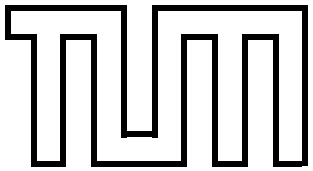


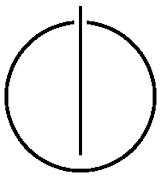
FAKULTÄT FÜR INFORMATIK

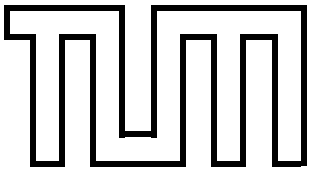
DER TECHNISCHEN UNIVERSITÄT MÜNCHEN

Master's Thesis in Informatik

Personalized Mass Email Communication

Baris Oztop





FAKULTÄT FÜR INFORMATIK

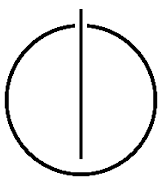
DER TECHNISCHEN UNIVERSITÄT MÜNCHEN

Master's Thesis in Informatik

Personalisierte Email Massenkommunikation

Personalized Mass Email Communication

Author:	Baris Oztop
Supervisor:	Prof. Dr. Johann Schlichter
Advisor:	Dr. Wolfgang Wörndl
Submission:	August 15, 2013



I assure the single handed composition of this master's thesis only supported by declared resources.

Munich, August 15, 2013

(Baris Oztop)

Acknowledgements

Acknowledgements goes here

Abstract

Reaching out to a large-scale audience via the Internet is a fast and cost-effective way compared with postal mail or telephone. Therefore, email has been used not just for research but also for marketing, customer support, and other data collection purposes. However, getting an acceptable response rate on the sent emails requires additional efforts on the researchers' side. This study investigates a communication system which contributes to increasing the response rate while minimizing the burden on the researchers' side.

To achieve this, the system constructs a workflow which supports the researchers in extracting information, providing a rule-based and automated decision-making mechanism for respondents' emails, and personalizes the content of the emails with the respondents' information which is extracted from either current or earlier conversations. It also provides an option to enable contributions from other researchers such as assistants to interact with the workflow with the permission of the initial researcher. Therefore, the distribution of the work can ease an individual's efforts of the mass email communication. This feature can be further extended by enabling crowd assistants to contribute to nearly all phases of the communication flow and getting guidance or assistance from the lead researcher when required.

This study demonstrates that by providing a proper workflow and enabling the possibility of an assistant's contribution, effective and efficient mass email communication can be achieved in a way each email was individually tailored for each recipient, which can contribute to higher response rate. As it minimizes the effort required to create emails, it maximizes the scale of people communicated with.

Contents

List of Figures

List of Tables

Listings

List of Acronyms

BCC Blind Carbon Copy

CC Carbon Copy

CRM Customer Relationship Management

CSS Cascading Style Sheets

HCI Human-Computer Interaction

HTML HyperText Markup Language

IMAP Internet Message Access Protocol

KVP Key-Value Pair

NDA Non-Disclosure Agreement

RoR Ruby on Rails

SaaS Software as a Service

SMTP Simple Mail Transfer Protocol

UI User Interface

UID Unique Identifier

URL Uniform Resource Locator

1 Introduction

Increased Internet usage turned email into a tool for communication, replacing telephone calls and regular mail (Norman and Lutz, 2000; Madden and Rainie, 2003). Email is used in many ways, proving that it plays a huge role in the communication world. Email is popularly used in; marketing, for engaging with clients; customer support, for offering aftersales assistance; research, on gathering the opinion of people on a certain topic; and many other cases showing that email has essentially become a part of our daily lives.

However, as the number of people you want to reach increases, the way on how you compose emails and extract information changes. Because, each email response should be uniquely composed based on the flow of the conversation to effectively deliver which require an increased personal effort and time on the part of the researchers. As a result, researchers tend to use online tools or third-party applications on sending out generic emails to their recipients with non-adequate personalization, which is known as one important factor needed to increase response rates (Dillman, 1991; Schaefer and Dillman, 1998). Such emails are treated with a low priority, which results to low response rates at the end (Dillman et al., 2009, page 272).

There are several products in the market focusing on email communication and data collection. A Customer Relationship Management (CRM) application keeps track of a company's communication with their clients. A help desk application offers a platform on helping solve customers' problems and provide guidance regarding products. Email marketing applications help on sending out commercial messages to groups of people. Finally, survey applications aid on conducting online surveys in getting people's views and opinions. The similarity of all these applications is that they all focus on email communications. However, none of these mentioned tools are capable of offering a complete

workflow in helping out a researcher communicate with a great amount of people on a personalized level, as easy as possible by email.

The goal of this study is to understand the possible workflow of a personalized mass email communication, and to show that it is possible to reach a great amount of people while keeping it personalized at the same time. A complete system named Myriad has been developed to demonstrate the practical aspects of this idea.

1.1 Email as a Data Collection Method

There is nearly a 600% growth rate in the world-wide Internet usage between the years 2000 to 2012 that makes Europe's 63% and North America's 80% over-all population Internet usage proportion (Group, 2012). Email is ranked as the most popular online activity, along with search engine usage with 92% of online adult users (Purcell, 2011). Also, the connectivity and the flexibility have increased with the introduction of smartphones and tablet devices (Madden and Jones, 2008). In addition to these facts, email is low cost and has a quick turnover compared to regular mail or telephone communication (Zikmund and Babin, 2006). Therefore, email as a part of communication is considered a viable option for data collection as well (Zikmund and Babin, 2006).

There are several reasons for data collection depending on the situation. However, purposes of data collection can be classified under the following categories (Sue and Ritter, 2011) (Babbie, 2012, pages 92–94):

1. To explore and get information about a topic
2. To describe the events and the situations
3. To explain things by questioning

To illustrate these purposes and to see how we can use email to explore, describe, and explain things, let's suppose that we have an online learning platform, offering various courses publicly:

Exploration Offering online courses is relatively a new trend; therefore we do not have much previous knowledge about the topic. To explore the popularity of the platform,

we would need to ask the platform's users questions such as: Why are they attending our online courses? Have they taken any online courses before? What are their income levels? Figuring out the answers to these questions will help us improve the system or to decide on its future. For example, the aggregated answers to the income level question will help us decide whether to charge the users for their usage or to offer it for free and find some sponsors to make it more viable.

Description Our goal is to describe the characteristics of the online learning platform's users. The questions that can help us to describe this can be: Where do they come from? What age range do they belong to? Were they able to attend college? With these questions, we will end up with a user profile like: between ages 16 – 22; have never attended college, and is coming from a less developed country. Knowing our users' portfolio according to this outcome can help us to attract organizations with necessary connections in supporting such countries' young population. Hence, they can take advantage of our platform as a tool in reaching those populations, in return they monetize our platform.

Explanation With our descriptive study, we discovered out that our platform's user's age range is between ages 16 – 22. The reasons on how our platform's users' age range turned out to be between 16 – 22 makes up our explanatory purpose. Questions about how often they are connected to the Internet or have they attended college or a similar high level of an education institute might help us to figure out the answer on why do young people use our platform more frequently compared to older people. Collecting such statistics may help us to develop an explanation to a topic.

Since all of our registered users provided their email addresses as the primary and mandatory contact medium, we can send them emails to conduct our data collection whether the reason is to explore, describe or explain the user trends on our online learning platform.

1.2 Problem Statement

To date, email, as a popular medium for communication, is utilized for many purposes such as reaching groups of people to explore, describe, and explain things. However, as the group's size gets larger, it becomes harder on the researchers' part to maintain the consistency and effectiveness of the flow of the exchange of emails as compared to that of small groups. Therefore, the researchers tend to write generic emails ignoring or using inadequate recipient specific information with the help of a third-party application or an online tool. This results into low response rates, since recipients do realize that because they are a part of a large group of people being responded to means that you will feel less important and less valued, and the chance of getting a reply is less likely to happen. On the other hand, if researchers individually tailor those emails according to their recipients, it will require a huge additional effort at an increased cost, hence reducing the advantage of using email as the primary communication medium.

Even though, there are many products available in the market supporting email communication, there is just no available product allowing anyone to reach larger groups via email, requiring minimum effort while keeping the communication personalized at the same time.

The main goal of this study is to show that a personalized email communication with large groups is possible if a proper workflow is provided. In order to achieve this goal, the study will:

1. Examine the workflow of an email communication with large groups and possible exceptional cases on this flow.
2. Investigate the effects of an email's content's personalization on the response rates.
3. Describe how an adequate amount of personalization in emails can be supplied.
4. Analyze the comparison of existing products claiming to provide solutions on email communication and collection of the respondents' information.
5. Describe the design and implementation of an application satisfying the mentioned workflow to aid the researchers, including the initial prototype.
6. Show how assistants can support the mentioned workflow.

7. Analyze real life usage of the application and its users' opinions about the application, and the latest statistical information giving an insight on how and in which way the application is used by its users.

This study also contributes on the following areas:

1. Email as a data collection method.
2. Conducting surveys with the use of email.
3. Defining a workflow on a mass email communication.
4. Possible crowd-sourced assistant usage.
5. Personalization of email content.

1.3 Outline

Chapter 1 The first chapter introduces the concept of the personalized mass email communication, defines email as a data collection medium, as well as its purposes and continues with the problem statement and the contributions of this study.

Chapter 2 The second chapter gives the necessary foundation on data collection by investigating related work about the email surveys, its errors, the factors affecting the response rates on a research and the studies on the personalization of emails.

Chapter 3 The third chapter is about existing applications available in the market and their connection with mass email communication as well as their features useful in reducing the efforts of a researcher on initiating a mass email campaign.

Chapter 4 The fourth chapter builds up a mass email communication scenario and introduces the prototype built upon to reflect the initial findings on a personalized mass email communication. The prototype will be reviewed, including its requirement analysis and architecture and finally, its evaluation.

Chapter 5 The fifth chapter will introduce the final solution, the developed and enhanced idea of the initial findings on a personalized mass email communication. The final

solution will be reviewed, including its improved requirement analysis and architecture, and finally, the benefits of the solution will be described together with the experience of users with it will be brought in, with the statistical results.

Chapter 6 The last chapter will summarize the findings according to the chapters, and mention about the future work for the provided solution.

2 Foundation and Related Work

This chapter presents the related work on the data collection domain. Even though technology is different for email surveys in collecting data from well-established regular mail-surveying methods, the nature of the communication is similar to self-administrated questionnaires (Schaefer and Dillman, 1998). In lieu with this, this chapter will also investigate regular mail surveys in emphasizing points which are also related with email communications, and the earlier studies about response rate influences.

2.1 Surveys and Data Collection

A Survey is defined as a system for collecting information (Sue and Ritter, 2011, page 3). It helps to learn about people's opinions and behaviors (Dillman et al., 2009). The produced data during or at the completion of the survey belongs to the data collection process. Therefore, data collection is a fundamental step in producing useful data to enable analysis on the researcher's part (Groves et al., 2009, page 149). These researches include – but not limited to – many disciplines like sociology, statistics, psychology, marketing, economics, and health sciences.

2.1.1 Email Surveys

Comparing many different characteristics of surveys and interviews, the concerns regarding speed and cost make the most powerful differences (Sproull, 1986; Schaefer and Dillman, 1998). Email surveys offer more rapid surveying than other methods including regular mail and telephone surveys. In addition to that, email surveys are inexpensive since it removes postage, paper and printing, and interview costs (Schaefer and Dillman, 1998).

Sproull (1986) identified the characteristics of email with organizational research, within a Fortune 500 office products and systems manufacturer, who were using email for 12 years in the organization and over 80 percent of all employees in the selected unit had email access at the time of the research. Selected candidates were separated into two groups. The data collection protocol within the organization asked each of the group's participants series of questions regarding their 3-day old email inbox. Both groups filled out the questionnaire and answered open-ended questions either electronically or in writing.

The result of the study indicated that the average duration of the data collection process for the email version was less than a week, which is half of the duration of the written version. While the response rate of the email version was 73 percent, the conventional written version's rate was 87. The percentage of missing data in the questionnaires was .2 percent in the written version, and 1.4 in the email version. There were no differences in the nature of answers in the email version compared with the written questionnaire.

In another study by Sheehan and Hoy (2006), they administered an email only survey to query individuals about their online behaviors, attitudes and opinions regarding their privacy concerns. They have reached the shortest response time of 3.65 days, compared with earlier studies conducted until that time (See table 2.1).

Table 2.1: Summary of Survey Research Methods Using E-mail (Sheehan and Hoy, 2006)

Author	Response Sample	Survey Topic	Sample Size	Usable Sample	Method	Response Rate	Time (days)
Kiesler & Sproull (1986)	Employees of a Fortune 500	Corporate Communication	115	77	Mail	67%	10.8
			115	86	Email	75%	9.6
Parker (1992)	Employees of AT&T	Internal Communication	70	27	Mail	38%	NA
			70	48	Email	68%	NA
Schuldt & Totten (1994)	Marketing & MIS Professors (US)	Shareware Copying	200	113	Mail	56.5%	NA
			218	42	Email	19.3%	NA
Mehta & Sivadas (1995)	Usenet Users	Internet Communication	309	173	Mail	56.5%*	NA
			182	99	Email	54.3%*	NA
Tse, et al (1995)	University Population (HK)	Business Ethics	200	54	Mail	27%	9.79
			200	12	Email	6%	8.09
Bachman, Elfrink & Vazzana (1996)	Business School Deans	TQM	224	147	Mail	65.6%	11.18
			224	117	Email	52.5%	4.68
Sheehan & Hoy (1997)	University Population (Southeast US)	Privacy and New Technology	580	274	Email	47.2%	4.7

Continued on next page

Table 2.1 – continued from previous page

Author	Response Sample	Survey Topic	Sample Size	Usable Sample	Method	Response Rate	Time (days)
Smith (1997)	Web presence	Business Activities	150	11	Email survey	8%	NA
			150	42	Email solicit	11.3%	NA
Schillewaert, Langerak and Duhamel (1998)	Web users in Belgium	Attitudes toward the Web	430	125	Email	31%	NA
			62.5M	110	Ad in magazine	0%	NA
			4000	67	USENET Posting	2%	NA
			7500	51	Hyperlinks	0.68%	NA
Weible and Wallace (1998)	MIS Professors (US)	Internet Use	200	70	Mail	35.7%	12.9
			200	50	Fax	30.9%	8.8
			200	48	Email	29.8%	6.1
			200	52	Web form	32.7%	7.4
Schaefer and Dillman (1998)	University Faculty	Unknown	226	130	Mail	57.5%*	14.39
			226	131	Email	58.0%*	9.16
*Differences not significant							

In addition to the speedy response time of the email surveys, cost benefits have been indicated in Sheehan and Hoy's (2006) study. They also concluded that email is considered as an extremely cost-efficient method for data collection, where the total cost estimated at $470 ($30 for printing out the responses, $440 for the 22-hour computer usage on downloading surveys for printing), while postal mail costs were estimated at $6,500 (printing, postage, survey, and reminder mailing).

In another study from Mavis and Brocato (1998), email survey was considered to be as nearly as seven times more cost efficient compared to a postal survey. This includes labor hours, survey materials like booklets, mailing labels, envelopes, and postage costs. Total time spent for the postal survey was 33 hours, but it only required 12 hours for the email survey. Final cost was $503.36 for postal survey – $305.36 of which for postage and the remaining $198 for student labor costs. Now on the other hand, email survey costs amounted to only $72 in total.

Moreover, Paolo et al. (2000) reported that their respondents made longer comments to open-ended questions for the email version of the survey, compared to the regular mail version. While the average number of words per comment was 58.33% in the mail version, the average for the email version was 75.40%. Bachmann et al. (1999) had similar

findings on their studies conducted on 1995 and 1998, where open-ended questions were more likely to be responded on by email recipients than by mail recipients. In the latter study conducted in 1998, the researchers also found out that email respondents were more likely to expand their answers, even if it was not suggested on the survey, resulting in a more candid set of responses compared to the set of responses on mail surveys. Responses to open-ended questions are one of most the important measures on determining the quality of the returned surveys.

Given these advantages and positive benefits of email surveys, the next section will provide information on survey errors.

2.1.2 Survey Errors

Sample surveys are quantitative estimations of the distribution of a characteristic in a population by obtaining this information from a small portion of the corresponding population (Dillman, 1991). To generalize results from a small portion, which is a sample, to a population, following sources of errors needs to be considered (Dillman, 2006, page 9; Dillman, 1991):

Sampling Error The greater number of people surveyed, the larger degree of precision can be achieved. Therefore, limitations on the number of people surveyed are considered under the sampling error. For example, while public opinion of 100 people results $\pm\$10\%$ of the true percent, 2,200 people results higher confidence with the percent of $\pm\$2\%$ (Dillman, 2006, page 9). Surveys relying on a predefined list of recipients considers that the list is randomly generated or with a systematic sampling. Hence, it has got little research to reduce sampling errors compared with face-to-face interviews in which multistage cluster designs[1] are used due to cost and time limitations (Groves et al., 2009, page 106; Dillman, 1991).

[1] Cluster sampling selects preexisting groups of population elements instead of a single element of the population (Groves et al., 2009, page 106). Departments of a university or households in a block represents clusters of people. When the allocation of those sampling resources are stratified and based on multiple stages, frequently three stages, it is called multistage cluster sampling. First step selects the sample of counties, followed by the blocks within those counties, and finally the dwellings from the chosen blocks (Scott and Smith, 1969).

Coverage Error When the list of surveyed people does not include all elements of the population, coverage error happens (Dillman, 2006, page 9). Coverage error is considered as one of the biggest issues of surveys ever since while surveying the general public (Dillman, 1991).

Measurement Error When a respondent's answer is hard to evaluate or cannot be compared with the other respondent's answers or there are inconsistencies between the observable variables like opinions, behaviors, or attributes and the survey responses, measurement error happens (Dillman, 2006, page 9; Dillman, 1991). The possible reason might depend on poor wording, wrong order of the questions or the characteristics of the surveyed person, such as incapability to provide correct answers or motivational factors (Dillman, 1991).

Nonresponse Error When there is a large amount of people who would not provide a response and their characteristics are different from the ones who responded, then it results to a nonresponse error (Dillman, 2006, page 9). Low responses are considered as a major problem, and many researches have focused on improving the response rates (Dillman, 1991).

2.2 Response Rate Influences

As mentioned in the previous section, one of the survey errors is the nonresponse error. Researchers have concerns regarding response rates, since responses coming from survey participants may be substantially different from those of non-respondents, which will result in a biased estimate of representation of the population (Bogen, 1996).

Low response rate was even considered a shortfall of the email methodology despite to its advantages (Bachmann et al., 1999). In table 2.1, there are nine studies, where both postal mail and email are compared side by side. Out of those nine studies, four of them showed a high response rate on the postal mail, three of them got a higher response on email and two studies did not show any significant differences. Parker's (1992) study of AT&T employees was the only study that was able to get an acceptable high response

rate by email. Schaefer and Dillman (1998) attributed this fact to the novelty of email and that sent emails were carefully examined instead of considered as a company junk email. Mavis and Brocato (1998) stated that studies cited by the others in support of email surveys, as also shown in table 2.1, was not able to compare email data collection with the more traditional methods, and their study design and analyses varied greatly. Sheehan and Hoy (2006) also focused the attention on many of these studies' small and homogeneous population; therefore, it may not be applicable to represent larger population groups' response tendencies.

Hence, researchers investigated on how to increase the response rates for email communications. Schaefer and Dillman (1998) concluded that even though the technology for email is quite different from the well-established postal mail surveying methods, the communication itself is considered to be similar to self-administrated questionnaires delivered by post. Hence, the techniques used in increasing response rates on postal mail can be applied to develop an email methodology. The following techniques indicated below are where the researchers focused on, to evaluate their effects on response rates.

2.2.1 Length

For many people, the total amount of time spent on conducting surveys is considered the biggest cost (Dillman et al., 2009, page 26). The study from Heberlein and Baumgartner (1978) also states that the length of the survey has a negative effect on mail survey response rates, where they stated that each additional question reduces responses by .05%. On the other hand, Bradburn (1978) suggested that the length of the survey is correlated with its importance; therefore, it will increase the efforts both on the researchers' and respondents' side, resulting to a higher response rate. Bogen (1996), in his literature review, concluded that the relationship between the interview length and the nonresponse rate is weak and inconsistent.

2.2.2 Multiple Contacts

The researchers found out that the number of attempts in contacting people increases the response rates (Heberlein and Baumgartner, 1978; Schaefer and Dillman, 1998). The sce-

narios for multiple contacts include pre-notification contact, which is a brief notice for the main request, and follow-up contacts, aimed for the people who did not respond upon the initial contact. Heberlein and Baumgertner (1978) showed that follow-up mailing has a mean return rate of 19.9% at the initial contact, and continued on with 11.9% and 10.0% for the second and third contacts, respectively (Heberlein and Baumgartner, 1978). Schaefer and Dillman (1998) also stated that the same conclusion applies for the multiple contacts for email in their literature research. According to this, the average response rate for email surveys with a single contact was 28.5% while 41% and 57% for two and more than two contacts, respectively (Schaefer and Dillman, 1998).

2.2.3 Personalization

Personalization has been addressed as an important factor in increasing response rates by many researchers (Dillman, 1991; Schaefer and Dillman, 1998). It builds a connection between the respondent and researcher, by making the respondent feel important and drawing the respondent from out of the group (Dillman et al., 2009, page 272). Dillman and Frey (1974) conducted a study to see the effects of personalization, where they reached half of a university's alumni sample via personalized cover letters, while the other half got impersonalized letters. The personalization treatment included personal salutations and real signatures affixed on the letters. They have achieved nearly 9% greater response rates for the personalized group. It is also stated that this type of personalization techniques can be also applied to emails (Schaefer and Dillman, 1998). In the next section, we will continue with the application of personalization in emails, and give the results of some studies.

2.3 Personalization of Emails

Studies on mail surveys showed that personalization helps increases the response rates (Dillman, 1991; Schaefer and Dillman, 1998). Personalization is also important for email communication since it builds a connection between the respondent and researcher as in the mail surveys studies, and make them feel more important and valued (Dillman et al., 2009, page 272). With this argument, Dillman et al. (2009), emphasized the social

exchange theory[2] of the personalization of the email.

