ETANA-CMV: A coordinated multiple view visual browsing interface for ETANA-DL

Johnny L. Sam Rajkumar

Thesis submitted to the faculty of
Virginia Polytechnic Institute and State University
in partial fulfillment of the requirements for the degree of

Master of Science

in

Computer Science and Application

Dr. Edward A. Fox - Chair
Dr. Weiguo Fan
Dr. Chris North

December 13, 2006
Blacksburg, Virginia

Keywords: Archaeology, Digital Libraries, ETANA, Data Cube, Information Visualization

ETANA-CMV: A coordinated multiple view visual browsing interface for ETANA-DL

Johnny L. Sam Rajkumar

Committee Chair: Dr. Edward A. Fox

Abstract

Archeological research embracing complex Information Technology techniques can result in vast quantities of heterogeneous information from different sites in different formats. ETANA-DL is an Archeological Digital Library (DL), providing services suited for the archeological domain. With a growing collection of records in the DL, it is a challenge to present them in an organized and meaningful way.

We have designed a new visual browsing interface called ETANA-CMV that aims to provide users a richer and more insightful browsing experience. ETANA-CMV allows users to navigate through the records in ETANA-DL that are multidimensional, hierarchical, and categorical in nature. ETANA-CMV was designed to be scalable, flexible, and easy to learn.

This interface employs a data cube based browsing index to counter performance issues that usually limit the interactivity of visual browsing interfaces to DLs. The interface has been integrated with the existing Browse Interface and the search service in ETANA-DL. Formative evaluation of the new visual interface led to several improvements in the interface. It appears that users were able to detect trends in the DL collections more accurately using visualization based strategies than with the existing textual browse interface.

To Sujatha

Acknowledgements

Special thanks go to my family. My mom inspires and motivates me. She has taught me a lot in life. Dad has given me a great deal of independence, for which I am grateful. I thank my lovely sister for the treats and sibling fights that keep me entertained, and my girlfriend Rasika for being ever so loving. I thank my grandparents for bringing me up with their blessings.

I am indebted to my three advisors. Many thanks to Dr. Fox for his guidance throughout my Masters program. His suggestions have helped me a lot in completing my thesis. The two courses that I have studied with him taught me a lot. I am indebted to Dr. Fan who molded my initial ideas into a concrete topic. The meetings with him were invaluable at every design iteration to improve my interface. Thanks to Dr. North for his suggestions during the evaluation phase and for making the "Information Visualization" course my favorite one at Virginia Tech! That was where I learnt about designing visual interfaces and was inspired to take this as a topic.

I thank Srinivas for helping me understand the code in ETANA and for answering every silly question I had about implementation. I thank the entire Digital Library Research Laboratory for their comments, and Sujatha, Uma, Ricardo, Rao Shen and Doug for their help and advice. I thank my friend Manjula Iyer for her insightful suggestions.

I thank the many teachers in my life so far who have shaped me into the person I am - Dr. Tatar and Dr. Santos at Virginia Tech and professors Hema, Kanimozhi, Balakrishnan, Nathan, SS and Jairam - who taught in my undergraduate program. I thank my school teachers Radhakrishnan, Jagdeesan, Helen Sudarsanam, Usha Rani, Philip, Sophia and Prema - I will always cherish my memories of them.

Thanks to IPO and John Talbott for funding me as a graduate student. I thank the Nutters who were wonderful hosts during my Blacksburg days. Many thanks to Dee Dee and Vidhya's family for providing me a nice place to work on my thesis. Thank you Ed Snow for repairing my car and making me feel safe during all the trips in the last one year.

I thank Sujatha, my best friend for all the fun we had. Thanks for being there during my best and worst moments at Virginia Tech. A big thanks to her family as well - now we are like family friends. I thank Azeem for being my best apartment-mate - we had endless nights talking about cricket and life back in India. Thanks to my boss Mohan for all the fun at work, the many walks and fun lunches. Without him, I would have probably finished six months earlier and missed a lot of fun this semester! I would like to express my warm regards to my close friends for always being there for me: Senthil, Natesan, Jothi, Crisly, Anil, Praveen, Kumar, Elsaba, Malu and Malarvizhi. Thanks to my other friends at Virginia Tech: CP, Jagan, Kamini, Matt, Manjula, Priya, Sreeram, Sridevi, Srinivas, Uma, Vijay, Vikram. It was a pleasure spending these two years with such wonderful people.

Most importantly, I want to thank God for always being there when I need Him.

Table of contents

List of Figures

List of tables

1. INTRODUCTION

1.1. Problem statement

In recent times, research in archeology is using Information Technology techniques for data collection, preservation, and dissemination. There is a vast amount of information available from different archaeological projects. There are few common standards among these sites regarding the ways they collect, organize, and store data. This leads to a great deal of heterogeneous data from different projects. Archeology is beginning to adopt Digital Libraries, henceforth called DLs in this document, to store their data, and make it available to other researchers and users. Archeological DLs manage this data and provide services that are relevant to archeologists and other interested users.

Browsing is one of the primary services provided by DLs. Due to the huge quantity of data and information, presenting the archeological records in the DL to users in a meaningful manner is a great challenge. All DL systems have bounds on the number of records that can be displayed, or browsed at any stage. Meta-data is used to group together records under categories that are meaningful in the archaeological domain. Textual browsing interfaces are the common way of presenting the records in the DL to users. Unfortunately, often they limit the users' ability to gain insight into the collections in the DL. Visual browsing interfaces are also found in a number of DLs. They approach the problem of presenting the vast number of records from the visualization perspective. However, they have several performance issues.

Archeological records should be presented in a browsing interface that provides users a rapid and efficient manner of retrieval and also insight into the archeological collections in the DL.

1.2. ETANA-DL

ETANA-DL is one such archeological digital library with initial data from eight sites in the Middle East. It serves as a global repository for records about different artifacts like bones, pottery, documents, etc. It combines data from several excavation projects like Nimrin [1], Megiddo [2], Lahav [3], Umayri [4], etc. Information access services are an important category of services provided by ETANA-DL. The ETANA Browse Service was designed to achieve the goal of enabling users to locate relevant information in the DL easily and accurately [5].

ETANA adopts the harvesting approach supported by the Open Archives Initiative [6] to handle heterogeneous archaeological data. The data from sites that want to be data providers to ETANA-DL is harvested into a central Union Catalog through OAI-PMH (OAI Protocol for Metadata Harvesting). ETANA-DL then provides its services on the local copy of the harvested data [5].

1.2.1. The nature of records in ETANA-DL

The "etanamultidibrowse" database provides the tables with the browsing index needed by the ETANA Browse Service. Figure 1 gives the tables that form the browsing index for ETANA-DL.

Idtable

Field
unionid
spaceid
objectid
timeid
datestamp

space_tbl

Field
spaceid
SITE
PARTITION
SUBPARTITION
LOCUS
CONTAINER

object_tbl

Field
objectid
OBJECTTYPE
NAME

time_tbl

Field
timeid
Period
Chronology

Figure 1: Tables in the etanamultidibrowse database: the browsing index for ETANA-DL

The *idtable* is the main table, with an entry for each record in the DL. Each record is associated with a *spaceid, objectid,* and *timeid*. The *datestamp* is simply a timestamp of when the record was harvested into the ETANA-DL. The *space_tbl*, *object_tbl,* and *time_tbl* describe the *spaceid*, *objectid,* and *timeid* attributes for each record. Each *spaceid* in the *space_tbl* has a categorical value for Site, Partition, Sub-partition, Locus, and Container. The *object_tbl* expands each *objectid* to a corresponding Object type and an Object name. The *time_tbl* expands each *timeid* to a corresponding Period and Chronology. These tables define the browsing structure that is used by the ETANA Browse Service.

We describe the records in ETANA-DL as **multidimensional**, since they can be browsed by space, time, or object type. Each dimension is **hierarchical**. For example users can browse through the Space dimension at any of the following levels, namely site, partition, sub-partition, locus, and container. Each level is **categorical**. For example, the site level in the Space dimension shows that records are from one of the following eight sites, namely Bab edh-Dhra, Lahav, Madaba, Megiddo, Mozan, Nimrin, Umayri, and Umm el-Jimal.

1.2.2. Data cube in DLs

The data cube [7], also known as multidimensional database [8], is a data analysis tool used in data warehouses. Users of data warehouses are presented data as a multidimensional data cube with 2-D, 3-D, or higher dimensional views. Users explore these views trying to discover interesting information. Each cell in the data cube has values representing some measures of interest [9].

These measure attributes are those whose values are of interest. Other attributes are selected as dimensional attributes and the measure attributes are aggregated with regard to the dimensions [10]. For example, consider a database containing a collection of cars. A data cube can be built with COUNT as a measure attribute, and MAKE and MODEL as dimensions. This data cube would provide aggregated total counts of all cars of a given make and/or model. There could be other dimensions in our example such as year and number of cylinders.

Browsing is an iterative process of users refining their browsing path. Summarization of intermediate result sets is important in providing the user with information that can help him navigate through the browsing structure and narrow down his result set interactively. Summarization refers to aggregating the items in the result set in some manner. In DLs with a large number of records, the categories in the browsing structure are usually associated with a count of records in them [10].

Data cubes have been used as the basis for efficient generation of summaries for query result sets. A technique for integrating hierarchical categories found in DLs into data cubes has been presented by Geffner [10]. The hierarchical categories can be modeled as a categorical tree. Each query has a range, which means the query contains a set of categories at a given level in the tree. The summary for a query with a given range of nodes is the aggregate of summarization associated with each node in the range. The hierarchy in the categories is addressed with the following mapping: The summarization value associated with a node is the aggregate of all summarization values associated with all its sub-categories. In terms of the query, the range assigned to a category should enclose the range associated with all its sub-categories. Also, the mapping ensures that two disjoint categories must have ranges that are disjoint.

1.2.3. ETANA Browse Service

The browse service is an XML based web service that returns the corresponding categories of records for the requested navigation path in each of the three dimensions. For example, when given the Navigation path SITE=Bab edh-Dhra; PARTITION=C in the space dimension, and OBJECTTYPE=Pottery in the object dimension, it returns the following XML result (see Figure 2).

```
- <Response>
    <requesttype>1</requesttype>
  - <dimension>
      <dname>space</dname>
      <navigationpath>SITE=Bab edh-Dhra;PARTITION=C</navigationpath>
      <levelname>SUBPARTITION</levelname>
      <value>001</value>
      <value>002</value>
      <value>003</value>
      <value>005</value>
      <value>006</value>
      <value>009</value>
      <value>010</value>
      <value>011</value>
      <value>012</value>
    </dimension>
  - <dimension>
      <dname>object</dname>
      <navigationpath>OBJECTTYPE=Pottery</navigationpath>
      <levelname>NAME</levelname>
      <value>Unclassified</value>
    </dimension>
  - <dimension>
      <dname>time</dname>
      <navigationpath />
      <levelname>Period</levelname>
      <value>EARLY BRONZE IA</value>
    </dimension>
  </Response>
```


Figure 2: XML result returned by the ETANA Browse Service for the Navigation path SITE=Bab edh-Dhra; PARTITION=C in the space dimension, and OBJECTTYPE=Pottery in the object dimension

The service is a dynamic multi-dimensional browsing system in that it filters off the categories that do not have any records for the given navigation path. For example the categories bone, locus-sheet, and seed are returned in the object dimension when the navigation path Site=Nimrin is passed to the Browse Service. This means that records from Nimrin are about bones, locus-sheets, or seeds. The partition NE in site Nimrin has only Bone records. On adding the Partition=NE to the navigation path, the XML result returned shows only the Bone category under the object dimension. The categories that are not present in partition NE, namely locus-sheet and seeds, are filtered out.

1.2.4. ETANA Browse Interface

The ETANA Browse Interface is the current browsing interface to the records in ETANA-DL, and it feeds from the ETANA Browse Service. It is a textual browsing system and allows users to browse through the records using hyperlinks. Figure 3 shows the initial screen of the browse interface.

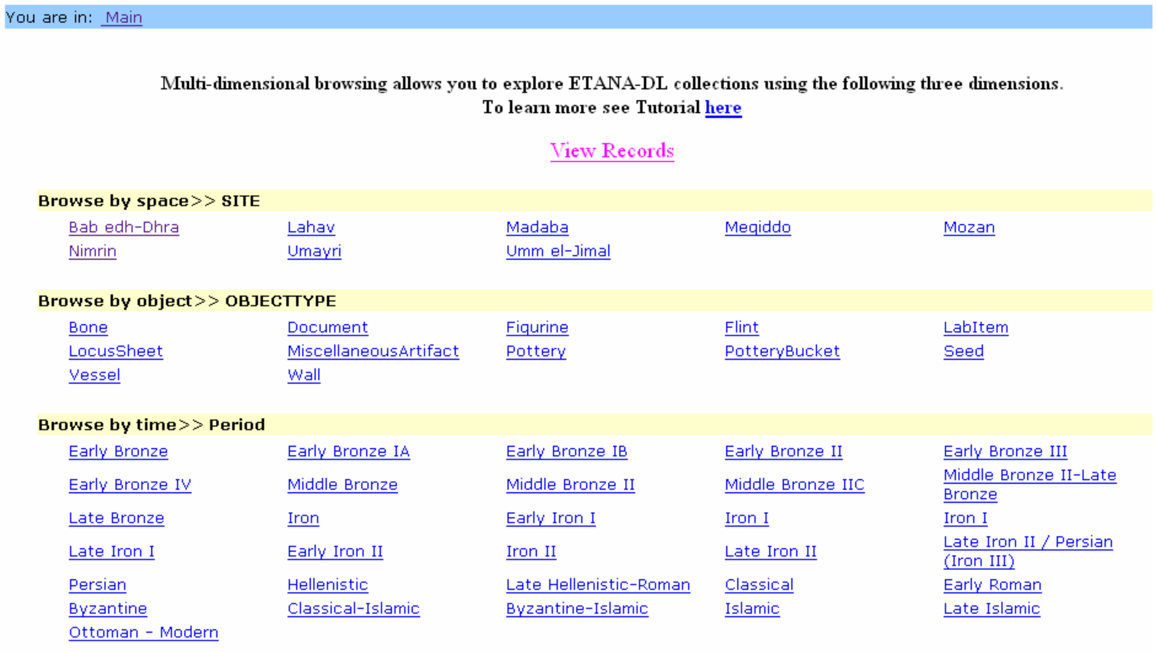

Figure 3: Initial screenshot of the ETANA Browse Interface

The three dimensions are specified and the categories in each dimension are listed in alphabetical order. Users can click on a category to add it to the navigation path and browse down the hierarchy in that particular dimension. For example, by clicking on Nimrin, users will get the following screen (see Figure 4).

Figure 4: ETANA Browse Interface after clicking the Nimrin Link – shows the partitions inside the site Nimrin

Note that Nimrin has been added to the navigation path. The partitions inside Nimrin are listed under the Browse by Space tab. The dynamic multi-dimensional nature of the browse service filters the object types that are not in Nimrin. Only those types of objects in Nimrin are listed under the Browse by Object tab and similarly only the time periods of records in Nimrin are listed under the Browse by Time tab. By clicking on the suitable hierarchical level in the navigation path, users can browse back to a higher hierarchical level.

The “View Records for the Context Below” link, on clicking, lists the records in the browsing context as a series of pages with each page having 10 records (see Figure 5).

Figure 5: Clicking the View Records shows the records in the browsing context in groups of ten

Some of the key features of the textual interface are:

Flexible: Users can seamlessly browse across dimensions. For example, a user can click the Nimrin link under the Browse by Space tab, and then click Bone under the Browse by Object tab. This is equivalent to users browsing through Bone records in Nimrin. Now, the user can click the NE link which is a partition in Site Nimrin. The NE link is listed in the Browse by Space tab. Then, the user ends up browsing through Bone records in the NE partition of Nimrin. This feature, that allows users to navigate the dimensions in the browsing structure in any order, makes this interface flexible.

Easy to learn: The complex ordering of archeological data in a multidimensional, hierarchical, and categorical format is a new concept to archeologists. The textual Browse interface is a very simple interface with minimal learning needed for new users with little experience using DL systems before. This, we believe, is due to the simple interface design and the uniform interaction provided across all dimensions. The interface treats all dimensions equally and provides uniform interaction to browse through all dimensions.

Scalable: The browse interface is not limited by the number of records in the ETANA-DL. Each time new collections of records are harvested into the system, the browsing index is updated. New categories are added to each hierarchical level in each dimension. The use of simple lists in a scrollable page makes the browsing interface scalable. When new dimensions are identified and added to the browsing structure, more scrolling will be involved. However, that seems to be acceptable; the Flamenco textual browsing system

[11] for the UC Berkeley Architecture Image Library displays nine dimensions for browsing, each having about 10 categories. The clicking and drilling down interaction for hierarchical navigation has no limitations on the number of hierarchical levels in any dimension.

1.2.5. Limitations of the current interface

The current Browsing Interface clearly reveals to the user the way in which records are organized in the DL. However it presents no insight into the collections. Based on interviews with archeologists, it was found that "the interface was restrictive." One archeologist complained that "it was not possible to browse the virtual stacks of the Digital Library and that the interface needs to be more visually stimulating". There is no clear way of letting the user know the distribution of records in the various categories in each dimension. A simple question like "Which is the site that has the most records?" involves users clicking each site and viewing the count of records in each site.

1.2.6. Information visualization in DL research

Card, Mackinlay, and Schneiderman define Information Visualization as "***The use of computer-supported, interactive, visual representations of abstract data to amplify cognition***" [12]. Information visualization as a field has a major impact on designing visual browsing interfaces to DLs. Visual interfaces to DLs serve users by providing services such as insightful browsing, domain specific analysis, and rapid information access. However, performance has been a major problem with such interfaces, when it comes to huge collections that are hierarchically classified in many dimensions.

1.2.7. Data cubes in visualization systems

The data cube provides different views of the data for user exploration and analysis. The summarized views in the data cube are sometimes presented in the form of interactive visual reports for users to explore, and analyze the data. Following are brief descriptions of some popular research initiatives combining visualizations and data cubes.

DIVE-ON [13] is one data cube visualization system. Users can interact with different views of a data warehouse in an immersive virtual environment using head and hand trackers. The data warehouse is built from one or more existing databases and a data cube is generated to feed the views to the immersive virtual environment rapidly.

Polaris [14] is a visualization system used for generating table based graphical displays of multidimensional data in quick time. It was designed to explore multidimensional data bases. It was later extended to support interactive visual exploration of hierarchically structured data sets in data warehouses that are represented as data cubes. The extended Polaris has been effectively applied to different datasets: a 12-week trace of mobile network usage, results from the 2000 presidential election, and historical business metrics for a hypothetical coffee chain.

Map cube [15] is another visualization tool for spatial data warehouses. It provides a cartographic visualization of the data cube by generating an album of related views in the cube. It is a combination of research in data cube, visualization and GIS. Map cube has been applied to the traffic data warehouse. It aids analysis and decision making in a warehouse storing aggregate planning data related to behavior of highway traffic.

1.3. Research questions

We believe that archaeological data when presented using a visual browsing interface to users can reveal more insights and patterns in the collections in the DL. The research questions are "How can we design a visual browsing interface for the collections in ETANA-DL while ensuring a flexible, easy to learn, and scalable browsing system? How can we provide more information to the users without limiting the interactivity of the browsing system?"

1.4. Hypothesis

We hypothesize that:

a) The problem of presenting heterogeneous archeological records to users in an insightful manner can be solved by our approach of designing a visual interface based on coordinated uniform multiple view design.

b) System performance when browsing through records in ETANA-DL can be enhanced by the use of a data cube based browsing index.

