The Alexandria Digital Earth Prototype System

Terence R. Smith[1], Greg Janée[2], James Frew[3], Anita Coleman[4]

Alexandria Project, University of California at Santa Barbara,
Santa Barbara, CA 93106

1. INTRODUCTION

This note summarizes the system development activities of the Alexandria Digital Earth Prototype (ADEPT) Project.[5] ADEPT and the Alexandria Digital Library (ADL) are, respectively, the research and operational components of the Alexandria Digital Library Project. The goal of ADEPT is to build a distributed digital library (DL) of personalized collections of geospatially referenced information. This DL is characterized by: (1) services for building, searching, and using personalized collections; (2) collections of georeferenced multimedia information, including dynamic simulation models of spatially distributed processes; and (3) user interfaces employing the concept of a "Digital Earth." Important near-term objectives for ADEPT are to build prototype collections that support undergraduate learning in physical, human, and cultural geography and related disciplines, and then to evaluate whether using such resources helps students learn to reason scientifically. Collections and services developed by ADEPT researchers will migrate to ADL as they mature.

2. ARCHITECTURE

The ADEPT architecture (Figure 1) extends the ADL architecture while diverging from the latter's traditional client-server roots. ADEPT is a set of distributed peers, each supporting one or more collections of objects, subsets of which may be "published" (made visible) to other peers. The "library" is the sum of all such collections.

Collections are the primary organizational metaphor in ADEPT. A collection is a set of independent, typed objects (more precisely, a set of references to objects) together with certain contextual and structural collection-level metadata. Objects are largely undefined, save for their ability to provide object-level metadata and to be uniquely identified. While the contents of the library may thus be utterly heterogeneous, three features integrate the library into a uniform whole: (1) common collection-level metadata; (2) the ADEPT "bucket" system (a common model for object-level metadata); and (3) a central collection discovery service.

ADEPT buckets differ notably from other metadata schemes such as detailed content standards (e.g., FGDC [2]) and high-level definitions (e.g., Dublin Core [3]). The ADEPT bucket system is a transparent metadata aggregation system in which semantically similar, quasi-strongly-typed metadata fields may be formally grouped to provide higher-level search and description capabilities. Objects map their relevant native metadata to buckets; significantly, both field names and values are recorded. Collections index, aggregate, and summarize this information. They allow clients to search by entire buckets and to retrieve bucket-level descriptions of objects, and also to "drill down" into buckets (i.e., to search by specific metadata fields that have been mapped to the buckets). Buckets are strongly typed, being either spatial, temporal, hierarchical, textual, qualified textual, or numeric. Associated with each bucket type are a standard representation (e.g., polygons and boxes for spatial buckets) and a set of standard operators (e.g., set algebra). This combination of a formal aggregation system with strong typing yields a metadata model that is flexible enough to accommodate a wide range of metadata, yet expressive enough to support powerful, targeted queries.

ADEPT collection metadata includes contextual information (scope, purpose, etc.) and automatically derived structural information such as the spatiotemporal distribution of objects, the types and counts of objects, and bucket metadata mappings. Metadata mappings are qualified with statistics to give clients an indication of how well represented metadata fields are within buckets.

New ADEPT services layered atop the existing ADL infrastructure support: (1) creating and managing local collections; (2) creating new objects from client-supplied metadata; (3) importing objects from remote collections into local collections; and (4) categorizing objects according to client-supplied thesauri. In creating new objects clients are free to provide any metadata, the only requirements being that the metadata be represented in XML and that the client provide an XSLT script mapping the object metadata to a standard bucket-level description. Remote objects are currently imported by simply copying them, but a future version of ADEPT will also support proxy objects and explicit linking. The ability to create new collections and categorize them with user-supplied thesauri effectively supports building personalized collections.

To facilitate including legacy content in the library, especially content managed with relational database technology, we have developed a generic collection driver for relational databases. The driver assumes only that it is connected to the collection via JDBC. The collection's schema, metadata bucket mappings, thesauri, etc. are all described in configuration files. Metadata reports are described by templates, which are interpreted by custom, query-driven report generation software. Query translation is controlled by a Python-based scripting system that sup-

[1] smithtr@cs.ucsb.edu [2] gjanee@alexandria.ucsb.edu
[3] frew@bren.ucsb.edu [4] asc@acm.org
[5] The research reported in this paper was partially supported under NSF IIS-9817432.

ports user-defined functions and customized query formation and optimization.

The ADEPT collection discovery service manages a registry of known collections and, more significantly, supports discovering collections probabilistically according to their relevance to a given query. Using the notion that relevance is determined by the numbers of objects satisfying a query, we are: (1) investigating new indexing techniques that support joint multidimensional indexing of collection contents using sparse, compact data structures; and (2) using the simple histogram and count information contained in collection-level metadata to derive relevance from typed spatiotemporal information density.

