Intelligent Information Dissemination

Harry Behrens

Dept. of Information Engineering
University of Tokyo

Hidehiko Tanaka

Dept. of Information Engineering
University of Tokyo

Abstract

Using autonomous intelligent agents that converse based on mutually accepted ontologies distributed entities can share and disseminate knowledge across system and organization boundaries. By using ontologies that are expressed in first order predicate-logic participating agents can query and declare knowledge in a very high-level fashion that goes beyond mere data sharing. This high-level conceptual layer of software sits on top of a sophisticated network programming protocol that is based on network streams. This paper will a prototype system currently being developed at our laboratory.

1 Introduction

The last years have brought dramatic developments to the global information infrastructure. The introduction of the HTTP protocol [?] and the HTML mark-up language [?] have brought forth explosive growth in the utilization of networked resources.

This has triggered massive development of commercial interest which in turn has brought about large-scale investments that are bringing higher and higher bandwidth to the individual's PC.

The basic protocol infrastructure on which all of the WWW is based has been remarkably constant however: On the transport and network layer the vast majority of all distributed systems are based on the TCP/IP protocol stack. Even though it is of course difficult to make predictions in such a fast moving sphere and there are efforts to change some aspects of the IP protocol [?], it doesn't look as if the basic transport layer architecture of the Internet will change.

All of this permits us, to think about applications and networked environments that would have beyond all practicality only a few years ago.

This research project deals with two major problems that have arisen due to the explosive growth of information resources:

1. **the** *Information Retrieval* **(IR) problem:** As almost all of us, there is information on almost anything imaginable and unimaginable available on the Net. The problem is to find specific information with relevance to clearly defined points of interest.

2. **the** *Information Placement* **problem:** This problem is essentially complementary to the IR problem. It is the situation an information source faces in sending its information to as many interested and relevant consumers (*information sinks* as possible. This is becoming especially urgent with dynamic publishing, the merging of broadcasting and Internet technology and the emergence of data channels [?], [?] that offer the means to publish on-line on the Internet.

The specific aim at this stage of the research project is to design an infrastructure that allows intelligent information multicasting.

This paper proceeds to describe this infrastructure in a top-down fashion.

2 Information Routing Agents

The whole system is based on set of distributed **sites** that constitute the market places for the whole system. A site is a machine or a network that is accessible over the network and uses the same basic protocols as the other sites. [1] Each site knows about its capabilities and its needs in terms of globally specified *ontologies*.

Ontologies are definitions of concepts that can be thought of as computer-processable encyclopedia entries. [?] In the context of this project they are defined in predicate logic using the PROLOG programming language.

---

[1] The underlying structure is described in [?]

Using a logic programming language and its inference engine has the benefit, that verification such as consistency checks can be done automatically.

Ontologies are used within a **context** which defines a basic semantic category within which the ontological entry should be interpreted.

Within a given context all participants use the same ontologies. Whenever entries are changed this is immediately propagated to all participants.

2.1 The Publisher's view

One idea behind this whole project is the fact, that more and more information providers will spend considerable efforts in pushing their information to interested clients.

Based on this assumption, we think that it is realistic to assume that information providers will spend some effort to annotate their data if seems probable that the annotation information helps in delivering the content to the client. [2] In the system we are building, information published is annotated with meta-information placing it within an ontological context. An example could be the the publishing of a memo on a corporate intranet, pertaining to an ongoing project and with reference to previous documents. In such a case, the client will be publishing the document within the context of his corporate groupware. By adding the ontological information at the administrator level, when new projects or new processes are created, the quality of the information can be kept relatively high with relatively little manpower effort.

The ontological information thus added, places the published memo within a semantic context. Within this semantic context classic indexing technology can be used to describe the document in terms of its raw textual content within the given semantic context.

3 The Consumer's view

When this article talks about consumers this does not necessarily mean the end client with his desktop interface. In a complex information dissemination network such as the one examined and implemented in this project, data and information travels along potentially many-stepped filters until it gets to its final destinations.

---

[2] This can already be witnessed on the WWW, where considerable effort is being spent to make Web pages accessible through the use of search engines. At this point these efforts are mainly centered on <META> tags and nifty use of comments that are seeded with keywords.

Sites such as departments or whole corporations will specify certain contexts within which they might be interested in receiving information. This can refer to an on-line Web-based news channel, an intra-corporate discussion forum or large corporations soliciting bids for component products along certain specifications.

The software mechanism used is analog to that of the publisher: ontologies are used to define semantic contexts. Within these semantic contexts, more syntax oriented queries, that correspond to the index information added by the producers, are added. These index queries can be SQL queries, SOIF queries such as those used by the Harvest system [?] and the Netscape Catalog Server or Z39.50 queries [?] such as those used by many libraries and the WAIS system or just plain textual searches.

4 Summary and Conclusion

The amount of raw information available on the Internet has been growing dramatically during the last few years and is bound to continue doing so. At the same time connectivity is becoming better due to major investments by governments and corporations.

One of the main problems that the Internet poses is how to route published information to the correct clients and how to retrieve desirable information. Using an approach in which both sides use logical meta-information that can be handled by autonomous network agents, we are currenly designing a prototype system, that should demonstrate how efficient use of distributed processing in combination with high-level semantic content description can help solve some of the problems the Internet is currently facing.

References

[1] T. Berners-Lee, R. Fielding, and H. Frystyk. Hypertext Transfer Protocol – HTTP/1.0. *IETF Internet Draft <draft-ietf-http-v10-spec-04.html>*, October 1995.

[2] T. Berners-Lee and D. Connoly. Hypertext markup language - 2.0. *IETF RFC 1866*, November 1995.

[3] Deering S. and Hinden H. Internet protocol, version 6 (ipv6) - specification. *IETF RFC 1883*, December 1995.

[4] Microsoft Inc. Channel Definition Format (CDF). *http://www.microsoft.com/standards/cdf.htm*, 1997.

[5] Marimba Inc. Castanet (tm) Developer Documentation. *http://www.marimba.com/doc/Developer_Documentation/Castanet_Developer_Docs/index.html*, 1997.

[6] Behrens H. and Tanaka H. Information Dissemination using Network Streams. *to be presented at the DiCoMo workshop of the Information Processing Society of Japan (*情報処理学会*) July 2nd, 1997*, 1997.

[7] Gruber T. Toward principles for the design of ontologies used for knowledge sharing. Technical Report KSL 93-4, Stanford University, Knowledge System Laboratory, 1992.

[8] Bowman C.M, Danzig P., Hardy D., Manber U., Schwartz M., and Wessels D. Harvest: A scalable, customizable discovery and access system. Technical Report CU-CS-732-94, University of Colorado - Boulder, 1995.

[9] American National Standards Institute (ANSI). Application service definition and protocol specification for open systems interconnection. *ANSI/NISO Z39.50-1992, ISSN: 1041-5653*, July 1992.