

GeoWorlds: A Geographically-based Information System for Situation Understanding and Management

Murilo Coutinho, Robert Neches, Alejandro Bugacov, Ke-Thia Yao, Vished Kumar, In-Young Ko, Ragy Eleish.

University of Southern California, Information Sciences Institute,
4676 Admiralty Way, suite 1001, Marina del Rey, CA 90292

{coutinho,bugacov,rneches,kyao,vkumar,iko,ragy}@isi.edu

Abstract

GeoWorlds (<http://www.isi.edu/geoworlds>) is a component-based information management system aimed at helping organizations to marshal, analyze, discuss, and act on *all* of the available information about a situation playing out over space and time. The system integrates digital library, geographic information systems (GIS), and remote sensor data management technologies, together with other information analysis, retrieval and collaboration tools. It shows how users can rapidly *assemble a custom repository* of geographic information about a region, *bi-directionally link* it to collections of document-based information from the World-Wide Web, and *monitor* real-time sensor data for information that might change conclusions or decisions formed on the basis of this rich information set. This functionality is provided within a framework that enables synchronous and asynchronous collaboration over finding, filtering, organizing and visualizing the needed information.

1 Introduction

GeoWorlds is an experimental system that demonstrates how carefully integrating three key technologies can provide teams of users with a sense of *shared regional vision* -- the ability to marshal and organize everything known about an area, displayed with respect to space and time. The system seeks to provide synergy between three technologies -- digital libraries, geographic information systems, and telecommunications of remote sensor data. It retrieves, organizes and displays available information about a region in rich displays, allowing teams of users in distributed locations to collaboratively assess situations, develop appropriate responses, and monitor the situation's evolution. The system integrates in-house tools, CoTS components and various products of other research institutes that are collaborating with the University of Southern California Information Sciences Institute (USC ISI) Distributed Collaborative Enterprises Group¹.

¹ USC Information Sciences Institute is developing GeoWorlds on behalf of the US Defense Advanced Research Projects Agency, with collaborative support from: the USC ISI DASHER

More specifically, GeoWorlds provides these teams the ability to rapidly assemble a *custom* repository of information about a geographic area. It enables them to select data sets from large samples of pre-determined information stored in GIS databases, to relate the GIS data to collections of document-based information from the World-Wide Web that have been found, filtered and organized on-the-fly, and to tie these to physical events monitored by real-time sensor feeds.

The geographic and document-based information are *bi-directionally* correlated, i.e., by selecting a geographic region on a GIS display GeoWorlds can retrieve, filter and organize sets of documents from the Web that are associated with that region, and conversely, we can visualize in the GIS display the locations that selected documents are geo-referencing. The system has a notion of a *service registry* which helps users find, select among, and initiate automated analysis tools. These tools can draw upon different sources of information (e.g., geographic and document-based information, numerical simulation of

Project, University of Southern California Department of Geography, UC Santa Barbara Alexandria Digital Library Project, University of Illinois at Urbana-Champaign Digital Library Initiative, University of Arizona Artificial Intelligence Laboratory, University of California at Berkeley Digital Library Project, and University of Illinois National Center for Supercomputing Applications.

physical processes). GeoWorlds also provides an initial level of support for establishing monitoring of real-time sensor feeds in order to compare them against predictions of the analysis packages and generate warnings when actual events are not playing out as predicted.

This functionality is provided in a framework that enables synchronous and asynchronous collaboration over finding, filtering, and organizing needed information and actions. The collaboration framework supports viewing and annotating the information to facilitate group decision making both to form and record conclusions about the interpretation of shared information, and to select and coordinate actions based upon those conclusions. The growing capability to establish monitors against the real-time data allows teams to rapidly assess situations, develop and execute responses, and -- when needed -- revisit their conclusions and decisions if the assumptions upon which they were based turn out to be invalid.

GeoWorlds is an investigation into multi-domain software frameworks for situation understanding and information management systems. Its current demonstration applications are in disaster response and consequence management, but the techniques are equally applicable to topics as disparate as search and rescue, weather tracking, urban planning, and natural resource management.

