ADEPT Research Proposal:
Personal, Distributed Collections

Greg Janée • September 22, 2000

Background

The deficiencies of the Web are well-known: lack of structure, poor searchability, extreme volatility, and so forth. Digital library projects have attempted to address these problems, but for the most part the projects (ADL in particular) have focused on the searchability problem alone. Left largely unaddressed are the issues of how information is created, stored, organized, and made searchable in the first place.

Proposal

We propose to develop a prototype digital library that addresses the full lifecycle of information, from creation and storage of information, to organization of that information, and finally to publishing and subsequent discovery and reuse of that information. The architecture of this new library is based on three core ideas:

- There is a single unit of information on which the library operates: the *digital object*. A digital object is a logical entity that is identified by a persistent identifier and that encompasses all of the data, metadata, and functional behavior of a piece of information. The granule size of a digital object is deliberately left unspecified, but it is safe to say that a single number ("the rainfall in Santa Barbara in 1998 in inches") is probably too small to comfortably be a digital object and a heterogeneous collection ("the contents of Davidson Library") is probably too large.
- The library provides principally one organizational structure for digital objects: the *collection*. A collection is an arbitrary set of digital objects that is tied together with contextual and structural metadata and that is identified by a persistent identifier. The collection structure provides the unit of searchability (the user finds digital objects by locating collections and then searching within collections) and the metaphor by which the user's workspace is organized (the user stores new and discovered digital objects in personal collections) and the mechanism for publishing (the user's collections may themselves be made searchable by others). The library may provide other organizational structures, most notably *series* of digital objects, but the collection is the principal structure.
- The library imposes no requirements on digital object data, metadata, or behavior; digital objects and collections may be (in fact, will be) quite heterogeneous. The only uniformity provided by the library is in the form of uniform views of metadata and uniform mechanisms by which digital object behavior is invoked.

The library's functionality includes:

- Creating new collections; managing collections; creating new digital objects; populating collections with new and discovered digital objects.
- Discovering relevant collections.

- Searching collections for digital objects.
- Viewing collection and digital object metadata; invoking digital object behaviors.

Development plan

We propose a one-year development plan broken into four, three-month stages.

In the first stage, the functional behavior of digital objects will be ignored; digital objects will be treated as consisting of metadata only, much as they currently are in ADL. The library's overall architecture, specifically its collection-related functionality, will be prototyped. A proof-of-concept, web-browser-based client that demonstrates library functionality will be developed.

In the second stage, the prototype client and architecture will be refined.

In the third stage, functional behavior will be added to digital objects.

The fourth stage is left open to address any significant problems that are bound to arise in the first three stages.

Required resources

Two software engineers.

A scenario to test the library on, i.e., a body of information suitable for organization into digital objects and collections.

A "power user" to exercise the system and provide feedback.

Evaluation; definitions of success

There are at least three kinds of evaluation applicable to this proposal, and correspondingly three notions of what constitutes success.

The first and most narrow definition of success is: Does the system work as advertised? Is it possible to find digital objects? To create new collections? And so forth.

A second and broader definition of success is: Is the system pragmatically useful? Does it support information storage, organization, and retrieval in ways that are superior to existing technologies? In particular, is it possible to use the system to store and organize instructional materials in a way that encourages reuse by others and that facilitates incorporation into instructional settings?

Third, and most broadly: Does the system have any impact outside the ADEPT project? Does it address any real-world needs? Do other digital library efforts build on the architecture or adopt aspects of the architecture? Is the system adopted by potential users of such technology such as DLESE?

Related work

This proposal is closely related to other ADEPT ideas and projects such as the Digital Earth metaphor, geospatial browsing, and tightly-integrated gazetteer services. These other efforts are necessary to this proposal and complement this proposal.

In addition, there are a number of other ideas that would nicely extend this proposal. For example, this proposal is focused mostly on architecture, and does not really address the user interface to the library beyond constructing a proof-of-concept interface. It would be very instructive and worthwhile to explore alternative interfaces to the library, specifically, from existing tools such as GIS systems like ARC/INFO and general productivity tools like Microsoft Office.