Exploring Digital Libraries: Integrating Browsing, Searching, and Visualization

Rao Shen[1], Naga Srinivas Vemuri[1], Weiguo Fan[1], Ricardo da S. Torres[2], Edward A. Fox[1]

Digital Library Research Laboratory[1]
Virginia Tech
Blacksburg, Virginia USA 24061
{rshen, nvemuri, wfan, fox}@vt.edu

State University of Campinas[2]
Institute of Computing
Campinas, SP, Brazil
rtorres@ic.unicamp.br

ABSTRACT

Exploring services for digital libraries (DLs) include two major paradigms, browsing and searching, as well as other services such as clustering and visualization. In this paper, we formalize and generalize DL exploring services within a DL theory. We develop theorems to indicate that browsing and searching can be converted or mapped to each other under certain conditions. The theorems guide the design and implementation of exploring services for an integrated archaeological DL, ETANA-DL. Its integrated browsing and searching can support users in moving seamlessly between these operations, minimizing context switching, and keeping users focused. It also integrates browsing and searching into a single visual interface for DL exploration. A user study to evaluate ETANA-DL's exploring services helped validate our hypotheses.

Categories and Subject Descriptors

H.3.7 [**Information Storage and Retrieval**]: Digital Libraries

General Terms

Design, Theory, Experimentation

Keywords

Integration, Exploring, Searching, Browsing, Visualization

1. INTRODUCTION

Browsing and searching are two major paradigms for exploring DLs. They are often provided by DLs as separate services. Developers commonly see these functions as having different underlying mechanisms, and they follow a functional, rather than a task-oriented approach to interaction design. While exhibiting complementary advantages, neither paradigm alone is adequate for complex information needs (e.g., that lend themselves partially to browsing and partially to searching [16]). Searching is popular because of its ability to identify information quickly. On the other hand, browsing is useful when appropriate search keywords are unavailable to users (e.g., a user may not be certain of what she is looking for until the available options are presented during browsing; certain criteria do not lend well to keyword search; the exact terminology used by the system may not be known). Browsing also is appropriate when a great deal of contextual information is obtained along the navigation path. Therefore, a synergy between searching and browsing often is required to support users' information-seeking goals [2, 3, 8, 12]. Accordingly, a panel at the World Wide Web Conference in 2005 brought together experts to discuss trends in the integration of searching and browsing, and in 1995 there was a panel on "Browsing vs. Search: Can We Find a Synergy?" at the Conference on Human Factors in Computing Systems.

Text mining and visualization techniques provide DLs additional powerful exploring services, with possible beneficial effects on browsing and searching. RB++ [22] provides visualized category overviews of an information space and allows dynamic filtering and exploration of the result set by tightly coupling the browsing and searching functions. Our user study of the CitiViz system [7], which combines browsing, searching, document clustering, and information visualization, showed its advantages, in user performance and preference.

Though many research projects have developed different interaction strategies allowing smooth transition between browsing and searching, to the best of our knowledge, none of them generalize these two predominant exploring services in DLs. Reflecting upon the current state of the art, and different types of exploring services for DLs, has led us to the following research questions:

- Are browsing and searching duals or can they be converted to each other when certain conditions are met?
- Can we generalize these DL exploring services within a formal DL framework?
- Can the formal generalization guide development of exploring services for domain focused DLs?

To address the above mentioned questions, we

- Generalize DL exploring services such as browsing, searching, clustering, and visualization in the context of the 5S DL theory [9, 10], and develop theorems and lemmas.
- Prove that browsing and searching can be converted and mapped to each other under certain conditions based on the theorems and lemmas developed.

JCDL'06, June 11–15, 2006, Chapel Hill, North Carolina, USA.

- Use an integrated archaeological DL, ETANA-DL (http://etana.dlib.vt.edu) [17, 18], as a case study to illustrate the application of our theoretical approach. We conducted a user study to evaluate ETANA-DL's exploring services. We found that users significantly prefer to integrate browsing and searching.

The remainder of this paper is structured as follows. Section 2 discusses related work. Section 3 formalizes DL exploring services. Section 4 describes the exploring services for our archaeological DL, developed based on the theorems and lemmas. Section 5 presents the user evaluation of those services. Conclusions are outlined in Section 6.

2. RELATED WORK

The idea of integrating searching and browsing can be found in some early systems in the 1980s, such as I^3R [5] and RABBIT [20]. Though I^3R had that idea, it did not implement it. While affording compelling browsing experiences, the interface to a database provided by RABBIT is based on the paradigm of 'retrieval by reformulation'.

About 10 years after RABBIT and I^3R appeared, searching and browsing integration resurfaced in many efforts, such as PESTO [4] and DataWeb [13]. PESTO integrated browsing and querying via a "query-in-place" paradigm for exploring the contents of object databases. It allowed a user to issue a query relative to the point that her navigation had reached. However, PESTO was not equipped for browsing semi-structured data.

Navigation is the primary mode for DataWeb to interact with the database. DataWeb viewed navigation as a process of query rewriting and query refinement. One can browse or search to attain a different hierarchy at any point while interacting with the DataWeb system. While in this context queries induce hierarchies, there is also an initial set of pre-existing hierarchies available as exemplars for a user to browse prior to querying. Thus, a user may begin an information-seeking activity in the DataWeb system with a query, or browse an extant hierarchy.

Typically, XML data elements are nested, making XML documents conducive to browsing hierarchically. Thus, interactively blending browsing and querying of XML is quite natural. The MIX project [14] provided virtual (i.e., non-materialized) integrated views of distributed XML sources and facilitates the interleaved browsing and querying of the views at both the front-end level and the programmatic level. At the front-end level it provided the BBQ GUI [15], which adopted PESTO's feature of "query-in-place". At the programmatic level MIX provided an API called QDOM (Querible Document Object Model) supporting interleaved querying and browsing of virtual XML views, specified in an XQuery-like language. The navigation commands are a subset of the navigation commands of the standard DOM API. QDOM allowed an "in-place-query" to be issued from any node in the result of previous queries. The query generates a new "answer" object from which a new series of navigation commands may start.

Though searching and browsing integration were embraced in the database area (as per above), the combined paradigm is exhibited by Web users during their information-seeking, and presented in many research efforts such as AMIT [21], WebGlimpse [11], ScentTrails [16], and SenseMaker [1]. AMIT (Animated Multiscale Interactive TreeViewer) [21] is a Java applet that integrates fisheye tree browsing with search and filtering techniques. WebGlimpse [11] allowed limiting the search to a neighborhood of the current document.

ScentTrails [16] annotated the hyperlinks of retrieved Web pages with search cues: indications that a link leads to content that matches the search query. The annotation was done by visually highlighting links to complement the browsing cues (textual or graphical indications of the content reachable via a link) already embedded in each page.

SenseMaker [1] increased the fluidity between browsing and searching DLs by introducing structure-based filtering and structure-based searching. In SenseMaker, a user issued a query and aggregated the retrieved results into bundles by "bundling criterion" (e.g., "same author"). Structure-based filtering allowed users to focus on selected bundles and to employ structure to limit a collection of results quickly and at a high level of granularity. The structure-based searching involves growing selected bundles or adding related bundles.