In the next section we give an overview of the software architecture of the GeoWorlds system, list the components that have been integrated to the system and describe the implementation methodology. In Section 3 we illustrate some capabilities of the system through an example: conducting emergency relief operations in response to a hazardous materials spill. In Section 4 we summarize the status of the current phase of work on the project, discuss its limitations and present our vision of the future for GeoWorlds.

2 Component-based Architecture Overview

GeoWorlds is targeted to a community with a broad and demanding range of functional requirements: the situation understanding and information management systems community. A major challenge in designing and building such a system is not only to develop basic system capabilities but to provide a framework where the best available tools can be integrated into the system with minimal effort and that each of these components can communicate with each other to create new applications. With this in mind, we designed GeoWorlds as a component-based system that will be able to support continuous increase of functionality and portability as new and more sophisticated tools become available.

We believe that there is a broad class of applications targeted at providing *total regional vision for situation understanding and management* and that these applications can be grouped based on the set of functional requirements that they have in common. HA/DR (Humanitarian Assistance and Disaster Relief) operations belong to this set of applications and we consider the current version of GeoWorlds to be a reference implementation prototype supporting this kind of applications.

The basic set of requirements that we have identified and considered during the design of the system is:

- dynamic storage/retrieval, processing and analysis of data and information available about the region/situation.
- simulation of physical or logistic problems involved in the situation.
- visualization and analysis of the information and results.
- marking, annotating and collaboration of the information between spatially distributed teams.

At the same time, we believe that information systems supporting these applications can be built by integrating components belonging to a common set of categories. In our view, there are five main categories that are general enough to enable the development of component-based situation understanding and management systems. They are:

- **Customized repository:** provide access to information repositories that may be highly structured (e.g., geographical digital libraries with elaborate meta-data, sensor data) or highly unstructured (e.g., the Web). Users can work with these components to rapidly create a customized repository of information (documents and data) restricted to those supporting a particular task at hand.
- **Characterization/partitioning:** provide access to meta-data if it exists, or otherwise provide mechanisms for clustering, analyzing, sorting and organizing the contents of information repositories. These components can be used to browse sets of documents and, in the process, develop characterizations of them. In doing so, they can simultaneously specify a taxonomy of topics of interest to them and use it to describe and organize the documents.
- **Processing/extraction/piping:** represent automated and interactive components that manipulate data and run numerical simulations. They allow performing specific, commonly used services such as extraction (or addition as annotations) of information about places and dates referenced in documents, and linking them to geo-locations.
- **Interactive analysis and view control:** provide mechanisms for exporting data out of objects in the repositories (or aggregate objects collected by characterization and/or partitioning) and importing them into Interactive Analysis and Viewing components. Through these components users can interact with documents and the extended information about them. This interaction may consist of exploring relationships between the documents and laying out presentations that will help rescue teams understand possible connections between the events and facts taken place during the situation being considered.
- **Collaboration and visualization:** all previous services run on this substrate allowing multiple participants to interactively explore and understand the objects being manipulated with (or output by) the other services. Components of this category are responsible for supporting user interaction with documents and geographic information using highly graphical techniques to represent relationships.

2.1 GeoWorlds components, features and implementation issues

Figure 1 sketches the organization of components in the current GeoWorlds system. A fundamental requirement of the system is to effectively provide full functional support to spatially distributed teams. The GeoWorlds system is implemented using the client-server architecture paradigm.

The client side is embedded in the Habanero collaborative framework, and therefore contains all applications and services that support collaboration. The client is used to log in to the server and start collaborative sessions to trade results in shared information spaces. The GeoWorlds server, contains all services and tools that are common to a GeoWorlds session and their results and information are available for all teams or clients collaborating in a session. The GeoWorlds system integrates technologies from several Universities and Research Labs, each of them contributing with the reference implementation of one or more components from the basic set of five components the architecture is made of, as follows:

- **Customized repository:** Components of this class include UCSB Alexandria Digital Library (gazetteer service), UIUC HDF² (structured data storage) and USC/ISI DASHER³ information space analysis tools (managing repositories and gazetteer access).
- **Characterization/partitioning:** Components of this class include University of Arizona neural-network clustering tools [3], and USC/ISI DASHER document partitioning services.
- **Interactive analysis and view control:** Components of this class include UC Berkeley Multi-Valent Documents [7] tools for marking and annotating.
- **Processing/extraction/piping:** Components of this class include all wrappers, scripts and routers available in the GeoWorlds system and used to pass information back and forth between system components with different data format requirements.
- **Collaboration and visualization:** GeoWorlds uses the NCSA Habanero collaboration framework to support collaboration among rescue teams.