1.5. Approach

Our approach to presenting the archeological records to users is through the design of a visual browsing interface to ETANA-DL, "ETANA-CMV" [16]. We use the existing browsing index database to build a data cube for the visual browsing index. The new data cube based browsing index provides more information to users, and better performance than the existing browsing index. The number of archeological records in ETANA-DL keeps growing as more collections are integrated. We have a script that updates the data cube based browsing index used by ETANA-CMV.

Existing visual browsing interfaces to DLs were studied. We outline three design principles that we use for the design of ETANA-CMV so that it meets the flexibility, scalability, and learnability requirements also offered by the current Browse Interface to ETANA, while providing more insight into the collections. ETANA-CMV is based on a coordinated uniform multiple view design. The design and implementation of ETANA-CMV is described along with challenges encountered during the process. A formative usability study was conducted to identify usability problems in the interface and identify its relative strength as compared to the current textual browsing interface.

1.6. Contributions

The main contributions of this work are

- A visual browsing interface design that presents the archeological records in ETANA-DL in a meaningful and insightful manner.
- An approach that combines visualization research and database techniques to create a powerful browsing interface in terms of performance.
- An automated and componentized data cube generator that could be used to update the visual browsing interface when new collections are added to ETANA-DL and also by the current textual browsing interface to present more information to users.

2. SURVEY OF DL VISUAL BROWSING INTERFACES

In this chapter, we explore the goals of visual browsing interfaces, and approaches to designing such interfaces. We also present some case studies of popular visual browsing interfaces to DLs and study how they can be applied to the design of ETANA-CMV.

2.1. Goals

Providing means to rapid and efficient information access, insightful browsing, presentation of search results analysis of collections, and information sharing and collaboration, are objectives of visual browsing interfaces to DLs [17].

2.1.1. Rapid and efficient information access

Due to the huge quantity of records in DLs, it is a challenge to organize these records, so that they can be retrieved in a rapid and efficient manner. Hyperlink based textual browsing systems [11, 18, 19] use the meta-data available to deduce a browsing structure which groups records into categories. They allow multidimensional and hierarchical browsing of record categories. It is essential that visual interfaces should provide a richer browsing experience without limiting the information access that is provided by textual browsing interfaces. Users should be able to access the records by efficient use of the browsing structure, in little time. Time is a concern because visual interfaces are computation intensive, and given the complex nature of DL records, they could take time to load, or have slow response to user interaction.

2.1.2. Insightful browsing

Browsing is not limited to navigation among records in information systems like DLs. It has been described by a multi-dimensional framework [20]. Browsing has the following dimensions: Context, Behavioral, Motivation, Cognitive, and Resource.

- The context dimension is affected by the way the information is organized and presented by the interface.
- The behavioral dimension deals with the user activities during the process of browsing, like scanning through categories and interaction with the interface through input devices.
- The motivation dimension describes the needs of the user while browsing. What is the purpose of browsing and what are the goals that need to be achieved?
- The cognitive dimension deals with the user's domain knowledge/experience with the information collection, and similar interfaces.
- The resource dimension focuses on the form of the records, their representation as images, textual information, and the focus of the browsing system; it may be the navigation path or the actual content that is the primary focus of the browsing system.

The above framework outlines the need to treat browsing from more than an information access perspective. Visual browsing interfaces derive a lot of their design from information visualization research, whose primary motive is to amplify cognition and insights into the abstract data that is being represented visually. Textual browsing systems primarily focus on the context and resource dimensions of the framework. Visual interfaces should be designed so that they amplify the insights into the collections and provide a richer and insightful browsing experience to the users. They should allow users to gain an overview of the collections in the DL.

2.1.3. Presentation of search results in a browsing platform and query refinement

The search service is regarded as the primary means to access and locate records in a DL. It is very rare to see a DL without a search interface. The underlying search engines behind the search interfaces provide relevant records for the user query. The result set returned is usually displayed in batches in decreasing order of relevance so that users can scan through them and access the records.

It is hard to understand the inter-relation of the retrieved records. When these search results can be mapped back to a visual browsing interface it is possible that users can understand the relations in the records based on the organization of the records in the browsing interface.

Further, most search interfaces lack the tight coupling of the query specification and results browsing. There is not much assistance from the interface and the results retrieved, to help the user refine his query and search more effectively. Visual interfaces are expected to enable iterative query refinement by helping users understand the relationship between the query and the results retrieved.

2.1.4. Analysis of collections

By visually representing the abstract structure in the records, visual interfaces also can help users gain domain specific insights into the collections in the DL. Visual interfaces should enable the users to see the collections in the DL from several perspectives simultaneously. Such interfaces use the semantic relationships in the meta-data structure.

Mathematical techniques like Factor Analysis, Multidimensional scaling, and Latent Semantic Analysis are used to extract the underlying semantic structure. Clustering algorithms are used to spatially visualize patterns and semantic networks of the records for further interactive exploration. The modified Boltzman algorithm that computes attraction and repulsion forces among nodes based on data analysis is also used in spatial visualizations [21].

2.1.5. Facilitate information sharing and collaboration

Most DL interfaces are designed for single users. They do not take advantage of the behavior of like-minded people who access the collections in the DL. Visual browsing interfaces can support collaboration by showing records that other users accessed, annotated, or were interested in [22].

2.2. Approaches

Information visualization research has impacted the design of DL interfaces and the interaction techniques used in these interfaces. In this section, we explore some common approaches used in the design of DL visual browsing interfaces.

2.2.1. Interface design

2 D scatter plots:

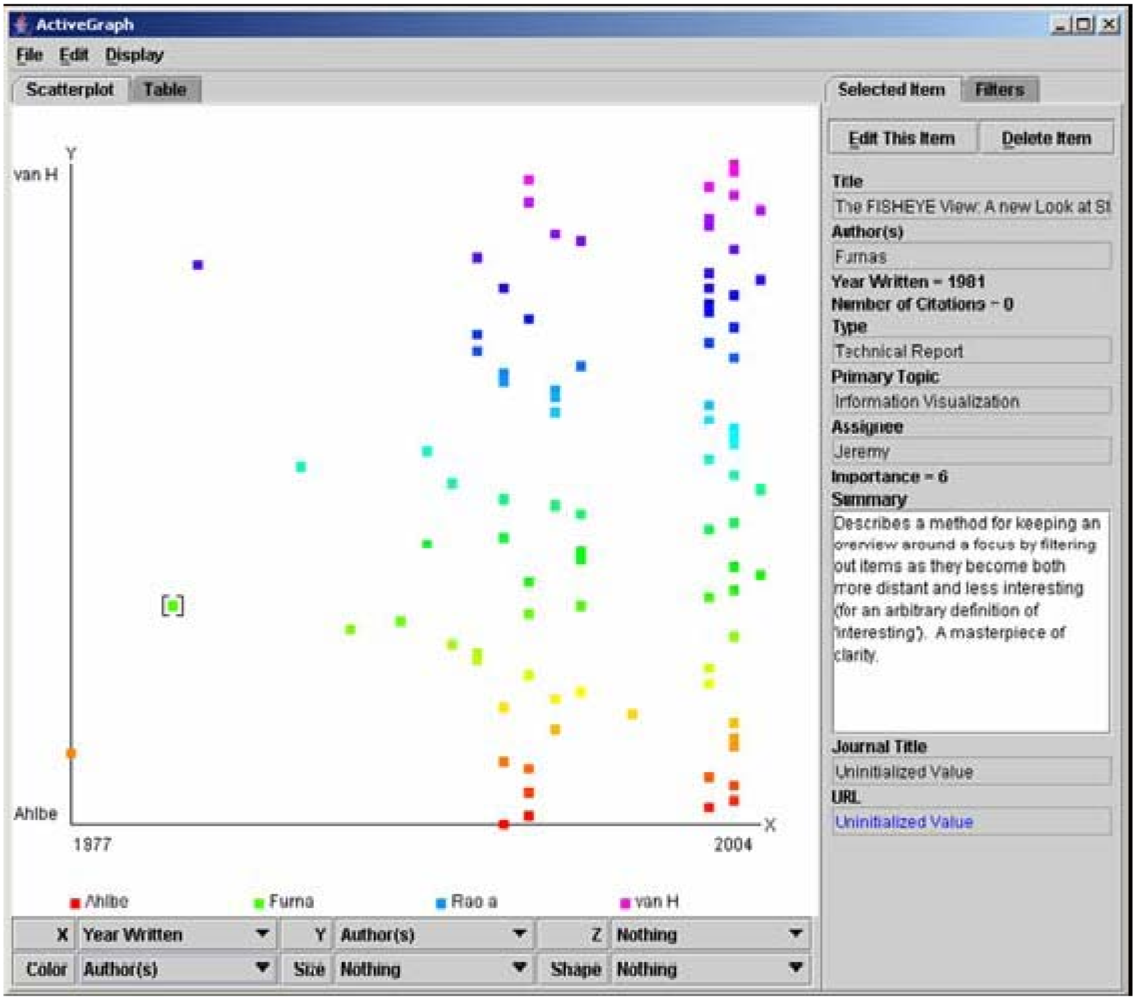

Figure 6: LibGraph visualization of the contents of a collaborative library using 2D scatter plot (adapted from [23])

Scatter plots are very effective in revealing insights into two dimensional data, spanning thousands of items. They give users a quick overview of the dataset and help them understand the distribution of data points, revealing clusters and outliers clearly. Two dimensional DL interfaces are based on scatter plots and have either continuous variables or categorical variables on the axes. The labels on the axes provide hints and valuable information to users about the distribution of records along the two variables on the axes. When categorical axes are used, the two dimensional space has to be considered as a coordinate space to map the records to the categories. Figure 6 shows a scatter plot based Lib Graph visualization of the contents of a collaborative library using the ActiveGraph visualization tool.

Based on similarities, clustering of records and displaying them in two-dimensional space is possible to represent more records, but sometimes can be confusing because of the spatial distances between clusters.

Three dimensional scatter plots are also possible. Scatter plots have six visual attributes and hence up to six dimensions can be mapped onto them. The six visual attributes of the scatter plot are X-, Y-, and Z- axes, color, size, and shape. Examples of DL visual interfaces that are based on 2D scatter plots are Envision [24], GRIDL [25], and ActiveGraph [23].

Hyperbolic trees:

Hyperbolic trees present a scalable solution to visualize and navigate through large hierarchies. A portion of the hierarchy occupies most of the display space; the context is preserved by distributing the hierarchy uniformly on a hyperbolic plane and then mapping this plan to a circular region [26]. They employ a focus + context technique called “fish eye”.

Hyperbolic trees are better than traditional tree representations in terms of the nodes they display at once, by an order of magnitude. Though they incur an initial learning time for new users, they are found to be 62.5% faster in enabling people to locate the information that they seek [27]

DL records that are arranged hierarchically can be presented well using hyperbolic trees because they are scalable and more suitable to categorical data than continuous data. The EtanaViz [28] browsing interface uses hyperbolic trees in one view. Other DL tools like the Schema Mapper [29] also use hyperbolic trees in their interfaces. Figure 7 shows a hyperbolic tree from the Schema Mapper interface.

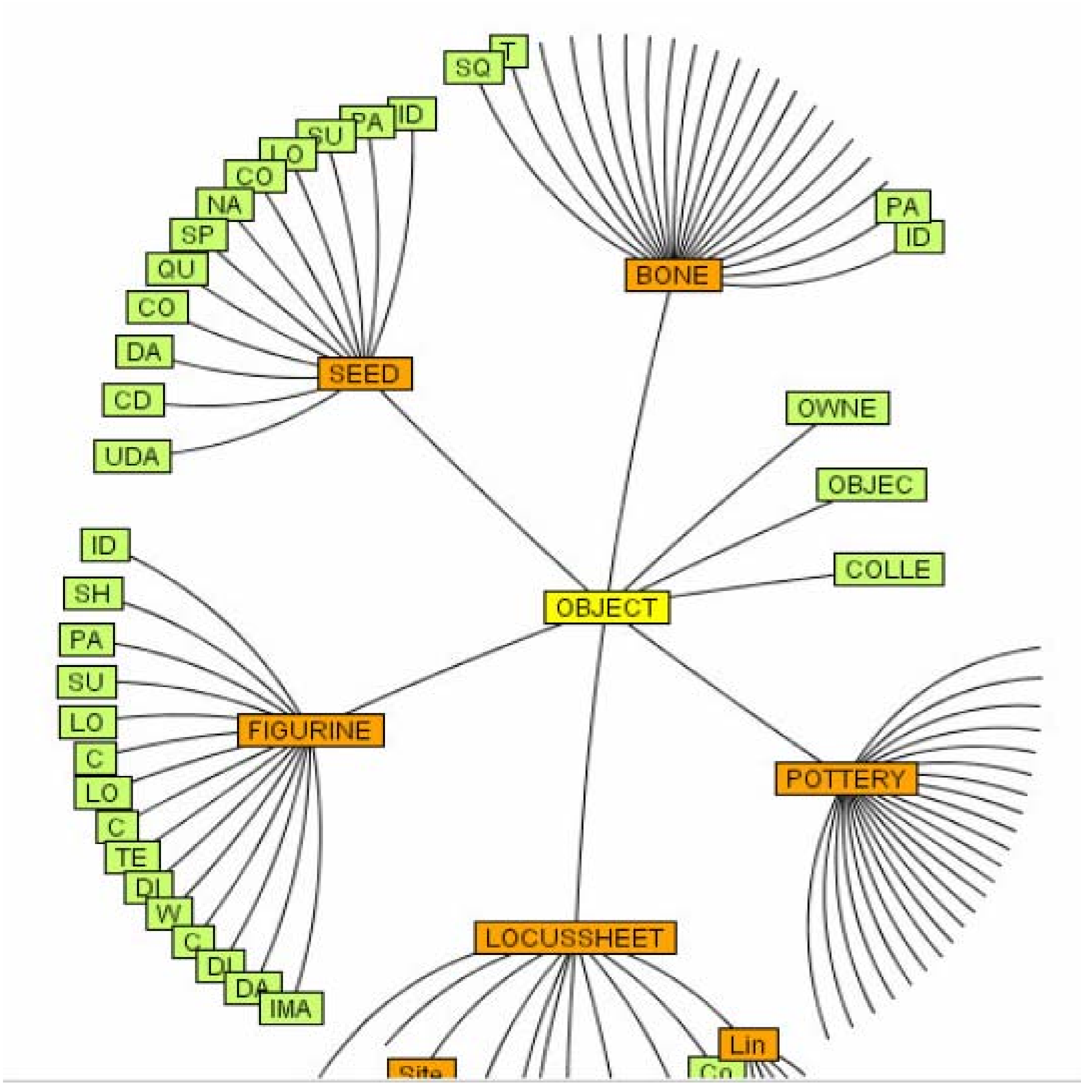

Figure 7: Hyperbolic tree from the Schema Mapper interface (adapted from [29])

Bar Charts:

Bar charts are simple visual mappings used to represent one dimensional data. The categories are represented as bars and the bar attributes like length indicate the number of records in each category.

In DL browsing interfaces, bar charts are used within multiple view browsing interfaces. They are helpful in analyzing the record distribution across categories in DL collections. Examples of DL visual interfaces using bar charts include Relational Browser ++ [30] and EtanaViz [28].

Multiple Views:

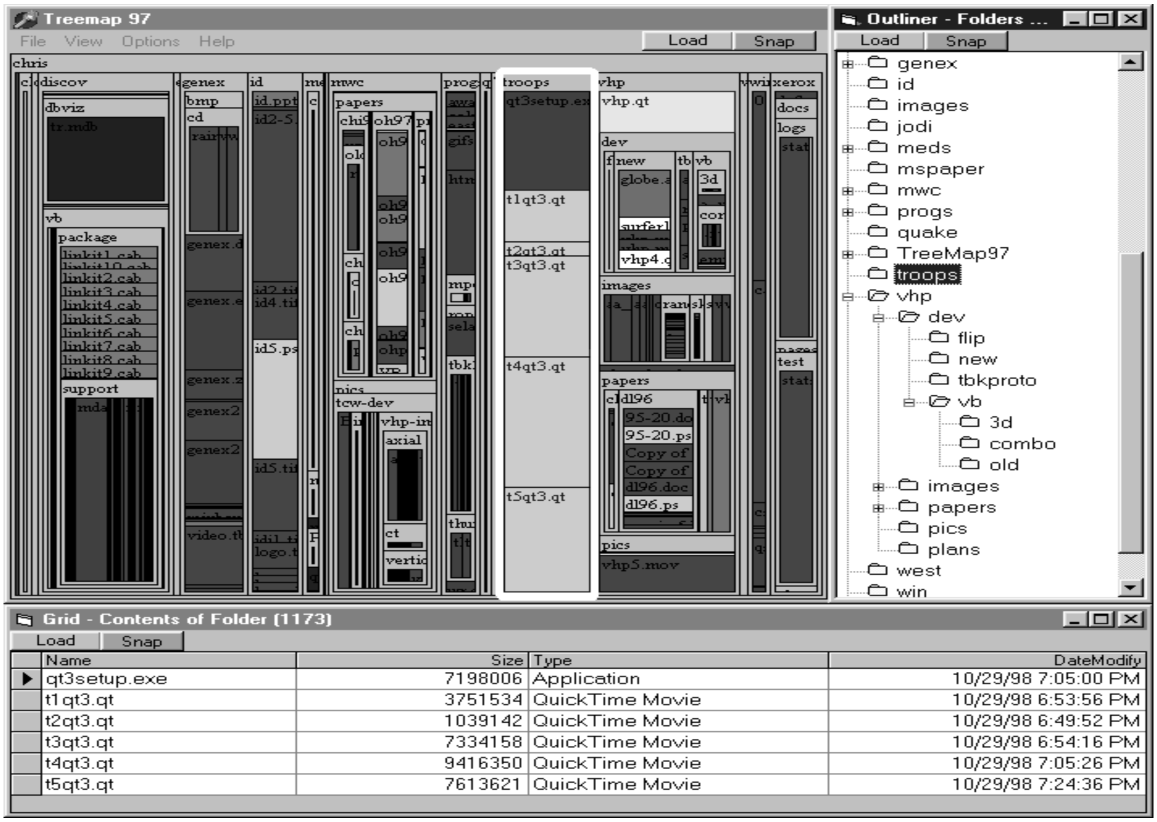

Figure 8: A coordinated multiple-view interface, created with Snap-Together Visualization, for exploring directory structures (adapted from [28])

Multiple views use two or more distinct views to project a single conceptual entity in a dataset. They are typically used to visualize multi-dimensional data. A single view describes a dataset and a visual mapping of how the dataset is represented visually. Each of the views in multiple view interfaces can employ the same or different visual mapping. Views also can differ in their data.

Coordination among multiple views is effective in designing information exploration interfaces [31]. Coordination allows tight coupling of the different views; user interaction in one window affects the visual mapping in other views. Figure 8 shows a coordinated multiple view interface, created with the Snap-Together Visualization to explore directory structures.

Meta-data in DL records are usually multi-dimensional and hence several DLs offer visual browsing interfaces based on multiple views. Examples include Relational Browser ++ [30] and Alexandria DL [19].

Outliners:

Outliners are the most common hierarchical browsing components. Outliners display the categories sequentially and are collapsible and expandable (see Figure 9). Expanding outliner categories shows lower level hierarchies; collapsing them enables navigation of higher level categories [32]. They resemble linear tree structures and are the easiest way to display the structure of data organization. Similar data values in a dataset are grouped together in hierarchical categories and then displayed through outliners.

Navigation in Windows Explorer is based on outliners. Outliners are used along with visual components like bar charts and histograms, which represent a quantitative attribute associated with each branch in the outliner tree. Hyperlink-based textual based interfaces resemble outliners in their collapsible and expandable nature. DL visual interfaces using outliners include WebTOC [33] (see Figure 9) and Greenstone [32].