Though many research projects have developed different interaction strategies allowing smooth transition between browsing and searching, to the best of our knowledge, none of them generalize these two predominant exploring services in DLs. In the next section, we will show that related works like those above can be viewed as cases of our theoretical approach. We first formalize the DL exploring services in the context of a DL theory, 5S [9, 10]; then we prove that, when certain conditions are met, searching and browsing are duals; thus, mapping or conversions between them are readily supported.

3. EXPLORING SERVICE FORMALIZATION

Notation:

Let C be a collection (a collection is a set of digital objects; see Def. 17 in [10] for details) and 2^C be the set of all subsets of C.

Let $HT = (H, Contents, P)$ be a hypertext, where

1) $H=((V_H, E_H), L_H, F_H)$ is a structure (i.e., a directed graph with vertices V_H and edges E_H, along with labels L_H and labeling function F_H on the graph; see Def. 2 in [10] for details)

2) $Contents \subseteq C \bigcup AllSubStreams \bigcup AllSubStructuredStreams$ can include digital objects of a collection C, all of their (sub)streams (a stream is a sequence whose codomain is a nonempty set; see Def. 1 in [10]) and all possible restrictions of the *StructuredStream* (see Def. 15. in [10] for details) functions of digital objects.

3) $P: V_H \rightarrow 2^{Contents}$ is a function which associates a node of the hypertext with the node content. Note that the range of P is $2^{Contents}$ instead of *Contents* as defined in Def. 22 in [10].

If $subC \in 2^C$ is not an empty set, *subC* can be partitioned into a set of (non)overlapping clusters (groups) $\{cluster_1, cluster_2, ..., cluster_i, ..., cluster_k\}$, where $cluster_i$ is denoted as a cluster belonging to *subC*, and $\bigcup_{i=1}^{k} cluster_i = subC$.

Contents of *subC* is denoted by $CluCon(subC) = \{cluCon_1, cluCon_2, ..., cluCon_i, ..., cluCon_k\}$, where $cluCon_i$ is the contents (contents can include digital objects of a collection and all of their streams; see Def. 22 in [10] for details) associated with $cluster_i$.

Let *VSpa* be a vector space and *Base* be a set of basis vectors in *VSpa*. Let *{VisualM}* be a set of visual marks (e.g., points, lines, areas, volumes, glyphs) and *{VisualMP}* be a set of visual properties (e.g., position, size, length, angle, slope, color, gray scale, texture, shape, animation, blink, motion) of visual marks.

- Definition 1: Let $Q = \{(H_q, Contents_q, P_q)\}$ be a set of conceptual representations for user information needs, where $H_q=((V_q, E_q), L_q, F_q)$ is a structure (i.e., a directed graph with vertices V_q and edges E_q, along with labels L_q and a labeling function F_q on the graph; see Def. 2 in [10] for details), $Contents_q$ includes digital objects and all of their streams, and P_q is a mapping function P_q: $V_q \rightarrow Contents_q$.
- Definition 2: An **Exploration Space** (***ESpa***) is a space (see Def. 4 in [10] for details). ***ESpa*** $= (Q, Contents, OP\_Set)$, where Q is a set of conceptual representations for user information needs (see Def. 1), *Contents* includes digital objects of a collection *C* and all of their streams, and *OP_Set* is a set of operations on *Q* and *Contents*. $\{OP_s, OP_b, OP_{clu}, OP_{viz}\} \subseteq OP\_Set$, where *s*, *b*, *clu*, and *viz* relate to search, browse, cluster, and visualization operations, respectively, and

  1) OP_s: $(Q \times C) \times Sim_s \rightarrow 2^{Contents}$, where

  $Sim_s = \{OP_q(q, do) | q \in Q, do \in C\}$, where OP_q: $Q \times C \rightarrow R$ is a matching function that associates a real number with $q \in Q$ and a digital object $do \in C$. The range of function OP_s is the *Contents* associated with collection *C*.

  2) OP_b: $E_H \rightarrow 2^{Contents}$ is a function which, given a link, retrieves the content of target node, where E_H is a set of edges of the digraph defined for a hypertext (see above).

  3) OP_{clu}: $(2^C \times 2^C) \times Sim_{clu} \rightarrow 2^{Contents}$, where

  $Sim_{clu}=\{OP_{clu1}(cluster_x, cluster_y) | cluster_x \in 2^C, cluster_y \in 2^C\}$, where OP_{clu1}: $2^C \times 2^C \rightarrow R$ is a matching function that associates a real number with a pair of subsets of *C*. Sim_{clu} is a set of numerical values measuring the similarity between each pair of subsets of *C*. The range of OP_{clu} is a set of the *Contents* associated with collection *C*.

  4) $OP_{viz} = (VisualMap_1, VisualMap_2, VisualMap_3)$, where

  $VisualMap_1$: $2^C \rightarrow VSpa$ associates a set of digital objects with a set of vectors;

  $VisualMap_2$: $2^C \rightarrow VisualM$ associates a set of digital objects with a visual mark;

  $VisualMap_3$: $Base \rightarrow VisualMP$ associates a basis vector with a visual property of a visual mark.

  A special case is that there is only one digital object *do* in the set. We give an example to illustrate OP_{viz} in such a situation. Example: Given a vector space *VSpa* of three dimensions, a document is mapped to a vector of three elements, i.e., its length, date published, and number of citations, by function $VisualMap_1$. It is mapped to a visual mark: a point in 2D space, by function $VisualMap_2$. The first two base vectors in *VSpa* are associated with the position of the point in 2D space, while the third base vector may be mapped to another visual property of the point, its gray scale (e.g., a document represented by a black point has more citations than a document represented by a gray point).
- Definition 3: An **Exploring Service** (***ESer***) is a set of scenarios $\{sc_1, ..., sc_n\}$ over an exploration space *ESpa*. Each scenario is a sequence of events. An event e_i is associated with one or more operations in *OP_Set* as defined in *ESpa*.

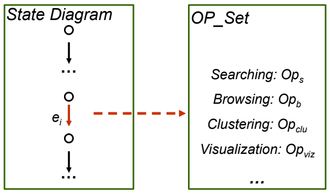


Fig. 1. Constructs of an exploring service

Fig. 1 shows two constructs of an exploring service. The left part of Fig. 1 is a state diagram, which consists of events. The dashed arrow means an event e_i has associated operations(s) in the set of operations, denoted by *OP_Set*. Characterized by its associated operations(s) in *ESpa*, an exploring service can be a searching, browsing, clustering, or visualization service as illustrated in the following theorems and lemmas according to Def. 1, Def. 2, and Def. 3.

- Theorem 1: If $\forall e_i$, the associated operation with e_i is OP_s, then an exploring service is a searching service.

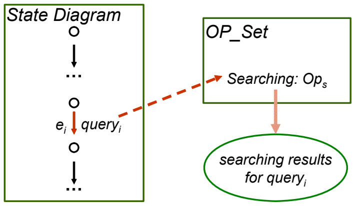


Fig. 2. An exploring service is a searching service.

The event e_i in Fig. 2 illustrates that a user issues a query $query_i$. The event then triggers operation OP_s, as indicated by the dashed arrow. The patterned arrow denotes the output of OP_s, i.e., searching results for $query_i$.

Proof: $q \in Q$, where Q is a set of conceptual representations for user information needs (see Def.1), there is a searching scenario having a final event of returning the matching function value $Sim_s = OP_q(q, do)$ for each digital object $do \in C$ and $\{OP_q((q, do), Sim_s)\}$, the contents of the retrieved digital objects for query *q*.

