. He is the author of more than 20 publications and has served on the program committees of some of the most important theoretical computer science conferences. He is a member of the executive committee of SIGACT, ACM's theoretical computer science group, and on the editorial board of Information Retrieval.

Karger's work on information retrieval includes the co-development of the Scatter/Gather information retrieval system at Xerox PARC, which suggested several novel approaches for efficiently retrieving information from massive corpora and presenting it effectively to users. He has received two patents related to his work on this project. He is presently co-PI on the Haystack personalized information management project.

Professor Karger's World Wide Web home page can be found at http://theory.lcs.mit.edu/~karger/

Publications relevant to proposed work:

- D. Cutting, D. R. Karger, J. Pedersen, and J. W. Tukey. ``Scatter/Gather: A Cluster-based Approach to Browsing Large Document Collections.'' In *Proceedings of the 15th* Annual International ACM SIGIR Conference on Research and Development in Information Retrieval}, pages 318--329, 1992.
- D. Cutting, D. R. Karger, and J. Pedersen. ``Constant Interaction-Time Scatter/Gather Browsing of Very Large Document Collections.'' In *Proceedings of the 16th* Annual International ACM SIGIR Conference on Research and Development in Information Retrieval}, July 1993. Pittsburgh, PA.

Adar, E., D. R. Karger, and L. A. Stein, "Haystack: Per-User Information Environments," *Conference on Information and Knowledge Management* , Kansas City, Missouri, November 1999, pp. 413-422.

Lynn Andrea Stein

Lynn Andrea Stein (A.B. cum laude in Computer Science, 1986, Harvard and Radcliffe Colleges; Sc.M., 1987, and Ph.D., 1990, in Computer Science, Brown University) is an Associate Professor of Computer Science and a member of the Artificial Intelligence Lab and the Lab for Computer Science at the Massachusetts Institute of Technology, where she heads the AP Group. She was a recipient of (among others) the Ruth and Joel Spira Teaching Award, the General Electric Faculty for the Future Award, and the National Science Foundation Young Investigator Award.

Professor Stein is an internationally known researcher whose work spans the fields of knowledge representation and reasoning, software agents, human-computer interaction and collaboration, information management, object-oriented programming, and cognitive robotics. She is author or co-author of more than 50 publications and has served on the program and advisory committees of numerous national and international conferences in these fields. She has recently given invited lectures at the national conference on artificial intelligence (AAAI) and the European Meeting on Cybernetics and Systems Research and will deliver plenary addresses at this spring's International Conference on Complex Systems. From 1991 to 1996 she served as Associate Chair and then Chair of the AAAI's Symposium Committee and from 1995 to 1998 as an Executive Councilor of the AAAI. Recent major publications include:

Stein, Lynn Andrea. *Introduction to Interactive Programming* (San Francisco: Morgan Kaufmann, 2001).

Stein, L. A., "Resolving Ambiguity in Nonmonotonic Inheritance Hierarchies," *Artificial Intelligence* **55** *(2-3):*259-310, June 1992.

Boddy, M., R. P. Goldman, K. Kanazawa, and L. A. Stein, "A Critical Examination of Model-Preference Defaults," *Fundamenta Informaticae,* **21** (1-2), July-August 1994.

Stein, L. A. and L. Morgenstern,, "Motivated Action Theory: A Formal Theory of Causal Reasoning," *Artificial Intelligence* **71** (1):1-42, November 1994.

Stein, L. A. and S. B. Zdonik, "Clovers: The Dynamic Behavior of Types and Instances," *International Journal of Computer Science and Information Management* **1** (3): 1-11, 1998.

Spertus, E., and L. A. Stein, "A Relational Database Interface to the World-Wide Web,"

ACM Conference on Digital Libraries, Berkeley, California, August 1999, 248-249.

Adar, E., D. R. Karger, and L. A. Stein, "Haystack: Per-User Information Environments," *Conference on Information and Knowledge Management* , Kansas City, Missouri, November 1999, pp. 413-422.

Ralph R. Swick

Ralph R. Swick (B.S. Physics and Mathematics, Carnegie-Mellon University, 1977) is a Research Associate in the Lab for Computer Science at the Massachusetts Institute of Technology where he is the Technical Director for the Technology and Society Domain of the World Wide Web Consortium and leads work on metadata within the W3C.