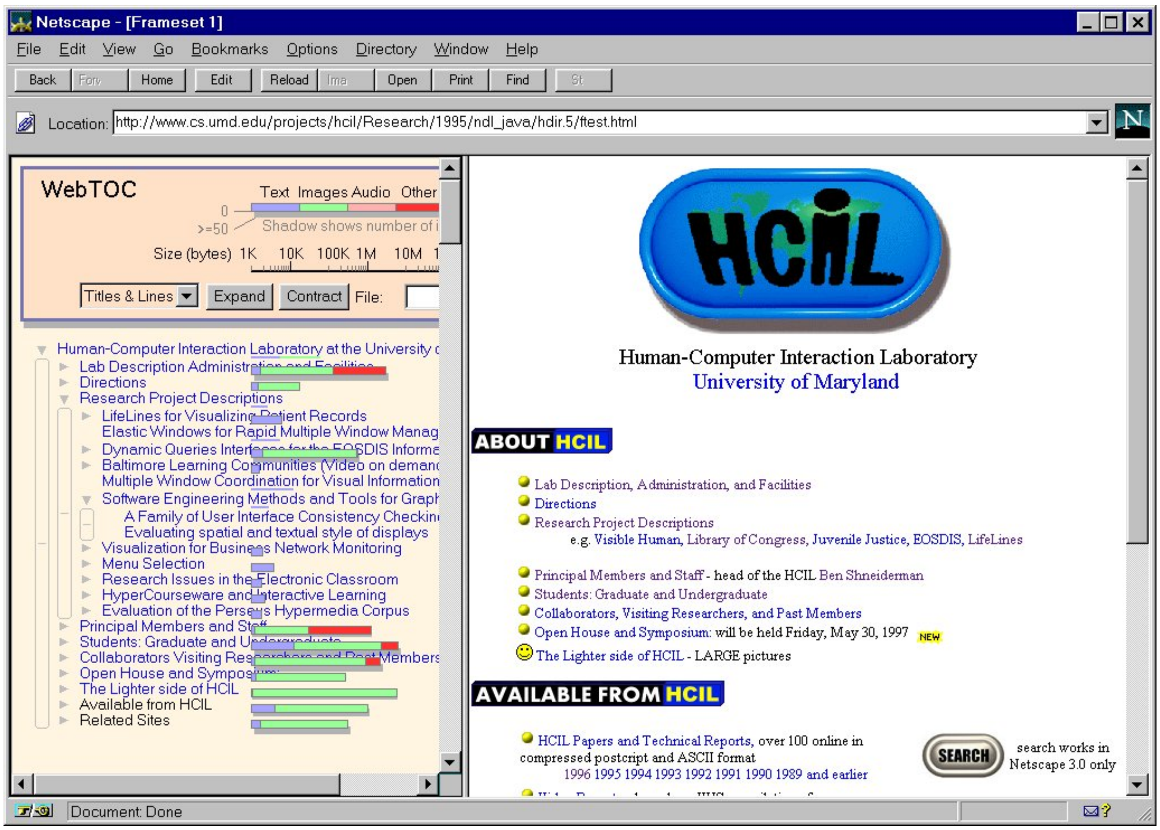

Figure 9: Outliners in the left view of the WebTOC interface (adapted from [33])

Others:

Other approaches to visualizing hierarchies include tree-maps [34], cone trees [35], and hieraxes [25]. Tree-maps provide a space filling, two dimensional space to view hierarchies (see Figure 10). Tree-maps are used to visualize document collections in the UC DL system [36]. Cone trees are hierarchies that are uniformly spread out in three dimensional spaces. The topmost hierarchical node or the root node forms the tip of the cone, and categories are placed around this tip based on their hierarchical level. Hieraxes use the outliners approach to display hierarchies on the axes of 2 dimensional scatter plot based interfaces.

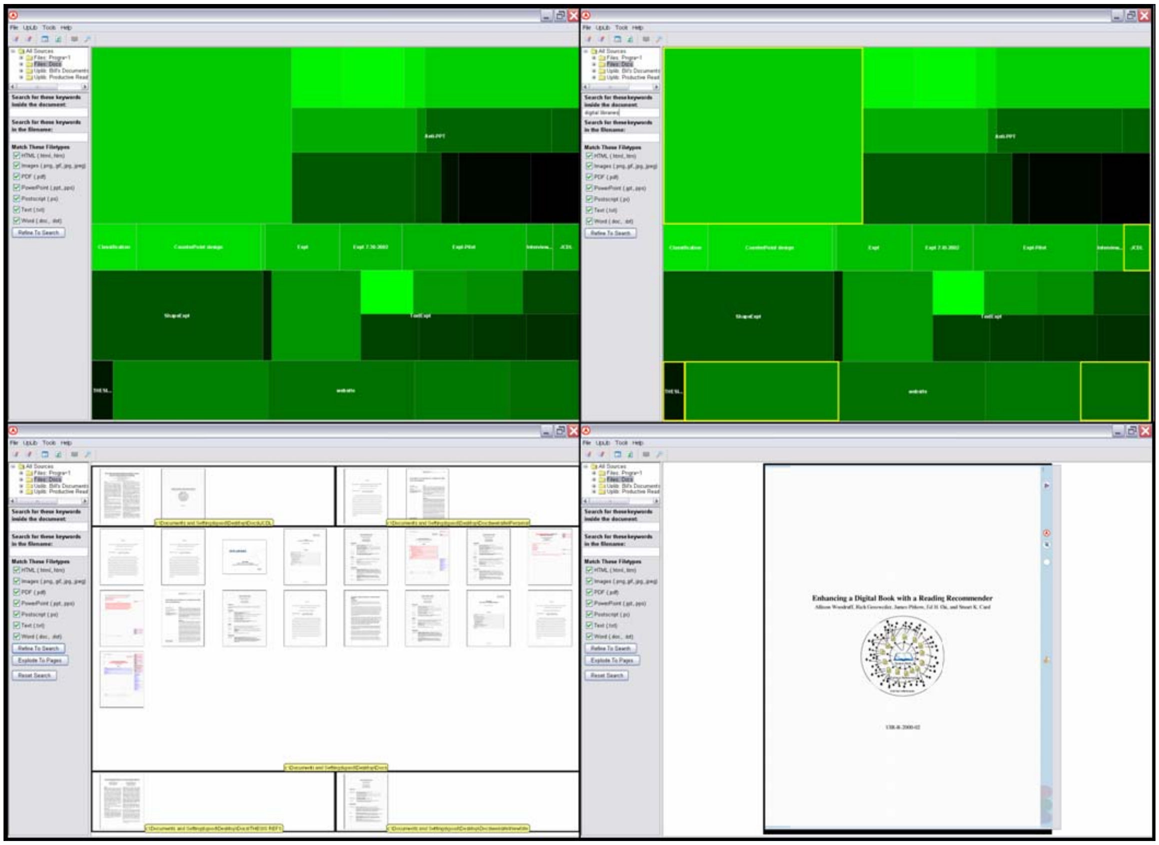

Figure 10: Treemap visualization of a document collection (top left) in the UC DL system (adapted from [36])

2.2.2. Interaction strategies

Visual interfaces allow users to get instant feedback based on their interaction with the interface. Interactions reveal relationships and more details about the collections. The strategies are oriented towards the fundamental mantra of information visualization of "Overview first, zoom and filter, details on demand" [31]. This section outlines some of the common interaction strategies used in DL interfaces.

Filtering:

Filtering is a querying mechanism where users specify a set of conditions depending on which information that matches the conditions is highlighted, or information that does not match is removed from the visualization [37]. Examples of visualization systems that use dynamic query sliders are the Dynamic Home Finder [38], where users were able to view available houses based on the upper and lower bounds they set on house prices. Sliders and buttons are usually used as filters.

Filtering is used in DL querying interfaces to show dependencies between the result set and the query terms or conditions. NCSTRL [39] is one DL that uses filtering mechanisms to help users refine their queries and reduce the number of records in the result set (see Figure 11).

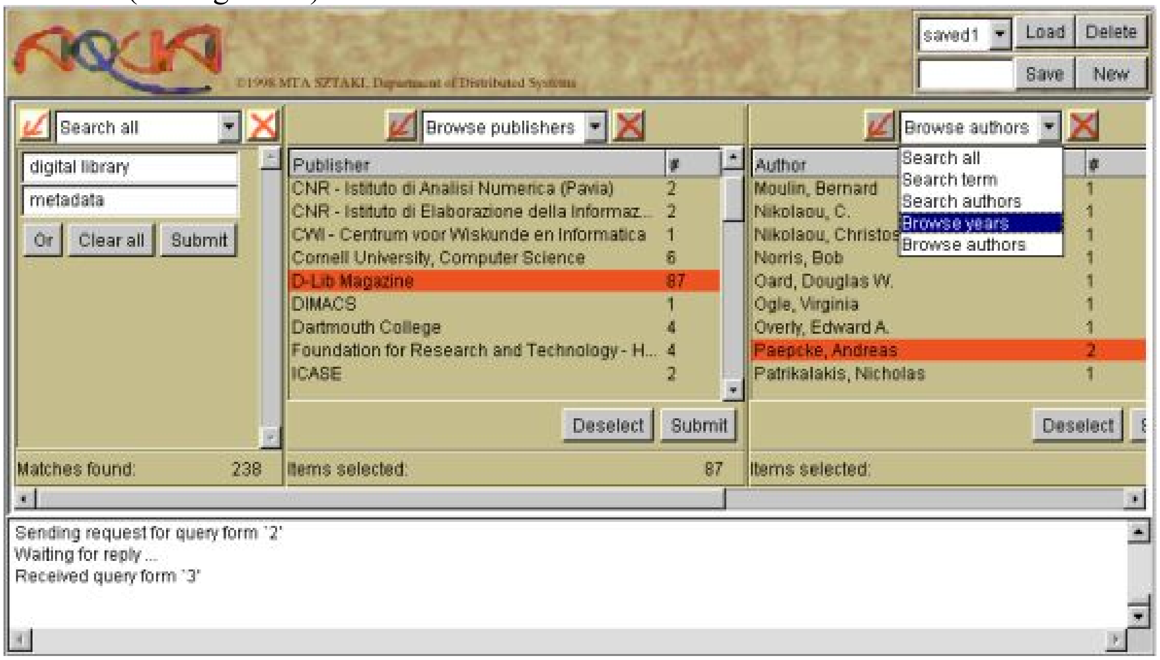

Figure 11: Filtering in the AQUA interface of the NCSTRL DL to help users refine their queries (adapted from [39])

Distortion:

Distortion is a transformation that creates focus + context views. The overview and details are shown in the same visual structure; portions of the data are shown in detail while others are shown with lesser detail. A hyperbolic tree is a visualization that uses this interaction technique [21].

Zooming:

Zooming allows users to have a higher resolution view of the data in which they are interested, by showing them details of the data points that they select. Distortion and zooming together are focus based strategies. Distortion shows more detail around a focal point whereas zooming is a uniform increase in detail which leads to lesser or no context/overview details in the same view [21].

Brushing and Linking:

This strategy is used commonly in multiple view interfaces. Selecting data points (brushing) in one view highlights the same data points in the other views (linking) to enable users to detect dependencies and correlations [21].

2.3. Case Studies

2.3.1. GRIDL

Primarily intended to visualize DL search results, the GRIDL [25] is based on two dimensional visualizations and browsers for hierarchical datasets (see Figure 12). The display is treated as a 2 dimensional grid. The axes, called hieraxes, are categorical and hierarchical. The two axes allow browsing in two dimensions. The two browsing dimensions can be chosen from drop down lists present on both the axes. Each grid point has a cluster of dots or a bar chart. The dots are color coded and the color attribute presents the third dimension of meta-data. Each dot represents a single record. If there are more than 49 items in a region of the grid, a bar chart replaces the cluster of dots.

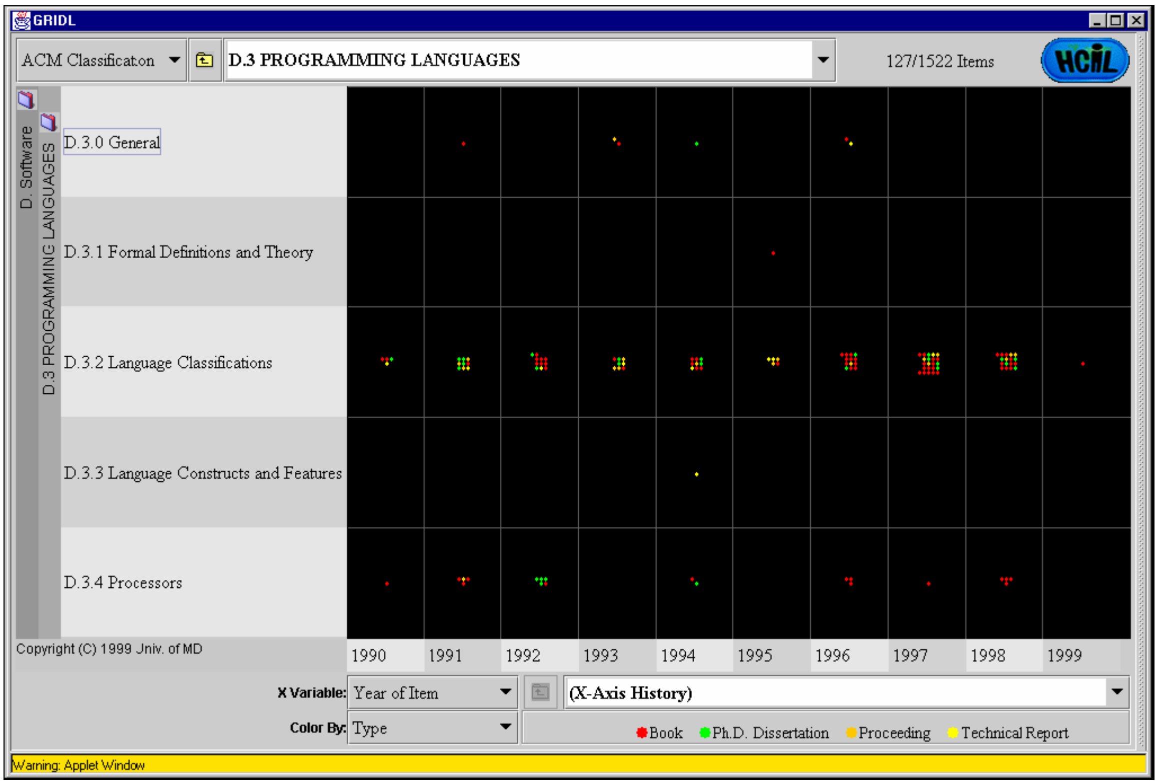

Figure 12: GRIDL interface showing documents categorized using the ACM classification system (adapted from [25])

Users can get an overview of the result set by the dots or bars arranged in a grid and organized by familiar labeled categories. Categorical zooming refers to drilling down a hierarchy for categorical variables. This is achieved by users clicking on the folder like icons to the left of an axis value. This expands the selected category and filters the display to show only the records in the category. The browsing context is maintained as an indented series of labeled folder icons. Clicking on the "Up level" icon allows users to navigate to a higher hierarchy. Hieraxes use tool tips, dynamically imposed hierarchies, paging, or scroll/zoom interaction to deal with crowded axes.

2.3.2. Relational Browser ++

RB++ [30], which is an enhanced version of the Relational Browser interface, provides a tightly coupled search and browse interface for large collections (see Figure 13). RB++ is based on uniform multiple view design, where all the different views have the same visual components. RB++ uses bars to represent categories and has multiple views, representing multiple dimensions. Each view has a set of labeled bars. The bar lengths are proportional to the number of records in the category, which are displayed to the left of the bars. The bars give a visualized category overview of the collection. By moving the mouse over a bar (brushing), users can see the conditional distribution of the records in the bar across other dimensions by proportionally highlighting bars belonging to other dimensions (linking). Users can select the dimension they want to browse through in each view by choosing from the drop down lists.

Figure 13: RB++ interface applied to EIA website (adapted from [30])

The bottom half of the interface has a table that shows users the records in the current browsing context. There are textboxes on top of each field in the results table. The string patterns in these textboxes are matched with the results in the table, and only those records matching the keyword entered are displayed. Moreover, the keyword entered is highlighted in each record.

2.3.3. Library of Congress American Memory

The Collections Browser for the Library of Congress American Memory [40] collections is also based on a multiple view design, but the different views do not have the same visual components (see Figure 14). One view, called the Collection Overview, consists of a zoomable timeline; bars represent collections and show their time coverage. Another view called the Collection Filters has three menus for location, topic, and format. The Collection List view shows the list of collections and a brief description of the collections.

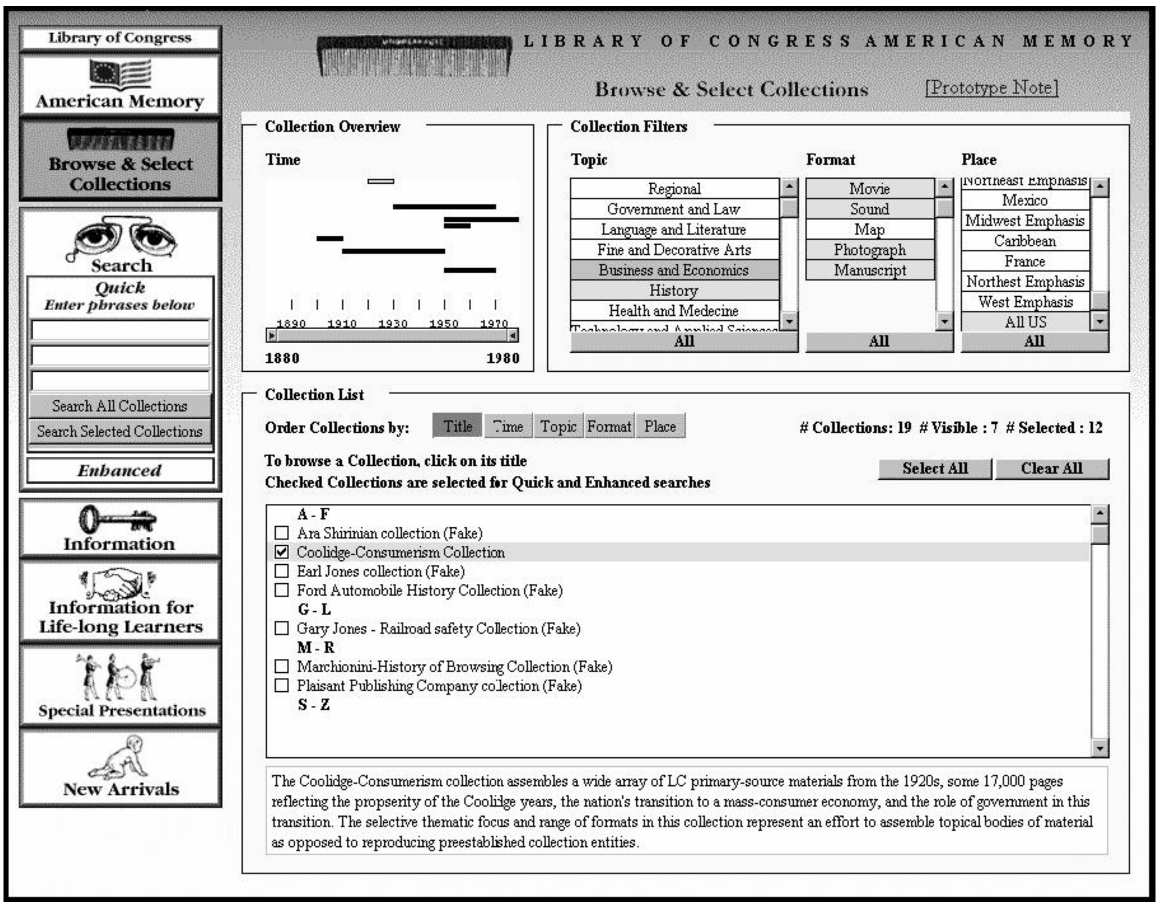

Figure 14: Overview and browsing of collections in the Library of Congress American Memory (adapted from [40])

The timeline Overview bars, menus in the Filters view, and the Checkboxes in the List view, are interactive filters. They act as dynamic query widgets. Deselecting menu items in the Filters view filters collections from the other two views. Mouse over (Brushing) of bars in the timeline view highlights (linking) collections in the other views. The collection list view displays all the collections that match the conditions in the Filters view. Clicking on any collection shows a brief description of the same. The checkboxes to the left of each collection in the list determine whether the collections are selected for quick and enhanced searches. Users also can order the collections by Title, Time, Format, Place, or Topic by clicking on the corresponding buttons in the List view.

