Towards an XML and agent-based framework for the distributed management and analysis of active data archives

Yanyan Yang, Omer F. Rana, and David W. Walker
Department of Computer Science,
Cardiff University, PO Box 916,
Cardiff CF24 3XF, UK
E-mail: {L.Yang, O.F.Rana, David.W.Walker}@ cs.cf.ac.uk

Abstract

Implementation and use of a prototype distributed active data archive system are outlined. This system is based on the Synthetic Aperture Radar Atlas (SARA) and utilises cooperative software agents for data access and analysis, and uses XML to model metadata and support agent communication.

Keywords: *Digital Libraries, XML, Mobile Agents, Parallel and Distributed Computing, Remote Sensing*

1. Introduction

The general problem of managing large digital data archives is particularly challenging when the system must cope with active data which is processed on demand [P.D. Coddington,1999]. Active data represents data that is dynamically generated by a scientific experiment, or may be obtained from a sensor or monitoring instrument. SARA is an active digital library of multi-spectral remote sensing images of the earth, and provides Web-based on-line access to such images. SARA has been online for over a year at the University of Lecce in Italy, at Caltech in California, and at SDSC in San Diego, California [R. D. Williams]. The SARA service is actually maintained by Caltech and the University of Lecce[G. Aloisio]. Scientists making use of the SARA archive often require integrated access to information combining retrieval, computation, and visualization of multiple images. Such an analysis can involve overlapping images. to detect change within a given area, to colour coding images based on electro-magnetic polarisation of radar frequencies. Generally, radar frequences vary over three bands, and this is used to identify the backscatter from land features over which the image is taken. A large scientific collaboration may generate many queries and the resulting analysis of images can lead to large quantities of data, some of which must be integrated with data from other systems such as a Geographic Information System for a given area/region. The data is maintained in different kinds of file systems such as Sun NFS and IBM/Livermore HPSS. These requirements have generated an urgent need for a more advanced software infrastructure to create, maintain, evolve, and federate these active digital libraries of scientific data. It will also be necessary for human interfaces to these archives to become simpler and more flexible, and to support both professional and casual users. As part of the on going SARA digital library project, this paper describes a collaborative effort to explore an XML and agents based framework for the distributed management and analysis of remote sensing archive.

2. The multi-agent system

A multi-agent system, which comprises both intelligent and mobile agent, has been developed to manage and analyze distributed multi-agency remote sensing data. Compared with other agent applications, our emphasis is on the use of the mobile agents to transfer analysis algorithms to image processing archives. We localize the most complex functionality in non-mobile Local Interface Agents (LIA), which remain at one location, communicate with the mobile User Interface Agents (UIA), and provide resources and facilities to lightweight mobile agents that require less processor time to be serialized, and are quicker to transmit. Each agent is responsible for offering a particular type of service, and the integration of services is based on a user specification. SARA mobile agents are persistent, and can wait for resources to become available. Agents allow the delivery and retrieval of data to complete without user monitoring or recovery actions.

There are two types of User Interface Agents: User Request Agents (URA) and User Assistant Agents (UAA). URA supports the user in creating a query or operation to perform on the SARA data. UAA manages the information of the user and provide control functions to the user, such as updating their file space on a remote server, and parameter settings for their visualization tool. There are many types of Local Interface Agents: a Local Assistant Agent (LAA) supports interaction with any visiting User Request Agents (URAs) by informing them about the available data and computing resources, and cooperating on the completion of the task carried by the URA. A Local Management Agent (LMA) coordinates access to other LAAs and supports negotiation among agents. It is responsible for optimizing itineraries of mobile URAs, to minimize the bottlenecks inherent in parallel processing and ensuring that the URA is transferred successfully. A Local InteGration Agent (LIGA) provides a gateway to a local workstation cluster, or a parallel machine. A Local Retrieval Agent (LRA) can translate query tasks and performs the actual information retrieval from the local archive. In addition to retrieval, a LRA may also perform other operations. For instance, it may save the results to a file before sending it to the user. A Local Security Agent (LSA) is responsible for authenticating and performing a validation check on the incoming URA. The URA will be allocated an access permission level. Agents from registered users may use, and have access to, more information resources than agents from unregistered users.

3. XML based data specification and agent communication

The eXtensible Markup Language (XML) is becoming the standard for data interchange on the internet, and enables a new generation of web services that are not meant for humans to use directly, but rather to be used by other software. Every XML document refers to a Data Type Definition (DTD), which is a grammar that defines the document syntax. In our system we use XML to encode system structure as metadata and user requests. When a user launches a query, an XML document is created. Every specific XML specification is based on a separate DTD that defines the names of tags , their structure and content model. While the XML specification contains the structured information, the DTD defines the semantics of that structure, effectively defining the meaning of the XML document. An agent can generate the processing programs representing the XML elements of interest according to the DTD with a parser, and then travel

across the internet to retrieve related information. For example, corresponding to tag TRACK in the XML document, a Java method getTrack can be generated to interpret the meaning of the tag. The resulting data can be returned to the user with an XSL style sheet.

We also use XML as an application-specific transport protocol, to enable agents within the system to communicate with each other. Autonomous agents cooperate by sending messages and using concepts from the SARA ontology – where the ontology describes terms and concepts (such as a Track, a Lattitude/Longitude coordinate etc) and their inter-relationships. We are defining a message which embeds such an ontology, to provide a mechanism for agents to exchange requests and message interpreters. Agents send and receive information through XML encoded messages. Based on pre-defined tags, agents may use different style DTDs to fit different mediation. Moreover, a mobile agent can carry an XML front-end to a remote data source for data exchange, where both queries and answers are XML encoded.

In our prototype, we use the JAXP (Java API for XML Parsing)[JAXP] interface to XML developed by SUN that supports SAX and the general purpose Document Object Model (DOM). We currently have a working prototype. Our future work is to extend the ontologies definition and investigate system scalability and response times when thousands of agents are hosted on a single data source.

Acknowledgements

This work has been partly supported by the Center for Advanced Computing Research (CACR) at Caltech and by the Cardiff University research fellowship scheme.

References

[1] P.D. Coddington, K.A. Hawick & H.A. James, Web-based access to distributed high-performance geographic information systems for decision support, *In Proc. HICSS, Vol.* 32, Maui, January 1999.

[2] R. D. Williams, G. Aloisio, M. Cafaro, G. Kremenek & P. Messina, *SARA: The Synthetic Aperture Radar Atlas*, http://www.cacr.caltech.edu/sara/ and http://sara.unile.it/sara

[3] G. Aloisio, G. Milillo & R. Williams, An XML Architecture for High Performance Web-Based Analysis of Remote Sensing Archives, *FGCS Int. Journal*, Vol. 16, No.1, 1-156, North Holland, 1999.

[4] JAXP, see web site at: http://java.sun.com/products/xml/index.html