- Theorem 2: If $\forall v \in V_q$, $v \in V_H$, and $\forall e_i$, the associated operation with event e_i is OP_b, then an exploring service is a browsing service.

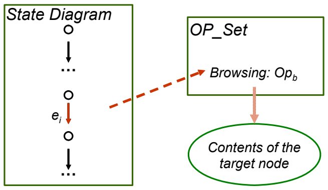


Fig. 3. An exploring service is a browsing service.

By Def. 23 of [10], a browsing service is associated with an underlying hypertext construct. Event e_i in Fig. 3 models a path through a website a user follows to access the target node. It invokes operation OP_b defined in Def. 2. The output of OP_b is the contents of the target node. A sequence of target nodes, v_{t_0}, v_{t_1}, …, v_{t_i}, …, v_{t_k}, associated with a sequence of events, e_0, e_1, …, e_i, …, e_k, is denoted as a user's navigation path π.

Proof: given a node v_s and a link (v_s, v_t), where $v_s, v_t \in V_q$ and $v_s, v_t \in V_H$, according to Def. 2, each link traversal event e_i is associated with a function OP_b: $E_H \rightarrow 2^{Contents}$, $OP_b(v_s, v_t)=P(v_t)$, and P is a function which associates a node of the hypertext with the node context, i.e., given a node v_s and a link (v_s, v_t) retrieves the contents of target node v_t. Therefore, the exploring service is a browsing service.

- Theorem 3: If $\forall\ e_i$, the associated operations with event e_i are OP_s followed by OP_{clu}, then an exploring service is a post retrieval clustering service.

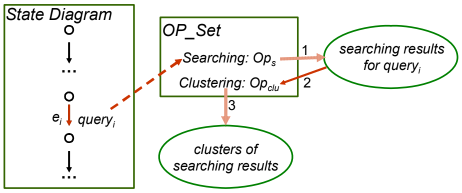


Fig. 4. An exploring service is post retrieval clustering.

The event e_i in Fig. 4 associates operation OP_s, as indicated by the one dashed arrow. The two patterned arrows (numbered 1 and 3) point to the output of OP_s and OP_{clu}, respectively. Searching results for $query_i$ are the input to OP_{clu}, (shown by the arrow numbered 2).

Proof: $\forall\ q \in Q$, there is a searching scenario returning C_{retr}, a set of retrieved digital objects, and a post retrieval clustering scenario having a final event of returning the matching function value $Sim_{clu} = OP_{clu1}(cluster_x\ ,\ cluster_y)$ for each pair of clusters and the contents of the clustering results $\{OP_{clu1}((cluster_x\ ,\ cluster_y),\ Sim_{clu})\}$, where $cluster_x$, $cluster_y \in C_{retr}$.

Note that if C_{retr}=C, then the exploration service also is a clustering service on a whole collection C.

- Lemma 1: Let $Espa_{browse}$=(Q_{browse}, $Contents_{browse}$, OP_Set_{browse}) be the exploration space of a browsing service $Eser_{browse}$, where $OP_b \in OP\_Set_{browse}$; let $Espa_{search}$= (Q_{search}, $Contents_{search}$, OP_Set_{search}) be the exploration space of a searching service $Eser_{search}$, where $OP_s \in OP\_Set_{search}$; let π be a user's navigation path, a sequence of target nodes consisting of v_{t_k-1} and v_{t_k} as the last two nodes; let $Л$ be a set of π, where π is a user's navigation path, a sequence of target nodes, v_{t_0}, v_{t_1}, …, v_{t_i}, …, v_{t_k}, associated with a sequence of events, e_0, e_1, …, e_i, …, e_k.

1) $Eser_{browse}$ can be converted to $Eser_{search}$, denoted $Eser_{browse} \Rightarrow Eser_{search}$, if

$\exists M_1$: $Л \rightarrow Q_{search}$, such that $\forall\ \pi \in Л$, $M_1(\pi) = q \in Q_{search}$, and $OP_b(v_{t\_k-1}, v_{t\_k})= P(v_{t\_k})= OP_s(q)$, where $P(v_{t\_k})$ is the contents associated with the last target node v_{t_k} and $OP_s\ (q)$ is the content associated with retrieved digital objects for query $q \in Q_{search}$.

2) $Eser_{search}$ can be converted to $Eser_{browse}$, denoted $Eser_{search} \Rightarrow Eser_{browse}$, if

$\exists M_2$: $Q_{search} \rightarrow Л$, such that $\forall\ q \in Q_{search}$, $M_2(q)= \pi \in Л$, and $OP_b\ (v_{t\_k-1}, v_{t\_k})= P(v_{t\_k})= OP_s(q)$, where $P(v_{t\_k})$ is the contents associated with the last target node v_{t_k} and $OP_s\ (q)$ is the content associated with retrieved digital objects for query $q \in Q_{search}$

Proof:

1) $\forall\ \pi \in Л$, $M_1(\pi) = q \in Q_{search}$, and the results of the operations associated with each link traversal event are the contents of retrieved digital objects for query q. Therefore, $Eser_{browse} \Rightarrow Eser_{search}$.

2) $\forall\ q \in Q_{search}$, $M_2(q) = \pi \in Л$, and the results of the operations associated with the event of issuing query q are the contents of the last target node v_{t_k} in the user's navigation path π. Therefore, $Eser_{search} \Rightarrow Eser_{browse}$.

- Lemma 2: Given Q_{search}=$\{q_1, q_2, ..., q_n\}$, $Л$=$\{\pi_1, \pi_2, ..., \pi_n\}$, where π_i is a user's navigation path, a sequence of target nodes consisting of $v_{i\_t\_k-1}$ and $v_{i\_t\_k}$ as the last two nodes, $OP_s(q_i)$=$OP_b(v_{i\_t\_k-1}, v_{i\_t\_k})$=$contents_i \in 2^{Contents}$ (see Def. 2), $OP_s^{-1}\ (contents_i)$=q_i, and $OP_b^{-1}\ (contents_i)$=π_i, then $\exists M_1$, $\exists M_2$, $Eser_{browse} \Rightarrow Eser_{search}$, and $Eser_{search} \Rightarrow Eser_{browse}$.

Proof:

1) $\exists M_1$, $\forall\ \pi_i \in Л$, $M_1(\pi_i) = OP_s^{-1}\ (OP_b(v_{i\_t\_k-1}, v_{i\_t\_k}))= OP_s^{-1}\ (contents_i)= q_i$, therefore, according to Lemma 1, $\exists M_1$: $Л \rightarrow Q_{search}$ and $Eser_{browse} \Rightarrow Eser_{search}$.

2) $\exists M_2$, $\forall\ q \in Q_{search}$, $M_2(q_i)= OP_b^{-1}\ (OP_s(q_i)) = OP_b^{-1}\ (contents_i)= \pi_i$, therefore, according to Lemma 1, $\exists M_2$: $Q_{search} \rightarrow Л$ and $Eser_{search} \Rightarrow Eser_{browse}$.

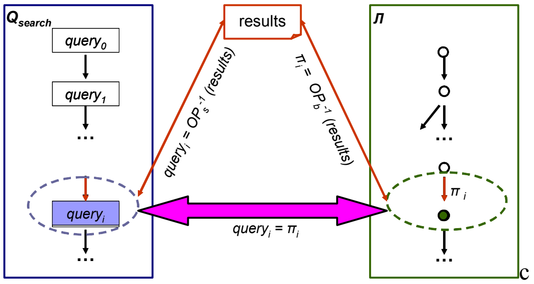


Fig. 5. "$query_i$" and "e_i" are associated with the same results.