On the other hand, Barron and Yechiam (2002) stressed on the socio-psychological phenomenon, the diffusion of responsibility, which is also an outcome of a volunteer's dilemma. With a volunteer's dilemma, one player is needed to volunteer in order to reach the outcome preferred by everyone else in the game. However, each person might be inclined on hoping that someone else will volunteer, resulting to a scenario of a higher instance of not volunteering, rather than volunteering. According to this, the greater the number of people in the group size, the lesser probability of volunteering will result, which will then produce the diffusion of responsibility effect. In order to experiment on the effect of diffusion of responsibility in the context of email requests, they sent several emails asking for help either to a single address, or to a list of five addresses. In the email body (see Appendix A), a fictitious graduate student asked a question to know if the university has a biology faculty, whose answer is actually a given to anyone familiar with the institute. The result of the study showed that the number of replies sent to a single email address per email got a 20% higher response rate than the number of replies sent to a group of email addresses per email. In addition to this, the study classified the given responses according to its level of helpfulness, and the rate of "very helpful" replies retrieved from the emails sent to a single address per instance was 187% higher compared to the responses retrieved from emails sent to a group of email addresses per instance.

Another outcome regarding the use of multiple email addresses in the "To" field resulted concerns from respondents in the study of Selm and Jankowski (2006). An introductory email including a link to a web-based questionnaire was sent to recipients to explore the opinions of elderly Internet users about an electronic political debate. One of the respondents raised his privacy and confidentiality concerns when the header of the email contained all the email addresses of all of the respondents, explicitly. His reaction was quoted in the study as in listing 2.1.

[2]Social exchange theory was considered as a frame of reference to other theories rather than a theory by itself. It implies a two-sided, mutually contingent and rewarding transactions or exchanges (Emerson, 1976).

"Well, it could be good (for you) to fill in this form, but I better not. Do you want to know why? 'All responses will be treated confidentially', but what do I see in the address column? I see all the email addresses of those you've sent this message to. Do you folks call that confidentiality !? I've decided not to participate in this 'carefully composed' study, although I do have an opinion on the subject matter."

Listing 2.1: A Respondent's Reaction Regarding Confidentiality (Selm and Jankowski, 2006)

Even though the authors believed that the person was just "skeptical" and his reaction displayed a "vivid skepticism". To this date, one of the biggest concerns involving the whole email medium is confidentiality, which might result into very embarrassing situations, including invasion of privacy involving anything, from doing research up to business perspectives. A very recent email message (See listing 2.2 for the excerpt) dropped to my email inbox verifies the importance of confidentiality.

Dear Valued Customer,

Earlier today the email seen bellow was inadvertently sent without utilizing 'Bcc' recipients .

Our sincerest apologies for any inconvenience this may have caused you.

Kind Regards

Listing 2.2: An Email Message Showing the Importance of Confidentiality

In another study by Heerwegh (2005), personalization is applied to salutations in emails. The randomly drawn 2,540 samples from the student database of Katholieke Universiteit Leuven, Belgium were separated into two equally sized groups. For the non-personalized group, the salutation of "Dear student" was used, while in the personalized group "Dear [First name] [Last name]" was used. The email content was an invitation

to a web survey which was about adolescent attitudes towards marriage and divorce. The result of the study showed that the personalization applied group got a 6.9% higher survey login rate than the non-personalized group. Therefore, they concluded that increased response rates were in line with the social exchange theory and the diffusion of responsibility theory.

In addition to the personalization of salutations on emails, Joinson and Reips (2007) stated the power of its combination with the power or status of the sender. In the study, groups of discussion panels composed of students from the Open University in UK were sent an email invitation to complete a survey. Panel members were assigned on one of the conditions, and salutations were modified to "Dear student", "Dear John Doe", and "Dear John". The sender power were manipulated on the first and last lines of the emails by assigning a neutral power, saying that "From <name> (Strategy, Planning, and Partnerships), The Open University" and a high power "From Professor <name>, Pro-vice chancellor (Strategy, Planning, and Partnerships), The Open University". The results showed that the highest response rate was achieved when a personalized invitation came from a high power source and the lowest when an impersonal one came from a neutral power source (See table 2.2). The possible reason for this was suggested that as personalized salutations increase one's sense of identifiability, its combination with a higher power audience increases, giving them a sense of being socially desirable, a strategic behavior.

Table 2.2: Power, salutation and response rates (raw and %) (Joinson and Reips, 2007)

	Dear Student	**Dear John Doe**	**Dear John**
Neutral power	143 (40.1)	158 (44.4)	166 (46.6)
High power	150 (42.1)	154 (43.3)	190 (53.4)

The aforementioned studies showed that different forms of personalization help increase the response rates on email communications. However, it has become very easy to add personalized information into emails thanks several third-party software. Dillman et al. (2009, page 237-238) stated that over-personalization using software tools can easily result to impersonal messages, similar to the example given below (See listing 2.3).

Dear Don Dillman,

I am writing to inform you and your wife Joye that the XYZ Company has created a new dog food that we are sure your Boston Terrier, Crickett, will find to be very tasty.

We would like to send a free sample to your home in Pullman, Washington.

Kind regards,
XYZ

Listing 2.3: A Sample for an Over-personalized Email (Dillman et al., 2009, page 237-238)

In this message, there is an overwhelmed personalization with the usage of person's wife, their dog's type and name, and their home address. Moreover, experienced email users can easily identify if a message is written by a person or if it is just a computer generated one by looking at the appearance of one's name on certain locations, and similar patterns for other information (Dillman et al., 2009, page 272). Therefore, it is difficult to have a correct amount and tone of personalization. The more daily interaction with digital devices there is, the more it will make true and authentic personalization rare, hence achieving such will make it more important and effective (Dillman et al., 2009, page 238).

2.4 Conculusion

In conclusion, the researchers conducted more mail survey studies than other survey methods to further investigate data collection (Dillman, 1991). Some of those studies tried to answer the question of nonresponse error, which has been considered as a major problem, compared to other survey errors as discussed in section 2.1.2. According to the mail survey studies, personalization has been addressed as an important factor in increasing response rates by many researchers in addition to other influences affecting response rates as identified in section 2.2. With the advance of world-wide Internet usage, many researchers started to consider email as a form of a data collection method, because of its cost and speed benefits compared to other data collection methods as discussed in

section 2.1.1. However, some studies showed that response rates on email surveys are lower than that of regular mail surveys despite of its advantages; in addition, it may pose as a burden to the researchers during the collation of responses since email communication does not emphasize on any structure, like in web forms or even respondents may come up with additional clarifying questions (Selm and Jankowski, 2006). Therefore, even if the technology for email is different from regular mail surveying methods, the researchers considered the response rate influences of regular mail surveys for email since the communication itself is the same in nature. In section 2.3, several studies applying different types of personalization were mentioned. Some of those studies modified the header of the emails to study the diffusion of responsibility. Other studies changed the salutations and signatures of the emails, which resulted to an increased response rates to the emails. On the other hand, those studies did not consider the increased awareness of their recipients to the possibility of computerized personalization techniques, which resulted in over-personalized emails. Also, none of the studies gave attention to the personal efforts of a researcher while extracting information from respondents' answers. This study will try to focus on the shortcomings of those studies as well, and provide a web application in attempt to overcome those issues.

In the next section, existing applications in the market, which leverages the email communication, will be evaluated. While some of them would just focus only on the email communication like email marketing applications, other applications like CRM and help desk applications helped this study on identifying useful features that can be deemed helpful in the area of personalized email communications.

3 Evaluation of Existing Applications

After building a foundation by giving out an overview about the related work on personalized mass email communication, this section will evaluate existing applications available in the market.

3.1 Application Categories and Their Relation with the Thesis

There are three different application categories related with this study that focuses on email communication either directly or indirectly. The following section will give a brief description of those categories, and their relation to this study:

3.1.1 Customer Relationship Management (CRM)

A CRM application helps manage customer relationships effectively, a topic studied both by the academia and industry in the recent years. Such applications play an important role in the marketing field, where organizations use a more customer oriented approach instead of a product or brand-oriented marketing strategies. Therefore, each customer's economic value is different to the company, and the organizations' customer relation strategies require adapting their customer offerings and communication strategy personalized, according to individual customers (Reinartz et al., 2004).

One of the reasons why this study considers on evaluating CRM applications is because of the communication aspect of a company with their clients. Another reason is that as mentioned on section 2.3, the adequate amount of personalization in emails is crucial on the response rates, and people's increased daily interactions with the digital world makes true and authentic personalization rarer. Achieving such level of personalization requires getting to know each recipient very well by considering not only the recent conversations,

but also earlier conversations. All the information that might be extracted from those conversations helps build a relationship with the respondents. Since a CRM application aims to keep track of each customer's history regarding a product or a brand, such data storage could be leveraged to add an adequate amount of personalized information to email conversations.

3.1.2 Help Desk

Another type of application that focuses on a company and its relationship with their clients is a help desk application. Its main purpose is to provide information and support related to a company's products and services offered to their customers. As a part of a knowledge acquisition, help desks support both sides of the communication in a way that customers or end users find the knowledge they need and the people who provide help by making the knowledge available and reusable (Halverson et al., 2004).

Reusing existing knowledge requires structuring of the captured knowledge. This is where it makes its connection to this study. A help desk application provides a workflow for both parties on developing an exchange of communication wherein a person who needs assistance describes his/her problem, while people who would provide help will then identify the solution to the problem by looking for similar earlier cases or by asking additional questions to clarify the initial problem. This also requires cooperation from assistants while providing help to a problem, at which one person might have a previous experience that can help guide the other assistants. As a result, a help desk application is similar to a mass email communication wherein a researcher initiates with an open-ended questionnaire, then extracts information from the coming replies, and organizes them according to the answers that he or she seeks for. In addition, respondents might also come up with additional questions to clarify things, where existing answers can easily be reused. Having such email conversations with large groups requires great effort from a researcher, so he might end up assigning tasks to distribute the efforts to other researchers in order to effectively deal with the demands of the large size of the group.

3.1.3 Email Marketing

Organizations and marketers use email on marketing for several reasons. Some of those purposes are for brand and customer loyalty building, acquiring or converting customers, advertising the brand or the product, solicit sales or donation, communicating for promotional offers and even educational purposes. At the end, these approaches can be grouped under the following categories (Eley and Tilley, 2009):

- **Educational Communication:** An educational message is given in the form of a newsletter, avoiding sale push, but it might still include some content encouraging recipients indirectly. For example, a free monthly newsletter which contains tips about digital photography, and photography accessories used in the tips might be linked to an online shopping website.
- **News and Updates:** Used to notify the customers about important updates or changes to a business. For an instance, the release of a new product, changes on contact details or major changes on a company's website information.
- **Direct Sales Messages:** Emails sent out by others consists of marketing ads, and clear messages on offers.
- **Housekeeping:** Emails such as subscriptions for confirmation messages or welcome emails. These messages are often to be system generated or automated messages. However, they can be used to promote messages as well as offering a discount code along with the registration of the confirmation email.

Since these categories consist of a communication with a large group of people, this study also evaluates existing tools available in the market for email marketing, including their technical aspects.

3.2 Methodology

The analysis examined two products from each of the categories – CRM, Help Desk, and Email Marketing. The selection of the products depends on several product comparison websites, including Toptenreviews.com[1], Softwareshortlist.com[2], as well as the sugges-

[1]http://{email-marketing-software-review, crm-software-review}.toptenreviews.com/

[2]http://www.softwareshortlist.com/crm/solutions/

tions of Stanford HCI group members[3]. In addition to those websites and suggestions, their demo or trial version availability was also considered, since some of the products actually require a certain fee before using them. After the products were shortlisted, the last filtering was done by getting their web traffic rankings from Compete.com[4], Alexa[5], and Google Trends[6]. Finally, the trial accounts of those applications were created, and a scenario was simulated to get the full insight from them.

3.3 Results

Evaluation of the products will be performed according to their respective categories. A brief description of the product will be presented, as a part of its evaluation. This description will mainly focus on the product's features, which is related to support email communication, as explained in section 3.1. Afterwards, each category will then be a conclusion, including a comparison matrix of the selected products.

3.3.1 CRM Applications

SugarCRM and Highrise are the two CRM applications that were analyzed in this study. Table 3.1 shows a summary of their features, and the following paragraphs will give a more in-depth exploration for these products.

[3]http://hci.stanford.edu/people/

[4]https://www.compete.com/

[5]http://www.alexa.com/

[6]http://www.google.com/trends/

Table 3.1: Comparison Matrix for CRM Applications

	SugarCRM	**Highrise**
Versions	On-premise and SaaS	SaaS
Pricing	$35 – $100 user/month, and a free community edition	$24 – $99/month, and a free plan with limitations
Task Management	Calendar based, no additional view	Individual module
Syncronization	Plugins are available for Outlook, Lotus Notes	Requires additional module installation
Email Client	Built-in, allows email marketing with variable insertions	No
Contact Importing	Via forwarding emails or plugins for Outlook, Lotus Notes	Outlook, Excel, vCard, or via forwarding emails
Mobile Support	Yes	No
Analytics	Marketing Analytics, sales forecasting and trends	No

SugarCRM SugerCRM comes in three different deployment versions: On-premise, Software as a Service (SaaS) and the free community edition. It has a clean User Interface (UI) with a single navigation menu. Its calendar view can be synchronized with Outlook's calendar or any other platform's, which supports iCalendar[7]. It has a built-in email management feature, as well as integrations with several platforms like Outlook and Gmail, or an Internet Message Access Protocol (IMAP) based email server. Users can archive emails in the SugarCRM application by adding a unique email address into the TO, Carbon Copy (CC) or Blind Carbon Copy (BCC) fields. This address can also be used to link an email recipients' information, including email attachments with SugarCRM by simply forwarding the emails. Therefore, it removes the additional effort on manually importing them into the SugarCRM application and reduces dependency on a platform. The SugarCRM also comes with a built-in email client. Even though its inbox view can

[7]iCalendar is the calendar data exchange standard (RFC 5545) having file extension of .ics, and it allows sending meeting requests or tasks via email.

only provide basic functions, its email creation view goes a little further in supporting email marketing by providing dynamic variables that can be embedded into an email's content that can be replaced with actual values available in the SugarCRM application. For example, a variable for "first name" will be replaced by a contact's actual first name while email is being sent (See figure 3.1).

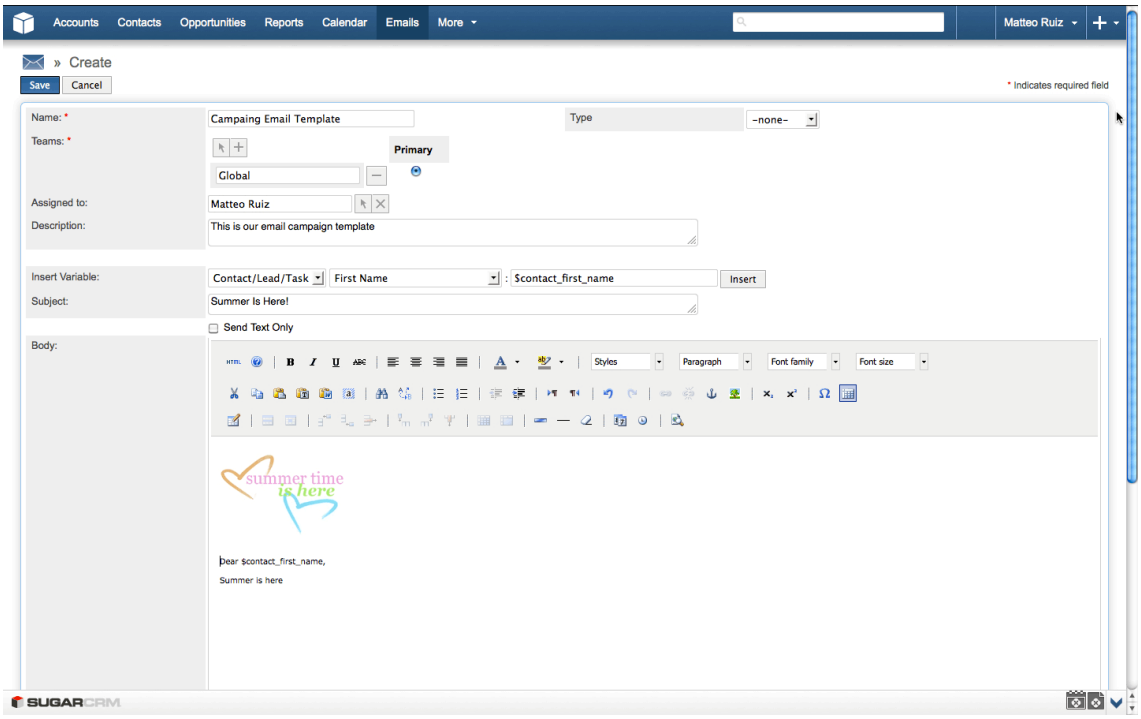


Figure 3.1: SugarCRM Email Composer with Embeded Variables (SugarCRM Inc., 2013)

Initiated email marketing can be monitored to track response rates, generated leads, and unsubscribed contacts. A marketing target lists can also be imported from third-party lists. The SugarCRM also let users save an email as a HyperText Markup Language (HTML) template, allowing the user to use it again within an email composer. Finally, it offers a mobile version, allowing users to easily access most of the application's features using their smartphones and tablet devices (SugarCRM Inc., 2013).

Highrise Another SaaS application available on the market is Highrise[8]. It offers several purchasing plans, with a 30-day free trial period. It has a simple UI like the SugarCRM, but it also has quick access buttons for adding a task or a contact. The task

[8]http://highrisehq.com/

management feature on the Highrise application makes it different from the SugarCRM, since unlike the latter, instead of a calendar view, it offers a task view, which can be synchronized with iCalendar as well. In addition, users can create tasks from emails by using one of the unique email addresses for several time slots provided by Highrise, and adding them into BCC, CC, or simply by forwarding an existing email, creating a task in Highrise. Contact information can be imported from Outlook or by uploading a vCard[9] file. It provides all the basic contact information fields, including social accounts; however, it does not offer custom field creation on those profiles. An email, including its attachments, can also be linked to a contact profile just by simply forwarding it to the provided unique email address. If a user does not exist in the Highrise contact database when an email from him or her was forwarded to link it, a contact profile is created using any available information in that email. Adding tags to contact profiles also makes it easier to organize contacts and browsing within them. However, the Highrise application does not offer an email composer to do an email marketing, as seen in the SugarCRM application. Therefore, users will have to depend on a different third-party application to do simple email campaigns. The provided activity view helps users on keeping track of their own or other users' recent actions within the Highrise application. Lastly, it offers options on customizing the look and feel of the application with the use of the system provided color schemes, depending on the user's preference (37signals LLC, 2013).

3.3.2 Help Desk Applications

The two help desk applications reviewed in this study are Zendesk and Kayako. Table C.1 provides a comparison matrix of their features, and the details are described in the following paragraphs.

[9] vCard is a file format standard for exchanging business contact information.

Table 3.2: Comparison Matrix for Help Desk Applications

	Zendesk	**Kayako**
Versions	SaaS	Software and SaaS
Pricing	$24 – $119 agent/month, with a limited free trial version	$29 – $49 user/month, with a limited free trial version
Channels	Website, email, phone, and social platforms	Website, email, and only the Fusion version supports phone
Macros	Yes, basic	Yes, advanced
Ticket Management	Groups and tags	Types, statuses, priorities, and tags
Mobile Support	Yes	No
Analytics	Yes	Yes

Zendesk Cloud-based customer service software Zendesk [10] provides a nice and clean UI. Zendesk has more than 30,000 businesses from a wide variety of industries. Zendesk offers one-on-one support through different communication channels including a company's website, email, phone, and social media platforms like Facebook and Twitter. Hence, support requests coming from those platforms can be turned in to a support ticket, and those support tickets can be grouped under categories, and further classification can be done via tags for each ticket. This feature also helps in finding related archived resolved tickets, so they can be reused for new tickets. Thanks to the automated process coming with macros, a combination of actions can be done with a one-click like setting status, priority, type of a ticket, and assign it to another person with a predefined comment for the ticket. A ticket can be merged with another one, or copied to the forum to make it available to the public, which helps creating a reliable knowledge base. Customer ticket histories and basic personal information are kept in the system. However, it does not allow adding additional fields on the customers' profiles. In addition to the desktop version, Zendesk has a mobile version for smartphones and tablet devices.

[10] http://www.zendesk.com

Therefore, the support team does not have to depend on the desktop platform, as long as they have it on their mobile devices as well. Lastly, the provided analytics view by reports gives an overview of customer satisfaction, and performance of the support team (Zendesk Inc., 2013b,a).

Kayako Kayako's[11] complete solution for customer support is named as Kayako Fusion. It comes as software and SaaS. Kayako has more than 30,000 clients within the last ten years. Unlike Zendesk, using its UI seemed to be a little more complicated than the latter, and does not have a social media integration, therefore support tickets are generated only over a company's website, or via email and phone. Tickets can have customized types, statuses, priorities, and tags. Similar to Zendesk, it also supports macros to assign tickets into a department, owner, type, priority, and provide canned responses for tickets with just a single click. Kayako also keeps basic information of customers, as long as they are registered in the system. Registered customers can also participate in building knowledge base in a forum-like environment by answering other customers' questions together with a support team. However, Kayako does not have a native app for smartphones and tablet devices like Zendesk. Finally, it has an analytic view to keep track of ticket reports, measuring customer satisfaction and support team's performance (Kayako Inc., 2013b,a).

[11] http://www.kayako.com/

3.3.3 Email Marketing Applications

MailChimp and Constant Contact are the chosen email marketing applications to be reviewed in this study. Table 3.3 shows an overview of their features on a side-by-side comparison to. Supporting details are provided in the following paragraphs.

Table 3.3: Comparison Matrix for Email Marketing Applications

	MailChimp	**Constant Contact**
Versions	SaaS	SaaS
Pricing	$10/month with max 500 subscribers – $240/month with max 50,000 subscribers. Pay as you go available	$15/month with max 500 subscribers – $75/month with max 10,000 subscribers.
Template Editor	Drag and drop including advanced photo editor	Drag and drop including basic photo editor
Recipients List	Conditional filtering	Grouping
Variable Support	Yes, advanced	No
Permissions	Admin, manager, author, and viewer account types	None
Mobile Support	Yes	No
Analytics	Yes	Yes

MailChimp MailChimp[12] comes as SaaS, and offers either a fixed monthly plan or a pay as you go plan. Along with its intuitive UI, it offers a drag-and-drop function on the email content creation. It supports email marketing processes ranging from designing the sign-up form so that users can add their desired fields and apply brandings on it to applying personalization on emails using dynamic variables. The recipients' list can also be filtered out according to several conditions like their campaign name, location or ratings, as assigned by the user. There are different user account types with different levels of privileges in accessing MailChimp. A person who has an "Admin" account is the only one capable of granting permissions to other users, as well as determining one's access

[12]http://mailchimp.com/

limitations on using MailChimp. This allows an efficient way for the distribution of mail marketing tasks. For an instance, while the assigned manager manages the recipients list, the author team can focus on the emails' content and design (The Rocket Science Group LLC, 2013e). To design an email content, users can either decide on picking an available template from a collection provided by the MailChimp application or to create their own HTML templates with its drag-and-drop editor (See figure 3.2).

