InfoATV: enhancing PIM systems to support re-finding to (re)discovery

Charisse Madlock-Brown
Health Informatics
The University of Iowa
Iowa City, Iowa 52242
charisse-madlock-brown@uiowa.edu

Si-Chi Chin
Information Science
The University of Iowa
Iowa City, Iowa 52242
si-chi-chin@uiowa.edu

David Eichmann
Institute for Clinical and Translational Science
The University of Iowa
Iowa City, Iowa 52242
david-eichmann@uiowa.edu

ABSTRACT

Personal information management (PIM) systems provide a means to organize and re-find information. An effective PIM model would allow individuals to make intelligent use of the information they possess and, prospectively, create knowledge from the items in their collections. This paper introduces InfoATV, a system that associates users' personal collections and the MEDLINE database, allowing them to re-find and (re) discover research documents. To facilitate re-finding our system allows users to navigate through re-creations of previous search and re-finding sessions—improving upon existing systems by allowing them to view their flow of activity to help them re-find. InfoATV facilitates re-finding for long-term management and for session-based re-finding activities. We identify two characteristics necessary for a successful search interface that help provide additional contextual data for future locating. Additionally, the system mines a user's personal document collection for materials related to currently viewed documents and displays a network showing the ways in which those documents relate. This can assist users to, sometimes unexpectedly, (re)discover previously retrieved documents and reuse them for new purposes.

Categories and Subject Descriptors

H.3 [**Information Storage and Retrieval**]: Information Storage and Information Search and Retrieval – *file organization, clustering, retrieval models, and search process.*

General Terms

Management, Design, Human Factors.

Keywords

Personal Information Management, Information Visualization, MEDLINE.

1. INTRODUCTION

This paper presents a system designed to enhance the process of re-finding research documents, and how to facilitate the research synthesis process by maintaining related material over time. We propose a PIM tool, InfoATV (*Info*rmation *A*ll *T*errain *V*isualizer) that has the potential to assist researchers to re-find and further (re) discover their research documents. InfoATV differs from existing systems in that it allows users to make connections between papers in their collection and provides a behavior-centric model for re-finding that allows users to re-create their flow of activity. While other systems show users the documents viewed during a given period or the bread crumb trail of pages touched, we allow them to browse through views of what the system looked like at each step in their search process. This will allow them to more easily re-create the mental model they previously had that was associated with each document in their collection, which, in turn, facilitates re-finding, (re)discovery, and knowledge creation.

Managing a collection of personal documents over time is an arduous task. Navigating the information space of such a collection can be like walking through a maze—where it's difficult to perceive where one is going, how much ground one is covering, and whether one is circling in the same space. Users of research management systems can benefit from the ability to tease out a relationship between sections of the information space built over time to make an informed decision about the next step in the re-finding process. Researchers may fail to recall details of the previously viewed documents or what drove them to download those articles. Revealing the connections between papers in the information space helps them discover novel and relevant documents. Our system searches the existing document collection and aims to discover the unknown connections between documents and re-use the acquired knowledge for new purposes.

To orient researchers in the information space they build over time, we introduce the notion of re-finding to (re)discovery. We define this as the process by which a system not only displays related material as users are re-finding, but also reveals the semantics of connections between items. For example, if a researcher has re-found a document discussing cardiac rehabilitation, and they viewed a document a year ago about a specific rehabilitation program, the system would display that other document—a document they may have forgotten about, or not thought relevant, thereby allowing them to (re) discover it. The automated process by which one can go from re-finding activities to an exploration of related previously viewed documents is what we are defining as re-finding to (re) discovery. As researchers are re-finding documents, InfoATV mines their personal collection for similar documents to help make connections in the collection. So little may be remembered from these papers that it is effectively finding them for the first time. Providing users with contextual information so that they can understand how related documents *are* related is necessary to fully facilitate (re)discovery. Our system provides users with a network showing users how these documents are conceptually related.

Furthermore, users need a system that helps them manage documents at the session level. When researchers are trying to resolve an information need it becomes necessary for them to identify and construct queries for several inter-related questions. As users perform several searches in a session with long result sets, it becomes difficult to re-find what seemed initially irrelevant but becomes relevant as the information need is reformulated when new evidence is discovered [1]. Researchers often have to repeat the previous search processes, which can be a time-consuming activity. It requires a system that will provide context to assist users in keeping track of research articles

We designed InfoATV, in part, to address this issue. We introduce the idea of local re-finding, which can be defined as the process of re-finding material from a previous search within the current search session. Our system maintains breadcrumbs of recent activities to allow users not only to see searches, but also their interactions with information visualizations and the documents they clicked, making it easier for them to re-locate results from previous searches.

The main contributions of this paper include:

- We introduce an implemented behavior-centric system that caches and visualizes all user search activities (e.g., queries and clicks) from the interface to support re-finding activities. The system allows users to both search their personal collection for items and view the flow of activity associated with their search results.
- We identify two characteristics for a system to improve search that actually facilitate the re-finding process.
- We propose the notion of re-finding to (re)discovery. As users are re-finding documents our system mines their personal collection for related material to reveal unknown connections and (re)discover papers about which they may have forgotten.

