ResearchExplorer: Gaining Insights through Exploration in Multimedia Scientific Data

Bo Gong
School of ECE
Georgia Institute of Technology
Atlanta, GA 30332
bgong@ece.gatech.edu

Rahul Singh
School of ECE
Georgia Institute of Technology
Atlanta, GA 30332
rsingh@ece.gatech.edu

Ramesh Jain
School of ECE
Georgia Institute of Technology
Atlanta, GA 30332
jain@ece.gatech.edu

ABSTRACT

An increasing amount of heterogeneous information about scientific research is becoming available on-line. This potentially allows users to explore the information from multiple perspectives and derive insights and not just raw data about a topic of interest. However, most current scientific information search systems lag behind this trend; being text-based, they are fundamentally incapable of dealing with multimedia data. An even more important limitation is that their information environments are information-centric and therefore are not suitable if insights are desired. Towards this goal, in this paper, we describe the design of a system, called ResearchExplorer, which facilitates exploring multimedia scientific data to gain insights. This is accomplished by providing an interaction environment for insights where users can explore multimedia scientific information sources. The multimedia information is united around the notion of research event and can be accessed in a unified way. Experiments are conducted to show how ResearchExplorer works and how it cardinally differs from other search systems.

Categories and Subject Descriptors

H.5.2 [**Information Interfaces and Presentation**]: User Interfaces – *graphical user interfaces*; H.2.8 [**Database Management**]: Database Applications – *scientific databases*; H.3.1 [**Information Storage and Retrieval**]: Content Analysis and Indexing – *indexing methods*; H.3.3 [**Information Storage and Retrieval]**: Information Search and Retrieval – *query formulation, search process*.

General Terms: Management, Design, Experimentation.

Keywords

Insight, Interaction Environment, Event, Research Event, Exploration, Multimedia Data, Spatio-Temporal Data.

1. INTRODUCTION

Current web search engines and bibliography systems are information-centric. Before searching for information, users need to construct a query typically, by using some keywords to represent the information they want. After the query is issued, the system retrieves all information relevant to the query. The results from such queries are usually presented to users by listing all relevant hits. Thus, with these information-centric systems, users can find information such as a person's homepage, a paper, a research project's web page, and so on. However, when users want to know the following types of things, they are unable to find answers easily with current search systems:

1) Evolution of a field
2) People working in the field
3) A person's contribution to the field
4) Classical papers (or readings) in the field
5) Conferences/journals in the field
6) How the research of a person or an organization (group, dept, university, etc) has evolved.

The reasons why current information-centric search systems have difficulty to help users to find answers to questions above are due to the limitations of their information environments.

First, some issues result from their data modeling. For example, to answer the question of "evolution of a field", the most important information components, which are time and location, need to be captured and appropriately presented or utilized. However, in typical bibliography systems such information is rigidly utilized (if at all available) in the time-stamping sense.

Second, many important issues arise due to the presentation methods utilized by such systems. For example, even though users can find all papers of a person with some systems, it is not easy for users to observe the trend if the results are just listed sequentially. As an alternative, presenting results in a visual form can make trend easier to identify.

Third, some of the questions listed above can not be answered directly by the system because the answers depend on individual person. For example, different users will have different judgments on a researcher's contribution to a field. To form their own thoughts, users may need to investigate and compare several factors many times. In this case, it is too tedious if each query is a new query. Thus, it is necessary that the system can maintain query and user states and allow users to refine queries dynamically. In other words, the user can not only query but also explore information.

For this study, we propose a bibliography system with novel interaction environment that aids not just in syntactic query retrieval but also aids in developing insights. The goal of this system is to provide users an interaction environment where information is modeled, accessed, and presented in such a way that users can gain insights easily through exploration. Specifically, in the interaction environment, scientific information is modeled around the notion of a *research event*, which brings together all semantically related information regardless of the media (text, image, or video), through which it is expressed. Thus, when users explore the information space, they can view research in multiple media formats. Further, the interaction environment presents information using multidimensional views, which include temporal and spatial views. At the same time, the interaction environment shows information of other attributes of research, like category and people information.

