

1st Annual SLIS Student Conference – Conference Abstracts

Table of contents

1. Presentations

Maggie Swan : Creative Women on the WWW: Visual Analysis of Representation, Symbolism and Ideology

Elijah Wright -- Credibility in Weblog War Coverage

Anis Pervez -- Film, Cognition and Information

Yueyu Fu -- Protein Association Discovery in Biomedical Literature

Ketan Mane -- Content Coverage of PNAS in 1982-2001

Aaron Loehrlein -- Systems for the Representation of Bio-Medical Knowledge (MedInfer)

Andrew Kurtz -- Using a Multilevel Interest Profile for Personalizing Dynamic Online News

John Fieber -- Feedback Interfaces for Information Filtering

Eric Meyer -- Communication Regimes and Digital Images: Understanding Communication Regimes, Digital Photography and the Internet

Yung-Rang (Laura) Cheng -- Qualitative Preparation for Quantitative Analysis of College Students' Information Seeking

2. Posters

Dan Albertson -- Video Searching and Browsing Using ViewFinder

Bi-Yun (Ruby) Huang -- Portraying Falun Gong: A Content Analysis of World News Media Coverage (March 1999-March 2003)

Dan Kutz -- 24 Hours of News: Micro -Longitudinal Content Analysis of CNN and BBC

Kazuhiro Seki -- A Hybrid Approach to Named-Entity Recognition in Biomedical Texts

Ning Yu -- Email Visualization Abstract

Maggie B. Swan
mbswan@indiana.edu

Creative Women on the WWW: Visual Analysis of Representation, Symbolism and Ideology

This paper reports the results of quantitative and qualitative content analysis of images from five Websites which aim to support women's creativity. Visual analysis methods were applied to images of women sampled from the sites in order to analyze denotative aspects and connotative levels of meaning. Denotative variables included representation, collectiveness, categorization, social distance and modality; the examined connotative layer of meaning encompassed both symbolism and ideology.

In terms of denotation, it was expected that social distance would be more intimate or close since the Websites presumably are attempting to appeal to a common community. It was also hypothesized that there would be a high occurrence of group images for this same reason. Regarding connotation, it was expected that the cultural symbolism and ideologies exuded by the Websites would vary significantly and be dissimilar across sites. That is, that there would be distinct creative woman themes attached to one site or another and that there would be little overlap between themes across sites.

Two denotation variables - representation and modality - were found to have significant differences in means across groups. Figurative representations, rather than abstract, were typically the norm, and across sites, modality tended to be less high than typical Website imagery. Also, contrary to hypotheses, results indicated that women were overwhelmingly depicted individually rather than in groups and further, that social distance tended to be social more often than personal or intimate.

In terms of connotation, results suggested that the WWW encompasses a wide spectrum of cultural, symbolic and ideological beliefs about creative women, and that there exists homogeneity of themes across sites which appear at first glance to have strikingly different ideologies. Seventeen symbolic themes were discerned from examination of the images and found to overlap somewhat across sites. Three major symbolic categories were extracted from these themes: Creative Self, Traditional Values and Higher Order. The ideology of each site was then determined by categorizing prevalence of the themes within each site. In general, the ideologies seem to promote traditional notions of femininity, depicting women in traditional roles such as mother and connecting with nature, although less traditional representations do appear, such as sexuality and anonymity.

These results have wide-ranging implications for future Web medium research, including areas such as personal identification on the WWW vs. offline, voicing strategies for underrepresented groups, and community-building as a WWW construct. In addition, the methodology used can be modified to better fit content analysis on the Web medium. Future analysis with regard to these particular sites and women's creativity in general should encompass analysis of non-women images to see if the same themes occur, and also additional women's creativity sites.

Elijah Wright
elw@stderr.org

Credibility in Weblog War Coverage

In the period preceding September 11 of 2001, weblogs – rapidly updated, chronological journals which often comment on current events or ideas encountered by the author - had enjoyed a rapid rise in popularity. With the events of 9/11, and the huge increase in use of the internet to seek out current news, a new subgenre of weblogs appeared. Known informally as "warblogs," these sites focused first on the events at the World Trade Center, later on al Qaeda and the United States invasion of Afghanistan, the "war on terror", and most recently on events surrounding the United States' invasion of Iraq. Characterized by close attention to the relationship between the United States and the rest of the international community, warblogs are most clearly a subgenre because of the near-complete lack of information that does not pertain to international events. Though calling them "war blogs" may be a misnomer - it might be more accurate to call them "socially aware international issues blogs" - the term "warblog" has captured the attention of the Internet and weblog communities.

