

# Mail Explorer - Spatial and Temporal Exploration of Electronic Mail

Daniel A. Keim, Florian Mansmann, Christian Panse, Jörn Schneidewind and Mike Sips

Department of Computer and Information Science, University of Konstanz, Germany

---

## Abstract

*In today's world, e-mail has become one of the most important means of communication in business and private lives due to its efficiency. However, the problems start as soon as mail volumes go beyond the scope of human information processing capabilities. Firstly, time does not allow for leaving certain messages unanswered for a long time, and in certain cases, for reading all messages. Secondly, the dilemma of electronic filters leaves a choice of too many junk mails getting through versus a risk of solicited mails being dumped.*

*In this paper we present a new interactive visual data mining approach for analyzing individual e-mail communication. It combines classical visual analytics (help to identify pattern such as peaks and trends over time) with geo-spatial map distortions (help to understand the routes of e-mails). Experiments show that our visual e-mail explorer produces useful and interesting visualizations of large collections of e-mail and is practical for exploring temporal and geo-spatial patterns hidden in the e-mail data.*

Categories and Subject Descriptors (according to ACM CCS): H.4.3 [Communications Applications]: Electronic Mail

---

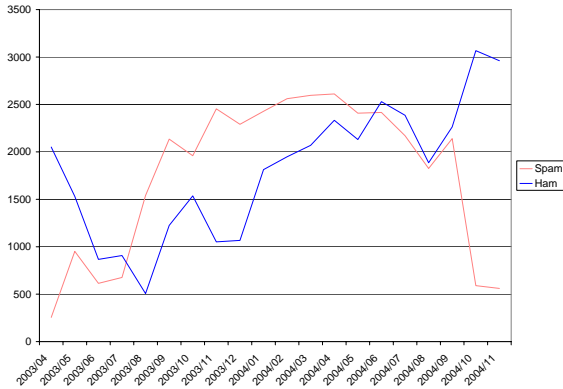
## 1. Introduction

E-mail is often described as the 'killer app' of the internet [Od10]. As Odlyzko points out the internet is mainly used for communication, not for obtaining content. On the one hand it becomes harder and harder to imagine everyday life without e-mail. According to the study [LV03], the daily flow of e-mail was estimated to be 1,829 TB in 2003 and will have doubled by 2006. On the other hand, everybody knows how difficult it is to manage a crowded personal inbox after some days of e-mail absence. Along with the rapid growth of this vast information flow comes an increased interest in the possibilities of analyzing these data, of extracting patterns and insights, as this might be interesting and useful to simplifying the management of the personal inbox. Without such possibilities, the personal frustration levels rise through each daily inbox clean-up.

Finding interesting large-scale patterns is an important task for tools which explore large collections of e-mails. It is important for personal use to respond quickly to major communication partners. Data analysts and business managers

in large companies especially, who receive thousands of e-mails every day, need to recognize and respond to changing conditions or major customers within minutes. In such large companies fraud detection becomes an important task in e-mail communication. The understanding of patterns which classify fraud, Spam and regular e-mail (ham) at full scale is still a challenging research area. Already in 1982, Peter J. Denning realized in his article [Den82] that electronic junk had become an ever-growing problem. In our department, one single user received 2308 mails in a busy month, out of which 1377 were spam. An automatic spam filter is used in order to filter out the unsolicited mails. Figure 1 shows some interesting statistics to illustrate these problems.

Nowadays there are many approaches from statistics, data mining and machine learning to separate Spam and regular e-mail communication. Such filters on specialized servers can help to sort e-mail automatically, but at the same time, many users feel uneasy with leaving decisions about their communication channels open to such clients. In principle, there are two important problems. The first one is that users



**Figure 1:** This chart shows the e-mail statistics of three of our IMAP users. From July 2003 to May 2004 we received more spam than solicited messages. Infrastructural changes in September 2004 decreased the spam volume considerably.

do not have insight into the filter and do not have the possibility to tune the filter in practice. The goal of our visual e-mail explorer at this point is to allow a faster, flexible and more intuitive way to explore very large collections of e-mail and therefore give the user insight into their e-mail communication pattern. Insight is the key to understanding and fine-tuning such filter algorithms. The user is still tightly integrated in the decision-making process. Privacy issues come up as many filter algorithms are based on certain statistics about the e-mail behaviour of the users, and it is very easy to use such data to apply data mining. Our visual mail explorer is a powerful and intuitive tool on the user's desktop and gives them the possibility to analyze their e-mail communication on IMAP e-mail servers. Interesting features are highlighted so that the users will recognize them easily.

The second problem involves the time needed to manually scan the daily e-mail amount. Smart visual analytics and interaction methods for fast recognition of important and unimportant messages are needed.