In summary, the contribution of this work is to propose a novel interaction environment for insights. Although the system is focused on scientific information, we believe the techniques developed in this work are applicable to other applications and can work as a framework guiding design of interaction environments for insights. The paper is structured as follows. We begin with an introduction of interaction environment for insights. Section 3 describes the system architecture. Section 4 explains data modeling of the interaction environment. Section 5 presents how the interaction environment is implemented. Section 6 discusses experiments and results. Section 7 gives a review of related work. Section 8 concludes.

2. INTERACTION ENVIRONMENT FOR INSIGHTS

2.1 Insight

Our goal in designing the system is to provide an interaction environment for users to explore multimedia scientific data to gain insights into research. Insight is commonly understood as follows.

Insight*: the clear (and often sudden) understanding of a complex situation* [21].

From the definition, we can see insight is different from information. If insight is gained, people should be able to understand the inner nature of things. To illustrate their difference, we refer the reader to Figure 1. In the figure, left part shows two columns of numbers. What these numbers convey to people is just information. It is very difficult for people to understand the relationship between numbers in these two columns by looking at numbers only. But if we show these numbers by a chart as in the right hand, people can easily tell and understand that the two columns have linear relationship. That is the insight. In this case, people gain insight by understanding relationship, which is visualized by a certain technique.

In the context of research, insights should include clear understanding of different situations. Examples of these situations are a research field, a person, an organization, and a specific research event which will be defined later.

1	51
5	255
3	153
9	459
8	408
4	204
2	102
6	306
7	357

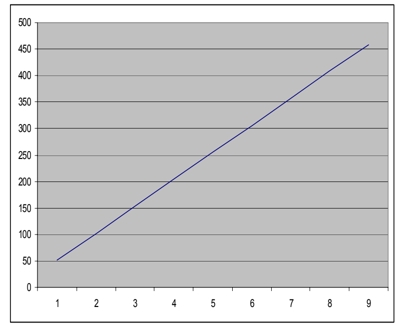


Figure 1. Information vs. insight

2.2 Key Characteristics of Interaction Environment for Insights

An interaction environment for insights is an environment that helps users to gain insights through exploration. It consists of a database to store data, and user interface to explore data. Such an environment has the following key characteristics.

1) Database to store information. As described in section 1, spatio-temporal characteristics of information are critical to present the evolution of a situation. Clear understanding a situation often requires understanding how the situation evolves. Therefore, spatio-temporal aspects of information should be captured in data modeling. In addition, the data modeling should be able to unify multimedia information. Multimedia enables users to observe things using different senses. Some media can help people to understand quickly. Figure 2 shows such an example. By looking at the text at the top, people may not be able to understand what the paper "Content Based Image Synthesis" talks about. But with the help of the images below, people can get an idea quickly what the paper is about. Further, multimedia provides users opportunity to view things from different perspectives. This is especially important when clear understanding a situation requires examine the situation from multiple angles.

2) User interface. As people gain insights by exploration, they may need to check into a situation repeatedly and from different viewpoints. Thus, interaction between a user and the environment becomes very important. The design of user interface should take this into consideration. We believe the key features of UI are as follows.

   a. The UI should support exploration of the spatial and temporal characteristics of information.

   b. The UI should support direct interactions between the users and the information. This requires the UI to have two characteristics: First, the UI should have the same query and presentation space. In other words, a window in the UI can not only be used to show information but also be used to specify queries. For example, time and location windows can show temporal and spatial information. At the same time, users can issue temporal and spatial queries in time and location windows. To specify a query, the operation should be simple and direct. The other characteristic is the reflective nature of the UI. This means once information in a window is changed, all other windows will be updated automatically. This helps users to interact with the environment directly and effectively.

c. The UI enables users to issue dynamic query. In some current interaction environments, users are constrained in forming queries. For example, users can only generate temporal query with one time interval. In an interaction environment for insights, users should be able to form a query with multiple choices. This provides users more flexibility to look into a situation of interest.

d. The UI maintains the query state. It should know which query whose results are used in the search condition of another query and which query is based on another query's results. This helps users not only to be aware to contexts but also to form complex queries.

e. The UI should have zoom-in/zoom-out functionality that allows examining the information at different resolutions. When large volume of data is retrieved, there is readability issue. To address this issue, zoom-in/zoom-out functionality is needed.

f. Different visualization techniques need to be used. As shown in figure 1, visualization techniques help users to understand relationships and gain insights. However, different relationships need different visualization techniques. For example, social relationships are of network structure. Temporal relationships are two dimensional. To visualize these two types of relationships effectively, it requires different techniques.