Warblogs are of interest primarily because of the attention currently being paid to Internet media, and especially the growth of Web-based information and resources. During the events of 9/11, many Web news resources were overloaded and inaccessible because of a sudden and unprecedented surge in demand for information. It seems likely that the perceived immediacy of the Internet had quite a bit to do with this - readers, accustomed to finding new information regularly posted to news sites, flocked to those same sites with the desire for instant gratification. Weblogs, which had become known as a source of endless trivia and much entertaining information, could not help but spawn a new genre to fill the temporal gap.

With increased timeliness, however, one must also begin to pay attention to issues of credibility. This is particularly crucial with amateur weblogs, which usually have no oversight and no (or few) established editorial standards. Many media consumers have come to believe that material produced by professional journalists is necessarily credible; likewise, it is often believed that material produced and posted via the internet must, de facto, be non-credible or unbelievable information. In view of the level of use of the Internet for scams, spam, and the perpetration of hoaxes, it is not difficult to understand how this viewpoint could come to exist. An emergent trend has been for a growing subset of the general population to depend on amateur-produced warblogs for the interpretation and aggregation of new information on international issues.

If warblogs are in fact less reliable sources of information and interpretations than the accounts provided by traditional journalists, it will be vital for us to know whether or not there are differences in the perceived credibility of the two sources. Knowing this will allow us to make more suitable recommendations to casual users of the internet that may not be able to evaluate information appropriately by themselves. Furthermore, this knowledge should be taken into consideration by users who are attempting to discern the difference between the quality of information and other factors that encourage readers to believe what they are being told.

In this study, criteria developed in B.J. Fogg's Stanford-Makovsky study of traditional web pages are applied to both amateur and professional weblogs, and an assessment is made of the effects that previous media forms (especially traditional web pages and newspaper traditions) have had on the reception of weblogs as a new, emergent format.

Anis Pervez
mpervez@indiana.edu

Film, Cognition and Information

Film is a mediated representation of the world as cognized by filmmakers and this form of representation is internalized by receptors in their ways of understanding film. It is a process of encoding and decoding audio-visual information. Cognitive theories of film offer the space to understand film as information. Bordwell's cognitive perspective of film together with a semantic theory of information, as developed by Dretske , are used in this paper to develop a theoretical ground to understanding film as information.

Yueyu Fu
yufu@indiana.edu

Protein Association Discovery in Biomedical Literature

Protein association discovery can directly contribute toward developing protein pathways; hence it is a significant problem in bioinformatics. LUCAS (Library of User-Oriented Concepts for Access Services) was designed to automatically extract and determine associations among proteins from biomedical literature. Such a tool has notable potential to automate database construction in biomedicine, instead of relying on experts' analysis. This paper reports on the mechanisms for automatically generating clusters of proteins. A formal evaluation of the system, based on a subset of 2000 MEDLINE titles and abstracts, has been conducted against Swiss-Prot database in which the associations among concepts are entered by experts manually.

Ketan Mane
kmane@indiana.edu

Content Coverage of PNAS in 1982-2001

This paper reports results of determining changes in content coverage of the Proceedings of the National Academy of Science (PNAS) for the years 1982-2001. Dominant keywords were extracted to visualize major research topics and their evolution in the 20-year time frame. Two time slices namely 1995 and 2001 are examined in detail. Latent Semantic Analysis was applied to determine the space of semantically similar documents for both years. The results of a co-word analysis of terms used in titles and keywords of articles were visualized using graph layout algorithms available in the graphical visualization software Pajek.

Aaron Loehrlein
aloehrle@indiana.edu

Systems for the Representation of Bio-Medical Knowledge (MedInfer)

The MedInfer project examines existing representational systems in the field of bio-medicine, particularly the sources of the Unified Medical Language System (UMLS). It analyzes the current impact of these systems as well as their potential to support machine-processable inferences in applications such as the Inference Web (i.e., the Semantic Web). It also takes into account tools designed to augment bio-medical representational systems. Because the field of bio-medicine already boasts many representational systems, MedInfer examines the effect that a large quantity of representational systems might have on a field of knowledge. It also examines the extent to which existing systems, or networks of systems, comprehensively represent the field. In development is an online database of representational systems that is designed to assist future academic research into systems for the representation of bio-medical knowledge.