The rest of the paper is organized as follows: Section 2 discusses related research from the fields of personal information management, information seeking behavior and information visualization. Section 3 relates our design rationale and describes our system. Section 4 discusses system features. We conclude the paper and describe future work in Section 5. This paper focuses on the needs of researchers, but the proposed framework is generalizable to other personal information management systems.

2. BACKGROUND

2.1 Personal Information Management

PIM systems provide a means to support users in organizing and finding information that has already been selected [30]. In a study by Elsweiler et al. the purpose of approximately 40% of web search activities were to re-find something, suggesting that re-finding is a very important issue [16]. As document collection size increases users increasingly experience difficulty in managing their acquired information [2, 5, 15]. The issues are even more complex for researchers, as they not only want to locate information, but to know how a document relates to similar research, who has worked on the same problem, and what are the most relevant papers [25].

Designing PIM systems to support re-finding has been the focus of numerous publications, as users frequently do not remember enough details about any given document to retrieve it successfully—particularly when they can only search based on location or document title [16, 19]. To resolve this problem, many information management researchers have designed systems that provide additional contextual meta-data associated with documents to search on. For instance, the attribute of time has frequently been used in personal information management tools to aid in the re-finding process. The LifeStreams project is amongst the earliest such systems [18]. This system allows users to navigate through a temporally ordered stream of documents and summarizes groups of related documents to generate overviews. The time attribute serves as a valuable pointer because it is a distinguishing trait used in daily life to provide context for experience; hence recollection of the ‘what’ can be triggered by knowing the ‘when’. The LifeStreams models as well as other systems including MyLifeBits [20] and Stuff I’ve Seen [11] allow users to view series of documents or search for documents based on temporal information. Users of the Stuff I’ve Seen system reported that they were able to find items relevant to current needs that they forgot about and for which they would not have perform an explicit search. Therefore, a system that explicitly supports re-finding to (re)discovery would be very useful to further assist them in stumbling upon relevant documents. However, these systems do not provide users with a means of recreating the context in which the documents were discovered or allow them to perceive the flow of activity from the time in question. In short they are archival-centric models that allow users to find collections of documents not representations of behavior. If they had they would provide the users with a better means of triggering their memory not just based contextual meta-data. Instead they would allow them to see what they were doing at the time they originally found the item and more aptly assist users in having that ‘ah ha’ moment where they can re-create their mental model associated said item and re-finding with greater ease.

Psychology literature suggests that users people remember things through a chain of associations [12, 29]. Chau et al. [9] indicated that a limitation of the LifeStreams model is that it does not provide users with a means making use of building such chains. Chau et al. built a system that allows users to create complex queries by providing them with various attributes on which to search such as associated events. For example, a user could do a search where the chain of association is the file type, date retrieved, person associated, etc. A limitation of this approach is that users have to specify the associations beforehand and users may not know which will trigger memory at an arbitrary point in the future. Other systems (e.g., Presto [13]) also force users to create their own meta-data, requiring more work from the user

The Implicit Queries system provides functionality that is somewhat similar to our re-finding to (re)discovery support [14]. When users of this system are composing or writing emails, the system automatically creates a query from keywords in the text of the email and searches their personal index of information for related material. The personal index includes mail, calendar items, web pages, files, IM Messages, etc. Relevant results are displayed in a list, which does not provide users with any contextual information that can be used to determine if it is, in fact, relevant.

Elsweiler et al.[16] explored the behavior of email users who are in the re-finding process. Their research showed that users in the process of re-finding often have difficulty, and end up performing several searches and looking through many documents until the needed item is found. They also found that users become disoriented and in one re-finding session often look through the same folder or at the same document multiple times. They suggest that users could benefit from a search-trail metaphor embedded within PIM systems so that users do not repeat the same search over and over within a session. Secondly systems must help orient users such that they would be able to determine the relevancy of specific documents. They advise building systems that orient users by automatically producing connections between items just as landmarks orient people in the physical environment. Additionally, they maintain that systems must highlight re-found items, as people tend to re-find many items multiple times. We have designed our system to, in part, respond to these suggestions, which we discuss further in section 4.

Vannevar Bush, famously, presented the idea of the breadcrumb trail metaphor as a tool to facilitate re-finding over 60 years ago [7]. He believed human knowledge would best be exploited when systems allowed us to find information similarly to how we remember—that is associatively. Breadcrumb trails provide users with an association between items based on the discovery process. Supporting this functionality in a way that is useful and meaning has been a goal for decades. Many systems such as Feldspar [8] and MyLifeBits have been—in part—designed to do just that. Our system differs from the aforementioned in that it does not merely provide one with the sequence of pages traveled to, as if the links are static unchanging items, but what that page dynamically looked like at the time—thereby giving the user a more complete picture of their activity.

2.2 Information Seeking Behavior

Information seeking behavior research attempts to identify the characteristics of users as they are finding information. We use research from this field to support the need to minimize the difficulty of local re-finding and aid in re-finding to (re)discovery activities.