By offering the possibility to browse through the folder hierarchy of an IMAP mail account, the user is equipped with a powerful tool that takes advantage of his very own way of organization. Our visualizations were designed to be highly scalable, which means that we are able to handle several thousand messages at a time.

## 2. Previous Work

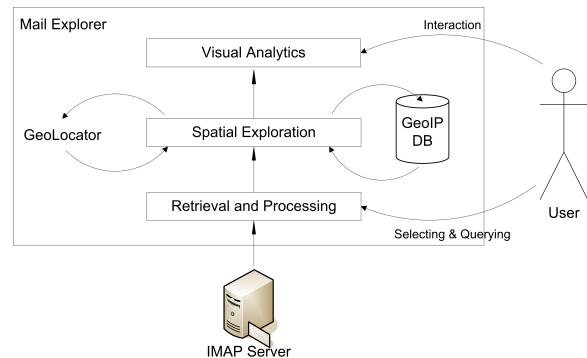
In 2003 Bernard Kerr presented a new visualization technique named Thread Arcs [Ker03]. As a small-scale visualization, its purpose was to give a visual representation of an e-mail thread within an e-mail client. Relationships among the postings are expressed through arcs. The technique preserves the chronology of the thread and gives an impression

of its size. Special attention was paid to the criteria of compactness and scale as only limited display space is available. Each message is to stay on the same spot within the visualization as the thread grows. Interaction techniques allow selecting and highlighting certain postings and their relationships. The study focuses on analyzing communication threads, whereas our study regards e-mail communication in a broader scope.

In another study in 2004 [KPSS04], we applied the classical visual analytics such as Parallel Coordinates, 2D Level Plots and Cartograms to large data cubes of e-mail. Some of these previously-used techniques lack a fast and intuitive way of interaction with the data and their application is closely related to data preprocessing. In our new application Mail Explorer, the preprocessing step is completely integrated and we utilize the commonly used IMAP protocol. In this way, any IMAP mailbox can be easily analyzed.

## 3. System Architecture

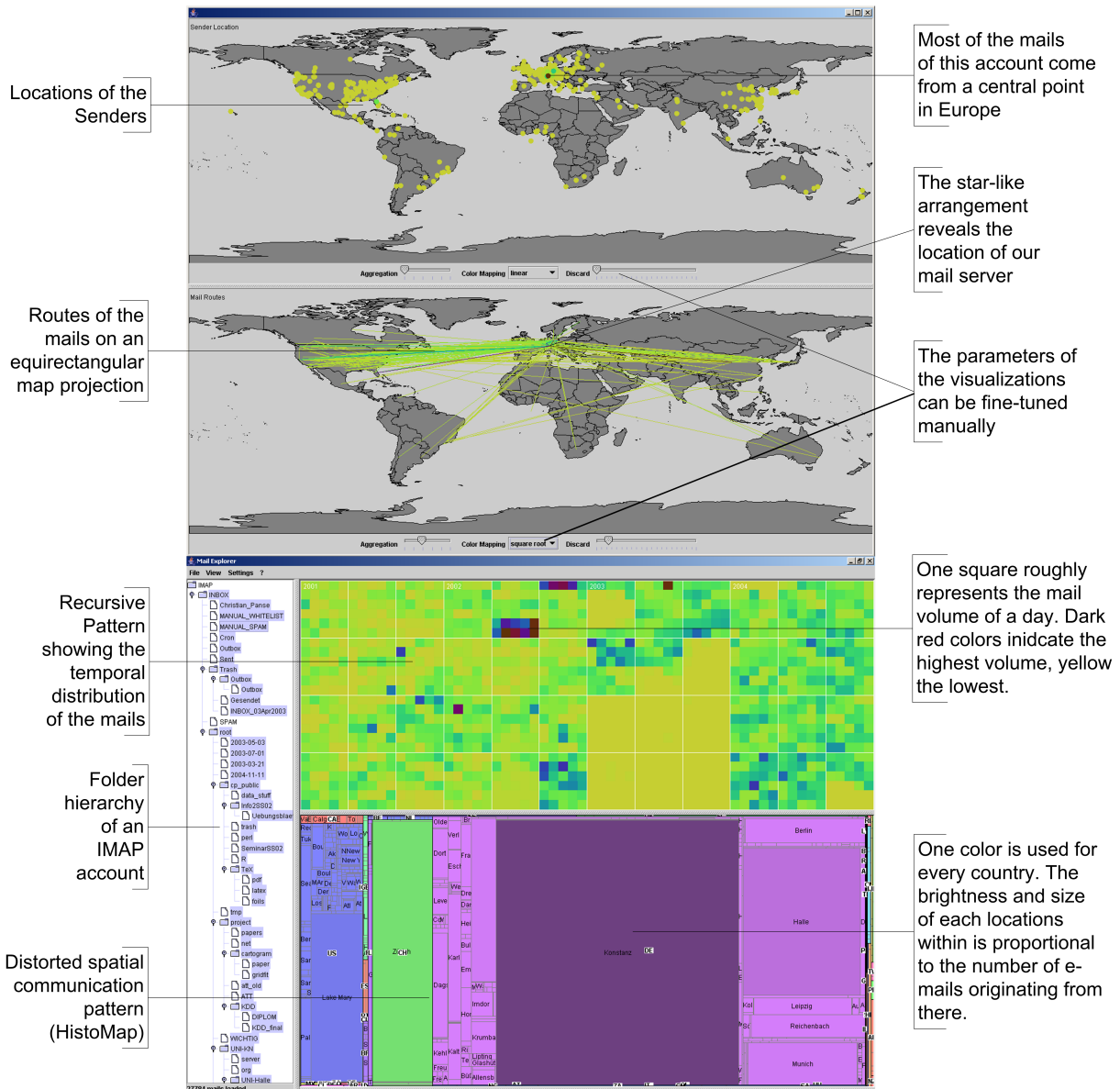
Our Mail Explorer is based on a three layer architecture to reduce dependencies to both machine architectures and data formats. Figure 2 illustrates the three layer architecture. The lowest layer provides a well-defined interface to the internet message address protocol (IMAP). IMAP provides a flexible mechanism to access e-mail from all over the world for users and a well-defined protocol to manage large collections of e-mail for system administrators.