The current version of GeoWorlds is implemented in Java JDK 1.1.6, with some wrappers written in C and Perl to encapsulate non-Java components and COTS software (such as ESRI ArcView.) Communications between components was achieved using Java RMI, TCP/IP sockets or Windows dynamic data exchange (DDE), depending on the component being integrated with the system. The following sections describe the set of reference implementations of the components integrated in the system so far.

2.1.1 The Registry Component

A valuable concept in component-based distributed systems is the possibility of integrating new components and registering with the system the services that they provide. This allows the users and/or the other components to become *aware* of the requirements and resources that this new component brings into the system. In GeoWorlds we support this functionality and have extended the concept to also register events or *incidents* that become part of the system and users can identify, track their evolution and store as part of their body of experience. Incidents and services are registered in the GeoWorlds server and their

information is available to all clients joining a GeoWorlds session.

The service registration forms the conceptual basis of the exchange of data and information between components. It defines the type of interface that the component implements (e.g., socket, local method invocation (LMI), remote method invocation (RMI), DDE, command line execution) and the names and types of the service inputs and outputs (which are then used to properly chain components when executing applications). For example, a tool for analyzing the spread of a toxic-gas release may require the ambient temperature in Kelvin degrees as input. The system can inspect the Service Registry to see if this variable matches the output of some other registered service (e.g., a wrapper to NCDC's web site with hourly surface meteorological conditions) and obtain this value from that service.

In certain applications, the registration of incidents or events may play an important role in the particular session that teams are working on (see Section 3) and provides the system with all the relevant properties and description of the incident. Incidents are described by their name, the type, the date the incident was registered with the system and the status (open or closed). The main property of the incident is the incident location which can be entered as the exact address or geographic location or by some description of the neighboring environment or locations. Once the incident location is defined, the system is able to map its location by querying the registered Data Warehouses to obtain maps having place names matching the incident location description. These maps or layers can then be displayed in the GIS component together with a mark highlighting the incident location geographic position.

2.1.2 The Geographic Information System (GIS) Component

The current version of GeoWorlds has ESRI's ArcView registered as its GIS component. This CoTS non-Java application has been integrated into the system and is used to display the geographic data obtained from registered Data Warehouses.

A key functionality of the system is to allow the user to interactively *move* back and forth between the document information space and the geographic (or physical) world. Within GeoWorlds, ArcView's functionality has been enhanced to *bi-directionally* interact with other system components such as the toxic-gas diffusion analysis tool and DASHER (the document management tool.) For example, from ArcView the user can extract the list of place names belonging to several Arc/Info layers (e.g., streets, hospitals, airports) whose location fall inside a marked region. This list of place names is then exported to DASHER where the user can search DASHER's information spaces to obtain documents associated to the selected place names. Flow of information in the opposite direction is also possible. The location information (e.g., street address) from documents in DASHER is sent to ArcView for geo-referencing the selected documents.

Following the component-based approach, GeoWorlds implements ESRI's ArcView through its GISViewer Interface. In this way, any application implementing this interface can then be plugged into GeoWorlds as a registered GIS component and be dynamically loaded when deemed appropriate. We could replace Arcview with any other GIS component, for example, UCB's Map Viewer, without having to modify any part of GeoWorlds code. Unfortunately, Arcview currently provides very basic support for communicating with external applications running either on a remote machine or locally. The

² Hierarchical Data Format

³ Defense Acquisition Services for High Performance Electronic Commerce.

mechanisms currently supported are RPC (Remote Procedural Calls) and Windows DDE (Dynamic Data Exchange). We use the

knowledge of the initial conditions. Providing the system with this functionality involves three mayor steps. First, the system must be able