Mr. Swick was the co-editor of the W3C Resource Description Framework Model and Syntax Specification and was one of the principals in the RDF design. Before joining W3C Mr. Swick was Technical Director of the X Consortium where he led the design and development of major portions of the X Window System. Mr. Swick was one of the named co-recipients of the 1999 Usenix Association Lifetime Achievement Awards presented to the X Window System Community. While a member of the corporate research staff of Digital Equipment Corporation, Mr. Swick was active in Project Athena at MIT where he worked on the X Window System and contributed to the Kerberos authentication system. Subsequent to Project Athena Mr. Swick was in the Office Systems Advanced Development Group at Digital where he prototyped several advanced information management and group collaboration tools.

Asente, Converse, Swick. The X Window System Toolkit (Digital Press/Butterworth-Heinemann, 1998).

Berners-Lee, Connolly, Swick, Web Architecture: Describing and Exchanging Data (World Wide Web Consortium Note, 1999) http://www.w3.org/1999/04/WebData

Lassila, Swick [eds], Resource Description Framework (RDF) Model and Syntax Specification (World Wide Web Consortium Recommendation, 1999) http://www.w3.org/TR/1999/REC-rdf-syntax-19990222

Daniel Weitzner

Daniel Weitzner is Director of the World Wide Web Consortium's Technology and Society activities. He was one of the first to propose user empowerment tools for the Web, such as content filtering and privacy protection technology, as means to address global policy challenges on the Internet. As Technology and Society Domain Leader, he is responsible for development of technology standards that enable the web to address social, legal, and public policy concerns such as privacy, free speech, protection of minors, authentication, intellectual property and identification. He is also the W3C's chief liaison to public policy communities around the world and a member of the ICANN Protocol Supporting Organization Protocol Council.

Weitzner holds a research appointment at MIT's Laboratory for Computer Science and teaches Internet public policy at MIT.

Before joining the W3C, Mr. Weitzner was co-founder and Deputy Director of the Center for Democracy and Technology, an Internet civil liberties organization in Washington, DC. At CDT he played a leading role in the Supreme Court litigation establishing comprehensive First Amendment rights on the Internet. He was also Deputy Policy Director of the Electronic Frontier Foundation. Mr. Weitzner has a degree in law from Buffalo Law School, and a B.A. in Philosophy from Swarthmore College.

Selected Publications:

Weitzner, D, Berman, J., "Abundance and Control: Renewing the Democratic Heart of the First Amendment in the Age of Interactive Media," *Yale Law Journal*, Vol. 104, No.7, May 1995

Weitzner, D, "Directing Policy-Making: Beyond the Net's Metaphor," *Communications of the ACM*, Vol. 40, No. 2, February 1997.

Weitzner, D., Berman, J., "Technology and Democracy, " *Social Research* (Volume 64 No. 3 (Fall 1997))

Weitzner, D., (ed.), Browning, G., *Electronic Democracy: Using the Internet to Influence American Politics*. Pemberton Press, 1996.

Weitzner, D. "The Clipper Chip Controversy," *Computerworld* July 25, 1994.

Weitzner, D., Kapor, M., "The EFF Open Platform Proposal," 1992.

http://www.eff.org/pub/Infrastructure/Old/overview_eff.op

He is also a commentator for NPR's Marketplace Radio.

R. Bibliography

Adar, Eytan. *Hybrid-Search and Storage of Semi-structured Information*. Master's Thesis, MIT, May 1998.

Adar, E., D. R. Karger, and L. A. Stein, "Haystack: Per-User Information Environments," *Conference on Information and Knowledge Management* , Kansas City, Missouri, November 1999, pp. 413-422.
http://www.acm.org/pubs/citations/proceedings/cikm/319950/p413-adar/

Adjie-Winoto, W., Schwartz, E.,Balakrishnan, H., Lilley, J., "The design and implementation of an intentional naming system," Proc. 17th ACM SOSP, Kiawah Island, SC, Dec. 1999.

Asoorian, Mark. *Data Manipulation Services in the Haystack IR System*. Master's Thesis, MIT, May 1998.

Baader, F. and Hollunder, B. (1991). KRIS: Knowledge representation and inference system. *ACM SIGART Bulletin* **2**(3) 8-14.