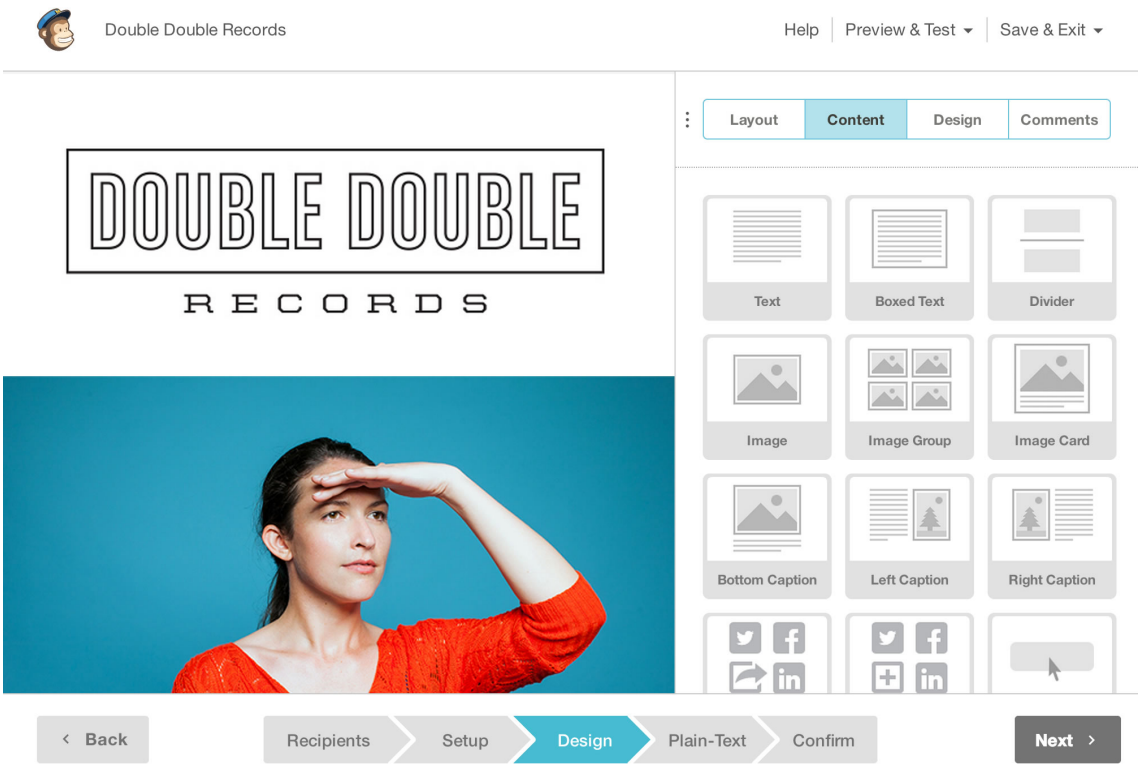


Figure 3.2: MailChimp Drag-and-Drop Content Editor (The Rocket Science Group LLC, 2013b)

The template editor has a photo editing feature, and authors can add comments to give a feedback on the content and design of the templates. Created templates can be previewed as if they are viewed on either a software email client, an web-based email client, or even on a mobile web browser (The Rocket Science Group LLC, 2013b). Similar to the SugarCRM application, MailChimp allows you to use dynamic variables – merge tags – in the email content. Therefore, sending out emails can be personalized with recipient-specific information. However, it provides more types of dynamic variables than SugarCRM, and

it is possible to add a conditional logic to them. For example, in listing 3.1, a custom discount message will be shown in the email depending on what state in the US the recipient is from.

```
*|IF:STATE=CA|*
    Save 20% on surf boards!
*|END:IF|*
*|IF:STATE=GA|*
    Save 20% on Mountain Bikes!
*|END:IF|*
*|IF:STATE=FL|*
    Save 40% on water skis!
*|END:IF|*
*|IF:STATE=CO|*
    Save 50% on ski gear
*|END:IF|*
```

Listing 3.1: MailChimp's Conditional Merge Tags (The Rocket Science Group LLC, 2013c)

MailChimp also offers an auto-response feature, based on a triggering event. These events can be a clickable link in the email, a specific date like birthday of a contact, or scheduled dates. Finally, an analytics dashboard is present, where users can track the amount of opened emails, or the click rates of the links in the emails (The Rocket Science Group LLC, 2013a,f,d).

Constant Contact Another email marketing SaaS is the Constant Contact Email Marketing[13], whose purchase plan depends on the number of contacts you have, but a free 60-day trial period is available upfront. It offers a drag-and-drop content creation on a clean UI like MailChimp's. It offers quite a number of templates to choose from and customize. Users can embed sign-up forms into their websites or Facebook accounts. The recipients list can also be imported from different sources like Microsoft Excel, Outlook, and Gmail. In addition, recipients can be grouped into sub-lists, which can also be merged into each other easily. An option of removing duplicate contacts from the lists or delete recipients who unsubscribed from the list is also available. Users can track how many opens, clicks, forwards, and social media platform shares are done for their email campaign. On the other hand, it does not offer any sophisticated email variables to be replaced by an actual content from the application (Constant Contact Inc., 2013b,a, 2011).

[13] http://www.constantcontact.com/email-marketing

3.4 Conclusion

In conclusion, the aforementioned applications in three different categories provide support for the email communication in several ways:

- Contact User profiles kept in the CRM applications can help a researcher to get to know their respondents better and to identify basic attributes like their name, gender, address, and phone numbers. However, fields are non-customizable, and are limited to the fields provided by the application.
- Importing contact information from other popular software, e.g. Outlook, can ease the time on creating recipient lists for email communication. SugarCRM and Highrise supports importing an email into the system by just simply forwarding it to a specific email address, making a researcher's life easier. Providing such flexibility will reduce the dependency on a platform, therefore, as the researchers continue on using the email clients they are familiar with, they can easily switch to another platform whenever it is necessary.
- Both CRM applications reviewed provides a module on creating tasks, which can be helpful on reminding researchers on the things they need to do, for their email campaign, depending on their priority level. It works like a task showing the next thing to do in an email campaign, as initiated by the researcher. For an instance, sending a reply to an email in which a respondent asked a question clarifying something about the initial campaign.
- Both CRM applications provide support on the archiving of emails by simply forwarding them to a provided unique email address, and linking those emails to the users' profile. This can be very helpful on looking for important conversations with respondents on an earlier time, to provide content or opinion on how to initiate upcoming conversations with those people. However, forwarding an email is an additional step, which requires additional effort and time.
- Reusability of earlier emails is important, so you do not have to write them again. As we have seen, the SugarCRM application also allows saving emails as a template for future use. However, no filtering mechanism or a similar function is available, but by just remembering the given name of the template helps the users find the corresponding template to use.
- Not all the time a researcher would be the first one to initiate the communication

via email. Sometimes, a number of emails might be sent to a researcher's email inbox for inquiries or whatnots. For example, students attending a course may ask questions or some clarifications regarding their homework. Given the scenario, similar or identical questions might be asked several times. A help desk application provides a ticketing system for customer related issues, which is also applicable to the above-mentioned scenario. Therefore, existing email replies can be reused for further recipients.

- Both help desk applications support tagging or grouping of incoming emails, which can be helpful in identifying conversations related to a specific situation where a researcher initiated more than one campaign. However, there was no available visual representation of the state of the communication of a support ticket, but just status labels like "resolved" or "assigned".
- Another feature of a help desk application is a support ticket that can be shared or assigned among the members of the support team, which helps decrease the time needed to answer those tickets. This can be also useful in a mass email communication to share the responsibility of replying or extracting information from incoming emails.
- The email marketing application MailChimp provides dynamic variables that let users add into their email content and its variable will be replaced with actual value. Such a feature can be helpful in a personalized mass email communication, where it is difficult and very time consuming on adding recipient specific personalized information into emails. However, there is no attached information on those variables showing users what part of the communication exchange they were extracted, and again they are separately created – an additional view where users are away from the actual emails where they can extract information.
- MailChimp also provides different types of permissions to leverage in an email marketing task. For an instance, as the author creates the email content, a viewer can just follow the reports to see what the success rate of an initiated campaign is. Such functionality can also be helpful in mass email communication, where some users can extract the information from the emails so others can easily reply to them.
- Both of the email marketing and help desk applications provide analytic reports to keep track of the success of a campaign or a support team. This is a very useful

function in a mass email communication, as well as to getting a quick overview of the current state of the communication.

As mentioned above, there are many useful features that can be helpful to ease a mass email communication. However, there is no one specific application capable of doing all of the mentioned features, or doing them in a way to support their main purposes, which are CRM, help desk, and email marketing.

In the next section, an initial prototype will be introduced to support the workflow of a personalized mass email communication.

4 The Initial Idea and the Prototype

In this chapter, a mass communication schema and possible exceptional cases will be introduced. Afterwards, the initial prototype will be reviewed and evaluated.

4.1 Mass Email Communication Concept

Whenever a researcher initiates a mass email communication, several unexpected outcomes cannot be predicted beforehand, and they can affect the flow of the communication. Respondents may come up with clarifications to an initial question, an email address considered current and active might not exist anymore, an auto responder might have been set since the respondent is not available at the moment are just some of the unexpected outcomes affecting the flow of the communication.

4.1.1 A Mass Email Communication Scenario

In section 1.1, an online learning platform scenario was used to illustrate the possible reasons explaining the need for an email data collection. Considering it once more, assuming that we are exploring our users in helping us improve our platform. Therefore, we started sending out emails to our registered users to obtain their permission for us to ask some questions. This initial email might be similar to listing 4.1.

Dear John,

you have recently attended the "Cryptography" and the "Natural Language Processing" courses. Do you mind if you answer some questions regarding these courses?

The questionnaire will take only 15 minutes, and it will help us a lot on improving our platform.

Kind regards,
your online course team.

Listing 4.1: Email Invitation to Questionnaire for Online Learning Platform

After the initial invitation email was sent to obtain their permission for the upcoming questionnaire, these are possible answers that we might get:

1. Yes, I would like to be involved.
2. No, I am quite busy.
3. I could be involved, but I am busy until the end of this month.
4. An automated message from the recipients, indicating that he or she is on vacation.

Figure 4.1 shows a simplified version of this scenario. The people who would accept our invitation will get a second email containing the questionnaire. As soon as they respond back with the answers to the questions, we will then send a "thank you" email for their participation. Those who opted not to accept the invitation will get a motivation email, encouraging them on joining next time. Next, we will classify the people who returned a conditional acceptance, under a "maybe" case. Again, there might be several reasons behind their conditional acceptance, such as they are willing to answer the questionnaire but they are busy that week. With this, we can email them back, wishing them luck if they are about to take an exam and sending them a reminder email for the questionnaire, depending on their availability. After the reminder email, we expect to get an email, confirming their participation for the questionnaire. Then, the email interaction will continue with the people who initially accepted invitation for our research. Cases of an invitation with an "auto-respond" reply is not clear, therefore the interaction will not go further, unless they send us a reply confirming otherwise in the near future. However, if the auto-respond email provides some information, we can set a reminder email for the date mentioned.

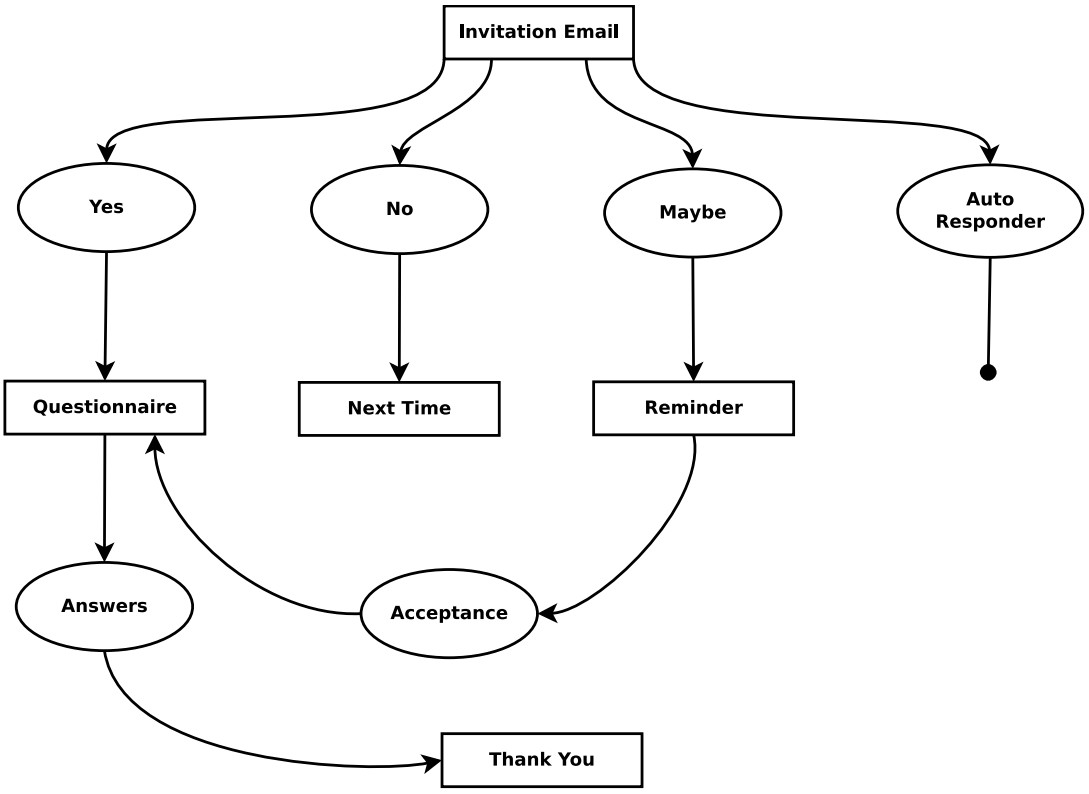


Figure 4.1: An Email Campaign Scenario

The above mentioned sample scenario can be handled manually with individual efforts for a small group of participants. These manual efforts will involve the following points:

- Figure out the courses for the participants to attended in our email list, and adds them to the email content for each person.
- Write a generic answer for those who accepted the invitation for the questionnaire, and write the questions.
- Write a generic motivation email to support those who did not accept the invitation.
- Go over the "maybe" cases, and record the date and other important information they mentioned involving their participation.
- Write a generic email to the "maybe" cases mention about the reminder that will be sent according to the time period they mentioned. This email may include a couple sentences encouraging them to give a feedback and share the reason why they are unable to accept the invitation for the questionnaire at the moment.
- Go over the "auto-respond" cases to figure out if they are related to a failed delivery, due to problematic email addresses, or if they are just automated vacation

responder, so we can set a reminder email later on.

- Send a thank you email to everyone who answered the questionnaire.

However, the larger the amount of participants there is, the harder it is to manually work on all of them, thus, the lesser personalization the emails will get. For example, the "maybe" cases might not get a personalized feedback, but instead just an email indication that they will be reminded on a later date. Moreover, this is just one of the possible cases. There are a lot more cases in real life situations, which might require the researchers to completely write a generic email by again ignoring the context personalizing the emails to be sent out according to each recipient, since manual efforts will not be enough to write custom emails for those cases. For extreme cases, we might even consider the need to terminate the communication forever, or results might give the side effect of a diffusion of responsibility theory as mentioned in section 2.3.

In figure 4.2, it is shown that the personalization of emails is directly proportional to the effort of composing those emails. The shaded part of the graph represents the situation, when the researchers used an application to create a mass email communication, while the area marked with a star represents the gold standard, which is really hard to reach with the available applications.

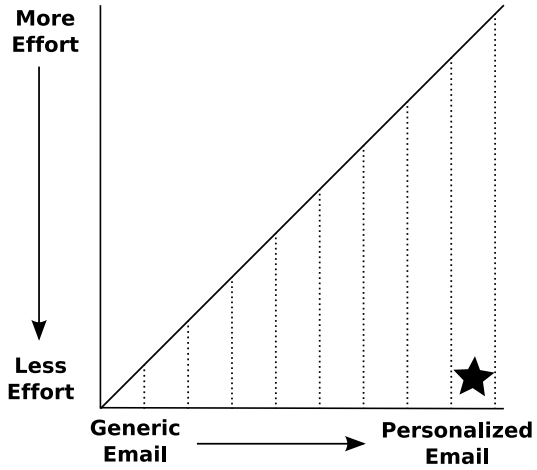


Figure 4.2: Effort vs Personalization on Composing Email

In the next section, we will see how we can help researchers to reduce the efforts while increasing the personalization of emails.

4.1.2 Possible Improvements to Reduce Efforts

After a mass email communication scenario was described in the previous section, we will see the possible improvements presented on this section to reduce the efforts on the researchers' side, while increasing the personalization of the content on some degree.

Contact Information The first step would be to import contacts and their relevant information. These will include some of the basic information we asked when they filled-up the registration form to our learning platform e.g. first and last name, email address and gender. Therefore, we can easily export these information into a different, easily-accessible file formats like Microsoft Excel or CSV[1] to make it ready for use to initiate our mass email conversation. The application that was discussed in section 3.3 offers several options in importing data onto the mentioned systems. Therefore, the researchers do not need to manually enter those information to the system.

Dynamic Variables The email invitation included user specific data, like the name of the recipient at the salutation and the courses he or she attended. As it is hard to write information manually into all of the emails, we can apply dynamic variables insertion approach that is used by both email marketing applications (MailChimp and Constant Contact) and one of the CRM application (SugarCRM), as analyzed in section 3.3. However, it is identified that SugarCRM has the limitation of allowing users to record only a limited amount of contact profile fields. Therefore, this can only help us apply the dynamic generation of the salutations in the emails, and not the course taken by the recipients. On the other hand, both Constant Contact and MailChimp support the feature of adding custom fields into a contact profile, and dynamically use them in the email's content. This can also help on removing the manual effort applied for the creation of the initial invitation email with the use of an email template with dynamic fields to be replaced by the application while sending them out. These dynamic fields can also add some degree of personalization into the emails.

[1] CSV stands for comma separated values, and is a plain-text file to store tabular data to often exchange data between disparate applications (Repici, 2010).

Information Extraction As we begin getting answers from the respondents, we might need to extract some information from those emails. For example, this could be a date that we need to return an email to the recipients with a reminder as discussed on section 4.1.1. Storing those information as we store contact information might allow us to easily access them again, and gives an opportunity to reuse them in the email content as a dynamic variable.

Reusability of Existing Emails After the initial invitation email, there will be many instances calling the need on reusing different versions of an email for further states of conversations. We mentioned the term "template" as an email with some dynamic variables in it. Later on, these variables will get their actual values, and it will be the actual message that we want to send. From here on, we will add another meaning to the term "template", to emphasize a message's reusability. Let us consider the "maybe" cases mentioned in section 4.1.1 to represent respondents who were too busy to participate on answering the questionnaire at that moment; but, maybe it is possible for them to participate on a later time (See figure 4.1). For instances like this, our recipients has to indicate a possible date for them to participate or provide an unclear answer about their availability. It is definite that we cannot ignore them if they provide us a date, and we should send them a reminder before that date to participate in the questionnaire in a generic email. This might give them a hint that the messages we send are actually automated, and it may result to some drawbacks that we discussed in chapter 2. However, if we create two different email templates; one with a dynamic variable in it to mention when we will send a reminder, and the other one with a more generic content, including a time far enough to encourage them to participate in the questionnaire. Therefore, the next time we encounter a "maybe" case, we have two ready-to-use email templates to send them.

Visualization of Conversation State At this point, we have begun on creating different email templates for different conditions, to ease a researcher's life by allowing reusability on the replies. However, it is also difficult to pick out a template since a flow of a certain conversation is different from another one, one way or the other, and we have different email templates for different conditions. This will result to an additional effort in choosing the appropriate template. Besides, we might not need some of the templates after conversation's state gets further, and just use the latest ones more frequently

compared to the the older ones. Therefore, we need an overview to show us what the current state of the conversation is, as well as to allow us to pick the proper template by leveraging this view.

Figure 4.3 represents a view in a directory tree structure for this purpose. All the possible email messages that we have written are collected in a hierarchical way, starting with the first invitation email at the root node. After the researcher began on obtaining messages from respondents, all of the possible answers we can write are at the second level of the tree, right after the invitation email node. Whenever we get an answer for the questionnaire, we would write an email to say thank you. This is also represented in the third level of the tree, right after the questionnaire node.

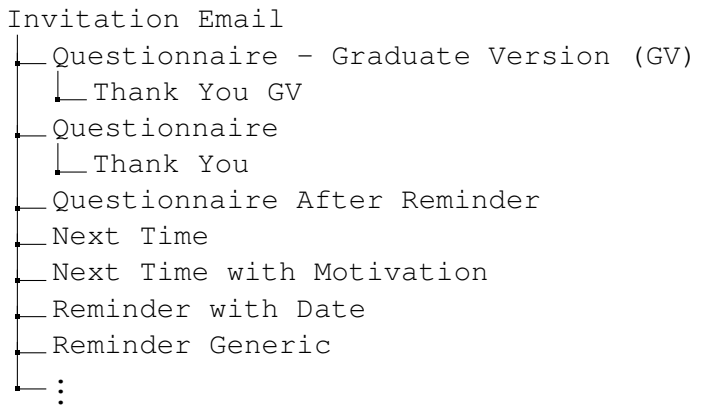


Figure 4.3: A Sample Conversation State and Template View

A view shown in figure 4.3 can also help the researchers pick out a template to reuse according to the email of the respondents. Therefore, they can easily see the previous answers they gave, and the level of the tree provides an idea on the overall state of the conversation. Considering existing applications in the market on section 3.3, none of them provides such conversation view and available templates had to be chosen from a mixed list, which includes messages belonging to other conversations as well.

Finally, after reviewing these initial information gathered to figure out how can we ease the efforts of the researchers while keeping the personalization in the emails high, the next section will discuss how this initial idea was applied into an application.

4.2 The Prototype

A software prototype provides developers a better understanding of the requirements, what could be developed with the existing technology, and by allowing client interaction, it demonstrates to them what is functionally feasible, and revise with their imagination accordingly, which results in better inputs from them and more forward looking systems (Bernstein, 1996). In this section, we will analyze the requirements, and the actual prototype.