In an effort to ensure that they can easily find articles they have deemed relevant, many researchers keep a personal index. Scientists not only search for information but need to synthesis and keep track of information for knowledge formation, maintaining a personal index helps them do this. Research by Jahoda et al. suggests that around 61% of scientists keep a personal index [22]. Organizing information into categories lets people make use of information in manageable pieces. However, people have a difficult time deciding upon which categories to choose and the categories considered most relevant for a given document change not only with time but as users consider different research problems [23]. This suggests that users need systems to dynamically categorize documents. Additionally, scientists experience difficulty remembering which file folder was used to archive a particular document and have to engage in extensive searching [28]. Users forget the names of folders they have created. Complex problem solving requires planning and an understanding of the relationship between many concepts, which is very difficult without external aides—without such tools users run the risk of not being able to even remember that they have a paper in their personal collection that might be relevant to a current research project let alone find it. For this reason we have extended our system to support re-finding to (re)discovery. This functionality allows users to locate related documents without having to create a complex index that changes over time.

Studies of the information seeking behavior of scientists suggest that researchers often conduct *weak problem solving,* associated with a vague understanding of the problem space, an inability to come up with a systematic plan to resolve information needs and a lack of prior domain knowledge [25]. As researchers perform several inter-related queries during the search process much of the information may not appear immediately relevant. Resolving information needs quickly becomes complex as the volume of information to manage and synthesize grows. For this reason we introduce the notion of local re-finding. Our system assists users in maintaining their flow of activity at the session level.

2.3 Information Visualization

Visualizations can enhance cognition, representing multiple data points in a single image. In designing InfoATV we made use of information visualization to accomplish the goal of facilitating re-finding. The advantage of using images to convey information is that they are instant, memorable, automatic, and energizing [6]. There are numerous features one could use for visualizations—tempting the conveyance of too much information and leading to difficulty in interpreting the data. If visualization tools are not matched to appropriate users, tasks and real-work problems, systems fail to be adopted when deployed in real-world settings [26]. Our system makes use of information visualization design principles that are as close to gold standards as is possible, so that our system will be as user friendly as is feasible.

3. InfoATV: SYSTEM DESCRIPTION

3.1 System Overview

InfoATV is a fully automated novel extension of the LifeStreams model embedded in a document search system. Our system caches the user interface after any action that fully repaints the screen--including re-finding activities—so that users can view the context in which documents were found. The system groups cached views by the session they were found in, alleviating the need for users to organize. To deal with information overload issues, we purposed InfoATV to allow users to search documents and previous queries and then browse the flow of activity associated with the results. Users can visually browse through system states grouped by session much like they would browse through an online photo gallery.

3.2 Materials and Implementation Details

We downloaded a copy of the 2011 version of MEDLINE, which includes citation information for over 20 million records, from the PubMed website[1] and stored the data in a PostgreSQL database[2]. We indexed the titles and abstracts of all records using Lucene for subsequent search. The desktop interface was implemented in Java 1.6 using the Swing API. The Lingo [24] clustering algorithm was used to cluster documents from searches for user navigation. Lingo was chosen because it emphasizes cluster label quality, making it easier for users to select the appropriate clusters to view. During each session our system maintains an array of containers associated with each search and document touched or visualization viewed. The array is stored in the database when users exit the system. During each search or re-finding session, additional information is stored in the database. Each time a user runs a query, we store the search string along with the associated clusters. Each time a paper is touched, or cluster content is viewed--whether this takes place during searching or re-finding activities--the database is updated for later retrieval.

[1] http://www.nlm.nih.gov/databases/journal.html

[2] http://www.postgresql.org/

4. InfoATV FEATURES

4.1. Local Re-finding and Information Retrieval

We designed InfoATV with reference to improving information retrieval that actually facilitates local re-finding. Local re-finding is the process of finding items within a given search session. As discussed earlier, when researchers have an unclear information need, they perform many inter-related searches, and some items may not be immediately relevant. Therefore, during a given session, users may need to backtrack to re-find local (recent) documents. Our search interface has two characteristics whose attributes the system makes use of to facilitate the re-finding process. The first characteristic to improve the search process is the automatic generation of overviews of search results using the Lingo algorithm so that users can get a sense of which topics are in the result set. An overview allows users to engage in "strategic reading" [27] the process by which researchers navigate through resources to reduce the number of clicks to get to relevant information so that they can scan more and read less. InfoATV users can not only view clusters but they can then view sub-clusters of clusters and drill down as many times as they want until they reach a one-cluster set.

The second characteristic is support of citation chaining, which accommodates user browsing of both citing and cited by documents for a particular publication. Citation linkage provides a very strong relationship between papers and is already in use by scientific researchers to find information [15]. Additionally, researchers interested in mapping scholarly fields and charting changes in those fields have used citation information for years to generate overviews [4].

Citation chaining, is a very useful method for finding new information that does not require the user to construct a new query. This is an important system feature, as we know users often have a difficult time coming up with the most appropriate queries for their information need. Therefore, allowing users to chain citations not only facilitates known behaviors of researchers, but also gives them the ability to navigate to new material easily. Citations are clustered with labels so that users can determine which papers are relevant.

We update our database each time a user views a set of clustered results, clicks on any cluster to view its content, or views any papers. The user can re-create their flow of activity later when they are re-finding providing a rich contextual information space to explore. Because InfoATV improves the search process by providing overviews and facilitating citation chaining, users have even more contextual data to help orient them to the documents they are viewing. All re-finding activity is updated so that users can see not only see what they have done during the search process, but their finding activities as well.