These characteristics will guide the interaction environment design of the system we are discussing.

3. SYSTEM ARCHITECTURE

Figure 3 shows the high level architecture of ResearchExplorer. There are three main components: Event Collector, Event Database and Interaction Environment. One of the functions of Event Collector is to gather data from different sources. Then it parses and assimilates information around the notion of research event. Finally, it sends these data to Event Database. Event Database is a database of events. It stores all information around events. ResearchExplorer uses a natural XML database for Event Database. The reasons will be explained in next section. In this database, all information about a research event is stored as an XML file. The schema will be defined in section 4.2. Interaction Environment consists of User Interface and Searcher. Through the UI, users form a query. The query is then converted into XPath format by the Searcher and sent to the Event Database. After the results are retrieved from the Event Database, the Searcher gets them back to the User Interface to present to users.

In this paper, our focus is not on how to collect data or unify multimedia information by event. Interested readers can refer to [5][6][14][15][18] for information gathering and [17] for multimedia information assimilation. What we focus is on the design of Interaction Environment based on research event.

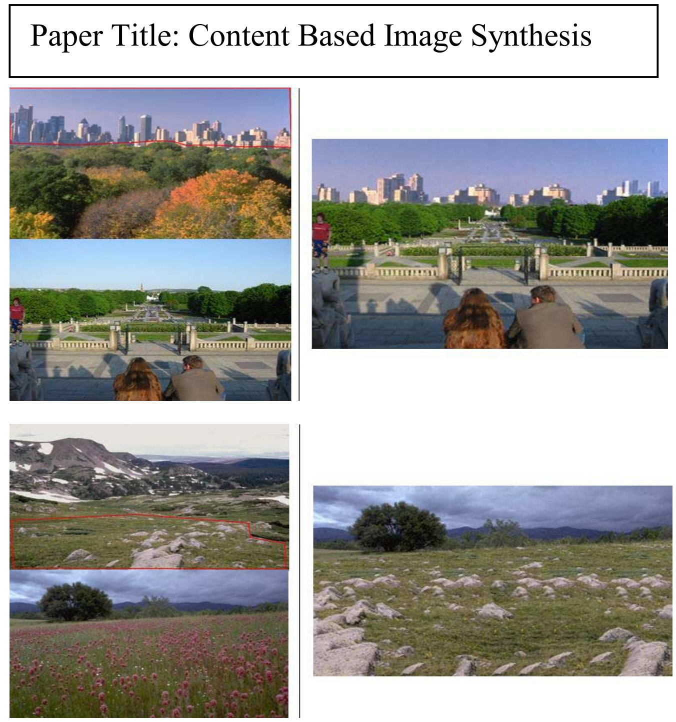


Figure 2. Multimedia helps understanding.

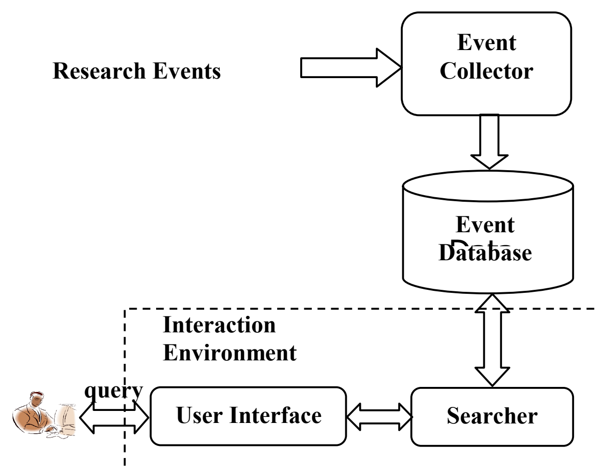


Figure 3. ResearchExplorer architecture.

4. EVENT DATABASE

4.1 Event-based Data Modeling

As described above, the interaction environment for insights needs data modeling which can capture temporal and spatial characteristics, and unify multimedia. A recent work in [17] has proposed a unified multimedia data model that is capable of describing spatial-temporal characteristics of information. This model is built on the notion of events. Its efficacy has been demonstrated in different domains including modeling of multimedia information related to meetings [17] and personal multimedia information management [16]. We base our current research on these ideas and extend them further to the specific domain of scientific information management. In [17], an event is defined as follows.

