

UNIVERSITY OF OKLAHOMA
GRADUATE COLLEGE

VERIVIS: EFFECTIVE SPAM VERIFICATION USING HIGHLY
INTERACTIVE MULTI-DIMENSIONAL VISUALIZATION

A THESIS
SUBMITTED TO THE GRADUATE FACULTY
in partial fulfillment of the requirements for the
Degree of
MASTER OF SCIENCE

By
MOHAMMAD DIBAYMOGHADAM
Norman, Oklahoma
2013

VERIVIS: EFFECTIVE SPAM VERIFICATION USING HIGHLY
INTERACTIVE MULTI-DIMENSIONAL VISUALIZATION

A THESIS APPROVED FOR THE
SCHOOL OF COMPUTER SCIENCE

BY

Dr. Christopher E. Weaver (Chair)

Dr. Amy McGovern

Dr. Dean F. Hougen

Acknowledgments

I would like to express my deepest gratitude to my committee chair, Prof. Chris Weaver, for supporting me throughout my studies. Working under Prof. Weaver was one of the best things to happen to me during my graduate studies at the University of Oklahoma.

I would also like to thank my other committee members, namely Prof. Amy McGovern for introducing me to machine learning. Machine learning is one of my favorite fields of study and I am happy that I took the course with her. Thank you Prof. McGovern for the Halloween candies. Thanks to coach Dean Hougen for supporting us at the ACM programming competition. I had great experience being his TA for the operating system course.

Thanks to Prof. K. Thulasiraman for being a cool professor in general and introducing Networks and optimization. Thank you Prof. K.T for making me feel like a friend.

Thanks to my parents and my lovely sister for encouraging me all the time to pursue my Masters Degree in Computer Science. Thank you for Skyping with me from thousands of miles away, and giving hope when I was really frustrated.

I would also like to thank to my fellow Lab members Maryam, Sayantani and Naveed for being supportive, including walking with me to Starbucks every time to get my latte. Finally, thanks to my friends Masoud, Hossein, Solmaz, Alireza, and others for giving me hope and energy.

Table of Contents

Acknowledgments	iv
List Of Figures	vii
Abstract	x
1 Introduction	1
1.1 Fundamental Concepts	1
1.1.1 Spam	1
1.1.2 Spam Filtering	2
1.1.2.1 Spam Filtering Errors	2
1.1.3 Spam Verification	3
1.2 Thesis Contribution	4
1.3 Organization of the Thesis	5
2 Background	6
2.1 Spam and Spam Filtering	6
2.2 Spam Visualization	7
2.3 Background Summary	11
3 Problem and General Approach	12
3.1 Information Visualization	13
3.1.1 Multi-view Visualization	15
3.1.2 Dynamic Queries	16
3.2 Email as Multi-Dimensional Data	17
3.2.1 Header Data Attributes	17
3.2.2 Content Data Attributes	19
4 Design and Implementation	21
4.1 Architecture	21
4.2 Processing Component	21
4.2.1 Trec07p Email Corpus	21
4.2.2 Trec07p Attribute Extraction	23
4.2.2.1 Email Attributes in VeriVis	24
4.3 Visualization Component	27
4.3.1 Views	27
4.3.1.1 Email Table Views	28
4.3.1.2 Calendar View	30
4.3.1.3 Scatter Plot of Size vs. Number of Lines	31

4.3.1.4	The Senders-Receivers Bipartite View	32
4.3.2	Filtering	36
4.3.3	Highlighting	40
5	Application	44
5.1	Non-spam Identification: Sample Activities	44
5.1.1	First Sample Activity	45
5.1.2	Second Sample Activity	54
5.2	Spam Exploration: Sample Activities	58
5.2.1	First Sample Activity	58
5.2.2	Visual Outlier Detection: Sample Activity	60
6	Future Work and Conclusion	68
	Reference List	71

List Of Figures

1.1	Spam filtering classification's errors	3
2.1	Thread Arcs visualization, redrawn from [11], here showing all sets of individual messages that are related to each other through the “reply” attribute.	8
2.2	Social Network Fragments visualization, redrawn from [28].	9
3.1	Example of a tabular view of messages.	12
3.2	The VeriVis user interface.	14
3.3	Interactive visualization of messages, with two coordinated views. . .	16
3.4	Sample raw email, reproduced from Trec07p email corpus [7].	18
4.1	The architecture of VeriVis.	22
4.2	Distribution probability of email attributes in a subset of the Trec07p corpus.	25
4.3	Input schema to the VeriVis visualization approach.	26
4.4	CSV files of email message and location records.	26
4.5	The junk box view, showing all received spam messages in the email server.	29
4.6	The verified view, showing all user-verified spam messages.	29
4.7	The calendar view, showing received spam messages over time.	30
4.8	The calendar view, showing total counts for each week.	31
4.9	The calendar view, showing total message counts for each day of the week.	32
4.10	Scatter plot of message size vs. number of lines: (a) Lighter circles show that a small number of messages have the same size and number of lines; (b) Darker circles show that a large number of messages have the same size and number of lines; (c) User selected spam messages are illustrated with blue (dark) edges.	33
4.11	The bipartite view, labels on the left are senders' email addresses; labels on the right are receivers' email addresses.	34
4.12	The bipartite view, here showing messages from sending top-level domains to receiving top-level domains. At bottom are controls to aggregate senders and receivers by domain (checkboxes), control line translucency (slider), and choose which attribute to color senders and receivers on (combo boxes).	35
4.13	The bipartite view, showing messages between senders' top-level domains and receivers' top-level domains for messages in French.	37
4.14	The bipartite view, revealing those top-level domains that were more involved in supposed spam messages in French than those in other languages.	38

4.15	The filtering control panel, here with filtering on attachments only. . .	39
4.16	The filtering control panel for Boolean attributes, with filtering on for messages that have attachments.	40
4.17	Attribute table views, showing the unique values and corresponding message counts for each attribute.	41
4.18	The highlighting control panel, with highlighting on for the “Language” attribute.	42
4.19	The bipartite view, with highlighting and filtering of messages in the sports category.	43
5.1	The non-spam identification first sample activity: starting state of the bipartite, scatter plot, and junk box views.	46
5.2	The non-spam identification first sample activity: selecting two weeks prior to the receipt date in the calendar view.	47
5.3	The non-spam identification first sample activity: searching for and selecting CNN-related subject line tokens in the token words table view.	47
5.4	The non-spam identification first sample activity: choosing to filter on Date in the filtering control panel.	48
5.5	The non-spam identification first sample activity: viewing filtered spam messages in the bipartite and scatterplot views.	48
5.6	The non-spam identification first sample activity: selecting “mail.cnn.com” as a sender domain address in the bipartite view.	49
5.7	The non-spam identification first sample activity: viewing highlighted non-spam messages from “mail.cnn.com” in the junk box view.	50
5.8	The non-spam identification first sample activity: recovering non-spam messages from “mail.cnn.com” to the verified view.	51
5.9	The non-spam identification first sample activity: selecting a sender domain address in the bipartite view.	52
5.10	The non-spam identification first sample activity: viewing a spam message from “gestyre.com” in the verified view.	53
5.11	The non-spam identification first sample activity: removing the received spam message from “gestyre.com” from the junk box view.	54
5.12	The non-spam identification second sample activity: initial visualization state for received spam messages in the “flax9.uwaterloo.ca” email server.	56
5.13	The non-spam identification second sample activity: selecting desired days of the week and token words in the corresponding table views.	56
5.14	The non-spam identification second sample activity: filtering data records in the bipartite, scatter plot, and junk box views.	57
5.15	The non-spam identification second sample activity: selecting sender domain address in the bipartite view.	58
5.16	The non-spam identification second sample activity: recovering email messages from the “shareholder.com” domain.	59
5.17	Spam exploration first sample activity: initial visualization state of the bipartite, scatter plot, and junk box views.	61

5.18 Spam exploration first sample activity: filtering on selected attribute values in the calendar and word token table views, using the filtering panel.	61
5.19 Spam exploration first sample activity: viewing patterns in VIAGRA spam messages received in April 2007.	62
5.20 Spam exploration first sample activity: using the highlighting control panel to color selected countries in the country table view.	62
5.21 Spam exploration first sample activity: viewing VIAGRA spam messages colored on their source country in the bipartite view.	63
5.22 Visual outlier detection activity: initial visualization state for messages from the “.ca” top-level domain.	64
5.23 Visual outlier detection activity: selecting outlier spam messages in the scatter plot.	65
5.24 Visual outlier detection activity: highlighting selected (outlier) spam messages in the junk box view.	65
5.25 Visual outlier detection activity: the verified view, showing three messages identified as outliers by mistake.	66
5.26 Visual outlier detection activity: spam messages mistakenly detected as outliers (circled) in the scatter plot.	67
5.27 Visual outlier detection activity: highlighting of spam messages in the verified view based on selections in the scatter plot in figure 5.26. . .	67

Abstract

Spam is one of the problems commonly encountered by users of email. *Spam* is a common term for unwanted, indiscriminate, and disingenuous messages delivered to a large number of recipients [8]. The purpose of spam is to stealthily deliver a payload to steal personal information, advertise a product, or infect a computer with a virus.

To protect users from spam messages, service providers apply filtering to identify spam messages at the server level. Spam filtering consists of automated techniques to identify spam messages and prevent them from getting into users' inboxes. The spam filtering process involves three steps: calculation of a *spamminess* score, classification of messages as spam or *ham* (non-spam) based on that score, and internal knowledge about similar messages previously reported as spam, and collection of messages identified as spam. Given individual differences between people and their situation-specific definitions of spam, existing spam filters are more effective than might be expected. But not all of them are able to identify all types of spam messages without misclassification errors.

Spam verification is a method to assess the accuracy of spam filtering. It allows people coming from different perspectives to apply their own spamminess criteria to the messages that spam filtering has detected as spam. It helps users to identify, mark, and recover non-spam messages among those classified as spam, and to remove the unmarked ones. Spam verification can be performed either by server administrators in email servers, or by end users on their personal junk box.

VeriVis is a highly interactive visualization tool for spam verification in email servers. The tool displays multiple interactive views of common email attributes used in spam filtering, such as “To,” “From,” “Date,” “Time,” and useful attributes derived from message content, such as “Category” and “Language”. VeriVis allows server administrators to observe and analyze patterns of message attributes. Using VeriVis, server administrators can interactively highlight, select, and filter spam messages in terms of multiple attributes. This helps server administrators identify and reclassify non-spam messages according to their situation-specific spamminess criteria. Finally, VeriVis allows server administrators to mark and recover non-spam messages among the messages that are classified as spam and delete the unmarked spam messages from email servers. *This thesis argues that VeriVis can help server administrators effectively verify received spam messages in email servers.*

Chapter 1

Introduction

1.1 Fundamental Concepts

Email is one of the most common forms of internet-mediated communications. Nowadays people from a wide variety of cultures use email for their communication needs. Email is a fast and cheap communication method compared to other methods and additionally makes it easy for users to share files and media.

1.1.1 Spam

One of the drawbacks of email, however, is the prevalence of unwanted, indiscriminate, and disingenuous message, known as spam. In general, spam is the result of an online social situation that was created through the deployment of communication technology in a community of users [35]. Email has been used as a rich communication channel by spammers to send unwanted, disingenuous messages indiscriminately to a large number of recipients without having any background relationship with them. Spammers transfer information which contains a payload that advertise a product (often illegal or non-existent products), set up computer malware to hijack victims' computers, or steal their personal information [8]. Spam can also cause problems for service providers, such as bandwidth consumption, storage consumption, and security issues.

1.1.2 Spam Filtering

According to MessageLabs, nearly three out of every four messages are classified as spam, and more than 75% of email traffic on the Internet is spam [6, 16]. The transmission of spam messages can have several negative consequences for both users and service providers.

Service providers use spam filtering as a largely automated means to identify and prevent spam messages from getting into users' inboxes. A spam message is effective if it can pass a spam filter's gates and get into a user's inbox. A spam filter is effective if it not only blocks spam messages, but also avoids misclassification of non-spam messages as spam [7]. To do their work, spam filtering approaches utilize the content of messages, their own knowledge and black-box algorithms, and their memory of previous messages. They also use feedback from users and a few external resources, such as DNS-based blacklist techniques, to identify a message as spam and protect users and service providers from possible negative consequences of spam (e.g., computer viruses and bandwidth consumption) [20].

1.1.2.1 Spam Filtering Errors

Spam filters are more effective than might be expected, given: (1) the different techniques that spammers apply to prevent spam filtering techniques from identifying them, and (2) individual differences between users and how they define a message as spam. Spam filtering techniques can accurately and rapidly eliminate up to 70% of spam messages in real time, using known classification algorithms [18]. Still, they cannot identify all types of spam messages in email servers and prevent them from getting into users' inboxes.

Spam filtering techniques suffer from *false positive* and *false negative* misclassification errors (Figure 2.1). Misclassifying a real spam message as non-spam is a false

	Detected as Spam	Detected as Ham
Spam	True Positive	False Negative
Ham	False Positive	True Negative

Figure 1.1: Spam filtering classification's errors

negative error. This type of error causes spam messages get into users' inboxes. Misclassifying a non-spam message as spam is a false positive error. This type of error can cause users to lose important non-spam messages.

1.1.3 Spam Verification

Because of spam misclassification errors, even popular email providers such as Gmail, Hotmail, and others need their users to check their junk box as well as their inbox to look for misclassified messages. This can help users assess the accuracy of spam filtering and identify and recover those messages that are misclassified as spam. These actions also send feedback to the servers of the email provider as part of the spam filtering system [8].

Spam verification can be useful for both users and service providers. For the purpose of this thesis, and because of a few limitations imposed by our input data, which is collected from an email server, we consider spam verification only in email servers. From now on in this thesis, "users" refers to "email server administrators" and "spam verification" refers to spam verification in email servers.

For the purpose of this thesis, we consider two different situations in which users perform spam verification in email servers. The first situation is when users have

a specification of what constitutes a non-spam message, possibly one misclassified as spam, and they use that knowledge to identify that and similar messages. The second situation is when users want to explore other aspects of spam to increase their knowledge of incoming spam messages in their email servers; they might use that knowledge to improve their spam filters, for instance.

1.2 Thesis Contribution

This thesis posits an affirmative answer for the following question: “Is there another way to represent all spam messages received by an email server to increase users’ awareness about spam messages and help them to perform effective spam verification?” VeriVis utilizes multiple visualization techniques and multiple coordinated views to help users perform effective spam verification in email servers.

This research described in this thesis contributes to the area of:

- applied visualization, by designing and implementing a highly interactive multi-dimensional visualization tool (VeriVis) for server level spam verification;
- spam verification at the server level, by designing a visualization tool for spam verification and studying how it can be effective for users to perform spam verification either for spam exploration or non-spam identification purposes in email servers; and
- visual outlier detection among received spam messages in an email server.

VeriVis allows users to visually identify those messages with different patterns of size (in bytes) and number of lines among the messages that are identified as spam, and mark them as outliers. It also highlights those marked messages throughout the visualization to help users get more descriptive information about those highlighted messages; they might use this contextual information to detect and evaluate outliers.

1.3 Organization of the Thesis

The remaining chapters in the thesis are organized as follows.

- **Chapter 2** presents background information on spam filtering and spam visualization. It also highlights the differences between VeriVis and other background research in both areas.
- **Chapter 3** Describes important concepts in information visualization as part of our approach. It also describes the attributes of messages as multi-dimensional data.
- **Chapter 4** describes the architecture of VeriVis and presents more detailed information about its design and implementation.
- **Chapter 5** defines multiple usage scenarios and assesses the degree to which VeriVis supports users to perform required actions in each scenario.
- **Chapter 6** concludes and lists potential directions for future work.

Chapter 2

Background

Email is one of the best-known internet-mediated communication tools. Email has been studied by researchers in different fields including computer science, communications, political science, etc. [25, 15, 10] Most of the research papers reviewed for this thesis are in computer science, and consider one or both of:

- spam and spam filtering, and
- email and spam visualization.

This chapter reviews research on these two topics. At the end, we highlight the differences between past work and the work described in this thesis.

2.1 Spam and Spam Filtering

To the best of our knowledge, there is no automated technique that can identify and classify all types of spam messages without misclassification errors. This makes it an interesting problem for researchers in information retrieval and machine learning [37], and results in debates on a formal definition of spam. A variety of research projects have applied both supervised and semi-supervised machine learning techniques to lower the incidences of false positive and false negative misclassification errors. Individual differences between users is a major factor in classification errors.

For example, SpamBayes is a Bayesian classifier for email classification which uses tiling unigrams and bigrams to produce better results than other heuristic and

classification algorithms. Using a combination of features such as a message scoring system, tokenization, and n-gram tailing, SpamBayes achieve high accuracy with few false positive errors [17].

Brodsky and Brodsky [6] has proposed a distributed spam classification system that operates independently of message content and can be used with other existing spam classifiers. This latter aspect may have increased usefulness as spam itself evolves. Previously, most spam messages were coming from marketing companies. Most of these messages were sent through a few fixed IP addresses. It was easy to identify spam by applying blacklisting methods. Nowadays, spammers try to hijack their victims' computers' IP addresses and use their computers to send their spam messages to other email receivers. For this reason, traditional blacklisting methods have become less effective and will eventually be considered obsolete. The distributed method proposed by Brodsky uses a combination of resource identification and peer-to-peer based distributed databases to identify and stop distributed spam messages.

SpamTracker is another spam filter system that uses behavioral blacklisting to defend against distributed techniques and botnets. Instead of maintaining a list of spam senders' IP addresses, SpamTracker tracks patterns in sending of spam messages. It can then use this behavioral data to distinguish spam from non-spam email and thus detect spammers that are missed by the blacklist. It is easy to integrate SpamTracker with current, deployed spam filtering techniques by making a few changes in the configuration of existing email servers [21].

2.2 Spam Visualization

Visualization researchers have designed single and multiple view visualizations to uncover latent patterns between multiple dimensions of message attributes in email

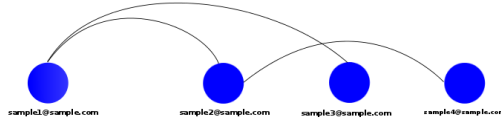


Figure 2.1: Thread Arcs visualization, redrawn from [11], here showing all sets of individual messages that are related to each other through the “reply” attribute.

archives and make it easier for regular users to analyze their message communications. Most email visualizations are designed to show the social network dimension of email archives using multiple message attributes such as “from,” “to,” “cc,” “subject,” and “date”. In this section, we review several email visualization tools and techniques. We also go over their similarities and differences in comparison to the multi-view visualization design used in VeriVis.

Thread Arcs [11] is a single view visualization technique similar to the Tree Diagrams and Tree Tables visualizations [9]. Thread Arcs helps users to find all sets of individual messages that are related to each other through the “reply to” attribute of messages, using message time stamps to determine the chronology and the sequence of received messages. Thread Arcs applies highlighting and attribute-shading features to visually separate nodes from each other in communication threads as a function of time, or of the roles of people in a thread (Figure 2.1). Thread Arcs serves to help people to see various attributes of conversations, find relevant messages between them, and organize their messages overall. This type of visualization can be useful for spam verification at the server level to group and categorize relevant spam messages that are in individual communication threads.

EzMail is a multi-view email visualization tool with the purpose of email management [23]. EzMail can help users manage and retrieve their information from an email archive using information visualization techniques. EzMail defines email threads as a collection of messages having the same topic, where the topic of a thread is the subject line of an email in that thread after removing “Re:” and “FWD:” from the

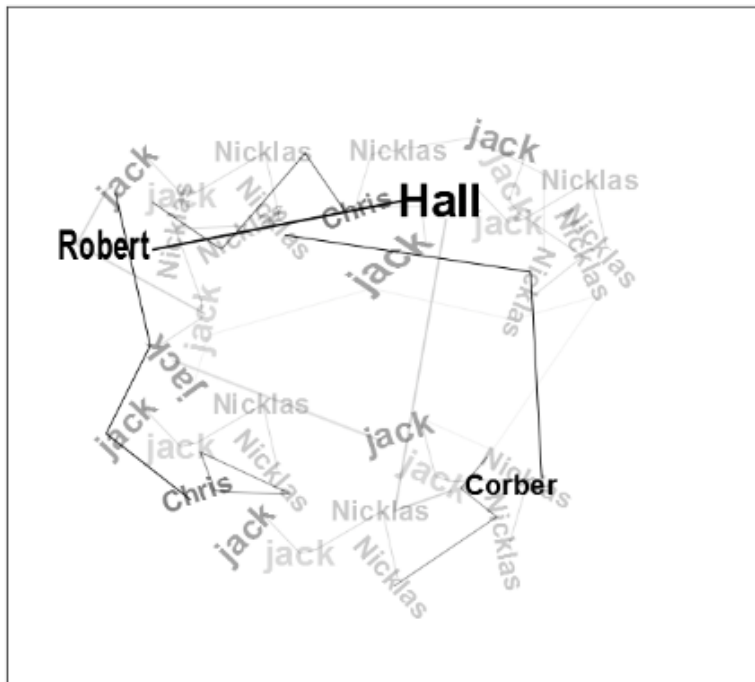


Figure 2.2: Social Network Fragments visualization, redrawn from [28].

beginning of its subject line. The email thread view in EzMail helps users to identify the sets of messages that constitute communication threads.

Social Network Fragments (Figure 2.2) is another type of visualization for highlighting the temporal patterns of interactions between individual clients and the social networks of senders and receivers. This type of visualization can help regular users to observe the visual structure of their social network and get more information about how they segment their social network into smaller clusters such as job, family communications, and relationships and differences between them [28].

After preprocessing our testing corpus, we could not find any message thread that involved both multiple senders and multiple receivers. Unlike the Thread Arcs and EzMail visualizations, VeriVis does not address analysis of the social network dimension of the email archive.

TimeStore is a multi-view visualization of a user’s inbox, which displays email communications between the receiver and various senders in a two-dimensional scatter plot. It uses the time of arrival as the principal attribute to view messages. TimeStore displays a list of senders on the y-axis and time on the x-axis. This type of visualization helps users to visually track their trends in real time to those senders [36]. Experience with TimeStore shows that displaying messages on a scatter plot can help users manage their received messages more effectively. Unlike TimeStore, VeriVis displays spam messages on a scatter plot based on other attributes of spam messages, such as number of lines and size of the email.

Ma and Muelder [19] present a set of visualization techniques that are designed to show patterns between incoming messages and that can be useful in revealing misidentified pieces of spam. They argue that, because of the nature of spam, it should be possible to distinguish them from non-spam messages regardless of their content. They used a bipartite view to display relations and communication patterns between groups of senders and receivers in an email server. We include a bipartite view in VeriVis, which can help users find possible spammers based on the one-to-many relationships between spammers and receivers while maintaining the privacy of email.

TRIB (Telescope for Responding comments for Internet Blogs) is a visualization system used to visualize blog articles and their related comments [12]. TRIB assumes that, based on the use of sequential list views in blogs currently, it is too complicated for users to search more than 10,000 messages to find which comments are useful. TRIB provides an interactive 2-D layout to show the whole collection as comment clusters. Unlike TRIB, which uses a user-defined keyword dictionary to classify comments and messages on a blog in terms of user interest, VeriVis uses the online Alchemy text categorization API (Application Programming Interface) [27] to

categorize messages into different categories, and treats the resulting email category as a derived data attribute for the purpose of visualization and filtering.

2.3 Background Summary

Spam filters consider multiple factors to classify a message either as spam or non-spam. One of these factors is the filtering system’s internal knowledge about similar messages formerly classified as spam. Part of the system’s internal knowledge comes from configurations that are applied by email server administrators.

Spam verification is an important part of the spam filtering process. Unlike most prior research in both spam filtering and email visualization, this thesis studies the effectiveness of applied multi-view visualization techniques on spam verification as a component of the spam filtering system in email servers. Nevertheless, there are similarities between VeriVis and earlier email visualization tools in terms of the visualization techniques and views that are common between them. VeriVis is designed to help users perform effective spam verification for the purpose of both non-spam identification and spam exploration. Both of these types of activities can increase user knowledge about received spam messages in the email server; users can also apply their knowledge to improve spam filters in the email server.

Chapter 3

Problem and General Approach

In this thesis, we consider two different situations in which users can perform spam verification in email servers. The first situation is when users have a specification about non-spam messages that are possibly misclassified as spam and they use that knowledge to identify those messages. The second situation is when users want to explore aspects of spam to increase their knowledge about incoming spam messages in their email servers; they might use that knowledge to improve their spam filters.

In both of these situations, being able to sort and filter spam messages based on their attributes can help users perform more effective spam verification. Most well-known email browsers display the full list of spam messages received by servers in chronological order and in a tabular format (e.g., Figure 3.1). Most browsers allow users to sort received spam messages based on their “title,” “subject,” or “time,” and filter messages using simple text searching or Boolean queries on attributes.