2.3.4. CitiViz

CitiViz [41] is a visual user interface toolkit for the Computing and Information Technology Interactive Digital Educational Library (CITIDEL, http://www.citidel.org). Users can search for documents and view the result set on a visual interface (see Figure 15). It reveals hierarchical classifications of documents in the ACM classification system and also distributions of document result sets by rank and date. Users can choose other attributes for the axes from a drop down list. Based on a multiple view approach, it has three views with different visual components. A hyperbolic tree on the top right of the screen allows users to drill down the hierarchical categories in the ACM system through focus + context navigation. The size of the bubble associated with each node in the hyperbolic tree indicates the number of documents clustered under that node category. The middle right of the screen has a 2 D scatter plot, where the user can see the documents distributed by rank and date. Each document is plotted as a tower of different colored categories to which it belongs. The clustering component in CitiViz is dynamic; towers in the scatter plot view are displayed such that there is minimal occlusion. The left view shows a legend of the different colors used to represent different categories. This also serves as a filter to remove the categories in which the user is not interested. The color of the categories is synchronized in the hyperbolic tree and the 2D scatter plots. The view on the bottom shows the detailed information about a document that is selected from the 2D scatter plot view. Brushing and linking is the prime interaction strategy used to coordinate between the views.

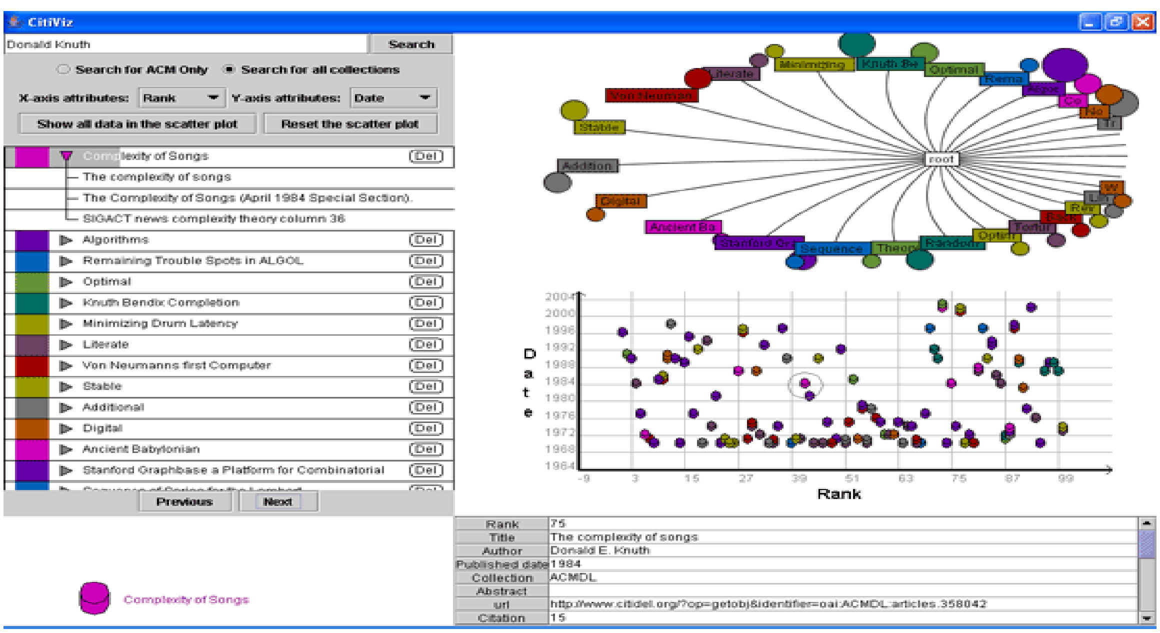

Figure 15: CitiViz showing retrieved results from ACM DL (adapted from [41])

CitiViz also helps users analyze document collections. Users can see all papers cited by a particular paper. When a user clicks on a document tower in the 2D scatter plot, links to the cited papers are displayed.

2.3.5. Generalized Query Previews

Generalized query previews [42] is a user interface architecture to facilitate efficient browsing of large online data, that is multidimensional and present across multiple tables (see Figure 16). Users accessing this huge data collection are shown data distribution information as an overview of the data. This overview is essential in helping users refine their query to obtain the desired dataset. Users receive continuous feedback about the size of the result set as the query is being formatted. This serves as a preview to the result set.

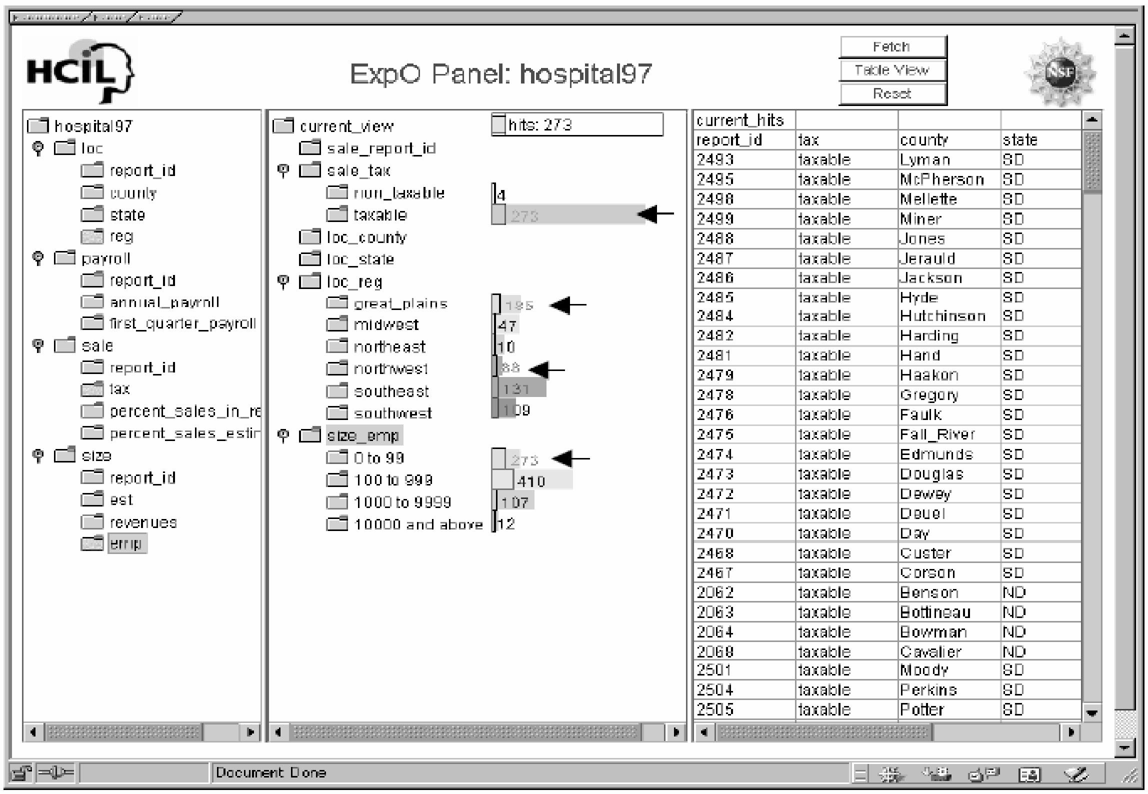

Figure 16: Query Previews in EXPO (adapted from [42])

The generalized query preview interface has been applied to a number of systems, one example being EXPO. This system is based on a huge dataset from the 1997 United States Economic Census Collections. The left panel shows the four tables in the database. The four tables and their attributes are presented using a hierarchical outliner representation. The root of the tree is the name of the database, the first level of hierarchy displays the tables in the database and the second level, the attributes in each table. Users can select some of the attributes to form a user view shown in the middle panel. Each attribute contains buckets of possible values for the attribute and a bar attached to each value. The lengths of the bars show the distribution of data over this attribute. Users can click on a bar to select/deselect the values represented by the bar. A separate bar, called the result bar on the top of the middle panel, shows the total number of items matching the selections/query specified by the user. After refining their queries based on the distribution and result set size, users can finally view the result set in the right side panel.

This interface architecture can also be applied to DL browsing interfaces to help users narrow down the result set.

2.4. Summary

With 2D scatter plot layouts, users are limited to browsing in two dimensions at any given instant though they can choose the dimensions on the axes. This affects the scalability of the interfaces to the number of dimensions, the flexibility, and the uniformity of browsing. However, they are well suited for analysis purposes along with bar charts. Hierarchical browsing is made possible through categorical expansion of axes, as in hieraxes, but usability studies have revealed that it is hard for users to maintain their navigation location while using hieraxes.

Multiple views, based on different components in different views, is the most common design used in DL visual interfaces. The components in the views are chosen so that the visual mapping correlates to the user's mental model of the dimension. For example, a view that facilitates browsing through the time dimension might use a component like timeline bars. A dimension that is hierarchical might be mapped to a hyperbolic tree or an outliners view. A pair of flat dimensions with no hierarchies is best viewed on a 2D scatter-plot. Brushing and linking and filtering are the primary interaction strategies.

The RB++ interface employs uniform multiple view layouts and is scalable to more dimensions; the limiting factor is the number of bars of any dimension on the screen. It allows flexible browsing across dimensions but does not deal with hierarchical categories.

Though multiple view design is commonly adapted across DL visual interfaces, they do have some drawbacks. Users require time and effort to learn multiple view systems. There is load on the user's working memory and effort is required for comparison and context switching between views. They also have system limitations. The display space limits the number of possible views, or the visual components used in the views. The computational requirements are enormous when compared to single view interfaces [43].

With DL records typically organized into hierarchical categories based on multiple attributes, much memory space is required to store the inter-dependencies between categories in different dimensions. Integration of services (searching and analysis) to the visual browsing interface increases the computational overhead. Limitations in the meta-data have to be considered. Use of advanced clustering techniques, and data mining techniques, is strongly recommended while designing and implementing visual interfaces to DLs.

In general the services provided by the visual interface, and interface design, are strongly dependent on the primary motives behind the system. Some goals need to be traded off for better benefits in the primary area of concern for that particular system. An overview of the various approaches and popular DL interfaces should be given thought when designing a new interface for a DL.

3. DESIGN OF ETANA-CMV

First, we describe three principles used in the design of the interface for ETANA-CMV. Then, the final design of ETANA-CMV is described. This is followed by a description of major design alternatives considered during the iterations leading to the final design.

3.1. Design Principles

The following three principles are derived from our survey of the literature on visual browsing interfaces to DLs.

3.1.1. Browsing structure, not individual records, is the primary data

There exists a tradeoff between being able to view more of the individual records and being able to view more of the browsing structure. Displaying each record as a separate visual mark leads to occlusion and less space for the underlying browsing structure. If the labeling of the category names is cut short due to limited space, navigation through the browsing structure becomes harder.

The first design principle is to display the categories in each dimension and present the records in each category as a cluster rather than attempt to display all individual records on the browsing interface.

3.1.2. Display all possible browsable dimensions

Textual browsing systems, like Flamenco [11] and ETANA Browse Interface, show users all the dimensions through which they can browse. The categories in each dimension are listed below the dimension name, usually along with the count of records in the category. When the number of dimensions or categories in a dimension is high, scroll bars are used. The objective is to give users an overview of all possible dimensions for browsing. Visual interfaces like GRIDL [25] and CitiViz [41] display a set number of dimensions that is restricted by the visualization component used. 2D scatter plots allow navigation through the two dimensions on the axes. Users are forced to choose the two dimensions they want to browse through. RB++ [30] provides flexible browsing, by having a separate view for each dimension.

The second design principle is to display all possible dimensions for browsing to enable flexible browsing across dimensions. Display size rather than the visualization component used should be the limiting factor for users to view the underlying browsing structure.

3.1.3. Treat dimensions uniformly

Textual browsing systems treat all dimensions uniformly; they are all based on hyperlink based browsing. With visualizations, it is common to find different visual components used to map different browsing dimensions to the display space. This helps particularly in analyzing collections. For example a set of document collections may be best viewed in a 2 D scatter plot distributed by rank and date. Depending on the domain and nature of the data, components can be chosen so as to maximize the number of insights into the collection. However, this leads to more complexity and learning time from the users who might be visiting the DLs rarely or sometimes only once. There is a tradeoff between the capabilities and goals intended of the interface, versus its complexity and ease of use.

The third design principle is to treat all dimensions uniformly to reduce complexity and enhance learnability of the interface.

3.2. Description of final design

Our design of ETANA-CMV is illustrated in Figure 17, and described in the following sub-sections.

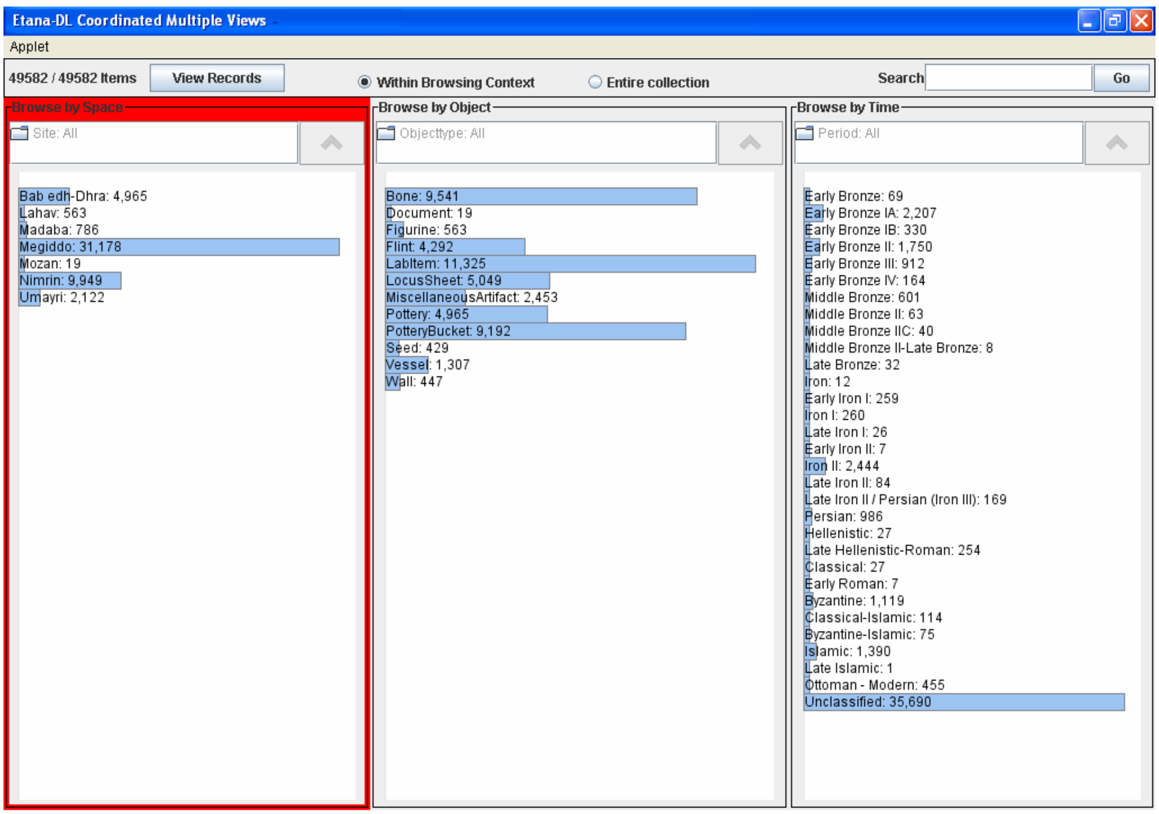

Figure 17: ETANA-CMV at startup

3.2.1. Panels

There are three panels, one for each dimension. Each panel is named after the dimension it represents. The three panels are labeled “Browse by Space”, “Browse by Object”, and “Browse by Time”. Each panel has two sections; the bottom section shows a number of bars and the upper section shows a tree-view. The three panels have the same visual components (the bar chart and the tree-view) and support the same interaction techniques.

Moving the mouse over a panel causes that panel to be highlighted, indicating the current browsing dimension.

3.2.2. Bar chart

The bars in each panel show the categories in each dimension. The lengths of the bars indicate the number of records in each of these categories. The bars are labeled with the category name and the number of records in the category. The panels are scrollable, depending on the number of bars (categories) in each (panel) dimension. The lengths of the bars are scaled relative to each other. A category with 1000 records would have a bar that is one tenth the length of a bar that denotes a category with 10000 records. However, the "Late Islamic" category in the Time dimension has 1 record while 35690 records belong to the "Unclassified" category, so relative scaling would make the "Late Islamic" bar too small to be observed. To avoid this problem, bars with very small relative lengths are assigned a fixed minimum length.

3.2.3. Mouse Move-over bar chart: Brushing and linking

Brushing and linking is used to coordinate the different panels (see Figure 18). Moving the mouse (brushing) over a bar highlights it and also pops up a tool-tip with the same label on the bar, just for clarity. It also highlights other bars in other panels (linking). For example, moving the mouse over the bar representing "Seed" category under the "Object" dimension causes the "Persian" bar to highlight under the "Time" dimension. This is because records of the seed object type are found in the Persian era.

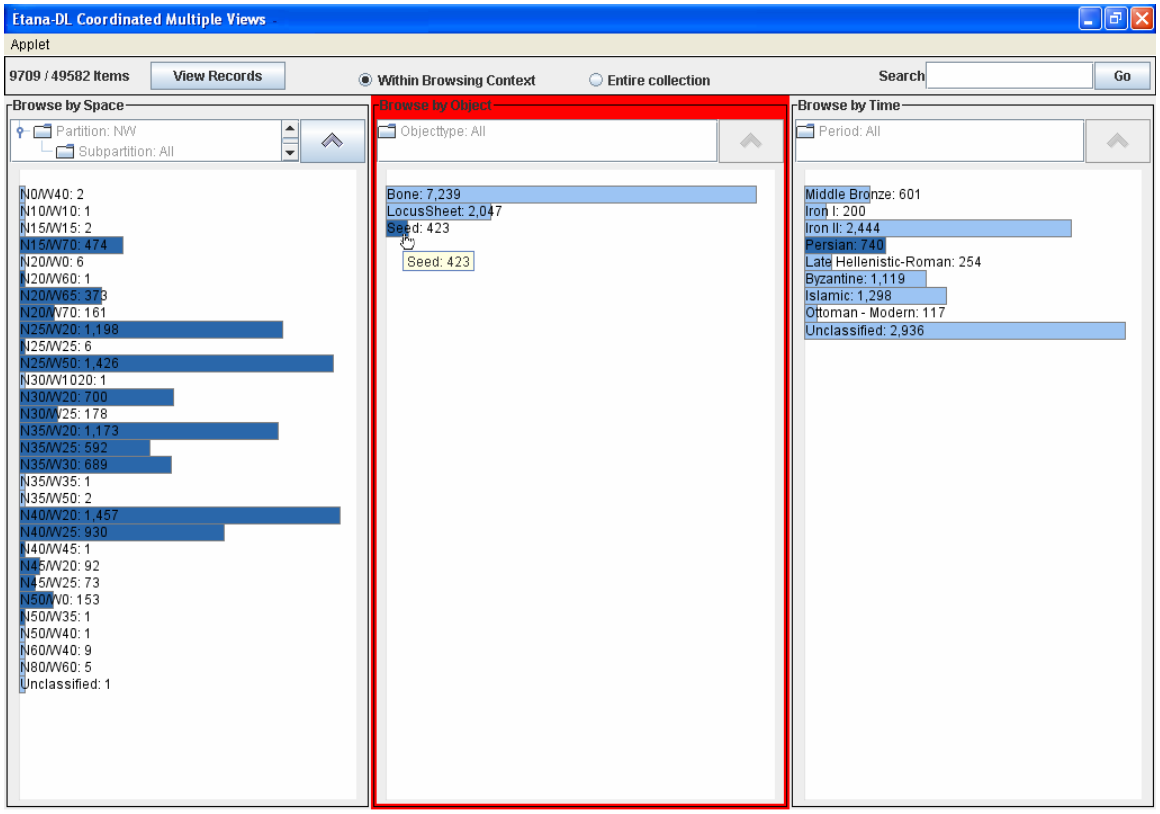

Figure 18: Brushing and linking in ETANA-CMV

When the number of categories in a panel is high, the panel becomes scrollable. In such cases, the linking effect might not be obvious, because a bar towards the bottom of the list might be highlighted and this bar will not be visible without scrolling in that panel. To reveal the linking, a portion of the bar chart in each panel is displayed such that at least one bar is highlighted during mouse move events.