Andrew Kurtz
ajkurtz@indiana.edu

Using a Multilevel Interest Profile for Personalizing Dynamic Online News

Online news data sources may contain frequently updated information. Providing personalized access to such dynamic resources is an important goal. This presentation investigates the use of a multilevel interest profile to track users' interests and personalize the content displayed. This research continues work presented at JCDL 2003 in a paper by Kurtz and Mostafa.

Many online personalized news interfaces provide the option to select broad topic categories of interest, such as Business or Technology, but these broad categories may not accurately reflect the specific interests of a user. Having a multilevel interest profile can allow for both course- and fine-grain tracking of interests and filtering based on those interests. The course grain categories, which we call channels, reflect general interests while tracking interest based on the specific topics within each channel provides a more targeted and fine-grain filter.

Tracking users' interests may be performed using explicit or implicit interest indicators. Explicit indicators may provide a more accurate initial view of the interest, but over time may disrupt the news reading process. Implicit indicators are accurate in tracking users' interest over time, but it may take some time for an accurate profile to develop. A better solution would be to track users' interest using both explicit and implicit interest indicators.

We have created an intelligent news-sifting system that initially creates an interest profile based on explicit indicators and, over time, changes in the interest profile are tracked through implicit indicators. In addition, a multilevel interest profile is maintained to more accurately represent the users' interests.

References

Kurtz, A. J., Mostafa, J. Topic Detection and Interest Tracking in a Dynamic Online News Source. 2003 Joint Conference on Digital Libraries, May 2003, Houston, TX.

John Fieber
jfeiber@slis.indiana.edu

Feedback Interfaces for Information Filtering

The notion of relevance feedback for information retrieval and filtering has a long history, but the realization of the technique in research and production filtering systems trivialize the subtle and complicated means by which people arrive at a judgment of document relevance. The end result a miscommunication between the user and system about what the feedback signifies. In the context machine learning algorithms designed to model a user's interests, this miscommunication becomes noise which diminishes if not destroys the effectiveness of the algorithm.

My dissertation research focuses on clarifying the human-computer dialog for relevance feedback in information filtering to improve the signal to noise ratio of the communication. I look at two approaches to improving the signal to noise ration of the human-computer dialog. On one side, the system needs to be more sensitive to a range factors that play into a user's judgment of the value or utility of a document. For example, a feedback may be based far more on author reputation than an interest in a statistical term distribution of the document content but a system based solely on term frequency will falsely attribute the feedback to term frequency. On the other side, the user interface to the system must make clear the limited types of feedback the system can make effective use of. Not making this clear wastes user effort in providing feedback that either cannot be used, or that is mistakenly used to the detriment of system performance.

To shed an empirical light on this, I will implement a simple news filtering system with three different feedback interfaces—a baseline case implementing the status-quo “is this interesting” feedback, and two cases with more precisely defined channels for feedback. Users will be randomly assigned to one of the three cases and will use the system over the period of approximately one month. Analysis of the data will reveal how more precisely defined channels of feedback affect: the difficulty for the user of providing feedback; the quality or signal to noise ration of the feedback, and the overall performance as judged by the user.

Eric Meyer and Rob Kling
etmeyer@indiana.edu

Communication Regimes and Digital Images: Understanding Communication Regimes, Digital Photography and the Internet

One of the promises of broadband Internet access is the increasing ease of capturing, using, analyzing, manipulating, storing, and sharing digital images. An information processing perspective emphasizes the extent to which new technologies can alter these processes -- such as allowing arbitrary image manipulation or the rapid access to vast digital corpuses. In practice, important social relationships constrain the extent to which new information processing capabilities are used, or are even legitimate! For example, artists can manipulate their digital images arbitrarily, while the pictures that can be used for evidence in courts or for "hard news" journalism are not subject to arbitrary manipulation.

This paper will examine changes in communication regimes with the introduction and adoption of digital photography and internet transmission of images into communication regimes with existing systems based on "wet" photographic technologies. The regimes we will discuss are medical researchers, police forensic photographers, political campaign photojournalism, and family photographers/hobbyists. Specifically, we are interested in the intersection between social networks in these regimes and information technology. Thompson (1995) argues that "mediated communication is always a contextualized social phenomenon: it is always embedded in social contexts which are structured in various ways and which, in turn, have a structuring impact on the communication that occurs." To what extent has the adoption of equipment for making digital images and the ability to quickly transmit the resulting images to others altered the structures of communication in specific communication regimes, and to what extent have the existing communication regimes shaped the methods of using these new technologies? How do theories of communication and collaboration such as Kling's STIN (Socio-Technical Interaction Networks) and Latour's Actor-Network Theory help understand these communication regimes?