Email servers classify a myriad of incoming messages as spam every day. Presenting spam messages visually with added contextual information allows users to

sample@sampledomain.com	Sample Title	Sep 30
sample@sampledomain.com	Sample Title	Sep 30
sample@sampledomain.com	Sample Title	Sep 28
sample@sampledomain.com	Sample Title	Sep 27

Figure 3.1: Example of a tabular view of messages.

identify non-spam messages more effectively, and to take action about those messages. As Thomas and Cook [26] argue, email is multi-dimensional data, meaning that it has a wide variety of attributes that can be used in analysis, and visual representation of these attributes is one of the best ways to effectively support discovery of insights about their relationships. Moreover, multiple views leverage perceptual capabilities to improve understanding of relationships among data dimensions [3]. VeriVis (Figure 3.2) utilizes multiple visualization techniques and coordinated views to represent multiple attributes of spam messages in a coherent way. It also allows them to observe latent patterns between different attributes of spam messages by differently projecting multiple attributes of a set of messages to multiple coordinated views.

We argue that using multi-view visualization techniques can increase users' knowledge about incoming spam messages in email servers and help them apply their knowledge to effectively identify non-spam messages among the messages that are classified as spam. These techniques let users perform spam verification at the server level without necessarily having access to receivers' email addresses, which can be considered as private information by users. The rest of this chapter reviews information visualization concepts and techniques that are used in VeriVis, and then discusses spam as multi-dimensional data.

3.1 Information Visualization

Information visualization is a way to map numerical and non-numerical data into graphical characteristics that users can observe to gain knowledge about patterns within and between multiple attributes of data [13, 26]. VeriVis utilizes a variety of visualization techniques and view types to represent multiple spam attributes with the goal of supporting a spam verification process.

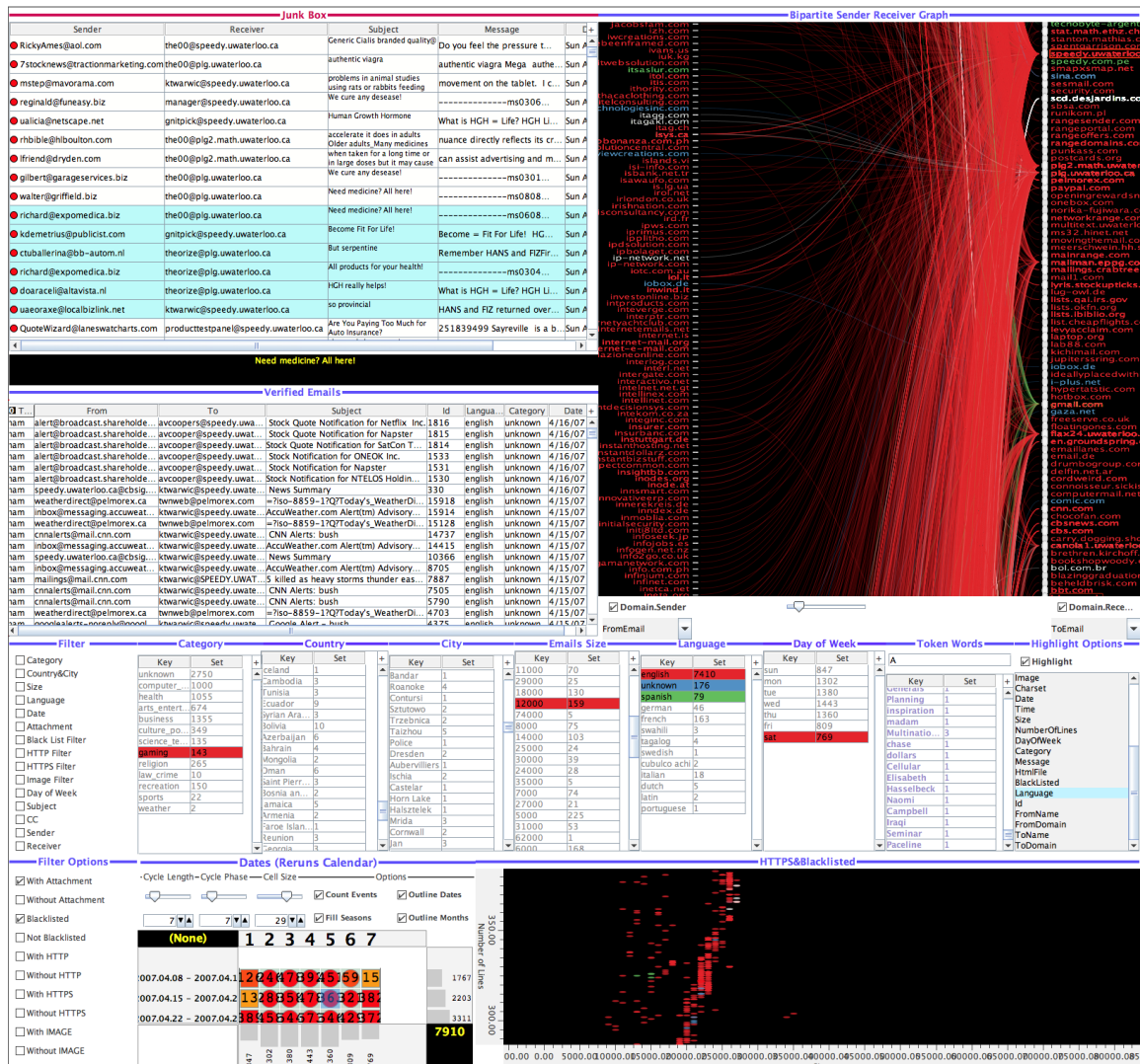


Figure 3.2: The VeriVis user interface.

To display data in views, we need to map data into visual representations using graphical characteristics such as position, size, shape, and color [14]. Some of these characteristics are useful to present quantitative information and some of them are good for qualitative information. For instance, area is one of the more effective characteristics for encoding quantitative information. But, it is also possible to (poorly) encode nominal information using area [14]. Multiple graphical characteristics can be combined to display different attributes of data together in a single view.

3.1.1 Multi-view Visualization

For many multidimensional datasets, a single view is insufficient to display all of the attributes needed for analysis. Presenting many attributes of data in a single view can also cause cognitive overload, impeding successful visualization [26]. Therefore, visualization designers often use two or more views to support investigation of a single dataset [3]. Moreover, coordination of the interaction between multiple views allows users to explore flexibly different combinations of data attributes [22]. Users can interact with one view that represents several data attributes, and observe the effects of their interaction on other views. For example, Figure 3.3 shows two coordinated views of a single underlying table of spam data. The scatter plot (top) plots the “size” attribute of messages; another scatter plot (bottom) plots the “number of lines” attribute. These two scatter plots allow users to compare size and number of lines in terms of the number of messages (left axis). A lens in the top scatter plot allows users to select messages within a range of sizes, then observe the number of messages that fall within the selected size range in the bottom scatter plot. Figure 3.3 reveal that messages under 2 KB in size have between 2,075 and 2,200 lines.

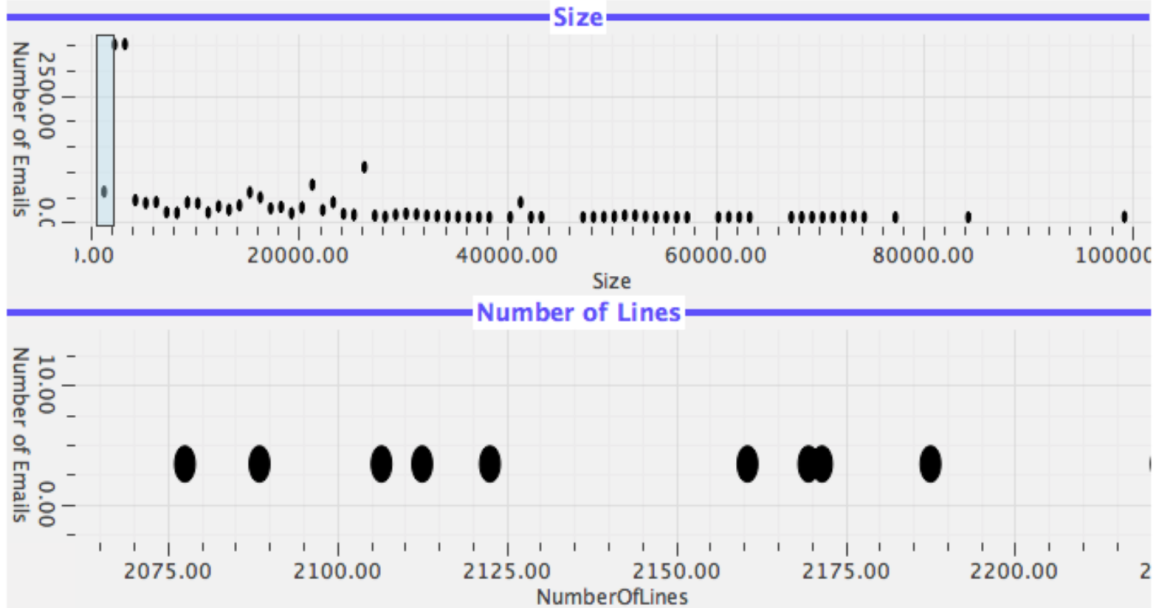


Figure 3.3: Interactive visualization of messages, with two coordinated views.

3.1.2 Dynamic Queries

Dynamic queries allow users to continuously update the value of multiple data attributes just by interacting with user interface components (e.g., sliders, buttons, etc.) [1]. Dynamic queries allow users to reverse changes they have applied to data by reversing their interaction with user interface components (e.g., by dragging the slider back to its former position). According to Ahlberg, et al. [2] “Being able to drag the slider left and right and get immediate updates of the query results, it is possible to do tens of queries in just a few seconds and it speeds up the querying process”. For example, to query data from databases such as MySQL, the user needs to understand the syntax of the SQL query language and needs to have knowledge about the data schema. Dynamic queries help users to rapidly query data from a database by interacting with graphical components and to observe the visual representation of the query result.

3.2 Email as Multi-Dimensional Data

VeriVis utilizes several visualization techniques to represent multiple attributes of spam messages. To fully understand this multi-dimensionality, it is necessary to understand all of the attributes inherent in a typical message. All attributes are found in either the header or the content of an email (Figure 3.4).

The header section contains primarily structured attributes such as sender and receiver email addresses, carbon copied (“cc’d”) email addresses, date, time, etc. The content section contains more semi-structured attributes such as the message text of an email, attached images, embedded HTML code, etc.

The remainder of this chapter gives more detailed information about the email attributes in the header and content of messages, and what kind of information they can provide for use in the spam verification process in VeriVis.

3.2.1 Header Data Attributes

Common attributes in the header section of the email include:

- **Sender email address:** This attribute contains the email address of the sender of the email.
- **Receiver email address:** This attribute contains the email address of the receiver of the email.
- **Date and time information:** This attribute contains the time and date that the email was received.
- **IP addresses:** The IP (“Internet Protocol”) address is the number used to identify a computer communicating with a network. Every email contains the IP addresses of all of the computers used to originate (and forward) it. This makes it possible to trace the route of any email.

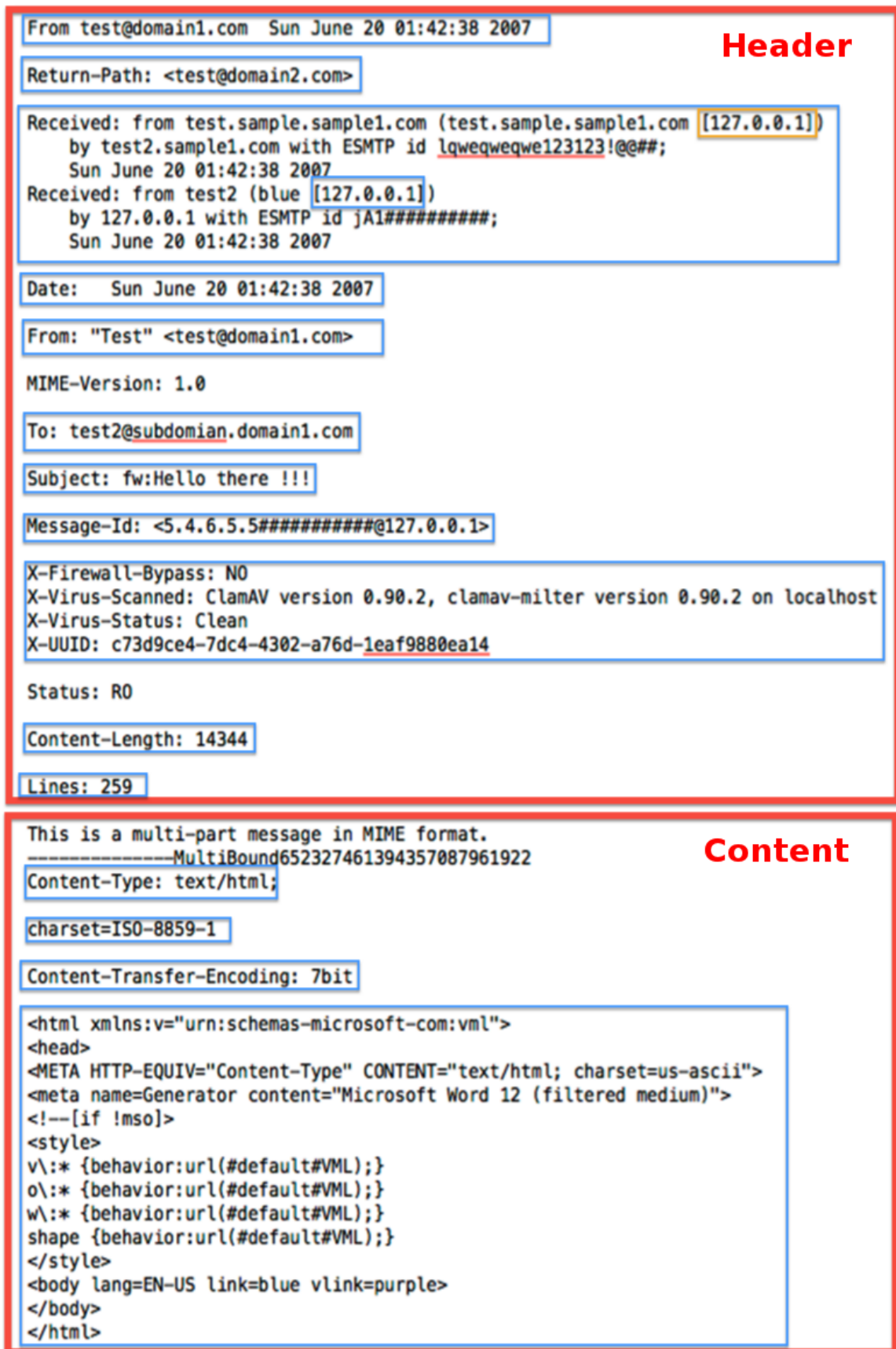


Figure 3.4: Sample raw email, reproduced from Trec07p email corpus [7].

- **Email size:** This attribute indicates the size of the email.
- **Number of lines:** This attribute indicates the number of lines of text in an email.
- **Virus detection flag(s):** Sometimes, service providers use antivirus applications at the server level to check if a message is infected with a virus. They include the result of their analysis in the header section of the email. Not all service providers provide this information for their end users.

3.2.2 Content Data Attributes

The following are email attributes that are either intrinsic to the content section of a message, or that can be computed as derived attributes:

- **Font:** If the message contains embedded HTML, the font attribute identifies the size and color of text.
- **Links (HTML and HTTPS hyperlinks):** This attribute indicates if message contains a link to another website.

The language and category are two derived attributes of each message. Using various machine learning and information retrieval techniques, it is possible to compute a wide variety of derived attributes that can be useful for both spam visualization and spam verification.

- **Language:** Using machine learning techniques, it is possible to detect the language of the message.
- **Category:** Using text categorization algorithms, it is possible to categorize message under multiple category labels such as “politics,” “art,” etc.

The next chapter presents the architecture of VeriVis and provides more information about the selection process of email attributes in the design of VeriVis and how it applies visualization techniques to display the selected attributes in multiple coordinated views.

Chapter 4

Design and Implementation

4.1 Architecture

The architecture of VeriVis consists of two fundamental components, shown in Figure 4.1. The first is the processing component, which is responsible for extracting multiple attributes from the whole input email corpus and for preparing the input data for the visualization component. The visualization component visually encodes input data records and displays them as graphical attributes in multiple coordinated views. These views allow users to interactively highlight, select, and filter spam messages according to their situation-specific spamminess criteria. It also allows people to mark and recover non-spam messages among those classified as spam and to delete the unmarked spam messages. The rest of this chapter explains these two components in more detail.

4.2 Processing Component

4.2.1 Trec07p Email Corpus

For this thesis, we seek an input email corpus that convincingly simulates a real spam folder for the purpose of spam verification. Ideally, we want a mixture of both spam and non-spam messages in our input email corpus. Finding a realistic input email

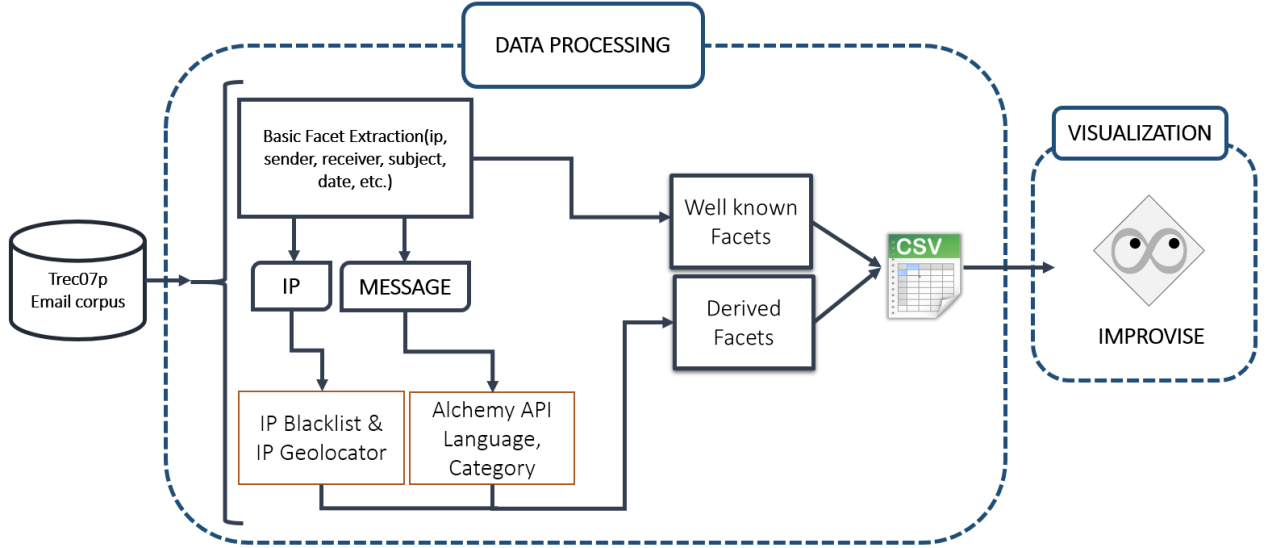


Figure 4.1: The architecture of VeriVis.

corpus is challenging, because most corpus creators choose to remove or alter user information to protect user privacy.

Trec07p is a popular public email corpus available for spam research. It is one of the most realistic email corpuses [8]. Trec07p contains 75,419 messages collected from a single email server between April 8 and July 6, 2007. The corpus is made up of 25,220 ham (non-spam) messages and 50,199 spam messages. This proportion of spam to ham simulates the snapshot of the contents of a realistic email server, as it contains both spam messages and misclassified non-spam messages. Trec07p also preserves most of the standard header and content attributes¹ such as “**senders and receivers email addresses,**” “**size,**” and “**number of lines**”. Still, because of user privacy issues, some common message attributes are removed or altered in the corpus.

¹<http://tools.ietf.org/html/rfc2822>

4.2.2 Trec07p Attribute Extraction

The processing component of VeriVis parses all messages in the Trec07p corpus and applies a combination of regular expressions to extract the email attributes. After extracting all attributes for each individual email, it uses an external API to extract five derived attributes for each email including “city,” “country,” “language,” “category,” and “blacklist” attributes.

There are many text categorization algorithms that can automatically categorize text into multiple categories. Since this thesis is not about implementing a machine-learning algorithm, our processing component uses Alchemy [27], an online natural language processing service, to derive semantic attributes from the content and header parts of each individual message. Alchemy’s text categorization service uses multiple statistical and text categorization algorithms to classify text into twelve different categories: Arts & Entertainment, Business, Computer & Internet, Culture & Politics, Gaming, Health, Law & Crime, Religion, Recreation, Science & Technology, Sports, and Weather. The processing component of VeriVis uses the Alchemy text categorization API to classify each message into one or more of those categories.

Because the Trec07p email corpus is collected from an email server, it is possible to have messages received from multiple geographical locations and in multiple languages. The processing component of VeriVis uses the Alchemy’s language detection API, which is able to recognize more than 95 different languages, to derive the language attribute value for each message. It also uses IPinforDB², a popular IP locator database, to find cities and countries related to the IP addresses extracted from each message.

Another derived attribute considered by VeriVis is whether the IP addresses in the messages in the corpus are related to spammers’ IP addresses or not. The processing

²http://ipinfodb.com/ip_location_api.php - September, 20, 2013

component of VeriVis uses Spamhaus³, SpamCop⁴, and abuseat⁵ blacklist databases to check if an IP address in an individual message has been already reported.

4.2.2.1 Email Attributes in VeriVis

To support analytical reasoning processes, visualization must enable the analyst to focus on the relevant attributes of any given data set. To put it another way, to look at all of one's data instead of focusing in on specific dimensions can make it harder for users to discover important and unexpected information [26].

Because of multiple processing limitations, such as the limitation on the number of free API calls to the Alchemy Server, the processing component of VeriVis randomly subsets the whole Trec07p corpus to a smaller email set, in our case, sets of 33,171 spam and 4,325 ham messages. It then calculates the distribution probability of each attribute for both spam and ham messages in the corpus. For example, the distribution probability of each “From” attribute for spam messages is the number of spam messages in Trec07p that contain that “From” attribute, divided by the total number of spam messages in the entire corpus.

As illustrated in Figure 4.2, attributes such as “From,” “Date,” “To,” “Subject,” “Received path,” “Return path,” “Http,” and “Content type” are available in both spam and ham messages with the same distribution probability. Attributes such as “Font-color” and “X-mailer,” on the other hand, are less equally probable between spam and ham messages and are therefore good candidates for use in differentiating spam and ham messages from each other. Since we do not know that such an attribute is removed from a message because of user privacy issues or that a message did not have that attribute from the beginning, we select attributes with

³<http://www.spamhaus.org/> - September, 20, 2013

⁴<http://www.spamcop.net> - September, 20, 2013

⁵<http://cbl.abuseat.org> - September, 20, 2013

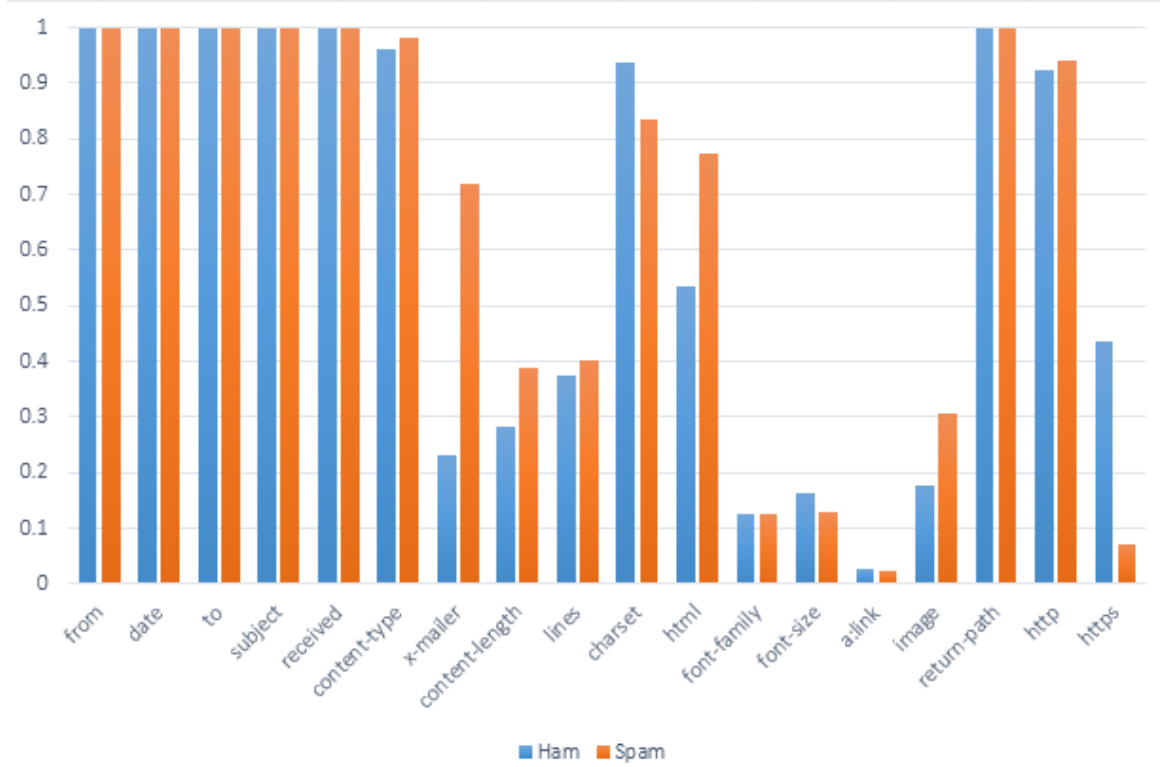


Figure 4.2: Distribution probability of email attributes in a subset of the Trec07p corpus.

higher and equal distribution probability for both spam and ham messages in the email corpus.

