

# EmVis – A Visual E-mail Analysis Tool

Bjoern HECKEL

Department of Computer Science/Center for Image  
Processing and Integrated Computing (CIPIC)  
University of California, Davis  
Davis, CA 95616-8562

heckel@cs.ucdavis.edu

Bernd HAMANN

Department of Computer Science/Center for Image  
Processing and Integrated Computing (CIPIC)  
University of California, Davis  
Davis, CA 95616-8562

hamann@cs.ucdavis.edu

## 1. ABSTRACT

**EmVis is an information visualization tool for the analysis of e-mail exchange among people in an organization. A company-specific organizational hierarchy is used to provide the basic layout for a graph-based visualization of e-mail exchange. Various parameters are mapped to the nodes and edges in this graph: rank within the company, frequency/amount of e-mail exchange between subjects, possibly information content. The system we describe also supports more general graph-based visualizations not necessarily based on a company's hierarchy; it is possible to use EmVis for queries regarding any group of employees and the e-mail exchange among them. We employ generalizations of traditional information visualization concepts, such as cone trees and hyperbolic space representations.**

### 1.1 Keywords

Analysis Methods: analysis of contents of particular domains. Domain Specific Designs: information retrieval. User-Interface Design Issues: visual output strategies, graphic representations.

## 2. INTRODUCTION

In our global workplace of today, people are communicating in different ways than ever before. E-mail, in particular, has taken on a new meaning. At the outset, e-mail was primarily used for personal communications. Today, businesses are using e-mail for all varieties of intra-organization messaging.

We know that e-mail is heavily used. But these questions remain: who is communicating with whom, and what are

they saying? An analysis tool can help executives understand how information moves throughout their organizations, as well as see patterns of communication between people. Through this, gaps can be found where more communication must occur. Further, the value-address of the communications can be examined. The implications are enormous, in that some people may not have the information they need, while others use their time unproductively. Upon analysis, the derived knowledge can be used for training purposes to improve interactions between people.

To understand communication in organizations, it is important not only to understand who communicates with whom, but also to know what is contained in the messages and how this information is spread throughout the organization. Visualization, coupled with document analysis, provides the foundation for the development of EmVis, an e-mail visualization tool, to examine these issues.

## 3. OVERVIEW

An e-mail data set is characterized by a hierarchically organized set of nodes that are partially connected. This hierarchy describes the structure and linkages within the organization. Each node in the hierarchy represents an individual. The links between these nodes, furthermore referred to as "streams," represent a set of messages exchanged between two people. The network of streams represents the entire pattern of communications in the organization.

Several methods (e.g. utilizing cone-trees [1], [2] and hyperbolic space [3], [4], [6]) have previously been developed to visualize hierarchies and networks. The focus of the current project is to combine and further develop these methods and integrate them with document/data analysis and data warehousing. The resulting tool will be used to unveil communication patterns by allowing the user to examine the involved objects and create different views on the data set. Furthermore, the tool will provide a platform for a more general exploratory tool.

## 4. FEATURES

The first prototype provides two representations, based upon cone-trees that lay out the structure of the organization. This approach allows the user to view a limited number of streams, so as to reduce potential confusion. But as will be described later, different visualization techniques will be implemented for future prototypes. This will allow a comprehensible display of a larger quantity of streams.

The first prototype focuses on the display of hierarchical and individual communications. In "chain-of-command mode," messages sent between a single person and his/her manager or between the person and his/her direct reports are displayed. In "individual mode," all communication streams that a single individual or a group of individuals generates or receives are visualized.

A colored node, representing a single person, shows that person's rank within the organization. By using white to indicate high rank through dark gray to indicate low rank, the user is able to maintain an orientation while manipulating the three-dimensional visualization.

The size of the node indicates the number of messages a person has received or sent, which are not displayed in the visualization. For example, in "chain-of-command mode," a person sending many e-mails outside his or her hierarchy would be represented by a large node. Conversely, a person who communicates only within the hierarchy would be represented by a small node.

The color of streams shows the number of messages sent in a certain direction and within a particular time interval. Red indicates very active communication from one person to another, while blue signals limited communication. Colored arrows on both ends of the stream indicate the direction of this communication.

Further information about the visualized objects can be retrieved by activating associated methods (e.g. "show" on a stream object to display a message list).

EmVis will be extended by the following features:

- Displays of all streams at one time, utilizing hyperbolic space: the geometric representation of the organizational structure will be transformed to hyperbolic space. The resulting graph will be projected onto a disc and displayed, focusing on a certain node. In addition, nodes can be elevated as a function of their rank. As a result, the data set will be projected onto a cylinder-shaped space.
- Graph-theory-oriented clustering on communication streams: in contrast to using a static layout of the graph that is derived from the structure of the organization, the geometrical representation can be determined by analyzing the e-mail messages and their contents. Workgroups are then identified and

can be used to validate the existing structure of the organization.

- Animation displaying the change of the communication flow over time.
- Document analysis, allowing message filtering and classification by content. Coupling document analysis with animation enables the display of how certain information spreads out over time. Normal or Body Text

## 5. SUMMARY

A first prototype has been developed to visualize e-mail data. This tool has proved to be useful in examining communications within organizations. It provides a method to analyze a large volume of heterogeneous data in a comprehensible manner.

The visualization tool will be extended to visualize more than e-mail data sets. E-mail data is seen only as an instance of information represented through a collection of objects, classifications and links. The generalization of EmVis will display nodes as representations of objects and object aggregates of various types (e.g. documents, video and sound files). Streams will represent link, similarities, or other objects connecting nodes, while the hierarchy describes a classification of the nodes. The object types will be organized into a class hierarchy. Operations such as clustering and filtering, as well as very type-specific algorithms, will be provided to organize and manipulate the representation of objects. Streams and hierarchy can be changed dynamically by user interaction, creating a different view on the data set. The tool will enable the user to explore the characteristics and functionality of objects through a visual interface.

## 6. ACKNOWLEDGMENTS

The development of EmVis has been supported by the IBM Almaden Research Center, Digital Library Group.

## 7. REFERENCES

- [1] J. Carriere, R. Kazman. Interacting with Huge Hierarchies: Beyond Cone Trees. In *Proceedings of IEEE Information Visualization '95 Symposium*, IEEE, 1995.
- [2] G. Robertson, S. Card, J. Mackinlay. Cone Trees: Animated 3D Visualizations of Hierarchical Information. In *Proceedings of the ACM SIGCHI '91 Conference on Human Factors in Computing Systems*, ACM Press, 1991.
- [3] C. Gunn. Visualizing Hyperbolic Geometry. In *Computer Graphics and Mathematics*. Eurographics, Springer Verlag, 1992.

[4] M. Phillips, C. Gunn. Visualizing Hyperbolic Space: Unusual Uses of 4x4 Matrices. In *Symposium on Interactive 3D Graphics '92*, volume 25, ACM Siggraph, 1992.

[5] J. Lamping, R. Rao, P. Pirolli. A Focus+Content Technique Based on Hyperbolic Geometry for Viewing Large Hierarchies. In *Proceedings of the ACM SIGCHI '95 Conference on Human Factors in Computing Systems*, ACM Press, 1995.

[6] T. Munzner, P. Burchard. Visualizing the Structure of the World Wide Web in 3D Hyperbolic Space. In *Proceedings of the VRML '95 Symposium*, ACM Siggraph, 1995.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

NPIV '97 Las Vegas Nevada USA

Copyright ACM 1998 1-58113-051-1...\$5.00