Communication regimes share some common characteristics but also show marked differences in how the regimes adopt and adapt digital photography and internet image transmission. For instance, all of the regimes mentioned use photography as a communication tool, but vary in level of training for professional technicians to make photographs, the need for the resulting image to be scientifically credible, and the extent to which audience for the images is a critical one.

One thematic issue is speed. In police forensic photography, the use of digital photography combined with the ability to securely transmit the digital images via the internet has been touted in the press as a new tool that allows police to prosecute formerly difficult cases. The New York Times (Sept. 03, 2002) reports that police are replacing low-resolution Polaroid images that took weeks to reach judges with high-resolution digital images that are sent directly to prosecutors and judges, which has reportedly begun to make a noticeable difference in domestic abuse cases.

In medical research, a different dimension of speed is the enhanced ability of researchers to engage in collaborative research between or among scientists who are not co-located. Photojournalists who are following political campaigns rely heavily on the internet to maintain connections with publication offices while on the road following political events. The pressures for speed are increased because of the highly visible web presence of news outlets. Similarly, family photographers who post "family pix" on web sites or send them as email attachments to

friends and family share the desire for quick turnaround with photojournalists, but have very different communication needs and limitations.

Even though the communication regimes being examined in this paper share some of the characteristics identified above, there are many differences beyond just the obvious field differences. Among the early differences observed include the degree of bureaucratization, the formality of training, the barriers to adoption of new methods, and a variety of other differences including the rules for digital manipulation (who can legitimately manipulate images and in what ways) and different types of collaborative and communicative networks in the fields. Additionally, the paper details how different regimes create photographs that are can be characterized by their degree of intimacy, type of production values, degree of manipulation, file size and detail quality to show ways in which the technologies themselves vary among regimes.

Note: The full version of this paper has been accepted for presentation at the Association of Internet Researchers 4.0 Annual Meeting, to be held October 16-19, 2003 in Toronto, ON, Canada.

Yungrang (Laura) Cheng
yungchen@indiana.edu

Qualitative Preparation for Quantitative Analysis of College Students' Information Seeking

Research on information seeking is an important area in library and information science. Knowing how people think, feel, and act can help us develop more usable information systems. Several components of information seeking have been identified in the literature. Among them, thoughts, feelings, and actions are commonly studied. Kuhlthau's Information Seeking Process (ISP) model is an example. Current information seeking literature revealed some problems: 1. The importance of factors has not been tested statistically. 2. Relationships and interactions among factors are unknown.

The presentation is my dissertation study which is in progress. The research questions that I hope to answer are: 1. What is the relative importance of thoughts, feelings, and actions in an information seeking process? 2. What is the strength of interactions or relationships among the factors? 3. How do the interactions or relative importance of the factors change between the early stage of an information seeking process and the later stage of the process?

Qualitative studies have not been able to answer these questions. Quantitative methods may be a possible way to provide insights into these questions. In this study, a questionnaire and confirmatory factor analysis are chosen as the quantitative methods to answer these questions. Measurable variables need to be identified in advance for questionnaire design. Since measurable variables are not available from literature, qualitative methods, i.e., interviews and content analysis, are being applied to identify these variables.

Interviews were conducted to collect qualitative data from college students in a Midwestern university while they were seeking information for their research papers. Students were interviewed twice at different stages of their information seeking processes. Interview results were transcribed through IBM ViaVoice 9, in April 2002. Based on interview results, content analysis was then used to identify variables related to students' thoughts, feelings and actions. Three inter-coders were recruited for the coding scheme reliability test. The average of an 89% match between the inter-coders' and the researcher's coding schemes demonstrated that the variables identified were stable. According to the frequency of variables being mentioned during interviewed and variables' relevancy to students' information seeking behaviors, 22 of 34 variables were chosen for designing a questionnaire. Forty-seven statements were developed based on students' answers from interviews. A 7-point Likert scale was used to measure these statements. A pre-test of the questionnaire was finished in summer 2003. The modified survey will be distributed in Fall 2003 and Spring 2004. The questionnaire will also be distributed at two different stages of students' information seeking processes when they are working on their research papers. Quantitative data collected will be used to perform confirmatory factor analysis and canonical correlations afterward.