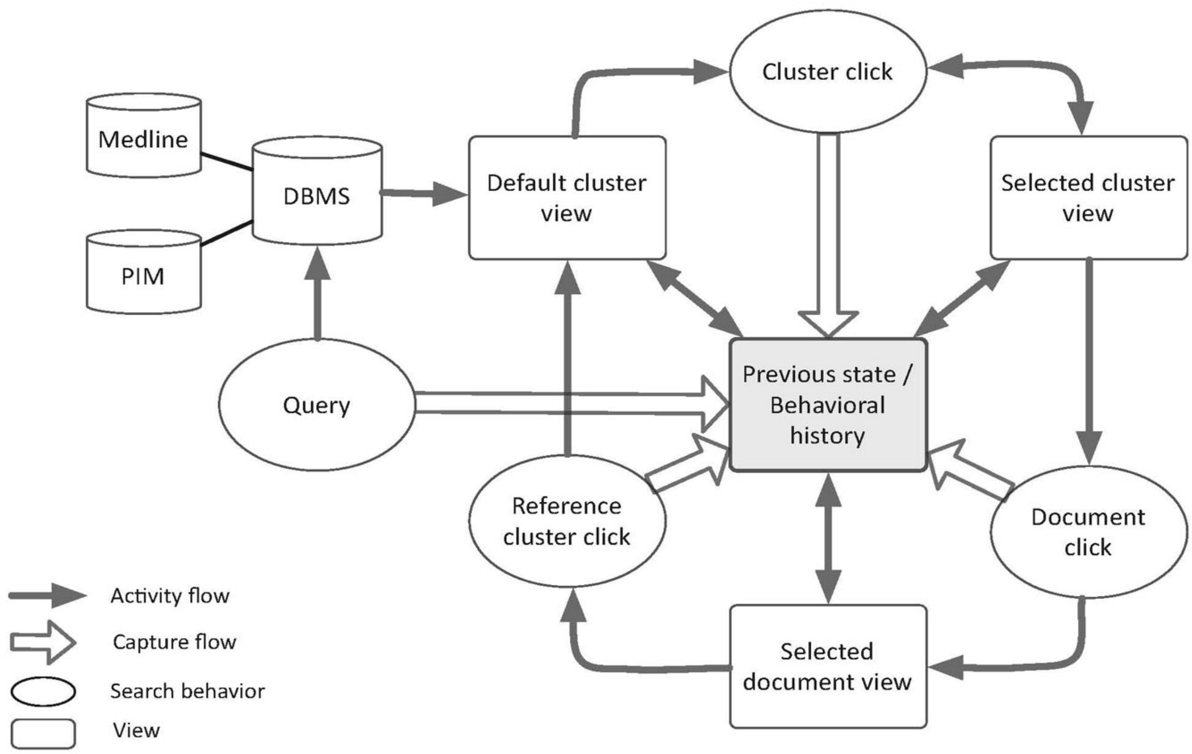

Figure 1 InfoATV flowchart. The system maintains a database of Medline and PIM system. In the flowchart, the system captures and visualize search activities (e.g. , queries and clicks)., allowing users to return to an arbitrary prior search status at any given search stage

These three characteristics allow the system to re-create a contextually rich flow of activity to support local re-finding. Context is not limited to search results, but cluster nodes, and citation linkage. As users do many inter-related searches, the need exists to find previously viewed documents aspect of managing research activities. As mentioned earlier, researchers often do not know if a particular item is relevant when they first come across it. This typically implies a need to shift perspectives between the current search activity and the results of previous searches within the session, Thus, the ability to backtrack and locate previously viewed items, even at the session level, is crucial. We refer to this as local re-finding. InfoATV users can not only back track to searches, they can view clusters they have clicked, document pages they have viewed, and see the sequence of their interactions. Figure 1. Details the flow of activity possible in rehabilitation and decides she wants to read more about this. She remembers she saw a few papers in one of the other searches about psychological aspects of rehabilitation, but at this point she has looked through so many papers she doesn't remember what it was called. So, she scrolls down to the recent activity section and hovers over several of the thumbnails to view a larger image. She finds that one of the clusters she viewed for the last search had the relevant paper she wants. She was able to easily find this item even though she did not click on its link and it would have been number 56 in the relevance ranked list. Had she looked at the relevance ranked list it would have taken her a while to scroll through the list to find what she needed. Because the system provides her with cluster labels and state views there is richer contextual data to trigger her memory.