Baldonado, M., C.-C. K. Chang, L. Gravano and A. Paepcke (1997). "The Stanford Digital Library Metadata Architecture." *International Journal of Digital Libraries* 1:2, September 1997, pp. 108-121.

Berners-Lee, T., 1999, *Weaving the Web*, HarperCollins Publishers, 10 East 53rd Street, New York, NY

Berners-Lee, T., and D. Connolly, 1998. *Web Architecture: Extensible Languages*, W3C Note
http://www.w3.org/TR/1998/NOTE-webarch-extlang-19980210

Boddy, M., R. P. Goldman, K. Kanazawa, and L. A. Stein, "A Critical Examination of Model-Preference Defaults," *Fundamenta Informaticae,* **21** (1-2), July-August 1994.
ftp://ftp.ai.mit.edu/pub/users/las/critical-examination-of-MPD.ps.Z

Borgida, A., Brachman, R.J., McGuiness, D.L. and Resnick, L.A. (1989). CLASSIC: a structural data model for objects. *Proceedings of 1989 SIGMOD Conference on the Management of Data*. pp.58-67. New York, ACM Press.

Brachman, R.J., Gilbert, V.P. and Levesque, H.J. (1985). An essential hybrid reasoning

system: knowledge and symbol level accounts of KRYPTON. *Proceedings of IJCAI85*. pp.547-551. San Mateo, California, Morgan Kauffman.

Brachman, R.J. and Schmolze, J. (1985). An overview of the KL-ONE knowledge representation system. *Cognitive Science* **9**(2) 171-216.

Bray, T., J. Paoli and C.M. Sperberg-McQueen. 1998. *Extensible Markup Language (XML)*. W3C (World-Wide Web Consortium) Recommendation.
http://www.w3.org/TR/1998/REC-xml-19980210

Burrows, M., Abadi, M., Needham, R. (1990) *A Logic of Authentication*, Technical Report, DEC System Reearch Center, Number 39
http://gatekeeper.dec.com/pub/DEC/SRC/research-reports/abstracts/src-rr-039.html

Bush, Vannevar. As We may Think. *Atlantic Monthly*, 176(1)641-649, January 1945.

Chawathe, S., H. Garcia-Molina, J. Hammer, K. Ireland, Y. Papakonstantinou, J. Ullman and J. Widom (1994). The TSIMMIS Project: Integration of Heterogenous Information Sources. *IPSJ Conference*, Tokyo, Japan.

Craven, M., D. DiPasquo, D. Freitag, A. McCallum, T. Mitchell, K. Nigram and S. Slattery. 1998. Learning to Extract Symbolic Knowledge from the World Wide Web. In *Proceedings of the AAAI-98 Conference on Artificial Intelligence*. AAAI/MIT Press.

Crawford, J. (1990), *Access-Limited Logic: A Language for Knowledge Representation*, Doctoral dissertation, Department of Computer Sciences, University of Texas at Austin, Austin, Texas. UT Artificial Intelligence TR AI90-141, October 1990. (Algernon and Access-Limited Logic)
ftp://ftp.cs.utexas.edu/pub/qsim/papers/Crawford-PhD-91.ps.Z
http://www.cs.utexas.edu/users/qr/algernon.html

Cutting, Douglass, David R. Karger, Jan Pedersen, and John W. Tukey. "Scatter/gather: A cluster-based approach to browsing large document collections." In *Proceedings of the 15th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 318-329, Copenhagen, Denmark, 1992.

Eastlake, D., Reagle, J, Solo, D. (Eds.), (2000), *XML-Signature Core Syntax and Processing*, W3C Working Draft
http://www.w3.org/TR/2000/WD-xmldsig-core-20000104/

Engelbart, Douglas C. *Augmenting Human Intellect: A Conceptual Framework*. Stanford

Research Institute Technical Report, Menlo Park, CA, October 1962.

Evett, M.P., W.A. Andersen, and J.A. Hendler. 1993. Providing Computational Effective Knowledge Representation via Massive Parallelism. In *Parallel Processing for Artificial Intelligence*. L. Kanal, V. Kumar, H. Kitano, and C. Suttner, Eds. Amsterdam: Elsevier Science Publishers.
http://www.cs.umd.edu/projects/plus/Parka/parka-kanal.ps

Farquhar, A., Fikes, R., Pratt, W., & Rice, J. (1995). *Collaborative ontology construction for information integration.* Technical Report KSL-95-63, Stanford University Knowledge Systems Laboratory.

