

DAML Intent of Work

UMBC, JHU APL and MIT Sloan

Tim Finin, James Mayfield and Benjamin Grosf

April 2001

Abstract

Our work during the next year will cover three areas: the integration of DAML into current multiagent system technologies, the exploration of hybrid web searches involving both DAML and text, and extending DAML to include support for rules. Much of this work will be organized around and tested using ITTalks (<http://ittalks.org/>) as a portal of information about information technology (IT) research talks. This portal will provide our team many opportunities to explore how DAML can be used to organize the information and drive associated services. In addition to basic portal services, we will also develop several agent-based services that use this information to notify users of talks which (1) match their interests expressed in one of several topic ontologies, (2) are appropriate to their location and (3) are compatible with their schedule. ITTalks will also provide the DAML and semantic web communities with a large volume of DAML enhanced web pages for analysis and experimentation.

This document describes our *intent to work* for the next year (March 2001 to February 2002). Our team includes researchers at UMBC, JHU APL and MIT Sloan. We have also planned collaborations with the DAML research groups at LMCO, USC/ISI, Teknowledge, Cycorp, Booz-Allen and Hamilton and BBN.

ITTalks is a portal of information about talks, seminars and colloquia related to Information Technology¹ (IT). We have developed and will continue to refine and extend DAML encoded ontologies for three main areas: seminar event information, people and places. Information on these is stored in a relational database (MySQL currently) and used to dynamically generate descriptions in DAML, HTML and XML using Java servlets and JSP files.

ITTalks is currently configured as a centralized service but this is mainly for ease of development and maintenance. Future versions can and should be distributed. Even in its centralized implementation there are many distributed components: each user's profile is kept and maintained by the user in his own file space, independent agents offer services with information obtained dynamically from the ITTalks portal via its programmatic interfaces, and users can have personal agents which mediate some of the communication with ITTalks. DAML is heavily used as a language for interoperability between the portal and the programs, services and agents which access and use its information.

Our work over the next year falls into two areas. First we will continue to develop the portal as a practical and useful application for DAML. Our goal here is to become for research oriented talks about IT what NEC's CiteSeer [2] is for computer science research papers – a comprehensive repository of information which is automatically generated and maintained. We will work to highlight the value that DAML adds to this portal and make it a resource for other researchers in the DAML program as well as the semantic web community at large. Second, we will use this portal to support our own advanced research topics, including the integration of agents and DAML (UMBC), extending information retrieval systems and techniques to

¹Limiting this to talks on IT is, of course, arbitrary and most of the structure we are building will generally support such talks in any area. We may expand our coverage to “academic” talks of all kinds in the future, but believe that it will be helpful to focus on a narrower domain for the scope of this project.

work in a DAML-enriched web (JHU), and extending DAML to include rules and to connect with practical rule-based languages and systems (MIT).

Making ITTalks practical

We are already in the midst of a major revision to the current ITTalks system (v. 2.4) to support our vision of it as a central repository of information about IT related talks for the international IT research community. This will result in a new version (v. 2.6) to be available in mid Q2 2001 which will support the ability to have talks from multiple domains roughly corresponding to different host institutions (e.g., UMBC, UMCP, MIT, Stanford, IBM, etc.). Each domain will have a moderator who can define the scope of the domain and delegate the ability to enter and edit entries to additional registered users. For example, the *stanford.ittalks.org* domain might be moderated by Stefan Decker who could configure it to include just talks hosted at Stanford University. The *sri.ittalks.org* domain (hosted by, say, Jerry Hobbs) might be configured to include talks held at SRI but also those at Stanford as well as any within 15 km. of the SRI facility. We will also support a user-defined customization, *my.ittalks.org*, in which a registered user can define constraints on the talks in which he is interested.

We have a number of additional design features which are under way, many of which depend on DAML to implement their added functionality. These include allowing users to browse the database based on one or more DAML-encoded ontologies, the ability to generate RDF Site Summary (RSS) files corresponding to standing queries, an expanded ability to capture and encode user-modeling information in DAML, and better support for notifying users of talk which match their interests, schedule and location.