3.2.4. Bar ordering

The bars (categories) are ordered in alphabetical order in the "Browse by Space" panel and "Browse by Object" panel, and chronologically in the "Browse by Time" panel. The ordering can be chosen by the user. The mouse right click event on a panel pop-ups the menu shown in Figure 19. The bars can be sorted in alphabetical order or based on the number of records they denote. In the "Browse by Time" panel, bars can be sorted chronologically based on the ordering provided by the Browsing Index database.

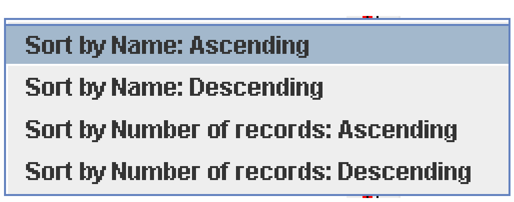

Figure 19: Pop-up menu in ETANA-CMV for ordering the categories

3.2.5. Hierarchical navigation: Drill down and rollup

The drill down approach found in hypertext-based textual interfaces is followed. Clicking on a bar drills down into the hierarchy. For example, clicking on the bar representing Nimrin in the space dimension displays a new set of bars representing the partitions inside the site Nimrin. It also dynamically filters out the categories in other dimensions that are not associated with records from the site Nimrin. The records from Nimrin are Bone, Locus-sheet, and Seed. The “Browse by Object” panel is refreshed with only these bars, with the counts of records in each of these categories within the site Nimrin. For the same reasons, the “Browse by Time” panel is also refreshed with a fewer set of bars.

Users can navigate to a higher hierarchical level by clicking on the “Up” button found above the bar-chart area to the right of the tree-view. Clicking on this button after the above mentioned operation of drilling down into Nimrin, will display the original screen where all bars are shown and the other two panels are also updated with the original set of bars.

3.2.6. Tree-view for Browse History: Maintaining context

The tree-view (see Figure 20) in each panel shows the browse history within the dimension. It shows the browsing context to the user. The tree-view in the “Browse by Space” panel in the initial screen shows ‘Site:All’, meaning that the bar chart in the same panel shows all categories at the Site level. When a particular site, say Nimrin, is part of the browsing context, it would show ‘Site: Nimrin, Partition: All’, meaning that the bar chart in the same panel shows all partitions belonging to the Nimrin site. The tree-view panel becomes scrollable when displaying more than two levels of hierarchy.

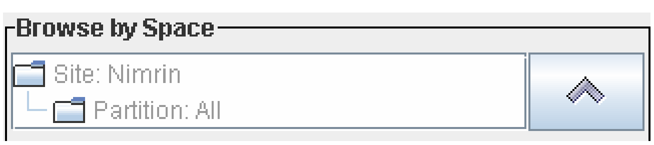

Figure 20: The tree-view in the “Browse by Space” panel showing the browsing context Site: Nimrin

3.2.7. Number of records in current browsing context

The count on the top left label shows the number of records within the current browsing context. In Figure 18, the label reads 9709/49582 items, meaning that there are 9709 records in the current browsing context which is ‘Site: Nimrin and Partition: NW’. 49582 represent the total number of records in ETANA-DL. In the initial screen (see Figure 17), the label shows 49582/49582 items.

3.2.8. Integration with other services in ETANA

Users can view the records within the current browsing context, or the entire set of records, by choosing the appropriate radio button of the two shown in the centre of the top panel and then clicking the "View Records" button to the left of the radio buttons. This would pop up the interface in ETANA that shows the records as a series of pages with ten records on each page.

Similarly, users can search for content in records through the "Search" textbox to the right of the radio buttons: entering a query and hitting go. The search can be limited to the current browsing context or to the entire records in ETANA-DL. The search result set is displayed by the Search interface provided by ETANA-DL as a series of pages with ten records on each page.

3.3. Design iterations

To arrive at the final design of ETANA-CMV, three design iterations were involved.

3.3.1. First Iteration Design: EtanaViz

A first design was made by Vidhya Vijayaraghavan and implemented as EtanaViz [28]. A study into her work is described as the first iteration to this work. Figure 21 below shows a snapshot of EtanaViz.

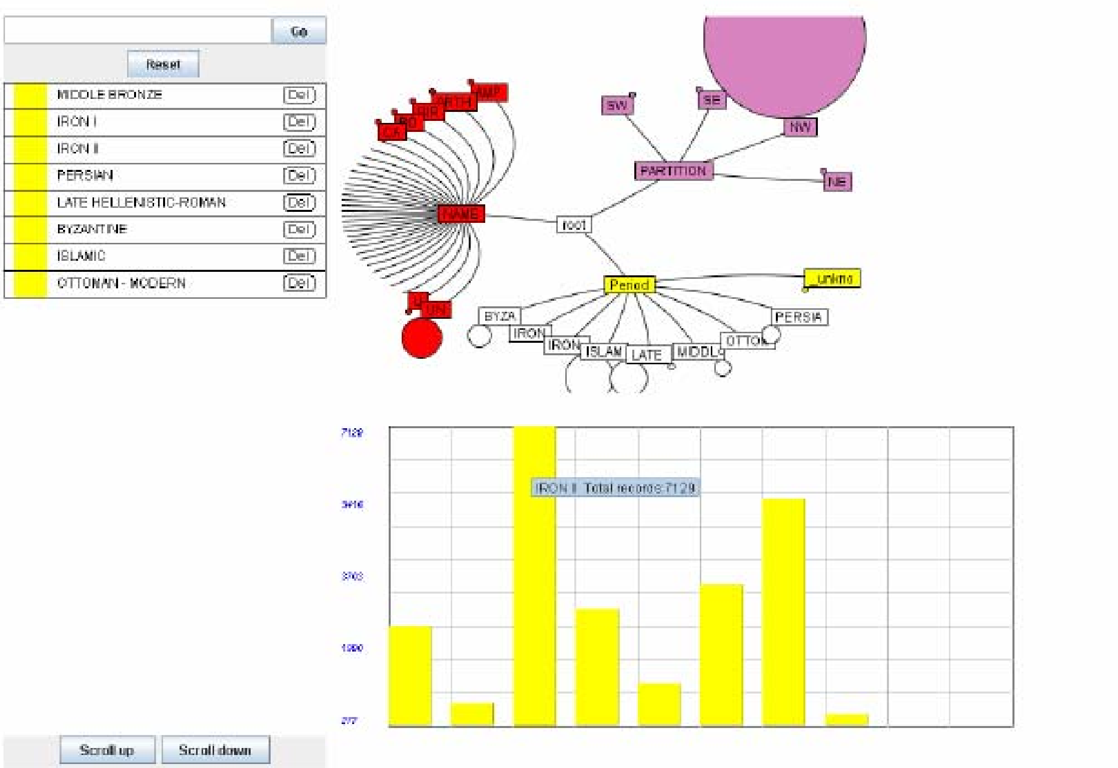

Figure 21: EtanaViz interface after adding categories from the time period dimension to the bar chart (adapted from [28])

The EtanaViz interface is a visual representation of the underlying browsing structure provided by the ETANA Browse service at a given hierarchical context. This instance was intended to help analyze bone records in the site Nimrin. Though the interface does not allow hierarchical navigation across the entire browsing structure, it can be loaded with the categorical values from the three dimensions at any hierarchical level. This design included the following features.

- The hyperbolic tree on the top right portion of the screen displays the categories in the three dimensions. The three branches to the root node denote the three dimensions. In the example, the Space (partition) branch has children denoting the partitions in Site Nimrin. The Object (Name) branch has children denoting the different bone names. The Time (Period) branch has children denoting the different time periods to which the bones in Nimrin belong.
- The size of the bubble attached to each node denotes the number of records associated with that node. This gives a quick overview of the distribution of records in the various categories in each dimension.
- Nodes from a branch of the hyperbolic tree can be right clicked upon and added to the bar chart below for comparison.
- Nodes from another branch can be right clicked and added to the bar chart to make it a stacked bar chart that shows the distribution.
- The color coding on the left helps identify the different categories in the stacked bar chart. The colors can be selected by the user.

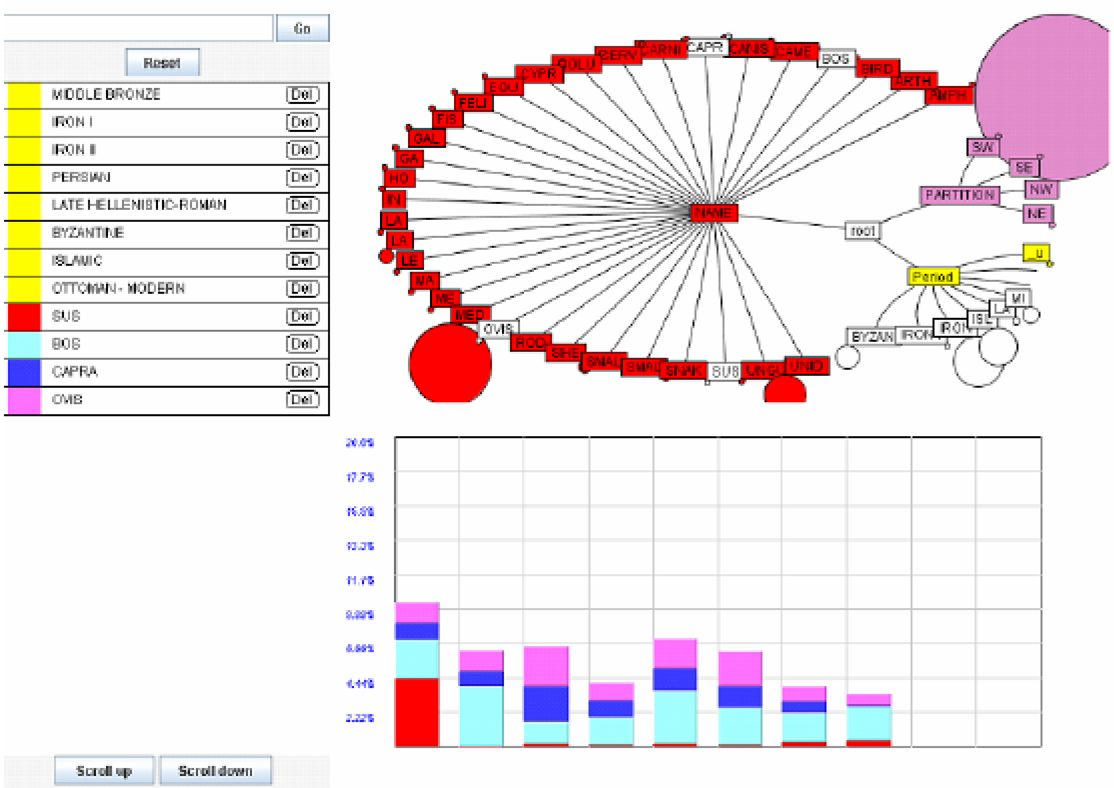

Figure 22: Stacked bar chart in EtanaViz showing percentages of Animal Bones across Nimrin (adapted from [28])

Table 1 lists pros and cons for each of these features. We use a claim analysis [44].

Table 1: Claims for design iteration 1

Design feature	Possible pros and cons
Hyperbolic tree on the top right with three branches for the three dimensions	+Scalable in terms of dimensions; since each dimension is added as a branch to the root node +No scrolling is required -but needs to be reoriented to visualize some regions of the tree -but has been known to have a significant learning curve for users to get used to the interaction provided by it
Bubbles attached to the hyperbolic tree nodes	+Gives overview of all categories and distribution of records in categories in each dimension -but takes up more space that could have been used to label the nodes; the labels are not complete -but the absence of label/tool-tip denoting the number of records represented by the bubble reduces its utility
Bar chart	+Provides an efficient way to compare categories with well labeled y-axis showing number of records and tool-tip showing category name and number of records in it -but bars could have been labeled better on the x-axis with the category name -but y-axis scale is not normalized
Stacked bar chart	+Displays distribution of records from categories in one dimension onto categories in another dimension +Use of percentages helps analyze distributions better -but bars could have been labeled better on the x-axis with the category name -but y-axis scale is not normalized
Color coding to differentiate categories	+Powerful way to visually differentiate categories -Not scalable; particularly when number of categories is around 20 at a given hierarchical level

A formative usability study was conducted on the entire ETANA-DL prototype by Rao Shen. Subjects were shown a video demonstration of EtanaViz and were asked to subjectively rate the interface design on a scale of 1 (Poor) to 5 (Excellent); the average score was 4.08 when compared to 3.60 for the ETANA Browse Interface. The median score was 4 in both interfaces.

Comments from subjects revealed the usefulness of visualizations for archeological data. However, users had mentioned that the hyperbolic tree could be confusing to use and a little difficult to navigate. They also wanted improvements in the labeling of bars in the bar chart.

The EtanaViz interface was successful in presenting a quick overview of the bone records in Nimrin. The interface is scalable with regard to the number of records, mainly because it is based on visualizing the browsing structure rather than the individual records. It is based on the first design principle. The design needed capabilities for hierarchical navigation to make it a complete visual browsing interface for the entire set of records in the DL. Though the hyperbolic tree presents the categories from the three dimensions, only two of them can be chosen at any stage for viewing on the stacked bar chart. Users could not view the distribution of records across categories in all dimensions at once. It thus lacked the flexibility provided by the ETANA Browse Interface.

3.3.2. Second design iteration

In this iteration, to provide for flexible browsing across all dimensions, the hyperbolic tree was replaced with separate views for each dimension. Figure 22 shows a snapshot of the new design prototype.

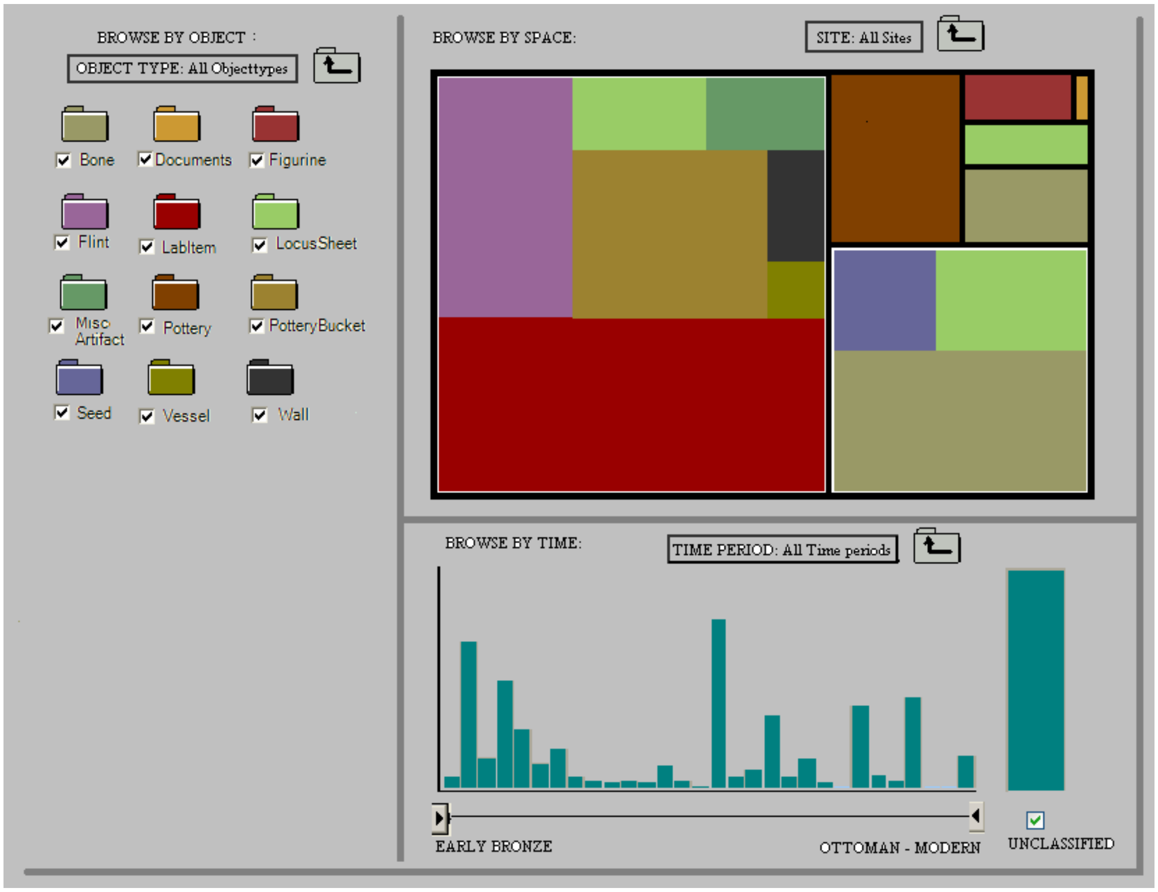

Figure 23: Initial display of the second design prototype with three views for browsing through the three dimensions in ETANA-DL

Three views are present, one for each dimension. The tree-map on the top right displays the categories in the space dimension, the view on the left with the folders displays the categories in the Object dimension, and the bar chart view in the bottom displays the categories in the time dimension. This design included the following features.

- A tree-map which is a space filling visualization is used to denote categories in the space dimension. Each block enclosed by bold black lines represents a category in the space dimension, initially a site. On clicking inside a block, the tree-map is reloaded with a new set of blocks that denote the partitions inside the clicked site. Now, the entire tree-map represents the clicked site. This facilitates hierarchical drill-down and the context is maintained in the label on the top of the tree-view. Users can click on the up arrow button to move up the hierarchy. The size of the blocks is proportional to the number of records inside the category denoted by the block.
- Colors are used to represent the different object types in the DL. These colors, on the folders view on the left, match the colors of rectangles inside the tree-map. For example, if red color denotes lab items, the presence of a red rectangle inside the largest block in the tree-map, indicates that lab items are present inside this site. Here again, the size of the colored rectangles is proportional to the number

of records of that object type present in the space category. The folders too are clickable to drill down into the hierarchy in the object dimension.