***Event**: An event is an observed physical reality parameterized by space and the time. The observations describing the event are defined by the nature or physics of the observable, the observation model, and the observer.*

This definition was given to events in general. In order to be concrete in research domain, a specific event definition to research is necessary. Therefore, based on their definition, we are defining an event in research domain, which is called research event, as follows.

***Research event:** A research event is a set of semantically correlated events within research domain, parameterized by time, location, participant, and content.*

Note that semantics is contextual. It depends on many factors like time, location, people etc. Thus, a research event is flexible. For instance, it can be a research paper, a thesis, a technical report, a book, a patent, a presentation, an image, a video, or a project combining part or all of aforementioned. Semantics also depends on domain level. It is generated differently at different domain levels even though from the same event. That is because different aspects of the event are emphasized at different domain levels. Thus, a research event could be part of another one. For example, someone is talking is an event by itself. At the same time, it is part of a seminar event as well.

The definition of a research event provides us with the central characteristics to meet the requirements of our application. By the definition, a research event is parameterized by time and location. It can capture the dynamics of itself. Thus, users can easily observe how events evolve, which is helpful to insight generation. Relationships between events can be shown in terms of attributes of an event. This will enable users to observe events in a big context and get deeper understanding. Further, all multimedia data is unified around the notion of a research event. Thus, a research event becomes an access point to multimedia data.

4.2 Semi-Structured Data

Multimedia data about scientific research does not follow a rigid structure. For example, research papers have reference while images do not have. Even for references, the number of citations varied over different papers. At the same time, these data do have some common information components such as time and location information. This semi-structured characteristic makes methods like relational database for storing structured data unsuitable. Techniques for storing semi-structured data are appropriate instead.

XML is one of the solutions to model semi-structured data. It has become very popular for introducing semantics in text. And it has rapidly replaced automatic approaches to deduce semantics from the data in text files. This approach to explicitly introduce tags to help processes compute semantics has been very successful so far [13]. Based on this, we choose XML to store research event information. Figure 4 shows the schema of XML files for research events.

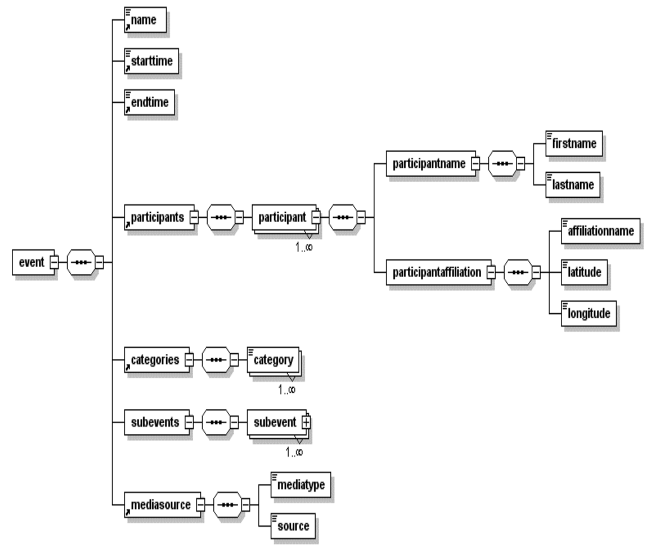


Figure 4. Schema of research event.

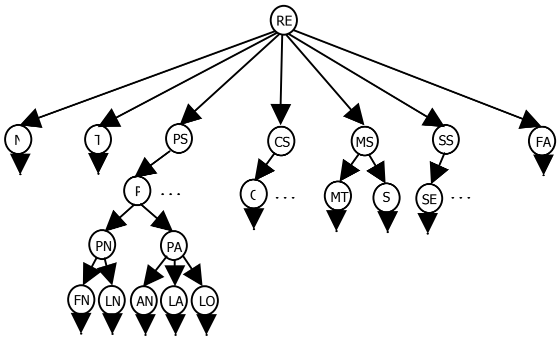
4.3 Description of the Data Model

Based on the definition of research event, four fundamental information components are needed to describe a research event. These components are: *when* the research event happens, *where* the research event occurs, *who* participates in the research event, and *what* the research event is about. Thus, a data model as follows is proposed to represent a research event. As shown in Figure 5, a research event is characterized by the following attributes: *Name, Time, Participant, Category, Mediasource, Subevents,* and *Free Attributes*. Here *Name* refers to the name of a research event, *Time* refers to the times when the research is done, *Participant* refers to people who do the research and their affiliations, *Category* refers to the ACM Classification of the research, which can belong to several categories, *Mediasource* contains media type and source (URL) of the media covering the research event, *Subevent* refers to part of the research event and has the same structure of a research event, *Free Attributes* are used to capture media specific characteristics when needed, for example, it refers to reference for a paper.