**Figure 2:** Three layer architecture of the Mail Explorer – Retrieving, exploring and browsing large collections of e-mail

The Mail Explorer browses through the user's e-mail folder hierarchy and retrieves the headers. The headers of the mails are a wealthy source of different information. A single header of an e-mail contains several heterogeneous attributes.

For instance, information about the current conditions of the network over which the e-mail was sent is contained in the header. The network consists of nodes, links, and statistics about the network itself. It is very easy to extract basic



**Figure 3: Overview first** – The Mail Explorer Interface has a number of configurable visual analytics. The user has four overview visualizations to analyze temporal and geo-spatial properties of the e-mail communication. He can identify interesting patterns: here, for example, most sender locations are in the US and Europe. This can be explained by the fact that most research partners of one of the authors are located in US and Europe. Some e-mails with exotic sender locations such as China are probably spam.

information about e-mails from the headers, such as the internet addresses of the nodes, the time at which the e-mail was sent, the load of the network and so on. These raw data associated with the nodes and links of the network are subject to changes over time. This data extraction is done in the retrieval layer of the Mail Explorer. The exploration of large

collections of e-mail is in practice a difficult task and important observations can be made.

The data listed in the headers are measured within the network under circumstances that cannot be controlled by the user. The information is often incomplete (missing values) or has some bias because large networks are highly non-

uniform real world applications. These problems are handled in the preprocessing layer. The challenge here is to address these problems in such a way that no information that was not included in the headers is added or that no relevant patterns are lost. The next layer extracts geo-spatial information contained in the e-mail headers. This layer starts DNS lookup queries to resolve the internet address of each node listed in the header. Then, the internet address is used to start queries on a geo-ip database. The geo-ip database contains the geographic locations of most of the internet addresses. The Mail Explorer uses the Maxmind GeoIP City Database [Max] (further reading about the basic ideas of GeoIP databases and GeoLocator see [HKP\*04]).

After applying the processing steps in these layers we have a well-defined set of information about a large collection of e-mails. Formally, this set consisting of a large collection of e-mail can be seen as a massive database  $DB$  with spatio-temporal, non-spatial and possibly geo-spatial information:  $DB = \{\{\lambda_1, \varphi_1, t_1, a_{1,1}, \dots, a_{1,m}\}, \dots, \{\lambda_n, \varphi_n, t_n, a_{n,1}, \dots, a_{n,m}\}\}$  where  $\lambda_i \in [-\pi; \pi]$  represents longitude,  $\varphi_i \in [-\pi/2; \pi/2]$  latitude,  $t_i$  the time and  $a_i \in \mathbf{R}^m, n, m \in \mathbf{N}$  the attributes of the  $i$ -th data item. All these data classes have specific characteristics and it is very important to take these characteristics into account when analyzing the e-mail data.

Our Mail Explorer offers several visualizations for each of these characteristics. Currently, scatter-plots on the coordinates, histograms on time zones, recursive patterns on time series and advanced geo-spatial visualizations called *HiToMap* are integrated. The open system architecture allows for further integration of visualization techniques.