Freeman, Eric and Scott Fertig. "Lifestreams: Organizing your Electronic Life" AAAI Fall Symposium: AI Applications in Knowledge Navigation and Retrieval, November 1995, Cambridge, MA.

Finin, T., D. McKay, R. Fritzson, and R. McEntire. 1994. KQML: An Information and Knowledge Exchange Protocol. In *Knowledge Building and Knowledge Sharing*. K. Fuchi and T. Yokoi, Eds. Ohmsha and IOS Press.
http://www.cs.umbc.edu/kqml/papers/kbks.ps

Finin, T., McKay, D., Fritzson, R. (Eds.). (1992) The KQML Advisory Group. *An overview of KQML: A Knowledge Query and Manipulation Language*
http://www-ksl.stanford.edu/knowledge-sharing/papers/index.html#kqml-overview

Finin, T., Weber, J., Wiederhold, G., Genesereth, M., Fritzson, R., McKay, D., McGuire, J., Shapiro, S. and Beck, C. (1992). Specification of the KQML Agent-Communication Language. The DARPA Knowledge Sharing Initiative External Interfaces Working Group.

Genesereth, M., and Fikes, R., Eds. (1992). *Knowledge Interchange Format*, version 3.0 reference manual. Technical Report Logic-92-1, Computer Science Department, Stanford University.
http://www-ksl.stanford.edu/knowledge-sharing/papers/index.html#kif

Gruber, T. R. (1992). *Ontolingua: A mechanism to support portable ontologies*. Stanford University, Knowledge Systems Laboratory, Technical Report KSL-91-66
http://www-ksl.stanford.edu/knowledge-sharing/papers/index.html#ontolingua-long

Gruber, T. (1993a). A translation approach to portable ontology specifications.

Knowledge Acquisition, 5(2):199-220.

Guha, R. V. (1992). Contexts: A formalization and some applications. Ph.D. Thesis, Dept. of Computer Science, Stanford.

Heflin, J., Hendler, J., and Luke, S. SHOE: A Knowledge Representation Language for Internet Applications. Technical Report CS-TR-4078 (UMIACS TR-99-71), Dept. of Computer Science, University of Maryland at College Park. 1999.
http://www.cs.umd.edu/projects/plus/SHOE/aij-shoe.ps

Heflin, J., Hendler, J., and Luke, S. Applying Ontology to the Web: A Case Study. In: J. Mira, J. Sanchez-Andres (Eds.), International Work-Conference on Artificial and Natural Neural Networks, IWANN'99. Proceedings, Volume II. Springer, Berlin, 1999. pp. 715-724.
http://www.cs.umd.edu/projects/plus/SHOE/iwann99.pdf

Hendler, J., Heflin, J., Luke, S., Spector, L., Rager, D. *Simple HTML Ontology Extensions*, University of Maryland
http://www.cs.umd.edu/projects/plus/SHOE/

ISO/IEC 13211-1:1995 Information technology -- Programming languages -- Prolog -- Part 1: General core

Knoblock, C. A. and J. L. Ambite (1997). Agents for Information Gathering. *Software Agents*. J. Bradshaw. Menlo Park, CA, AAAI/MIT Press: 1-27.

Karger, D. R., Klein, P. N., and Tarjan, R. E., "A Randomized Linear-Time Algorithm to Find Minimum Spanning Trees." *Journal of the ACM*, 42(2):321--328, 1995.

Karger, D. R., and Stein, C., "A New Approach to the Minimum Cut Problem." *Journal of the ACM*, 43(4):601--640, July 1996. Preliminary portions appeared in SODA 1992 and STOC 1993.

Karger, D. R., Motwani, R., and Sudan, M., "Approximate Graph Coloring by Semidefinite Programming." *Journal of the ACM*, to appear. Preliminary version in *Proceedings of the 35th Annual Symposium on the Foundations of Computer Science*, pages 2--13. IEEE, IEEE Computer Society Press, November 1994.