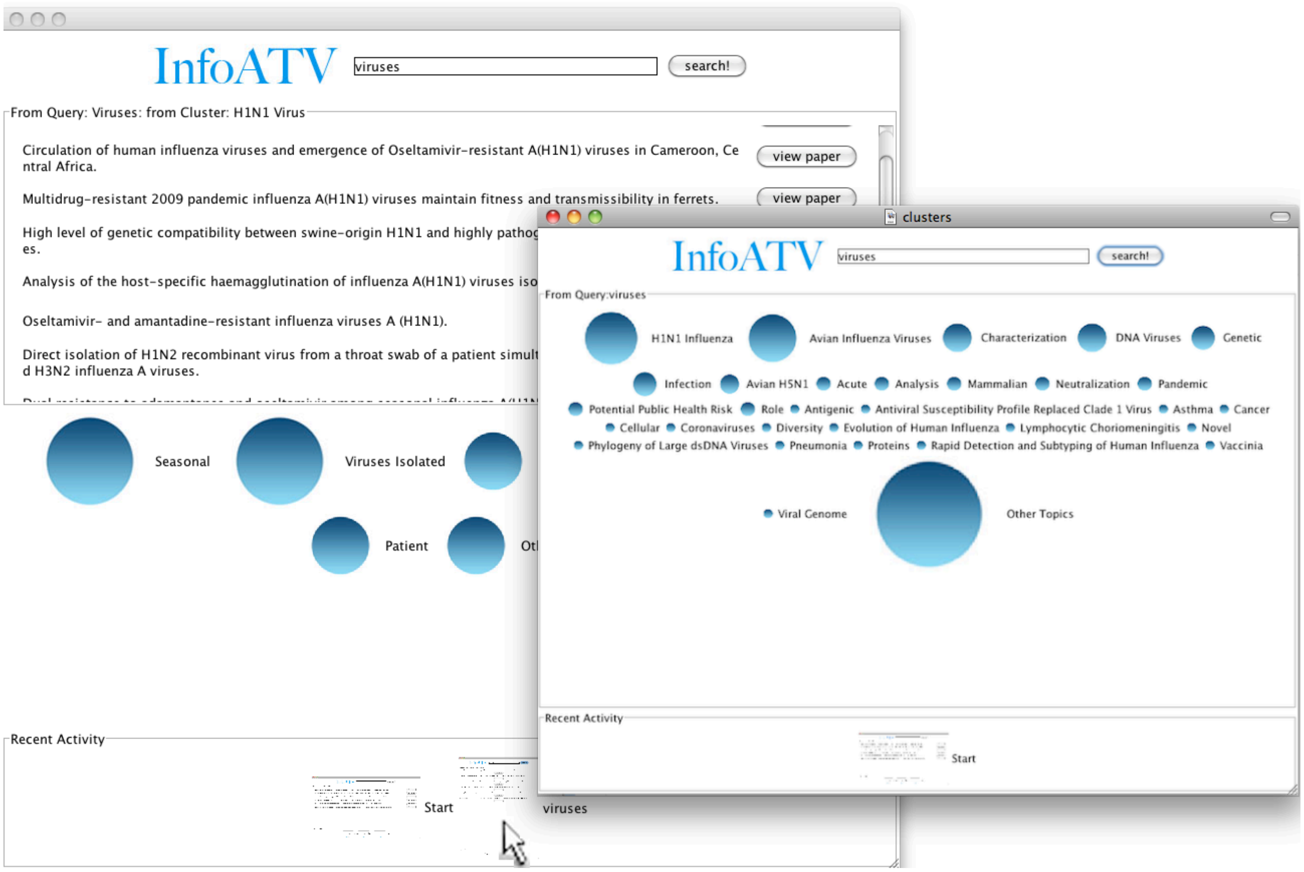

Figure 2 InfoATV interface. This is a snapshot of the interface showing a view of cluster content. At the top of the larger image is the search bar. Next down is a list of the documents with a label displaying the query and cluster the results came from. At the bottom of the page is a breadcrumb trail of clickable thumbnails of session states, each of which enlarges when the mouse hovers over it.

InfoATV. Figures 2 and 3 are snapshots of the interface and show the way local re-finding activities can be executed.

We will illustrate the usefulness of these features by describing a fictional use-case. In this narrative, Dr. Black will be performing a series of searches related to cardiac rehabilitation.

Dr. Black performs three related searches: weight-loss intervention, cardiac rehabilitation, and exercise-based rehabilitation. When she performs the last search, and the results are displayed, she clicks on a cluster mental health. She sees that there are many papers about psychological issues pertaining to

4.2 Re-finding to (re)discovery

Our system facilitates re-finding to (re)discovery. As Engelbert and others have noted, over time, the way people categorize information changes [17]. What may have classified under one heading may not seem appropriate as more items are collected and new or different conceptual groupings emerge. Many documents collected at one point may be relevant for categories that were not determined until much later. As the information landscape changes, managing connections between documents become difficult. As noted earlier, in reference to the Stuff I've Seen system, when users are given the opportunity to explore their information space they often find information that they did not realize was relevant to a current information need. Our application explicitly supports the unexpected finding of relevant information, or, as we name it, re-finding to (re)discovery. This feature is very useful for two main reasons. First, it eases the burden of remembering every relevant paper in the collection off the user. Merely by re-finding one needed document others can be found.

conceptual overlap between the paper selected by the user and the other papers using the MeSH terms they have in common as well as queries the documents are associated with. We implement the network using the Fireworks model[10], dynamically generating the network using JUNG[4]. In this model each document is represented as a floral break where the center node is the document and it is connected by a set of nodes representing either MeSH terms or queries. as shown in Figure 3. Users can not