As shown in Fig. 5, both "$query_i$" and "π_i" are associated with the same results, therefore, $\exists M_1$: $M_1(query_i)= \pi_i$, $\exists M_2$: $M_2(\pi_i)= query_i$, $Eser_{browse} \Rightarrow Eser_{search}$ and $Eser_{search} \Rightarrow Eser_{browse}$.

PESTO [4], DataWeb [13], and MIX [14] are cases where browsing can be converted to searching. Because of PESTO's "query-in-place" paradigm, DataWeb's hierarchically browsing, and MIX's navigation commands of the standard DOM API, the navigation paths of each of them can be mapped to queries. Therefore $Eser_{browse} \Rightarrow Eser_{search}$.

- Lemma 3: Let $Espa_{postBrowse}= (Q_{postBrowse}, Contents_{postBrowse}, OP\_Set_{postBrowse})$ be the exploration space of an exploring service $Eser_{postBrowse}$ occurring ***after*** $Eser_{browse}$, where $Contents_{postBrowse}=OP_b(v_{t\_i-1}, v_{t\_i})$ is the contents associated with edge (v_{t_i-1}, v_{t_i}), v_{t_i-1} and v_{t_i} are the last two nodes of a user's navigation path $\pi_i \in \Pi$ in $Eser_{browse}$, $C_{postBrowse}$ is a set of digital objects associated with $Contents_{postBrowse}$, and $OP_b \in OP\_Set_{postBrowse}$. According to Theorem 1, $Eser_{postBrowse}$ is a searching service (i.e., browsing service $Eser_{browse}$ leads to searching service $Eser_{postBrowse}$), if $\exists OP_s$: $(Q_{postBrowse} \times C_{postBrowse}) \times Sim_s \rightarrow 2^{Contents\ postBrowse}$, where $Sim_s= \{OP_q(q, do) \mid q \in Q_{postBrowse}, do \in C_{postBrowse}\}$, where OP_q: $Q_{postBrowse} \times C_{postBrowse} \rightarrow R$ is a matching function that associates a real number with $q \in Q_{postBrowse}$ and a digital object $do \in C_{postBrowse}$.

  Proof: $\forall q \in Q_{postBrowse}$, $\{OP_s((q, do), Sim_s)\}$ is the contents of the retrieved digital objects for query q, where $Sim_s= OP_q(q, do)$, therefore, by Theorem 1, $Eser_{postBrowse}$ is a searching service.

  The switch from browsing to searching in PESTO [4], DataWeb [13], and MIX [14] can be generalized as shown in Fig. 6. The arrow numbered 1 points to the browsing results associated with the link traversal event π_i. Since π_i and $query_i$ can be mapped to each other in these systems as discussed before (indicated by the arrow numbered 3), they are associated with the same results, $Contents_{postBrowse}$. Therefore, the arrow numbered 2 also points to $Contents_{postBrowse}$. After browsing, a user searches $Contents_{postBrowse}$ for a new query $query_{i+1}$. Searching results for $query_{i+1}$ then are a subset of $Contents_{postBrowse}$. This is illustrated as the circle and pointed to by the arrow numbered 4 in Fig. 6. Therefore, $query_{i+1}$ is a new query refined from query$_i$ as indicated by the arrow numbered 5. So switching from browsing to searching in this situation is a query refining or expansion process.

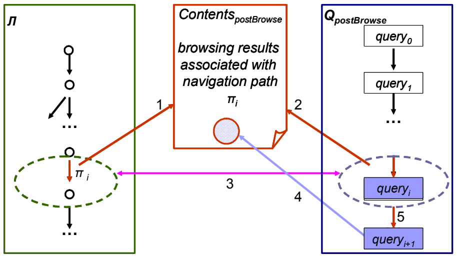


Fig. 6. *query*$_{i+1}$ is refined from *query*$_i$ after browsing.

- Lemma 4: Let $Espa_{postRetr} = (Q_{postRetr}, Contents_{postRetr}, OP\_Set_{postRetr})$ be the exploration space of an exploring service $Eser_{postRetr}$ occurring **after** $Eser_{search}$, where $Q_{postRetr} = \{((V_{postRetr}, E_{postRetr}), L_{postRetr}, F_{postRetr}\}$ (see Def. 1), $Contents_{postRetr}$ is associated with C_{retr}, a set of retrieved digital objects for query $q \in Q_{search}$ in $Eser_{search}$. According to Theorem 2, Lemma 1, and Lemma 2, $Espa_{postRetr}$ is a browsing service (i.e., searching service $Eser_{search}$ leads to browsing service $Espa_{postRetr}$), if $OP\_Set_{postRetr} = \{OP_s, OP_{clu}\}$, $cluCon_{retr} = \{OP_{clu}((cluster_x, cluster_y), sim_{clu}) \mid cluster_x, cluster_y \subseteq C_{retr}\} =\{cluCon_{retr\_1}, cluCon_{retr\_2}, \ldots, cluCon_{retr\_i}, \ldots, cluCon_{retr\_z}\}$ is the contents of clustered retrieved results, where $sim_{clu} = OP_{clu1}(cluster_x, cluster_y)$ (see Def. 2), $\Pi = \{\pi_1, \pi_2, \ldots, \pi_i, \ldots, \pi_z\}$, where $\pi_i= (v_0, v_i)$ is a navigation path consisting of only two nodes, v_0, $v_i \in V_{postRetr}$, and $\Pi \rightarrow cluCon_{retr}$.

  The event e_i of issuing $query_i$ triggers the operation OP_s, as indicated by the dashed arrow numbered 1 in Fig. 7. The patterned arrow numbered 2 denotes the output of OP_s, i.e., $Contents_{postRetr}$ (searching results for $query_i$). OP_{clu} takes $Contents_{postRetr}$ as input and yields as output the contents of clusters as shown by the arrows numbered 3 and 4. The arrow numbered 5 represents the mapping from each navigation path to the contents of a cluster. Therefore, the contents of the last target nodes of these navigation paths are the contents of clusters and the mapping function $M_{b\_cluster}$ can be viewed to be OP_b for browsing.

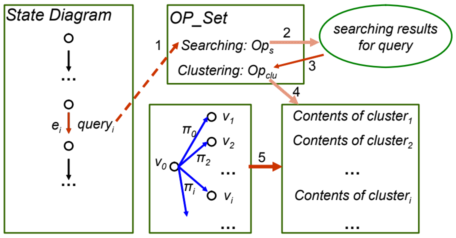


Fig. 7. Switch from browsing to searching.

  Proof: $\forall v \in V_{postRetr}$, $v \in V_H$, and $\forall e_i$, the associated operation with event e_i is $OP_b((v_0, v_i))=M_{b\_cluster}(\pi_i)= cluCon_{retr\_i}$, where v_i is the target node of π_i, therefore by Theorem 2, $Eser_{postRetr}$ is a browsing service.

  Categorizing or clustering searching results is a case of switching searching to browsing. ScentTrails [16] can be viewed as a special case as $|cluCon_{retr}|=1$, i.e., each cluster is a singleton having one item from the retrieved result list.

- Theorem 4: If $\forall e_i$, the associated operations with e_i are OP_s followed by OP_{viz}, then an exploring service is a post retrieval visualization service.