Topic ontologies

We have develop an general ontology for describing the topics of arbitrary talks and papers. Using this, we implemented an ontology to describe IT related talks based on ACM's Computer Classification System [7]. We are planning to develop at least one additional DAML ontology for IT talks based on a portion of the Open Directory [16]. These same topic ontologies are used to describe a user's interests. These ontologies will be used in a number of ways:

- We will develop a system which can be used to classify a talk (based on its title and abstract) or a user's interests (based on her web pages and papers) with respect to a target ontology. We have obtained a training collection for the ACM CCS and can easily generate one from the Open Directory. These will be used to train a text classification system. The results of such classification will be used as an initial "guess" which can be further refined or modified by the user.
- We will refine and extend our simple web-based HCI so that users can define, edit and modify their interests in terms of a DAML encoding of a topic ontology. Users will also have the ability to add additional assertions in DAML to further characterize their interests.
- We will develop components which can map topics in one topic ontology into those in another by taking advantage of the fact that nodes in each have associated collections of text as well as DAML encoded information
- We will allow users to select any of the DAML topic ontologies and then use it to browse the underlying database of talks.

Scaling up ITTalks

Currently information about talks has to be manually entered into ITTalks via a web form interface. We have made this as simple as possible by supporting automatic form completion using information from the

database and user's DAML profile. However, this remains a time consuming process and presents a barrier to scaling up ITTalks to include a significant number of talks. Moreover, some organizations already have their talk announcements in a database and generate the web versions using some form of active server pages (e.g., PHP, Coldfusion, JSP, ASP, etc) and it should be possible to acquire information about these talks without manual entry.

We plan to build a a "focused" search or web spider to collect talk announcements from open sources on the web on an ongoing and continuous manner. Some of these will contain markup in appropriate ontologies that will be sufficient for us to enter them into the ITTalks database. Others will contain only partial markup or even no semantic markup at all. For these we will use the Lockheed Martin Aerotext text extraction system to identify the key information in these talk announcements and to add this information to the ITTalks database for subsequent use in generating DAML descriptions of the talks. One of the research issues is exploring how this DAML markup can be used to (i) guide the focused search and (ii) enhance the text extraction.

DAML and user modeling

ITTalks maintains a *user model* for each of its registered users. Currently this is partly kept in the relational database and partly in a *DAML profile* which is stored in the user's web-accessible public file space. Our notion here is that the DAML component represents a set of fact that the user is willing to share about herself with the ITTalks system as well as with other systems and agents. This information should be encoded in a general fashion, making it useful for more than just ITTalks. We plan to expand on this approach, drawing on past experience in developing general user modeling components [12]. One of the issues we will address is how a user can maintain some control over who has access to what information.

DAML and security

In earlier work [14, 15, 10] we have developed a model in which to represent and reason about the delegation of belief and authorization in an open and distributed environment. In this model, agents make (possibly conditional) statements governing such delegations and digitally sign them (e.g., "*I hereby authorize John to be able to edit this record*"). Agents can also define, via rules, security policy (e.g., "*Users can edit this record if they are UMBC graduate students or UMBC faculty and they are members of the ebiquity research group*"). These statements can then be shared with other agents (e.g., via an INFORM speech act) or published in an open manner (e.g. via a web page). Agents which control resources can use these statements to prove that an agent holds a certain belief (in the case of distributed belief) or has a certain rights, privileges or obligations (in the case of delegation of authority). Initial work was done in the context of a agent communication language (KQML) using X509 certificates. We are extending this model to work in the web, basing it on DAML and XML signature [17]. We will use this to augment the security model of ITTalks portal and also explore its use in pervasive computing environments [11, 3, 4].

DAML and multiagent systems

We will explore the use of DAML in multiagent systems by developing a DAML ontology based on the abstract model of the FIPA agent communication language [8] and using DAML as a content language. We will adapt one or more software packages such as Jackal [6, 5] or JADE [1] to accept and generate such DAML encoded messages. We will also develop a practical DAML reasoning engine in XSB and incorporate this into this framework. This will be used to develop a tool that can be used to create an internet agent that can answer queries about a set of DAML pages.

