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 4, 2013

(Baris Oztop)

Acknowledgements

Acknowledgements goes here

Abstract

Reaching out to large-scale of people via Internet is a fast and cost efficient 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 out emails requires additional efforts from the researchers' side. This thesis 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 supporting researchers to extract information, providing rule based automated decision making mechanism on respondents' emails, and personalize the content of the emails with the respondents' information which is extracted from the current state or earlier conversations. It also provides an option to enable contribution of other researchers as assistants to interact with the workflow under the permission of the initial researcher. Therefore, distribution of the work can ease individual's efforts on 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 by the initial researcher when it requires.

This thesis demonstrates that providing a proper workflow and the possibility of an assistant contribution, a mass email communication can be achieved as if each email is individually tailored to each recipient, which contributes to high response rates. Therefore, while it minimizes the efforts on the creation of emails, it maximizes the scale on the number of people communicated to.

Contents

Appendices   **67**

Appendix A An Email to Study Volunteer's Dilemma   **67**

Appendix B An Email Message Format   **69**

Appendix C Visualization of the Communication State   **71**

Bibliography   **73**

List of Figures

List of Tables

Listings

List of Acronyms

BCC Blind Carbon Copy

CC Carbon Copy

CRM Customer Relationship Management

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

1 Introduction

Increased Internet usage turned email into a tool for communication replacing telephone and regular mail (Norman and Lutz, 2000; Madden and Rainie, 2003). There are many use cases showing that email plays a huge role as a communication tool. Some of them include marketing for engaging clients, customer support for offering assistance after sale, surveying people to get their opinion on a topic, and many other cases showing that email become essential part of our daily life.

However, when the amount of people you want to reach increases, the way how you compose the emails and extract the information changes. Because, the personal effort will not be enough anymore to individually tailor the emails according to each recipient or reading all the respondent's emails to extract the answers that you seek for. As a result, researchers tend to use online or software tools to send out generic emails to recipients with a non-adequate personalization, which is known as one of the important factor to increase response rates (Dillman, 1991; Schaefer and Dillman, 1998). Such emails are treated with low priority, which results 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 to solve customers' problems or provide guidance regarding products. Email marketing applications help out sending commercial messages to groups of people. Finally, survey applications aid to conduct online surveys to get people's opinions and behaviors. One of the common properties of all these applications is their dependency on email communication. However, none of these mentioned tools offers a complete workflow to help out

a researcher to communicate by email with a great amount of people in a personalized manner and as easy as possible like communicating with an individual.

The goal of this thesis 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 by keeping the communication 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

Nearly 600% growth rate on world-wide internet usage between 2000 to 2012 makes Europe's 63% and North America's 80% overall 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 been increased with the introduction of smart phones and tablet devices (Madden and Jones, 2008). In addition to these facts, email has low cost and quick turnover compared to regular mail or telephone communication (Zikmund and Babin, 2006). Therefore, email as a part of communication is considered as 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 group under the following three 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 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 a relatively new trend; therefore we do not have much previous knowledge about the topic. To explore the popularity of the platform, we

need to ask the platform's users questions: 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 to improve the system or to decide its future. For example, the aggregated answers to the income level question will make us decide whether to charge the users for their usage or offer it for free and find some sponsors to make it viable.

Description   Our goal can be to describe characteristics of the online learning platform's users. The questions helping us to describe this can be: Where do they come from? What are their age ranges? Have they attended a college? At the end, we might end up with a user profile like: at the age of 16 – 22, who have never attended to a college, and coming from less developed countries. Knowing our users' portfolio according to this outcome can help us to attract organizations who have already had engagements to support those countries' young population. Hence, they can leverage our platform as a tool to reach those populations.

Explanation   We figured out that our platform's users' age range is between 16 – 22 in our descriptive study. The reasons of why this ended up like that make our explanatory purpose. The questions like how often they are connected online or have they attended a college or a similar high level education institute might help us to find out the answer of why young people use our platform more frequently than 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 a primary and mandatory contact medium, we can use email 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 has many use cases including to reach groups of people to explore, describe, and explain things. However, when the group's size gets larger, it becomes difficult from the researchers' perspective to manage

the state of the communication as in small groups. Therefore, researchers tend to write generic emails ignoring or using inadequate recipient specific information with the help of a software or online tool in the emails. This results low response rates since recipients become aware of being part of a large group, hence feeling less important and valued, as well as the chance to volunteer to reply the email gets less. On the other hand, if researchers individually tailor those emails according to recipients, it will require much more additional efforts and as a result costs, hence reducing the advantages of using email as a communication medium.

Even though, there are many solutions in the market to support email communication, there is no individual product allowing researchers to reach larger groups with minimum effort and keeping the communication personalized at the same time.