- The categories in the time dimension are represented as bars in a bar chart. The height of these vertical bars denotes the number of records in each time period. Upon clicking a bar, a new set of bars is loaded to denote drilling down into the hierarchy in the time dimension. The dynamic query slider to the bottom of the bars can be moved to reveal the distribution of records from various time periods in the different categories in the space dimension. The order of bars is based on chronological ordering of time periods. A separate bar called "Unclassified" is used to denote all records that are not classified into a distinct time period.
- Coordination between views differs between each pair of views. There is no direct coordination between the object view and the time view. The coordination between the time view and space view is based on filtering. On moving the dynamic query slider such that certain time periods are deselected in the time view, certain portions of the tree-map get whitened (filtered), meaning that these categories in the space dimension had records belonging to the time periods that were deselected. The space view and the object view are linked only by the colors.

Table 2 lists pros and cons for each of these features.

Table 2: Claims for design iteration 2

Design feature	Possible pros and cons
Tree-map for viewing space categories	+The space filling approach of tree-maps correlates to the spatial distribution of records across various spatial categories. -but it is a complex visualization with significant initial learning time -but labeling of blocks inside a tree-map is tricky
Color coding for object categories	+Powerful way to visually differentiate categories -but not scalable; particularly when number of categories is around 20 at a given hierarchical level
Bar charts and dynamic query slider for time category	+The ordering of bars and the filtering mechanism provided by the slider suit tasks related to temporal visualizations -but the number of bars is not scalable when used along with a slider
Specific coordination for each pair of views	+provides interaction suited to the data -but difficult for users to understand -but inconsistent and difficult to extend beyond three dimensions

Upon discussion of this interface with the ETANA focus group, the complexity of the interface and its interaction mechanisms were identified. The ETANA focus group consisted of co-PIs who have been involved with the ETANA research project since 2003 and graduate students who developed the ETANA-DL prototype. An argument was made that it would be difficult to use, for archaeologists, visitors to the DL, and users with little exposure to visualizations. It was thus decided that this design not be implemented. However, this design was a major step from EtanaViz in that it aimed to provide users with the ability to switch dimensions while browsing. The use of separate views could ensure the ability for flexible browsing. This is the second design principle. Drilling down into the hierarchy was provided by clicking on the visual mapping of the categories; rolling up to a higher hierarchy was provided by the "Up" button.

3.3.3. Final design iteration and rationale

The final iteration integrated the third design principle into the interface: treating the dimensions uniformly using uniform views. Thus the three design principles described in section 3.1 have been used in the final design called ETANA-CMV. The rationale behind the three different principles is that they provide scalability, flexibility, and learnability to ETANA-CMV, the new visual browsing interface. These three features are integral to the existing ETANA Browse Interface as seen in section 1.2.4.

In ETANA-CMV, the power of visualization and coordination should provide much more insight into the collections in the DL than the existing textual browsing interface without limiting the degree of scalability, flexibility, and learnability. The following table describes how the design elements that differ in the final design from the previous iterations contribute to the design rationale.

Table 3: Design goals of ETANA-CMV and design features that contribute to the goals

Design feature	Goal
Thee separate panels for the three dimensions	Flexibility
Use of bar chart, and tree-view in each panel instead of different visualization components	Learnability
Simple brushing and linking across all Views	Learnability
Displaying the categories instead of records	Scalability
Scrollable bar chart area and tree-view area	Scalability
Absence of colors for categories	Scalability
Automatic scrolling to reveal linking in a long list of categories	Scalability

A detailed description of the design has already been given in section 3.2.

3.4. Summary

Three design principles were derived from the survey on visual browsing interfaces to DLs. They were applied in the design of ETANA-CMV. The third version seemed to address all goals and objectives and was fully implemented, following the architecture described in the next chapter, so it would be efficient as well as effective to use.

4. ARCHITECTURE

In this chapter, we describe the overall architecture of ETANA-CMV. The major concepts that make up each component of this architecture are presented along with the implementation details. This chapter also presents the issues confronted in the software development process.

4.1. ETANA-CMV architecture

Figure 24 below gives a conceptual overview of the architecture of ETANA-CMV. The functionality of each component is described below the figure.

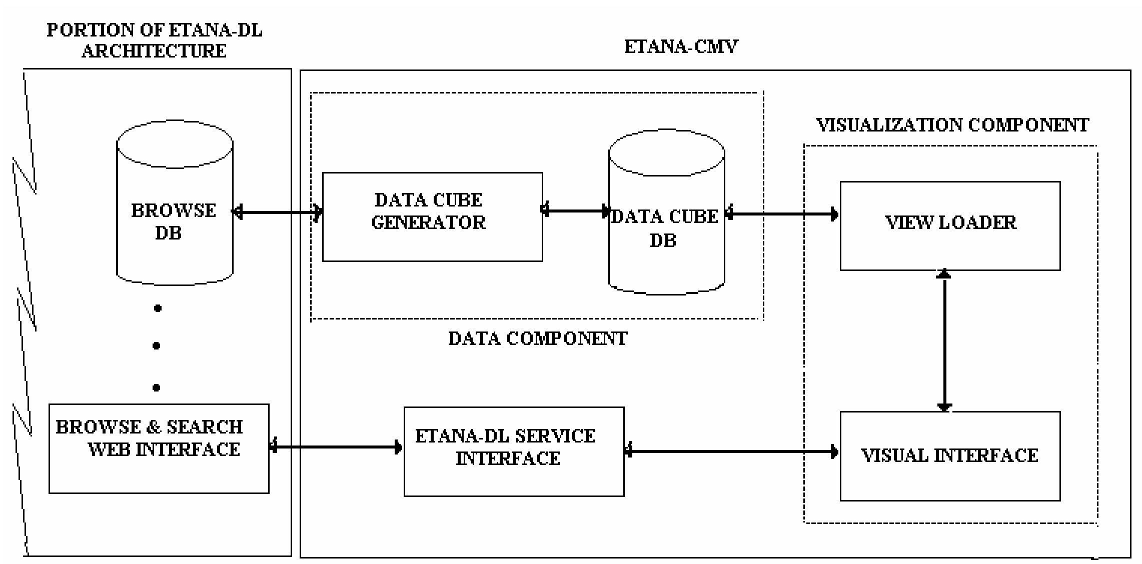

Figure 24: ETANA-CMV Architecture

4.2. Data component

This section outlines the new visual browsing index for ETANA-CMV, and the data cube that forms this new index.

4.2.1. Existing browsing index

The “etanamultidibrowse” database described in section 1.2.1 is the browsing index for the ETANA Browse Service, and the textual ETANA Browse Interface is based on this service. As stated earlier, the idtable is the main table, that has an entry for each record in the DL and maps the record described by its unique OAI Id to the location in the browsing structure by associating it with a *spaceid*, *objectid,* and *timeid* [5]. These three map the record to categorical values for Site, Partition, Sub-partition, Locus, and Container in the space dimension; Object type and an Object name in the object dimension; and Period and Chronology in the time dimension.

To get the count of all records in Nimrin, a linear traversal of all the records in the *idtable* of the "etanamultidibrowse" database is needed. The browsing structure of ETANA-DL being hierarchical, the browsing interface will have to display the count of records in each category at each hierarchical level. The number of record aggregates needed at each level of browsing and the size of the browsing index table, *idtable,* rules out the possibility of dynamically computing the aggregates each time the interface is loaded for a given browsing path. The browsing index database "etanamultidibrowse" is a set of tables and a join operation on these tables is needed to compute each aggregate; this further aggravates the problem of dynamic summarization for records in ETANA-DL.

The initial implementation of ETANA-CMV used dynamic summarization of ETANA-DL records along with the ETANA Browse service. The time taken to compute the aggregates before refreshing the visual interface was much beyond the unprepared response time constant of ~ 1 second. The unprepared response time as referred to by Card, and the operation time of Newell, determine the necessary rate of response to simple user actions, like clicking on a web link or clicking on a bar in ETANA-CMV. Clicking on a bar should load the next set of bars within 1 second, to be effective. Else, feedback has to be provided to the user within the 1 second to reassure him that something is happening [45].

Performance has been a major issue with DL visual browsing interfaces. Frequent interaction with the server to fetch the aggregates showing the distribution of records in categories limits the interactivity of the interface. To minimize waiting on the server for the aggregates, partial distribution aggregates can be preloaded in the working memory. These preloaded aggregates would typically be those that are frequently displayed in the browsing interface. However, this leads to some data being unavailable when needed and requires interaction with the server. The RB++ interface [30] downloads all the aggregates onto the client side memory before loading the visual interface but does not support any hierarchical navigation through the records. To address the performance problem with the existing "etanamultidibrowse" browsing index, we use a data cube based browsing index for ETANA-CMV.

4.2.2. Data cube based browsing index

The data cube for records in ETANA-DL has a view for each possible level of the browsing tree. The browsing tree in ETANA-DL has three dimensions, each of which is hierarchical. As already stated, the space dimension can be browsed at 5 possible hierarchical levels, namely: site, partition, sub-partition, locus, and container. The object dimension can be browsed at 2 possible hierarchical levels, namely: object type and object name. The time dimension has 2 levels, namely: period and chronology. The total possible number of levels across these three dimensions is 5*2*2=20. Thus our data cube has 20 views. Count of records is the measure attribute and the hierarchical levels form the dimensions of the different views.

Depending on the level of browsing in each dimension the corresponding view is used to summarize the intermediate result set. The initial level of browsing is: site in the space dimension, object type in the object dimension, and period in the time dimension. The levels are labeled 1 to 5 in the space dimension, and 1 to 2 in both the object dimension and the time dimension. 1 represents the top most level in each dimension, 2 represents the second hierarchical level, and so on. So, the initial levels in the three dimensions are 1, 1, and 1. The following two examples illustrate when each view is queried.

There is a view called view_111 corresponding to the top most level of browsing in each dimension. It contains the count of records grouped by each site, object type, and period. To get the number of bone records in Nimrin belonging to the Iron I period, this view is queried. To find the total number of records in Nimrin, counts of all rows in this view with site=Nimrin are aggregated.

To summarize the result sets in site: Nimrin, partition: NW, sub partition: N15/W70, object type: bone and period: Persian, we need to view the distribution of records in the given query across the various loci in sub-partition: N15/W70, the object names in object-type: seed, and chronologies in the period: Persian. Essentially the interface will be loaded with the levels 4, 2, and 2 in the three dimensions. For this case, view_422 is used. It contains the count of records grouped by site, partition, sub partition, locus, object type, object name, period, and chronology. The rows with values matching the given values in the query are aggregated to produce the required summarization.

Essentially these 20 views store pre-computed summarizations of records across various dimensions and browsing levels. Based on the query, the appropriate view is selected, the range of categories is selected from the view, and the aggregates are computed from the view. By initially loading a set of views to the interface memory, frequent interaction with the server can be avoided.

4.2.3. Data cube generator

This module builds/updates the data cube based on the existing browsing index of ETANA-DL. Each view in the data cube is created from the existing browsing index “etanamultidibrowse” database, whose schema is described in the ETANA Browse schema. Counts of records grouped by the dimensional attributes in the data cube are computed from the etanamultidibrowse database and stored in the corresponding views of the data cube.

With newer collections being added to the ETANA-DL union catalog, the “etanamultidibrowse” browsing index gets updated frequently. This module automatically updates the data cube when the corresponding script is run.

4.2.4. Implementation

The data cube generator has been implemented in Java. The database “etanavizbrowse” that serves as the data cube based browsing index for ETANA-CMV is a MySQL database. It has 20 tables, each representing a view in the data cube. Following are the important classes in the data component of ETANA-CMV. A brief explanation about the functionality of each of them is provided.

CubeViewCreator: This class creates each view in the data cube. Its constructor accepts the name of the view and the schema of the view and creates the table in the "etanavizbrowse" database. It then populates the table with the count of records in the various categories present in the hierarchical levels represented by the view. These counts are computed from view_522 which contains the counts of records in each category across each dimension. It is the view with the highest granularity of all the views in the data cube. This view is created by the EMDBExtractor class.

DBManager: This class extracts the dimensions of the cube, and determines the schema of the tables that represent the views of the cube. It interfaces with the Browsing schema of ETANA-DL stored in the etanabrowse.xml and extracts the above information. It then instantiates objects of the CubeViewCreator class. Each object of the CubeViewCreator class creates a view of the data cube.

EMDBExtractor: This class creates the view_522 table from the idtable of the "etanamultidibrowse" database. The spaceid, objectid, and timeid are replaced with the actual expanded hierarchical categories in the view_522 table. This table is used by the CubeViewCreator class to compute the aggregates in the higher hierarchies of the browsing tree stored in the other tables of the "etanavizbrowse" database. This class also has the additional functionality of formatting the category names while extracting them from the "etanamultidibrowse" database.

4.2.5. Issues

Incomplete meta-data: Some of the records in ETANA-DL have no meta-data available for some of the levels in the three dimensions. For example, records from Mozan do not have any values for the period and chronology fields. We need to classify these records under some category so that aggregates can be correctly computed. Some other records that do not fit under any category have the value unclassified. We decided to mark the category values as unclassified for records with incomplete/missing meta-data, just like those records that have the value unclassified for some fields.

Space: The data cube is a space consuming tool that provides quick summarization without any computation at the time of the query. The space needed by our data cube is 70 MB. There are three approaches to implementing the data cube. The entire data cube can be materialized, nothing can be materialized, or a part of the data cube can be materialized [9]. We have chosen to materialize the entire data cube. In the future, when the browsing structure of ETANA-DL becomes huge, other strategies to implement the data cube can be considered.

Time taken to build/update the data cube: It takes about 10-15 minutes to build/update the current data cube with the present records in ETANA-DL. This update needs to be initiated each time the browsing index is updated with new records in the union catalog. The current approach to the update is to rebuild the entire data cube from scratch after dropping all the tables in the "etanavizbrowse" database. In the future, when the browsing structure becomes huge, an incremental update can be applied to reduce the time taken to update the data cube.

4.3. Visualization component

This section outlines the two main modules in the visualization component. The visual interface is responsible for generating the GUI. The view loader interacts with the data cube and makes the corresponding view and data available to the visual interface depending on the browsing context/query.

4.3.1. Visual interface

This contains the logic for: generating the bars in the three panels, as well as the different colors of the bars during brushing and linking; maintaining the hierarchical browsing context in the tree view; and loading a new set of bars during hierarchical navigation in any of the three dimensions. It also contains the logic for the context menus in the GUI and for sorting the bars in the three panels based on user preferences.

4.3.2. View loader

The view loader ensures that the view that is needed by the visual interface is present in the working memory so that the response delay is reduced. Based on the browsing level, the view loader provides the categories, their counts, and the inter-dependencies between the categories in the three dimensions to the visual interface. It filters the categories that do not fall under the browsing context/query range, and computes the aggregate of each category in the three dimensions.

4.3.3. Implementation

Java Swing has been used for implementing the GUI in the visual interface. The view loader has been implemented in Java. The bar charts library used for visualizing the categories in the three dimensions is an open source library called JFreeChart library, available from jfree.org. Following are the important classes in the visualization component of ETANA-CMV and a brief explanation about the functionality of each.

DBExtractor: This class connects to the "etanavizbrowse" database on the server and extracts the views needed by the visual interface to the applet client memory. This class runs in a separate thread from the GUI thread, so that the GUI can be interactive while views are being downloaded to the client memory.

QueryProcessor: This class determines which view needs to be used for computing the aggregates based on the browsing level. It also computes the aggregates of records in the categories from the view based on the query range.

DimensionValues: This class contains the hashtable needed to store the categories along with the count of records in them and their interdependencies in the three dimensions.

EtanaVisualViews: This class uses the DimensionValues object to render the bar charts and the other elements in the GUI. It also contains logic to sort the bars based on user preferences, and supports user interaction with the GUI through mouse events and listeners.

4.3.4. Issues

Client side memory: The JVM in the applets has a default size of 64 MB. The entire data cube cannot be loaded at once into the JVM. Though the user can reset the size of the JVM, it is not a very straightforward process and is unfair to expect users to do the same. We address this problem currently by loading the views on a separate thread after discarding the views that are further from the browsing levels. The interactivity provided by applets can be provided by DHTML and AJAX technologies that have no such limitations on client side memory.

4.4. ETANA-DL Service Interface

ETANA-DL provides four categories of services. Information satisfaction services include searching, browsing, and recommending objects of similar interests by other users. Domain (Archeology) specific services are digital object comparison based on parameters and marking items as objects of interests to other users. Annotations, recent search history, and personal lists of items of interest are value added services. User management and Collection description are miscellaneous services [5].

This component integrates ETANA-CMV with the ETANA-DL portal. At present ETANA-CMV communicates with the browsing and searching services, and the interfaces provided by ETANA-DL. The browse service in ETANA-DL is an XML based web service described in section 1.2.3 whereas the search service in ETANA-DL uses the Lucene search engine [46]. To view the records from ETANA-CMV within the browsing context, the browse query is passed to the ETANA Browse Interface, which displays the corresponding records. If users wish to see all the records in ETANA-DL, the ETANA Browse Interface is called with no query. Similarly, to search for records with a given term, the search query is passed to the ETANA Search Interface, which displays the result set to the user. Depending on the user preferences, the browsing context too can be passed along with the search query. This would limit the result set to records present within the browsing context.

4.4.1. Implementation

The ETANA Browse Interface and Search Interface are Java servlets that use the ETANA Browse Service and the ODL search component, respectively. The URL of the browse/ search servlets with the query is constructed and a new browser window pointing to the constructed URL is opened.

4.4.2. Issues

Integration with other services: At present ETANA-CMV interfaces with only two services provided by ETANA-DL. It needs to be integrated more with the ETANA web portal so that users also can access these services from ETANA-CMV. There should be a protocol to switch between the ETANA-CMV interface and the ETANA web portal.

4.5. Summary

The architecture of ETANA-CMV is componentized. There are three components: the data component, the visualization component, and the ETANA service interface component. We believe that this architecture is extensible and can be adopted to develop similar visual browsing interfaces to other DLs.

5. FORMATIVE EVALUATION AND RESULTS

After the implementation of ETANA-CMV, as per the design described in section 3.2, was complete, it was decided to explore the usability of the system. In this chapter we explain the formative usability study that was conducted. The objectives and methodology of the study and the results from it are elaborated. The critical usability issues identified through this study were fixed/incorporated into ETANA-CMV.

5.1. Objectives of the study

Feedback from the focus group consisting of co-PIs and graduate students involved in the research project on ETANA-DL was used to refine the interface design and improve its usability leading to the designs described in sections 3.3.1 and 3.3.2 concerning iterations 2 and 3. At the end of the software development of ETANA-CMV, a formal usability study was designed. Formative evaluation, which is used to guide the design process and help re-design features of the user interface based on feedback from the users, was used to make further improvements to the ETANA-CMV GUI. The fundamental objective of this study was to identify key usability issues in the interface. The study would collect general user reactions about the ETANA-CMV interface design. The feedback from users would help guide research towards improvement of the tool.

5.2. Procedure

Archeologists and people with interest in archaeology would be the potential users of this tool. The participants for this study were those who had used DLs, like those from ACM or IEEE. They would understand the utility of browsing through records in ETANA-DL. Sixteen participants were recruited in total. All the participants were students from the CS4624 class at Virginia Tech, covering “Multimedia, Hypertext, and Information Access”. Participation in this study was voluntary.

The study was divided into four sections. The four sections were:

- User demographics
- Tutorial and exploration of content and features of two interfaces
- Task set using ETANA-CMV and ETANA Browse Interface
- Subjective rating and overall impression using post-evaluation survey forms

The study was conducted using an online survey instrument. Fourteen users performed the study simultaneously and the remaining two participated separately, at a convenient time. In the user demographics section, users were asked to rate their familiarity with Addison, ACM, IEEE Explore, and other DLs. A Likert scale (range: 1 – 9) was chosen for the rating questions. A tutorial explained the various features of the two interfaces. After reading the tutorial, users explored the two interfaces till they were comfortable with their features. Each participant then used both interfaces to perform ten benchmark tasks, five on each interface. Correctness of the answers, and the time taken to complete the tasks, including time taken to read the question, were measured. Users were asked to express their subjective satisfaction rating on the following four measures:

Comprehensibility (confusing to clear), Ease of use (difficult to easy), Speed of use (slow to fast), and Overall satisfaction (terrible to wonderful).