DAML and hybrid search

One of the main goals of our team is to provide hybrid search over text and DAML. Unfortunately, there is not yet enough DAML-enabled text to pursue our hypothesis that DAML will confer significant advantages in traditional ad hoc search. Whether sufficient page volume will become available within the next year or two will depend (among other factors) on whether DAML develops a reputation as a markup language, and not just a knowledge representation language. In light of this constraint, we will restrict our current focus to two areas for which we can be assured that appropriate DAML will be available: ontology mapping, and domain-restricted search.

Ontology Mapping

In its simplest form, ontology mapping is the alignment of nodes in one ontology with those of another ontology, so as to capture shared meaning between the ontologies. Effective ontology mapping is (in our opinion) critical to the success of DAML. We believe that hybrid search can play an important role in ontology mapping by using the text associated with nodes in the ontologies to suggest possible pairings. This technique might be used to constrain or inform a separate process that uses other methods (such as structure mapping) for ontology mapping, or, in a pinch, to serve as the entirety of the mapping process. We therefore intend to build the following DAML services which will be available to both people and programs:

- a service that, given an ontology node as input, returns the most closely related nodes in other ontologies;
- a service that, given an ontology node and a target ontology as input, ranks the nodes in the target ontology for similarity with the input node;
- a service that returns a set of words that describe a given ontology node; and
- a service that returns those ontology nodes that are best described by a given input text (which may itself contain DAML statements).

Three teams have already expressed interest in using such services: Teknowledge (Pease), USC, and Cycorp. We will work with these teams both to tailor the query and results ontologies to their needs, and to increase the usefulness of the matching results to their systems. In addition, the team from Booz-Allen and Hamilton has expressed interest in fielding services that perform similarly to the first two listed above, but that exploit pathways through DAML ontologies directly. We will coordinate with them to ensure that our services exhibit a common interface.

We see two pathways for these services. Programatically, systems are likely to need them to constrain the mappings they consider, and possibly to provide evidence for those mappings they do consider. For people trying to develop their own DAML ontologies and data, these services can also be useful by suggesting ontologies that might serve as templates (or perhaps obviate the need for a new ontology). To support these two lifecycle pathways, we will provide two interfaces to the services. Both will be Web-based, one using DAML, the other, HTML. Both will run initially as servlets on daml.org.

Domain-Restricted Search

To explore broader issues of the influence of DAML on ad hoc search, we will contribute to the UMBC ITTalks project. This domain features talk announcements, some of which may be DAML-enhanced, as well as expressions of user interest, which again may be some combination of text and DAML. Furthermore, the project will use DAMLized versions of two existing hierarchies: the ACM topics

hierarchy, and the Open Directory hierarchy. Alignment of the two resulting ontologies will exercise our work in ontology mapping. Furthermore, the needs of the system to match people to talks will provide a good incubator for our hybrid retrieval work. We will demonstrate matching of talk abstracts to the two ontologies, as well as directly to user profiles. We anticipate that this work will result in anecdotal evidence that a search (such as from a user profile to talk descriptions) that combines text and DAML will outperform either component individually.

This approach should bring some of the benefits of DAML to DAML-ignorant users of the ITTalks site. Furthermore, it has the potential to convey added benefit to users who embed DAML either in their user profiles or in their talk announcements.

DAML and rules

We will continue our work [9] investigating how DAML can be extended to include rules and how these can be mapped to any of several families of practical rule-based systems. This will be done in the context of the RuleML rule markup effort.

Collaborations and potential military users

Our effort represents joint work of researchers at UMBC, JHU APL and MIT Sloan School. We have begun or anticipate collaborations with a number of other groups within the DAML program. Teknowledge (Pease), USC, and Cycorp have all expressed an interest in using one or more of our ontology mapping services. Booz-Allen and Hamilton will be developing some similar services and we expect to collaborate in order to develop consistent and compatible interfaces to these services. We are experimenting with LMCO's Aerotext text extraction system as a component which can ease the entering of new talks into the ITTalks portal. We will investigate using USC/ISI's web scraping technology [13] as another technique that can be used to facilitate adding information to the portal. Both BBN and USC/ISI have been identified as potential users of the DAML encoded web pages we generate.