4.2.1 Requirements Analysis

In section 4.1.2, we initially considered some features on improving a researcher's personalized mass email communication. The following points will put those findings into a summary to describe them as requirements of the prototype:

1. Contact information can come from different sources. Assuming we have at least the recipients' first names, last names, and email addresses, the application must store these information under a contact list permanently.
2. The application user can use variables in the email message for a recipient's first name and last name. Therefore, the application must allow a user to add different dynamic variables into an email's content to personalize the salutation of the emails.
3. The application user can extract information from an email message into a KVP[2] while reading responses. Therefore, the application must offer an option on obtaining KVPs from users. At the end, each contact's record will have enough information saved in KVP to describe all details we know about them.
4. The application must allow users to use the extracted KVPs as dynamic variables in the email content. This will allow us to personalize the email with the information specific to each recipient.
5. The user can reuse a sent email as a template to reply to respondents or initiate

[2] A key-value pair (KVP) represents two conntected data items, where a key is a unique identifier for some data items, and the value can be again a data item or a pointer to the location of data (Rouse, 2008). In this context, this thesis uses KVP in the context of extracted information as values, and a unique key to represent its class. For example, "first name" can be a key, and each recipients first names will be a value for the "first name" key.

a conversation for a group of people. Therefore, the application must offer an overview for picking out templates from a list.

6. The application must show the whole flow of the conversation. Therefore, the template list can be showed in a directory tree structure as in figure 4.3. Each node must represent the templates the user created, and let the user pick out a custom name for the given template.
7. The application must offer a template selection and conversation flow representation under a specific single view. Therefore, the user should be allowed to pick which templates to reuse by looking at the nodes in the tree to figure out the current flow of the conversation.

The next section will introduce the implemented prototype to satisfy the initial idea of a personalized mass email communication system and analyze how the aforementioned requirements are satisfied – or not – during the evaluation of the prototype.

4.2.2 System

The initial system was built on top of an existing project named as EmailValet[3] also known as GmailValet in the Human-Computer Interaction (HCI) group[4] at the Stanford University. EmailValet is a SaaS application that combines an email client with a task manager. Remote assistants from an expert crowdsourcing marketplace extract tasks from incoming emails, and annotate them in a task list separated from the inbox (Kokkalis et al., 2013).

Having a ready-to-use email client to implement the first version of the personalized mass email communication system helped us to save time and effort, and make us focus more on the initial requirements, as discussed in section 4.2.1. From the technology point of view, we were also limited by EmailValet's existing architecture, which was developed in a Web framework Ruby on Rails (RoR)[5], a JavaScript framework jQuery[6], and

[3]https://www.gmailvalet.com/

[4]https://hci.stanford.edu/

[5]http://rubyonrails.org/

[6]http://jquery.com/

a Cascading Style Sheets (CSS) framework Twitter Bootstrap[7].

EmailValet requires a Gmail account to register and sign in[8] to the system, therefore the application's users have to grant access to EmailValet before using it. Figure 4.4 shows EmailValet's inbox view after a user signed in and browse to a campaign[9]. On the left hand side, there is the email list pane, and a selected email can be read on the reading pane on the right.

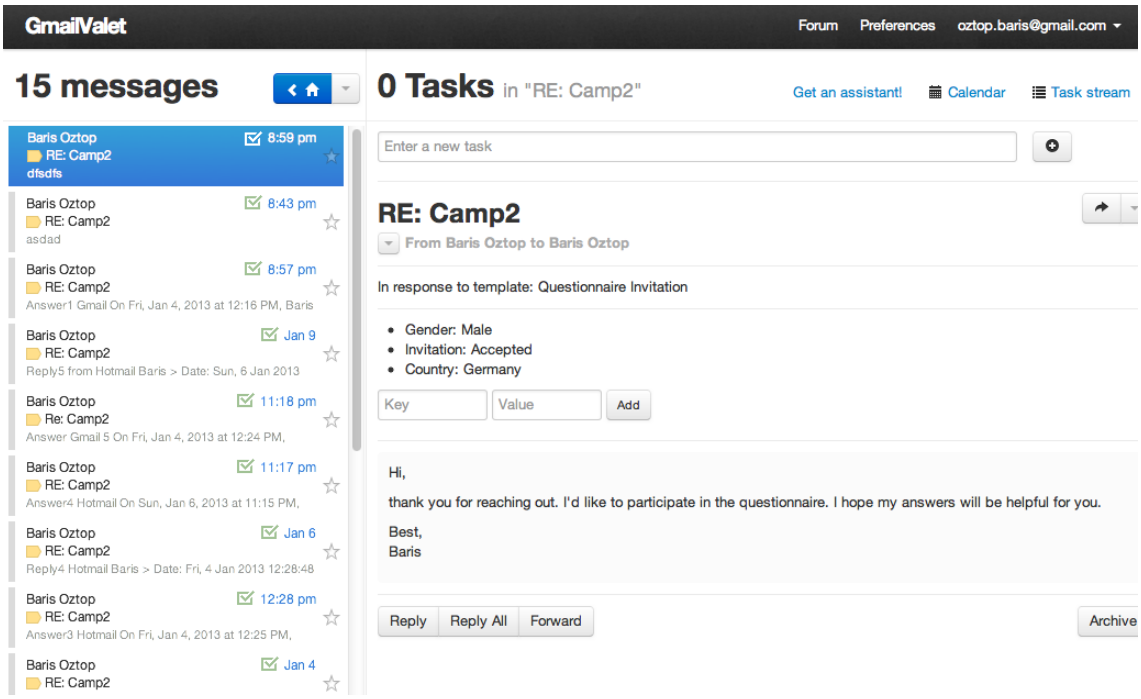


Figure 4.4: EmailValet Inbox View and Message Reading Pane

Starting a campaign is done exactly the same way as in composing an email. However, the corresponding view has two additional input fields to get the campaign's and the template's names from the user. A campaign name will help on identifying groups of conversations from other campaigns, or from regular email conversations in the inbox. A template name will help us identify which respondent's answers correspond to which emails that we sent. For example, in figure 4.4, a recipient's answer is the response to

[7] http://getbootstrap.com/

[8] EmailValet uses an open protocol named OAuth to provide credentials for authentication with Gmail. OAuth's community webpage is at http://oauth.net/

[9] In this study, the word campaign is used as a shorhand for a mass email communication campaign, wherein a researcher initiates a conversation with groups of people related to one specific topic.

the questionnaire invitation that we sent earlier. A template name will also help us represent the template in a tree structure to pick from and reuse, and show the state of the conversation in the same tree structure.

The researchers can add their recipients list into the "Merge"[10] input field as shown in figure 4.5, which corresponds to the "To" field in a regular email client. However, the format should be similar to the shown figure – first name, last name, and email address written in angle brackets. Hence, those fields can be used dynamically in the salutation of the emails by writing one of the variables of *{{first_name}}* and *{{last_name}}*. Once the recipients list is entered, and the email is sent, these contacts will be recorded on Gmail's contacts book.

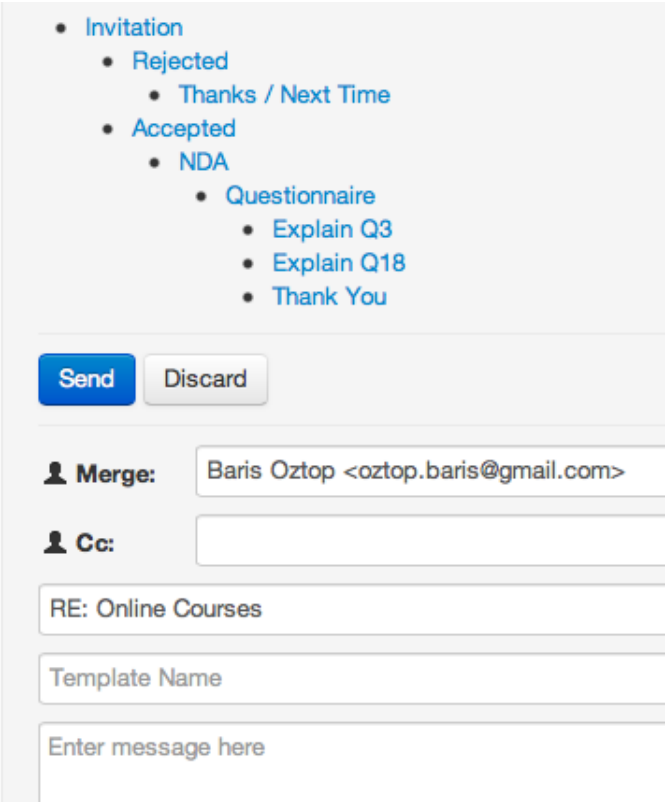


Figure 4.5: Email Template Selection to Reuse Earlier Replies

[10]The word "merge" comes from the term "mail merge", which means a procedure to enable to combine a document with data files consisting a list of names. Therefore, the copies of the document will be different for each person it is sent to (Collins English Dictionary, 2013).

Figure 4.5 also shows the template tree used to represent the state of the conversation, and allow users to select a template from. The indentation at the nodes of the tree helps on identifying which templates are used after which one and therefore it gives an idea about the current state and flow of the conversation. According to this sample scenario in the figure, right after we sent the invitation email and started on getting responses, there were two possible answers to give the recipients. For those who rejected the invitation, we can select the "Rejected" node from the tree to reuse an answer we gave before. This can be a motivation email as discussed in section 4.1.2. On the same note, the people who accepted the invitation would receive a Non-Disclosure Agreement (NDA), and afterwards the questionnaire as the conversation with them carries on. After sending the questionnaire, there were cases where some respondents sent an email clarifying some questions found in the questionnaire. According to figure 4.5, these cases are related to the questions number 3 and 18, and we explained those questions, to clear off the confusion or issues they are having. Again, since the system allow us to reuse our previous answers as a template; we do not need to rewrite our explanation for those questions not clear for future queries. On the other hand, if the researchers wish to create a new template, he or she can simply do so by adding the name of the template into the corresponding input field, and send the email. The next time the researcher fires up the application, he or she will find it in the template tree under the corresponding level of the node. Also, it is possible to select a template, made some modifications on it and save it as a new template. Therefore, a slightly different version of the templates can be reused during the communication as well.

Finally, there is an option to add KVPs while reading the respondents' answers. In the reading pane, two input fields are present, allowing users to enter a key and its value corresponding to the extracted information from those emails while reading them. As shown on figure 4.6, as a user adds a new KVP, they can also see the existing ones added earlier to that email message.

In the next section, we will address some implementation details about EmailValet's mass email communication modifications.

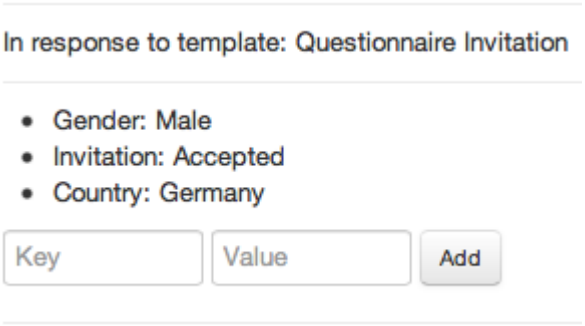


Figure 4.6: Extracted Key-Value Pairs and Input Fields to Add New Ones

4.2.3 Architecture

Even though, we have a ready-to-use email client with EmailValet, there were quite many changes to reflect the requirements that were mentioned in section 4.2.1. Since EmailValet dealt with single email messages, and not a group of them as an email client, there were also some modifications needed to be done in order to make the prototype understand whether a message belongs to the campaign or not. In this section, we will briefly see how the changes for the prototype fit into the existing EmailValet's design, and make a technology overview to fully understand how a mass email communication is done.

Identification of a Campaign Message

An email message consists of a header and body sections as defined by RFC 5322. While a message's header keeps structured information such as "From", "To", "CC", Subject, Date, and other information with a special syntax, a message's body contains the content, which is an unstructured text and it is optional to have (Resnick, 2008). An example email message with its header fields and body can be seen in Appendix B.

One of the fields in an email message's header is called Message-ID[11], which is a unique message identifier, and set by email clients. The value of a Message-ID is used in another header field, named "In-Reply-To", when a message is a reply to that message.

[11]The message identifier, Message-ID, is enclosed in the angle bracket characters, "<" and ">", an its syntax only permits the dot-atom-text form, which is `1*atext *("." 1*atext)`, on the left-hand side of the "@" and a domain name is recommended for the right-hand side of the "@" (Resnick, 2008). For example: `<CAF2E4bfH4+GAYHcJFZJ6dTJJ+pux4mTjff2neCS_VR_zVCUY9g@mail.gmail.com>`

EmailValet had already the data model to store an email message's related data when it is synchronized[12] with Gmail. However, creating a mass email campaign requires to identify if the synchronized email belongs to a mass email campaign created or if it is an answer to a campaign message that was sent before. Because EmailValet's initial design was able to fetch emails from Gmail, then assigning of different properties should follow. The emails composed by EmailValet comes to an existence in the EmailValet database after recipients send a reply to them, since EmailValet fetches them from Gmail's inbox as part of the thread of messages. Therefore, Message-ID and In-Reply-To fields of an email message are leveraged to keep track of a campaign message by setting a Message-ID into it, and store it in a "Campaign" data model as illustrated on 4.7. With this, we were able to identify the type of messages during the synchronization with Gmail, whether the message is an initial campaign message, or if it is an answer to a campaign message that was created earlier. Before this, setting a Message-ID was done by EmailValet's "mail" gem[13] during its execution time without allowing to get its value.

[12] Fetching the emails from Gmail is done via the IMAP extension that Gmail provides. IMAP is a Internet message access standard defined by RFC3501. It allows a client to access and manipulate mail messages on server (Crispin, 2003). On the other hand, EmailValet uses Simple Mail Transfer Protocol (SMTP) protocol to connect to the Google's SMTP servers to send a composed email in EmailValet. SMTP is a mail transport and delivery protocol defined by RFC5321 (Klensin, 2008).

[13] A Gem or a RubyGem is a software package, containing a packaged Ruby application or library (RubyGems Guides, 2013).

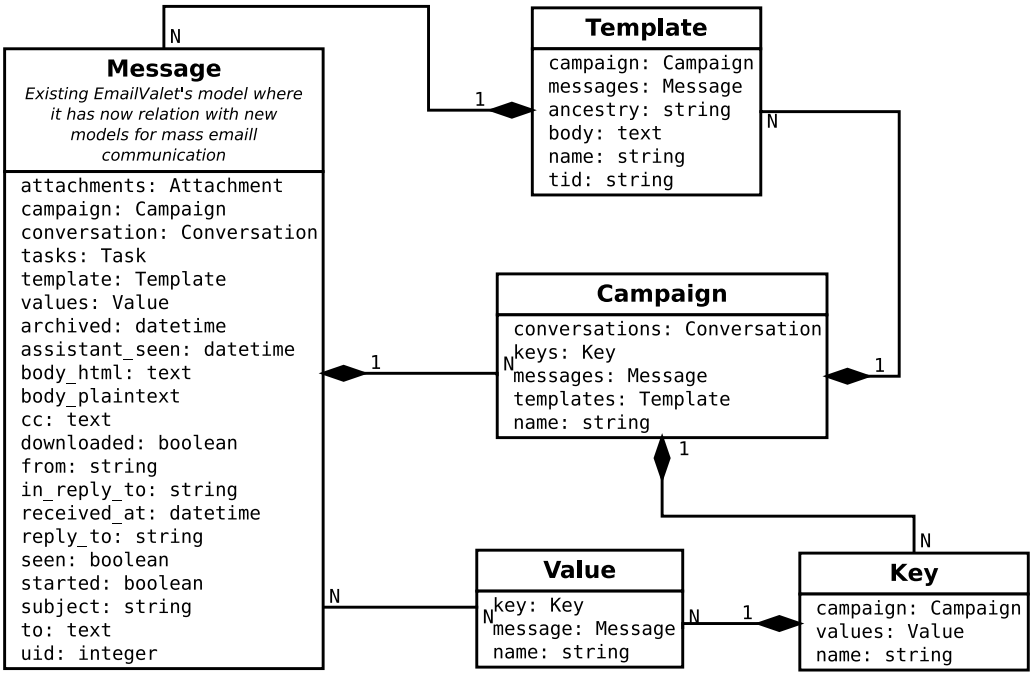


Figure 4.7: Model Dependency of the New Features of EmailValet

Template and Email Message Relation

Each campaign message composed in EmaiValet is actually a template with dynamic variables in it. Therefore, the campaign messages fetched from Gmail corresponds to a template in EmailValet. This interaction can be seen in figure 4.7. As mentioned in the previous section, there is an available tree view when a user replies to a respondent's answer. This tree view's purpose is to let users choose earlier templates to reuse, and give an overview of the latest state of the conversation. The hierarchy between the nodes of the tree is a nested set model, in which each node has a left and a right integer values. The left and right values in a node represent a set of child nodes. The same concept applies for the each child nodes as well; hence, each child node has a left and right value that represents another set of children, as a part of outer nested set. Figure 4.8 depicts a tree in a nested set model.

Nested set model helps us on finding a node's children faster than a usual adjacency set model, where each node only keeps information about its parent node, which is the ID (Celko, 2004, pages 45–47). For example, in order to find all the nodes of the root node

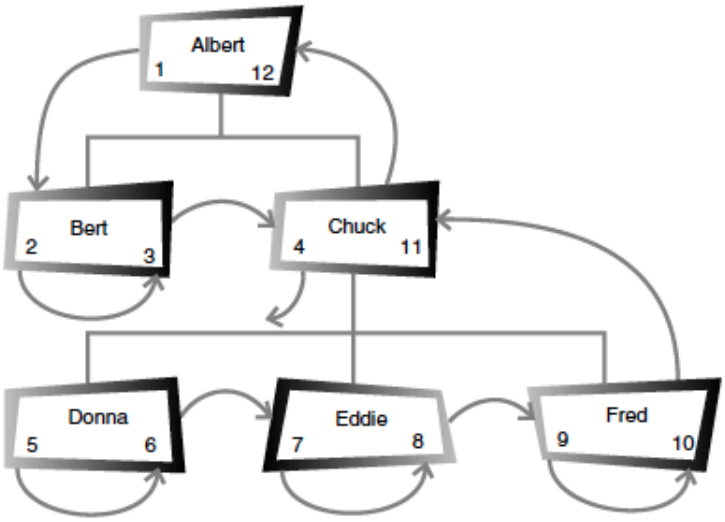


Figure 4.8: A Nested Set Model Tree Hiererchy (Celko, 2004, page 46)

named "Albert" in figure 4.8, we just need to find all nodes whose left value is greater than 1, and right value is less than 12. In the adjacency list, however, we first need to find all the nodes whose parent IDs are equal to "Albert's", then for those nodes' IDs, we need to find other nodes whose parent IDs are equal to them, and it continues on until all nodes are finished in the tree in the same way.

The template data model in 4.7 keeps those left and right number information in one column named "ancestry", which is populated by a gem named Ancestry[14].

KVPs Relation with a Campaign and a Message

As shown in figure 4.7, each email message can have more than one value corresponding to a key of KVP. This allows us to identify when a KVP is added to a message that belongs to a respondent. Keys are organized according to campaigns; therefore, same key's value can be used in different campaigns for different purposes. This is mainly because a word chosen as a key can have different meanings, depending on context.

After getting an overview of the technical implementation of the prototype, the next section will focus on the evaluation of the prototype by pointing out the missing features of initial requirements due to the limitations on the existing email client, and what new

[14]https://github.com/stefankroes/ancestry

ideas were emerged after the informal user test within the Stanford University.

4.3 Evaluation of the Prototype

In this section, we will analyze the limitations of the prototype, despite the fact that it was able to fulfill most of the requirements mentioned in section 4.2.1. Some of these limitations are related with the development of a prototype from an existing application, whose main purpose is different than that of a regular email client. One of the reasons for making a prototype is to let developers and users conceptualize different requirements or review the existing ones. In here is where we will also mention new and different requirements we deducted as we went through the whole process, as well as after the informal user test within the Stanford HCI group and with an organization at Stanford University conducting mass email communication regularly to reach their community. As an informal user test, we asked the participants to use the prototype instead of their regular applications or methods to conduct a mass email communication, and collected the feedback.

4.3.1 Limitation of the Existing Application

As mentioned, using EmailValet as an email client on developing a prototype helped us to save time and effort. However, there were some limitations present with it since it focuses on crowd-assisted task manager actions, rather than a simple email client.

Number of Message Limitation One of the problems encountered with EmailValet was the limit on the amount of messages synchronized from Gmail. It is limited to the latest 100 messages, to maintain EmailValet's privacy and accountability (Kokkalis et al., 2013). However, a mass email campaign will quickly reach that given limit, which may ignore older emails.

Different Inbox Purpose EmailValet offers an email client with a task manager feature, as a replacement to everyone's regular email client. However, a mass email communication is not something a person needs to resort as frequent as a personal email communication. Therefore, a mass email communication application should be separated from

a person's regular inbox, since it does not need to show all their emails, except those related to an email campaign. Combining both inboxes also resulted to some drawbacks on the performance while it was synchronizing with Gmail to fetch new emails. It is because EmailValet was also performing some additional steps during the synchronization on the side, such as checking and comparing message headers for changes in the earlier messages up to a date.

4.3.2 Limitation of the Prototype

KVPs Relation In the prototype design, each value of a key belongs to an email message of a respondent. However, we need to see that all KVPs are aggregated in a certain way when we browse a set of messages belonging to a specific person. Therefore, assigning KVPs to an email message actually means assigning them to a recipient in a mass email campaign. This requires keeping each individual recipient in a separate data model, and relating it with the existing EmailValet's design. Since we have already identified the limitations of EmailValet in the previous section, the decision was to add those enhancements into a new application.

Getting Contact List into the System During the informal user test at the Stanford HCI group, we realized that the people who performs such mass email communication import their recipient list from a spreadsheet, or a similar data file. Since, the prototype requires them to input a recipient list in a special syntax as mentioned in section 4.2.2, this requires the users to change their existing contact list in a format that the prototype will be able to parse with. This reduces flexibility and adds additional effort on the researchers, whose contact list is already in a data file. With this said, it is a feature considered to be implemented on a different project, separate from EmailValet.

4.4 Conclusion

In this section, a mass email communication concept was introduced with an illustrative scenario. We pointed out how it could be possible to reduce the efforts on the researchers' side while increasing the personalization of the emails at the same time. Later on, we converted those ideas into software requirements for the prospective prototype, where

we introduced its features and architecture. Finally, the informal user test helped us to evaluate the prototype to find out its limitations and to emerge new requirements, which made the basis to the next chapter, where the mentioned limitations are removed and the new requirements are reflected during the final application's iterative design cycles.

5 The Final Solution

In this chapter, a final design of a mass communication system named Myriad will be introduced. To begin with, improvements on the requirements will be compared to those defined for the prototype from the previous chapter. Later on, the actual product features, its architecture, and the benefits of the final solution will be discussed in the conclusion.