As described above, the data model encapsulates all information components of a research event by one or more attributes. *When* component is captured by *Time*, *where* is captured by *Participant Affiliation*, *who* is captured by *Participant*, and *what* is captured by *Name* and *Categories*. Multimedia supporting the research event is brought in by *Mediasource* attribute.

4.4 XML Database

In ResearchExplorer, Berkeley DB XML [3] is chosen for Event Database. Berkeley DB XML is an application-specific native XML data manager. It is supplied as a library that links directly into the application's address space. Berkeley DB XML provides storage and retrieval for native XML data and semi-structured data. So it can meet the requirements of Event Database.



RE: Research Event
N: Name
T: Time
PS: Participants
P: Participant
PN: Participant's Name
FN: First Name
LN: Last Name
PA: Participant Affiliation
AN: Affiliation Name
LA: Latitude
LO: Longitude
CS: Categories
C: Category
MS: Media Source
MT: Media Type
S: Source
SS: Subevents
SE: Subevent
FA: Free Attribute

Figure 5. Graphical representation of a research event.

5. USER INTERFACE

5.1 Windows

In ResearchExplorer, a unified presentation-browsing-querying interface is used. Research events are shown in multidimensional views. As multimedia data is organized around research event, the data is presented by fundamental components of research event, i.e., *When*, *Where*, *Who*, and *What*. Figure 7(a) shows a screenshot of the user interface we developed. There are totally five windows plus a text box. In the upper are timeline and map windows showing time and location information of research events. In the lower right, there are two windows showing people and category information. The window in the lower left is different from those windows aforementioned. It is used to show multimedia data of research events. Once a research event is selected, multimedia data like papers, images, and videos are presented in this window and they are presented according to the event-subevent structure. Clicking on a specific media instance label will lead users to the original source of the media and trigger appropriate application for that particular kind of media. So users can view original media as they want. The text box is designed for keyword-based searching. It enables users to search information in traditional way.

5.2 Research Representation

Time and location are the primary parameters based on which dynamics is captured. Therefore, they are depicted as the primary exploration dimensions. The way to represent research events in these windows is critical. In ResearchExplorer, two different representation methods are used. We borrowed idea of representing research events from [16] in the timeline window where research events are represented by rectangles. A rectangle spans the duration of a research event. Within each rectangle, there may be smaller rectangles. These smaller ones represent subevents of the research event. All rectangles for one research event are nested according to the event-subevent structure. The media, presenting the research event, are represented by icons in the rectangles. Icons are chosen intuitively for users to recognize easily. They are specific to each media. Icons belong to the same research event are grouped together in chronological order. The fidelity of such a representation is maintained during temporal zoom-in/zoom-out operations as described later. The recursive nature of the representation is used to capture aggregate relationships where a research event may comprise of other events. The primary purpose of such a representation is to provide users with a structural and temporal view of research events. In the map window, research events are represented by "dot maps" [19]. Each dot in the map shows a research event at the location of the dot. By means of dot maps, the precision of location information is high, and the variable density of dots conveys information about the amount of research events at a location.

5.3 What-You-See-Is-What-You-Get (WYSIWYG)

In the system, WYSIWYG search is employed – the query and presentation spaces are the same. As described above, windows serve as a way to display information and relationships of research events. These windows, except details window, serve another function in specifying queries as well. Contrary to many search interfaces where users specify several properties and then press a button to issue a query, users can issue a query by a simple operation in this user interface. For example, users can launch a query by specifying a time interval, a location region, a person's name, or a research category. Figure 6 shows examples of these methods. In ResearchExplorer, exploration is based on sessions. Each session consists of one or more queries. A query is either a new session query or a refine query. A new session query is the first query of each session. All other queries in a session are refine queries. For new session query, the system retrieves results from the database. If it is a refine query, the query will not be sent to the database. It will be executed based on the results set of the new session query of that session. With this method, users can choose a broad set of results first, and then observe any subset of the results of interest. This is very important because knowledge is accumulated as users manipulate the results by choosing different perspectives. Once a refine query is posed, results of the query will be highlighted in all windows.