Although ITTalks does not have direct military applications, it is easy to imagine similar systems being used to coordinate the scheduling and announcements of meetings, presentations and training seminars in a military setting. As we develop ITTalks and its associated agents-based applications we will anticipate such possible future generalizations in our design.

References

- [1] Fabio Bellifemine, Agostino Poggi, and Giovanni Rimassa. Developing multi agent systems with a fipa-compliant agent framework. *Software - Practice And Experience*, (3), 2001.
- [2] Kurt D. Bollacker, Steve Lawrence, and C. Lee Giles. Citeseer: An autonomous web agent for automatic retrieval and identification of interesting publications. In *Proceedings of the Second International Conference on Autonomous Agents (Agents '98)*, Minneapolis, 1998. ACM Press.
- [3] Harry Lik Chen, D. Chakraborty, L. Xu, Anupam Joshi, and Tim Finin. Service discovery in the future electronic market. In tim finin and Benjamin Grosz, editors, *Proceedings of the AAI Workshop on Knowledge Based Electronic Markets*. AAAI Press, August 2000.
- [4] Harry Lik Chen, Anupam Joshi, Tim Finin, and D. Chakraborty. Dynamic service discovery for mobile computing: Intelligent agents meet jini in the ether. *Baltzer Science Journal on Cluster Computing*, 2001.
- [5] R. Scott Cost, Tim Finin, Yannis Labrou, Xiaocheng Luan, Yun Peng, Ian Soboroff, James Mayfield, and Akram Boughannam. Jackal: A java-based tool for agent development. In Jeremy Baxter and Chairs Brian Logan, editors, *Working Notes of the Workshop on Tools for Developing Agents, AAAI '98*, number WS-98-10 in AAAI Technical Reports, pages 73–82, Minneapolis, Minnesota, July 1998. AAAI, AAAI Press.

- [6] R. Scott Cost, Tim Finin, Yannis Labrou, Xiaocheng Luan, Ian Soboroff Yun Peng, James Mayfield, and Akram Boughannam. An agent-based infrastructure for enterprise integration. In *First International Symposium on Agent Systems and Applications*, October 1999.
- [7] Neal Coulter et. al. Computing classification system 1998: Current status and future maintenance, report of the ccs update committee. *COMputing Reviews*, January 1998.
- [8] FIPA. FIPA 97 specification part 2: Agent communication language. Technical report, FIPA - Foundation for Intelligent Physical Agents, october 1997.
- [9] Benjamin Grosf, Yannis Labrou, and Hoi Chan. A declarative approach to business rules in contracts: Courteous logic programs in xml. In *First ACM Conference on Electronic Commerce*, November 1999.
- [10] Lalana Kagal, Harry Chen, Scott Cost, Timothy Finin, and Yun Peng. An infrastructure for distributed trust management. submitted to Autonomous Agents Workshop on Norms and Institutions in Multiagent Systems, March 2001.
- [11] Lalana Kagal, Vlad Korolev, Harry Chen, Anupam Joshi, and Timothy Finin. Centaurus: A framework for intelligent services in a mobile environment. In *Proceedings of the International Workshop on Smart Appliances and Wearable Computing (IWSAWC)*, April 2001.
- [12] Robert Kass and Tim Finin. General user modelling: A facility to support intelligent interaction. In J. Sullivan and S. Tyler, editors, *Architectures for Intelligent Interfaces: Elements and Prototypes*. Addison-Wesley, 1990.
- [13] Craig A. Knoblock, Kristina Lerman, Steven Minton, and Ion Muslea. Accurately and reliably extracting data from the web: A machine learning approach. *Data Engineering Bulletin*.
- [14] N. Li, J. Feigenbaum, and B. Grosf. A logic-based knowledge representation for authorization with delegation. In *IEEE Computer Security Foundations Workshop. Extended version is IBM Research Report RC 21492*, 1999.
- [15] N. Li and B. Grosf. A practically implementable and tractable delegation logic. In *IEEE Symposium on Security and Privacy*, May 2000.
- [16] The open directory project.
- [17] W3c. Xml-signature syntax and processing.