The main goal of this thesis is to show that personalized communication with large groups is possible when a proper workflow is provided. To achieve this goal:

1. Examine the workflow of an email communication with large groups and possible exceptional cases on this flow
2. Investigate the effects of email content's personalization on the response rates
3. Describe how an adequate amount of personalization in emails can be supplied
4. Comparison of existing products claiming to provide solutions on email communication and collection of respondents' information
5. Describe the design and implementation of an application satisfying the mentioned workflow to aid researchers including the initial prototype
6. Show how assistants can support the mentioned workflow
7. Real life use cases of the application and its users opinions about the application, and latest statistical information giving insight about how and in which way the application is used by its users.

This thesis also contributes on the following areas:

1. Email as a data collection method
2. Surveying with email

3. Defining a workflow on a mass email communication

4. Possible crowd sourced assistant usage

5. Personalization of email content

1.3 Outline

Outline goes here

1.3 Outline

2 Foundation and Related Work

This chapter presents the related work on the data collection domain. Even though, the technology is different for email surveys to collect data from well-established regular mail surveying methods, the nature of the communication is similar to self-administrated questionnaires (Schaefer and Dillman, 1998). Therefore, the chapter will also investigate the mail surveys in a way to emphasize the points which are also related with email communication, and the earlier studies on 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 belong to the data collection process. Therefore, data collection is a fundamental step to produce useful data to enable analyzes on researches (Groves et al., 2009, page 149). These researches include but not limited to many disciplines like sociology, statistics, psychology, marketing, economics, and heath 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 the postage, paper and printing, and interview costs (Schaefer and Dillman, 1998).

2.1 Surveys and Data Collection

Sproull (1986) identified the characteristics of email with an 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 are 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 answer open-ended questions either electronically or in writing.

The result of the study indicated that the average duration of data collection time 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 comparing with the written questionnaire.

In another study from Sheehan and Hoy (2006), where they administered only an email survey to query individuals about their online behaviors and their attitudes and opinions regarding privacy. They have reached the shortest response time with 3.65 days comparing 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 speed of the email surveys, cost benefits have been indicated in Sheehan and Hoy's (2006) study also concluded that email is an extremely cost-efficient method for data collection, where the total cost estimated at $470 ($30 for printing out the responses, $440 for 22 hours computer time to download surveys for printing) while postal mail is estimated at $6,500 (printing, postage, survey, and reminder mailing).

In another study from Mavis and Brocato (1998), the email survey was nearly seven times cost efficient than postal survey. This includes labor hours, survey materials like booklets, mailing labels, envelopes, and postage costs. Total time spent into postal survey was 33 hours, but it only required 12 hours for the email survey. Final cost was $503.36 for postal survey, whose $305.36 was spent for postage part, and remaining $198 was spent for student labor cost. The only cost resulted from email survey was student labor cost, which was total $72.

Moreover, Paolo et al. (2000) reported that people made longer open-ended response comments in email version of the survey compared to the mail version. While the average number of words per comment was 58.33% in the mail version, it was 75.40% in the email version. Bachmann et al. (1999) had the same finding in 1995 and 1998, where open-ended questions were responded more likely by email recipients than mail recipients. In

the latter study conducted in 1998, researches also found that email respondents were more likely to expand their answers, even it was not suggested by the survey, resulting in more candid responses than mail surveys. Responses to open-ended questions are one of the important measure to determine the quality of the returned surveys.

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

2.1.2 Survey Errors

Sample surveys are quantitative estimation 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 more number of people surveyed, the larger degree of precision can be achieved. Therefore, the 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). The surveys relying on predefined list of recipients considered that the list is randomly generated or with a systematic sampling. Hence, it has got little research to reduce sampling errors comparing 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).

Coverage Error   When the list of surveyed people does not include all the elements of the population, coverage error happens (Dillman, 2006, page 9). Coverage error is consid-

---

[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).

ered one of the biggest issues of surveys since while surveying general public (Dillman, 1991).

Measurement Error   When a respondent's answer is hard to evaluate or cannot be compared with 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 reasons might depend on poor wording or 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 are large amount of people who do not response, and their characteristics are different from the ones who responded, then it results nonresponse error (Dillman, 2006, page 9). Low response has been considered 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 nonrespondents, which will result in a biased estimate of representation of the population (Bogen, 1996).

Low response rate was even considered 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 show high response rate on postal mail, three of them got 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 which got an acceptably high response rate by email. Schaefer and Dillman (1998) attributed this fact to the novelty of email and sent emails were carefully examined instead of considered company junk email. Mavis and Brocato (1998) stated that studies cited by others in support of email surveys, also shown in table 2.1, did not

compare email data collection with more traditional methods, and their study design and analyses varied greatly. Sheehan and Hoy (2006) also take the attention to many of these studies' small and homogeneous population, therefore it may not represent larger population groups' response tendencies.