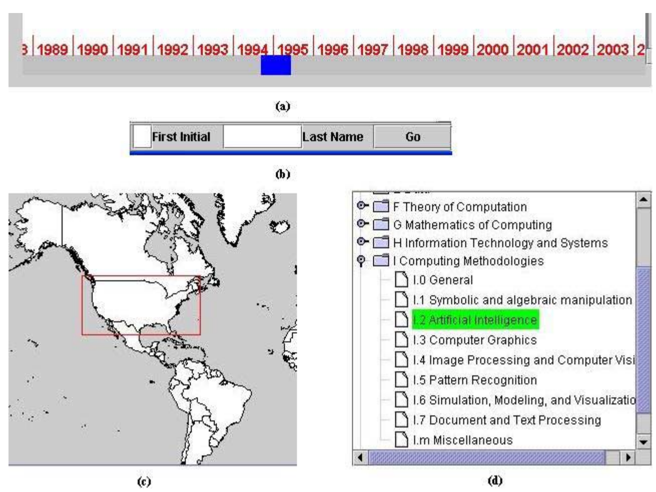


Figure 6. Different query methods. (a) shows a query by time. (b) shows a query by a person. (c) shows query by specifying a region of locations. (d) shows query by category.

5.4 Reflective UI

In designing the user interface, multiple window coordination strategy is used. By means of this strategy, components of the user interface are coupled tightly. The windows respond to user activity in a unified manner such that user interaction in one window is reflected instantly in other windows. For example, when the user selects a research event in timeline window, this research event will be highlighted in the map and other windows. This cooperative visualization is effective in information exploration as it maintains the context as the user interacts with the data. Figure 7(a) shows an example where a research event is selected and its information in other windows is highlighted.

5.5 Interactions with Time and Location Information

In ResearchExplorer, both timeline and map provide zoom-in/zoom-out function. This makes it possible for users to look at how a research event evolves in details. The timeline has year as the highest level of temporal representation. So it's likely that two subevents of a research event are overlapped when they are shown in year level. With zoom-in functionality, users can zoom into finer level to see the temporal relationship between these two subevents. The location window in ResearchExplorer has been implemented using an open source JavaBeans™ based package called OpenMap™ [2]. Research events are presented as dots on the map. Due to the size limitation, it is hard for users to differentiate events when they are close to each other. Similarly as in timeline, users can zoom into that area and see the events at finer resolution. Further, panning of the entire map is also supported.

6. EXPERIMENTS

In this section, we conduct some experiments as case studies. These case studies will show how users can look into details of research events and observe relationships between research events.

6.1 Experiment I: Exploring Information

Exploring information with context is one of important features of ResearchExplorer. With this function, users can refine retrieved results to check into different aspects of a situation. In this experiment, we are interested in how users can refine results as they explore information. Assume following information is of interest:

- Show all research events on AI during the time period from 1989 through 2004?
- Out of the results above, show the part done in CA
- For each person, show all research events he/she participated in.

To find answers to the first query, users can select the time interval from 1989 to 2004 and then choose AI category only. Figure 7(a) shows all results. Note that the results consist of all research events in AI, but they do not include all in the world. As shown in the figure, timeline window shows the temporal information and temporal relationships between research events, map window shows the distribution of location, category window shows what all categories these research events belong to, participant window shows all people being involved in these research events. The details window shows all multimedia data about research events. In this window, a research event named "UMN MegaScout" is placed at the top after the rectangle representing this event is clicked in the timeline window. If we click the image thumbnails, the original images will be shown. Figures 7(b) shows the three images and four videos. When we look at these images and video frames, we can have a better understanding about this research. In other words, direct observation of the multimedia data of a research event helps users to gain insights about the research. In order to answer the second query, we specify a region on the map which encloses all dots on CA. The research events are then highlighted in timeline windows as shown in Figure 7(c). To check research events a person participated in, we only need to move the mouse cursor to be over the person's name. Similarly, all research events he/she participated in will be highlighted.