5.1 The Improved Requirements

Seeing the prototype in action made us realize the need to revise the requirements and bring new ones. Some of those requirements are shaped according to the feedback we got during our research at the Stanford HCI group, as well as from a Stanford organization who conduct mass email communication regularly in reaching their community. The following section will discuss those new features for the final product.

5.1.1 Assistant Support

As described in section 4.1.1, the standard we want to achieve within a mass email communication is the most adequately personalized emails for every individual recipient, with a minimum effort on the researcher's side. The initial idea, as well as the prototype, included several features for this purpose, as discussed in chapter 4. However, considering the gold standard we would like to achieve as shown in figure 4.2, the prototype manages to still leave an amount of effort on the researcher's side in accomplishing a successful mass email campaign.

In order to reduce the effort as much as possible during a mass email communication, the addition of an assistants' involvement was considered. Therefore, a primary researcher will be able to share tasks with permitted assistants. These might be tasks such as extracting information from the incoming answers, proofreading the primary researcher's

replies before sending them, or even writing replies to those answers. Therefore, the primary researcher will only need to interact with the flow of a mass email campaign when the situation calls for it. However, the system would still need to provide necessary features that will be seen in the next sections in support for the workflow in a mass email campaign, hence, assistants will only need to interact with this workflow in letting the email campaign carry on by providing answers with using the provided email templates and extract information as KVPs.

5.1.2 Dynamic Variables and KVPs

We introduced the use of dynamic variables at the initial prototype; however, it was only limited to the salutation of the email. As described in section 3.3.3, email marketing applications support this feature. And because it makes the personalization of emails easier, the final product will also include this feature. As a result, application users can create KVPs and use the said keys in the content of an email message to be replaced dynamically by its value, according to the recipient. Therefore, the extracted information from emails will not just help us gather information in an organized way, but also in personalizing the emails.

As a result, instead of storing the KVPs in the system according to their respective responses, the system should store them according to the recipients itself. This is where the KVP-idea differs from the prototype. With this, we now have profiles of contacts for a campaign having all KVPs of a recipient visible during the whole state of the conversation. However, the system should offer an option to hide individual KVPs to avoid cluttering the view, and make the actively used KVP list the default view.

Importing KVPs can be done in several ways. One option is that the system should be able to synchronize with an online spreadsheet, e.g. Google Spreadsheet, in obtaining the KVPs at the beginning of a campaign. This is a convenient way for researchers, since they are already familiar with spreadsheet environments. Other options should include a campaign-wide view and a contact specific view in the system. Also, it should have an editable keys and values in a campaign.

5.1.3 Importing and Exporting Contacts and Their Information

In the prototype, the application user has to enter all the basic information of the recipients such as first name, last name, and email address into the system manually. However, as they perform this, they use a spreadsheet and copy the contact information from there. This was also done for their regular email client used for email campaigns. The applications that were reviewed in section 3.3 all had an option to import data from spreadsheets to ease the process. Therefore, the system should also offer an option similar to this in importing contact information.

However, importing should not only be limited to the basic information of the recipients. Since we already mentioned about importing KVPs from a spreadsheet in the previous section, the system should also be able to detect and import contact information and KVPs related to a specific recipient, granting that they are available in the provided spreadsheet.

The system should provide a bi-directional synchronization, and not just to import data from the spreadsheet. Therefore, the system should provide an option to export contact information and their created KVPs from the system to the spreadsheet as well. This gives a reporting functionality to the application users, where they can see all of their recipients, and the extracted information from a specific campaign, all in one view.

5.1.4 Interoperability with Other Email Clients

Even though we provided a new system for the users in initiating their mass email communication, there might be other cases wherein a mass email communication was initiated using a regular email client, therefore the system is not aware such campaign, since it was not created with it. The application should be able to provide an option in importing email messages created with a different email client into the system by recognizing specific messages, as annotated by the user.

Allowing the system to import email conversations from other email clients reduces the

need for dependency on the application, and while researchers continue on using their own email clients, assigned assistants can take care of those emails, as imported by the researchers. We saw the same import feature in CRM applications present in section 3.3.1 as well, where a user can easily forward an email to the CRM application's provided unique email addresses, and the system then takes care of assigning the imported emails to the corresponding recipients in the system.

5.1.5 Automated Decision-Making and Notifications

Even though the involvement of assistants helps make the primary researcher's life easier, the system still needs to provide an automated approach in answering emails whose statuses are clear in terms of its mass email communication flow. Therefore, a rule based decision-making mechanism should be used, at which a user sets the value of the keys from KVPs, and it triggers the action of sending emails to qualified respondents whose KVPs satisfies the provided condition.

Since the only purpose of the application is managing mass email communication and each campaign results to a great amount of messages in the inbox, the system should be able to provide notifications regarding what should be done next for each recipient. Labels should be added to email conversations to indicate whose turn is the next in the communication. For an instance, by proving a label saying "You need to reply" to the application user, the state of the communication awaits for an ample action from a researcher, or an assistant. In the same way, unread conversations and conversations waiting for an answer from the recipients should also be annotated in the appropriately. The system should also provide email notifications to the assistants' email addresses, to notify them whenever there is an action waiting to be taken care of.

5.2 Final System

In this section, we will see how the revised and improved requirements from the previous section reflected on the final product, named "Myriad".

5.2.1 Log-In and Campaigns Overview

Similar to the prototype featured in section 4.2.2, Myriad also requires a Gmail account to work with. The reason behind this is not just the popularity of Gmail, but also given the fact that Stanford University uses Google Apps by default, university wide. Therefore, each member of the university has a Google account ready to use with Myriad. This also provides flexibility, because Myriad does not have a registration form or a sign in screen. All the requested Google's permissions from a Myriad's user and their descriptions can be found in appendix C.

After users sign in to the system, the first screen they will see is the campaign overview screen. All created campaigns, including the ones assigned by other user's as an assistant are shown as in figure 5.1. It has a simple and clean UI, compared to regular email clients, emphasizing its focus on mass email communication.

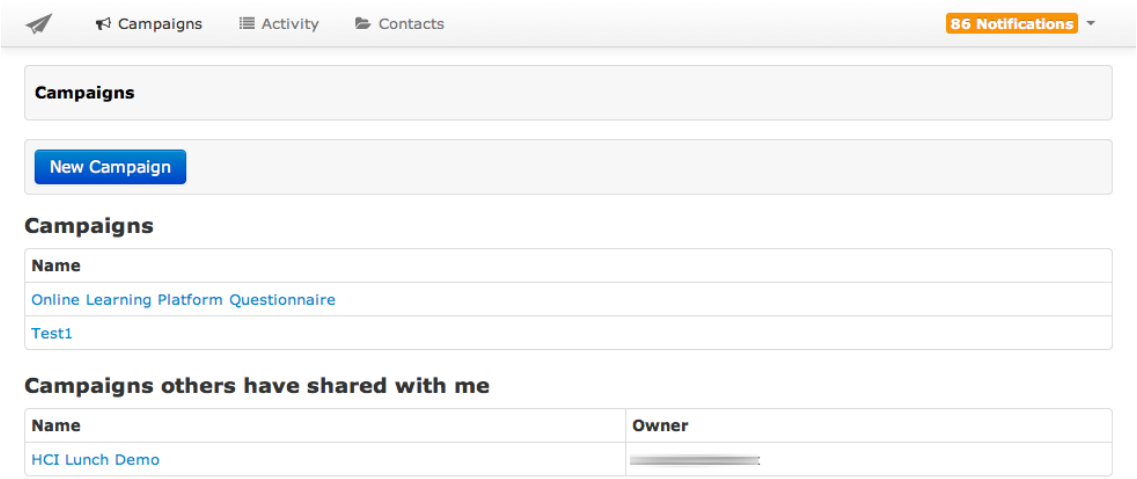


Figure 5.1: Myriad's Campaigns Overview Screen

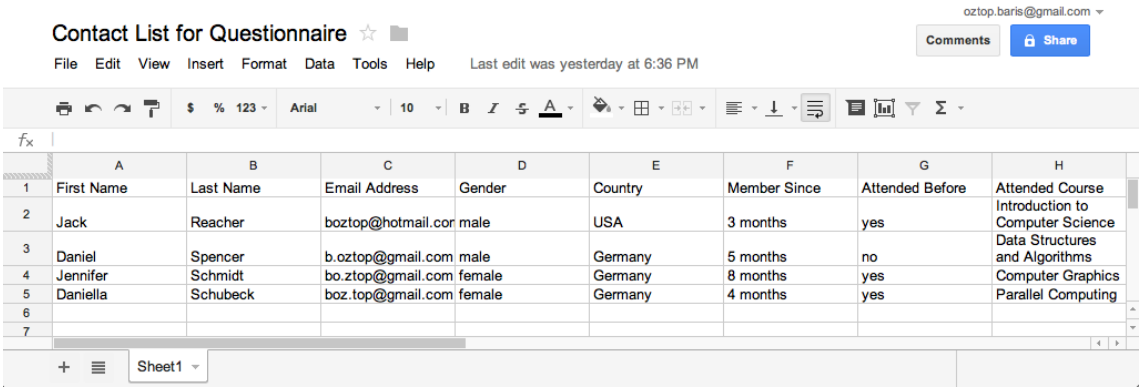
5.2.2 Synchronization with Other Systems

Myriad has the ability to get the recipients' information and their KVPs from Google Spreadsheet[1]. This is a convenient way in importing recipients' information into the system, since many people already keep their recipients' and related information in a spreadsheet environment as discussed in section 5.1.3. Therefore, Myriad offers a bi-

[1]https://docs.google.com/spreadsheet/

directional syncing from and to a spreadsheet, as defined at beginning of the creation of a campaign. Besides this, Myriad also has an option to enter a recipient's first name, last name, and email address into the system directly.

The corresponding columns in a Google Spreadsheet start with "first name", "last name", and "email address" as shown in figure 5.2. The rest of the columns will be filled out as a KVP, and imported into the system as well.



	A	B	C	D	E	F	G	H
1	First Name	Last Name	Email Address	Gender	Country	Member Since	Attended Before	Attended Course
2	Jack	Reacher	boztop@hotmail.cor	male	USA	3 months	yes	Introduction to Computer Science
3	Daniel	Spencer	b.oztop@gmail.com	male	Germany	5 months	no	Data Structures and Algorithms
4	Jennifer	Schmidt	bo.ztop@gmail.com	female	Germany	8 months	yes	Computer Graphics
5	Daniella	Schubeck	boz.top@gmail.com	female	Germany	4 months	yes	Parallel Computing
6								
7								

Figure 5.2: A Google Spreadsheet to Import Recipients' Information into Myriad

As mentioned in section 5.1.4, importing existing email conversations as a campaign message into the system is an important feature. Researchers might have initiated conversations using their regular email client, and if later on they decide to make them more manageable, they can import them to Myriad. Myriad takes advantage of Gmail's labeling feature, which is equivalent to the IMAP protocol's folders for the same purpose (Google Inc., 2013b). Myriad creates a Gmail label in the user's account, and the only thing a user needs to do is the enable the label syncing feature at the campaign creation screen. Next, there will be a label similar to the campaign's name, and grouped under the root label, "myriad" in the Gmail's inbox (see figure 5.3). Hence, researchers will able to import email messages not considered a part of a campaign for some reason, into Myriad.

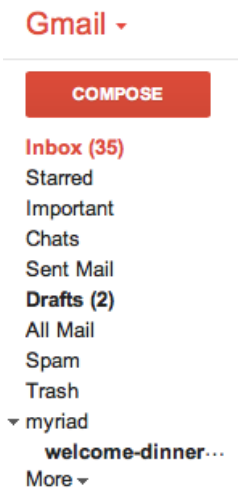


Figure 5.3: Gmail's Labels and Myriad's Campaigns Under the Myriad Label

5.2.3 Creating an Email Campaign

The campaign creation screen (see figure 5.4) has input fields for a campaign name and Google Spreadsheet's Uniform Resource Locator (URL) to synchronize with. Other two checkboxes for the spreadsheet are used to manage the frequency of synchronization and disabling the cascading warning messages in case of an erroneous data in the spreadsheet such as an empty email address field for a contact. The option for importing emails from Gmail's inbox is also in this screen, labeled as "synching section".

After the campaign is created, all the contacts and their KVPs will be imported as long as the user provided a Google Spreadsheet URL like in figure 5.5.

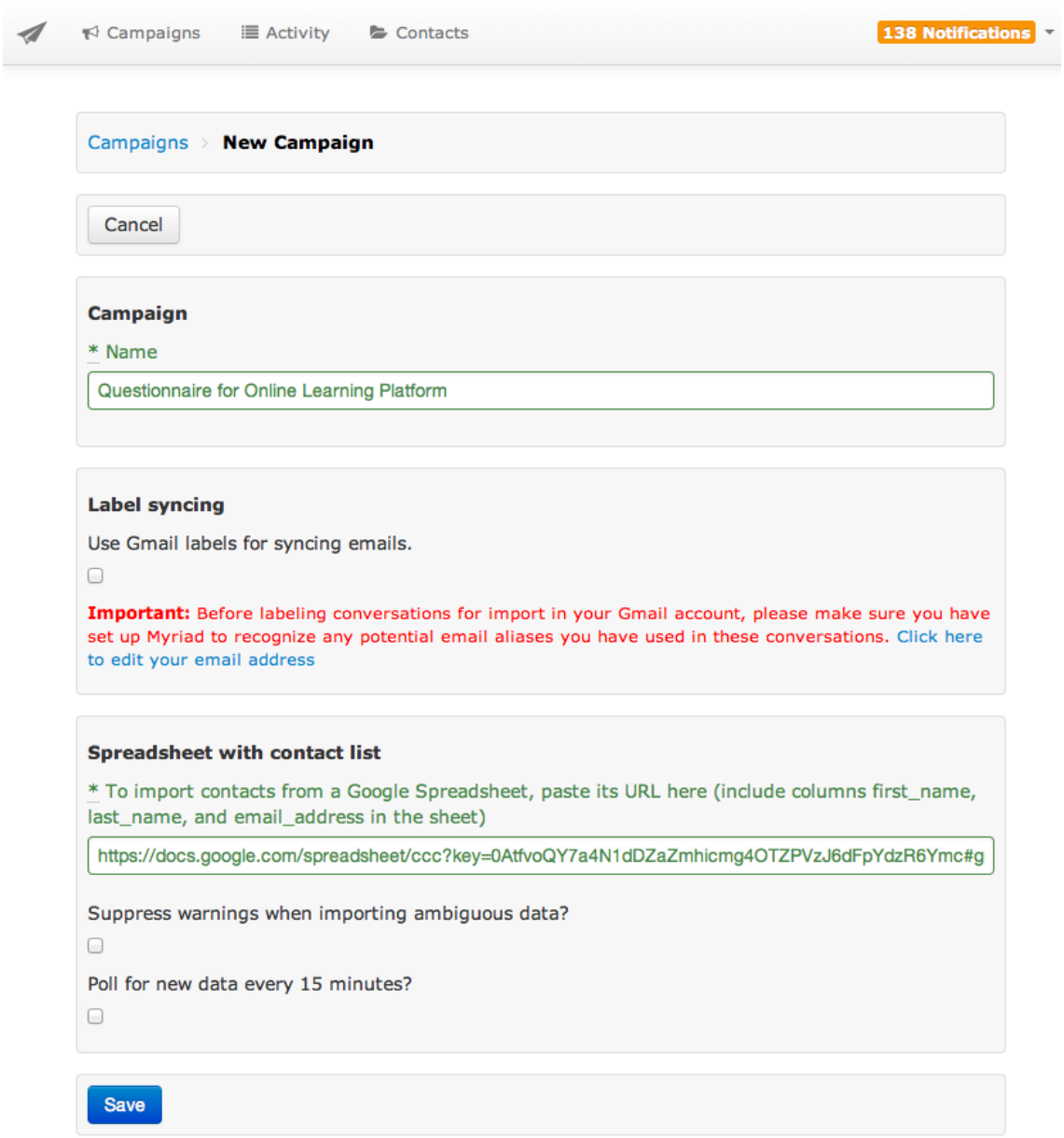


Figure 5.4: Creating a Campaing in Myriad

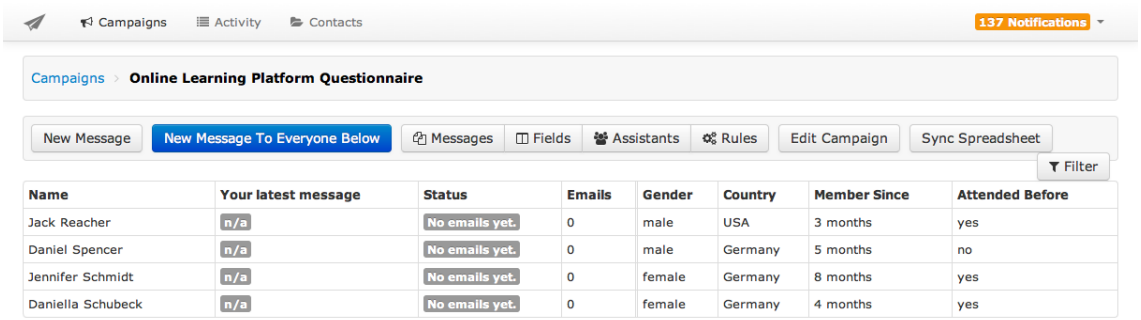


Figure 5.5: Contacts and Their KVPs After Synchronization in Myriad

5.2.4 Composing an Email Message

Users can send emails to all of the contacts they entered or imported into the system, or select a subset of them by using the filtering function as shown in figure 5.6. Provided filtering options are according to the values of KVPs, the conversation statuses such as unread, replied, or unreplied, or the message's template names.

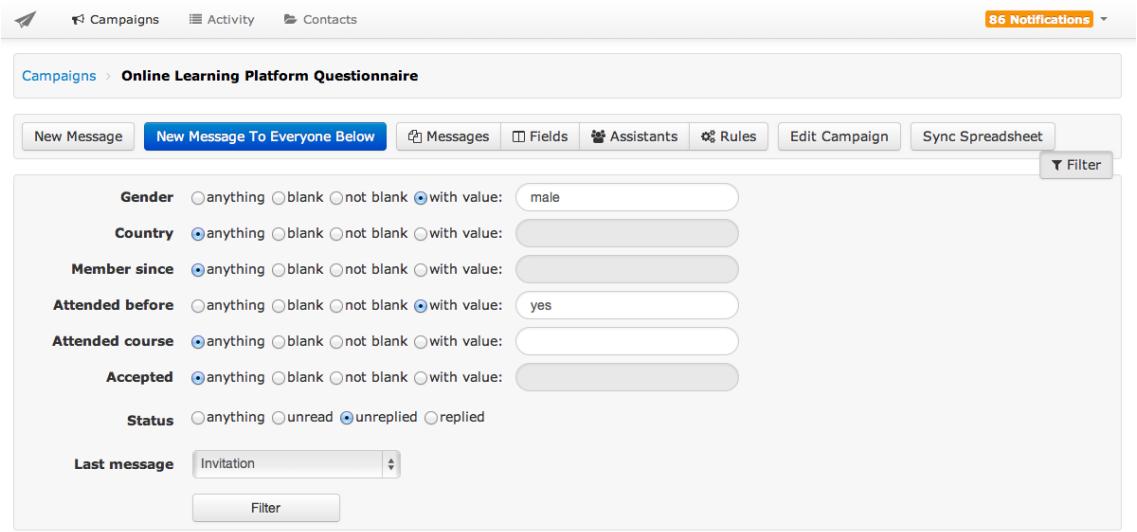


Figure 5.6: Filtering the Contact List in Myriad

By pressing the corresponding button, users can now readily compose emails for their filtered recipients at the compose screen. The compose email pane (see figure 5.7) contains a section that lists previously sent emails available for reuse. This is the same template concept that was introduced for the prototype in Chapter 4. Section 4.2.2. It also shows the visualization of the flow of the communication by using a tree structure, as well as the number of messages sent by using the corresponding template. A more long term campaign's visualization tree can be found in appendix D. The system also suggests an email template while composing a reply by considering the nodes at the same tree structure.

The compose pane allows users to add dynamic variables in the content of messages to personalize them according to the recipients. These variables are not just limited to the first name and last name, as they can also be of any keys from the assigned KVPs in the campaign.

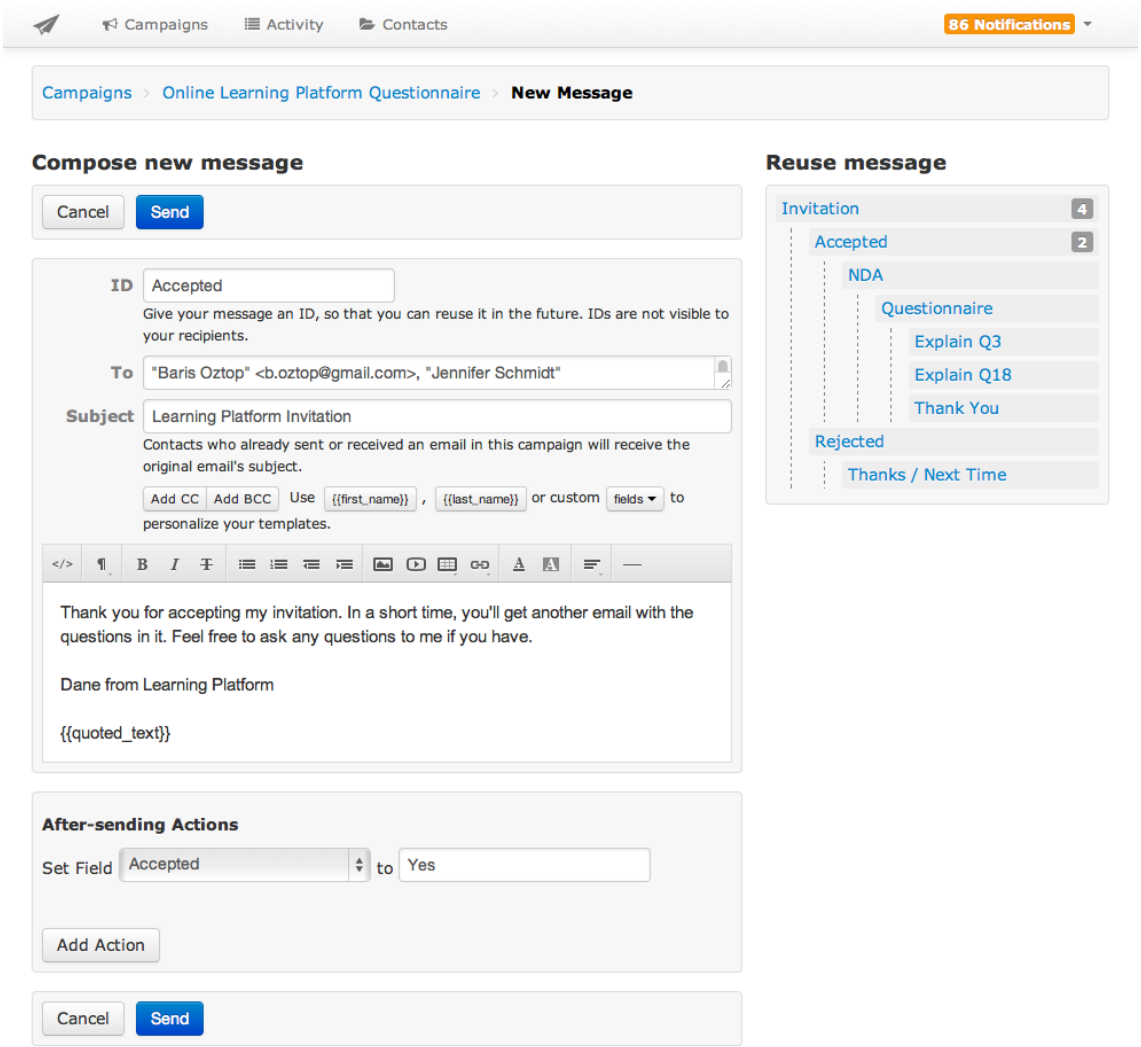


Figure 5.7: Compose an Email in Myriad by Reusing Earlier Messages

Each campaign has a link to the other campaigns, as ling as the recipient is involved in both campaigns, as featured in the next section. A researcher can access conversations and KVPs from other campaigns by switching to the campaign via the provided link. Providing an option to see the earlier campaigns a recipient got involved in gives broader knowledge about a recipient that may help researchers in personalizing the content of the emails more easily and properly. For example, if we have extracted information regarding which sports did the recipients got involved with in an earlier campaign. We can use the same information in making a friendlier and to frankly start a new campaign by mentioning details about the latest events of those sports areas in the country. Such a technique also supports the social exchange theory and the diffusion of responsibility theory as discussed in section 2.3. There is also an option to hide the KVPs in a campaign if they are not related at all or have become obsolete during the flow of a campaign to avoid cluttering the application's view.