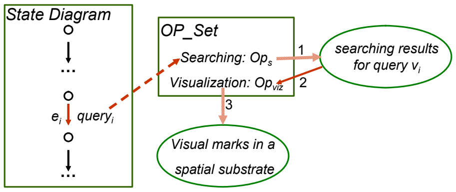


Fig. 8. An exploring service is visualization service.

The event e_i in Fig. 8 associates operation OP_s and OP_{viz}, as indicated by the dashed arrow. The two patterned arrows (numbered 1 and 3, respectively) point to the output of OP_s and OP_{viz}, respectively. Searching results for $query_i$ are the input to OP_{viz} (shown by the arrow numbered 2)

Proof: $\forall q \in Q$, there is a searching scenario returning a set of retrieved digital objects C_{retr} and a post retrieval visualization scenario having a final event of visually mapping a set of digital objects (or each digital object) of C_{retr}

to a visual mark with visual properties in a spatial substrate of *n* dimensions.

If n=2, it is 2-D visualization; if n=3, it is 3-D visualization. If $C_{retr}=C$, the exploring service also is a visualization service for a whole collection. If $\exists M_2(q)$, the exploring service is a visualization service for browsing. Vector graphics and raster display are two different types of display used for representation. Virtually all modern computer video displays translate vector representations to a raster format.

4. Exploring Services in ETANA-DL

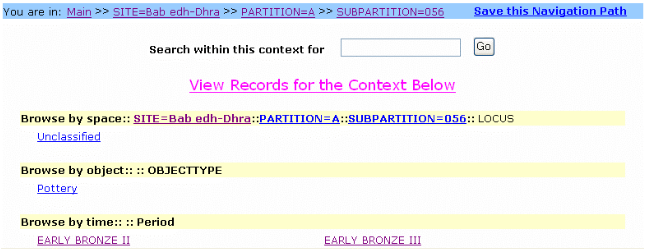
Our theory-based approach to describing DL exploring services allows us to understand browsing and searching in a new way. It guides us to design and implement exploring services for an archaeological DL, ETANA-DL [17, 18]. ETANA-DL is an integrated archaeological DL supporting integration of a number of (ETANA) sites in the Near East. It integrates searching and browsing, allowing users to browse at will and shift between browsing and searching seamlessly. It also provides a visual interface applying data analysis and information visualization techniques to help archaeologists test hypotheses and extend the understanding of past (material) cultures and environments.

In this section, we first introduce a multi-dimensional browsing service, which can actually be considered as a searching service according to Lemma 2. We then illustrate how ETANA-DL combines browsing and searching in two ways. The first way extends and empowers the multi-dimensional browsing. It can be viewed as query refining and extension based on Lemma 3. Organizing searching results hierarchically is the second way. Both ways allow seamless transition between browsing and searching, as suggested by Lemma 4. We finally describe the visualization service, which integrates browsing and searching into a single visual interface, as suggested by Theorem 4.

4.1 Multi-dimensional browsing

Multi-dimensional browsing allows users to move along any of the navigational dimensions, or a combination thereof. By navigational dimension we mean a hierarchical structure used to browse digital objects. Digital objects in ETANA-DL are various archaeological data, e.g., figurine images, bone records, locus sheets, and site plans. They are organized by different hierarchical structures (e.g., animal bone records are organized based on sites where they are excavated, temporal sequence, and animal names). These hierarchical structures contain one or more hierarchically arranged categories that are determined by the elements of the global schema of ETANA-DL. In addition to this, they can be refined based on taxonomies existing in botany and zoology, or from classification and description of artifacts by archaeologists.

Typical DLs provide a directory-style browsing interface (as in Yahoo! or Open Directory), with levels in the hierarchy displayed as clickable category names and DL items in that category shown below. Though some DLs (such as CITIDEL [7]) allow users to browse through several dimensions, they are limited in that users cannot navigate through all dimensions simultaneously, or across different dimensions.



You are in: Main >> SITE=Bab edh-Dhra >> PARTITION=A >> SUBPARTITION=056 **Save this Navigation Path**

Search within this context for Go

View Records for the Context Below

Browse by space:: SITE=Bab edh-Dhra::PARTITION=A::SUBPARTITION=056:: LOCUS
Unclassified

Browse by object:: :: OBJECTTYPE
Pottery

Browse by time:: :: Period
EARLY BRONZE II EARLY BRONZE III

Fig. 9. Multi-dimensional browsing interface

In ETANA-DL, a user can browse through three dimensions: space, object, and time. She can start from any of these dimensions and move along by clicking. The scenario shown in Fig. 9 tells that she is interested in the artifact records from the tomb numbered 056 in area A of the Bab edh-Dhra site. The clickstream representing her navigation path is denoted 'Site=Bab edh-Dhra >> PARTITION=A >> SUPARTITION=056'. While the navigation path is within the first dimension, it is associated with the other dimensions. The second dimension shows there is only one type of object, i.e., pottery, from that particular location. The third dimension presents the two time periods associated with those pottery records. Hence, the dynamic coverage and hierarchical structure of those dimensions yields a learning and exploration tool. The user can navigate across dimensions. By clicking "EARLY BRONZE II" in the third dimension, she can view all the interesting artifact records from the EARLY BRONZE II period. Her current navigation path (see the top of Fig. 10) can be saved for later use. It can be considered as a surrogate for a query for the records in that particular location and time period. Therefore, according to Lemma 2, the multi-dimensional browsing service can be viewed as searching, i.e., browsing service $\Rightarrow$ searching service.

Tomb #056 in Area A of Bab edh-Dhra, Time Period: EARLY BRONZE III

Save navigation path

You are in: Main >> SITE=Bab edh-Dhra >> PARTITION=A >> SUBPARTITION=056 >> Period=EARLY BRONZE III **Save this Navigation Path**

Search within this context for Go

View Records for the Context Below

View Records

Browse by space:: SITE=Bab edh-Dhra::PARTITION=A::SUBPARTITION=056:: LOCUS
Unclassified

Browse by object:: :: OBJECTTYPE
Pottery

Browse by time:: Period=EARLY BRONZE III:: Chronology
No SubCategories Present

Showing **1-1** out of 1 records
Page 1

☐ Bab edh-Dhra **Vessel Number** 029 Tomb Area A Tomb Number 056
Ages EARLY BRONZE III **Basic Category** Small bowls and Saucers
Rim Treatment unavailable **Handle Type** unavailabe **Mouth Width** 104 **Base Width** 44
[View complete record] [Add to Items of Interest] [Share Item]

Fig. 10. Save current navigation path for later use and view records

4.2 Browsing and Searching Integration

4.2.1 Search within browsing context

Searching within a browsing context blends querying and browsing and is reminiscent of IBM's PESTO GUI for "in-place querying" [4]. The main idea is that browsing will present a useful starting point for active exploration of an answer space. Subsequent browsing and searching is employed to refine or enhance users' initial, possibly under-specified, information needs.