6.2 Experiment II: Comparisons with Other Systems

In this section, we compare ResearchExplorer with other systems in terms of functionalities. Without loss of generality, we choose Google [9], CiteSeer [7], and ACM Digital Library [1] as examples. First, we compare the presentation methods. Figure 8 shows the screen shots of these systems after a query of "artificial intelligence" is issued. Compared with ResearchExplorer, these systems are unable to show how AI research evolves. Users thus can not get a whole picture of AI research, which otherwise is important to understand this area and conduct research in AI. Another comparison is done on query and explore functions. The results are shown in Table 1.

7. RELATED WORK

There are many systems which can search for scientific information. These systems can be classified into two categories. The first are bibliographical systems developed especially for scientific information searching. ACM Digital Library, IEEE Xplore [10], and INSPEC [11] are good examples of this class. These systems store information about publications, which are from some pre-selected sources, in their repositories. They organize data by using structural information of publications like title, author, etc. CiteSeer is another well-known system of this kind. Compared with the aforementioned, it collects publications from the web and performs citation indexing in addition to full-text indexing [8]. Another class is web search engine for general information. The most well-known of this type is Google. Systems of this class index the text contained in documents, allowing users to find information using keyword search. Our work differs significantly from these systems. These systems are designed to concentrating in information providing. Our work is focused on providing an interaction environment for insights into research.

Other related work comes from research on multimedia experience. Boll and Westermann [4] presented the MediÆther multimedia event space, a decentralized peer-to-peer infrastructure that allows to publish, to find and to be notified about multimedia events of interest. Our focus was not to create a multimedia event space but rather to develop an interaction environment for users to experience multimedia events. The Informedia group at CMU has also worked on multimedia

Figure 7. ResearchExplorer UI. (a) shows screen shot of the UI. (b) shows the images and videos of a research event. (c) shows the highlighted results when a spatial refinement is made.

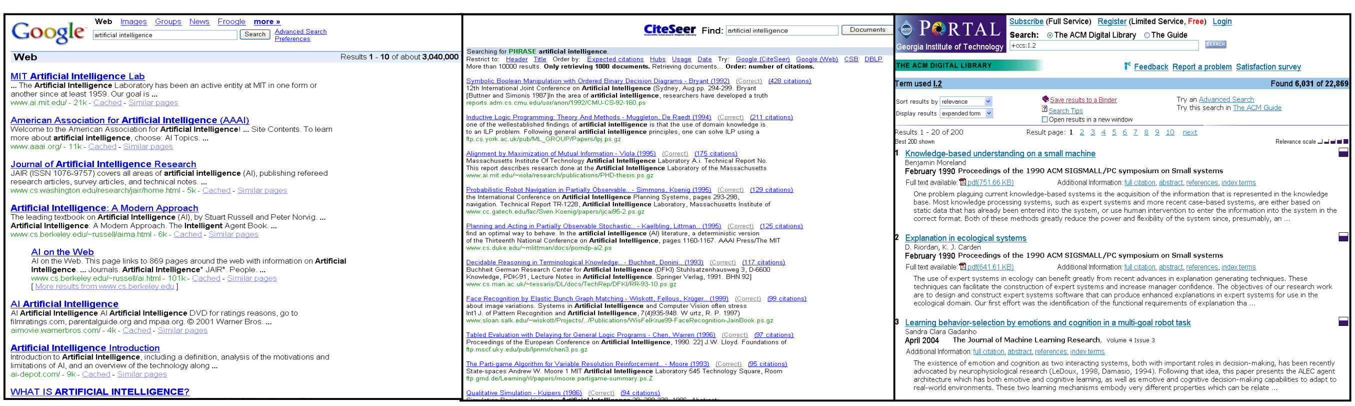


Figure 8. Screen shots of other search systems.

Table 1. Comparisons of ResearchExplorer and other systems

Functions	ResearchExplorer	Google	CiteSeer	ACM Digital Library
Show spatio-temporal relationships	Yes	No	No	Can only list results by date order.
Same query and presentation space?	Yes	No	No	No
Dynamic query	Yes	No	No	No
Maintain query state	Yes	No	No	No
Zoom-in/zoom-out	Yes	No	No	No
Visualization techniques	Multiple	Listing only	Listing only	Listing only

experiences not create an environment for experiencing multimedia personally.