#### 4. Exploratory Data Analysis

Our Mail Explorer follows a three step process: overview first, zooms and filters, and details on demand called the information seeking mantra [Shn96]. At first, the user gets an overview of the complete e-mail communication in the inbox (see Figure 3). He can also browse through his personal mail hierarchy on the IMAP server. In this overview the user is able to identify interesting communication patterns or some groups of data. Take, for example, a business manager in a large company. Such a user receives a lot of important business e-mails over one day. If such a person is away from the office for some days, the inbox is flooded with important e-mails. The task is now to get first a visual overview of the inbox communication, then to prioritize the e-mails that belong to the different business relations and in case of discussions to find the last entry of such communication threads. A second important issue is spam and fraudulent e-mails within his/her inbox. With the help of our tool, the business manager is now capable of easily exploring exotic sender locations from where he/she has never received any business mail for example. The spatial distribution of the e-mail communication helps to identify individual business

customers very quickly (e.g. a gold customer in the US, a silver customer in China) or to detect falsely classified messages in a dense e-mail folder as they are often spatial outliers. Time properties of the e-mail facilitate the identification of communication threads and assist the user in clearing the backlog in a certain business discussion.

To focus on one or more of these patterns, the Mail Explorer provides a flexible focus on highly-populated areas of e-mails rather than on areas with low communication appearance. The user can choose familiar land covering world maps with e-mail paths from source to receiver and also distorted maps which show the distribution of the e-mail communication around the world. Some classic statistics help the user to explore the temporal properties of his communication. These tasks are pretty different from classical statistics and data mining approaches in which patterns are often identified with some lower user satisfaction. Remember the business manager in the large company. He/she can apply filter mechanisms to show only gold customers and can zoom into the communication of some of them. After the identification of some gold customers in Manhattan who have sent some business e-mails, for example, he/she applies a filter process to remove all other customers and zooms in on the communication of this specific customer. The manager can immediately see that the e-mails are all from the customer's employees. The recursive pattern indicates that a lot of those messages arrived two days ago. It seems like something happened that day.

For analyzing interesting patterns in detail, the user can drill down to see details of individual e-mails. In the case of our business manager, he/she accesses the individual e-mails of the communication thread. By reading the important mails first, he obtains an idea of the problem and is capable of reacting quickly to it.

The personalized visualization techniques do not only provide the basic techniques to present an overview to the user, but they also bridge the gap between the three steps, using on the one hand linking and brushing and on the other hand intuitive interaction methods. The analysis of large collections of e-mail can be seen as a hypotheses generation process. First the user gets an overview of the temporal and geo-spatial properties of his e-mail communication. The visualization of the data with linked temporal and geo-spatial visual analytics allows the user to identify interesting communication patterns. After identifying such patterns the Mail Explorer allows the user to gain a deeper insight in the data. Thereby he confirms or rejects his/her hypotheses. In the following we present the exploratory analysis belonging to temporal and geo-spatial attributes in detail.

#### 5. Gain Insight: Temporal Properties of E-Mail Communication

A main aspect of the Mail Explorer concerns the investigation of the temporal attributes of large collections of e-mail

data. This is information such as the time an e-mail was sent or received, the number of received e-mails per hour or per day or the time zone an e-mail was sent from. The goal is to analyze e-mail behaviour over time, e.g. the detection of peak hours when most e-mails arrive or the growth of the amount of spam over time. By detecting temporal patterns in his e-mails, the user may be able to classify his e-mails as important or unimportant or to fine-tune his e-mail filter. A number of techniques for analyzing and visualizing temporal e-mail information are employed in the Mail Explorer.

Year	Q1	Jan	Feb	Mrch	1 .. 6 1 .. 6 1 .. 6 12 .. 7 12 .. 7 12 .. 7 18..13 18..13 18..13 19..24 19..24 19..24 30..25 30..25 30..25	
		Q2	Apr	May	June	1 .. 6 1 .. 6 1 .. 6 12 .. 7 12 .. 7 12 .. 7 18..13 18..13 18..13 19..24 19..24 19..24 30..25 30..25 30..25
			Q3	July	Aug	Spt
	Q4			Oct	Nov	Dec

**Figure 4:** Scheme of the recursive pattern used. Four years are placed next to each other. They are vertically subdivided into four quarters with three months each. A month consists of a matrix of the size 5x6.

Besides common techniques such as line-, bar- and pie charts, the Mail Explorer provides the recursive pattern visualization technique (for details see [AKK95]) which allows for the detection of the temporal aspects of the electronic messages. This technique is based on a recursive two dimensional arrangement of the underlying data points and is especially useful if the data points have an ordering according to one dimension, in our case for example the received-time attribute of an e-mail. Figure 4 illustrates the basic idea of our recursive pattern scheme.