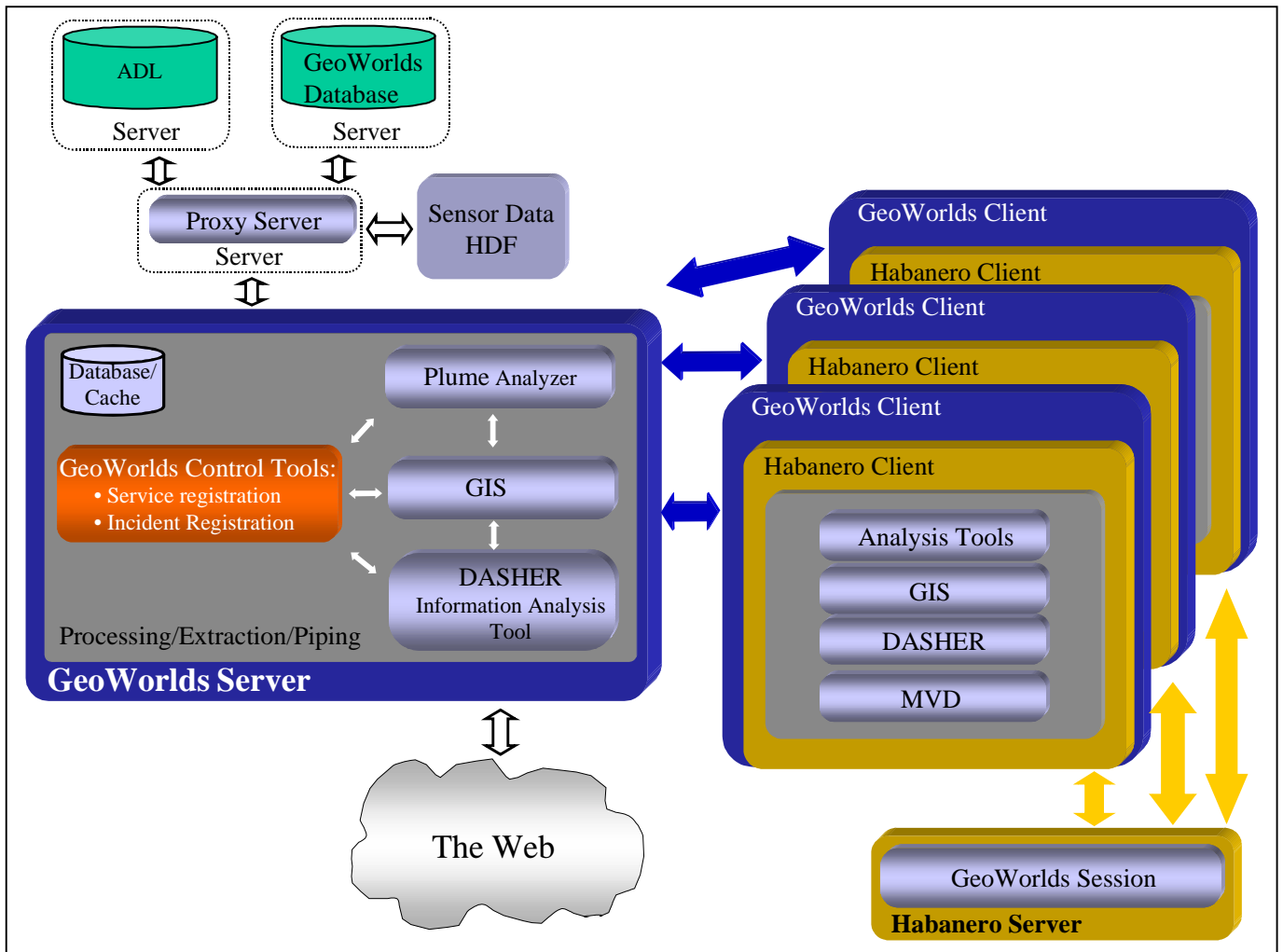


Figure 1: Schematic block diagram of GeoWorlds' high-level system architecture

latter as the mode of communication between GeoWorlds and Arcview. A Java wrapper that implements the GISViewer interface acts as the receiver for all requests to Arcview from GeoWorlds. The implementation consists of DDE calls to invoke scripts, written in Avenue (Arcview's own scripting Language), via a Java-DDE bridge. The ArcView scripts are bundled in an Arcview project that is loaded in when the GeoWorlds Server launches Arcview for the first time.