Karger, D. R., "Minimum Cuts in Near-Linear Time." In *Proceedings of the 28th ACM Symposium on Theory of Computing*, pages 56--63. ACM, ACM Press, May 1996.

Karger, D. R., Lehman, E., Leighton, T., Levine, M., Lewin, D., and Panigrahy, R., "Relieving Hot Spots on the World Wide Web." In *Proceedings of the 29th ACM Symposium on Theory of Computing*, ACM Press, May 1997.

Koller, D., A. Levy and A. Pfeffer (1997). P-Classic: A tractable probabilistic description logic. Fourteenth National Conference on Artificial Intelligence (AAAI-97), Providence, RI.

Kumar, D., Ed. (1990). *Current Trends in SNePS--Semantic Network Processing System*. Berlin, Springer.

Lassila, O., and Swick, R. [Eds.], (1999), *Resource Description Framework (RDF) Model and Syntax Specification* http://www.w3.org/TR/1999/REC-rdf-syntax-19990222

Levesque, H.J. and Brachman, R.J. (1987). Expressiveness and tractability in knowledge representation and reasoning. *Computational Intelligence* **3** 78-93.

Lieberman, H., L. A. Stein, and D. Ungar, "Of Types and Prototypes: The Treaty of Orlando," *SIGPLAN Notices* **23** *(5):43-44*, May 1988.

Low, Aidan. A Folder-Based Graphical Interface for an Information Retrieval System. Master's Thesis, MIT, May 1999.

Luke, S. and J. Heflin. 1997. SHOE 1.0, Proposed Specification. http://www.cs.umd.edu/projects/plus/SHOE/spec.html

Luke, S., L. Spector, and D. Rager. 1996. Ontology-Based Knowledge Discovery on the World-Wide Web. In Working Notes of the Workshop on Internet-Based Information Systems at the 13th National Conference on Artificial Intelligence (AAAI96). A. Franz and H. Kitano, eds. AAAI. 96-102. http://www.cs.umd.edu/projects/plus/SHOE/aaai-paper.html

MacGregor, R. M. (1991). The Evolving Technology of Classificiation-based Knowledge Representation Systems. *Principles of Semantic Networks*. J. F. Sowa. San Mateo, California, Morgan Kauffman: 385-400.

MacGregor, R. & Bates, R. (1987). The LOOM knowledge representation language. Tech Report ISI-RS-87-188, ISI, CA.

McHugh, Jason, Serge Abiteboul, Roy Goldman, Dallan Quass, and Jennifer Widom.

Lore: A Database Management System for Semistructured Data." SIGMOD Record, 26(3):54-66, September 1997.

Marchiori, M., Ed. (1999). Platform for Privacy Preferences 1.0 (P3P1.0). World-Wide Web Consortium. http://www.w3.org/TR/P3P

Neches, R., Fikes, R., Finin, T., Gruber, T., Patil, R., Senator, T., & Swartout, W. (1991). Enabling technology for knowledge sharing. *AI Magazine*, pages 36-56.

Raggett, D., Le Hors, A., and Jacobs I., Eds. 1999, *HTML 4.01 Specification*, W3C Recommendation
http://www.w3.org/TR/1999/REC-html401-19991224

Roscheisen, M., M. Baldonado, K. Chang, L. Gravano, S. Ketchpel, and A. Paepcke (1998). "The Stanford InfoBus and Its Service Layers: Augmenting the Internet with Higher-Level Information Management Protocols." Digital Libraries in Computer Science: the MeDoc Approach, eds. Barht, A., M. Breu, A. Endres and A. de Kemp. Berlin, Germany: Springer-Verlag, pp. 213-230.

Reiter, R. (1980). A logic for default reasoning. *Artificial Intelligence* **13** 81-132.

Salton, Gerard, James Allan, and Chris Buckley. Automatic Structuring and Retrieval of Large Text Files. Communications of the ACM, 37(2):97-108, February 1994.

Schmolze, J.G. and Woods, W.A. (1992). The KL-ONE family. Lehmann, F., Ed. *Semantic Networks in Artificial Intelligence*. pp.133-177. Oxford, Pergamon Press.

Sowa, J.F., Ed. (1991). *Principles of Semantic Networks: Explorations in the Representation of Knowledge*. San Mateo, California, Morgan-Kaufman.