The resolution of the data displayed is only limited by the amount of pixels on the screen.

Figure 5 a) shows the folder for project e-mails. Several days are colored in a bright color, which means that none or only a small number of e-mails came in on those days. The red and darker colored days mark high traffic volumes. The lower right corner is always left blank as all three measurements were made at the beginning of December.

The second image (b) shows all solicited e-mails of another mailbox in 2004. There were almost no messages in the first few days of the month January, probably due to extended Christmas vacations. Overall, the distribution of e-mails over time seems to be fairly constant in this mailbox.

The third image (c) is from a re-opened mailbox that was closed for five years. As it was used for many newsgroup

postings that are still harvested for e-mail addresses by many spammers, the mailbox filled up with 2200 junk mails within the last four months. A clear upward trend is visible from August to the beginning of December. It is very interesting that the trend has stopped and the traffic has decreased considerably in the last few days. We'll keep an eye on that mailbox.

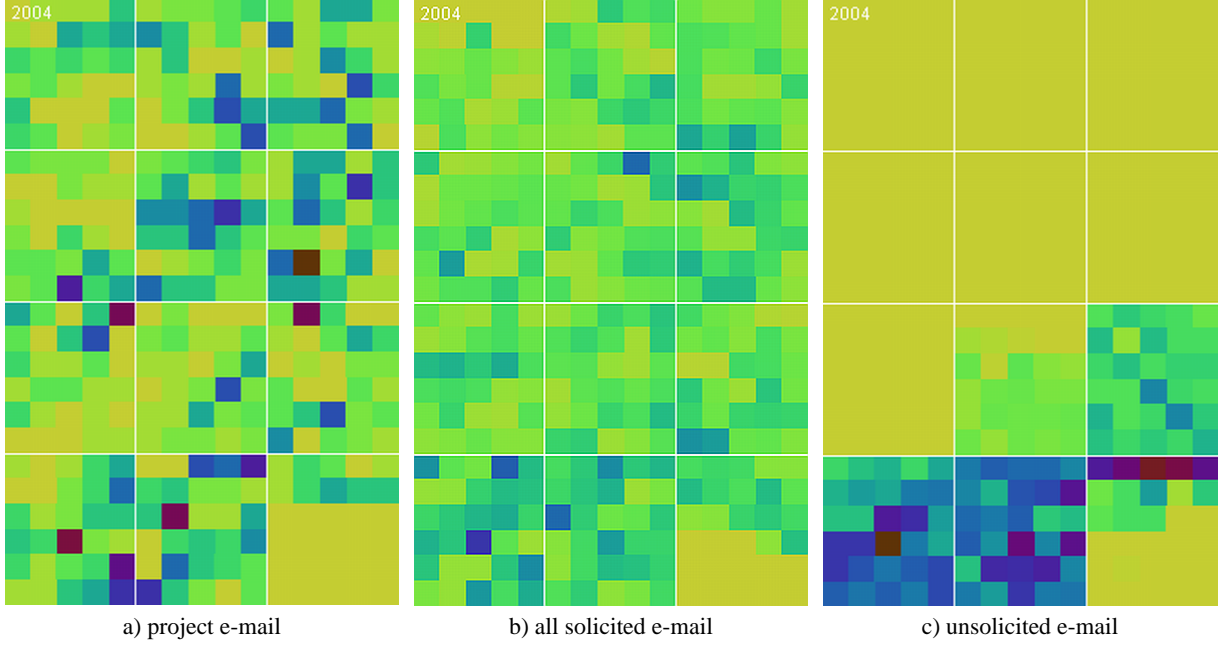
## 6. Gain Insight: Spatial Patterns in E-Mail Communication

Besides the temporal aspects of e-mail traffic, it is possible to use the Mail Explorer to analyze and visualize the spatial aspects of e-mail data. Since a geographic location can be extracted from an e-mail header either by analyzing the time zone or the ip addresses of the sender, the question is how to visualize this geographic information on the screen.

A common way to visualize spatial data is to use a two dimensional map and plot the data points to x/y locations on this map corresponding to the spatial location of each data point. The problem is that the overlapping of data points, which may occur at dense point areas, is not seriously addressed through this technique. On the other hand, a lot of screen space stays unused, since only few data points are located there.

To address these problems, we utilized the HistoMap algorithm which is a tree-map-like [Shn92] space-filling visualization technique. Given a data set of spatial data, a rectangle is assigned to each tuple. The size of this rectangle has to be proportional to the statistical value of the data, in this case the number of e-mails from that location. At the same time, some constraints have to be held up. The geographic relationship of those rectangles should reflect the spatial relationship of the spatial data points. At the same time, all display space should be used.