In figure 5.7, at the pane of "after-sending actions", users are able to set the values of the keys right after an email is sent. Therefore, a user does not need to browse to another screen in order to change or set KVPs whose values depends on the email that was recently sent.

At the beginning of this section it was mentioned that a user can filter the recipients list, and then send an email to a subset, according to the filter's criteria. Myriad saves those filtered criteria under the "Rules" menu as shown in figure 5.8 after the user sent the email. Therefore, the next time the system receives new emails with the same conditions as before, user will see them under the rules menu, and the earlier sent email messages can be sent to those new matching recipients automatically, if user opt to enable the automated response feature or a user can simply press the send button manually in the same screen. For example, in figure 5.8, there are three recipients whose "attending" key were set to "yes", and since the system already recognized that we have already sent an email to those attending, new matches for the same condition applies. In this case, the user can simply press the send button, or automate the process by enabling the provided option.

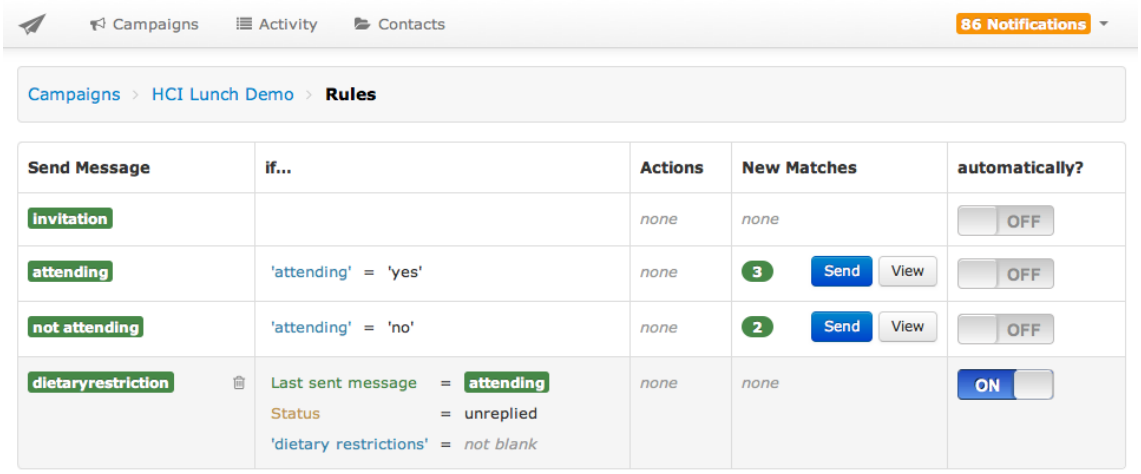


Figure 5.8: Rules to Automate the Sending Process of the Emails in Myriad

5.2.5 Extracting Information from Email Messages

Myriad's reading pane (see figure 5.9) offers a threaded view wherein all of the messages between a specific recipient and a researcher are visually grouped together. The advantage of a threaded view is that it allows a researcher to get a quick overview of the whole state of a conversation, therefore a researcher can write a customized message more easily by focusing on the specific personality of the individual being responded to, considering earlier conversations with him or her at a single glance.

Each of the researcher's message is annotated with the name of the message template that was used with, making the latest state of the communication easily recognizable without having to look for its content. In figure 5.9, emails sent by the user have green labels with indicators of the name of the template used, such as "Accepted" and "Invitation".

While reading a recipient's answer, the extracted information can be recorded as KVPs at the right-hand side of the reading pane (see figure 5.9). Earlier recorded keys' values can also be updated at the same pane. Having KVPs along with the conversation thread gives the researcher necessary information about the person being replied to. Under the KVPs pane, the researcher can also see if the recipient is involved with another campaign, and a link to the involved campaigns are provided with the number of emails exchanged,

next to it. This is a helpful feature in reminding researchers about the existence of previous interactions with the recipient, and if necessary, the researcher can switch to other campaigns to get an overview of extracted information as KVPs.

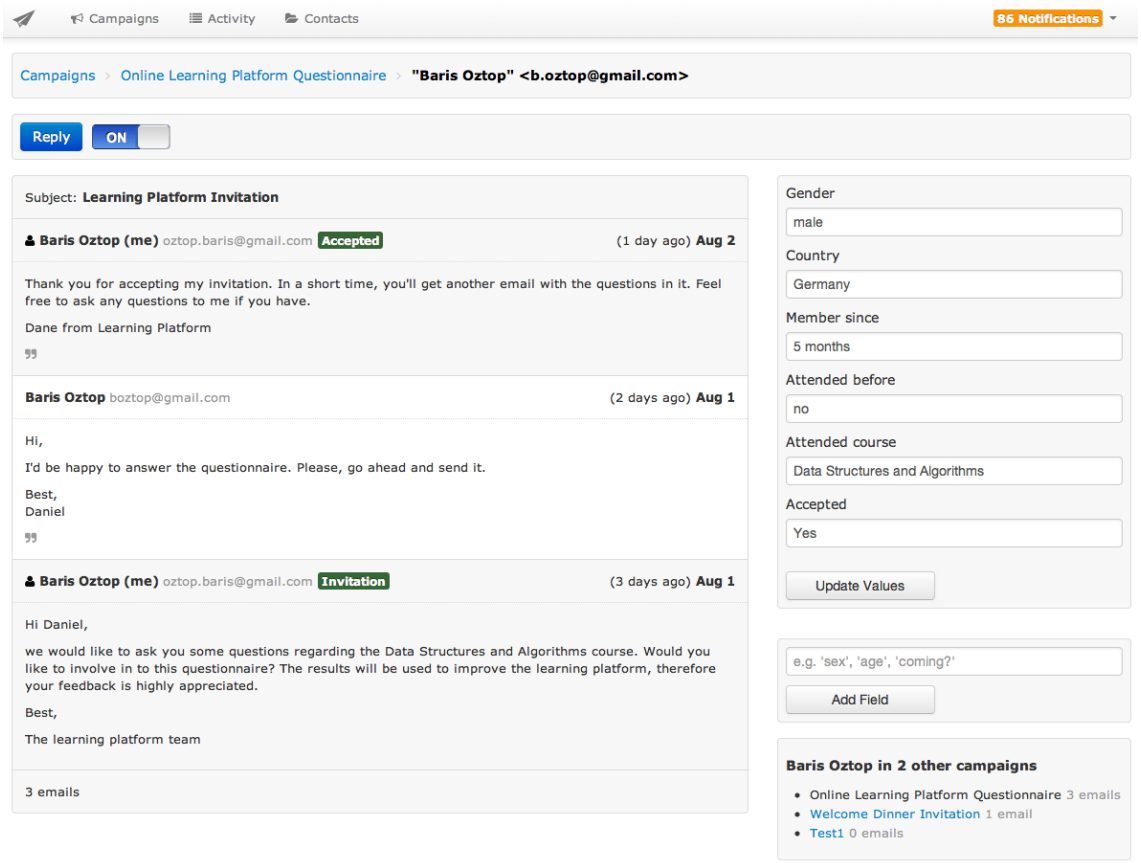


Figure 5.9: The Reading Pane and the Extracted KVPs in Myriad

5.2.6 Enabling Assistants

A researcher can add other researchers as assistants into a campaign by adding their Google account associated email addresses into Myriad, as shown in figure 5.10. The task of the assistant can range from extracting information from emails, writing answers to the recipients, to proofreading a researcher's emails before sending them, and many other tasks that can be generated according to the type of help a researcher needs.

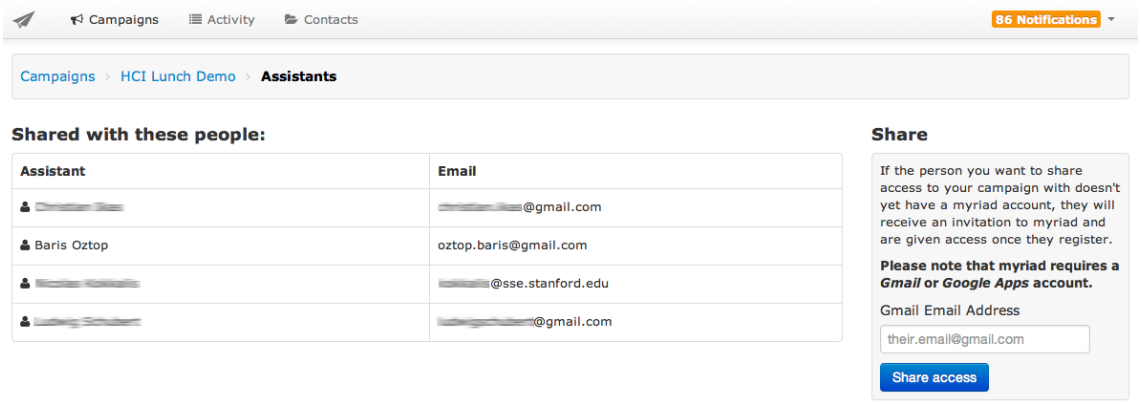


Figure 5.10: Rules to Automate the Sending Process of the Emails in Myriad

After an assistant is assigned to a campaign, he or she will get a notification email including the link to the campaign. Again, there will be a notification email for each email sent by a respondent to the assistants' email address to let them know that they need to act on the situation.

Myriad provides status labels for each received or sent email, giving a hint on the next awaiting action similar to what is shown in figure 5.11. These status labels give hints about the next action to be taken according to the state of the conversation, such as if the user needs to read or to send a reply to a message, or the conversation awaits for an answer from the recipient's side to continue to the communication. There are also status labels related with Myriad's internal state regarding to an email message, such as if Myriad was able to send the messages successfully, or if there was a failure encountered while sending them. The same view also provides a column showing last message sent by the user. With this, the researchers and the assistants will easily realize what should be done next, and see the status of the communication for each recipient.

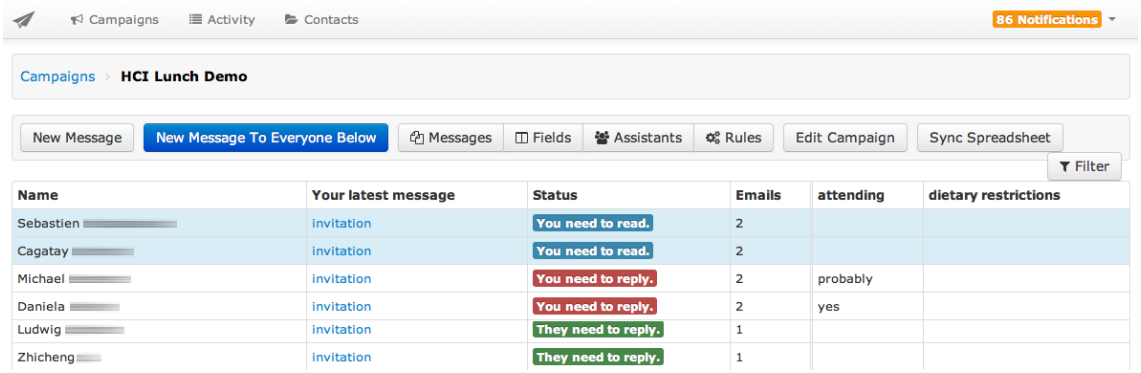


Figure 5.11: Status Labels for Each Email in Myriad

5.3 Architecture

Similar to the prototype, Myriad was also developed using a framework stack of RoR, jQuery, Bootstrap, and OAuth[2] protocol was used for the authentication with Gmail. The initial project structure was generated by a gem named Rails Apps Composer[3], consisting a collection of Rails application templates to start a project. Thanks to the other gems utilized for the project development and completing it on time in a robust way. The following section will discuss the implementation details of Myriad, how it makes mass email communication easier.

5.3.1 Keeping Track of Recipients

As mentioned in section 4.2, the prototype was not able to keep the basic contact information of the recipients, only the messages of them. Therefore, extracted information such as KVPs are only limited to email messages related to the recipients. The downside of this approach is its inability in obtaining all the KVPs of the recipients all at one glance. Instead, we need to go through earlier emails of the recipients each time we need to create the same key, since the system was not aware of an existing KVP created for a different email in the same campaign.

[2]OAuth is leveraged by using OmniAuth library (https://github.com/intridea/omniauth), and its Google authentication strategy is implementation by OmniAuth Google OAuth2. Details can be found at https://github.com/zquestz/omniauth-google-oauth2

[3]https://github.com/RailsApps/rails_apps_composer

In Myriad's case, this issue was solved by keeping all the recipients' information in a separate data model named "contact" as shown in figure 5.12. With this, we were able to keep track of the recipients among different campaigns through different conversations involving them, which is another set of a data model in the Myriad system, as each campaign involves many conversations with the recipient involved with them.

Myriad keeps all of the basic profile information of the recipients, including their first names, last names, and email addresses. These information can be imported into Myriad either by spreadsheet synchronization, by adding them into "To" filed when a user composes an email, or by importing from Gmail's inbox via the label synchronization. The recipients' email addresses are unique identifiers for Myriad in order to compare already existing recipient information with the synchronized ones.

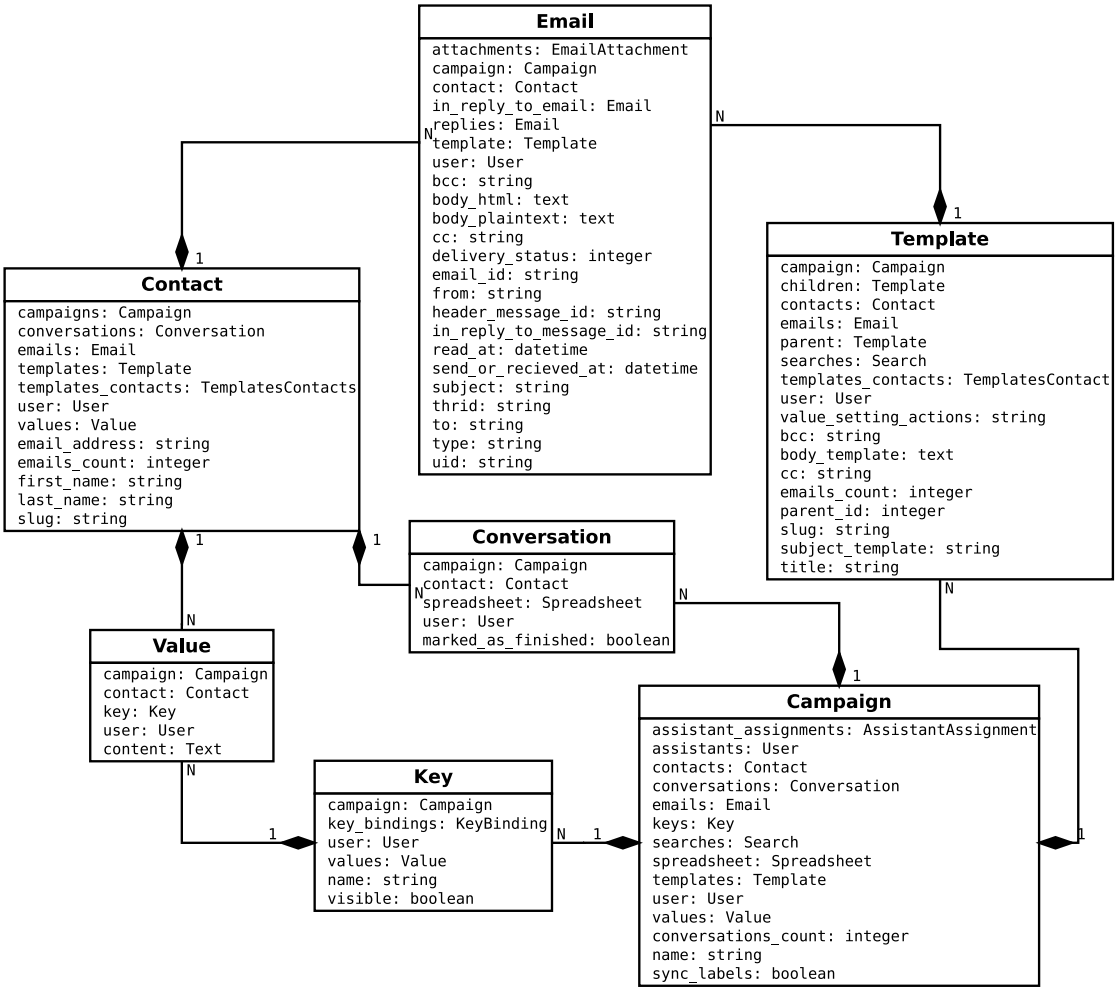


Figure 5.12: Model Dependency of the Myriad Compared with the Prototype

5.3.2 The KVPs and the Templates

Having a separate contact data model in keeping track of the recipients also helped us relate extracted information, specifically KVPs of the recipients, instead of their message in a campaign as shown in figure 5.12. Each campaign has many keys, and as well as their corresponding values, according to the recipients. One advantage of this is that as we browse a specific conversation with a recipient, we will be able to see all KVPs of the specific recipient under one view, which is the KVP pane next to the email threads, as shown in figure 5.7. The other benefit is its ability to synchronize with spreadsheets. When a new key is added to a recipient or a value of a key is updated via Myriad, it will be reflected on the spreadsheet as well. Therefore, the spreadsheet will always contain the latest changes done by the user in Myriad.

Similar to the prototype, templates contain emails messages with dynamic variables in it, and the actual emails whose dynamic variables are filled with values are kept in the Email data model as shown in figure 5.12, after they are fetched from Gmail.

To get an overview of the conversation and let the users select templates, Myriad offers a tree structure similar to the prototype. However, the execution of the tree structure is different from the prototype. In section 4.2.3, it was mentioned that the hierarchy between the nodes of the tree was a nested set model. We decided to use an adjacency list model instead of a nested set model to increase the readability of the templates' relationship with the data model level from the developers' perspective. Since the relation of the templates is not as complicated compared to hierarchical relations, the performance drawbacks on this decision was negligible.

5.3.3 A Campaign Message Identification

Due to the prototype's dependency on the existing EmailValet's email fetching process, we identify emails if they belong to a specific campaign or are user's regular emails, and then inserting them into the same inbox in EmailValet. With Myriad, we removed this dependency, and considered showing emails belonging to a campaign.

The identification of the emails depends on the three different conditions:

1. The emails fetched from Gmail and responses to one of to the campaign message.
2. The emails composed in Myriad as a part of the campaign, whether they are the emails to initiate a campaign or replies to recipients' emails.
3. The emails imported into Myriad from Gmail's inbox via label synchronization (as shown in section 5.2.2 for label synchronization)

Myriad initially retrieves all the email's Unique Identifier (UID)s[4] from Gmail's inbox for the last 14 days, instead of all the emails in the inbox to reduce the time needed for the fetching process. For those UIDs not stored in the Email data model of Myriad, Myriad retrieves the actual email along with the metadata, such as Gmail thread ID[5] and Gmail labels.

For the prototype, we have set a specific message-ID to the emails composed in Myriad. With this, we are able to identify whether an email from Gmail was composed to initiate a campaign or is a reply to one of the recipients of a campaign by a Myriad user during the aforementioned fetching process. Message-ID field is again leveraged for identifying a recipients' message belonging to a campaign. Myriad retrieves the Message-ID written in the In-Reply-To, and checks if it matches one of the campaign messages in Myriad. Another field that Myriad checks is the "References" field[6] to identify messages as a part of a campaign, but was forwarded to another person by a recipient in order to reply the message. Since the forwarded endpoint's email client sets the Message-ID of the forwarded message into the In-Reply-To header field and sets a new Message-ID, Myriad is not familiar with neither of the Message-IDs at all after it got the forwarded or replied message. Myriad checks the "References" field to identify a Message-ID belonging to a

[4]UID (Unique Identifier) is 32-bit integer value to uniquely identify emails in a mailbox in IMAP. Each email added into mailbox will have a higher value than the ones added before, however they are not necessarily contiguous (Crispin, 2003). UID is used as it was in the prototype by the implementation of the EmailValet's part. It is used to identify if a message is already fetched from Gmail, if it is not, it is fetched since it is a new message.

[5]Gmail thread ID is a 64-bit integer that is part of the provided Gmail's IMAP extension. It helps to associate groups of messages in the same manner as in the Gmail web interface (Google Inc., 2013a).

[6]Reference field in another identification field along with Message-ID defined by RFC 5322. It contains one or more Message-IDs used when creating a reply to a message to identify the original message or the other messages when a reply to a message that was itself a reply (Resnick, 2008).

campaign in this case. This scenario is illustrated in figure 5.13.