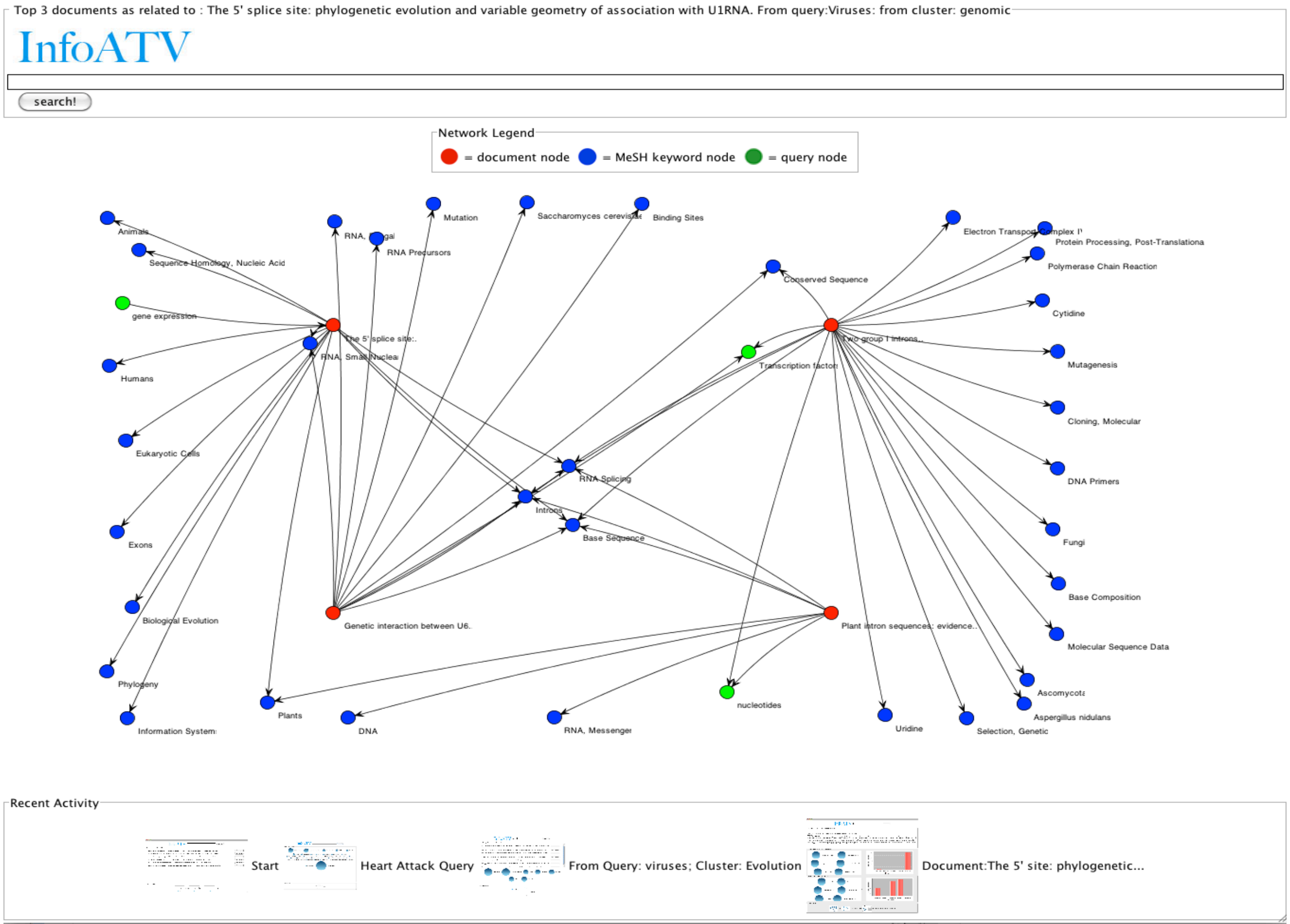

Figure 3 Re-finding to (re)discovery. Each red document node is linked to its MESH term nodes as well as query nodes. Users can see the overlap between the most similar papers to get a sense of how they related and explore the system by clicking on the nodes. Hovering over document nodes reveals the full citation.

Our implementation of this feature also facilitates the knowledge creation process for researchers. By showing users not only related papers, but how they are conceptually related, users can gain a better understanding of the research areas they are interested in.

To facilitate re-finding to re-discovery we allow users to view automatically generated connections between papers they have viewed using InfoATV. When users are viewing a document, either when they are re-finding or searching for a new document in the MEDLINE database, they can view a network of the most relevant documents and how they are connected. We use Lucene to find the documents that are most similar based on title and abstract words as well as MeSH subject headings[3]. We show the

only view the papers that are highly similar, they can see how they are similar and navigate away from the original paper to not only documents from their collection that are related and their contextual information, but any of the term or query nodes can be used for further exploration. This model allows users to make more intelligent use their document collection by showing them its complexity and connectedness.

Coming back to Dr. Black where we left her off will help demonstrate how (re)discovery fits into the re-finding process. The title of the paper she found when she was in the course of local re-finding was ‘Post myocardial infarction depression: Increased hospital admissions and reduced adoption of secondary prevention measures- A longitudinal study’. She views the paper

[3] MeSH terms are subject headings generated by the National Library of medicine for the MEDLINE database.

[4] Java Universal Network/Graph Framework (http://jung.sourceforge.net/)

and decides it is useful to her current information need. She clicks on the (re)discovery link. She sees a network much like figure 3. There are two documents she thinks might be relevant to her based on their title and the keywords she sees they are both associated with. She also sees both are linked to a query she did a week and a half ago using the phrases “cardiac rehabilitation” and “quality of life”. She forgot about that search as well as the papers she is viewing and deems them relevant. She decides that she will click on both document nodes and the associated query node for further exploration of her personal collection.