You are in: Main >> SITE=Bab edh-Dhra >> PARTITION=A >> SUBPARTITION=056 **Save this Navigation Path**

Search within this context for saucer Go

Search Saucer Records within the Context

View Records for the Context Below

Browse by space:: SITE=Bab edh-Dhra::PARTITION=A::SUBPARTITION=056:: LOCUS

Unclassified

Browse by object:: :: OBJECTTYPE

Pottery

Browse by time:: :: Period

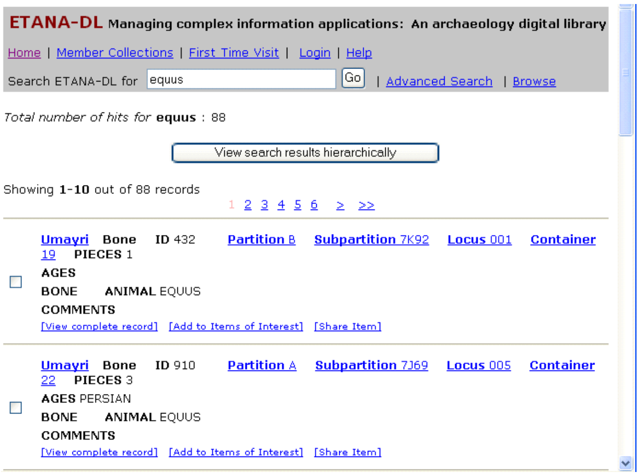
EARLY BRONZE II | EARLY BRONZE III

Fig. 11. Search saucer records

Browsing context is associated with a user's navigation path. Browsing results within a certain browsing context are defined as a set of records (web pages), e.g., there are 35 pottery records within the browsing context represented by the navigation path 'Site=Bab edh-Dhra >> PARTITION=A >> SUPARTITION=056'. Assume a user wants to find saucer records in the set of 35 pottery records. She types "saucer" in the search box as shown in Fig. 11. According to Lemma 3, she switches from browsing to searching, and searching then is a natural extension of browsing. Since the navigation path is a surrogate of a query, searching within a browsing context can be viewed as query refinement.

4.2.2 *Organize searching results hierarchically*

Eighty eight equus records are retrieved through the basic searching service (see a query named "equus" in Fig. 12). They are organized into three dimensions after the user clicks the button "View search results hierarchically" (see Fig. 13). The user starts browsing and then selects "Nimrin" in the first category to view the records. Thirty six equus records are displayed as shown in Fig. 14. According to Lemma 4, she switches from searching to browsing. During the next exploring stage of browsing, she can search as illustrated in section 4.2.1. Therefore, she switches seamlessly between browsing and searching, to specify her information needs.


ETANA-DL **Managing complex information applications: An archaeology digital library**

Home | Member Collections | First Time Visit | Login | Help

Search ETANA-DL for equus Go | Advanced Search | Browse

Total number of hits for **equus** : 88

View search results hierarchically

Showing **1-10** out of 88 records

1 2 3 4 5 6 > >>

☐ **Umayri** **Bone** **ID** 432 **Partition** B **Subpartition** 7K92 **Locus** 001 **Container** 19 **PIECES** 1
AGES
BONE **ANIMAL** EQUUS
COMMENTS
[View complete record] [Add to Items of Interest] [Share Item]

☐ **Umayri** **Bone** **ID** 910 **Partition** A **Subpartition** 7J69 **Locus** 005 **Container** 22 **PIECES** 3
AGES PERSIAN
BONE **ANIMAL** EQUUS
COMMENTS
[View complete record] [Add to Items of Interest] [Share Item]

Fig. 12. Equus records are retrieved through basic searching


Search within this context for Go

View Records for the Context Below

Browse by space>> SITE

Nimrin | Umayri

Browse by object>> OBJECTTYPE

Bone

Browse by time>> Period

Iron II	Islamic	Late Hellenistic-Roman	Persian	Middle Bronze
Iron I	Byzantine	Early Bronze	Late Iron II	Late Iron II / Persian (Iron III)
Middle Bronze IIC	Early Bronze III	Early Iron I	Classical-Islamic	Ottoman - Modern
Middle Bronze II	Late Bronze			

Fig. 13. Retrieved equus records are organized into three dimensions


You are in: >> query=equus >> SITE=Nimrin

Showing **1-10** out of 36 records

Page 1 2 3 4

☐ **Nimrin** **Bone** **ID** 1472 **Partition** NW **Subpartition** N40/W25 **Locus** 184 **Container** 252 **PIECES** 1
AGES IRON II **AGE** 900-800 BC
BONE TOOTH **ANIMAL** EQUUS
COMMENTS
[View complete record] [Add to Items of Interest] [Share Item]

Fig. 14. Browse the 36 equus records from the Nimrin site after searching

4.3 Visualization

While the searching and browsing services provided by ETANA-DL allow users to access primary archaeological data, their help with comprehending specific archaeological DL phenomena is limited when vast quantities of data are harvested into ETANA-DL. Fortunately, visual interfaces to DLs enable powerful data analysis and information visualization techniques to help archaeologists test hypotheses and extend the understanding of past (material) cultures and environments. Data generated from the sites' interpretation then provides a basis for future work, including publication, museum displays, and, in due course, input into future project planning. Thus, we developed EtanaGIS and EtanaViz to support visually exploring archaeological DLs. EtanaGIS allows integration of Geographic Information System (GIS) data for related archaeological sites into ETANA-DL. It provides a web-based GIS portal to allow users to spatially explore ETANA-DL. Details of EtanaGIS can be found at http://etana.dlib.vt.edu/~etana/Viz/EtanaGIS.pdf.

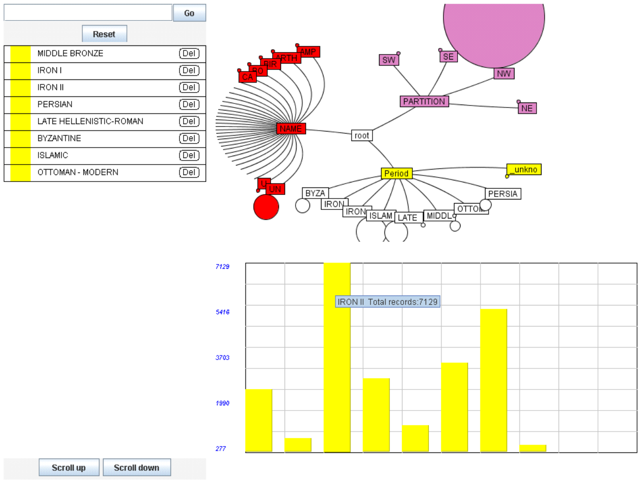


Fig. 15. Total number of animal bones across Nimrin culture phrases

In this section, we focus on EtanaViz. It integrates searching, browsing, clustering, and visualization into a single interface. The top left of the screen (see Fig. 15) is a query box. On the top right is a hyperbolic tree showing hierarchical relationships among excavation data based on spatial, temporal, and artifact-related taxonomies. A node name represents a category, and a bubble attached to a node represents a set of archaeological records. The size of a bubble attached to a node reflects the number of records belonging to that category. The hyperbolic tree supports "focus + context" navigation; it also provides an overview of records organized in ETANA-DL.

According to Def. 1, a cluster (group) of records is associated with a vector of two elements, i.e., name and size of the cluster; a cluster is mapped to a visual mark: bubble (circle); the name and size of the cluster are mapped to two visual properties: label and size of the bubble, respectively.

EtanaViz supports exploring to gain insights, as is illustrated in the following example scenarios.