Figure 4.3 is an entity relationship diagram showing all attributes for both entities used in the visualization component in VeriVis. Each message has a unique identifier in the email table, which can be used as a foreign key to find locations involved in that message in the location table. For example, suppose that someone sends a message from the United States to Germany, and that the person in Germany forwards the message to someone else in France. In this case, the message contains three different IP addresses located in three different countries. The processing component of VeriVis separates data attributes according to this relationship diagram and stores them in two CSV (Comma-Separated Values) files (see Figure 4.4). These two CSV files become the source of input data records to the visualization component.

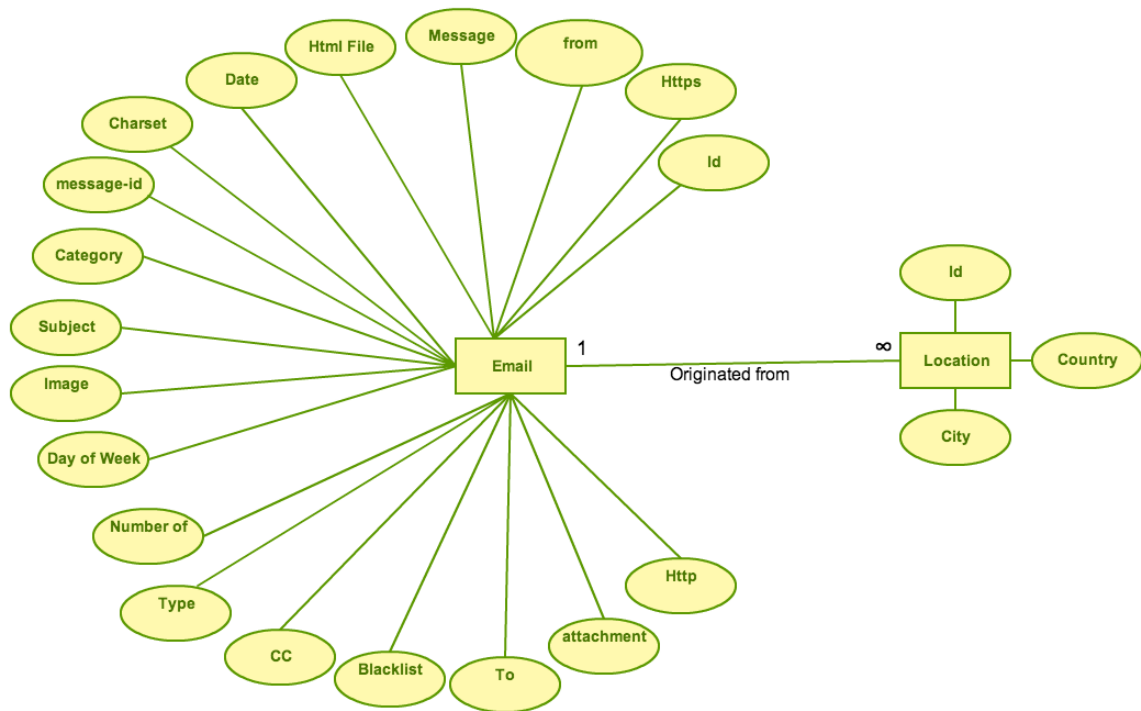


Figure 4.3: Input schema to the VeriVis visualization approach.

Email.csv							
ID	sender	receiver	message-id	subject	type	attach	
1	/Ames@ao	the00@sp	WYADCKPDF	Generic Cialis br	spam	FA	
2	@tractionm	the00@plg	20070408171	authentic viagra	spam	FA	
3	@mavoram	ktwarwic@	001301c77a1	problems in anim	spam	FA	
4	ald@funear	manager@	000001c77a0	We cure any dese	spam	FA	
5	a@netscap	gnitpick@	460301c77a0	Human Growth Hc	spam	FA	
6	@hlboulto	the00@plg	001601c77a1	accelerate it does	spam	FA	
7	id@dryden	the00@plg	001501c77a1	when taken for a l	spam	FA	
8	garagesen	the00@plg	000001c41d9	We cure any dese	ham	FA	
9	er@griffiel	the00@plg	000001c77a0	Need medicine?	spam	FA	
10	@expomec	the00@plg	000001c77a0	Need medicine?	ham	FA	
11	rius@public	gnitpick@	4d9c01c77a0	Become Fit For L	spam	FA	
12	rina@bb-a	theorize@	LVKTURBY.6	But serpentine	spam	FA	
13	@expomec	theorize@	000001c77a0	All products for yc	spam	FA	
14	icell@altavi	theorize@	78fc01c77a07	HGH really helps!	spam	FA	
15	ce@localbiz	ktwarwic@	LJFOTQWP.3	so provincial	spam	FA	
16	d@laneswz	producttes	20070408181	Are You Paying T	spam	FA	
17	abnamrofi	theorize@	000b01c77a0	chto sushchestvo	spam	FA	
18	ys@acces	producttes	20070408181	Davison can help	spam	FA	
19	una@graffit	the00@plg	e63001c77a0	Become Fit For L	spam	FA	
20	@sammim	gnitpick@	60e801c77a0	Effective Diet	spam	FA	

Location.csv		
Emial ID	City	Country
1	Seoul	Korea Republic of
1	unknown	Russian Federation
2	Oneonta	United States
2	Xian	China
3	Seoul	Korea Republic of
4	New York	United States
4	Geneva	United States
4	ountain Vie	United States
5	state Colleg	United States
6	arlos De Ba	Argentina
7	unknown	United States
8	Waterloo	Canada
8	Newberry	United States
9	Waterloo	Canada
9	Calgary	Canada
10	unknown	United States
11	Waterloo	Canada
11	Wiesbaden	Germany
12	Phoenix	United States

Figure 4.4: CSV files of email message and location records.

4.3 Visualization Component

The visualization component of VeriVis gets the processed data records as input from the processing component, visually encodes those data records into graphical attributes, and then displays them in multiple coordinated views. There are a wide variety of desktop and web-based visualization tools and toolkits for helping analysts perceive and make sense of complex data. Some of these, like D3 [5], Many Eyes [29], and others, allow users to apply visualization techniques to their data and make it available online, regardless of platform. For some of these visualization tools, a visualization designer needs to have knowledge of the syntax and library features to describe the desired data processing, view choices, and visual encodings. For example, if a designer wants to use D3 for visualization, they need to be familiar with JavaScript and the Document Object Model (DOM). Most of these visualization tools are limited in the composability of visualization techniques, which limits designers in their design decisions [4].

The visualization component of VeriVis is designed using *Improvise* [30], a desktop visualization software based on separating data and visualization models. This separation of tasks makes it a good choice for use with the two-part architecture of VeriVis. The advantages of Improvise for live designing multi-view visualizations including an integrated meta-visualization system [31, 32, 22], making it unique compared to other visualization tools. Improvise uses coordinated queries [33], a visual abstraction language based on the relational database model, to support live design of richly and highly interactive visualization tools having multiple coordinated views.

4.3.1 Views

VeriVis utilizes fourteen coordinated views to display the attributes of email messages contained in the subsetted Trec07p corpus.

4.3.1.1 Email Table Views

VeriVis contains two tabular views of spam messages: the junk box view, and the verified view. After identifying non-spam messages in the junk box view, a user can shift them to the verified view. The junk box view displays six attributes of all received spam messages, including “**sender email address**,” “**receiver domain address**,” “**subject**,” “**message**,” “**date**,” and “**time**”. To differentiate user-verified messages from other spam messages in the junk box view, VeriVis marks each cell in the sender column with a small circle. By default, all circles are black. When the user recovers a non-spam message in the junk box view, its circle turns to red. VeriVis uses a sky blue color to highlight selected messages in the junk box view. The user can select multiple spam messages by clicking rows in the junk box view, then use buttons to delete them or restore all removed spam messages back to the view (see Figure 4.5).

To check the application of VeriVis as a spam verification tool, we display all verified spam messages in the verified view (Figure 4.6). This view displays eight attributes of data including “**type**” of message, which is defined in our corpus either as spam or ham. As with the junk box view, the verified view uses a sky blue color to highlight user-selected messages among those verified as non-spam messages. As part of the spam verification process, the server administrator can, after the identification stage, recover messages or move them to another location for later users’ interventions and their own ad hoc verification.

The junk box view and verified view are connected to each other based on users’ interactions. Users can select spam messages individually or as a group from the junk box view and recover them to the verified view (by pressing the “A” key). They can also reverse the verification process by first selecting verified messages in the verified view individually or as a group, then removing them from the list of verified messages

Junk Box

Sender	Receiver	Subject	Message	Date	Time	
accsupport-4758111...	speedy.uwaterloo.ca	security maintenance.	-----	Fri Apr 20 00:00:00...	3:35	
manager7181391103...	speedy.uwaterloo.ca	BB&T - security maintenance	-----	Fri Apr 20 00:00:00...	3:35	
manchester2016.com...	plg.uwaterloo.ca	Avoid enhancement pills	Hi In 196...	Fri Apr 20 00:00:00...	3:52	
stephen@garageservic...	plg.uwaterloo.ca	She wants a better sex? All you need's here!	-----	Fri Apr 20 00:00:00...	3:52	
sScottish@veryspeedy...	plg.uwaterloo.ca	Fwd: Pharmacy bulletin	Dear value...	Fri Apr 20 00:00:00...	4:24	
krissysjbun@hkbmwgr...	speedy.uwaterloo.ca	I'm sorry to hear that	It struct...	Fri Apr 20 00:00:00...	4:28	
qodrake@wlf.de	speedy.uwaterloo.ca	I am 79 years young!	A few = w...	Fri Apr 20 00:00:00...	4:35	
CholesterolWatchers@j...	speedy.uwaterloo.ca	Cholesterol patients - what Dr. didn't tell you	Don't Fi...	Fri Apr 20 00:00:00...	4:35	
dwcomputerexchange...	speedy.uwaterloo.ca	re: Can you imagine that you are healthy	Dear custo...	Fri Apr 20 00:00:00...	4:46	
qbrde.2lp@greatbutte...	speedy.uwaterloo.ca	New Update to fix Windows File Errors in registry	Warning=2...	Fri Apr 20 00:00:00...	4:47	
shornless@123mail.org	speedy.uwaterloo.ca	Fwd: Pharmacy bulletin	Dear value...	Fri Apr 20 00:00:00...	4:53	
dwaexm@maex.net	speedy.uwaterloo.ca	Healthy life is your dream?	Dear custo...	Fri Apr 20 00:00:00...	4:53	
bujicichudi@aplfab.c...	speedy.uwaterloo.ca	Sorry i did forgot	*keep it a...	Fri Apr 20 00:00:00...	4:58	
parkvipcrckrp@ocn.ne...	plg2.math.uwaterloo.ca	Morning	By rose...	Fri Apr 20 00:00:00...	5:03	
gene74@mackone.fre...	speedy.uwaterloo.ca	Slots Roulette & Blackjack	Online Ga...	Fri Apr 20 00:00:00...	5:09	
4remarkable.com@lov...	speedy.uwaterloo.ca	Software At Low Price	Hi In 1963...	Fri Apr 20 00:00:00...	5:13	
plyjnh@terra.es	speedy.uwaterloo.ca	Critpick_h_k_0_0_0_d...	TH...	Fri Apr 20 00:00:00...	5:25	
mailings@mail.cnn.com	SPEEDY.UWATERLOO.CA	Preacher's wife found guilty in husband's death		Fri Apr 20 00:00:00...	5:37	
garfield@brainpod.com	speedy.uwaterloo.ca	Slots Roulette & Blackjack	Online Ga...	Fri Apr 20 00:00:00...	5:37	
FROM USER@snrav se...	in.uwaterloo.ca	s_00@020_05_0...	071 In KO...	Fri Apr 20 00:00:00...	5:41	

Buttons: Delete, Restore

Figure 4.5: The junk box view, showing all received spam messages in the email server.

Verified Emails

Type	From	To	Subject	Id	Language	Category	Date
spam	gilbert@pellicano.biz	the00@plg.uwaterloo.ca	All love enhancers on one portal!	11853	english	unknown	4/19/01
spam	theproductrocket.com@famil...	smiles@speedy.uwaterloo.ca	Beware of fake pills	12836	english	business	4/19/01
spam	dwconcordtkdm@concordtk...	ktwarwic@speedy.uwaterloo.ca	re: Need some help	11850	english	health	4/19/01
spam	dwlancafan@allanfan.c...	ktwarwic@speedy.uwaterloo.ca	re: Can't be a lover anymore?	10477	english	health	4/20/01
spam	Trotter774@gmail.com	josh@speedy.uwaterloo.ca	Wed like to offer you the vacant posit...	10475	english	unknown	4/20/01
spam	pcollins@hamptonroads.com	manager@speedy.uwaterloo.ca	VIAGRA	2175	english	unknown	4/20/01
spam	Drew2399@verizon.com	josh@speedy.uwaterloo.ca	Notification about the vacancy [letter...	10476	english	unknown	4/20/01
spam	euphemismsalphnumeric@...	catchall@speedy.uwaterloo.ca	Fwd: Pharmacy bulletin	2174	english	health	4/20/01
spam	accsupport-475811174ib@...	manager@speedy.uwaterloo.ca	security maintenance.	2172	english	unknown	4/20/01
ham	speedy.uwaterloo.ca@cbsig...	ktwarwic@speedy.uwaterloo.ca	News Summary	13308	english	unknown	4/26/01
ham	speedy.uwaterloo.ca@cbsig...	ktwarwic@speedy.uwaterloo.ca	News Summary	2653	english	unknown	4/27/01
ham	speedy.uwaterloo.ca@cbsig...	ktwarwic@speedy.uwaterloo.ca	News Summary	10775	english	unknown	4/27/01
ham	speedy.uwaterloo.ca@cbsig...	ktwarwic@speedy.uwaterloo.ca	CBS News Sunday Morning: Royal attr...	2488	english	unknown	4/27/01
ham	speedy.uwaterloo.ca@cbsig...	ktwarwic@speedy.uwaterloo.ca	60 Minutes E-mail Alert	16064	english	unknown	4/27/01
ham	speedy.uwaterloo.ca@cbsig...	ktwarwic@speedy.uwaterloo.ca	News Summary	8846	english	unknown	4/27/01
ham	speedy.uwaterloo.ca@cbsig...	ktwarwic@speedy.uwaterloo.ca	News Summary	6007	english	unknown	4/28/01
ham	speedy.uwaterloo.ca@cbsig...	ktwarwic@speedy.uwaterloo.ca	News Summary	9684	english	unknown	4/29/01
ham	speedy.uwaterloo.ca@cbsig...	ktwarwic@speedy.uwaterloo.ca	News Summary	1601	english	unknown	4/30/01
ham	speedy.uwaterloo.ca@cbsig...	ktwarwic@speedy.uwaterloo.ca	News Summary	7311	english	unknown	4/26/01
ham	speedy.uwaterloo.ca@cbsig...	ktwarwic@speedy.uwaterloo.ca	News Summary	10525	english	unknown	4/26/01
ham	speedy.uwaterloo.ca@cbsig...	ktwarwic@speedy.uwaterloo.ca	News Summary	5346	english	unknown	4/25/01
ham	speedy.uwaterloo.ca@cbsig...	ktwarwic@speedy.uwaterloo.ca	News Summary	15770	english	unknown	4/25/01
ham	speedy.uwaterloo.ca@cbsig...	ktwarwic@speedy.uwaterloo.ca	News Summary	10279	english	unknown	4/25/01
ham	speedy.uwaterloo.ca@cbsig...	ktwarwic@speedy.uwaterloo.ca	News Summary	831	english	unknown	4/24/01
ham	speedy.uwaterloo.ca@cbsig...	ktwarwic@speedy.uwaterloo.ca	News Summary	2089	english	unknown	4/24/01
ham	speedy.uwaterloo.ca@cbsig...	ktwarwic@speedy.uwaterloo.ca	News Summary	8919	english	unknown	4/23/01
ham	speedy.uwaterloo.ca@cbsig...	ktwarwic@speedy.uwaterloo.ca	News Summary	8078	english	unknown	4/23/01

Figure 4.6: The verified view, showing all user-verified spam messages.

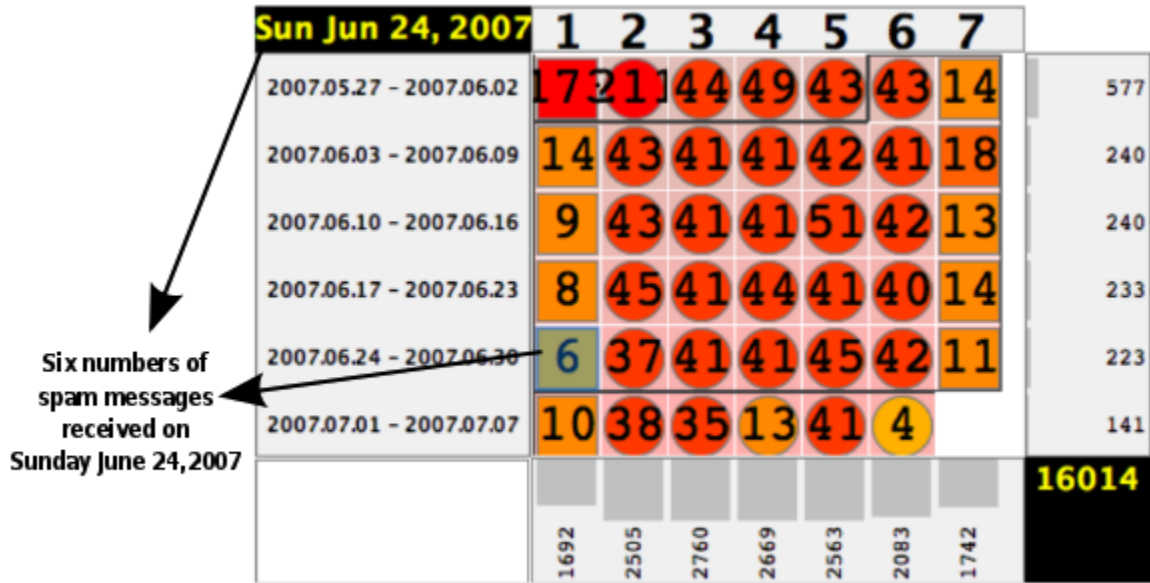


Figure 4.7: The calendar view, showing received spam messages over time.

(by pressing the “D” key). VeriVis also allows users to sort (and sub-sort) data items in multiple columns in tabular views, either in increasing or decreasing order.

4.3.1.2 Calendar View

The calendar view [34] in VeriVis allows server administrators to view messages over time and to investigate messages on specific dates. The calendar view illustrates each day’s cell using a colored circle (see Figure 4.7). It uses rectangles instead of circles to differentiate weekends from business days. The calendar view highlights each cell with a different color, in a scale from yellow to red, according to the number of spam messages received on that day. This makes it possible for server administrators to compare the number of messages received on different days. Each cell also has a label that represents the number of spam messages received on that day.

The bar chart at the bottom of the calendar view allows users to compare the total number of received spam messages on each day of a week. The bar chart on the right side of the calendar view allows users to compare the total number of received spam messages for each week of any given month. For example, in Figure 4.8, the total

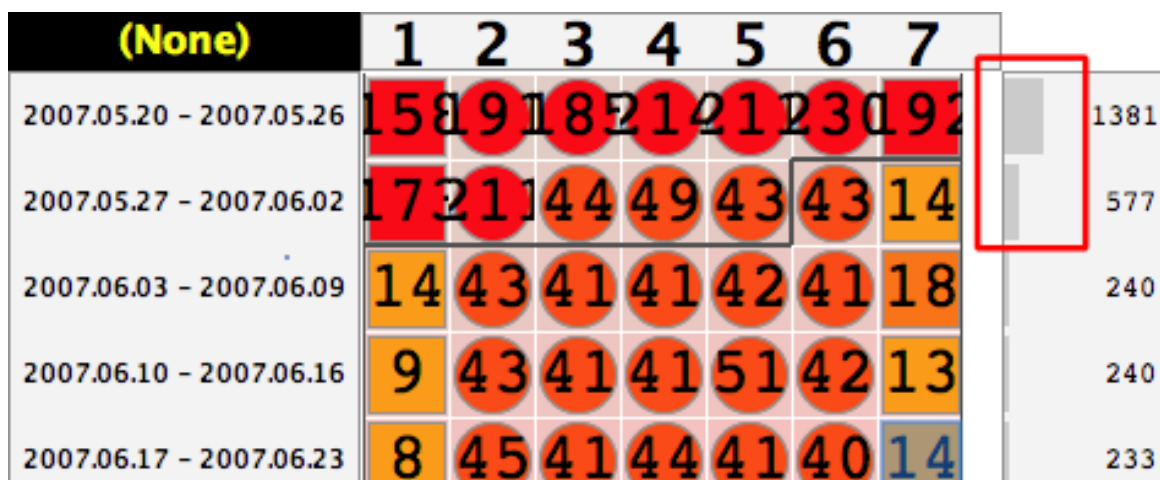


Figure 4.8: The calendar view, showing total counts for each week.

number of spam messages received in the last two weeks of May 2007 is more than any week in June 2007. Figure 4.9 reveals that the total number of spam messages received by the email server on weekends is less than on weekdays.

4.3.1.3 Scatter Plot of Size vs. Number of Lines

VeriVis can display spam messages based on their “size” and “number of lines” in a two-dimensional scatter plot (see Figure 4.10). The horizontal plot dimension encodes the “size” attribute of the spam message. The horizontal range is determined by the largest and smallest values of the “size” attribute. The vertical plot dimension encodes the “number of lines” attribute. The maximum and minimum values of “size” and the “number of lines” attributes vary over time as new messages are received.

It is common to have multiple spam messages with the same (or nearly the same) size and number of lines. In this case, the circles that encode those messages will overlap. VeriVis uses translucency to make it easier for users to observe and identify overlaps in the scatter plot. For example, a lighter circle indicates the presence of a large number of overlaid spam messages with the same size and number of lines; a

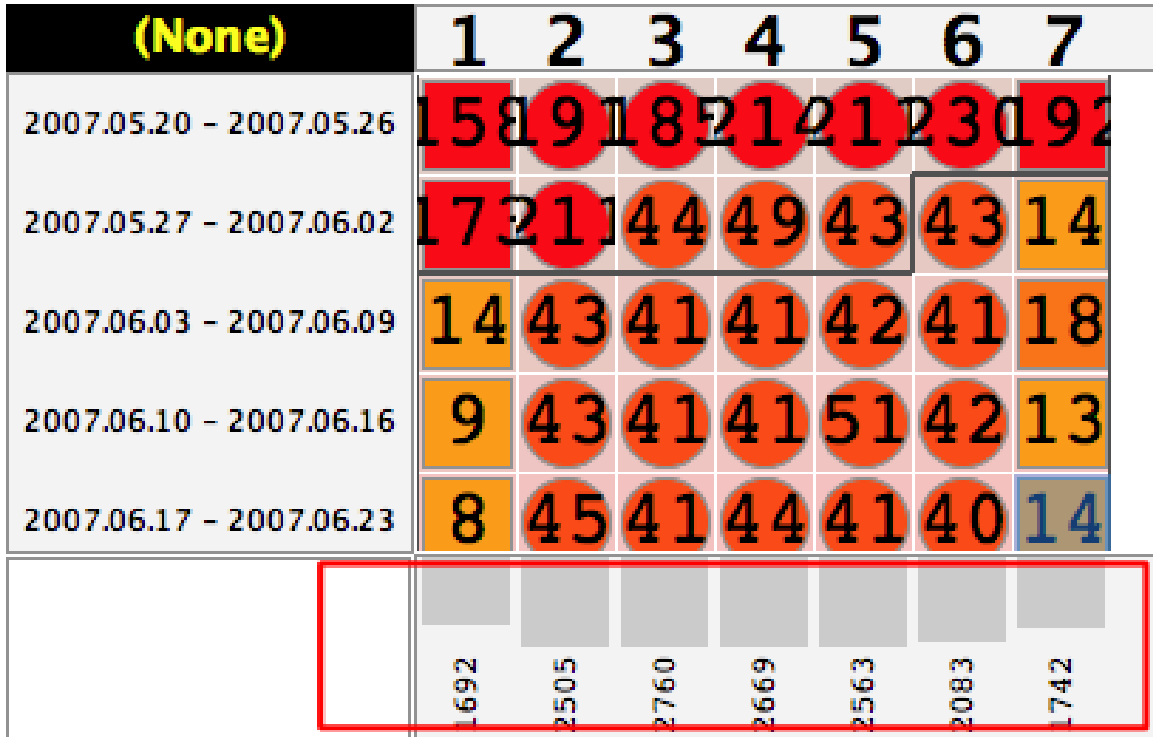


Figure 4.9: The calendar view, showing total message counts for each day of the week.

dark grayish circle indicates a low number of overlapped spam messages with that specific size and number of lines (Figure 4.10a and 4.10b, respectively).