In [12], Jain envisioned the essence of an experiential environment. The main goal of experiential environment is for insights. Following this work, there are some other work on experiential environment [13][16]. However, in these work, there is little discussion on experiential environments. Our work developed further some ideas in [12] and concretized the design framework of an interaction environment for insights.

8. CONCLUSION

We have described a novel system which helps users to gain insights through exploring multimedia scientific data. Although framework for designing an interaction environment for insights is identified, the implementation is a first step towards a mature system for insights. In future work, we will build on the methods described here. Also, we will investigate more on relationships between research events and methodologies to present these relationships. We believe some of the more interesting research problems will be identified when new relationships between research events are used to help users to gain insights.

9. ACKNOWLEDGMENTS

We would thank Punit Gupta and Rachel L. Knickmeyer for their help on timeline and map components of ResearchExplorer.

10. REFERENCES

[1] ACM Digital Library, http://portal.acm.org/portal.cfm.

[2] BBN Technologies (1999), OpenMap™ – Open Systems Mapping Technoloy, http://openmap.bbn.com/.

[3] Berkeley DB XML, http://www.sleepycat.com/products/xml.shtml.

[4] Boll, S., and Westermann, U. MediÆther — an Event Space for Context-Aware Multimedia Experiences. *Proc. of the 2003 ACM SIGMM Workshop on Experiential Telepresence* (ETP '03), 21-30.

[5] Brin, S., and Page, L. The Anatomy of a Large-Scale Hypertextual Web Search Engine. *Proc. of 7th International World Wide Web Conference* (WWW '98), 107-117.

[6] Cho, J., Garcia-Monila, H., and Page, L. Efficient Crawling through URL Ordering. *Proc. of 7th WWW Conference* (1998), 161-172.

[7] CiteSeer, http://citeseer.ist.psu.edu/cis.

[8] Giles, C. L., Bollacker, K. D., and Lawrence, S. CiteSeer: An Automatic Citation Indexing System. *The Third ACM Conference on Digital Libraries* (1998), 89-98.

[9] Google, http://www.google.com.

[10] IEEE Xplore, http://ieeexplore.ieee.org/Xplore/DynWel.jsp.

[11] INSPEC, http://www.iee.org/Publish/INSPEC/.

[12] Jain, R. Experiential Computing. *Communications of the ACM, 46, 7* (July 2003), 48-54.

[13] Jain, R., Kim, P., and Li, Z. Experiential Meeting System. *Proc. of the 2003 ACM SIGMM Workshop on Experiential Telepresence* (ETP '03), 1-12.

[14] Rowe, N. C. Marie-4: A High-Recall, Self-Improving Web Crawler that Finds Images using Captions. *IEEE Intelligent Systems, 17, 4* (2002), 8-14.

[15] Shkapenyuk, V., and Suel, T. Design and Implementation of a High-Performance Distributed Web Crawler. *Proc. of the Intl. Conf. on Data Engineering* (ICDE '02).

[16] Singh, R., Knickmeyer, R. L., Gupta, P., and Jain, R. Designing Experiential Environments for Management of Personal Multimedia. *ACM Multimedia 2004*. To Appear.

[17] Singh, R., Li, Z., Kim, P., and Jain, R. Event-Based Modeling and Processing of Digital Media. *1st ACM SIGMOD Workshop on Computer Vision Meets Databases*, Paris, France, 2004.

[18] Teng, S-H., Lu, Q., and Eichstaedt, M. Collaborative Web Crawling: Information Gathering/Processing over Internet. *Proc. Of the 32nd Hawaii Intl. Conf. on System Sciences* (1999).

[19] Toyama, K., Logan, R., Roseway, A., and Anandan, P. Geographic Location Tags on Digital Images. *ACM Multimedia* (2003), 156-166.

[20] Wactlar, H. D., Christel, M. G., Hauptmann A. G., and Gong, Y. Informedia Experience-on-Demand: Capturing, Integrating and Communicating Experiences across People, Time and Space. *ACM Computing Surveys 31*(1999).

[21] WordNet 2.0 http://www.cogsci.princeton.edu/cgi-bin/webwn.