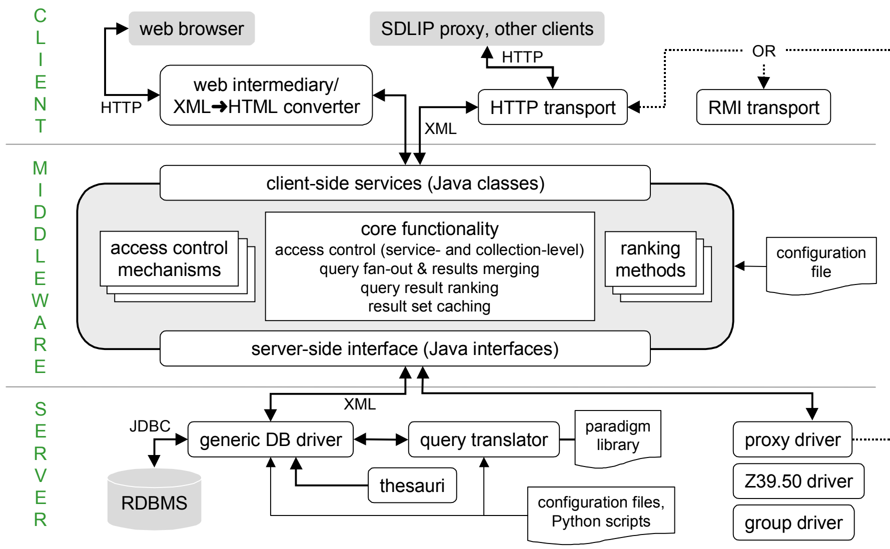

Figure 1. ADEPT architecture and implementation.

Future directions for the ADEPT architecture include supporting a richer information model, facilitating interoperability, and supporting dynamic content such as models of Earth processes and other executable programs. ADEPT's current information model (collections of independent objects) does not capture key relationships of benefit to users. For example, objects that are members of a series share a relationship with each other and to the series as a whole; we plan to make such relationships explicit. With regard to interoperability, we plan to facilitate searching over other types of collections by supporting connections to Z39.50 servers, and by extending the SDLIP protocol. With regard to dynamic content, our focus is on defining an architectural layer that delivers executable content to clients in a platform-neutral fashion, and that allows programs to interact with library contents and to populate the library with new objects. We are designing a suite of additional services that support building personalized collections by, for example, allowing users to: (1) build collections according to personal preferences; (2) build models representing concepts in different ways; and (3) build concept maps.

3. CATALOG AND COLLECTIONS

Three prototypical ADEPT collections are being built to aid in designing future collections and services, and to serve as a basis for evaluating the educational impacts of ADEPT services. The collections focus on processes and forms in (1) physical geography (e.g., landscapes and tectonic phenomena); (2) human geography (e.g., diffusion phenomena); and (3) cultural geography (e.g., religious sites). Initial collections for physical and human geography have already been tested in electronic classroom presentations. The collection building process involves (1) acquisition; (2) cataloging; (3) ingesting; (4) evaluating with respect to user scenarios; and (5) building ISO-standard topic maps. The collections may be searched by concept and reorganized with different levels of granularity into personal collections for various pedagogic purposes (e.g., lectures, self-paced labs, and individual and collaborative learning sessions.) Catalog descriptions of objects employ the ADEPT Learning Objects Metadata Model (ADEPT LOMM) standard, which is based on the IMS [4], DLESE [5], and IEEE LOM [6] metadata standards, and an ADEPT-developed standard for model metadata [7]. The model's key educational and pedagogical elements include (1) type of learning resource; (2) learning context; (3) interactivity level; and (4) description.

4. REFERENCES

[1] *The Alexandria Digital Earth Modeling System (ADEPT)*. http://www.alexandria.ucsb.edu/adept/proposal.pdf.

[2] Federal Geographic Data Committee. *Content Standard for Digital Geospatial Metadata*. http://www.fgdc.gov/metadata/csdgm/.

[3] Dublin Core Metadata Initiative. *Dublin Core Metadata Element Set*. http://dublincore.org/documents/dces/.

[4] IMS Global Learning Consortium. *IMS Learning Resource Meta-Data Information Model*. http://www.imsproject.org/metadata/mdinfov1p1.html.

[5] DLESE. http://www.dlese.org/.

[6] IEEE Learning Technology Standards Committee. *LOM: Base Scheme - v3.5*. http://ltsc.ieee.org/doc/wg12/scheme.html.

[7] Crosier, S. *Content Standard for Computer Model Metadata*. http://www.geog.ucsb.edu/~scott/metadata/standard/.