Users can zoom in and out in the scatter plot using the keyboard, mouse, or scroll wheel, thus interactively manipulating the size and number of lines dimensions to investigate trends. Users can also select arbitrary subsets of plotted messages by clicking circles. By default, messages are displayed using white-edged circles in the scatter plot. Circles with blue edges (Figure 4.10c) distinguish selected messages in the scatter plot from unselected ones.

4.3.1.4 The Senders-Receivers Bipartite View

The bipartite view in VeriVis shows email communications between different senders and receivers (see Figure 4.11). In this view, nodes on the left side of the view are senders' email addresses and the nodes on the right side are receivers' domain

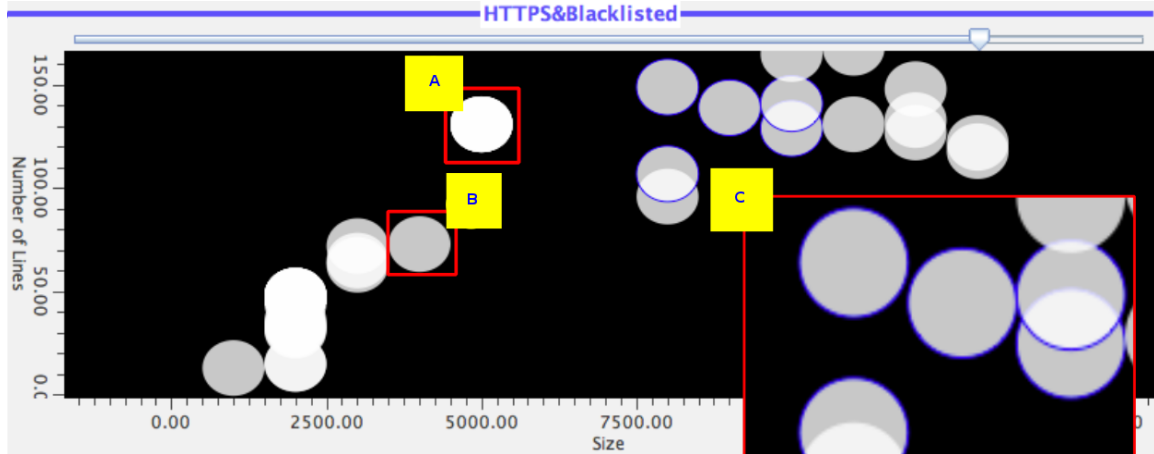


Figure 4.10: Scatter plot of message size vs. number of lines: (a) Lighter circles show that a small number of messages have the same size and number of lines; (b) Darker circles show that a large number of messages have the same size and number of lines; (c) User selected spam messages are illustrated with blue (dark) edges.

addresses. The line (or lines) connecting a sender to a receiver represents a message which is sent from that sender to that receiver.

VeriVis provides a customization panel for the bipartite view. Using the customization panel, users can select other email attributes for coloring the labels on either side of the bipartite view. It also allows users to observe patterns between senders and receivers at the domain and top-level domain levels by checking the related checkboxes on each side. To limit exposure of individually identifiable information, by default the right side of the bipartite view displays receivers' domain level addresses instead of their full email addresses. For example, if users check the “Receiver Top Domain address” checkbox on the right side and select “Category” in the combo box (on the left side), they can observe the exact same data as figure 4.11, but with the connections representing the relationship between receivers' top domain level and the senders' email categories (Figure 4.12).

It is possible to have multiple messages between one sender and one receiver. The lines connecting them will overlap. The slider in the bipartite view allows users to

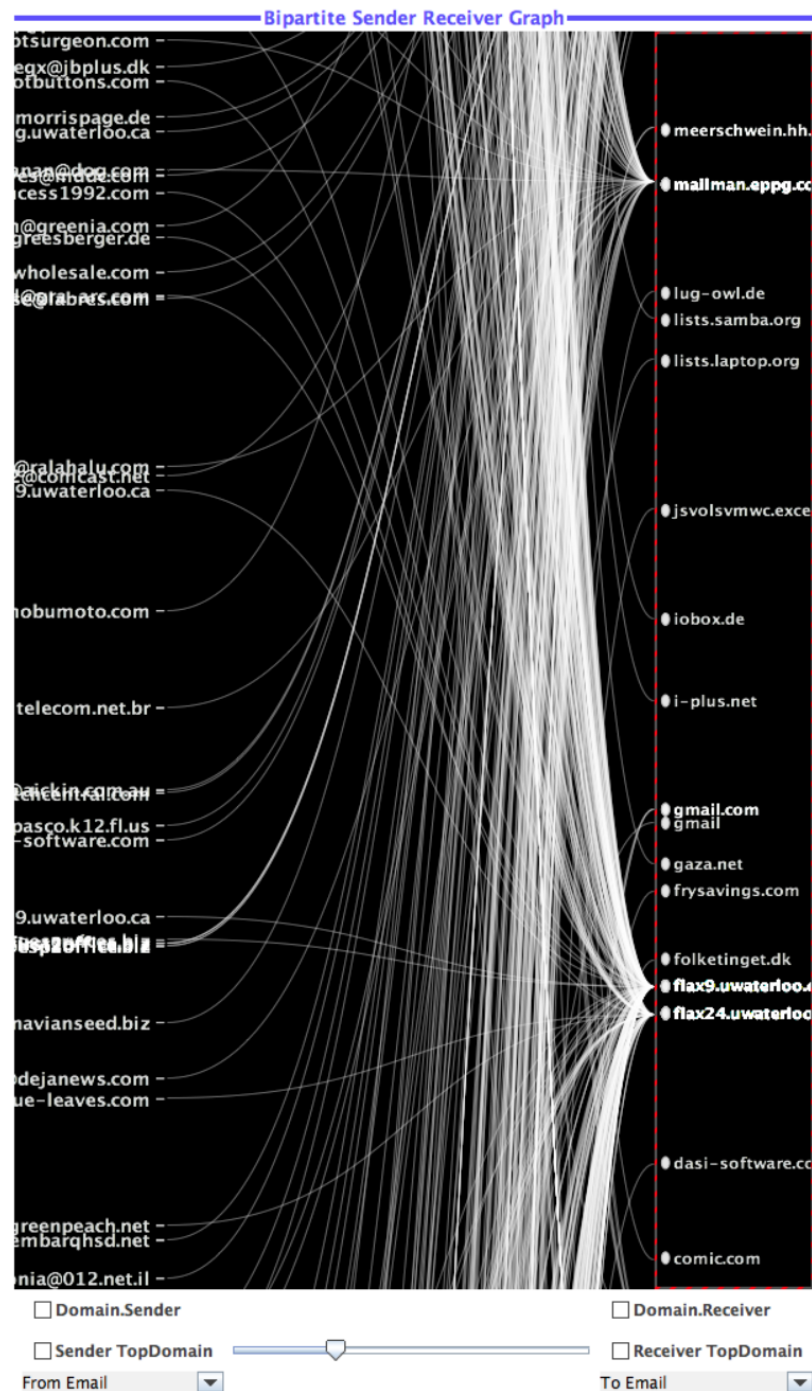


Figure 4.11: The bipartite view, labels on the left are senders' email addresses; labels on the right are receivers' email addresses.

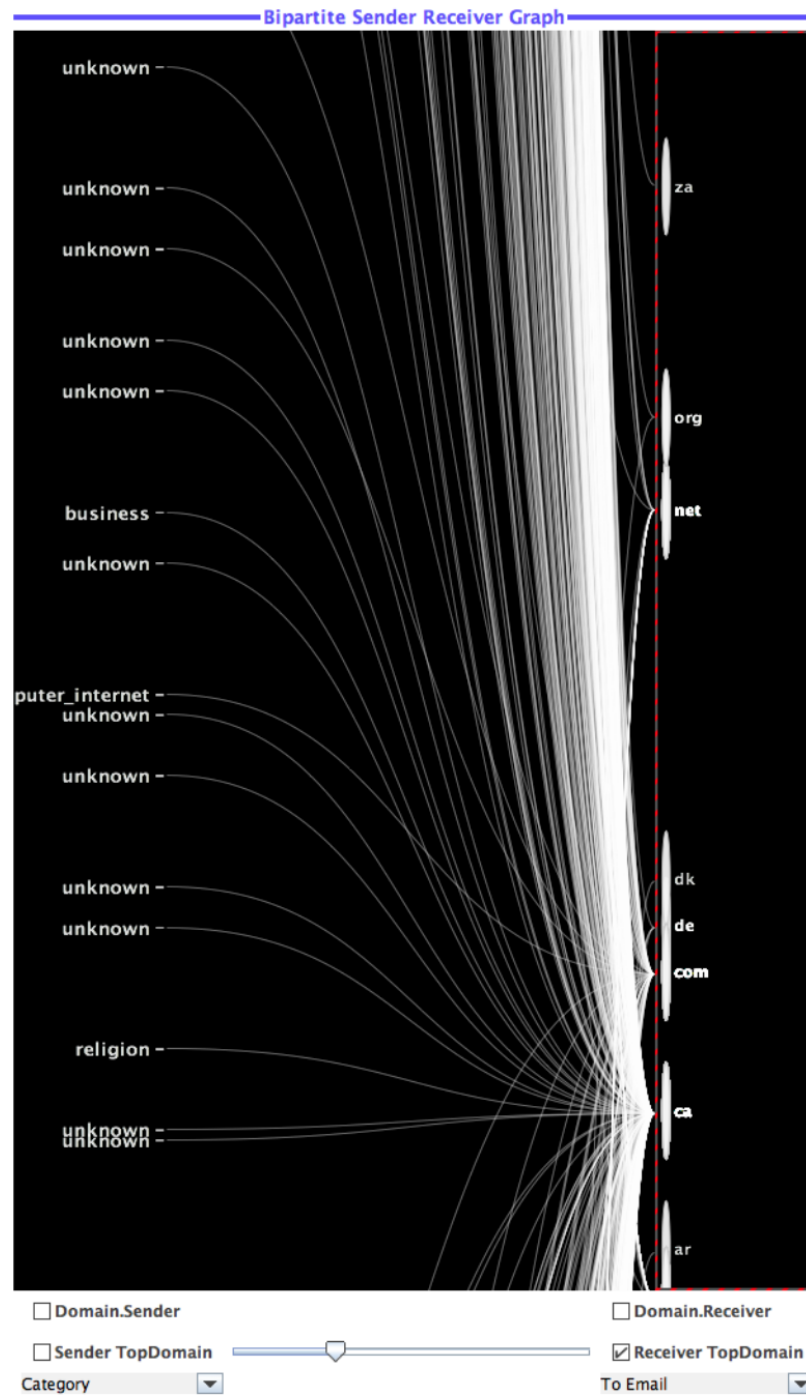


Figure 4.12: The bipartite view, here showing messages from sending top-level domains to receiving top-level domains. At bottom are controls to aggregate senders and receivers by domain (checkboxes), control line translucency (slider), and choose which attribute to color senders and receivers on (combo boxes).

observe differences between the overlapping lines, as well as other communication connections in the bipartite view, by changing their opacity. (The slider in the bipartite view changes the alpha channel of connector colors.) For example, Figure 4.13 shows the relationships between senders' top-level domains and receivers' top-level domains for 451 messages received in French. In this view, the high color opacity makes it hard to identify those top domain levels that were more involved for supposed spam communications in the French language. Figure 4.14 displays the exact same data records as in Figure 4.13, but in this screenshot the user has decreased the color opacity for connectors' colors using the opacity slider, thereby showing that the top-level domains of ".ca," ".com," and ".info" were common in spam communications in the French language compared to other top-level domains.

4.3.2 Filtering

Filtering is one of the most important concepts in visualization, and a key part of Shneiderman's mantra [24]. VeriVis allows users to filter their spam folders based on a combination of message attributes such as "sender email address," "receiver email address," "category," "language," "source country," etc.

VeriVis provides a filtering control panel (Figure 4.15) in which users can select different attributes on which to filter spam messages. Users can check or uncheck boxes of each attribute and then immediately observe the updated, filtered results in the various visualization views. Some email attributes are of Boolean type; for example, "attachment" is a Boolean-typed attribute. True and false values of the "attachment" attribute indicate whether or not a given message has an attachment.

For each Boolean attribute, users can filter messages to show only "true" or "false" cases. Figure 4.16 shows the panel used to select values of Boolean attributes for filtering. Suppose the user wants to filter all spam messages based on the "attachment"

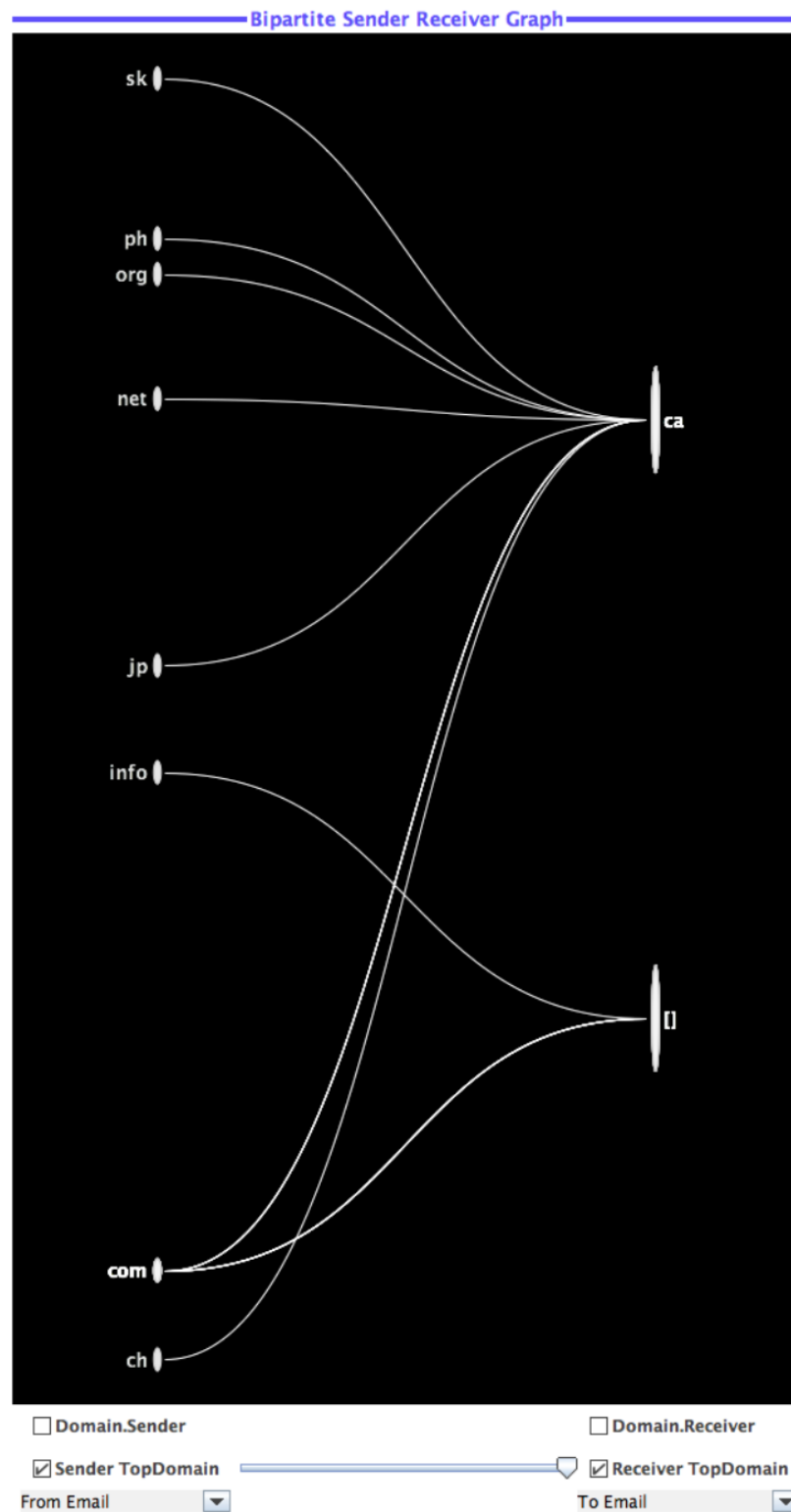


Figure 4.13: The bipartite view, showing messages between senders' top-level domains and receivers' top-level domains for messages in French.

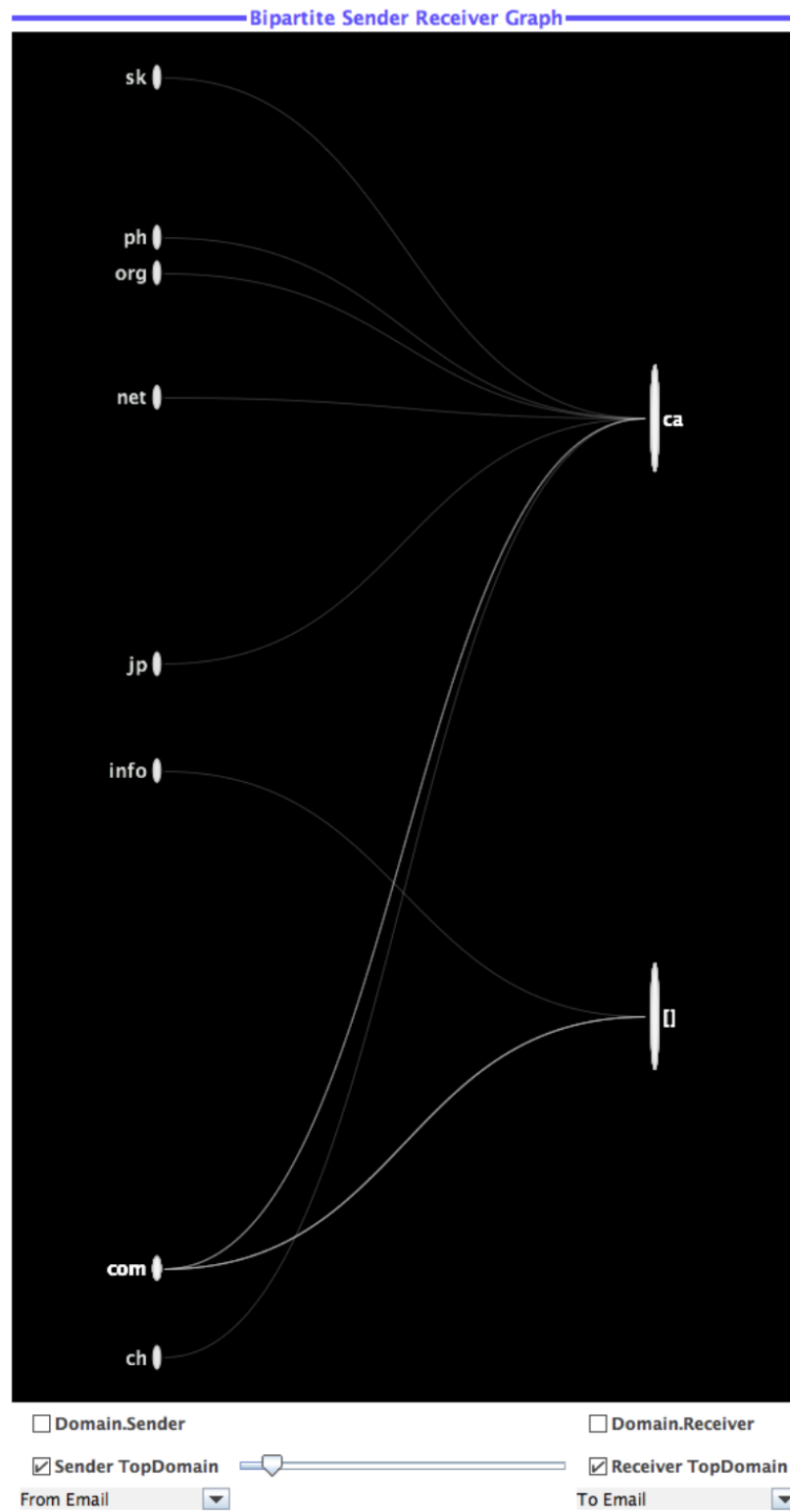


Figure 4.14: The bipartite view, revealing those top-level domains that were more involved in supposed spam messages in French than those in other languages.

Filter

- ☐ Category
- ☐ Country&City
- ☐ Size
- ☐ Language
- ☐ Date
- ☒ Attachment
- ☐ Black List Filter
- ☐ HTTP Filter
- ☐ HTTPS Filter
- ☐ Image Filter
- ☐ Day of Week
- ☐ Subject
- ☐ CC
- ☐ Sender
- ☐ Receiver

Figure 4.15: The filtering control panel, here with filtering on attachments only.

attribute. VeriVis allows users to filter either messages that have attachments, or messages without attachments.

VeriVis provides multiple table views to show the values of non-Boolean attributes such as “language,” “category,” and “size” (see Figure 4.17). The first column shows attribute values. The second column shows how many messages have each value. For example, the category table view in Figure 4.17 has two columns. The first column displays all categories of messages derived using the Alchemy API (by the processing component). The second column represents the number of messages for each category. Similarly, the language table view displays multiple languages in its first column and the number of messages in each language in its second column. All occurring data values of each attribute are extracted in the preprocessing phase from all messages in the corpus. As an exception, the token words table view is a list of words that appear in the subject line of messages. The token list is extracted by tokenizing the subject line of messages in the corpus (i.e., currently available in email server). Since this table view can contain a long list of tokens, VeriVis provides

Filter Options

- ☒ With Attachment
- ☐ Without Attachment
- ☐ Blacklisted
- ☐ Not Blacklisted
- ☐ With HTTP
- ☐ Without HTTP
- ☐ With HTTPS
- ☐ Without HTTPS
- ☐ With IMAGE
- ☐ Without IMAGE

Figure 4.16: The filtering control panel for Boolean attributes, with filtering on for messages that have attachments.

a text field for users to search for a specific word or group of words. The token words table view filters all messages that contain any of those words in their subject line. VeriVis also allows users to sort these table views on attribute value (first column) or message count (second column) in increasing or decreasing order.

4.3.3 Highlighting

VeriVis assigns different colors to selected data items in each attribute table view. A control panel (Figure 4.18) allows users to select one attribute to highlight on. Highlighting of data points also happens in other views, such as scatter plots and the bipartite view. For example, in Figure 4.19 the bipartite view highlights all messages that are in the “sports” category as a function of the language attribute. Spanish, German, English, French, and “Unknown” are selected items in the language table view; highlighting colors are automatically assigned to each of these selected languages. The view reveals that most of the received messages in the sports category are in English (highlighted in red).

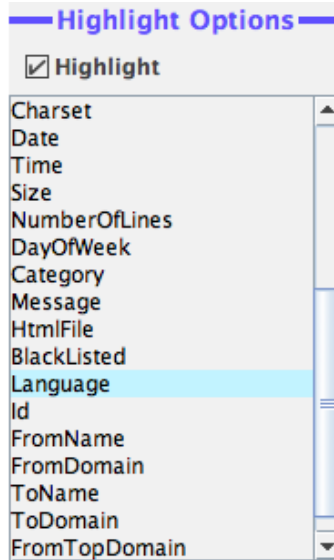


Figure 4.18: The highlighting control panel, with highlighting on for the “Language” attribute.

In designing VeriVis, we tried to select and arrange views such that, regardless of data types and format of the original data, the user can understand the purpose of each view and what information it represents. For instance, following Gestalt Principles of Design, we used similarity of views and visual encodings to help users analyze multidimensional attribute relationships and consequently perform effective spam verification. We also considered the logical order of interactions, such as in the checkboxes in the filtering control panel (Figure 4.15) and their related views, to determine view and control positions. Attention to cognitive and perceptual factors like these can help to increase the effectiveness of a visualization [2]. Our assessment of VeriVis is, however, based on utility rather than usability.