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

2.2.1 Length

For many people the time required to spend on survey 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) suggests that the length of the survey is correlated with its importance, therefore it will increase the efforts both on researchers and respondents side resulting a higher response rate. Bogen (1996), in his literature review, concluded that the relationship between interview length and nonresponse is weak and inconsistent.

2.2.2 Multiple Contacts

Researchers found that the number of attempts to contact people increases the response rates (Heberlein and Baumgartner, 1978; Schaefer and Dillman, 1998). The scenarios for multiple contacts include pre-notification contact, which is a brief notice for the main request, and follow-up contacts aiming to the people who did not respondent at 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 with 11.9% and 10.0% for the second and third contacts, respectively (Heberlein and Baumgartner, 1978). Schaefer and

Dillman (1998) also stated the same conclusion 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 to increase 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 alumni sample via personalized cover letters, while the other half got impersonalized letters. The personalization treatment included personal salutations and real signatures on the mails. They 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 applications of personalization in emails, and give the results of some studies.

2.3 Personalization of Emails

Studies on mail surveys showed that personalization 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

---

[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).

phenomenon, the diffusion of responsibility, which is also an outcome of volunteer's dilemma. In the volunteer's dilemma one player is needed to volunteer in order to reach the outcome preferred by all the others in the game. However, each person might be inclined to hoping that somebody else will volunteer, resulting in a higher utility of not volunteering than volunteering. According to this, the more people in the group size, the less probability of volunteering will result, which produce the diffusion of responsibility effect. In order to experiment the effect of diffusion of responsibility in the context of email requests, they sent emails asking for help either to single addresses 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 was well known to anyone familiar with the institute. The result of the study showed that the proportion of replies where they used single email address in the "To" field got 20% higher response than the replies where they used groups of email addresses. In addition, the study qualified the given responses according to its helpful level, and the proportion of "very helpful" replies in the single email address condition was 187% higher than the groups of email addresses condition.

Another outcome regarding using multiple email addresses in "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 remarked his concerns regarding the confidentiality when the header of the email contained all the email addresses of the other respondents explicitly. His reaction was quoted in the study as in listing 2.1.

"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 entitled that person as "skeptical" and his reaction as a "vivid skepticism", today it is one of the biggest concerns regarding email confidentiality, and it might result in embarrassing situations from the research or the business perspective. A very recent email message (See listing 2.2 for the excerpt) dropped in 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 was applied to the salutations in the emails. The randomly drawn 2,540 samples from the student database of Katholieke Universiteit Leuven, Belgium were separated into equally sized two groups. In 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 6.9% higher login rate to the survey than the unpersonalized group. Therefore, they concluded that increased response rates were in line with social exchange theory and with the diffusion of responsibility theory.

In addition to personalization of salutations on the emails, Joinson and Reips (2007) stated the power of its combination with the power or status of the sender. In the study, a group of discussion panel students of Open University UK were sent an email invitation to complete a survey. Panel members were assigned to one of the conditions where salutation was modified in "Dear student", "Dear John Doe", and "Dear John". The sender power was 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 lowest when an impersonal one came from a neutral power source (See table 2.2). The possible reason for this was stated as personalized salutations increase people's sense of identifiability, and its combination with a high power audience increase socially desirable, 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)

As aforementioned studies show different forms of personalization increase the response rates in email communication. However, it has become very easy to add personalized information into email thanks to the software. Dillman et al. (2009, page 237-238) stated that over-personalization using software tools might easily result impersonal messages, and gave an example (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 overwhelmed personalization with the usage of person's wife, their dog's type and name, and their home address. Moreover, experienced email users can identify if a message is written by a person or computer generated by looking appearance of one's name in certain locations, and similar patterns for other information (Dillman et al., 2009, page 272). Therefore, it becomes difficult to have a correct amount and tone of personalization. The more daily interaction with digital devices will make the true authentic personalization more rare, hence achieving it will make it more important and effective (Dillman et al., 2009, page 238).