4.3 Re-finding long after the fact

InfoATV supports re-finding not just at the session level, but for any item in the collection that was added at any time. Our support of both local re-finding and re-finding to (re)discovery allows for additional contextual data to trigger a user’s memory. If users want to find documents from a previous period there are many paths by which they may stumble upon it. When a user opens the application they can view what looks like a list of search paths for each query in the last session. The queries are from the most recent search session, orienting the user to their most recent behavior. The search paths only show the first few steps the user previously took in the process. Users can click on any query to view what documents and clusters were touched, as well as its associated breadcrumb trail. Users can also browse through sessions via a series of views as described above. By clicking on any query in the list, one can view the flow of activity associated with it by browsing its breadcrumb trail. Because we have designed our system to facilitate local re-finding the system is able to naturally group activities so that users do not have to view their entire stream of activity to find an item, thereby reducing information overload. Also, as users are performing local re-finding they will become more familiar with our interaction model. When they are engaged in re-finding from previous sessions they do not have to shift gears and perform vastly different behaviors to re-find. They will be merely engaging in activities that have become a part of their search process.

Users can also search their personal collection for queries or documents based on keywords. They will be presented with the documents and queries associated with those terms and can browse the associated cached states. As they are re-finding the breadcrumb trail is updated with those activities.

When Dr. Black loads InfoATV, after having used it for a year, she sees information concerning her trail of activity for her most recent session. Seeing that the last thing she did was query the system for documents about cardiac rehabilitation’, view the cluster ‘weight-loss’ and selected the paper about the ‘LEARN program’ on weight-loss. Trying to remember why she clicked on that paper she clicks on the link and is able to see what the system looked like when she first viewed it much like figure 2. She can view the abstract and title, or click on a cluster. She can see thumbnails of the flow of activity leading up to her first viewing this item. As she hovers over each successive node and views the enlarged representation, she begins to remember what it was she was thinking when she viewed this paper over three weeks ago. She now remembers there was a particular paper she wanted to re-read, but it was not from this search session. She clicks on the first thumbnail, which corresponds to the interface at start-up. She sees the thumbnails for the last few sessions and begins to hover over those one after the other. She clicks on the second to last thumbnail, having deemed this the one she needs. She looks at the name of several of the documents she touched, locating the one she needs.

4.4 System Discussion

Several aspects of our system make it novel and help resolve issues inherent in the re-finding process. Our system is behavior-centric as opposed to many other approaches that are archive-centric. We do not use contextual information merely as meta-data to augment search, and then simply display a list or show a desktop-like view of folders or items touched during a specific period. Instead, InfoATV allows users to re-experience previous sessions. Our model reveals user *behavior*, not just additional contextual data--making ours a behavior-centric approach. Local re-finding and re-finding to (re)discovery address some of the suggestions from Elsweiler, et al.’s [17] study of re-finding behavior. The authors suggest that systems need to provide users with a search trail of their behavior so that they do not repeat the same searches, look through the same folder, or view the same documents over and over in an effort to re-find. Our system maintains a breadcrumb trail of not only search activities but re-finding activities as well. By providing thumbnails of activity users can quickly view what they have recently done. The authors also suggest that systems need to facilitate orienting users to the items they are viewing when re-finding so that they understand whether or not it is relevant. By showing users their flow of activity, we are able to facilitate orienteering so that the system provides them with plentiful contextual information. As noted earlier, people remember through a chain of associations [12, 29] and we are, therefore, providing them with a chain of associated activity. The re-finding study also recommended that systems highlight frequently re-found items, as people look for the same item multiple times. Though our system does not overtly highlight re-found items, it does update the bread-crumb trail with each re-finding activity, so that items found multiple times will be in the bread-crumb trails of numerous sessions, in the context of other related activities. Those items are hence more likely to be found. Because the items will be found in context, users will be better able to determine the relevancy of those documents based on their information needs.

Usability was a key interface design concern. When data points are numerous, readability becomes an issue with many information visualizations [21]. To alleviate this problem, many information visualization systems make use of the principles of: *overview first, zoom, filter and details-on-demand* to lessen this issue [3]. Overviews allow users to identify complex relationships between items. Zooming in allows users to find the details of the patterns without being overloaded with too much information. Our system allows users to see a clustered overview of the data. They can zoom into a cluster, select a citation and view details about it, making it easier for them to make sense of the data.

We also try to keep our network view of information when the system is in re-finding to (re)discovery mode as uncluttered as possible, by showing only a small portion of the network of documents in the collection. This helps to minimize information overload as it reduces the number of connections to interpret and nodes do not occlude other nodes.

5. CONCLUSION/FUTURE WORK

By leveraging the current research on information seeking behavior, personal information management and information visualization, InfoATV provides users with better means to search, re-find and (re)discover. We have designed a behavior-centric model to assist in the re-finding process by re-creating prior search states and allowing users to browse through their flow of activity. InfoATV supports local re-finding, the processes of trying to find items from the current search session. The re-finding to (re)discovery mode reveals connections between documents and further facilitates the knowledge creation process.

Future work will include performing user studies to determine which aspects of the system are useful to users and gathering user behavior data to improve our framework. We also plan to extend our model to one that encompasses additional activities associated with the research process, such as re-finding notes.