Note: The study has been presented in the Annual Connection Conference (Connections 8: Quest for Knowledge, 6/6-8, 2003), at Long Island University, Long Island, New York on June 7, 2003. The abstract will be published in the Canadian Journal of Information and Library Science 27(4), 2003.

Dan Albertson
daalbert@indiana.edu

Video Searching and Browsing Using ViewFinder

ViewFinder is a system developed for the purpose of video retrieval. Throughout this research, we are exploring many issues surround video retrieval systems. These issues include user-interface design, data indexing and representation, and query modeling.

ViewFinder is currently being developed and evaluated in participation with the Text REtrieval Conference (TREC) and its video track. We first entered results of ViewFinder in the interactive search task of last year's (2002) TREC video track. Here, we took an existing version of ViewFinder and applied it to the video, image, and textual data provided by the conference. Our entry consisted of 1 interactive search run, where all 25 search topics were completed. After submitting the results, each shot was compared to the relevant shots (which were manually identified by human assessors from TREC), and deemed to be relevant or not. This allowed us to compare ViewFinder with other video retrieval systems developed by highly regarded researchers worldwide.

We intend on continuing participation in TREC and are currently preparing for TREC-VID 2003 (now a separate workshop offered through TREC). This year proves to be even more challenging than the previous years due to an expanded data set and stricter task definitions. Our tentative plan still consists of participating in the interactive search task, where we will enter one search run based solely on content-based retrieval and a second which consists of searching by closed caption transcripts and/or automated speech recognition outputs.

Aside from TREC, previous research using Viewfinder consisted of exploring ways to incorporate it into the CLIOH project at IUPUI. CLIOH is a digital archive project, which features digital content of the artifacts and treasures of the ancient world (e.g. Mayan civilization). The content in the CLIOH archive includes digitized video, and this is where Viewfinder was intended to be used. Viewfinder would allow users to retrieve video content from the archive.

Bi-Yun (Ruby) Huang
bihuang@indiana.edu

**Portraying Falun Gong: A Content Analysis of
World News Media Coverage
(March 1999-March 2003)**

This study presents the results of a content analysis of news stories related to Falun Gong, an ancient form of qigong banned by the Chinese government in July 1999, as reported by six online global news media sources, CNN (United States), CBC (Canada), BBC (United Kingdom), theage (Australia), Voice of Democratic (Hongkong), and TaipeiTimes (Taiwan). A total of 120 news items randomly selected from news media between March 1999 and March 2003 were systematically examined through quantitative content analysis. Scholars and media have perceived Falun Gong as a “spiritual movement”, “religion” or even a cult but practitioners claim that Falun Gong is not a religion. Thus, the goal of this paper is to examine how the news media define Falun Gong and the depictions used in these news sites.

The findings indicate that there is a variation existing in the terminology used by these six news media sources in their stories about Falun Gong. Several categories have low frequencies such as meditation, qigong, truthfulness, compassion & forbearance, which are emphasized by Falun Gong. Some categories indicate that there is a difference in understanding between media and Falun Gong. It is concluded from this study that Falun Gong is portrayed as a religion-like movement by these media sources.

Daniel O. Kutz
dokutz@indiana.edu

24 Hours of News: Micro -Longitudinal Content Analysis of CNN and BBC

People are increasingly depending on the World Wide Web to be the first accessible medium for obtaining news on current local and international events. Questions that arise are, what kind of material gets reported and presented on the website of a major news outlet? Do news agencies offer up-to-date information while providing a wealth of news that is not otherwise available through traditional media resources such as television and newspapers? How often are news webpages updated? Does the editing process occur online or offline? To answer these questions, a variation of longitudinal content analysis is proposed, which aims toward a better understanding of the chronological events that occur as a frequently-updated website undergoes a standard 24-hour time period.

A spider program was used to collect 24 hours of data on April 24, 2003 from the Cable News Network (CNN) and British Broadcasting Corporation (BBC) websites. Data was compiled in 10-minute intervals. During this period, 270 pages were retrieved from these sites (135 per domain). A total of 270 pages were analyzed for both websites. Out of the 135 pages analyzed on each website, changes were identified on 54 pages for BBC and on 74 pages for CNN.