2.1.3 The Analytical Component

To coordinate operations, users of GeoWorlds must be able to obtain *predictions* of the physical processes involved in the situation. The system needs direct interaction with analytical and simulation tools that can properly model these processes in space and time and compute predictions. Usually, these simulations involve the solution of boundary-value problems that require the

to evaluate the parameters representing the initial conditions of the numerical simulation. Second, it needs computational power to run the application in a timely manner and third, it needs visualization capabilities to display and analyze the results.

So far, we have integrated into GeoWorlds two different toxic-gas plume analysis tools, Aloha [1] and Slab [4], for modeling the dispersion of heavier-than-air gases. These models calculate the dispersion of the cloud by solving the conservation equations (mass, momentum, energy) and take into account the atmospheric conditions and the properties of the terrain. They compute the time-averaged concentration of the toxic substance as a function of time and downwind distance.

2.1.4 The Document Management Component

GeoWorlds uses the USC/ISI DASHER (Defense Acquisition Services for High Performance Electronic Commerce) Information Space Analysis Tools [6] as its document management component. DASHER helps users quickly gain knowledge or expertise on or about a given region of space or topic by collecting sets of documents or data from *information spaces* that were previously originated and are being maintained by specialized groups. At the same time, users can create their own custom organization of these large collections of documents.

In contrast to many other services available for information search and retrieval, DASHER helps users in *making sense* of the sets of data sources, by providing tools for characterizing, partitioning, sorting and filtering them. It combines multiple methods to assist in understanding, particularly natural-language text extraction techniques and ontology-based categorization. It provides graphic interfaces for set-manipulation operations and for visual display of partitions, giving users means of identifying topics of interest to them and sorting the space of data sources accordingly.

One of the most significant features of the tool set is that it gives the user multiple means of characterizing the contents and topics of the document set. This includes the ability to wrap net services, such as Yahoo! or some Internet Yellow Pages, to extract explicit formal taxonomies that have been developed as a common way of structuring the world. The methods also include the ability to extract implicit organizing features of the world, such as terms that appear to represent the hot topics in a community of interest (indicated, for example, by frequency of usage). Finally, the tools help users define their own categories, and sort a set of documents according to those categories, by mapping users' structuring of the information onto the explicit and implicit characterizations that have been extracted from the document set.

2.1.5 The Collaboration and Annotation Components

GeoWorlds features synchronous collaboration and joint information exploration capabilities by incorporating University of Illinois NCSA's Habanero [2] and UC Berkeley's Multivalent Document (MVD) [7] technologies as components. Habanero provides the mechanisms by which teams of spatially distributed users of GeoWorlds can *see* in real-time what each of the remote users are doing in terms of exploring and analyzing the different information spaces that they are working on. They not only share the content of the information but also the way they operate on it.

Habanero is a client/server collaborative framework for remote users to share tasks and information live over the Internet. The GeoWorlds client is embedded in the Habanero framework and uses the Habanero API to allow GeoWorlds client sessions exchange mouse events, keyboard events, and computational results.

Typically, the Habanero server is run on the same machine running the GeoWorlds server, but this condition is not restrictive. Habanero clients can define a new collaborative session or join an existing one. For example, if in the DASHER component one user clicks on a category tree node to expand it to view its subnodes, then all the users get to view the expanded

sub-nodes. Or, if one user creates new category tree node, then this new node is created for all the users.

Through Habanero, GeoWorlds incorporate the annotation layer (and its corresponding behaviors) from the MVD system. This layer allows users to mark-up documents by highlighting word phrases, adding hyperlinks, and attaching notes. These capabilities allow users to emphasize key points in the document, provide supporting documents, and attach comments and critiques to the document. With the incorporation of Habanero, this annotation facility becomes even more powerful. With both components in place, remote users are able to share the mark-ups, to carry on live discussions, and save the content of the discussion as annotation.

2.1.6 The Information Repository Component

GeoWorlds' data warehouse is a key component of the framework. It gives the system the possibility of accessing large and diverse collections of data and information from registered digital libraries or repositories and acquiring real-time sensor data from spatially distributed instruments. This feature drastically extends the amount and type of geographic data and information that GeoWorlds users have access to as oppose to normal GIS systems that usually provide information contained within a single organization.