The benchmark tasks were designed based on a similar set of browsing tasks used by Professor Chris North in a study evaluating the value of coordinated visualizations over independent or single visualizations [47]. The different types of browsing tasks and the task questions used in the study are as follows.

1. **Lookup**: **"**How many records belong to the Byzantine time period?"

2. **Search for target value:** "Which site has 563 records?"

3. **Scan all:** "Which object-type has the most records?"

4. **Overview patterns**: "How many sites have names that begin with the letter M?"

5. **Ordering patterns**: "How many time periods are there between the Classical and Classical-Islamic time periods?"

6. **Cross Dimensional Coverage-yes lookup**: "Which sites have bones?"

7. **Cross Dimensional Coverage-no lookup**: "Which is the only site whose all records are classified into distinct time periods? (None of its records should belong to the time period – 'Unclassified')"

8. **Hierarchical navigation followed by Cross Dimensional scan**: "Which is the most predominant object-type from the site Nimrin?"

9. a) **Hierarchical navigation followed by Cross Dimensional lookup**: "What is the total number of pottery records in Site: Bab edh-Dhra->Partition: A-> Sub partition: 007?"

   b) **View records**: "Using the View records, list the vessel numbers of the first two pottery records found in the above browsing context."

10. **Search records: "**Using the search service, how many records in the entire digital library are about equus (horse)?"

The tasks were grouped into two task sets with equal number of tasks in each set. The tasks in each task set were picked randomly. Task set 1 consisted of tasks 1, 4, 5, 9 and 10 while task set 2 consisted of tasks 2, 3, 6, 7, and 8. Another variable is whether the textual ETANA Browse Interface was presented first or ETANA-CMV is presented first. So that gives us four groups:

1) Task set 1, ETANA Browse Interface first
2) Task set 1, ETANA-CMV first
3) Task set 2, ETANA Browse Interface first
4) Task set 2, ETANA-CMV first

Sixteen users were present; four were randomly assigned to each group. Overall comments and suggestions about the two interfaces were also obtained from users.

5.3. Results

Three charts are plotted from the data obtained in the study. Please see Figure 25 through Figure 27, along with the discussion in the following subsections.

5.3.1. Accuracy of answers

The first chart (Figure 25) shows the number of correct answers for each task, by eight users of each interface.

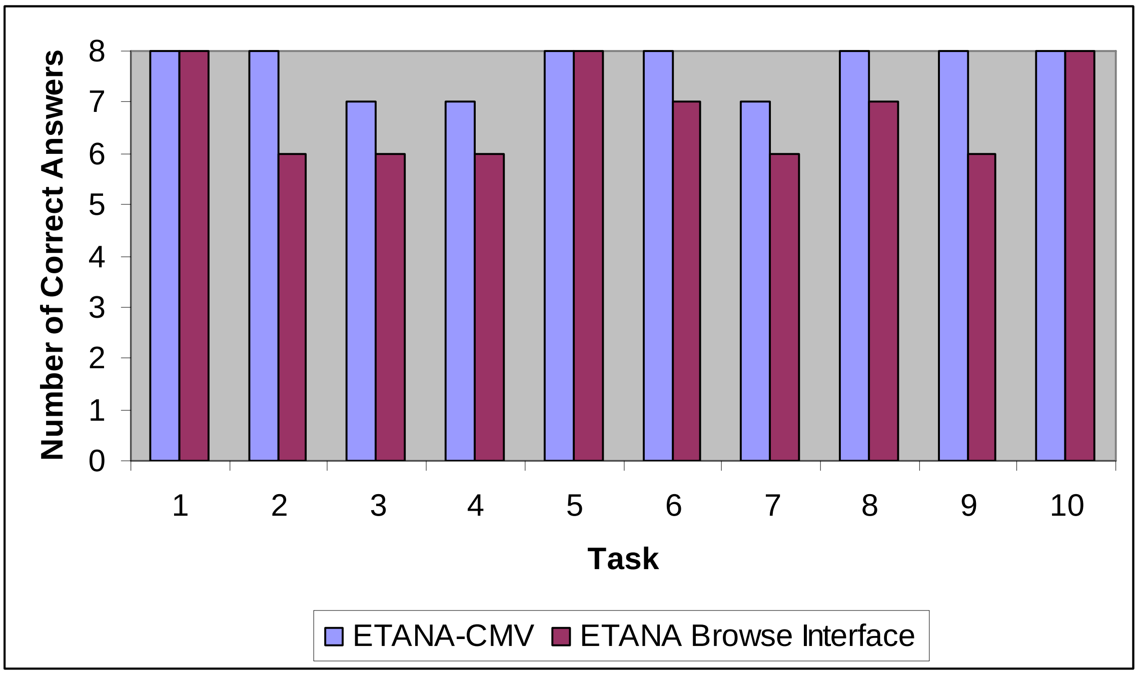

Figure 25: Average user accuracy for tasks, showing number of correct answers for each task

5.3.2. User performance time

Figure 26 shows the average user performance time for the ten tasks using both of the interfaces.

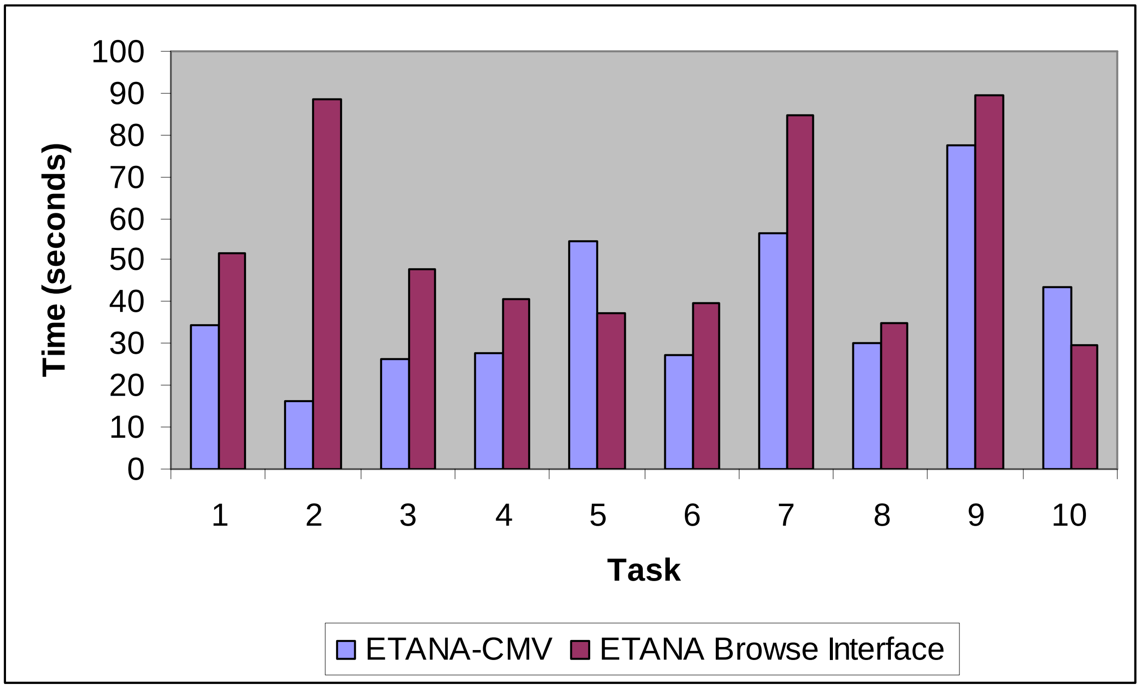

Figure 26: Average user performance time for tasks

5.3.3. User subjective satisfaction

Figure 27 shows the average subjective rating for comprehensibility, ease of use, speed of use, and overall satisfaction, on a 1-9 scale.

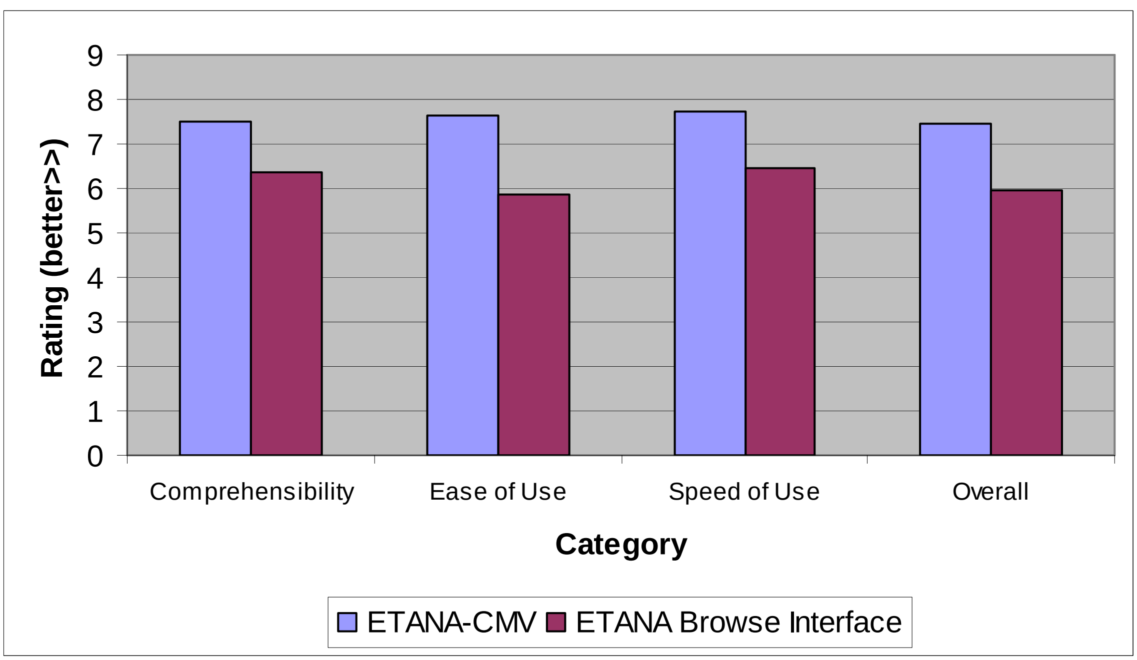

Figure 27: Average user subjective satisfaction

5.4. Discussion

5.4.1. Accuracy of answers

The total number of correct answers for all tasks using ETANA-CMV was 77 out of a possible 80, whereas the corresponding total for the ETANA Browse Interface was 68 out of a possible 80. These numbers map to 96% and 85%, respectively. This suggests that users were able to perform the browsing tasks more accurately using the ETANA-CMV interface.

All eight users working on each interface answered Tasks 1, 5, and 10 correctly. Visualization based strategies are not very helpful for answering Tasks 1, 4, 5, and 10. Task 1 was a simple lookup task to find the count of records in a category; users had to read the values on the bar labels to answer this question. Though the textual ETANA Browse Interface does not provide a count of records in each category, by clicking on the View Records link, users can see how many records are in the current category in the browsing level. Task 5 was designed to ensure that users were able to understand the ordering followed in the three panels of ETANA-CMV. Task 10 ensured that users were able to understand the integration of ETANA-CMV and the Search service provided in ETANA-DL.

In the other seven tasks, the chart in Figure 25 suggests that more users were able to correctly answer the task questions with ETANA-CMV than with the ETANA Browse Interface. The three inaccurate answers using the ETANA-CMV interface were one each for Tasks 3, 4, and 7.

5.4.2. User performance time

The chart in Figure 26 shows that for eight out of ten tasks, user performance time was less for ETANA-CMV. In task 1, ETANA-CMV achieved a 33.5% speedup over the ETANA Browse Interface. This task does not need any understanding of the visual mapping; it was a simple lookup task. The speedup reveals the importance of having counts along with the category names. With ETANA-CMV the count of records in each category can be viewed on the same screen, whereas users are required to navigate to the records page for each category to view the count with the ETANA Browse Interface.

In task 2, ETANA-CMV achieved around 81% speedup over the ETANA Browse Interface. The chart shows that this was the task with the biggest difference in average user performance time between the two interfaces. To complete this task using ETANA-CMV, users have to go through all the bar labels in the space panel in a linear fashion. However, due to the absence of count labels in the textual interface, users have to navigate to the view records page for each category to view the count. The results from this task re-emphasized the value of having counts along with category names.

In task 3, ETANA-CMV achieved around 46% speedup over the ETANA Browse Interface. This task requires an understanding of the mapping of the counts associated with each category to the length of the bars representing the categories. The speedup suggests that users are able to understand the bar chart visualization.

In task 4, ETANA-CMV achieved around 31% speedup over the ETANA Browse Interface. This task requires a scan of the labels on the bars in the space panel. Since the categories are sorted based on names in both the interfaces, users have to count the number of adjacent categories starting with M. Speedup was not expected for this task, because it requires no visualization based strategy to perform this task faster.

In task 5, users took more time using ETANA-CMV than using the ETANA Browse Interface. About 46% slowdown using ETANA-CMV was noticed. This task was to find the number of time periods between the Classical and Classical-Islamic time periods. This suggests that users took more time to understand the ordering of the bars in ETANA-CMV.

In task 6, ETANA-CMV achieved around 32% speedup over the ETANA Browse Interface. This task required users to list sites having bones. The brushing and linking interaction mechanism in ETANA-CMV is the ideal strategy to perform this task. It can also be performed by hierarchically navigating into the bones category in the Object dimension; the dynamic filtering feature shows only sites having bones in the Space dimension.

In task 7, ETANA-CMV achieved around 33% speedup over the ETANA Browse Interface. This task is similar to task 6 and requires the same interaction strategy required to perform task 6.

In task 8, ETANA-CMV achieved around 14% speedup over the ETANA Browse Interface. This task requires users to navigate to the Site Nimrin and then scan through the bars in the Object dimension. This task needs an understanding of the hierarchical navigation in ETANA-CMV along with the mapping of counts to length of bars.

In task 9, ETANA-CMV achieved around 13% speedup over the ETANA Browse Interface. This task has two parts; the first part is similar to task 8 and the second part requires users to click the View records button in ETANA-CMV and read the first two records from the Records page of the ETANA-DL portal interface. In the ETANA Browse Interface, the second part does not require a separate operation. Users have to click the view records link to see the count of records asked for in the task, and the answer for the second part is found on the same page. This task as a whole needs understanding of the visual mapping, hierarchical navigation, and integration of ETANA-CMV with the ETANA Web portal.

In task 10, users took more time using ETANA-CMV than using the ETANA-Browse Interface. About 45% slowdown using ETANA-CMV was noticed. This task involves usage of the Search integration feature in ETANA-CMV.

5.4.3. User subjective satisfaction

The differences in rating of user subjective satisfaction between the two interfaces for the four categories were 1.1, 1.8, 1.3, and 1.5 respectively in favor of ETANA-CMV.

5.4.4. Positive comments about features

Overall, in their written comments, users had expressed a better opinion about ETANA-CMV. There were specific comments that ETANA-CMV was much faster, and easier to learn as well as use. One user had mentioned that "having visual references and numbers available before viewing records was very helpful" and another user had mentioned that "the highlighting of categories was very useful and effective". One user had mentioned that "the visual interface's clear presentation of the categories and data helped a lot in accessing the database quicker and intuitive."

5.4.5. Usability issues and improvements

Following is a list of usability issues and improvements in ETANA-CMV identified through user comments.

1. Three users mentioned that clicking and mouse move over on the bars with very small lengths was difficult. They suggested using the entire horizontal area to the right of each bar, as a listener to user interaction.
2. One user suggested navigation through keyboard and zooming techniques to remove the scroll bars in the panels.
3. Three users had mentioned the need for a reset/home sort of button to navigate to the starting browsing context.
4. Two users remarked that they needed feedback when they reached the finest hierarchical level - that this was indeed the finest level and that they could not go any deeper.
5. One user had suggested highlighting only the part of the bar that relates to the percentage in that category to the one you are moving the mouse over.
6. Two users had suggested reducing the intensity in the color of the bars to make the labels on the bars more readable.

5.5. Summary

The general user subjective reaction and general comments about ETANA-CMV was positive. Key usability issues were identified. ETANA-CMV showed promise in helping users gain more insight into the collections in the DL by means of them being able to perform various types of browsing tasks more accurately, and faster than by using the existing ETANA Browse Interface.

Usability issues and suggestions 1, 3, 4, and 6 from section 5.4.5 were fixed and incorporated in ETANA-CMV by way of the following changes.

- The area to the right of each horizontal bar is also interactive and listens to mouse move over and click events.

- A Home button was added to allow users to browse back to the highest browsing level in a single click.
- A status bar was added in the bottom to give feedback to users when the lowest level in each hierarchy was reached.
- The intensity in the color of the bars was reduced to make the labels more readable.
- Additionally, the ordering of the bars was revealed on each panel through a drop down list and users could choose the ordering of bars from the drop down list. The context menu that pops up during the right click event on the panels was removed.
- It was also decided to show the percentage of records in each category along with the counts and category names based on feedback from discussion of the final interface design with the focus group for ETANA-DL.

Keyboard based navigation and use of zooming techniques are left for future work. Highlighting portions of the bar during mouse move over events will be very helpful in revealing the distribution of records from one category, over categories in other dimension. This feature too is left for future work on the interface.

6. CONCLUSION AND FUTURE WORK

In this chapter, we discuss the conclusions of this thesis work. We show how our hypothesis is demonstrated/ proven. The Future Work section details potential extensions to this thesis.

6.1. Conclusion

Our hypothesis is that:

a) The problem of presenting heterogeneous archeological records to users in an insightful manner can be solved by our approach of designing a visual interface based on coordinated uniform multiple view design.

b) System performance when browsing through records in ETANA-DL can be enhanced by the use of a data cube based browsing index.

We prove this hypothesis by describing the interface "ETANA-CMV" with the following features.

The visual interface of ETANA-CMV is designed to be:

- Scalable: Demonstrated by integrating the first design principle which is to visualize the browsing structure rather than the individual records. (described in section 3.1.1)
- Flexible: Demonstrated by integrating the second design principle of displaying all browsable dimensions to users so that they can seamlessly browse across dimensions. (described in section 3.1.2)
- Easy to learn: Demonstrated by integrating the third principle of using uniform multiple views instead of using different visualization components for each dimension. (described in section 3.1.2)

The data component of ETANA-CMV is designed to be:

- Efficient: Demonstrated by integrating a data cube based browsing index that enhances the performance of ETANA-CMV. (described in section 4.2.2)
- Semi-automated: Demonstrated by the script to automatically rebuild the data cube whenever the browsing index for ETANA-DL is updated due to addition of new collections. (described in section 4.2.3)

6.2. Future work

There are many potential extensions to this work that are worth considering.

Visualization component

Improvement of visualization tools happens over alternating stages of user studies and development. While mousing over any bar, other bars can be proportionally highlighted to show conditional distribution across all dimensions. Zooming techniques may be used to avoid scrolling in the three panels. Design features to include ontology, and personalization in the browsing interface, need to be considered when ETANA-DL has a well defined ontology in place.

Data component

In the current data cube for ETANA-CMV, we compute aggregates for all the views in the cube. A common alternative is to materialize only those views that are useful for frequently used queries. Harinarayanan [9] proposes a framework to infer the views that need to materialize for an efficient data warehouse with partial views. Such techniques can be used to build an efficient data cube based index. This would enhance the scalability of the system when more collections are added to the DL. The current data cube generator rebuilds the entire cube when new collections are added and the browsing index for ETANA-CMV needs to be rebuilt. Methods to incrementally update the data cube will reduce the updating time required.