We have an efficient heuristic to this problem. Our algorithm splits the world into six horizontal bins, assigning an equal span of longitude to each one. Then, each bin is resized according to the sum of the statistical values of all data elements contained in that bin. The same procedure is applied recursively alternating between vertical and horizontal bins until there is only one data point left. To obtain better results, we aggregated all data points of a country into a big rectangle. Then we applied the same algorithm again within the rectangle of the country. We fine-tuned the algorithm within the country rectangles in such a way that the direction of the next split depends on the resulting rectangle. The optimization criterion was to have more quadratic rectangles as those are visually easier to compare than very long and thin rectangles. Several numbers of bins have been tested and the number six has shown the most realistic geographic layout.



**Figure 5: Details on Demand** – shows three plots of different mail traffic of the year 2004 in a Recursive Pattern. Each square represents the mail volume of one day. Dark colors indicate high e-mail traffic.

## 7. HistoMap: A Space-filling Value-by-Area Technique

HistoMap combines *PixelMap*, *HistoScale*, and *RecMap* which were currently introduced (see [KNPS03, KNP\*03, HKPS04] for more details). HistoMap is different to the described approaches in so far that we have x-y-location and a statistical value as input and polygonal (rectangles) data as output.

The visualization goals are as follows: (a) using the full screen size for the visualization (no holes), (b) each location should be realized as a rectangle or at least as a display pixel, (c) the relative geographic position of each location should be preserved, and (d) the aspect ratio of each box should be equalized.

More formally: As *input* we have a point set  $\mathcal{P} = \{p_1, \dots, p_n\}$  and  $(p_i)_{i=1, \dots, n} \in \mathbf{R}^2$  and a vector  $\mathcal{X}$  of statistical values where  $\mathcal{X} = (x_i)_{i=1, \dots, n}$  with  $x_i > 0$  and  $x_i \in \mathbf{N}$ . The quality of  $\bar{\mathcal{P}}$  depends on two aspects. First, we want a space filling visualization where the area of each rectangle corresponds to the statistical value of each location and second, the “map” should be easily recognized as in  $\mathcal{P}$ .

These goals can be achieved using the following three objective functions. The *absolute point position* distance measures the difference between the x-y-location and the center of the resulting rectangle. The *absolute point position*  $d_{AP}$

can be expressed by

$$d_{AP} = d_{AP}(\mathcal{P}, \bar{\mathcal{P}}) \quad (1)$$

$$= \sum_{i=0}^{n-1} |p_i - \tilde{p}_i| \quad (2)$$

and the *relative point position*

$$d_{RP} = d_{RP}(\mathcal{P}, \bar{\mathcal{P}}) \quad (3)$$

$$= \sum_{i=0}^{n-1} \sum_{j=0, j \neq i}^{n-1} (|p_i - p_j| - |\tilde{p}_i - \tilde{p}_j|)^2 \quad (4)$$

which may be used as measures for the achievement of the neighborhood. The *aspect ratio error*  $d_{AR}$  reflects the average relative deviation of the aspect ratios of the rectangles in  $\bar{\mathcal{P}}$  and can be determined as follows:

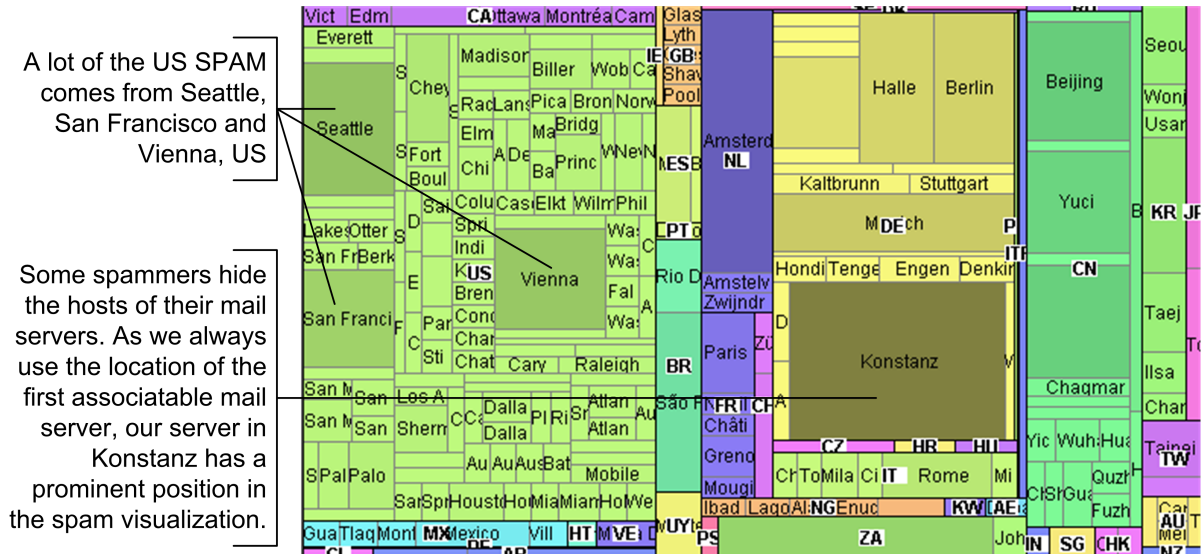
$$d_{AR} = d_{AR}(\mathcal{P}, \bar{\mathcal{P}}) \quad (5)$$