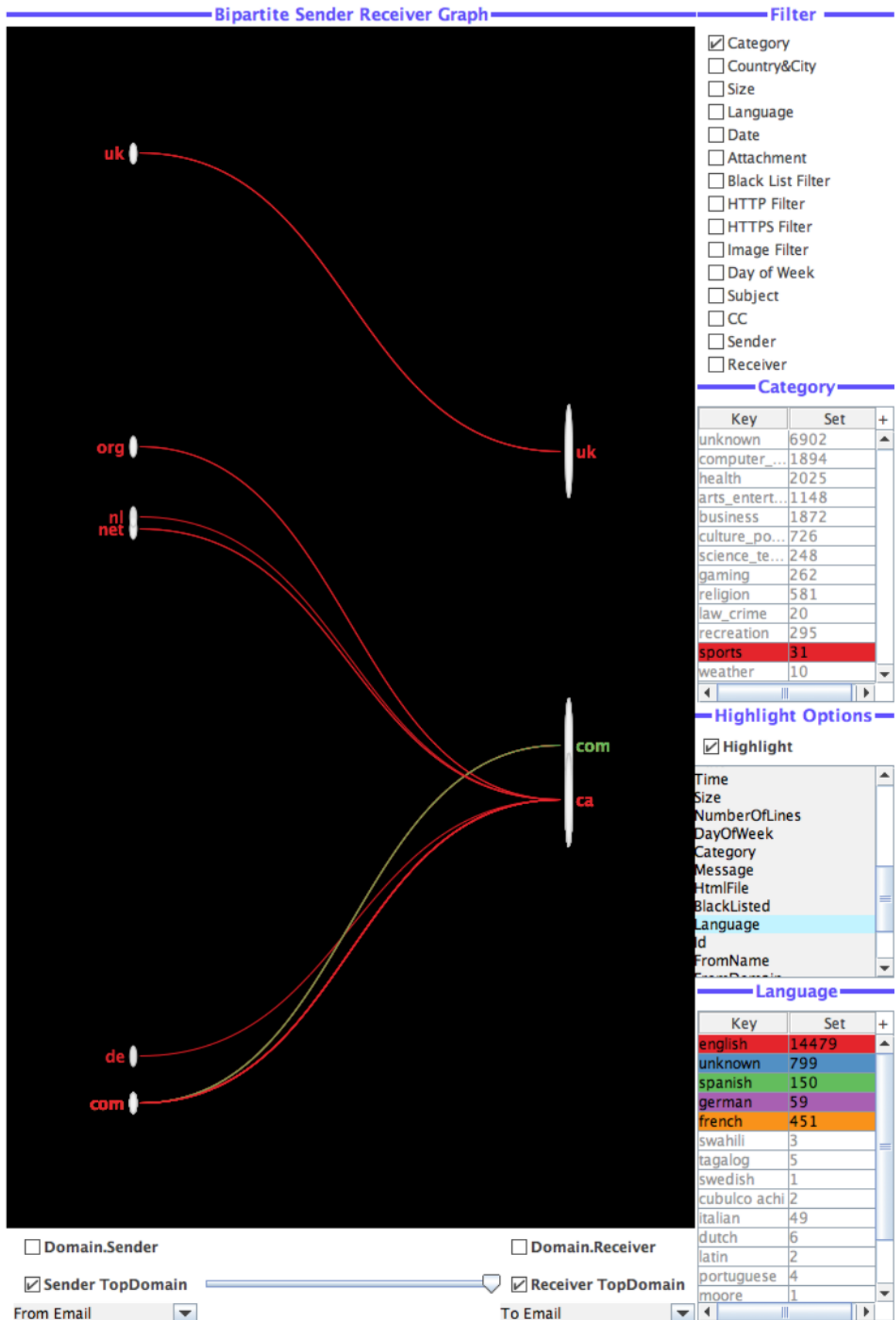


Figure 4.19: The bipartite view, with highlighting and filtering of messages in the sports category.

Chapter 5

Application

VeriVis is designed to help server administrators perform effective spam verification in two different situations. There are many activities that can be performed using a server-level spam verification tool like VeriVis. Because of individual differences and needs of people, as well as ongoing debate regarding the definition of spam, however, it is impractical to check the effectiveness of VeriVis for all reasonable, expected activities. We limit our study of the effectiveness of VeriVis as a spam verification tool to two types of activities: “non-spam identification” and “spam exploration”. We examine several examples of each type. These activities are hypothetical as constrained by the structure and contents of the Trec07p email corpus. We assess how well VeriVis supports users in performing the actions required to accomplish each of these activities.

5.1 Non-spam Identification: Sample Activities

For non-spam identification purposes, users typically have a specification about non-spam messages that are possibly misclassified as spam. Users apply their knowledge about non-spam messages to identify them among the messages classified as spam in an email server. This section considers two non-spam identification sample activities and checks if VeriVis allows users to perform the actions required for each of them.

5.1.1 First Sample Activity

Consider the following scenario. A server administrator at one of the departments of the University of Waterloo receives the following report from a number of staff members on Sunday, May 06, 2007: “We used to receive daily news from CNN in our inbox. For the last two weeks we have not received any email from CNN. We have also checked our junk box folder and looked for any misclassified messages from CNN, but we could not find anything useful.”

The server administrator knows that spam filters in their department’s email server identify spam messages and stop them in the email server. He also knows that sometimes their spam filters quarantine a group of the received messages that are possibly spam, send them to end users’ junk box folders, and ask them to perform ad hoc spam verification.

Based on the report, the server administrator knows that end users have already checked their junk box folders and they did not find any messages from CNN. Therefore, the server administrator wants to check and verify if any messages with those specifications have been detected as spam and stopped in the email server. The following is a list of actions users need to take to accomplish this specific activity, based on their knowledge of non-spam messages:

- Filter all spam messages in the email server based on dates of receipt.
- Select any spam messages received in the past two weeks.
- For the filtered data records, mark those messages sent from CNN domains.
- Recover marked spam messages to users’ inbox folders.
- Delete unmarked spam messages from the email server.

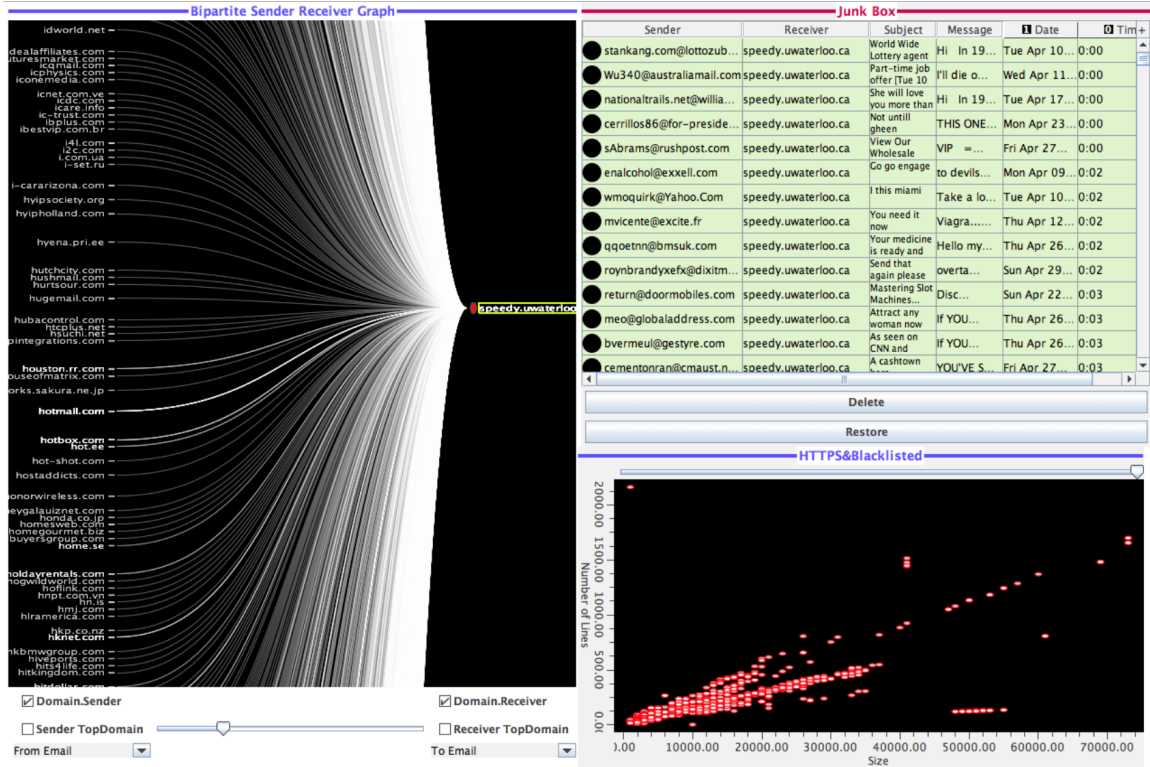


Figure 5.1: The non-spam identification first sample activity: starting state of the bipartite, scatter plot, and junk box views.

We know that any message meeting the above criteria is classified as ham in our email corpus. Having this knowledge, we check how VeriVis can help users perform all required actions for this non-spam identification sample activity.

Figure 5.1 shows the initial visualization state of the bipartite, scatter plot, and junk box views in VeriVis before the user starts looking for non-spam messages from CNN. Figure 5.1 shows all messages that are classified as spam in the “speedy.uwaterloo.ca” email server.

The server administrator receives the report on May 06, 2007. Because the exact time slice is one of the known specifications of this message, the user can select all of the days during the past two weeks in the calendar view (see Figure 5.2). The user also knows that these lost messages are daily news from CNN. Therefore they can search for any message that contains “CNN” in its subject line (see Figure 5.3). At this point the user can filter all received messages in the email server based on

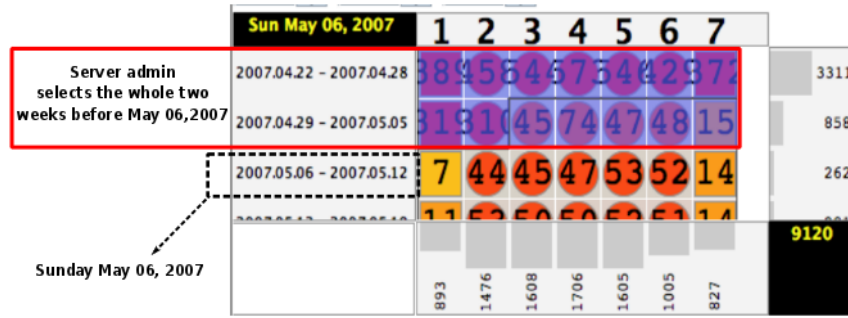


Figure 5.2: The non-spam identification first sample activity: selecting two weeks prior to the receipt date in the calendar view.

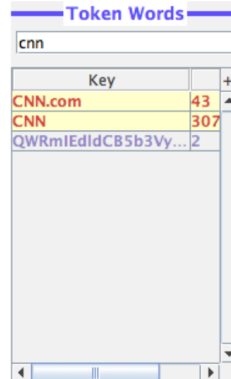


Figure 5.3: The non-spam identification first sample activity: searching for and selecting CNN-related subject line tokens in the token words table view.

both selected dates (in the calendar view) and selected tokens related to “CNN” (in the token words table view) (see Figure 5.4). After applying these filters on received messages in the email server, the user can immediately observe a smaller number of messages in other views (see Figure 5.5).

Now, using the bipartite view, the user can select any domain related to CNN. Of two senders’ domains in the bipartite view, “mail.cnn.com” is the only domain apparently related to CNN (see Figure 5.6). As soon as the user selects this sender domain address in the bipartite view, any received message from that sender domain address is highlighted in green in the junk box view (see Figure 5.7).

Filter

- ☐ Category
- ☐ Country&City
- ☐ Size
- ☐ Language
- ☒ Date
- ☐ Attachment
- ☐ Black List Filter
- ☐ HTTP Filter
- ☐ HTTPS Filter
- ☐ Image Filter
- ☐ Day of Week
- ☒ Subject
- ☐ CC
- ☐ Sender
- ☐ Receiver

Figure 5.4: The non-spam identification first sample activity: choosing to filter on Date in the filtering control panel.

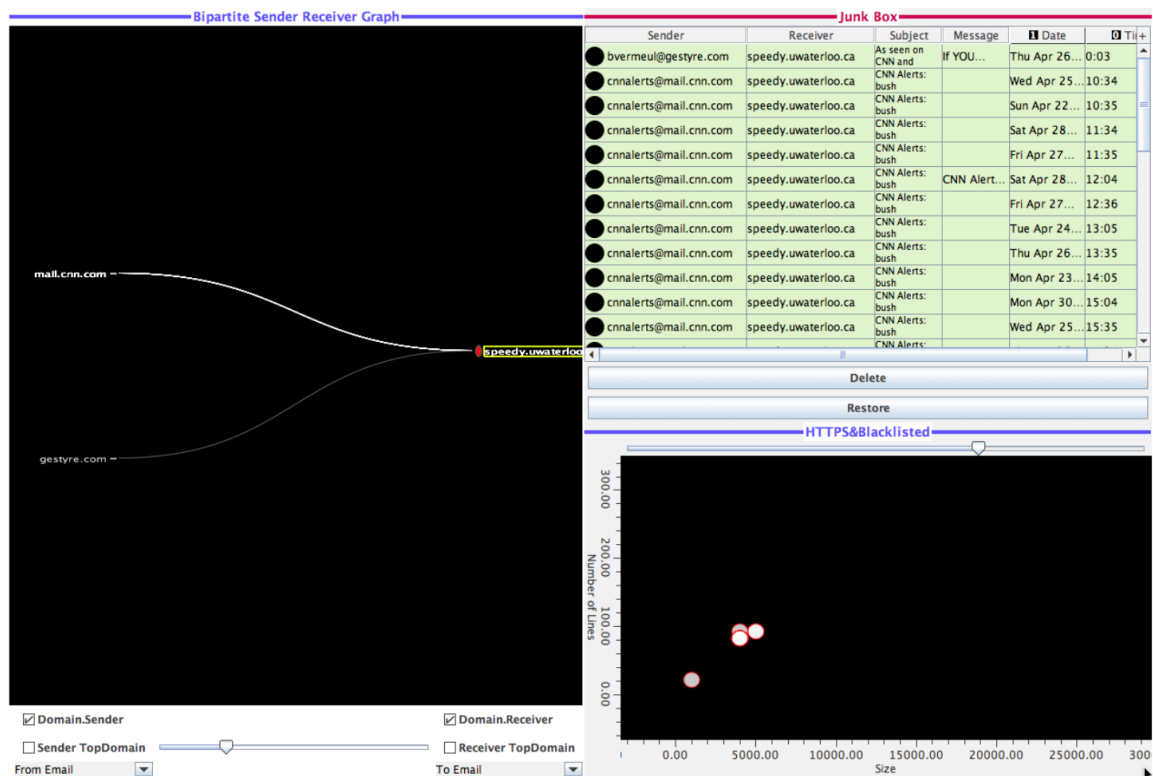


Figure 5.5: The non-spam identification first sample activity: viewing filtered spam messages in the bipartite and scatterplot views.

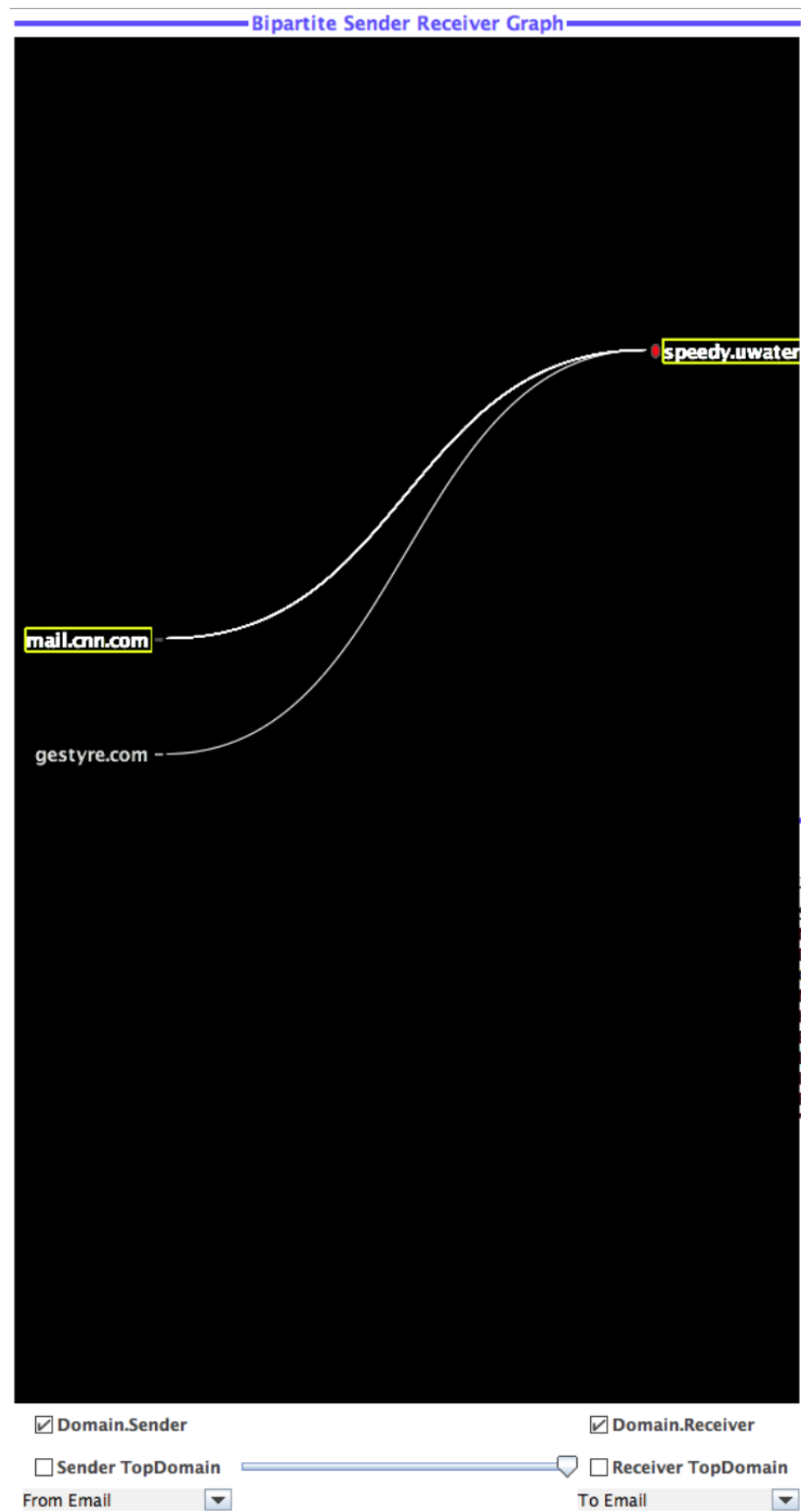


Figure 5.6: The non-spam identification first sample activity: selecting “mail.cnn.com” as a sender domain address in the bipartite view.

Junk Box						
Sender	Receiver	Subject	Message	Date	Time	^
● cnnalerts@mail.cnn.com	speedy.uwaterloo.ca	CNN Alerts: bush		Tue Apr 24 00:00:0...	13:05	▲
● cnnalerts@mail.cnn.com	speedy.uwaterloo.ca	CNN Alerts: bush	CNN Alerts:...	Tue Apr 24 00:00:0...	21:34	
● cnnalerts@mail.cnn.com	speedy.uwaterloo.ca	CNN Alerts: bush		Wed Apr 25 00:00:0...	9:34	
● cnnalerts@mail.cnn.com	speedy.uwaterloo.ca	CNN Alerts: bush		Wed Apr 25 00:00:0...	10:34	
● cnnalerts@mail.cnn.com	speedy.uwaterloo.ca	CNN Alerts: bush		Wed Apr 25 00:00:0...	15:35	
● cnnalerts@mail.cnn.com	speedy.uwaterloo.ca	CNN Alerts: bush		Wed Apr 25 00:00:0...	17:35	
● cnnalerts@mail.cnn.com	speedy.uwaterloo.ca	CNN Alerts: bush	CNN Alerts:...	Wed Apr 25 00:00:0...	21:34	
● cnnalerts@mail.cnn.com	speedy.uwaterloo.ca	CNN Alerts: bush etc.		Wed Apr 25 00:00:0...	22:35	
● cnnalerts@mail.cnn.com	speedy.uwaterloo.ca	CNN Alerts: bush		Wed Apr 25 00:00:0...	23:34	
● bvermeul@gestyre.com	speedy.uwaterloo.ca	As seen on CNN and ABCnews!	If YOU w...	Thu Apr 26 00:00:0...	0:03	▼
Delete						
Restore						

Figure 5.7: The non-spam identification first sample activity: viewing highlighted non-spam messages from “mail.cnn.com” in the junk box view.

The user can then mark all highlighted data records in the junk box view and recover them into users’ inboxes. VeriVis does not provide an inbox for end users. Instead, the server administrator can recover all marked spam messages to the verified view by pressing the “A” key on the keyboard (see Figure 5.8).

The “type” column in the verified view (leftmost column in Figure 5.8, bottom) shows that by using VeriVis, the user correctly identified misclassified messages from “mail.cnn.com”. Finally, by using the bipartite view, the user can select the other sender domain addresses related to CNN; the user might use this information for future configuration of spam filters in the email server (see Figure 5.9).

The user can check the junk box view to get more information about the highlighted received messages from the selected sender domain address. Based on our background knowledge of the corpus data set, we know that any message from “gestyre.com” is spam. The type column in Figure 5.10 shows that this message has been marked as spam. The user can gather contextual information about this specific message using the verified view.

Junk Box						
Sender	Receiver	Subject	Message	Date	Time	
bvermeul@gestyre.com	speedy.uwaterloo.ca	As seen on CNN and ABCnews!	If YOU...	Thu Apr 26 00:00:...	0:03	
cnnalerts@mail.cnn.com	speedy.uwaterloo.ca	CNN Alerts: bush	CNN Alert...	Sun Apr 22 00:00:...	19:04	
cnnalerts@mail.cnn.com	speedy.uwaterloo.ca	CNN Alerts: bush	CNN Alert...	Wed Apr 25 00:00:...	21:34	
cnnalerts@mail.cnn.com	speedy.uwaterloo.ca	CNN Alerts: bush		Fri Apr 27 00:00:0...	11:35	
cnnalerts@mail.cnn.com	speedy.uwaterloo.ca	CNN Alerts: bush		Tue Apr 24 00:00:...	8:35	
cnnalerts@mail.cnn.com	speedy.uwaterloo.ca	CNN Alerts: bush		Mon Apr 30 00:00:...	15:04	
cnnalerts@mail.cnn.com	speedy.uwaterloo.ca	CNN Alerts: bush		Sun Apr 29 00:00:...	9:34	
cnnalerts@mail.cnn.com	speedy.uwaterloo.ca	CNN Alerts: bush		Sun Apr 29 00:00:...	20:04	
cnnalerts@mail.cnn.com	speedy.uwaterloo.ca	CNN Alerts: bush		Wed Apr 25 00:00:...	17:35	
cnnalerts@mail.cnn.com	speedy.uwaterloo.ca	CNN Alerts: bush		Mon Apr 23 00:00:...	14:05	
cnnalerts@mail.cnn.com	speedy.uwaterloo.ca	CNN Alerts: bush	CNN Alert...	Sat Apr 28 00:00:...	12:04	
cnnalerts@mail.cnn.com	speedy.uwaterloo.ca	CNN Alerts: bush		Mon Apr 23 00:00:...	9:06	
		CNN Alerts: bush				
Delete						
Restore						
Verified Emails						
Type	From	To	Subject	Id	Langua...	Category
ham	cnnalerts@mail.cnn.com	5u0tf5\$117uu0@cnnima...	CNN Alerts: bush etc.	3850	english	unknown
ham	cnnalerts@mail.cnn.com	5u0tf5\$1j8hm7@cnnim...	CNN Alerts: bush etc.	14395	english	unknown
ham	cnnalerts@mail.cnn.com	5pu1v2\$8iqtdr@cnnima...	CNN Alerts: bush etc.	6651	english	unknown
ham	cnnalerts@mail.cnn.com	5u0tf5\$112d2l@cnnimai...	CNN Alerts: bush	870	english	unknown
ham	cnnalerts@mail.cnn.com	5u0tf5\$1j4nev@cnnimai...	CNN Alerts: bush	13883	english	unknown
ham	cnnalerts@mail.cnn.com	5u0tf5\$1jali9@cnnimai...	CNN Alerts: bush	2537	english	unknown
ham	cnnalerts@mail.cnn.com	5pu2lq\$1hgk8@cnnim...	CNN Alerts: bush	12654	english	unknown
ham	cnnalerts@mail.cnn.com	5pu1v2\$8ine5t@cnnima...	CNN Alerts: bush	1065	english	unknown
ham	cnnalerts@mail.cnn.com	5pu2lq\$1hkikh@cnnima...	CNN Alerts: bush	13824	english	unknown
ham	cnnalerts@mail.cnn.com	5pu2lq\$1ifhs9@cnnimai...	CNN Alerts: bush	7395	english	unknown
ham	cnnalerts@mail.cnn.com	5u0tf5\$1ivf23@cnnimai...	CNN Alerts: bush	13154	english	unknown
ham	cnnalerts@mail.cnn.com	5pu1v2\$8ir7t5@cnnima...	CNN Alerts: bush	9850	english	unknown
ham	cnnalerts@mail.cnn.com	5u0tf5\$1k2au6@cnnim...	CNN Alerts: bush	15678	english	unknown
ham	cnnalerts@mail.cnn.com	5pu2lq\$1icj7s@cnnimai...	CNN Alerts: bush	2651	english	unknown
ham	cnnalerts@mail.cnn.com	5u0tf5\$1k9rv6@cnnima...	CNN Alerts: bush	5027	english	unknown
ham	cnnalerts@mail.cnn.com	5u0tf5\$1kopud@cnnim...	CNN Alerts: bush	11939	english	unknown
ham	cnnalerts@mail.cnn.com	5pu2lq\$1i7bsg@cnnima...	CNN Alerts: bush	11117	english	unknown
ham	cnnalerts@mail.cnn.com	5u0tf5\$1k98jj@cnnimai...	CNN Alerts: bush	6107	english	unknown
ham	cnnalerts@mail.cnn.com	5pu1v2\$8hlqui@cnnima...	CNN Alerts: bush	5121	english	compute...