Integrating ontology

Ontology for archeological records in ETANA-DL is under development. With the different schemas and naming conventions followed across the different research projects in ETANA, the ontology would help bridge this gap. Once this ontology is in place, design features in the visual interface and new data structures and indices will be needed to handle the ontology. Ontologies are complex and often represented as a graph. The ontology graph is incremental; it grows when new collections are added to the DL.

ETANA-CMV facilitates browsing through a hierarchical browsing structure. There is no provision to visualize complex graph-like browsing paths. The ontology can serve as a foundation for various taxonomies. Depending on the user and his/her goals, the ontology can be partitioned into different taxonomies. These taxonomies are hierarchical trees, which can then be mapped onto a browsing structure. A different data cube for each of these taxonomies/ hierarchical trees will be sufficient to provide the efficient and rapid access to the information needed by the visual interface.

Implementation technologies

ETANA-CMV has been implemented using Java applets. Applets are restricted by the default Java Virtual Memory size for the different operating systems. This limits the number of views that can be loaded into the memory of the client running the applet. This problem can be countered by the use of Ajax (Asynchronous JavaScript and XML) technologies. Ajax provides better bandwidth utilization; the HTML rendering is done on the client side and data is exchanged asynchronously with the web server through the

XMLHttpRequest object. The use of Ajax will enhance the interactivity of browsing interfaces.

Integration with ETANA-DL

The current Browse Interface in the web portal is linked to the search service and interface. It allows users to save their navigation paths, view records within the browsing context, and browse through search results. Currently ETANA-CMV communicates only with the View Records and the Search interface in the web portal of ETANA-DL. A method to integrate ETANA-CMV with the other services in a seamless manner needs to be devised. One way of doing this would be to replace the current textual browsing interface with an Ajax version of ETANA-CMV and have it as a part of the web portal itself. This would enable users to log into the system and use ETANA-CMV. It is then possible to save navigation paths while browsing. Search results can also be visualized in ETANA-CMV; Envision and GRIDL have been successful at presenting search results in a browsing platform.

Applications

The VIDI protocol [48] is a light weight version of the Open Archives Initiative Protocol designed to achieve interoperability between visualization systems and digital libraries. It guides design of visualization systems for DLs as services rather than rigid interface components. The VIDI protocol can be integrated into the current architecture of ETANA-CMV to add flexibility and extensibility. An extensible architecture would assist development of similar interfaces for other DLs without much time and effort. We will then be able to use the tool as a browsing interface generator to DLs.

The interface design proposed in ETANA-CMV can be used to visualize the records in other DLs like CITIDEL [49] and NDLTD [50]. CITIDEL is a digital library of documents from different collections like the ACM DL. Users can browse these documents by subject or by source collection. The NDLTD is a digital library of electronic theses and dissertations. Users can browse the ETDs by year or institution. Essentially, the nature of records in NDLTD is multidimensional and categorical. Following are some ways to hierarchically organize these records: months within a year, alphabetical ordering of author names, departments within an institution. Our visual interface design can be applied to create visual browsing interfaces for the records in these DLs.

The data cube based index can also be used in the textual ETANA Browse Interface and in other DLs to support more interaction.

Summative evaluation

A Summative Evaluation that compares ETANA-CMV with other similar visual browsing interfaces in terms of performance, taking into account time to complete benchmark tasks as well as comparison of errors and learnability issues, could be conducted with a larger user base. This will help us to make statistical claims about the performance of ETANA-CMV. Participants for the summative evaluation can be archeologists, or students from archeology departments. It will also be helpful to have archeologists and participants in the various research projects in ETANA use this interface and to collect their feedback.

7. REFERENCES

1. *Tell Nimrin - A virtual archaeological site*. 2006, http://www.case.edu/affil/nimrin/ (as of November 2006).
2. *Megiddo*. 2006, http://www.tau.ac.il/humanities/archaeology/megiddo/index.html (as of November 2006).
3. *Lahav Research Project*. 1999, http://www.cobb.msstate.edu/dig/LRP-1999-01/ (as of November 2006).
4. *Madaba Plains Project – Tall al-'Umayri*. 2006, http://www.wwc.edu/academics/departments/theology/mpp (as of November 2006).
5. Ravindranathan, U., *Prototyping Digital Libraries, Handling Heterogeneous Data Sources - An ETANA-DL Case Study. Master's Thesis*. In *Department of Computer Science* 2004, Virginia Polytechnic Institute and State University: Blacksburg, VA.
6. *Open Archives Initiative*. 2006, www.openarchives.org (as of November 2006).
7. Gray, J., Bosworth, A., Layman, A., & Pirahesh, H., *Data Cube: A relational aggregation operator generalizing group-by, cross-tabs and sub-totals*, In *12th Int'l Conference on Data Engineering*. 1996: New Orleans. p. 152-159.
8. Agrawal, R., Gupta, A., & Sarawagi, S., *Modeling multidimensional databases*, In *13th Int'l Conference on Data Engineering*. 1997: Birmingham, U.K. p. 232-243.
9. Harinarayan, H., Rajaraman, A., & Ullman, J. D., *Implementing data cubes efficiently*, In *ACM SIGMOD '96*. 1996: Montreal. p. 205-216.
10. Geffner, S., Agrawal, D., Abbadi, A., & Smith, T., *Browsing large digital library collections using classification hierarchies.* In *8th International Conference on Information Knowledge Management, CIKM'99*. 1999: Kansas City. p. 195-201.
11. Yee, K.-P., Swearingen, Kirsten., Li, Kevin., & Hearst, Marti, *Faceted metadata for image search and browsing*. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, Editor. 2003: Ft. Lauderdale, Florida, USA. p. 401-408.
12. Card, S.K., Mackinlay, Jock D., & Shneiderman, Ben, *Readings in Information Visualization: Using Vision to Think*. 1999, San Francisco, CA: Morgan Kaufmann.
13. Ammoura, A., *DIVE-ON: From Databases to Virtual Reality.* ACM Crossroads Database Special Edition, 2001. **7**(3): p. 4.
14. Stolte, C., Tang, D., & Hanrahan, P., *Polaris: A System for Query, Analysis, and Visualization of Multidimensional Relational Databases.* IEEE Transactions on Visualization and Computer Graphics, 2002. 8(1): p. 52-65.
15. Shekhar, S., Lu, C., Tan, X., Chawla, S., & Vatsavai, R, *Mapcubes: A visualization tool for spatial data warehouses.* Geographic data mining and Knowledge Discovery (GKD), 1999.
16. Sam-Rajkumar, J.L., Shen, Rao, Vemuri, Naga Srinivas., Fan, Weiguo., and Fox, Edward A. , *EtanaCMV: A Visual Browsing Interface for ETANA-DL Based on Coordinated Multiple Views*. In *ECDL*. 2006: Alicante, Spain. p. 492-494.

17. Börner, K., & Chen, Chaomei *Visual Interfaces to Digital Libraries: Motivation, Utilization, and Socio-technical Challenges*. In *Second ACM+IEEE Joint Conference on Digital Libraries*. 2002, Springer: Portland, Oregon, USA. p. 1-12.
18. Ravindranathan, U., Rao, Shen., Goncalves, Marcos A., Weiguo, Fan., Edward A., Fox., & Flanagan, James W., *Prototyping Digital Libraries Handling Heterogeneous Data Sources - The ETANA-DL Case Study*. In *European Conference on Digital Libraries*. 2004: Bath, UK. p. 186-197.
19. Smith, T.R., Janee, Greg, Frew, James, & Coleman, Anita. *The Alexandria digital earth prototype*. In *International Conference on Digital Libraries*. 2001, Proceedings of the 1st ACM/IEEE-CS joint conference on Digital libraries: Roanoke, Virginia, United States. p. 118 - 119
20. Chang, S.J., & Rice, R.E., *Annual Review of Information Science and Technology*. 1993. p. 231-276.
21. Bertini, E., Catarci, Tiziana, Bello, Lucia Di & Kimani, Stephen, *Visualization in Digital Libraries.* From Integrated Publication and Information Systems to Virtual Information and Knowledge Environments, 2005: p. 183-196.
22. Börner, K., Feng, Ying., & McMahon, Tamara., *Collaborative Visual Interfaces to Digital Libraries*. In *Second ACM+IEEE Joint Conference on Digital Libraries*. 2002, Springer: Portland, Oregon, USA. p. 279-280.
23. Marks, L., Hussell, Jeremy A.T., Mcmahon, Tamara M., & Luce, Richard E. , *ActiveGraph: A Digital Library Visualization Tool.* International Journal on Digital Libraries, Special Issue on Information Visualization Interfaces for Retrieval and Analysis, 5(1), 2005.
24. Heath, L.S., Hix, Deborah, Nowell, Lucy T., Wake, William C., Averboch, Guillermo A., Labow, Eric, Guyer, Scott A., Brueni, Denis J., France, Robert K., Dalal, Kaushai & Fox, Edward A., *Envision: A user-centered database of computer science literature.* Communications of the ACM, 38(4), 1995: p. 52-53.
25. Shneiderman, B., Feldman, David, Rose, Anne, & Grau, Xavier Ferré, *Visualizing Digital Library Search Results with Categorical and Hierarchical Axes.* Proceedings of DL-00, 5th ACM Digital Library Conference, 1999: p. 57-66.
26. Lamping, J., Rao, Ramana, & Pirolli, Peter, *A Focus+Context Technique Based on Hyperbolic Geometry for Visualizing Large Hierarchies*. 1995, Xerox Palo Alto Research Center.
27. Zylab. *Visualization Module*. 2005 [cited; Available from: http://www.zylab.com/products_technology/productsheets/Visualization.pdf.
28. Shen, R., Vemuri, N. S., Vijayaraghavan, V., Fan, W., & Fox, E. A., *EtanaViz: A Visual User Interface to Archaeological Digital Libraries*. In *Technical Report TR-05-14*. 2005, Computer Science, Virginia Tech.
29. Raghavan, A., Vemuri, Naga Srinivas, Shen, Rao, Goncalves Marcos, Fan, Weiguo, & Fox, Edward A., *Incremental, Semi-automatic, Mapping-Based Integration of Heterogeneous Collections into Archaeological Digital Libraries: Megiddo Case Study*. In *ECDL* 2005. p. 139-150.
30. Zhang, J., & Marchionini, Gary, *Evaluation and evolution of a browse and search interface: Relation Browser++*. In *ACM International Conference Proceeding Series; Vol. 89*. 2005: Atlanta, Georgia. p. 179 - 188.

31. Shneiderman, B., *The Eyes have it: A Task by Data Type Taxonomy*. Proceedings of IEEE Symposium. Visual Languages 96, 1996: p. 336-343.
32. Buchanan, G., Jones, Matt., & Marsden, Gary. *Exploring Small Screen Digital Library Access with the Greenstone Digital Library*. In *2nd South African Conference on Human-Computer Interaction*. 2001. South Africa.
33. Nation, D.A., *WebTOC: a tool to visualize and quantify Web sites using a hierarchical table of contents browser*. In *Conference on Human Factors in Computing Systems*. 1998: Los Angeles, California, United States. p. 185 - 186.
34. Shneiderman, B., *Tree visualization with tree-maps: 2-d space-filling approach.* ACM Transactions on Graphics (TOG), 1992. **11**(1): p. 92-99.
35. Robertson, G.G., Mackinlay, Jock D., & Card, Stuart K., *Cone Trees: animated 3D visualizations of hierarchical information*. In *Conference on Human Factors in Computing Systems*. 1991, ACM Press: New Orleans, Louisiana, United States. p. 189 - 194.
36. Good, L.E., Popot, A. C., Janssen, W. C., & Bier, E. A., *Fluid Interface for Personal Digital Libraries*. In *ECDL*. 2005: Vienna, Austria. p. 162-173.
37. Graham, M., Kennedy, Jessie B., & Hand, Chris, *The Challenge of Visualising Multiple Overlapping Classification Hierarchies.* User Interfaces to Data Intensive Systems, 1999: p. 42-51.
38. Williamson, C., & Shneiderman, B., *The Dynamic HomeFinder: Evaluating dynamic queries in a real-estate information exploration system.* Proceedings of the 15th annual international ACM SIGIR conference on research and development in information retrieval, 1992: p. 339-346.
39. Kovács, L., Micsik, András., & Pataki, Balázs. *AQUA: query visualization for the NCSTRL digital library*. In *International Conference on Digital Libraries*. 1999: Berkeley, California, United States. p. 230-231.
40. Marchionini, G., Plaisant, Catherine, & Komlodi, Anita, *Interfaces and tools for the Library of Congress National Digital Library Program.* Information Processing and Management 1998. **34**(5): p. 535-555.
41. Kampanya, N., Rao, Shen, Kim, Seonho, North, Chris, & Fox, Edward A., *CitiViz: A Visual User Interface to the CITIDEL System.* In *European Conference on Digital Libraries*. 2004: Bath, UK. p. 122-133.
42. Tanin, E., *Browsing Large Online Data Using Generalized Query Previews*. In *Department of Computer Science* 2001, University of Maryland: College Park, Maryland.
43. Baldonado, M.Q.W., Woodruff, Allison., & Kuchinsky, Allan, *Guidelines for using multiple views in information visualization.* Proceedings of the working conference on advanced visual interfaces, 2000: p. 110 - 119
44. Rosson, M.B., & Carroll, J. M., *Usability engineering: scenario based development of human-computer interaction*. 2002, San Francisco, CA: Academic Press.
45. Robertson, G.G., Cameron, K., Czerwinski, M., & Robbins, D., *Animated Visualization of Multiple Intersecting Hierarchies.* Information Visualization, 2002. **1**: p. 50-65.
46. *Lucene* 2006, http://lucene.apache.org (as of November 2006).

47. North, C., & Shneiderman, B, *Snap-Together Visualization: Can Users Construct and Operate Coordinated Visualizations?* International Journal of Human Computer Studies, 2000. **53**(5): p. 715-739.
48. Wang, J., *VIDI: A Lightweight Protocol between Digital Libraries and Visualization Systems*. *Master's Thesis.* In *Department of Computer Science* 2002, Virginia Polytechnic Institute and State University: Blacksburg, VA.
49. *CITIDEL*. 2006, http://www.citidel.org (as of November 2006).
50. *NDLTD*. 2006, http://www.ndltd.org (as of November 2006).

8. APPENDIX A: QUESTIONNAIRE

Study ID:

Pre-evaluation questionnaire

1) What is your gender?
   - ○ Male
   - ○ Female

2) Please specify the following:

   Major or Area of Study:

   ____

   Level (freshman, sophomore, MS, PhD, etc.):

   ____

   Research interests (Networking, Graphics, HCI, etc.):

   ____

3) Rate your familiarity with digital library (DL) systems, e.g., Addison reached from www.lib.vt.edu, or any other system such as ACM Digital Library and IEEE Xplore.

	Not Familiar				Very Familiar
	1	2	3	4	5
Addison	○	○	○	○	○
ACM	○	○	○	○	○
IEEE Xplore	○	○	○	○	○
Others					
__________	○	○	○	○	○
__________	○	○	○	○	○
__________	○	○	○	○	○

Study ID:

Please attempt to the best of your ability to address each of the questions that follow. Please provide your answer along with any comments you have about the task, its feasibility, or the system behavior.

Tasks

1) How many records belong to the Byzantine time period?

2) Which site has 563 records?

3) Which object-type has the most records?

4) How many sites have names that begin with the letter "M"?

5) How many time periods are there between the Classical and Classical-Islamic time periods?

6) Which sites have bones?

7) Which is the only site where all records are classified into distinct time periods? (None of its records should belong to the time period "Unclassified".)

8) Which is the most predominant object-type from the site Nimrin?

9) What is the total number of pottery records in Site: Bab edh-Dhra->Partition: A-> Sub partition: 007?

    a. Using the View records feature, list the vessel numbers of the first two pottery records found in the above browsing context.

10) Using the search service, how many records in the entire digital library are about equus (horse)?

Study ID:

11) The ETANA DL has information about archeological records from Middle East archeological sites.

    a. List some analysis type of questions that you would like to ask about the collections in ETANA DL, or that you expect others might commonly ask.

    b. Please list observations, inferences, conclusions, and any other comments about the DL that relate to the questions you listed above.

9. APPENDIX B: INFORMED CONSENT FORM

VIRGINIA POLYTECHNIC INSTITUTE AND STATE UNIVERSITY
Informed Consent for Participants
in Research Projects Involving Human Subjects

Title of Project: **ETANA-CMV Usability Evaluation**
Investigator(s): Dr. Edward A. Fox, Johnny L. Sam Rajkumar, Srinivas Vemuri

I. Purpose of this Research/Project
We have two browsing interfaces to the digital library associated with the ETANA project (ETANA-DL). The first interface is a conventional text-based browsing interface, while the second one is a visual browsing interface (ETANA-CMV) which we believe will give more insight into the collection while browsing. We want to measure the effectiveness of the two browsing interfaces given a set of benchmark tasks, and to get subjective feedback on them. We expect to identify key usability problems in addition to evaluating the effectiveness of the visual browsing interface. The results of the study will be used to further refine ETANA-CMV.

The participants of this study will involve students from the CS4624 course at Virginia Tech who are more than 18 years old. We expect around 30 participants. The participants are from both genders and are healthy individuals.

II. Procedures
The experiment involves four phases during a single session. The whole procedure may last about 75 minutes. Each user is required to

1. View the online video tutorial of ETANA-CMV at
   http://feathers.dlib.vt.edu:8080/etana/Tutorials/Index.html
2. Complete a pre-evaluation questionnaire and explore the capabilities of the system.
3. Perform a series of tasks using ETANA-CMV and ETANA - Browse Interface. After completion of each task, users will be asked to fill-in a task-related questionnaire.
4. Provide subjective reactions using post-evaluation survey forms.

III. Risks
There are no more than minimal risks involved, and participants are free to discontinue the study at any time.

IV. Benefits
Participants will help identify usability problems which will help improve the design of browsing interfaces used in archeological digital libraries and other digital libraries in related disciplines.

V. Extent of Anonymity and Confidentiality
The performance on this study will not affect any individual in any manner. The personal information collected using the study will be used only for understanding the varied population using which the system has been studied.

VI. Compensation
In order to get 2 points extra credit toward the final class grade, participants will have the option of attending a class session in which the activity will be to participate in the study, or they will be able to undertake a related computer writing assignment. This is explained in the course outline page.

VII. Freedom to Withdraw
Participants are free to withdraw from the study any time they wish.

VIII. Subject's Responsibilities
I voluntarily agree to participate in this study. I have the following responsibilities:
Complete the information gathering document, pre-evaluation questionnaire, usability tasks, and post evaluation questionnaire.

IX. Subject's Permission
I have read the Consent Form and conditions of this project. I have had all my questions answered. I hereby acknowledge the above and give my voluntary consent:

____ Date__________
Subject signature

Should I have any pertinent questions about this research or its conduct, and research subjects' rights, and whom to contact in the event of a research-related injury to the subject, I may contact:

Johnny L. Sam Rajkumar — 540 818 6292 / johnny@vt.edu
Investigator(s) — Telephone/e-mail

Dr. Edward A. Fox — 540 231 5113/ fox@vt.edu
Faculty Advisor — Telephone/e-mail

Dr. Edward A. Fox — 540 231 5113/ fox@vt.edu
Departmental Reviewer/Department Head — Telephone/e-mail

David M. Moore — 540 231 4991/ moored@vt.edu
Chair, Virginia Tech Institutional Review Board for the Protection of Human Subjects — Telephone/e-mail
Office of Research Compliance
1880 Pratt Drive, Suite 2006 (0497)
Blacksburg, VA 24061