A user is interested in excavated animal bones from site Nimrin, located in the Jordan Valley. She inputs query "SITE=Nimrin&OBJECTTYPE=Bone". The results are displayed as a hyperbolic tree, as illustrated in Fig. 16. All excavation bone records are grouped into cultural phases (time periods). They are Middle Bronze, Iron I, Iron II, Persian, Late Hellenistic/Roman, Byzantine, Islamic, and Ottoman-Modern. The records also are classified by archaeological site organization and animal categories. The user wants to know the number of bone records for each period. She left clicks a node labeled "MIDDLE BRONZE" in the hyperbolic tree and selects the "add to compare…" option to view total bones throughout the Middle Bronze Age. This causes a bar to be displayed in a chart below the hyperbolic tree and an entry to be listed on the left. She continues to add more bars to view bones throughout the entire time sequence of Tell Nimrin occupation. When she moves the mouse over a bar, a tool tip shows the number of animal bones for the corresponding culture phrase.

She continues navigating the hyperbolic tree. She left clicks a node labeled "SUS" and selects the "add to view distribution…" option. She then left clicks the "BOS", "CAPRA", and "OVIS" nodes to show how they constitute the identified animal bones in each culture phrase. Eight stacked bars representing percentages of those bones are displayed, and four entries with different colors are included in the list on the left (see Fig. 16).

The color of the entry can be changed to help distinguish different categories. It is always synchronized with the color in the stacked bars. The red bars (at the bottom of the stacked bars), representing sus (pig) bones, show that sus constitute 4.71% of the Middle Bronze Age faunal assemblage, but less than 1% at the beginning of the Iron Age. The user is wondering why the percentage for pig bones drops dramatically over time at Tell Nimrin. She may hypothesize that the reasons are probably twofold: 1) the introduction of religious taboos against eating pork, and 2) increased demand for clean water sources as human populations grew at Nimrin [19].

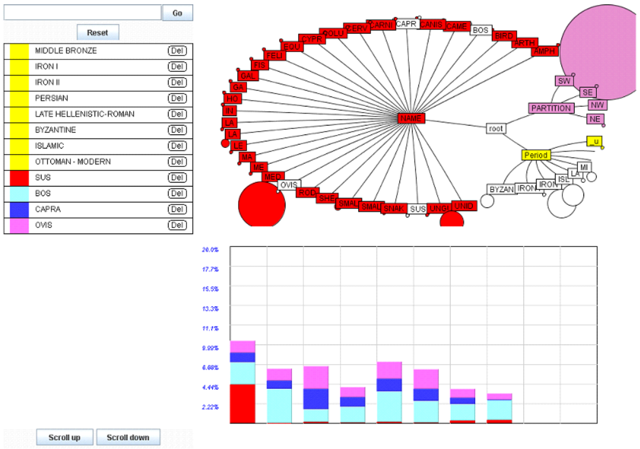


Fig. 16. Percentages of animal bones across Nimrin culture phrases

Light blue bars (on top of the red sus bars) represent bos (cattle) bones percentages. Two light blue bars are higher than the others. They correspond to the Iron II and Late Hellenistic/ Roman culture phrases. The user, considering that cattle figure most prominently during these periods, may suggest improved grazing conditions in the Jordan Valley during that time.

Pink bars and blue bars (the top two of the stacked bars) represent ovis (sheep) bones and capra (goat) bones, respectively. Pink bars are slightly higher than blue bars. This means that ovis bones slightly outnumber capra bones across culture phrases of Tell Nimrin. This would suggest that past environmental conditions in the Jordan Valley provided enhanced forage for sheep while goats would have been employed as browsers on drier vegetation. Relatively stable percentages of slightly higher sheep populations versus those of goats may indicate that favorable environmental conditions, and environmental or cultural desertification, did not greatly impact the agrarian way of life at Tell Nimrin on the banks of the Jordan, over time [19].

The user may be interested in animal bones excavated from other sites. By repeating the interaction with EtanaViz, as described above, she starts to analyze animal bones excavated from the Umayri site. She also can make inter-site comparisons.

5. EVALUATION

In fall 2005, we conducted a formative user evaluation for ETANA-DL. Many findings from the usability study are already influencing the iterative design and implementation of ETANA-DL to achieve our usability goals. In this section, instead of listing all findings, we focus only on those that help validate the hypotheses related to browsing, searching, and visualization.

5.1 Evaluation Methods and Procedure

Twenty eight graduate students from the computer science department at Virginia Tech participated in the evaluation experiment, which was posted with instructions online at http://etana.vt.edu:8080/etana/servlet/surveyTasks?submit_start. The experiment was conducted through four sessions. Each user was required to

1) learn the online tutorial of ETANA-DL;

2) complete a pre-evaluation questionnaire;

3) perform tasks using ETANA-DL. After completion of each task, he(she) was asked to fill out a task-related questionnaire and give comments.

4) provide subjective reactions using post-evaluation survey forms.

Users' interactions with ETANA-DL were logged. Time to com-plete, and the error rate for each task, were measured automa-tically. Upon completion of all tasks, users were asked to rate the exploring services on a 5-point scale, where 1= poor, and 5=excellent. Our reason for measuring users' impressions about ETANA-DL services (five of them are listed in Table 1) stems from the following two pre-experimental hypotheses:

- Users significantly prefer integrated browsing and searching to browsing.
- Users significantly prefer integrated browsing and searching to searching.

5.2 Results and Discussion

The median values for measuring users' impressions regarding five of the ETANA-DL services are shown in Table 1. Browsing, searching, and EtanaViz received four points on a 5-point scale, while searching within browsing context (abbreviated as SWBC) and saving navigation path (abbreviated as SNP) services received 4.5. Users commented that they appreciated SWBC and SNP because "SWBC is simple enough to understand and an excellent way of narrowing down a search…browsing through the different levels can be time consuming, so if we know that we will want to go to a given context a lot, it is useful to just be able to click on a link of SNP to get back to our context of interest…"

Table 1. Impression about ETANA-DL services (mean value)

Browse	Search	EtanaViz	Save navigation path (SNP)	Search within browsing context (SWBC)
4.0	4.0	4.0	4.5	4.5

We also did *t*-tests on the following four hypotheses.

- H_1: Mean value of impression about SWBC is larger than that for browsing at significance level 0.05.
- H_1: Mean value of impression about SWBC is larger than that for searching at significance level 0.05.
- H_1: Mean value of impression about SNP is larger than that for browsing at significance level 0.05.
- H_1: Mean value of impression about SNP is larger than that for searching at significance level 0.05.

The above four hypotheses were all accepted. The first two accepted hypotheses are associated with the two pre-experimental hypotheses mentioned above, i.e., users significantly prefer integrated browsing and searching to browsing (or to searching). These results agree with comments found in the related literature as described in section 2. To probe the last two hypotheses, we analyzed four of the 17 tasks performed by users. For four tasks, users were asked to give the number of records retrieved, for specific information needs. The followings are those four tasks.

1. Use browsing to give the total number of pottery records excavated from tomb 007 in area A of site Bab edh-Dhra.
2. Use searching to tell how many equus bones are from the Umayri site.
3. Use browsing to tell how many equus bones are from the Nimrin site.
4. Use saved navigation paths to give the total number of pottery records excavated from tomb 056 in area A of the Bab edh-Dhra site.

Fig. 17 shows the average time for each of the four tasks.