Spertus, E., and L. A. Stein, "Just-In-Time Databases and the World-Wide Web," *Conference on Information and Knowledge Management* , Washington, DC, November 1998, pp. 30-37.
http://www.mills.edu/ACAD_INFO/MCS/SPERTUS/cikm98.pdf

Spertus, E., and L. A. Stein, "A Relational Database Interface to the World-Wide Web," *ACM Conference on Digital Libraries*, Berkeley, California, August 1999, pp. 248--249.
http://www.acm.org/pubs/citations/proceedings/dl/313238/p248-spertus/

Stein, L. A., "Delegation is Inheritance," *Proceedings of the Conference on Object Oriented Programming Systems, Languages, and Applications*, Orlando, Florida, October 1987, pp.

138-146. Also appears as Brown University Technical Report *CS-87-15*, July 1987.

Stein, L. A., "Compound Type Expressions: Flexible Types in Object Oriented Programming," Panel Position Paper, *Proceedings of the Conference on Object Oriented Programming Systems, Languages, and Applications*, San Diego, California, September 1988, pp. 360-361.

Stein, L. A., *Resolving Ambiguity in Nonmonotonic Reasoning*. Ph.D. Thesis, Brown University Department of Computer Science, 1990.

Stein, L. A., "Resolving Ambiguity in Nonmonotonic Inheritance Hierarchies," *Artificial Intelligence* **55** *(2-3):*259-310, June 1992. Earlier version appears as MIT AI Lab Memo *1316*, August 1991.

Stein, L. A., "Science and Engineering in Knowledge Representation and Reasoning," *AI Magazine* **17** (4):77-83, Winter 1996.
ftp://ftp.ai.mit.edu/pub/users/las/science-and-engineering-in-KR.ps.gz

Stein, L. A., H. Lieberman, and D. Ungar, "A Shared View of Sharing: The Treaty of Orlando," in *Object-Oriented Concepts, Databases, and Applications*, W. Kim and F. Lochovsky, eds., A.C.M. Press, 1989, pp. 31-48.

Stein, L. A. and L. Morgenstern,, "Motivated Action Theory: A Formal Theory of Causal Reasoning," *Artificial Intelligence* **71 (1)**:1-42, November 1994. Earlier versions appear as MIT AI Lab Memo *1338*, December 1991 (postscript, 3.6M); and as Brown University Technical Report *CS-89-12*, March 1989.
ftp://ftp.ai.mit.edu/pub/users/las/motivated-action-theory.ps.Z

Stein, L. A. and S. B. Zdonik, "Clovers: The Dynamic Behavior of Types and Instances," *International Journal of Computer Science and Information Management* **1** (3): 1-11, 1998.
http://www.ai.mit.edu/people/las/papers/Clovers/Clovers_Final.pdf

Stoffel, K., M. Taylor and J. Hendler. 1997. Efficient Management of Very Large Ontologies. In Proceedings of American Association for Artificial Intelligence Conference (AAAI-97). AAAI/MIT Press.

Swartout, Bill, Ramesh Patil, Kevin Knight and Tom Russ. "Toward Distributed Use of Large-Scale Ontologies". In *Proceedings of the Tenth Knowledge Acquisition for Knowledge-Based Systems Workshop* , November 9-14, 1996. Banff, Alberta, Canada.

http://ksi.cpsc.ucalgary.ca/KAW/KAW96/swartout/Banff_96_final_2.html

Vilain, M. (1991). Deduction as parsing: tractable classification in the KL-ONE framework.

Wobber, E., Abadi, M., Burrows, M., Lampson, B. (1993) *Authentication in the Taos Operating System*, Technical Report, DEC Systems Research Center, Number 117, http://gatekeeper.dec.com/pub/DEC/SRC/research-reports/abstracts/src-rr-117.html

Woods, W.A. (1975). What's in a link: Foundations for semantic networks. Bobrow, D.G. and Collins, A.M., Ed. *Representation and Understanding: Studies in Cognitive Science.* pp.35-82. New York, Academic Press.

Yen, John, Robert Neches, and Robert MacGregor. "CLASP: Integrating Term Subsumption Systems and Production Systems." In *IEEE Transactions on Knowledge and Data Engineering* , Vol. 3, No. 1, pp. 25-32, 1991.