Figure 5.8: The non-spam identification first sample activity: recovering non-spam messages from “mail.cnn.com” to the verified view.

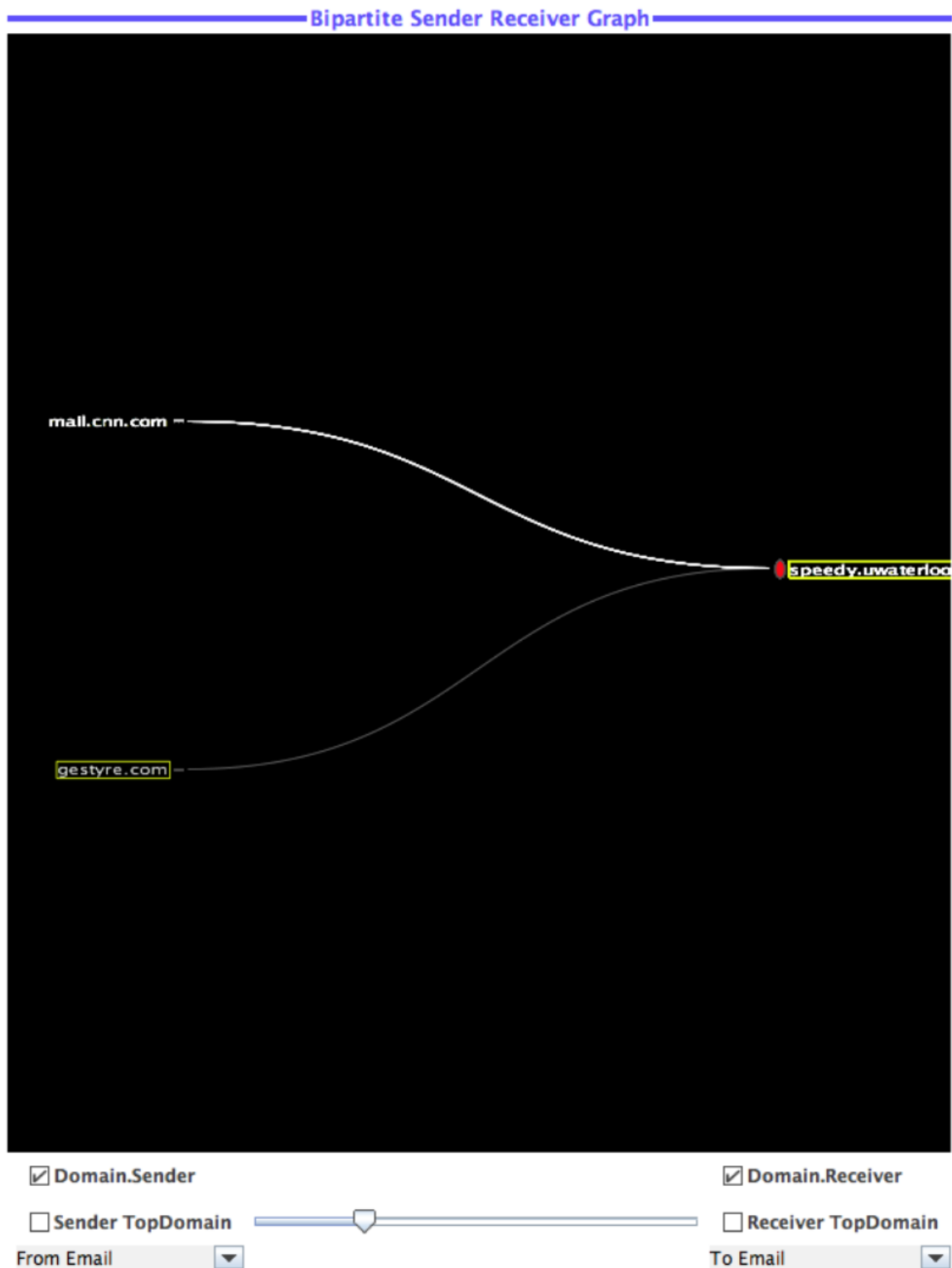


Figure 5.9: The non-spam identification first sample activity: selecting a sender domain address in the bipartite view.














Junk Box						
Sender	Receiver	Subject	Message	Date	Time	
 bvermeul@gestyre.com	speedy.uwaterloo.ca	As seen on CNN and ABCnews!	If YOU...	Thu Apr 26 00:00:...	0:03	▲
 cnnalerts@mail.cnn.com	speedy.uwaterloo.ca	CNN Alerts: bush	CNN Alert...	Sun Apr 22 00:00:...	19:04	
 cnnalerts@mail.cnn.com	speedy.uwaterloo.ca	CNN Alerts: bush	CNN Alert...	Wed Apr 25 00:00:...	21:34	≡
 cnnalerts@mail.cnn.com	speedy.uwaterloo.ca	CNN Alerts: bush		Fri Apr 27 00:00:0...	11:35	
 cnnalerts@mail.cnn.com	speedy.uwaterloo.ca	CNN Alerts: bush		Tue Apr 24 00:00:...	8:35	
 cnnalerts@mail.cnn.com	speedy.uwaterloo.ca	CNN Alerts: bush		Mon Apr 30 00:00:...	15:04	
 cnnalerts@mail.cnn.com	speedy.uwaterloo.ca	CNN Alerts: bush		Sun Apr 29 00:00:...	9:34	
 cnnalerts@mail.cnn.com	speedy.uwaterloo.ca	CNN Alerts: bush		Sun Apr 29 00:00:...	20:04	
 cnnalerts@mail.cnn.com	speedy.uwaterloo.ca	CNN Alerts: bush		Wed Apr 25 00:00:...	17:35	
 cnnalerts@mail.cnn.com	speedy.uwaterloo.ca	CNN Alerts: bush		Mon Apr 23 00:00:...	14:05	
 cnnalerts@mail.cnn.com	speedy.uwaterloo.ca	CNN Alerts: bush	CNN Alert...	Sat Apr 28 00:00:...	12:04	
 cnnalerts@mail.cnn.com	speedy.uwaterloo.ca	CNN Alerts: bush		Mon Apr 23 00:00:...	9:06	
		CNN Alerts: bush				▼
Delete						
Restore						
Verified Emails						
Type	From	To	Subject	Id	Langua...	Category
spam	bvermeul@gestyre.com	111001c787b8520fe0...	As seen on CNN and ABCnews!	10419	english	culture_... 4/2

Figure 5.10: The non-spam identification first sample activity: viewing a spam message from “gestyre.com” in the verified view.

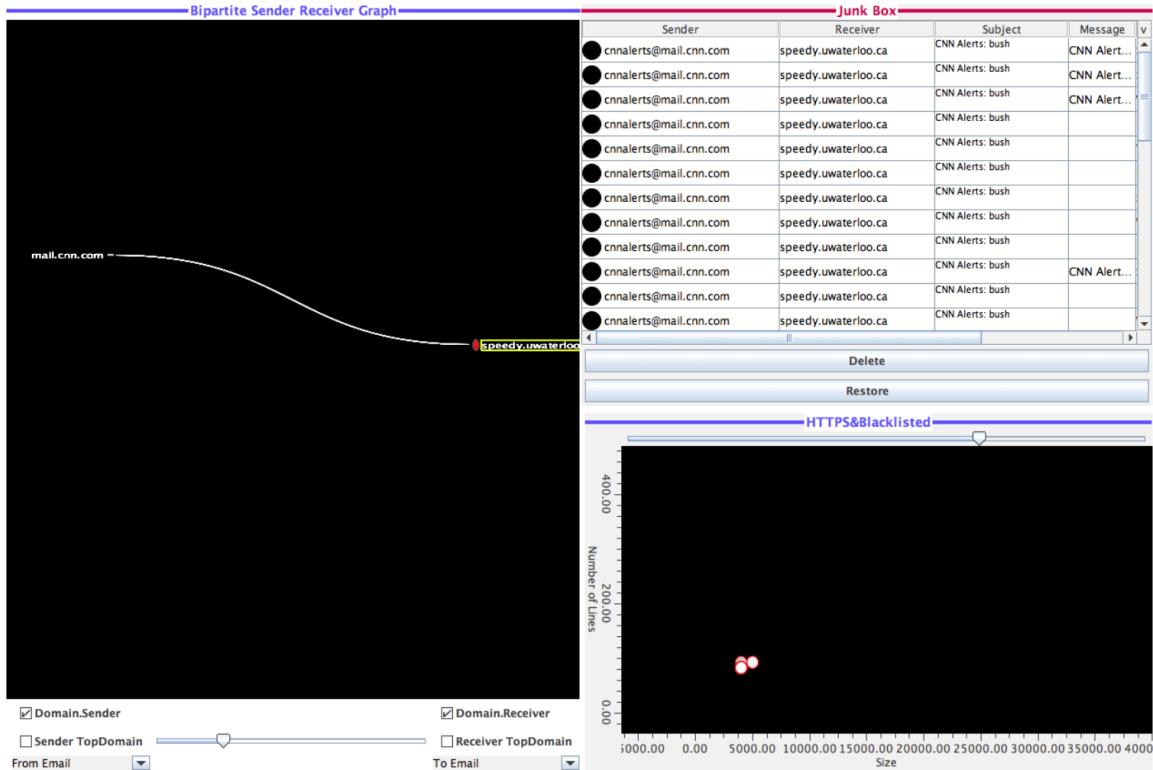


Figure 5.11: The non-spam identification first sample activity: removing the received spam message from “gestyre.com” from the junk box view.

Finally, the user removes the selected spam message received from the “gystyre.com” domain from the email server by pressing the delete button. Figure 5.11 shows that there is no spam message from “gystyre.com” in the email server after deletion.

5.1.2 Second Sample Activity

A server administrator at the Business Department at the University of Waterloo received the following report from a faculty member about losing important messages: “I have stocks from three companies. I usually check my stocks’ balance information using daily reports from the “shareholder.com” website. I have not received any emails from this website over the weekends. I have also checked my junk box folder, but I did not find any related email.”

Based on this report, the server administrator knows that the faculty member has already checked their junk box folder and did not find any message from “shareholder.com”. Therefore, the server administrator wants to check and verify if any message with those specifications has been detected as spam and stopped in the email server. The following is a list of actions that the user needs to take to accomplish this specific activity based on their knowledge of non-spam messages:

- The user knows that the faculty member cannot receive messages from that website on weekends (Sunday and Saturday). The user can therefore filter spam messages received on weekends in the email server.
- The user knows the sender domain address (“shareholder.com”), therefore they can search among a list of senders’ domains and mark senders with that domain address.
- The user can recover marked spam messages to end-users’ inboxes.
- The user can remove unmarked spam messages from the email server.

Figure 5.12 shows the initial visualization state of the bipartite, scatter plot, and junk box views in VeriVis, before the user starts looking for non-spam messages from “shareholder.com”. Figure 5.12 shows all messages classified as spam in the “flax9.uwaterloo.ca” email server.

The user selects Saturday and Sunday from the day of the week table view. Based on the information about non-spam messages in the report, the user knows that these target non-spam messages contain stock information. The user selects “information” and “stock” from the token words table view to minimize the number of spam messages for spam verification (see Figure 5.13).

At this time, the user filters all received spam messages in the email server based on the attribute values selected in the day of the week and token words table views.

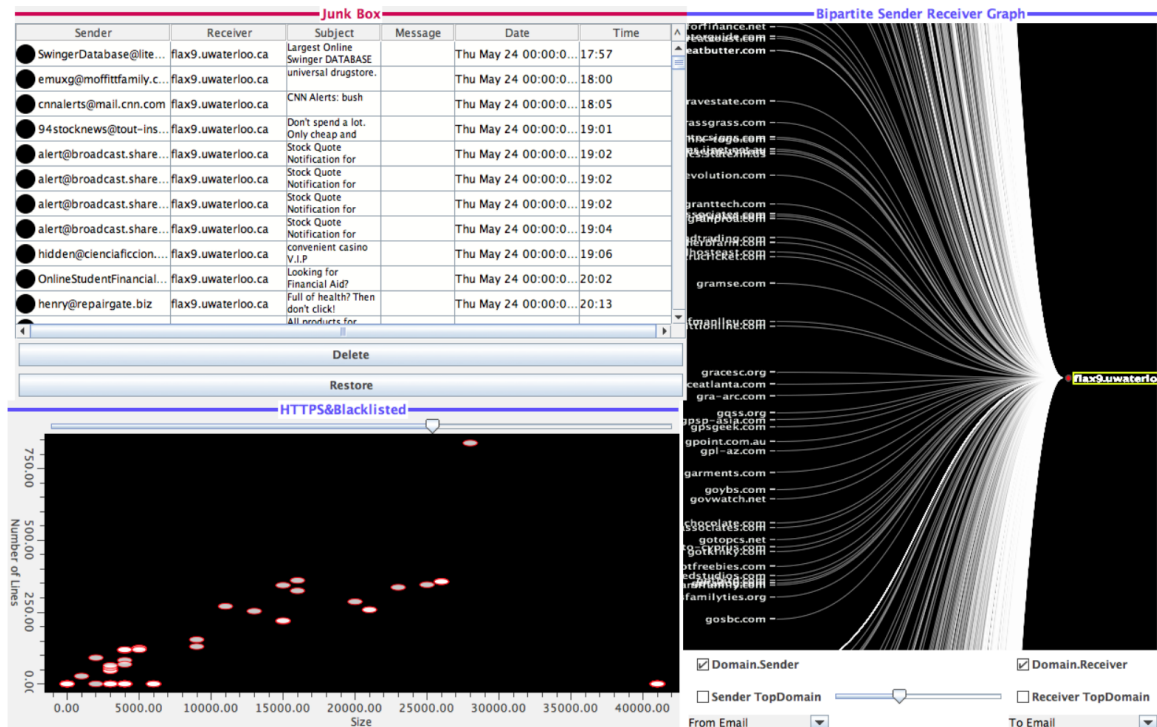


Figure 5.12: The non-spam identification second sample activity: initial visualization state for received spam messages in the “flax9.uwaterloo.ca” email server.

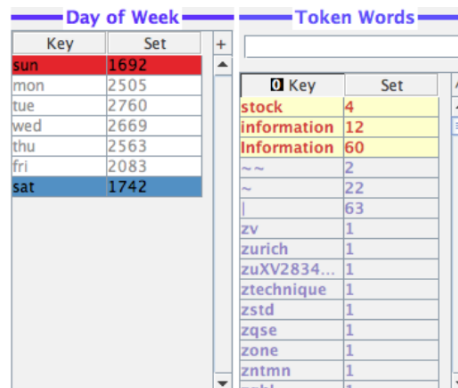


Figure 5.13: The non-spam identification second sample activity: selecting desired days of the week and token words in the corresponding table views.

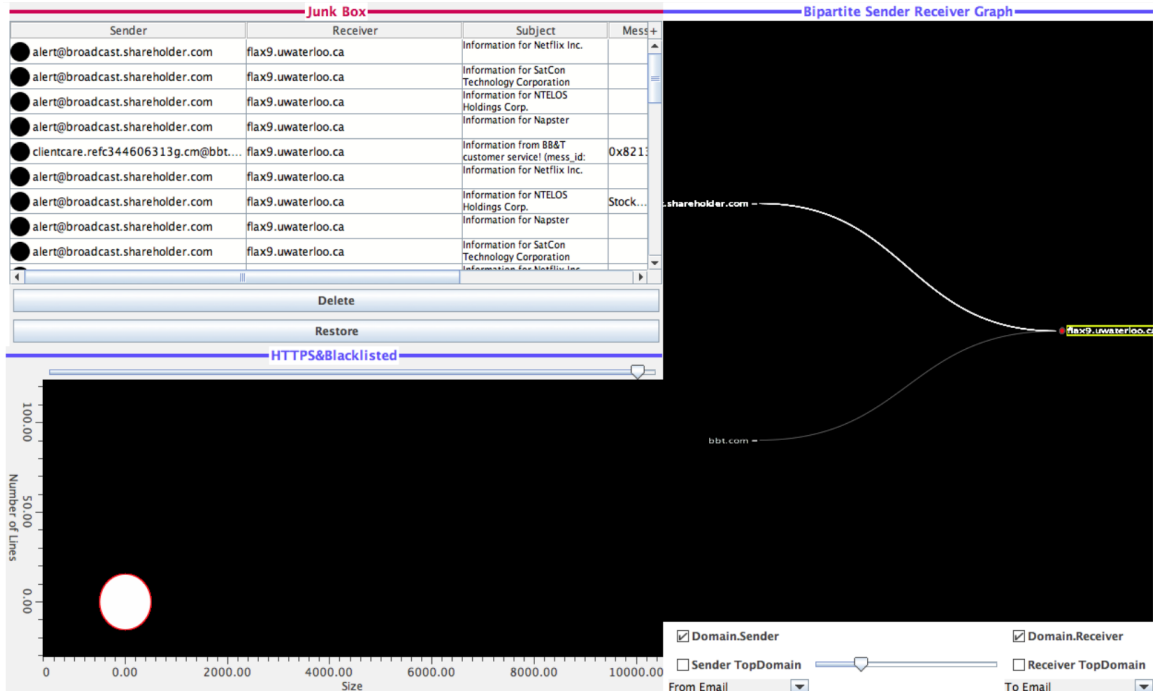


Figure 5.14: The non-spam identification second sample activity: filtering data records in the bipartite, scatter plot, and junk box views.

After filtering spam messages based on these two attributes, the user can explore the filtered results in other views (see Figure 5.14) to look for any email received from the “shareholder.com” domain. The user selects “shareholder.com” as the sender domain address in the bipartite view and subsequently highlights all received spam messages from that sender domain address in the junk box view (see Figure 5.15). Finally, the user marks all highlighted spam messages in the junk box view. In this case, the user recovers all marked spam messages to the verified view (see Figure 5.16). The “type” column in the verified view of Figure 5.16 shows that by using VeriVis, the user correctly identifies misclassified email messages from “shareholder.com” in the Trec07p corpus.

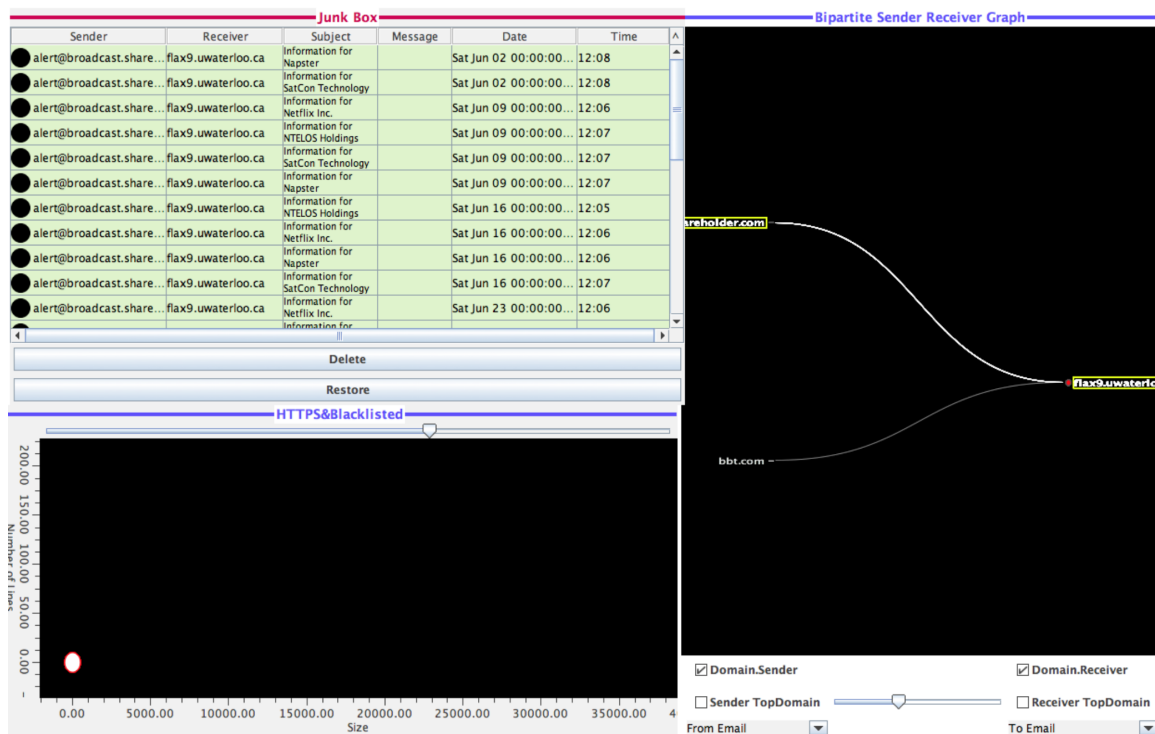


Figure 5.15: The non-spam identification second sample activity: selecting sender domain address in the bipartite view.

5.2 Spam Exploration: Sample Activities

5.2.1 First Sample Activity

Consider an Internet Service Provider (ISP) that provides email service for multiple organizations in Canada. A server administrator at the ISP receives the following report from several of the organizations: “In April 2007 we received a lot of VIAGRA spam messages. Please protect us from these types of messages.”

The user wants to identify the countries that most often originated messages related to VIAGRA and sent to the “.ca” top-level domain in April 2007. They might use this information to reconfigure the spam filters for the organizations. To explore patterns among source countries of messages related to VIAGRA, the user needs to take the following actions:

- Filter spam messages received in April 2007.

Junk Box				
Sender	Receiver	Subject	Mess	
alert@broadcast.shareholder.com	flax9.uwaterloo.ca	Information for Napster		
alert@broadcast.shareholder.com	flax9.uwaterloo.ca	Information for SatCon Technology Corporation		
alert@broadcast.shareholder.com	flax9.uwaterloo.ca	Information for Netflix Inc.		
alert@broadcast.shareholder.com	flax9.uwaterloo.ca	Information for NTELOS Holdings Corp.		
alert@broadcast.shareholder.com	flax9.uwaterloo.ca	Information for SatCon Technology Corporation		
alert@broadcast.shareholder.com	flax9.uwaterloo.ca	Information for Napster		
alert@broadcast.shareholder.com	flax9.uwaterloo.ca	Information for NTELOS Holdings Corp.		
alert@broadcast.shareholder.com	flax9.uwaterloo.ca	Information for Netflix Inc.		
alert@broadcast.shareholder.com	flax9.uwaterloo.ca	Information for Napster		
Delete				
Restore				
Verified Emails				
Type	From	To	Subject	Id
ham	alert@broadcast.shareholder.com	avcoopers@...	Information for Netflix Inc.	6934
ham	alert@broadcast.shareholder.com	avcooper@fl...	Information for NTELOS Holdings Corp.	524
ham	alert@broadcast.shareholder.com	avcooper@fl...	Information for SatCon Technology C...	7934
ham	alert@broadcast.shareholder.com	avcooper@fl...	Information for Napster	15725
ham	alert@broadcast.shareholder.com	avcooper@fl...	Information for NTELOS Holdings Corp.	12613
ham	alert@broadcast.shareholder.com	avcoopers@...	Information for Netflix Inc.	12614
ham	alert@broadcast.shareholder.com	avcooper@fl...	Information for Napster	12615
ham	alert@broadcast.shareholder.com	avcooper@fl...	Information for SatCon Technology C...	8753
ham	alert@broadcast.shareholder.com	avcoopers@...	Information for Netflix Inc.	12002
ham	alert@broadcast.shareholder.com	avcooper@fl...	Information for NTELOS Holdings Corp.	12376
ham	alert@broadcast.shareholder.com	avcooper@fl...	Information for Napster	13516
ham	alert@broadcast.shareholder.com	avcooper@fl...	Information for SatCon Technology C...	13519
ham	alert@broadcast.shareholder.com	avcoopers@...	Information for Netflix Inc.	7146
ham	alert@broadcast.shareholder.com	avcooper@fl...	Information for NTELOS Holdings Corp.	12114
ham	alert@broadcast.shareholder.com	avcooper@fl...	Information for Napster	7147
ham	alert@broadcast.shareholder.com	avcooper@fl...	Information for SatCon Technology C...	12190
ham	alert@broadcast.shareholder.com	avcoopers@...	Information for Netflix Inc.	1934
ham	alert@broadcast.shareholder.com	avcooper@fl...	Information for SatCon Technology C...	1931

Figure 5.16: The non-spam identification second sample activity: recovering email messages from the “shareholder.com” domain.