$$= \frac{1}{n} \cdot \sum_{i=1}^{n-1} \left| 1 - \frac{dy_i}{dx_i} \right| \quad (6)$$

The distance function  $|p_i - p_j|$  can be defined by an  $L^m$ -norm ( $m = 1$  or  $2$ )

$$|p_i - p_j| = \sqrt[m]{(p_i^x - p_j^x)^m + (p_i^y - p_j^y)^m}. \quad (7)$$

The *output* can be defined as an optimization problem. The *output* is a non overlapping, planar, and space filling map partition  $\bar{\mathcal{P}}$  where the weights  $\{a, b, c\} \in \mathbf{R}$ ,  $a + b + c =$



**Figure 6: Zoom and Filter** – Gain insight into spam. The HistoMap visualization shows a typical spam folder in our department. Colors are used in order to distinguish between the countries; brightness and size indicate the number of e-mails from the corresponding spatial location .

1.  $\{a, b, c\} \geq 0$ , and

$$\sum_{i=1}^{n-1} A(\vec{p}) = x_i \quad (8)$$

$$a \cdot d_{RP} + b \cdot d_{AP} + c \cdot d_{AR} \text{ is minimized.} \quad (9)$$

Our solution is based on the approaches mentioned above. The *HistoScale* algorithm is extended in such a way that we have an arbitrary number of histogram bins which divide the screen space into rectangular partitions. Similar to the *RecMap* approach for each bin we alternate between horizontal and vertical direction of the binning until each bin corresponds to one x-y-location and the area to the statistical value. A meta heuristic is used to compute several candidate transformations which are evaluated by the weighted objective functions. For each candidate transformation we use a different number of bins. We can also use a flexible number of bin in each binning step. Using an arbitrary meta heuristic [Neu96], e.g. a genetic algorithm, it is possible to minimize to the weighted objective functions. Since a fast computational time is important for an interactive visualization it is also possible to allow just one candidate transformation. The best results were achieved when we used a binning of 6 for 30000 emails. An example can be seen in Figure 6.

## 8. Results

After the visual exploration of several mailboxes, we made some interesting discoveries. Certain characters such as business partners, colleagues and friends can be easily associated to spatial locations. Once the name of the city they

live in appears in our visualization on the screen, it becomes clear from whom those e-mails are. Combined with background knowledge, conclusions about the sender and even about the content of those messages can be drawn.

When dealing with huge amounts of e-mails, forwarder often play an important role. The visualization of the routes in combination with a zoom has revealed that a lot of mails of one of our IMAP users is forwarded from Halle. As spam volumes are incredible large and a burden to almost every mail user, we took a closer look at the phenomenon. Indeed, our HistoMap visualization displayed a very international spatial pattern of the spam senders (see figure 6). Solicited e-mail usually comes from a few countries, and depends on the internationality of the user's e-mail communication pattern. However, our spam folder's traffic came from more than 40 different countries.

Concerning the temporal aspects, we confirmed the hypothesis that different e-mail shows different temporal patterns. Private e-mail and business e-mail seems to have a rather uniform distribution (Figure 5 a) and b)). Through our recursive pattern visualization, we were able to easily make out the dates of high and low e-mail volumes. In one of our mailboxes which had been closed for five years, we run an interesting experiment. Within four months the re-opened mailbox filled up with 2200 spam mails as the address was previously used in many newsgroups and on several webpages. The recursive pattern shows (Fig 5 c)) a strong upward trend from August to the beginning of December 2004. Afterwards, the volume fell back to the traffic level of mid-

August. It seems like infrastructural changes had taken place to centrally stop junk mail.