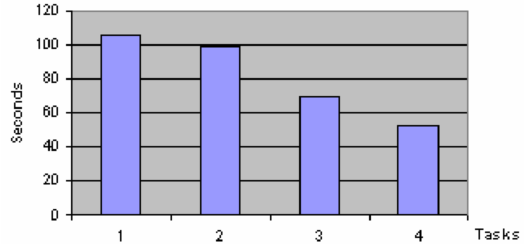


Fig. 17. Average time on tasks

Task 4 was completed significantly faster than either task 1 or task 2, at significance level 0.05. This showed that reusing saved navigation paths really improves users' performance. It saved users time during exploration. While similar information needs (e.g., task 1, 2, and 4) can be achieved through different ways (browsing, searching, or SNP), SNP keeps track of users' navigation history and helps reduce time and effort to achieve information seeking goals.

We expected that users would complete task 4 significantly faster than task 3. We also thought users would spend about the same time to complete similar tasks, i.e., 2 and 3. However, our experimental results were somewhat surprising in that the average time on task 4 was not significantly less than that on task 3, and the average time on tasks 2 and 3 was different. We did some follow-up interviews to probe the reasons. Our log file indicated that one user spent more than five minutes to complete task 4. We found that he was disconnected during the online experiment for task 4. Though tasks 2 and 3 have similar information needs, users found it was difficult to find appropriate keywords to complete task 2, therefore, more time was needed to try more queries. We believe that since users gained experience and developed a searching strategy when doing task 2, they completed task 3 faster than task 2 (since task 3 was performed after task 2).

Because our new service to organize searching results hierarchically was not implemented before we conducted the evaluation, we cannot yet report data about its efficiency and effectiveness. However, there is already evidence that information access is improved by posting search hits against an interactive tree structure [6].

6. CONCLUSIONS

To the best of our knowledge, we are the first to approach DL exploring services based on a DL theory. Studying DL exploring from this viewpoint has provided several insights. For instance, the formalisms bring a theoretical approach to the subject and the theorems we developed indicate browsing and searching can be converted and switched to each other under certain conditions. In addition, the theoretical approach provides a systematic and functional method to design and implement DL exploring services. An integrated archaeological DL, ETANA-DL, was used as a case study, where blending browsing and searching is achieved precisely. A formative user study to evaluate ETANA-DL exploring services was conducted. Some results were reported in this paper to validate our hypotheses.

We think our work has made contributions to aid both users and developers of DLs. For users, fluidity between browsing and searching supports them in achieving their information-seeking goals, thus helps bridge their mental model of an information space with the information systems' representation. For DL developers, we suggest some new possibilities for blurring the dividing line between browsing and searching. If these two services are not considered to have different underlying mechanisms, they will not be provided as separate functions in DLs, and may be better integrated.

7. ACKNOWLEDGMENTS

This work was funded in part by the National Science Foundation (ITR-0325579). Ricardo Torres is funded in part by CNPq, CAPES, FAPESP, FAEPEX, CNPq WebMaps, AgroFlow projects, and by the Microsoft Escience grant. We thank Douglas Clark, Joanne Eustis, James W. Flanagan, Paul Gherman, Andrew Graham, Larry Herr, Paul Jacobs, Douglas Knight, Oystein LaBianca, David McCreery, and Randall Younker for their support. We also thank all our colleagues in the Digital Library Research Laboratory at Virginia Tech.

8. REFERENCES

[1] Baldonado, M.Q.W. A User-Centered Interface for Information Exploration in a Heterogeneous Digital Library. *JASIS*, *51* (3). 297-310, 2000.

[2] Belkin, N.J. Anomalous states of knowledge as the basis for information retrieval. *Canadian Journal of Inf. Sci.*, *5*. 133–143, 1980.

[3] Belkin, N.J., Marchetti, P.G. and Cool, C. BRAQUE: Design of an interface to support user interaction in information retrieval. *Information Processing & Management*, *29* (3). 325–344, 1993.

[4] Carey, M., Haas, L., Maganty, V. and J.Williams PESTO: An integrated query/browser for object databases. In Proc. VLDB, 1996.

[5] Croft, W.B. and Thompson, R.H. I3R: A New Approach to the Design of Document Retrieval Systems. *JASIS*, *38* (6). 389–404, 1987.

[6] Egan, D.E., Remde, J.R., Gomez, L.M., Landauer, T.K., Eberhardt, J. and Lochbaum, C.C. Formative design-evaluation of SuperBook. *ACM Trans. Inf. Syst*, *7* (1). 30-57, 1989.

[7] Fox, E.A., Neves, F.D., Yu, X., Shen, R., Kim, S. and Fan, W. Exploring the computing literature with visualization and stepping stones & pathways. *CACM 49(4): 52-58, 2006*.

[8] Furnas, G.W. and Rauch, S.J. Considerations for Information Environments and the NaviQue Workspace. In Proc. DL 98: 79-88.

[9] Gonçalves, M.A. Streams, Structures, Spaces, Scenarios, and Societies (5S): A Formal Digital Library Framework and Its Applications, Ph.D. Dissertation, Virginia Tech, http://scholar.lib.vt.edu/theses/available/etd-12052004-135923, 2004.

[10] Gonçalves, M.A., Fox, E.A., Watson, L.T. and Kipp, N.A. Streams, structures, spaces, scenarios, societies (5s): A formal model for digital libraries. ACM Trans. Inf. Syst. 22(2): 270-312 (2004).

[11] Manber, U., Smith, M. and Gopal, B. Webglimpse: Combining browsing and searching. In Proc. Usenix Technical Conference, 1997: 195-206.

[12] Marchionini, G. Information seeking in electronic environments. Cambridge, England: Cambridge University Press. 1995.

[13] Miller, R.J., Tsatalos, O.G. and Williams, J.H. DataWeb: Customizable Database Publishing for the Web. *IEEE Multimedia, 4* (4). 14-21, 1997.

[14] Mukhopadhyay, P. and Papakonstantinou, Y. Mixing Querying and Navigation in MIX. In Proc. ICDE 2002: 245-254.

[15] Munroe, K.D. and Papakonstantinou, Y. BBQ: A Visual Interface for Integrated Browsing and Querying of XML. In Proc. of VLDB 2000: 277-296.

[16] Olston, C. and Chi, E.H. ScentTrails: Integrating browsing and searching on the Web. *ACM Trans. Comput.-Hum. Interact*, *10* (3). 177-197, 2003.

[17] Ravindranathan, U. Prototyping Digital Libraries Handling Heterogeneous Data Sources - An ETANA-DL Case Study, Masters Thesis, Virginia Tech, http://scholar.lib.vt.edu/theses/available/etd-04262004-153555/, 2004.

[18] Shen, R., Gonçalves, M.A., Fan, W. and Fox, E.A. Requirements Gathering and Modeling of Domain-Specific Digital Libraries with the 5S Framework: An Archaeological Case Study with ETANA. In Proc. ECDL 2005: 1-12.

[19] West, D., Finnegan, M., Lane, R.W. and Kysar, D.A. Analysis of Faunal Remains Recovered from Tell Nimrin, Dead Sea Valley, Jordan, final report, 1996.

[20] Williams, M.D. What makes RABBIT run? *International Journal of Man-Machine Studies*, *21* (4). 333-352, 1984.

[21] Wittenburg, K. and Sigman, E. Integration of browsing, searching, and filtering in an applet for Web information access. In Proc. CHI 1997: 293-294.

[22] Zhang, J. and Marchionini, G. Evaluation and evolution of a browse and search interface: relation browser. In Proc. of the 2005 national conference on digital government research, 179-188.