- Identify all spam messages related to VIAGRA.
- Extract their countries of origin and compare the number of VIAGRA-related spam messages received from each of the countries.

Figure 5.17 shows the initial visualization state of the bipartite, scatter plot, and junk box views before the user starts exploration. Figure 5.17 shows all spam messages in the email server sent to the “.ca” top-level domain. First, the user selects all days in April 2007 in the calendar view. They then select all tokens that contain the word “VIAGRA” in the token words table view. At this point, the user filters all received spam messages on selected “**subject**” and “**date**” attributes (see Figure 5.18). Figure 5.19 shows the resulting filtered data records in the bipartite, scatter plot, and junk box views.

Next, the user decides to use the highlighting feature to compare the source countries based on their assigned colors. The user selects “**Country**” to be a highlighting attribute in the highlighting control panel, then selects desired countries such as “Korea,” “Russia,” and “Argentina” in the country table view (see Figure 5.20).

Figure 5.21 shows VIAGRA spam messages received in April 2007, highlighted by source country. The bipartite view in Figure 5.21 shows that most of the received spam messages related to VIAGRA in April 2007 were sent from the Russian Federation (blue), followed by Argentina (orange), and then Korea (red).

5.2.2 Visual Outlier Detection: Sample Activity

A service provider wants to identify spam messages in the email server that look significantly different from the majority of received spam messages as a function of one or more message attributes. Figure 5.22 shows the initial visualization state for all received spam messages from the “.ca” top-level domain.

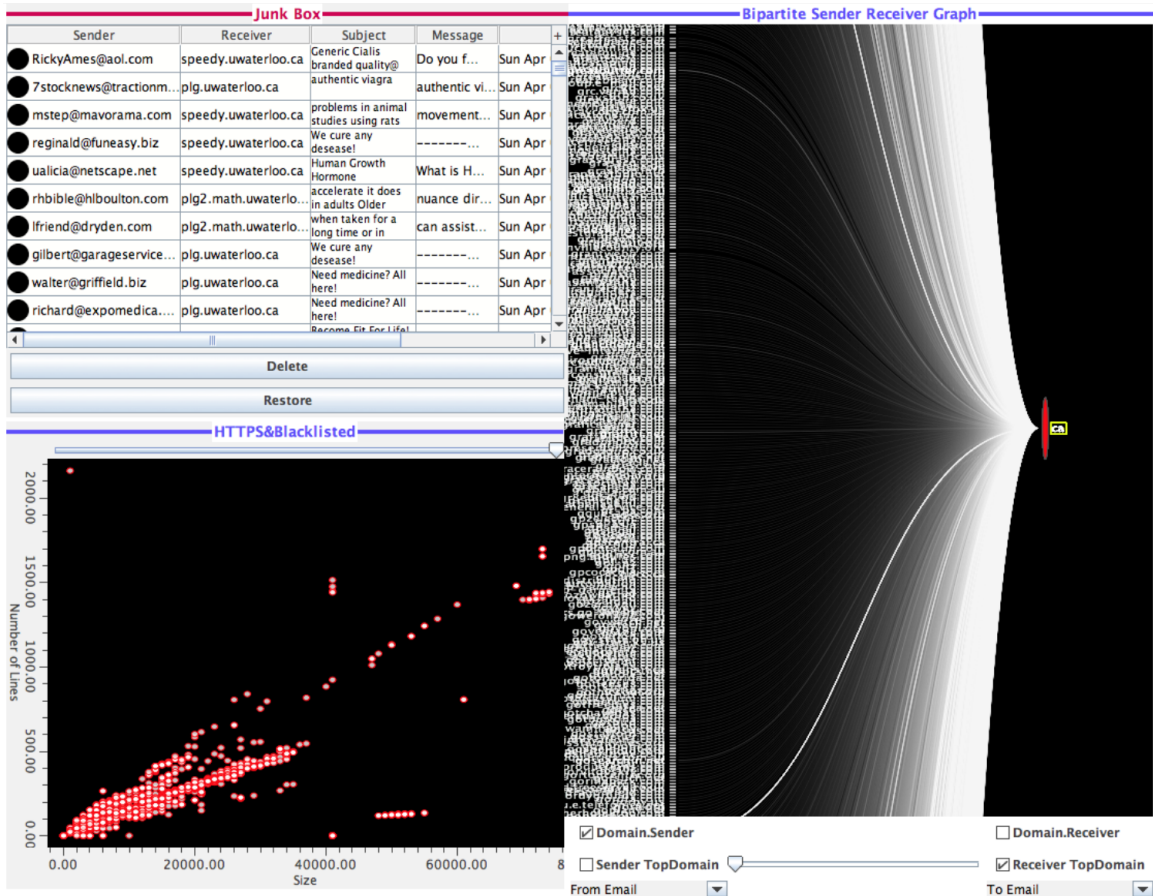


Figure 5.17: Spam exploration first sample activity: initial visualization state of the bipartite, scatter plot, and junk box views.

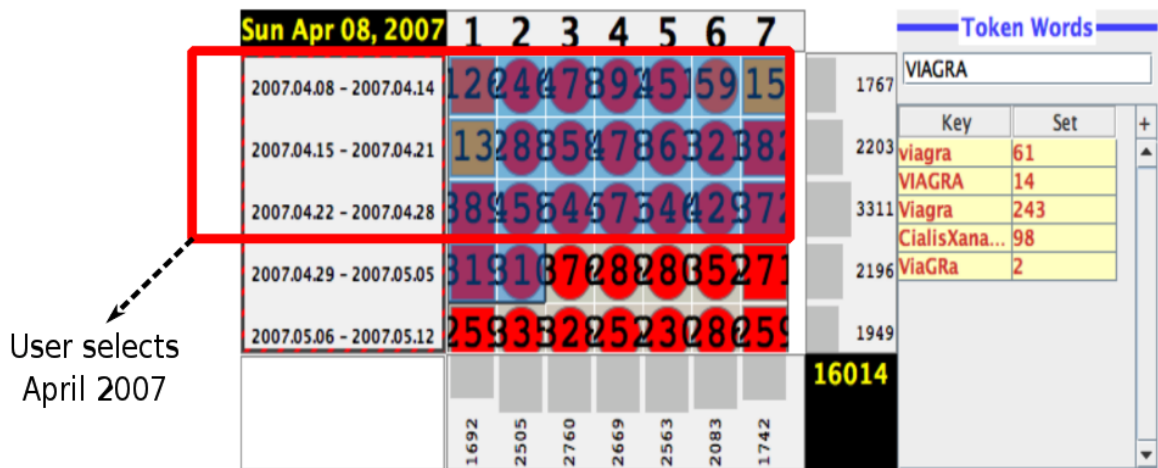


Figure 5.18: Spam exploration first sample activity: filtering on selected attribute values in the calendar and word token table views, using the filtering panel.

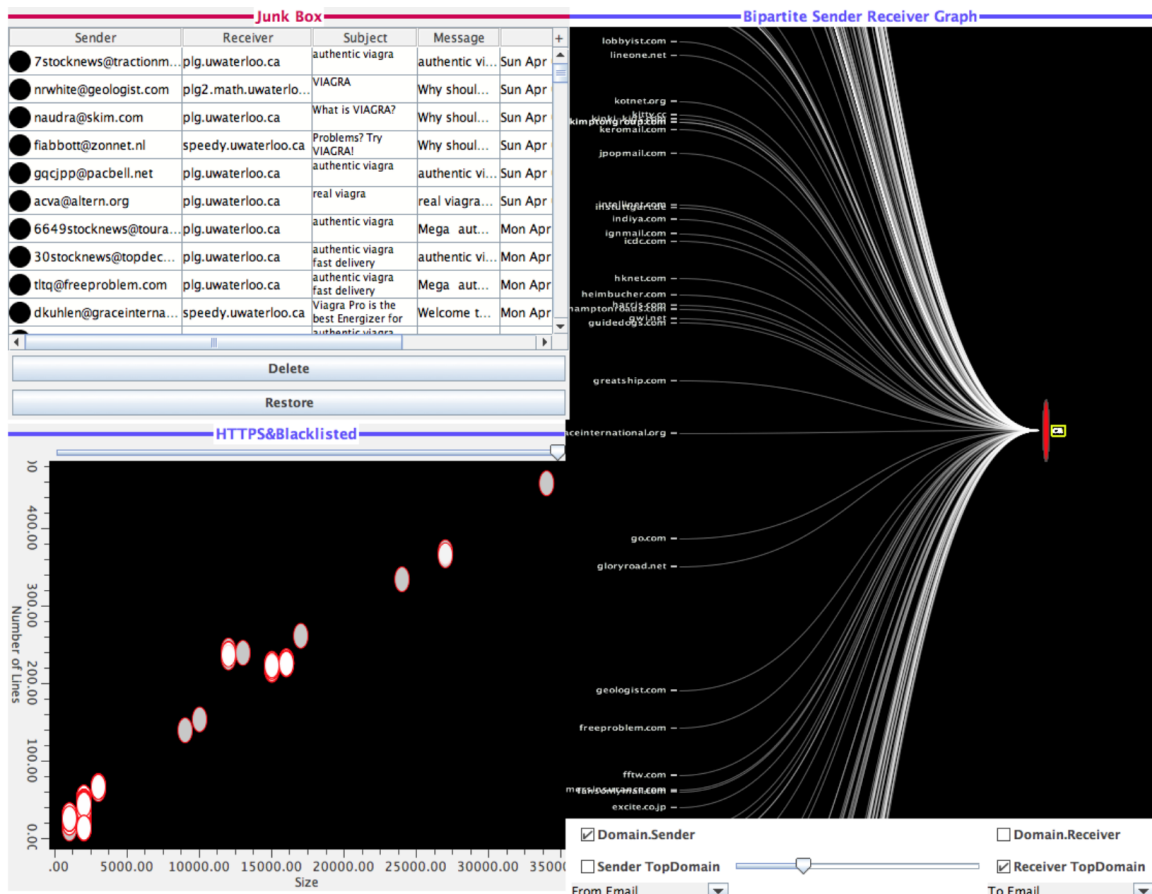


Figure 5.19: Spam exploration first sample activity: viewing patterns in VIAGRA spam messages received in April 2007.

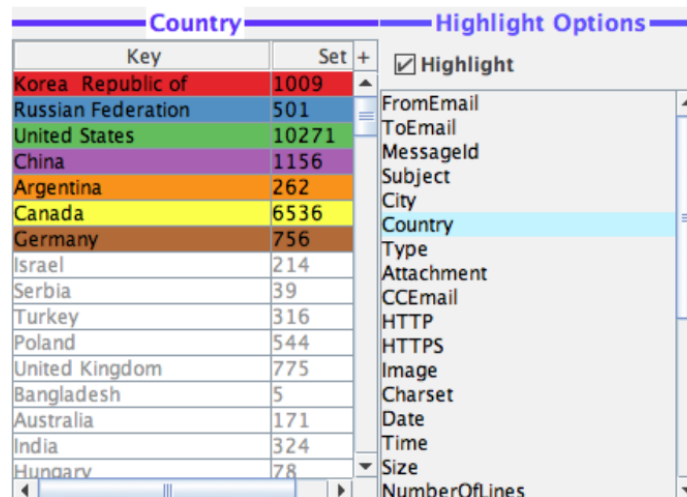


Figure 5.20: Spam exploration first sample activity: using the highlighting control panel to color selected countries in the country table view.

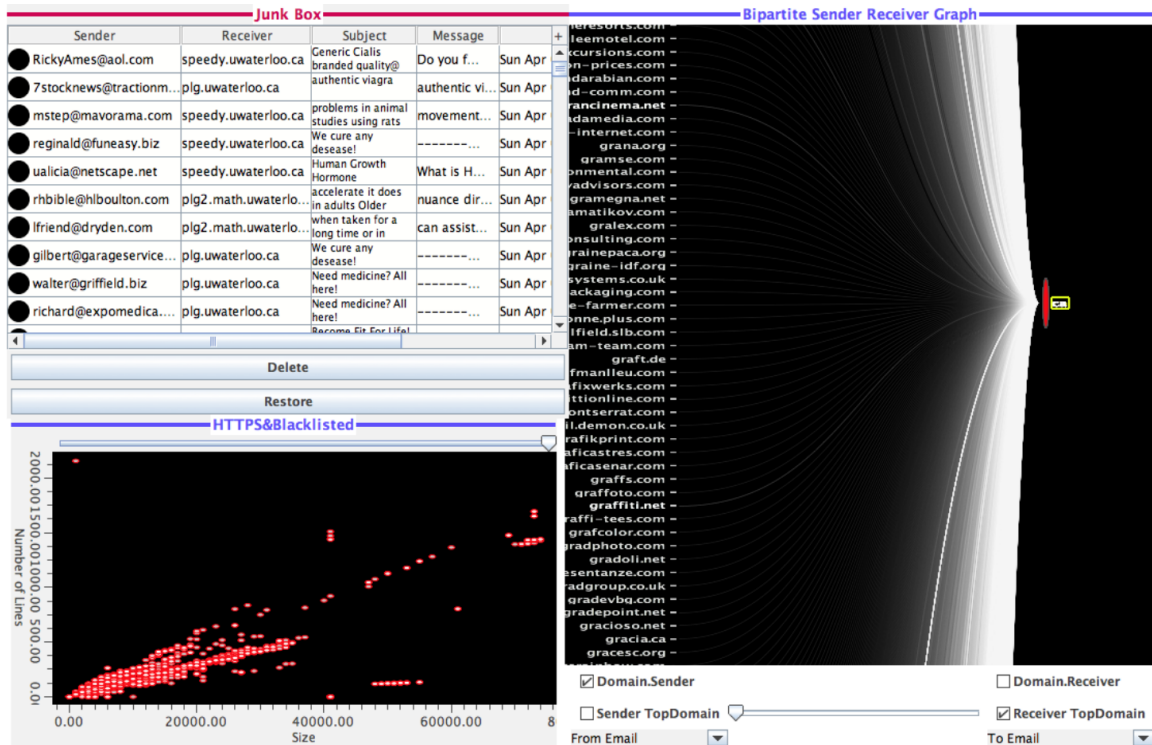


Figure 5.22: Visual outlier detection activity: initial visualization state for messages from the “.ca” top-level domain.

The scatter plot in VeriVis allows the user to compare all received spam messages in the email server based on their “size” and “number of lines” attributes. Using the initial visualization state, the user identifies a number of spam messages that vary from the norm in terms of those attributes (see Figure 5.22). The user selects these messages in the scatter plot, then looks for contextual information about them in other views (see Figure 5.23). As soon as the user selects those messages in the scatter plot, the same data records get highlighted in the junk box view (see Figure 5.24).

Now, the user recovers all highlighted messages from the junk box view to the verified view. This helps the user to check if those highlighted spam messages were really outliers or not. Figure 5.25 shows that most of the spam messages that were visually identified as outliers by the user using VeriVis are marked as non-spam messages in the Trec07p corpus. In this case, only three messages were identified as outliers by

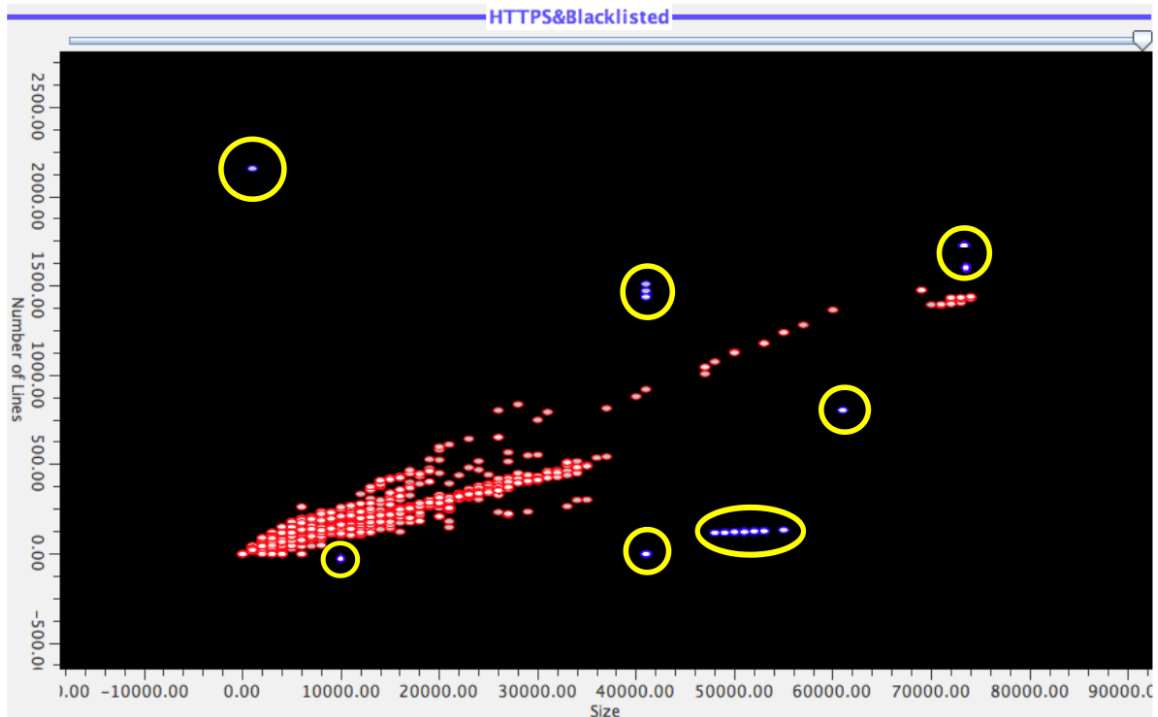


Figure 5.23: Visual outlier detection activity: selecting outlier spam messages in the scatter plot.

Junk Box						
Sender	Receiver	Subject	Message	Date	Time	
speedy.uwaterloo.ca@...	flax9.uwaterloo.ca	Chris Wragge: Come Celebrate With Mom		Fri May 11 00:00:00...	15:25	
speedy.uwaterloo.ca@...	speedy.uwaterloo.ca	60 Minutes E-mail Alert		Fri Apr 20 00:00:00...	16:18	
speedy.uwaterloo.ca@...	speedy.uwaterloo.ca	News Summary		Fri Apr 20 00:00:00...	16:50	
speedy.uwaterloo.ca@...	speedy.uwaterloo.ca	News Summary		Mon Apr 09 00:00:0...	12:06	
RickyAmes@aol.com	speedy.uwaterloo.ca	Generic Cialis branded quality@	Do you f...	Sun Apr 08 00:00:0...	13:07	
speedy.uwaterloo.ca@...	speedy.uwaterloo.ca	News Summary		Wed Apr 25 00:00:0...	9:17	
speedy.uwaterloo.ca@...	speedy.uwaterloo.ca	News Summary		Sun Apr 22 00:00:0...	16:38	
speedy.uwaterloo.ca@...	speedy.uwaterloo.ca	News Summary		Tue Apr 24 00:00:0...	9:18	
speedy.uwaterloo.ca@...	speedy.uwaterloo.ca	News Summary	CBSNews.c...	Tue Apr 17 00:00:0...	12:05	
speedy.uwaterloo.ca@...	speedy.uwaterloo.ca	News Summary		Tue Apr 24 00:00:0...	12:03	
speedy.uwaterloo.ca@...	speedy.uwaterloo.ca	News Summary		Thu Apr 26 00:00:0...	16:58	
speedy.uwaterloo.ca@...	speedy.uwaterloo.ca	60 Minutes E-mail Alert		Fri Apr 27 00:00:0...	16:08	
speedy.uwaterloo.ca@...	speedy.uwaterloo.ca	News Summary		Fri Apr 20 00:00:0...	12:06	
speedy.uwaterloo.ca@...	speedy.uwaterloo.ca	News Summary		Mon Apr 30 00:00:0...	9:17	
speedy.uwaterloo.ca@...	speedy.uwaterloo.ca	CBS News Sunday Morning: Royal attraction		Fri Apr 27 00:00:0...	13:04	
speedy.uwaterloo.ca@...	speedy.uwaterloo.ca	News Summary		Fri Apr 27 00:00:0...	9:18	
speedy.uwaterloo.ca@...	speedy.uwaterloo.ca	News Summary		Thu Apr 26 00:00:0...	12:04	
ispytraffic.com@grandl...	plg2.uwaterloo.ca	Separate yourself from other men	Hi She wa...	Tue May 08 00:00:0...	23:28	
nilsnf@plg2.math.uw...	plg2.math.uwaterlo...	pleasure more climax	Want to be...	Wed May 09 00:00:0...	3:03	
simon@repairnet.biz	flax9.uwaterloo.ca	All products for your health!	-----	Tue May 08 00:00:0...	23:24	

Figure 5.24: Visual outlier detection activity: highlighting selected (outlier) spam messages in the junk box view.