6. ACKNOWLEDGMENTS

This publication was made possible by Grant Number UL1RR024979 from the National Center for Research Resources (NCRR), a part of the National Institutes of Health (NIH). Its contents are solely the responsibility of the authors and do not necessarily represent the views of the CTSA or NIH

7. REFERENCES

[1] Barreau, D. 2008. The persistence of behavior and form in the organization of personal information. *Journal of the American Society for Information Science and Technology*. 59, 2 (2008), 307-317.

[2] Barreau, D.K. 1995. Context as a factor in personal information management systems. *Journal of the American Society for Information Science*. 46, 5 (1995).

[3] Bederson, B.B. and Shneiderman, B. 2003. *The Craft of Information Visualization: Readings and Reflections*. Morgan Kaufmann.

[4] Bellis, N.D. 2009. *Bibliometrics and Citation Analysis: From the Science Citation Index to Cybermetrics*. The Scarecrow Press, Inc.

[5] Boardman, R. and Sasse, M.A. 2004. "Stuff goes into the computer and doesn't come out": a cross-tool study of personal information management. (Vienna, Austria, 2004), 583-590.

[6] Burkhard, R.A. Learning from architects: the difference between knowledge visualization and information visualization. *Proceedings of the Eight International Conference on Information Visualisation*.

[7] Bush, V. 1996. As we may think. *interactions*. 3, 2 (1996), 35-46.

[8] Chau, D.H. et al. 2008. Feldspar: A system for finding information by association. *PIM 2008-Proceedings of Personal Information Management, Workshop at CHI 2008* (2008).

[9] Chau, D.H. et al. 2008. What to do when search fails: finding information by association. *Proceeding of the twenty-sixth annual SIGCHI conference on Human factors in computing systems* (2008), 999–1008.

[10] Chin, S.-C. et al. Firework visualization: a model for local citation analyisis. *Proceedings of the ACM SIGHIT International Health Informatics Symposium*.

[11] Cutrell, E. et al. 2006. Searching to eliminate personal information management. *Communications of the ACM*. (2006).

[12] Davies, G.M. and Thomson, D.M. 1988. *Memory in context: Context in memory*. John Wiley & Sons.

[13] Dourish, P. et al. 1999. Presto: an experimental architecture for fluid interactive document spaces. *ACM Transactions on Computer-Human Interaction (TOCHI)*. 6, 2 (1999), 133–161.

[14] Dumais, S. et al. 2004. Implicit queries (IQ) for contextualized search. *Proceedings of the 27th annual international ACM SIGIR conference on Research and development in information retrieval* (2004), 594–594.

[15] Ellis, D. and Haugan, M. 1997. Modelling the information seeking patterns of engineers and research scientists in an industrial environment. *Journal of documentation*. 53, 4 (1997), 384–403.

[16] Elsweiler, D. et al. 2011. Understanding re-finding behavior in naturalistic email interaction logs. *Proceedings of the 34th international ACM SIGIR conference on Research and development in Information* (2011), 35–44.

[17] Engelbart, D.C. and English, W.K. 1968. A research center for augmenting human intellect. (San Francisco, California, 1968), 395-410.

[18] Freeman, E. and Gelernter, D. 1996. Lifestreams: a storage model for personal data. *ACM SIGMOD Record*. 25, 1 (1996), 80–86.

[19] Fuller, M. et al. 2008. Applying contextual memory cues for retrieval from personal information archives. *PIM 2008-Proceedings of Personal Information Management, Workshop at CHI 2008* (2008).

[20] Gemmell, J. et al. 2002. MyLifeBits: fulfilling the Memex vision. *Proceedings of the tenth ACM international conference on Multimedia* (2002), 238.

[21] Herman, I. et al. 2000. Graph visualization and navigation in information visualization: A survey. *Visualization and Computer Graphics, IEEE Transactions on*. 6, 1 (2000), 24-43.

[22] Jahoda, G. et al. 1966. Characteristics and use of personal indexes maintained by scientists and engineers in one university. *American Documentation*. 17, 2 (1966), 71–75.

[23] Malone, T.W. 1983. How do people organize their desks?: Implications for the design of office information systems. *ACM Trans. Inf. Syst.* 1, 1 (Jan. 1983), 99–112.

[24] Osiriski, S. et al. 2004. Lingo: Search results clustering algorithm based on singular value decomposition. *Intelligent information processing and web mining: proceedings of the International IIS: IIPWM'04 Conference held in Zakopane, Poland, May 17-20, 2004* (2004), 359.

[25] Palmer, C.L. et al. 2007. Weak information work in scientific discovery. *Information Processing & Management*. 43, 3 (2007), 808–820.

[26] Plaisant, C. 2004. The challenge of information visualization evaluation. *Proceedings of the working conference on Advanced visual interfaces* (New York, NY, USA, 2004), 109–116.

[27] Renear, A.H. and Palmer, C.L. 2009. Strategic Reading, Ontologies, and the Future of Scientific Publishing. *Science*. 325, 5942 (Aug. 2009), 828-832.

[28] Stibic, V. 1980. *Professional Documentation for Professionals: Means and Methods*. North-Holland Publishing Co., Amsterdam.

[29] Tulving, E. and Thomson, D.M. 1973. Encoding specificity and retrieval processes in episodic memory. *Psychological Review*. 80, 5 (1973), 352-373.

[30] 2007. *Personal Information Management*. University of Washington Press.