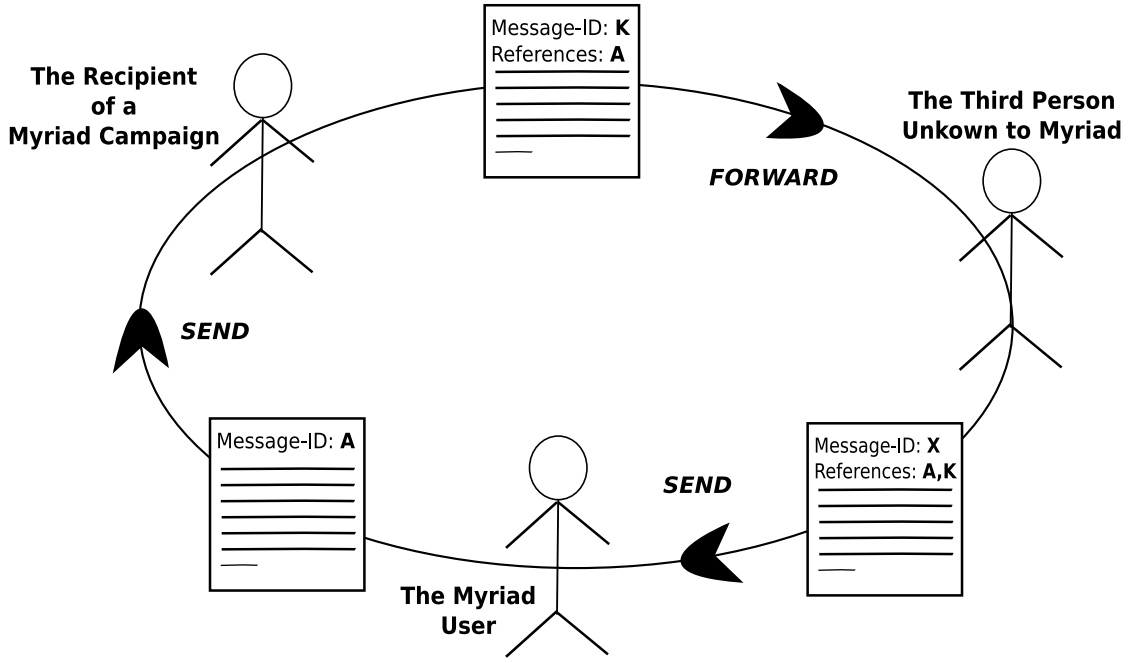


Figure 5.13: A Scenario Using References Field in an Email

If an email is not identified as a part of a campaign message by comparing the Message-IDs with the ones in Myriad's Email data model, the next step is to evaluate those emails in case they are labeled as part of a Myriad campaign in Gmail's inbox. In this case, if an email has a label in the pattern of "myriad/campaign-name", and there is already a created campaign in Myriad with the name "Campaign Name", that email will be stored in the data model. Finally, an identified email will be assigned to a contact by looking at its parent email's Message-ID if it was a reply or if it fails, the contact information will be recorded from the email's "From" header field. The rest of the emails will be ignored if no matches were found in the mentioned email header fields, specifically Message-ID, In-Reply-To, and References, or Gmail's Label.

5.4 Experiences

To date, Myriad has 67 users since its beta launch[7] last January 30, 2013. Among those 67 users, 30 of them are actively using it, since last month. In this section, statistics will be given about the created campaigns, and continued with some user's testimonials.

[7]Beta version of Myriad was launched at http://gv.stanford.edu/ and still accessible from this address.

5.4.1 Statistics

The beta launch of Myriad has been advertised at the Stanford HCI group and to some of Stanford's organizations who performs mass email communication regularly with their communities. Some of those campaigns opted to use the following features of Myriad, such as:

- **Invitation to an Event:** User created an initial invitation template and another template for those who rejected the invitation to motivate them for other upcoming events. Recipient list is imported via spreadsheet. User also created KVPs to easily identify recipients needing an individual follow up. Salutations of the emails are personalized by their first name through KVPs.
- **Survey via an External Link:** Users send an initial template including a link to an external survey. A detailed list of contact information was imported from a Google spreadsheet. Later on, users continued with a reminder email in a separate campaign.
- **Importing via Gmail Labels** One of the user imported campaign emails from his Gmail account to Myriad via a label synchronization, and continued the campaign in Myriad.
- **Interview Results** One user reached nearly 500 recipients through Myriad in releasing the result of an interview process. User created a rejection template for those who were rejected, and for the ones who were accepted, he created several templates depending on the invitation date, including the specific schedule related to that date. The user also automated the email sending process by the created rules on the filtered search results.
- **Getting User Details via Form** During one of the live demonstrations of Myriad at the Stanford HCI group, we asked to the attendees to browse to a link and fill up a Google Form[8] to obtain their contact information in a Google Spreadsheet, where the gathered data is recorded in Google Form by default. Therefore, we were able to easily import contact information into Myriad, and initiate a campaign. There were also four assistants extracting information during the live demonstration while the recipients were sending answers.

[8] https://docs.google.com/forms

After reviewing some cases using Myriad in real time, the following paragraphs will provide usage statistics regarding Myriad, starting from its beta launch date, January 30, 2013 until August 7, 2013.

Campaigns and Their Number of Recipients Figure 5.14 shows campaigns[9] created by Myriad, and their respective number of recipients. The average number of recipients was 105.32, and the total number of recipients was 9585. While the maximum number of recipients for one campaign was 752, the minimum amount was only one recipient.

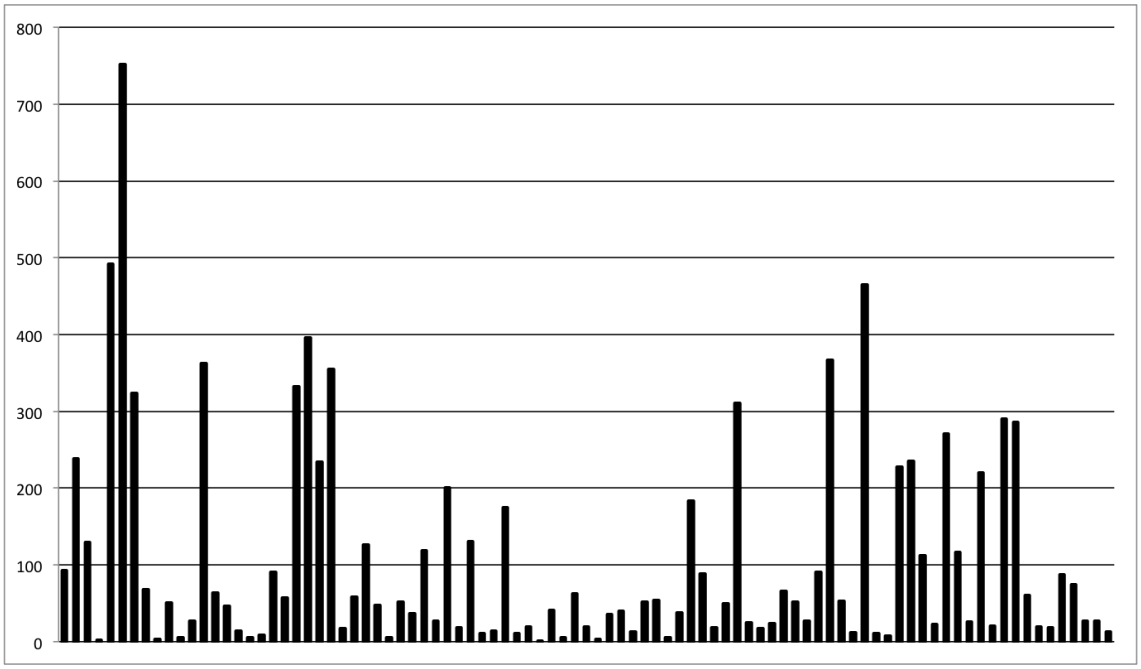


Figure 5.14: Campaigns and Their Number of Recipients in Myriad

Myriad Users and Their Number of Campaigns Figure 5.15 shows Myriad's total number of users[10], and the number of campaigns that they created with Myriad. There are a total 96 campaigns, after the ones created for test purposes were ignored. The average was 3.92 campaigns per user. While the maximum amount of campaigns created by one user was 13, the minimum amount was one campaign.

The Number of Campaigns per Month Figure 5.16 shows the ratio of created campaigns per month. 35 campaigns out of 96 were created in April 2013, as the highest

[9]Campaigns' names are removed due to privacy reasons.

[10]Users' names are removed due to privacy reasons.

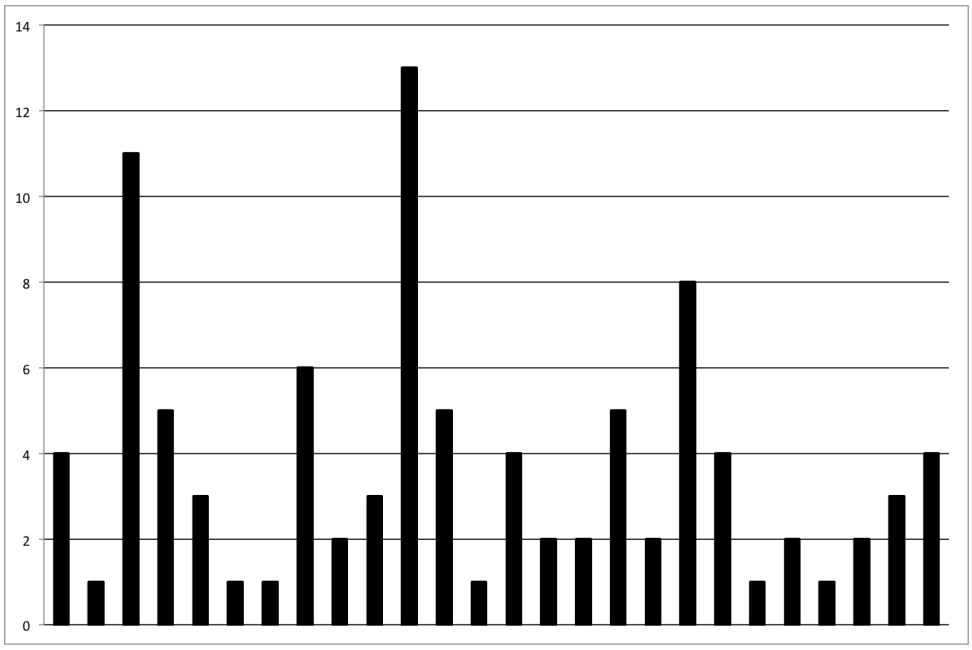


Figure 5.15: Myriad Users and Their Number of Campaigns

amount compared to other months. It might be because it is the middle of spring quarter at Stanford University, where lecturers and organizations need to reach their recipients more often. There were an average of 15.5 campaigns create per month if the launch month January is ignored.

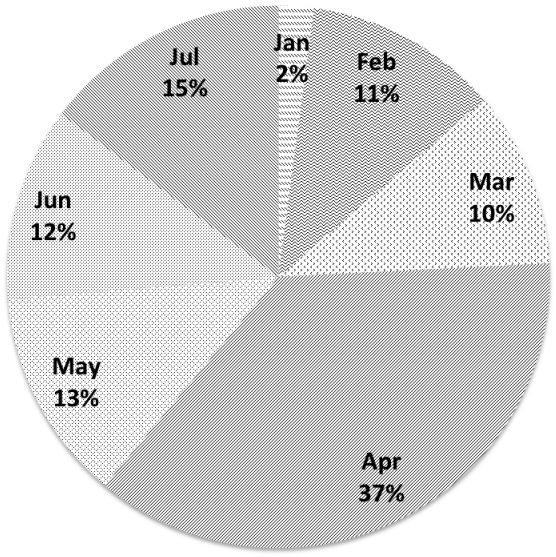


Figure 5.16: The Number of Campaigns per Month

5.4.2 User Testimonials

In this section, testimonials from the very active users of Myriad will be evaluated. Users were asked to evaluate Myriad, keeping the following four questions in mind:

1. What are the scenarios you use Myriad for?
2. What were you using before Myriad?
3. How satisfied are you with Myriad?
4. How regularly do you use Myriad?

Testimonials written in listing 5.1 and 5.2 belongs to lecturers from the Stanford HCI group. The last testimonial belongs to a Stanford organization who performs mass email communication very often to reach their community.

> My main use: handling a special issue. Several hundred emails (literally) meant that an email view was impossible. Myriad's structured view makes everything a whole lot easier. It still needs some polishing to be ready for prime time (like it's workflow emphasis is a little constrained, and it needs better handing of formatting and attachments), but as a first draft, it's amazing. I hope it continues so I can use it more!
>
> – Scott R. Klemmer, Stanford HCI Group

Listing 5.1: User Testimonial about Myriad #1

> I use Myriad to manage a large volume of emails about the courses that I'm teaching. I tag the email with the campaign, and the system+assistant help me respond with one of a number of common responses.
>
> – Michael S. Bernstein, Stanford HCI Group

Listing 5.2: User Testimonial about Myriad #2

> I use Myriad when I'm mass−mailing a group of people and want to personalize the email. Before, I was using Mailchimp or BCC everyone. It's a great tool to send personalized emails to the network! I use Myriad whenever I need to email more than 10 people and want to personalize the emails. For less than 10, the set up for Myriad may take longer than sending out the emails individually (this was discovered through trial and error).

Listing 5.3: User Testimonial about Myriad #3

5.5 Conclusion

In this section, the final solution, Myriad, for mass email communication was investigated. We first investigated the improved requirements, comparing them with the initial requirements, and the ones already applied to the initial prototype. Enabling assistant support, using KVPs as dynamic variables in the email, importing recipients list from a Google Spreadsheet, the ability to import email from Gmail's inbox by annotating emails using Myriad's campaign labels, and automated decision-making through previously applied search filters were some of the main improvements done in Myriad.

Previous sections continued showing how those features were applied for Myriad's UI, followed by introducing its architecture on how it was possible to enable the said features in support for mass email communication.

Finally, user and campaign statistics were given to show how actively Myriad was used during its beta stage, and the section was followed by testimonials from the active users of Myriad.

After discussing what Myriad is, its features, and users' testimonials, let us see how these features fit into a personalized mass email communication. A personalized mass email communication can be illustrated as shown in figure 5.17.

The information will be extracted from the recipients' emails and then a researcher will compose personalized emails for each recipient, according to the extracted information.

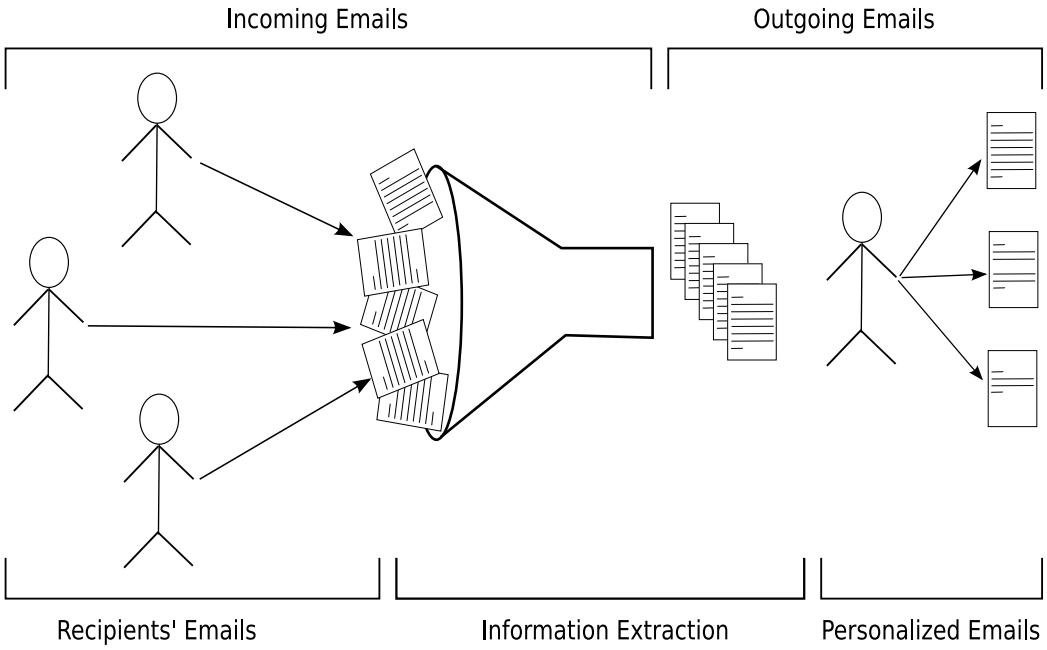


Figure 5.17: Illustration of a Personalized Mass Email Communication

Therefore, the discussed features of Myriad and their benefits can be grouped under the handling of the incoming emails and the outgoing email as follows:

Handling The Incoming Emails

Myriad has a campaign inbox view to avoid cluttering of unrelated messages in the inbox, fetching only the campaign related emails from the user's email account by ignoring private emails, and grouping each email according to the campaign it belongs to.

As a researcher reads the incoming emails, he or she can also record the extracted information from those emails into KVPs at the same time. This avoids the need on using a third-party application like a spreadsheet. However, if a researcher prefers a traditional way in recording information like a spreadsheet, Myriad still offers an option in synchronizing those KVPs with a Google Spreadsheet.

Assistants can help a researcher with appropriate given permission on extraction of in-

formation from incoming emails.

Handling The Outgoing Emails

Myriad provides a threaded view of all the email conversations of a recipient in a campaign letting researchers consider the earlier conversations with a recipient, without having the need to browse through other views. This saves time and helps on getting a quick overview of the conversation history to get a better context on personalizing the succeeding new messages.

The researchers can continue on using his or her own Gmail's account on answering or initiating a conversation. Myriad offers Gmail label synchronization to import those emails into a campaign, and make them more manageable along with other campaign emails.

The researchers can opt to reuse previously written emails as a template in composing new emails. The provided template tree also gives a direct visualization of the current campaign state and flow to let the researcher easily go through earlier given answers.

The researchers can use the recorded KVPs to personalize the outgoing emails. Since personalization has been addressed as an important factor in increasing response rates, the initiated campaign will end up with better results compared to non-personalized emails.

Myriad records the previous actions of a user as a rule in the system to notify the automation of the email sending process with the matching conditions of a group of recipients. This helps to save time and increase awareness of the researcher to previously taken actions for repeated or similar situations.

Assigned assistants can also help the primary researcher in composing emails or proof-

reading the composed emails before they get sent out to their respective recipients.

In the next section, we will see the conclusion of this study and the possible extensions on the requirements of Myriad as a future work.

6 Conclusion and Future Work

In this final chapter, the study will be finalized by summarizing it under the conclusion section, and its possible extensions will be discussed in the future work section.

6.1 Conclusion

Increased Internet usage made email a popular medium for communication because of its low cost, quick turnover and the flexibility of connectivity through the use of mobile devices. The benefits of email communication also attracted researchers to use it as a data collection medium to explore, describe, and explain things by communicating with a large group of people.

However, the nature of communication with large groups is different compared to small ones due to the required effort in personalizing messages according to its respective recipient. As a result, researchers tend to write more generic emails, ignoring recipient-specific details. Researchers investigated many response rate influences, and addressed that personalization of the message content is an important factor in increasing the response rate (Dillman, 1991; Schaefer and Dillman, 1998). Messages that are not personalized results in a low response rate for the answers expected from the recipients. As a result, this may end up with a nonresponse error.

Nonresponse error has been considered as a major problem by many researchers, because of the tendency of some of the respondents as a part of a large group who fails to respond might result in a biased estimate for the proper representation of the population. Hence, it affects the outcome of a research (Bogen, 1996).

For this reason, the researchers investigated possible theories regarding the reason behind proper personalization's effect on the response rate. While Dillman et al. (2009) emphasized the social exchange theory since the personalization of emails helps build a connection between the respondents and the researcher, Barron and Yechiam (2002) stressed on the diffusion of responsibility theory, which talks about the awareness of the availability of the other volunteers will result in a higher utility of not volunteering.

The researchers conducted studies on investigating the diffusion of responsibility. Barron and Yechiam (2002) showed that the number of replies where they used a single email address in the "To" field got a 20% higher response rate and the number of "very helpful" replies was 187% higher compared to the replies they got when they used groups of email addresses. In the study of Selm and Jankowski (2006), the recipients showed privacy concerns when the header of the email contained all the email addresses of the other respondents explicitly.

Other researchers investigated the social exchange theory. Heerwegh (2005) applied personalized salutations in the the emails, and got a 6.9% higher login rate for the provided survey link in the email compared to the non-personalized group in their study. In the study of Joinson and Reips (2007), when the level of authority or status of the sender was used with the personalized salutation, they got a 53.4% response rate while non-personalized salutations with a neutral power of the sender status got a 40.1% response rate.

Even though personalization of emails helps increase the response rates, Dillman et al. (2009) emphasized that an overwhelmed personalization can also result to a low response rate. Hence, only the adequate amount of personalization and its appropriateness should be considered for personalizing emails. In addition to this, experienced email users can easily notice if a message is manually written by a person, or if it is computer generated. Therefore, it is hard for the digital world to achieve authentic personalization, and achieving such level of personalization requires getting to know each recipient very well.

To understand what existing applications offer in support for personalized mass email communication, CRM, Help Desk, and Email Marketing applications were evaluated in chapter 3. Several features of these applications are considered to be helpful to researchers for their mass email communication. Some of these features are the following:

- Storing client related information extracted from conversations in KVPs.
- Integration with popular email clients such as Gmail.
- Importing recipient lists from third-party applications.
- Assigning emails to a recipient profile.
- Using dynamic variables in the emails to be replaced by their values.
- Assigning a help desk ticket to an assistant.
- Reusability of an existing email as a template.

Moreover, none of the available products indicated above offers these mentioned features in helping a personalized mass email communication all under a single product and some of these features needs additional effort to use since their application focus is not a mass email communication. Therefore, this study came up with an initial solution as a prototype featured in chapter 4. The prototype supported following features to help a researcher reach large groups of people in a personalized way with less effort:

- Kept extracted information in KVPs along with the recipient messages.
- Used dynamic variables to personalize the salutations of emails.
- Allowed to reuse earlier sent messages as a template.
- Provided a tree structure to get an overview of the state and flow of the conversation with the used templates.

Even though the prototype provided the above mentioned features in support to a personalized mass email communication, we moved the prototype from its origin product, EmailValet, to a new project after the informal user test with an organization who performs mass email communication regularly at Stanford University. The user test ended up with new features, and it helped us realize the drawbacks of the existing EmailValet as an email client we used for the prototype.