At the current state of the project, the data warehouse has access to one digital library, UCSB's Alexandria project (ADL), to the GeoWorlds Database and to sensor data collected in HDF format. The data warehouse is written in Java and implements the Java interfaces wrapping the repositories specifying the available methods to query and retrieve the information.

The UCSB's ADL project [5] maintains a distributed digital library for geographically-reference information. The library holds different type of objects (e.g., maps, photo images, satellite images, streets, hospitals) that can be accessed by some description and its geographic location. The communication between GeoWorlds and ADL is done through the GeoWorlds' DataWarehouse Interface which is written in Java.

The GeoWorlds Database is an Oracle relational database that teams can use to store any type of data and information that they may need to have special access to. For example, it can be implemented locally to be used as a cache where accessed material is stored and can later be searched for quick retrieval. The database interfaces with the data warehouse through the JDBC drivers.

Sensor data in HDF format can be directly accessed by GeoWorlds using normal HTTP or FTP protocols. The data can then be analyzed using NCSA's JHV [8] viewer. In addition, the data uploaded into JHV can be exported as a raster image that is then imported into ArcView as an additional layer.

3 Illustration of prototype capabilities: Conducting emergency relief operations for the release of a toxic gas.

To show how GeoWorlds can currently be used in a real-world situation and to illustrate the flow of information and interplay between the implemented components, we consider a HA/DR team acting on the following scenario: a massive amount of a toxic gas is

released in an urban area. GeoWorlds helps in assessing the impact on the affected region by laying out and discussing candidate resources for responding to the situation and monitoring for events that require re-evaluating previous conclusions.

We consider the release of a heavier-than-air toxic gas that quickly begins to spread impacting a heavily populated region. Using either the precise geographic location or informal description (e.g., “between the north runway of the Washington Building and the entrance to the County Hospital” or the street address) of the release location and the characteristics of the situation, a user proceeds to register the incident with GeoWorlds. Using the incident location information, GeoWorlds queries the registered datawarehouses to retrieve a list of maps with information matching the description of the incident location or whose geographic extent contains or overlaps the incident’s geographic location. The list of available information is presented in a data browser where the user can visually select the files wanted for download. In our current test scenario we focus on the San Diego/Point Loma (California) region and a typical query returns about 30 – 40 files of geographic information and satellite images in ESRI’s shape format and jpg, respectively. The typical download size (compressed) of the shape files is about 100KB. However, the size of some shape files (e.g., those containing street information of a large urban area) and of the digital images is in the order of 20MB or more. During our tests, the typical downloading time (using a 10 baseT connection) is usually below 1 minute.

Once the maps and digital images of the region are downloaded, they get automatically loaded into GeoWorlds’ GIS component and the rescue team can start planning their operations with the support provided by the GIS. For example, to obtain an estimate of the geographic spread of the toxic material, the plume analysis tool can be launched to compute the region of space covered by the toxic gas. The result from the analytical tool is then imported into ArcView as an additional layer. If real-time sensor data from instruments measuring the concentration of toxic gas would be available in HDF format, then the users can also add this information to the GIS display to contrast it with the results of the model prediction. The differences can be used to re-evaluate the assumptions in the simulations and to compute a newly and more accurate prediction. At the same time, a list of places (e.g., schools, hospitals, airports) and geographic points of interest can be generated by intersecting the region marked by the toxic gas with the rest of the layers loaded in the GIS. This list of places of interest can then be transferred to DASHER to obtain information associated to some of these names.

Different teams can use DASHER collaborative capabilities to find and retrieve specific information on the incident and operations. For example, if a transportation expert team would have a Habanero session running, then the team using GeoWorlds can join they session to find transportation equipment to evacuate perishable goods. Through the collaboration environment, the transportation experts can guide the users of GeoWorlds to find internet pages of candidate providers. If those companies are listed in some internet Yellow Pages server, DASHER can parse their results, extract the street address of the companies and communicate with the GIS component to display their geographic location relative to the affected area. We have tested this functionality with about four DASHER clients joining at least two different Habanero sessions and the performance was good in

Windows NT machines with sufficient amount of RAM memory (~128 MB). The overall performance of the collaborative environment and the scaling with the number of users is mostly determined by the characteristics of the machine running the Habanero server and the quality of the connection.