Verified Emails								
Type	From	To	Subject	Id	Language	Category	Date	
spam	wonderful@economysource...	l@speedy.uwaterloo.ca	SoftLaurie Offers Office 2007 for 795...	12184	english	compute	4/10/07	
spam	Editor@WindowsSecrets.com	langua2@speedy.uwaterl...	Help Fred Langa discover North Ame...	3477	english	compute	4/19/07	
spam	Editor@WindowsSecrets.com	langua@speedy.uwaterlo...	Help Fred Langa discover North Ame...	3479	english	compute	4/19/07	
ham	speedy.uwaterloo.ca@cbsig...	ktwarwic@speedy.uwate...	News Summary	3346	english	unknown	4/23/07	
ham	speedy.uwaterloo.ca@cbsig...	ktwarwic@speedy.uwate...	News Summary	10525	english	unknown	4/26/07	
ham	speedy.uwaterloo.ca@cbsig...	ktwarwic@speedy.uwate...	News Summary	7311	english	unknown	4/26/07	
ham	speedy.uwaterloo.ca@cbsig...	ktwarwic@speedy.uwate...	News Summary	13308	english	unknown	4/26/07	
ham	speedy.uwaterloo.ca@cbsig...	ktwarwic@speedy.uwate...	News Summary	2653	english	unknown	4/27/07	
ham	speedy.uwaterloo.ca@cbsig...	ktwarwic@speedy.uwate...	News Summary	10775	english	unknown	4/27/07	
ham	speedy.uwaterloo.ca@cbsig...	ktwarwic@speedy.uwate...	CBS News Sunday Morning: Royal attr...	2488	english	unknown	4/27/07	
ham	speedy.uwaterloo.ca@cbsig...	ktwarwic@speedy.uwate...	60 Minutes E-mail Alert	16064	english	unknown	4/27/07	
ham	speedy.uwaterloo.ca@cbsig...	ktwarwic@speedy.uwate...	News Summary	8846	english	unknown	4/27/07	
ham	speedy.uwaterloo.ca@cbsig...	ktwarwic@speedy.uwate...	News Summary	6007	english	unknown	4/28/07	
ham	speedy.uwaterloo.ca@cbsig...	ktwarwic@speedy.uwate...	News Summary	9684	english	unknown	4/29/07	
ham	speedy.uwaterloo.ca@cbsig...	ktwarwic@speedy.uwate...	News Summary	1601	english	unknown	4/30/07	
ham	speedy.uwaterloo.ca@cbsig...	ktwarwic@speedy.uwate...	News Summary	7883	english	unknown	4/30/07	
ham	speedy.uwaterloo.ca@cbsig...	ktwarwic@flax9.uwaterl...	48 Hours E-mail Alert	14523	english	unknown	5/4/07	
ham	speedy.uwaterloo.ca@cbsig...	ktwarwic@flax9.uwaterl...	48 Hours E-mail Alert	12311	english	unknown	5/11/07	
ham	speedy.uwaterloo.ca@cbsig...	ktwarwic@flax9.uwaterl...	CBS News Sunday Morning: Close loo...	1246	english	unknown	5/11/07	
ham	speedy.uwaterloo.ca@cbsig...	ktwarwic@flax9.uwaterl...	Chris Wragge: Come Celebrate With...	8406	english	unknown	5/11/07	
ham	speedy.uwaterloo.ca@cbsig...	ktwarwic@flax9.uwaterl...	60 Minutes E-mail Alert	8408	english	unknown	5/11/07	
ham	speedy.uwaterloo.ca@cbsig...	ktwarwic@flax9.uwaterl...	CBS News Sunday Morning: Designin...	14865	english	unknown	5/18/07	
ham	speedy.uwaterloo.ca@cbsig...	ktwarwic@speedy.uwate...	News Summary	5284	english	unknown	4/9/07	
ham	speedy.uwaterloo.ca@cbsig...	ktwarwic@speedy.uwate...	News Summary	12503	english	unknown	4/9/07	
ham	speedy.uwaterloo.ca@cbsig...	ktwarwic@speedy.uwate...	News Summary	6169	english	unknown	4/16/07	
ham	speedy.uwaterloo.ca@cbsig...	ktwarwic@speedy.uwate...	News Summary	12422	english	unknown	4/17/07	
ham	speedy.uwaterloo.ca@cbsig...	ktwarwic@speedy.uwate...	News Summary	6321	english	unknown	4/17/07	
ham	speedy.uwaterloo.ca@cbsig...	ktwarwic@speedy.uwate...	News Summary	14541	english	unknown	4/18/07	
ham	speedy.uwaterloo.ca@cbsig...	ktwarwic@speedy.uwate...	News Summary	6204	english	unknown	4/18/07	
ham	Editor@WindowsSecrets.com	langua3@speedy.uwaterl...	Help Fred Langa discover North Ame...	3480	english	unknown	4/19/07	
ham	speedy.uwaterloo.ca@cbsig...	ktwarwic@speedy.uwate...	News Summary	15542	english	unknown	4/19/07	
ham	speedy.uwaterloo.ca@cbsig...	ktwarwic@speedy.uwate...	News Summary	15068	english	unknown	4/20/07	
ham	speedy.uwaterloo.ca@cbsig...	ktwarwic@speedy.uwate...	60 Minutes E-mail Alert	14498	english	unknown	4/20/07	
ham	speedy.uwaterloo.ca@cbsig...	ktwarwic@speedy.uwate...	News Summary	9280	english	unknown	4/20/07	
ham	speedy.uwaterloo.ca@cbsig...	ktwarwic@speedy.uwate...	News Summary	9722	english	unknown	4/22/07	
ham	speedy.uwaterloo.ca@cbsig...	ktwarwic@speedy.uwate...	News Summary	8078	english	unknown	4/23/07	
ham	speedy.uwaterloo.ca@cbsig...	ktwarwic@speedy.uwate...	News Summary	8919	english	unknown	4/23/07	
ham	speedy.uwaterloo.ca@cbsig...	ktwarwic@speedy.uwate...	News Summary	2089	english	unknown	4/24/07	
ham	speedy.uwaterloo.ca@cbsig...	ktwarwic@speedy.uwate...	News Summary	831	english	unknown	4/24/07	
ham	speedy.uwaterloo.ca@cbsig...	ktwarwic@speedy.uwate...	News Summary	10279	english	unknown	4/25/07	
ham	speedy.uwaterloo.ca@cbsig...	ktwarwic@speedy.uwate...	News Summary	15770	english	unknown	4/25/07	

Figure 5.25: Visual outlier detection activity: the verified view, showing three messages identified as outliers by mistake.

mistake. In the real world, if a user were to recover these messages to end users' inboxes, it could increase the number of false negative errors.

The user wants to identify those three messages in the scatter plot that were mistakenly identified as outliers. For this purpose, the user repeats all of their previous interactions in the scatter plot. This time, as soon as the user selects any outlier items in the scatter plot, the same items get highlighted in the verified view as well. Figure 5.26 shows the messages mistakenly detected as outliers; Figure 5.27 shows the same messages highlighted in the verified view. As shown in Figure 5.27, one of the highlighted messages is a non-spam message that has the same size and number of lines as the real spam messages mistakenly detected as outliers.

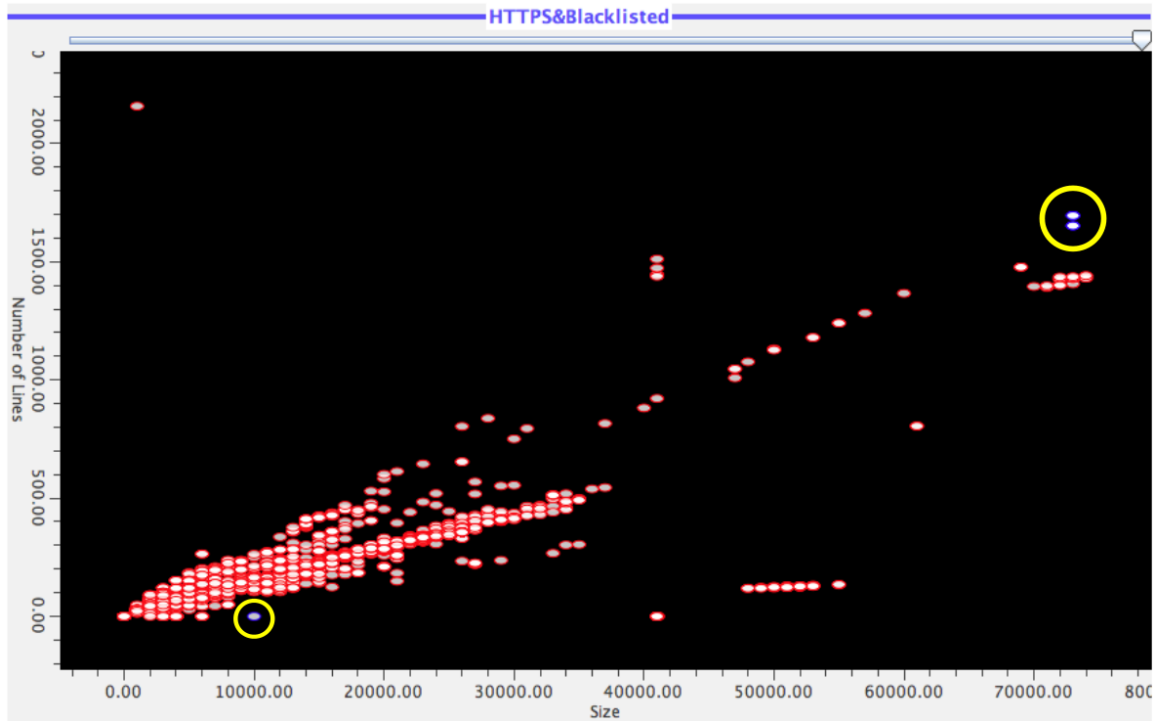


Figure 5.26: Visual outlier detection activity: spam messages mistakenly detected as outliers (circled) in the scatter plot.

Verified Emails									
Type	From	To	Subject	Id	Language	Category	Date		
spam	wonderful@economysource...	1@speedy.uwaterloo.ca	SoftLaurie Offers Office 2007 for 795...	12184	english	compute...	4/10/07		
spam	Editor@WindowsSecrets.com	langa2@speedy.uwaterloo.ca	Help Fred Langa discover North Ame...	3477	english	compute...	4/19/07		
spam	Editor@WindowsSecrets.com	langa@speedy.uwaterloo.ca	Help Fred Langa discover North Ame...	3479	english	compute...	4/19/07		
ham	speedy.uwaterloo.ca@cbisg...	ktwarwic@speedy.uwaterloo.ca	News Summary	5346	english	unknown	4/25/07		
ham	speedy.uwaterloo.ca@cbisg...	ktwarwic@speedy.uwaterloo.ca	News Summary	10525	english	unknown	4/26/07		
ham	speedy.uwaterloo.ca@cbisg...	ktwarwic@speedy.uwaterloo.ca	News Summary	7311	english	unknown	4/26/07		
ham	speedy.uwaterloo.ca@cbisg...	ktwarwic@speedy.uwaterloo.ca	News Summary	13308	english	unknown	4/26/07		
ham	speedy.uwaterloo.ca@cbisg...	ktwarwic@speedy.uwaterloo.ca	News Summary	2653	english	unknown	4/27/07		
ham	speedy.uwaterloo.ca@cbisg...	ktwarwic@speedy.uwaterloo.ca	News Summary	10775	english	unknown	4/27/07		
ham	speedy.uwaterloo.ca@cbisg...	ktwarwic@speedy.uwaterloo.ca	CBS News Sunday Morning: Royal attr...	2488	english	unknown	4/27/07		
ham	speedy.uwaterloo.ca@cbisg...	ktwarwic@speedy.uwaterloo.ca	60 Minutes E-mail Alert	16064	english	unknown	4/27/07		
ham	speedy.uwaterloo.ca@cbisg...	ktwarwic@speedy.uwaterloo.ca	News Summary	8846	english	unknown	4/27/07		
ham	speedy.uwaterloo.ca@cbisg...	ktwarwic@speedy.uwaterloo.ca	News Summary	6007	english	unknown	4/28/07		
ham	speedy.uwaterloo.ca@cbisg...	ktwarwic@speedy.uwaterloo.ca	News Summary	9684	english	unknown	4/29/07		
ham	speedy.uwaterloo.ca@cbisg...	ktwarwic@speedy.uwaterloo.ca	News Summary	1601	english	unknown	4/30/07		
ham	speedy.uwaterloo.ca@cbisg...	ktwarwic@speedy.uwaterloo.ca	News Summary	7883	english	unknown	4/30/07		
ham	speedy.uwaterloo.ca@cbisg...	ktwarwic@flax9.uwaterloo.ca	48 Hours E-mail Alert	14523	english	unknown	5/4/07		
ham	speedy.uwaterloo.ca@cbisg...	ktwarwic@flax9.uwaterloo.ca	48 Hours E-mail Alert	12311	english	unknown	5/11/07		
ham	speedy.uwaterloo.ca@cbisg...	ktwarwic@flax9.uwaterloo.ca	CBS News Sunday Morning: Close loo...	1246	english	unknown	5/11/07		
ham	speedy.uwaterloo.ca@cbisg...	ktwarwic@flax9.uwaterloo.ca	Chris Wragge: Come Celebrate With...	8406	english	unknown	5/11/07		
ham	speedy.uwaterloo.ca@cbisg...	ktwarwic@flax9.uwaterloo.ca	60 Minutes E-mail Alert	8408	english	unknown	5/11/07		
ham	speedy.uwaterloo.ca@cbisg...	ktwarwic@flax9.uwaterloo.ca	CBS News Sunday Morning: Designin...	14865	english	unknown	5/18/07		
ham	speedy.uwaterloo.ca@cbisg...	ktwarwic@speedy.uwaterloo.ca	News Summary	5284	english	unknown	4/9/07		
ham	speedy.uwaterloo.ca@cbisg...	ktwarwic@speedy.uwaterloo.ca	News Summary	12503	english	unknown	4/9/07		
ham	speedy.uwaterloo.ca@cbisg...	ktwarwic@speedy.uwaterloo.ca	News Summary	6169	english	unknown	4/16/07		
ham	speedy.uwaterloo.ca@cbisg...	ktwarwic@speedy.uwaterloo.ca	News Summary	12422	english	unknown	4/17/07		
ham	speedy.uwaterloo.ca@cbisg...	ktwarwic@speedy.uwaterloo.ca	News Summary	6321	english	unknown	4/17/07		
ham	speedy.uwaterloo.ca@cbisg...	ktwarwic@speedy.uwaterloo.ca	News Summary	14541	english	unknown	4/18/07		
ham	speedy.uwaterloo.ca@cbisg...	ktwarwic@speedy.uwaterloo.ca	News Summary	6204	english	unknown	4/18/07		
ham	Editor@WindowsSecrets.com	langa3@speedy.uwaterloo.ca	Help Fred Langa discover North Ame...	3480	english	unknown	4/19/07		
ham	speedy.uwaterloo.ca@cbisg...	ktwarwic@speedy.uwaterloo.ca	News Summary	15342	english	unknown	4/19/07		
ham	speedy.uwaterloo.ca@cbisg...	ktwarwic@speedy.uwaterloo.ca	News Summary	15068	english	unknown	4/20/07		
ham	speedy.uwaterloo.ca@cbisg...	ktwarwic@speedy.uwaterloo.ca	60 Minutes E-mail Alert	14498	english	unknown	4/20/07		
ham	speedy.uwaterloo.ca@cbisg...	ktwarwic@speedy.uwaterloo.ca	News Summary	9280	english	unknown	4/20/07		
ham	speedy.uwaterloo.ca@cbisg...	ktwarwic@speedy.uwaterloo.ca	News Summary	9722	english	unknown	4/22/07		
ham	speedy.uwaterloo.ca@cbisg...	ktwarwic@speedy.uwaterloo.ca	News Summary	8078	english	unknown	4/23/07		
ham	speedy.uwaterloo.ca@cbisg...	ktwarwic@speedy.uwaterloo.ca	News Summary	8919	english	unknown	4/23/07		
ham	speedy.uwaterloo.ca@cbisg...	ktwarwic@speedy.uwaterloo.ca	News Summary	2089	english	unknown	4/24/07		
ham	speedy.uwaterloo.ca@cbisg...	ktwarwic@speedy.uwaterloo.ca	News Summary	831	english	unknown	4/24/07		
ham	speedy.uwaterloo.ca@cbisg...	ktwarwic@speedy.uwaterloo.ca	News Summary	10279	english	unknown	4/25/07		
ham	speedy.uwaterloo.ca@cbisg...	ktwarwic@speedy.uwaterloo.ca	News Summary	15770	english	unknown	4/25/07		

Figure 5.27: Visual outlier detection activity: highlighting of spam messages in the verified view based on selections in the scatter plot in figure 5.26.

Chapter 6

Future Work and Conclusion

Spam verification in email servers is a way for server administrators to assess the accuracy of spam filtering. It allows them to identify, mark, and recover non-spam messages among those classified as spam, and send them to the users' inboxes if desired.

Given individual differences among people and their situation-specific definitions of spam, existing spam filters could be more effective. To the best of our knowledge, there is no spam filtering technique that can identify all types of spam messages without any misclassification errors. Spam verification alleviates this problem.

This thesis focuses on non-spam identification and spam exploration as two types of activities that can be performed using a spam verification tool. Non-spam identification consists of multiple steps. The first step in this type of activity is the identification of non-spam messages. The user first gets information about current spam messages, and then applies situation-specific criteria about non-spam messages to filter and identify possibly misclassified emails among those classified as spam.

VeriVis is a highly interactive visualization tool for spam verification in email servers. The tool allows users to drill down into the shared characteristics of spam messages in their email servers. The similarity, connectivity, and alignment of multiple coordinated views and the filtering and highlighting functionalities in VeriVis allow users to observe and identify patterns in multiple attributes of spam messages.

Based on our usage scenarios described in chapter 5, VeriVis can effectively support server administrators in performing spam verification activities for the purpose

of both spam exploration and non-spam identification. It also allows users to visually identify outlier messages among those classified as spam based on their “size” and “number of line” attributes. This type of visual outlier detection is distinct from statistical outlier detection. It would be interesting to study the integration of this type of visual outlier detection with a statistical outlier detection system in the future.

Design and implementation of a tool like VeriVis using an online visualization toolkit is a practical avenue to consider in the future. Currently, query calculation in VeriVis scales poorly with the number of spam messages being visualized. This may affect the speed of analysis in VeriVis, which is an important factor for highly interactive situations. Given a small enough input dataset—such as the 20,000 spam messages in our example corpus—VeriVis is suitably responsive to user interaction.

Currently, VeriVis does not consider user privacy as a concern for spam verification in email servers. Designing a high-level visualization that can allow email server administrators to perform spam verification while protecting access to users’ private information could also be considered in future work.

Choosing the right combination of views and visualization techniques can drastically affect the effectiveness of a visualization tool. For example, a set of views and visualization techniques that can be used for effective spam verification based on top-level domain information can be different from spam verification based on domain-level information. There is a vast design space of possibilities available to visualization designers to develop and apply other types of visualization techniques or other combinations of email attributes to the problem of spam verification at the server level. Finally, having access to a more comprehensive, yet still realistic spam corpus would help us to understand and exploit new derived email attributes that could be useful in improving the spam filtering and verification processes.

Reference List

- [1] Christopher Ahlberg and Ben Shneiderman. Visual information seeking: Tight coupling of dynamic query filters with starfield displays. In *Proceedings of the Conference on Human Factors in Computing Systems (CHI)*, pages 313–317, 479–480, Boston, MA, April 1994. ACM.
- [2] Christopher Ahlberg, Christopher Williamson, and Ben Shneiderman. Dynamic queries for information exploration: An implementation and evaluation. In *Proceedings of the Conference on Human Factors in Computing Systems (CHI)*, pages 619–626, Monterey, CA, May 1992. ACM.
- [3] Michelle Q. Wang Baldonado, Allison Woodruff, and Allan Kuchinsky. Guidelines for using multiple views in information visualization. In *Proceedings of the Working Conference on Advanced Visual Interfaces (AVI)*, pages 110–119, Palermo, Italy, May 2000. ACM.
- [4] Michael Bostock and Jeffrey Heer. Protovis: A graphical toolkit for visualization. *IEEE Transactions on Visualization and Computer Graphics (Proceedings of Visualization / Information Visualization 2009)*, 15(6):1121–1128, November–December 2009.
- [5] Michael Bostock, Vadim Ogievetsky, and Jeffrey Heer. D3: Data-driven documents. *IEEE Transactions on Visualization and Computer Graphics*, 17(12):2301–2309, December 2011.
- [6] Alex Brodsky and Dmitry Brodsky. A distributed content independent method for spam detection. In *Proceedings of the Workshop on Hot Topics in Understanding Botnets (HotBots)*, Cambridge, MA, April 2007. USENIX.

- [7] Gordon V. Cormack. TREC 2007 spam track overview. In *Proceedings of the Text Retrieval Conference (TREC)*, 2007.
- [8] Gordon V. Cormack. Email spam filtering: A systematic review. *Foundations and Trends in Information Retrieval*, 1(4):335–455, June 2008.
- [9] Leonie Bosveld de Smet and Mark de Vries. Visualizing non-subordination and multidominance in tree diagrams: Testing five syntax tree variants. In *Diagrammatic Representation and Inference*, volume 5223 of *Lecture Notes in Computer Science*, pages 308–320. Springer, 2008.
- [10] JooHyuk Jeon, Jihwan Song, JeongEun Kwon, YoonJoon Lee, ManHo Park, and MyoungHo Kim. An efficient and spam-robust proximity measure between communication entities. *Journal of Computer Science and Technology*, 28(2):394–400, 2013.
- [11] Bernard Kerr. THREAD ARCS: An email thread visualization. In *Proceedings of the IEEE Symposium on Information Visualization (InfoVis)*, pages 211–218, Seattle, WA, October 2003. IEEE.
- [12] Yun-Jung Lee, Min-Jung Bae, Gyun Woo, and Hwan-Gue Cho. A personalized visualizing and filtering system for a large set of responding messages on internet discussion forums. In *Proceedings of the International Conference on Computer and Information Technology (CIT)*, volume 2, pages 160–165, Xiamen, China, October 2009.
- [13] Gerald L. Lohse, Kevin Biolsi, Neff Walker, and Henry H. Rueter. A classification of visual representations. *Communications of the ACM*, 37(12):36–49, December 1994.
- [14] Jock D. Mackinlay. Automating the design of graphical presentations of relational information. *Transactions on Graphics*, 5(2):110–141, April 1986.

- [15] Amgad Madkour, Tarek Hefni, Ahmed Hefny, and Khaled S. Refaat. Using semantic features to detect spamming in social bookmarking systems. In *Supplemental Proceedings of the European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (ECML PKDD) (Discovery Challenge)*, pages 55–62, Antwerp, Belgium, September 2008.
- [16] MessageLabs. MessageLabs intelligence annual email security report. Technical report, MessageLabs, 2004.
- [17] Tony A. Meyer and Brendon Whateley. SpamBayes: Effective open-source, bayesian based, email classification system. In *Proceedings of the Conference on Email and Anti-Spam (CEAS)*, July 2004.
- [18] Samir A. Elzagheer Mohamed. Efficient spam filtering system based on smart cooperative subjective and objective methods. *International Journal of Communications, Network and System Sciences*, 6:88–99, 2013.
- [19] Chris Muelder and Kwan-Liu Ma. Visualization of sanitized email logs for spam analysis. In *Proceedings of the 6th International Asia-Pacific Symposium on Visualization (APVIS)*, pages 9–16, Sydney, Australia, February 2007.
- [20] Anirudh Ramachandran, David Dagon, and Nick Feamster. Can DNS-based blacklists keep up with bots? In *Proceedings of the Conference on Email and Anti-Spam (CEAS)*, Mountain View, CA, July 2006.
- [21] Anirudh Ramachandran, Nick Feamster, and Santosh Vempala. Filtering spam with behavioral blacklisting. In *Proceedings of the Conference on Computer and Communications Security*, pages 342–351, Alexandria, VA, October 29–November 2 2007. ACM.

- [22] Anthony C. Robinson and Chris Weaver. Re-visualization: Interactive visualization of the process of visual analysis. In *Proceedings of GIScience Workshop on Visual Analytics & Spatial Decision Support*, Münster, DE, September 2006.
- [23] Maryam Samiei, John Dill, and Arthur Kirkpatrick. EzMail: Using information visualization techniques to help manage email. In *Proceedings of the International Conference on Information Visualisation (IV)*, pages 477–482. IEEE Computer Society, 2004.
- [24] Ben Shneiderman. The eyes have it: A task by data type taxonomy for information visualizations. In *Proceedings of the IEEE Symposium on Visual Languages*, pages 336–343, Boulder, CO, September 1996. IEEE.
- [25] Mark Sweet. Political e-mail: Protected speech or unwelcome spam? *Duke Law & Technology Review*, 1(1), January 2003.
- [26] James J. Thomas and Kristin A. Cook, editors. *Illuminating the Path: The Research and Development Agenda for Visual Analytics*. IEEE Computer Society, August 2005.
- [27] Joseph Turian. Using AlchemyAPI for enterprise-grade text analysis. Technical report, AlchemyAPI, August 2013.
- [28] Fernanda B. Viégas, Danah Boyd, David H. Nguyen, Jeffrey Potter, and Judith Donath. Digital artifacts for remembering and storytelling: Posthistory and social network fragments. In *Proceedings of the Hawaii International Conference on System Sciences (HICSS)*, January 2004.
- [29] Fernanda B. Viégas, Martin Wattenberg, Frank van Ham, Jesse Kriss, and Matt McKeon. Many Eyes: A site for visualization at internet scale. *IEEE Transactions on Visualization and Computer Graphics*, 13(6):1121–1128, November/December 2007.

- [30] Chris Weaver. Building highly-coordinated visualizations in Improvise. In *Proceedings of the IEEE Symposium on Information Visualization (InfoVis)*, pages 159–166, Austin, TX, October 2004. IEEE Computer Society.
- [31] Chris Weaver. Visualizing coordination in situ. In *Proceedings of the IEEE Symposium on Information Visualization (InfoVis)*, pages 165–172, Minneapolis, MN, October 2005. IEEE Computer Society.
- [32] Chris Weaver. Metavisual exploration and analysis of DEVisé coordination in Improvise. In *Proceedings of the International Conference on Coordinated & Multiple Views in Exploratory Visualization (CMV)*, pages 79–90, London, UK, July 2006. IEEE Computer Society.
- [33] Chris Weaver. Coordinated queries: A domain specific language for exploratory development of multiview visualizations. In *Proceedings of the IEEE Symposium on Visual Languages and Human-Centric Computing (VL/HCC)*, pages 197–200, Herrsching am Ammersee, Germany, September 2008. IEEE Computer Society.
- [34] Chris Weaver, David Fyfe, Anthony Robinson, Deryck W. Holdsworth, Donna J. Peuquet, and Alan M. MacEachren. Visual analysis of historic hotel visitation patterns. In *Proceedings of the IEEE Symposium on Visual Analytics Science and Technology (VAST)*, pages 35–42, Baltimore, MD, October 2006. IEEE.
- [35] Brian Whitworth and Elizabeth Whitworth. Spam and the social-technical gap. *Computer*, 37(10):38–45, October 2004.
- [36] Kelvin S. Yiu, Ronald Baecker, Nancy Silver, and Byron Long. A time-based interface for electronic mail and task management. In *Proceedings of HCI International*, volume 2, pages 19–22, 1997.

- [37] Seongwook Youn and Dennis McLeod. Efficient spam email filtering using adaptive ontology. In *Proceedings of the International Conference on Information Technology (ITNG)*, pages 249–254, April 2007.