The improved requirements after user test brought up the final solution in chapter 5, and its features in support for mass email communication as follows:

- Recipients' information can be synchronized to the Myriad system through a Google Spreadsheet to remove the effort on importing and exporting recipients' information from the researchers.
- A researcher can assign assistants to an email campaign for them to interact with the flow of a mass email communication with tasks such as extracting information from incoming answers, proofreading the primary researcher's replies before sending them or even composing replies to those answers.
- Extracted respondents' information can be recorded to the recipients' respective profiles in the KVPs during the whole state of a campaign.
- Those recorded KVPs can be used as dynamic variables in personalizing the content of the emails.
- Reusability of exiting emails and getting an overview of the state of the conversation.
- Provided a decision-making mechanism to automate earlier decisions on sending out emails in a user verified manner.
- Provided visual cues to notify users on the current state of communication with individual recipients, and suggesting potential action based on the previous decisions of the user.
- All conversations and extracted information of each recipient are provided under a single view respectively to conveniently reach all the necessary information to write recipient specific personalized emails.

Considering the features gathered in the final solution, we provided a workflow to the researchers aimed to help reduce the amount of time needed in personalizing the content of the emails. Considering the Effort vs. Personalization chart from figure 4.2 after the final solution, it could be replaced in a way as shown in figure 6.1, compared with other products in the market focusing on email communication and data collection. Considering the user feedbacks and the amount of people actively using Myriad during its beta stage, we believe that Myriad was able to bridge the gap in achieving the gold standard of effort vs. personalization, as annotated in the figure.

Finally, possible future work on enhancing the existing solution and remove the known existing restrictions are to be discussed in the next section.

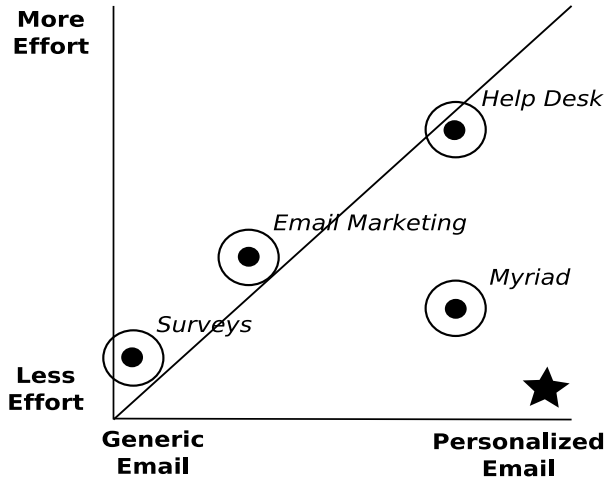


Figure 6.1: Effort vs Personalization and the Comparison of Available Solutions

6.2 Future Work

The provided assistant support in the final solution helps researchers share certain tasks in extracting information from emails, proofreading the primary researcher's replies before sending them, compose replies to recipients' emails, and to verify the rule-based actions before they are acted upon.

Along with the assistant support, the system could also provide an option to allow a crowd of anonymous workers to perform the tasks of assigned assistants, as a crowd assistant. If the system would be able to provide the required functions to allow the involvement of crowd assistants, the decisions on the possible tasks in a mass email communication can be done by anonymous crowd workers.

Surowiecki (2005) stated that under the right conditions, groups can be remarkably smart, and even smarter than the smartest person within them. Therefore, if you try to solve a complicated problem or try to make a decision, the best thing to do is to ask a group instead of trying to find an expert. Therefore, taking advantage of crowd assistants can

minimize the work a researcher needs to do in a personalized mass email communication.

Another improvement could be about the dynamic variables in the messages. KVPs were used in the messages as dynamic variables to personalize the content of the messages as described in section 5.2.4. However, when communicating with larger groups of people, some KVPs might not be applicable to some recipients, or other irregular KVPs needs to be used to personalize messages for some recipients in a large group. When we reviewed available products in the market as featured in chapter 3, Zendesk solved this issue by adding conditional logic to the dynamic variables, as illustrated in listing 3.1. Such extension would be an option for Myriad to avoid creating additional email templates for closely similar emails, with minor changes on KVPs.

In chapter 3, we saw that that CRM applications provide task management options to remind users about upcoming tasks. For example, Highrise allows association of a task to an email to let users easily browse through the source of the task in the provided task list. Currently, Myriad users can maximize the use of the KVPs for the same purpose by adding a key named "task" and a value according to the recipients. However, this contradicts with one design principle, separation of concerns, since the logic behind the KVPs is to store extracted information from the recipients' emails and use them to personalize emails. Besides, KVPs connected to the recipients' profile in a campaign is not useful in adding task messages related to a message of a recipient. Therefore, the task management feature in Myriad could be useful for researchers in attaching reminder messages as a task along with the recipients' messages, and a list of these tasks are shown at the main page of a campaign.

One Myriad user mentioned about the email editor and its attachment handling in the user testimonials, as shown in listing 5.1. Currently, Myriad provides an HTML editor to compose an email; however, it does not provide an option for adding attachments. This is due to the limited time for development, and adding this feature was not considered as a priority at the time of the development since users can still use Gmail in composing

an email, and Myriad will able to import it, together with its attachment. Therefore, a better HTML editor with an email attachment feature would necessary to accomplish all email composing tasks inside the Myriad system.

Next improvement could be on the visualization tree of the communication state and flow. The current visualization tree's nodes are arranged according to email templates and their order in the communication. However, a user can opt to create all email templates at the beginning of a campaign and use them according to the recipients answer later on. In this case, all the created templates at the beginning of the campaign will be considered as individual root nodes of a tree, losing the hierarchical structure in visualizing the communication state. Therefore, the possible improvement should let Myriad allow the rearrangement of those nodes in the visualization tree, together with a drag-and-drop ability for each node.

Lastly, a better workflow can be implemented in the future by the provided rules created according to the decision on sending emails. Currently, Myriad saves filtering conditions of a recipient search as a rule in sending emails the next time, without having the need to search for matching conditions of the recipients again. This feature is quite unintuitive, since when user decides to browse through the rules view, no defined rules will be shown, unless the user uses the provided filtering options in getting a subset of the recipient list and then send an email to the recipients afterwards. To make this feature more intuitive and better define the workflow of a mass communication before-hand, Myriad could offer an option on creating rules right under the rules view. With this, a researcher can easily create the rules and assistants can verify and send emails according to the matching rules, or even opt to automate the process by the provided option under the rules view.

A An Email to Study Volunteer's Dilemma

Single Condition

Date: Mon, 7 May 2001 19:26:21-0400

From: Sarah Feldman <feldman@yahoo.com>

To: harry@bt.technion.ac.il

Subject: please help

Quintuple Institutional Condition

Date: Mon, 7 May 2001 19:26:21-0400

From: Sarah Feldman <feldman@yahoo.com>

To: harry@bt.technion.ac.il,fredy@bt.technion.ac.il,jannet@bt.technion.ac.il,frieda@bt.technion.ac.il,susan@bt.technion.ac.il

Subject: please help

Quintuple Generic Condition

Date: Mon, 7 May 2001 19:26:21-0400

From: Sarah Feldman <feldman@yahoo.com>

*To:*harry@bt.technion.ac.il,david_87_5@zahav.net.il,opary76@hotmail.com,38Labovitz@hotmail.com,gilad_H_G@yahoo.com

Subject: please help

I am a graduate student of biology and am considering continuing my studies at the Technion. Do you know if there is a biology faculty at the Technion?

Thanks in advance,

Sarah Feldman

In the interest of privacy, email addresses have been changed. Source: Barron and Yechiam (2002)

B An Email Message Format

```
Delivered-To: oztop.baris@gmail.com
Received: by 10.112.143.99 with SMTP id sd3csp47394lbb;
        Sun, 28 Jul 2013 03:35:18 -0700 (PDT)
Return-Path: <boztop@gmail.com>
Received-SPF: pass (google.com: domain of boztop@gmail.com designates 10.182.186.66 as
permitted sender) client-ip=10.182.186.66
Authentication-Results: mr.google.com;
       spf=pass (google.com: domain of boztop@gmail.com designates 10.182.186.66 as permitted
sender) smtp.mail=boztop@gmail.com;
       dkim=pass header.i=@gmail.com
X-Received: from mr.google.com ([10.182.186.66])
        by 10.182.186.66 with SMTP id fi2mr49092899obc.98.1375007717111 (num_hops = 1);
        Sun, 28 Jul 2013 03:35:17 -0700 (PDT)
X-Received: by 10.182.186.66 with SMTP id fi2mr49092899obc.98.1375007717106;
 Sun, 28 Jul 2013 03:35:17 -0700 (PDT)
MIME-Version: 1.0
Received: by 10.76.120.234 with HTTP; Sun, 28 Jul 2013 03:34:57 -0700 (PDT)
From: Baris Oztop <boztop@gmail.com>
Date: Sun, 28 Jul 2013 12:34:57 +0200
Message-ID: <CAF2E4bfH4+GAYHcJFZJ6dTJJ+pux4mTjff2neCS_VR_zVCUY9g@mail.gmail.com
>
Subject: Questionnaire
To: Baris Oztop <oztop.baris@gmail.com>
Content-Type: multipart/alternative; boundary=089e0153688e8aabe504e28feec4

--089e0153688e8aabe504e28feec4
Content-Type: text/plain; charset=UTF-8

Hi,

thank you for reaching out. I'd like to participate in the questionnaire. I
hope my answers will be helpful for you.

Best,
Baris

--089e0153688e8aabe504e28feec4
Content-Type: text/html; charset=UTF-8

<div dir="ltr">Hi,<br><br>thank you for reaching out. I&#39;d like to participate in the
questionnaire. I hope my answers will be helpful for you.<br><br>Best, <br>Baris<br></div>

--089e0153688e8aabe504e28feec4--
```

Listing B.1: An Email Message with Header and Body Sections

In the interest of privacy, email addresses have been changed.

C The Necessary Google Permissions

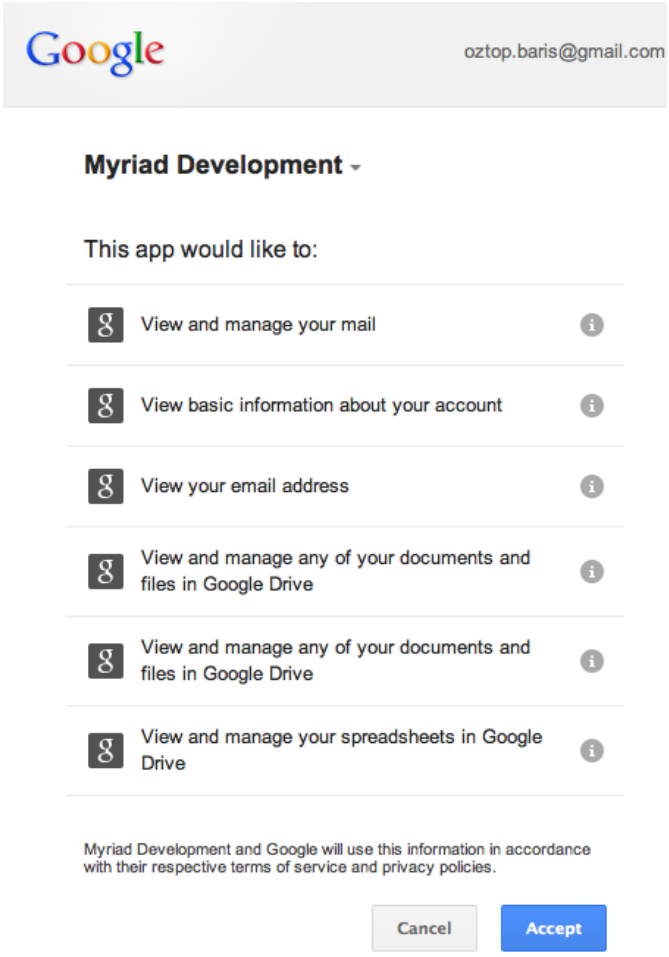


Figure C.1: The Necessary Permissions From Google to User Myriad

Table C.1: Comparison Matrix for Help Desk Applications

Permission	Description
View and manage your mail	• View and manage your mail in Google Mail
View basic information about your account	• View your name, public profile URL, and photo • View your gender and birthdate • View your country, language, and time-zone
View your email address	• View the email address associated with your account
View and manage any of your documents and files in Google Drive	• List and search all of your documents and files in Google Drive • Download any of your documents and files in Google Drive • Create, move, copy, edit, or delete any of your documents and files in Google Drive • Share or unshare any of your documents and files in Google Drive

Continued on next page

Table C.1 – continued from previous page

Permission	Description
View and manage any of your documents and files in Google Drive	• List and search all of your documents and files in Google Drive • Download any of your documents and files in Google Drive • Create, move, copy, edit, or delete any of your documents and files in Google Drive • Share or unshare any of your documents and files in Google Drive
View and manage your spreadsheets in Google Drive	• Create new spreadsheets • View and modify existing spreadsheets • Share spreadsheets with others

D Visualization of the Communication State

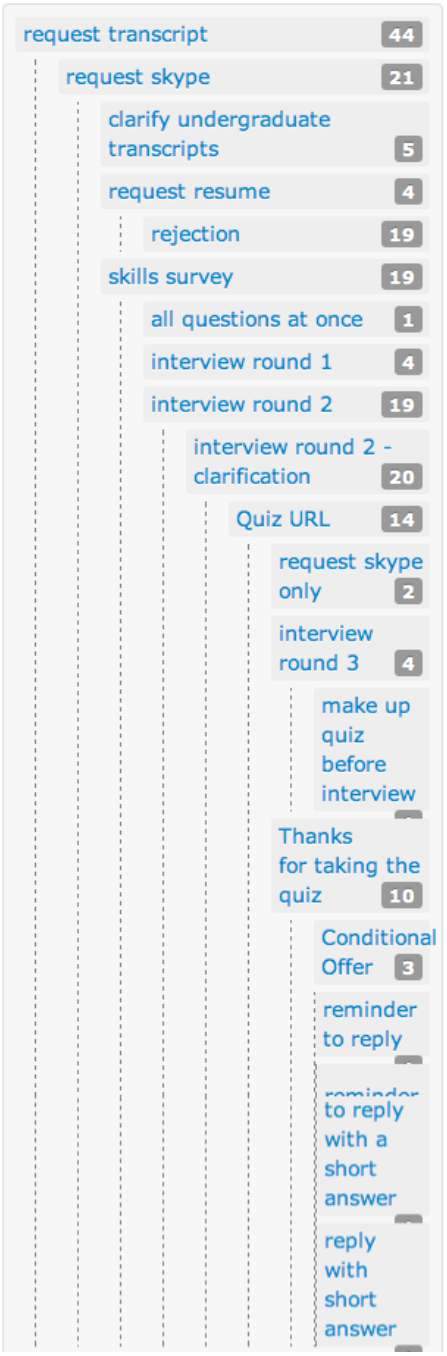


Figure D.1: A Tree Structure to Visualize a Mass Email Communication's State

Bibliography

37signals LLC (2013). Highrise tour. http://highrisehq.com/tour.

Babbie, E. (2012). *The practice of social research*. Cengage Learning, 12th edition.

Bachmann, D., Elfrink, J., and Vazzana, G. (1999). E-mail and snail mail face off in rematch. *Marketing Research*, 11(4):11–15.

Barron, G. and Yechiam, E. (2002). Private e-mail requests and the diffusion of responsibility. *Computers in Human Behavior*, 18(5):507–520.

Bernstein, L. (1996). Foreword: Importance of software prototyping. *Journal of Systems Integration*, 14:9–14.

Bogen, K. (1996). The effect of questionnaire length on response rates: A review of the literature. *Proceedings of the Section on Survey Research Methods*, pages 1020–1025.

Bradburn, N. (1978). Respondent burden. *Proceedings of the Survey Research Methods Section of the American Statistical Association*, pages 35–40.

Celko, J. (2004). *Joe Celkos Trees & Hierarchies in SQL for Smarties*. Morgan Kaufmann Publishers, 1st edition.

Collins English Dictionary (2013). Definition of mail merge. http://www.collinsdictionary.com/dictionary/english/mail-merge_1.

Constant Contact Inc. (2011). Building your email witht the advanced editor tool.

Constant Contact Inc. (2013a). Email marketing lists. http://www.constantcontact.com/email-marketing/features/email-list.

Constant Contact Inc. (2013b). Email marketing software from constant contact. http://www.constantcontact.com/email-marketing/features/email-creation.

Crispin, M. (2003). INTERNET MESSAGE ACCESS PROTOCOL - VERSION 4rev1. http://www.ietf.org/rfc/rfc3501.txt.

Dillman, D. (1991). The design and administration of mail surveys. *Annual review of sociology*, 17(1991):225–249.

Dillman, D. (2006). *Mail and Internet Surveys: The Tailored Design Method*. Wiley, 2nd edition.

Dillman, D. A. and Frey, J. H. (1974). Contribution of personalization to mail questionnaire response as an element of previous tested method. *Journal of Applied Psychology*, 59(3):297–301.

Dillman, D. A., Smyth, J. D., and Christian, L. M. (2009). *Internet, mail, and mixed-mode surveys*. John Wiley & Sons, 3rd edition.

Eley, B. Y. B. and Tilley, S. (2009). *Online Marketing Inside Out*. SitePoint, 1st edition.

Emerson, R. (1976). Social exchange theory. *Annual review of sociology*, 2(1976):335–362.

Google Inc. (2013a). Gmail APIs – Google Developers. https://developers.google.com/gmail/imap_extensions.

Google Inc. (2013b). How actions sync in IMAP - Gmail Help. https://support.google.com/mail/answer/77657?hl=en.

Group, M. M. (2012). World internet users statistics usage and world population stats. http://www.internetworldstats.com/stats.htm.

Groves, R., Jr, F. F., and Couper, M. (2009). *Survey Methodology*. Wiley, 2nd edition.

Halverson, C., Erickson, T., and Ackerman, M. (2004). Behind the help desk: Evolution of a knowledge management system in a large organization. *Proceedings of the 2004 ACM conference on Computer supported cooperative work*.

Heberlein, T. and Baumgartner, R. (1978). Factors affecting response rates to mailed questionnaires: A quantitative analysis of the published literature. *American Sociological Review*, 43(4):447–462.

Heerwegh, D. (2005). Effects of personal salutations in e-mail invitations to participate in a web survey. *Public Opinion Quarterly*, 69(4):588–598.

Joinson, A. N. and Reips, U.-D. (2007). Personalized salutation, power of sender and response rates to web-based surveys. *Computers in Human Behavior*, 23(3):1372–1383.

Kayako Inc. (2013a). Help desk software, live chat software and customer service software. http://www.kayako.com/.

Kayako Inc. (2013b). Tour. http://www.kayako.com/tour/.

Klensin, J. (2008). Simple Mail Transfer Protocol. http://www.ietf.org/rfc/rfc5321.txt.

Kokkalis, N., Köhn, T., Pfeiffer, C., Chornyi, D., Bernstein, M. S., and Klemmer, S. R. (2013). Emailvalet: managing email overload through private, accountable crowdsourcing. In *Proceedings of the 2013 conference on Computer supported cooperative work*, CSCW '13, pages 1291–1300. ACM.

Madden, M. and Jones, S. (2008). Networked workers. *Pew Internet & American Life Project*.

Madden, M. and Rainie, L. (2003). America's online pursuits: The changing picture of who's online and what they do. *Pew Internet & American Life Project*, (December).

Mavis, B. E. and Brocato, J. J. (1998). Postal surveys versus electronic mail surveys: The tortoise and the hare revisited. *Evaluation & the Health Professions*, 21(3):395–408.

Norman, N. and Lutz, E. (2000). Internet and society: A preliminary report. *Stanford Institute for the Quantitative Study of Society. . . .*.

Paolo, A. M., Bonaminio, G. A., Gibson, C., Partridge, T., and Kallail, K. (2000). Response rate comparisons of e-mail-and mail-distributed student evaluations. *Teaching and Learning in Medicine*, 12(2):81–84.

Purcell, K. (2011). Search and email still top the list of most popular online activities. *Pew Internet & American Life Project*.

Reinartz, W., , Krafft, M., and Hoyer, W. (2004). The customer relationship management process: Its measurement and impact on performance. *Journal of marketing research*, 41(3):293–305.

Repici, D. J. (2010). CSV Comma Separated Value File Format. http://www.creativyst.com/Doc/Articles/CSV/CSV01.htm.

Resnick, P. (2008). Internet Message Format. http://www.ietf.org/rfc/rfc5322.txt.

Rouse, M. (2008). What is key-value pair? http://searchenterprisedesktop.techtarget.com/definition/key-value-pair.

RubyGems Guides (2013). What is a gem? http://guides.rubygems.org/what-is-a-gem/.

Schaefer, D. and Dillman, D. (1998). Development of a standard e-mail methodology: Results of an experiment. *Public opinion quarterly*, 62(3):378–397.

Scott, A. and Smith, T. (1969). Estimation in multi-stage surveys. *Journal of the American Statistical Association*, 64(327):830–840.

Selm, M. and Jankowski, N. W. (2006). Conducting online surveys. *Quality & Quantity*, 40(3):435–456.

Sheehan, K. B. and Hoy, M. G. (2006). Using e-mail to survey internet users in the united states: Methodology and assessment. *Journal of Computer-Mediated Communication*, 4(3).

Sproull, L. (1986). Using electronic mail for data collection in organizational research. *Academy of Management Journal*, 29(1):159–169.

Sue, V. and Ritter, L. (2011). Conducting online surveys. *Sage*.

SugarCRM Inc. (2013). Email integration - gmail, outlook, lotus notes, yahoo, imap. http://www.sugarcrm.com/feature/email-integration.

Surowiecki, J. (2005). *The Wisdom of Crowds*. Anchor Books, 1st edition.

The Rocket Science Group LLC (2013a). Autoresponders. http://mailchimp.com/features/autoresponders/.

The Rocket Science Group LLC (2013b). Drag and drop editor. http://mailchimp.com/features/drag-and-drop-editor/.

The Rocket Science Group LLC (2013c). How do conditional or smart merge tags work? http://kb.mailchimp.com/article/how-do-conditional-smart-merge-tags-work/.

The Rocket Science Group LLC (2013d). How to use merge tags.

The Rocket Science Group LLC (2013e). Multi-user accounts. http://mailchimp.com/features/multi-user-accounts/.

The Rocket Science Group LLC (2013f). Reports. http://mailchimp.com/features/reports/.

Zendesk Inc. (2013a). Tour zendesk. http://www.zendesk.com/product/key-features.

Zendesk Inc. (2013b). Why zendesk? http://www.zendesk.com/why-zendesk.

Zikmund, W. and Babin, B. (2006). *Exploring Marketing Research*. South-Western College Pub, 9th edition.