4 Conclusions and Future Vision

The GeoWorlds system, as it stands today, is best understood as an initial experimental effort. Although not ready for general use, we believe it demonstrates the desirability of integrating a range of functionality. In particular, there is a great deal of utility in having ability to coordinate geographic information with document collections, and to collaboratively view and discuss the related information.

The current version of the system aims at helping teams performing humanitarian assistant/disaster relief operation to *understand* their situations in space and time. Users can rapidly create a *custom* repository of information about a geographic area of interest and associate them to collections of documents from the World-Wide Web. Geographic and document-based information are *bi-directionally* correlated. The user can move from geographic to document space to learn facts about a region or to see the locations that documents are geo-referencing. These capabilities are embedded in a framework that enables synchronous and asynchronous collaboration between teams of users.

The system is designed following a component-based approach to enable continuous increase of functionality and portability. The current version of the system is implemented in Java and uses some wrappers, written in C and Perl, to encapsulate non-Java components and COTS software. This imposes a somewhat limited communication capability between components and we are therefore working towards a pure Java environment that will enable a seamless interaction between components with minimum user interaction. Our goal is to have a system where new available components can be added as *plug-ins* in a plug-and-play fashion and where components with similar functionality can be swapped at run-time and have the system to automatically perform the necessary interconnections. We are also working on enhancing the interaction between documents information and geographic views and giving the user an immersive 3D environment containing digital terrain information, realistic textures, results from numerical simulations and mapping of the document space.

GeoWorlds' current capabilities for switching over to new geographic areas are subject to the same limitations as all other efforts in this field: finding, registering, and fusing geospatial data is a very complex and difficult process with significant manual elements. The system is more supportive with respect to switching over to new document collections or new topic categories, although this is by no means an automatic process. The current incarnation of GeoWorlds is best positioned to support a group which uses the system's capabilities to do advance preparation and structuring of the information. That way, work under time pressure can focus on fine-tuning the information collection and on utilizing information that is (mostly) already in place. However, this is not an unreasonable restriction for a useful system. Just as firefighters today spend their spare time maintaining their equipment in preparedness for action, the information specialists of the future can reasonably be expected to do the same. As this paper is being written, we are preparing for experimental use of the system at

the Department of Defense US Pacific Command Headquarters in Honolulu, Hawaii. We expect in future reports to be able to provide more information on empirical usage issues.

5 References

- [1] Aloha (Areal Locations of Hazardous Atmospheres): Chemical Air Dispersion Model. Distributed and supported by the National Safety Council (www.nsc.org/ehc/cam/alohapg.htm).
- [2] Chabert, A., Grossman, E., Jackson, L., S., Pietrowiz, S., R., and Seguin, C., "Java object-sharing in Habanero," Commun. ACM 41, 6 (Jun. 1998), Pages 69 - 76.
- [3] Chen, H., Schuffels, C., and Orwig, R., "Internet Categorization and Search: A Machine Learning Approach," Journal of Visual Communication and Image Representation, Special Issue on Digital Libraries, Volume 7, Number 1, Pages 88-102, 1996.
- [4] Ermak, D., L., User's Manual for Slab: An Atmospheric dispersion model for denser-than-air releases. June 1990. Lawrence Livermore National Laboratory and Lakes Environmental (www.lakes-environmental.com).
- [5] Frew, J., Freeston, M., Kemp, R., Simpson, J., & Smith, T. (1996). The Alexandria Digital Library testbed. D-Lib Magazine (July/August 1996). www.dlib.org/dlib/july96/alexandria/07frew.html
- [6] Neches, R., Abhinkar, S., Hu, F., Eleish, R., Ko, I.-Y., Yao, K.-T., Zhu, Q., Will, P., Collaborative Information Space Analysis Tools, D-Lib Magazine, October 1998. For additional information see www.isi.edu/dasher.
- [7] Phelps, T., A., and Wilensky, R., "Toward Active, Extensible, Networked Documents: Multivalent Architecture and Applications", Proceedings of Digital Libraries '96.
- [8] see: <http://heineken.gsfc.nasa.gov/eosgis/eghome.html>