Analysis showed that content on a news site is updated on an average of every 10 minutes during normal business hours (7:00am to 5:00pm), with changes occurring less frequently in the late evening and overnight hours (9:00pm to 6:00am). When a webpage is updated, 2.5 news items are generally added or modified on the page. These values fluctuate however, as the number of changes to a single page in a 10-minute period have been as high as 7 (BBC, 11:43am) or as low as 1 (BBC, 8:10pm).

A primary advantage of conducting micro-longitudinal content analysis is that it facilitates understanding of webpage evolution during a one-day time period. Furthermore, the editorial process is more apparent in the case of news sites. In traditional media outlets such as cable television and newspapers, the viewer is confronted with a final product, which has previously undergone fact-checking and editorial processes. The web, on the other hand, is a more porous medium with emphasis on speed. Such a process requires first sending the message, then refining it once in the public eye; a process that may be unconsciously perceived by readers that visit news websites only once or twice per day. A micro-longitudinal study captures minute changes that may only exist for short amounts of time; by viewing these small changes, one can attain a better understanding of functions, ideology, and the editorial processes applicable to news organizations.

Kazuhiro Seki
kseki@indiana.edu

A Hybrid Approach to Named-Entity Recognition in Biomedical Texts

We propose a novel method for named-entity recognition, in particular for identifying protein names in biomedical texts. Named-entity recognition is fundamental in, for example, information retrieval and text mining but not trivial in the biomedical domain. We use a set of heuristics for the initial detection of protein names and use a probabilistic model for determining their name boundaries, which incorporates suffix-based word classes to generalize the model. In contrast to previously proposed methods, our approach does not rely on natural language processing tools such as part-of-speech taggers and syntactic parsers so as to reduce processing overhead and the potential number of probabilistic parameters to estimate. A notion of certainty is also proposed to improve precision for identification. Based on our proposed method, we implemented a protein name identification system and evaluated the system on real-world biomedical texts in conjunction with the state-of-the-art protein name identification system developed by other researchers. The results showed that overall both systems performed comparably and that our system outperformed the other for compound names. In addition, it is demonstrated that our system can further improve precision by restricting the system output to those with high certainties.

Ning Yu
nyu@indiana.edu

Email Visualization Abstract

Currently, email has become not only an important individual communication tool, but also the most widely used business productivity application. On the other hand, email overload is a growing problem for many users in their workplaces, mainly because the filing and maintenance of emails are very time consuming and cognitively intensive. Some research has been done in this area. Unfortunately, progresses still need to be made.

I am primarily analyzing email in education and research communities, as it has become one of the most important communication tools and information sources in these communities. To be more specific, everyday a professor will get large amounts of emails from other researchers, students or mail-lists which cover several topics. Some of the emails are of immediate importance; while some may be needed later and some may be trash mails. An email inbox is a *repository* of “to-dos”, “to-reads”, and “research discussion” items not only for the researchers, but for the students as well. In education, especially remote education, students and teachers need to keep track of their progress by email. The instructor may need to answer the same questions from time to time for different students. Even if there is a mail-list, which focuses on a particular area, not all the emails in it will interest everybody who subscribes to this list.

I am conducting a series of visualizations on a dataset of email. My goal is to test whether visualization techniques can help email management. So far, I am more interested in the static (history) email analysis than the real-time email analysis, although dynamic visualization may be imported later. Further step will be taken in using visualization to reveal the email thread and email burst in order to detect the social network.

Three related visualizations will be presented.

Visualization for Email Research. This is a visualization for my literature review of the Email Research area. The dataset is the 86 papers from the bibliography list of emailsearch.org. Latent Semantic Analysis (LSA) is applied to the mined dataset (title and abstracts of each entry) to determine the degree of similarity among documents, as well as Force Directed Placement (FDP) for final results visualizations. Pathfinder Network Scaling is chosen for the resulting similarity matrix visualization which will provide an overview of the email research area and help to determine the future direction of this visualization.

Mailbox Hierarchy Structure Visualization. This project aims to speed up email retrieval by employing hyperbolic tree as a new navigation way. Also, color coding is applied to help processing the pending tasks (refers to the unread email). The dataset is the one week transaction folder in the mailbox of Walden Virtual Library (<http://www.lib.waldenu.edu>).

Time-Series Statistic Visualization. The main goal is to visualize the overall email transaction flow during the year 2002 in the Walden Virtual Library, checking the overall mail transaction flow based on each month, day of week or time in a day, and checking the number and/or length of the email messages based on a given period of time. Moreover, based on the statistics, librarians can also get an idea of how much interaction they have had with some particular student and then take this as a clue for offering further assistance.