## 9. Conclusions

Due to increasing e-mail traffic caused by the growing communication over the internet, the analysis of e-mail traffic is becoming more and more important for each single user. Problems like SPAM and malicious e-mails make the analysis of e-mail databases even more important. Additionally, due to the high number of received e-mails, many users experience difficulties in keeping track of their e-mail inboxes and e-mail folders. Since typical e-mail clients primarily concentrate on the sending, receiving and storing of e-mails, they do not adequately address all these problems.

In this paper we introduced a powerful visualization tool that offers novel possibilities for visual analytics of large e-mail databases, called *Mail Explorer*. The *Mail Explorer* adapts multiple visualization techniques from the fields of visual data mining and visual data analysis. Using these techniques, the user is able to discover and gain insight into the different aspects of his e-mail data, including temporal and spatial patterns. We showed that the *Mail Explorer* provides techniques, such as the *Recursive Pattern* for the analysis of temporal e-mail aspects as well as techniques for the investigation of spatial characteristics, such as the *HistoMap* approach. Due to the modular design of the *Mail Explorer* other visualization and interaction techniques may be integrated as well.

The presented experimental results, e.g. the analysis of our working group's IMAP accounts including more than 100 000 e-mails, show the usefulness of our exploration tool. Our *Mail Explorer* provides a user-friendly interface and may be used with any IMAP server. First analysis results with the *Mail Explorer* as well as the user feedback were very promising and show the potential of this novel visualization tool.

## 10. Acknowledgements

This work has been funded by the German Research Society (DFG) under the grant GK-1042, Explorative Analysis and Visualization of Large Information Spaces, Konstanz as well as FP06/03 Network Profiling, Network Intrusion Visualization and Network Security, University of Konstanz. The authors also thank Roland Heilmann for his great support.

## References

- [AKK95] ANKERST M., KEIM D. A., KRIEGEL H.-P.: Recursive pattern: A technique for visualizing very large amounts of data. In *Proc. Visualization '95, Atlanta, GA* (1995), pp. 279–286.
- [Den82] DENNING P. J.: Electronic junk. *Communications of the ACM* 25, 3 (1982), 163–165.
- [HKP\*04] HEILMANN R., KEIM D. A., PANSE C., SCHNEIDEWIND J., SIPS M.: Finding spatial patterns in network data. In *From Data to Patterns: Int'l Workshop on Pattern Representation and Management, PaRMA04, Heraklion - Crete, Greece, March 18* (2004).
- [HKPS04] HEILMANN R., KEIM D. A., PANSE C., SIPS M.: RecMap: Rectangular Map Approximations. In *InfoVis 2004, IEEE Symposium on Information Visualization, Austin, Texas* (October 2004), pp. 33–40.
- [Ker03] KERR B.: Thread arcs: an email thread visualization. In *IEEE Symposium on Information Visualization* (October 2003), pp. 211–218. IBM Research.
- [KNP\*03] KEIM D. A., NORTH S. C., PANSE C., SCHÄFER M., SIPS M.: HistoScale: An efficient approach for computing pseudo-cartograms. In *IEEE Visualization 2003 DVD-ROM, Seattle, Washington, USA* (October 2003), pp. 28–29.
- [KNPS03] KEIM D. A., NORTH S. C., PANSE C., SIPS M.: PixelMaps: A new visual data mining approach for analyzing large spatial data sets. In *The Third IEEE International Conference on Data Mining (ICDM03), Melbourne, Florida, USA* (November 2003).
- [KPSS04] KEIM D. A., PANSE C., SCHNEIDEWIND J., SIPS M.: Analyzing large collections of e-mail. In *IKE'04 - The 2004 International Conference on Information and Knowledge Engineering, June 21 - 24, 2004, Las Vegas, Nevada, USA* (2004), CSREA Press.
- [LV03] LYMAN P., VARIAN H. R.: How much information, 2003. Retrieved from <http://www.sims.berkeley.edu/how-much-info-2003> on 12/06/2004.
- [Max] MAXMIND: Geoip city database. MaxMind LLC, <http://www.maxmind.com>.
- [Neu96] NEUMANN K.: *Produktions- und Operations-Management*. Springer, Berlin, 1996.
- [Odl01] ODLYZKO A.: Content is not king. *First Monday* 6, 2 (February 2001). AT&T Labs - Research <http://www.dtc.umn.edu/~odlyzko/doc/history.communications2.pdf>.
- [Shn92] SHNEIDERMAN B.: Tree visualization with treemaps: 2-d space-filling approach. *ACM Transactions on Graphics* 11, 1 (January 1992), 92–99.
- [Shn96] SHNEIDERMAN B.: The eyes have it: A task by data type taxonomy for information visualizations. In *IEEE Visual Languages* (1996), pp. 336–343.