2.4 Conculusion

In conclusion, researchers conducted more mail survey studies than other survey methods to investigate data collection (Dillman, 1991). Some of those studies tried to answer the question of the nonresponse error, which has been considered one a major problem compared to other survey errors as discussed in section 2.1.2. According to those mail survey studies, personalization has been addressed as an important factor to increase the 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, researchers has been started to consider email as a data collection method, because of its cost and speed benefits compared to other data collection methods as discussed in sec-

tion 2.1.1. However, some studies showed that response rates on email surveys are lower than on regular mail surveys despite to its advantages; in addition, it may pose a burden to 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 the technology for email is different from mail surveying methods, researchers considered the response rate influences of mail surveys for email since the communication itself is the same. In section 2.3, several studies applying different types of personalization are 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 in increased on response rates to the emails. On the other hand, those studies did not consider the increased awareness of recipients to the possibility of computerized personalization techniques, which results in over-personalized emails. As well as, none of the studies has taken the attention to the personal efforts of a researcher while extracting information from respondents' answers. This thesis will try to focus on those shortcomings of those studies as well, and provide a web application to overcome those issues.

In the next section, existing applications in the market, which leverages the email communication, will be evaluated. While some of those are focusing only email communication, as in email marketing applications, other applications like CRM and help desk applications helped this thesis to identify the useful features that can be helpful in the area of personalized email communication.

3 Evaluation of Existing Applications

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

3.1 Application Categories and Their Relation with the Thesis

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

3.1.1 Customer Relationship Management (CRM)

A CRM application helps to manage customer relationships effectively, which is a topic studied both by academia and industry in recent years. Such applications play an important role in the marketing where organizations use more customer oriented instead of product or brand oriented marketing strategies. Therefore, each customer's economic value is different to the company, and 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 thesis considers evaluating CRM applications is its communication aspect of a company with their clients. Another reason is, as it is mentioned at section 2.3, the adequate amount of personalization in emails is crucial on response rates, and people's increased daily interactions with the digital world make the true authentic personalization more rare. Achieving such a level of personalization requires getting to know each recipient very well by considering not only the recent conversation, but also

earlier conversations. All the information that might be extracted from those conversations helps to build a relation with the respondents. Since a CRM system aims to keep track of each customer history regarding a product or a brand, such a data store could be leveraged to add an adequate amount of personalized information to email conversation.

3.1.2 Help Desk

Another application form that focuses on a company and its relation with their clients is help desk applications. Its main purpose is to provide information and support related to a company's products and services to their customers. As a part of knowledge acquisition, help desks support both sides of the communication in a way that while 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 the existing knowledge requires to structure the captured knowledge. This is where it makes the relation with this thesis. Because, a help desk application provides a workflow where both parties develop a communication where a person who needs assistance describes his/her problem while people who provide help identify the problem by looking earlier cases or asking questions to clarify the initial question. This also requires the cooperation of assistants while providing help to a problem at which one person might have previous experience to guide other assistants. As a result, a help desk application is similar to a mass email communication where a researcher initiates an open ended questionnaire, extracting 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 some questions to clarify things, where existing answers can easily be reused. Having such a email conversation with large groups requires great effort from a researcher, so he might assign tasks to distribute the efforts to other researchers to deal with the large size of the group.

3.1.3 Email Marketing

Organizations and marketers use email marketing for several reasons. Some of those purposes are 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, 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:** To notify the customers about important updates or changes to a business. For instance, the release of a new product, changes on contact details or major changes on a company's website

- **Direct Sales Messages:** Emails sent by others consists of marketing ads, and clear messages on offers.

- **Housekeeping:** Emails such as subscription confirmation messages or welcome emails. These messages are often system generated automated messages. However, they can be used to promote a message as well like offering a discount code along with the registration confirmation email.

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

3.2 Methodology

The analysis examined two products from each of the categories that are 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 suggestions 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 required a fee before using them. After the products were shortlisted, the last

---

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

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

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

filtering was done by getting their web traffic rankings from Compete.com[4], Alexa[5], and Google Trends[6]. Finally, 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 done according to their category. A brief description of the products will be presented. This description will mainly focus on the features, which are related to support email communication as explained in section 3.1. After that each category will be concluded with a comparison matrix of the selected products.

3.3.1 CRM Applications

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

---

[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 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	Require additional module installation
Email Client	Build in, allowing 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. These are 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 platform, which supports iCalendar[7]. It provides email management right in the application, and integrates with several platforms like Outlook and Gmail or an Internet Message Access Protocol (IMAP) based email server. Users can archive emails in the SugarCRM by adding a unique email address into TO, Carbon Copy (CC), or Blind Carbon Copy (BCC) fields. This address can also be used to link email recipients information including email attachments with SugarCRM by simply forwarding the emails, therefore it removes the additional effort to import them into SugarCRM and reduces dependency on a platform. SugarCRM also comes with a build

---

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

in email client. Even though, its inbox view only provides basic functionalities, its email creation view goes a little further to support email marketing by providing dynamic variables that can be embed into an email's content, and can be replaced with actual values available in SugarCRM. For example, a variable for first name will be replaced by contact's actual first name while email is being sent (See figure 3.1).
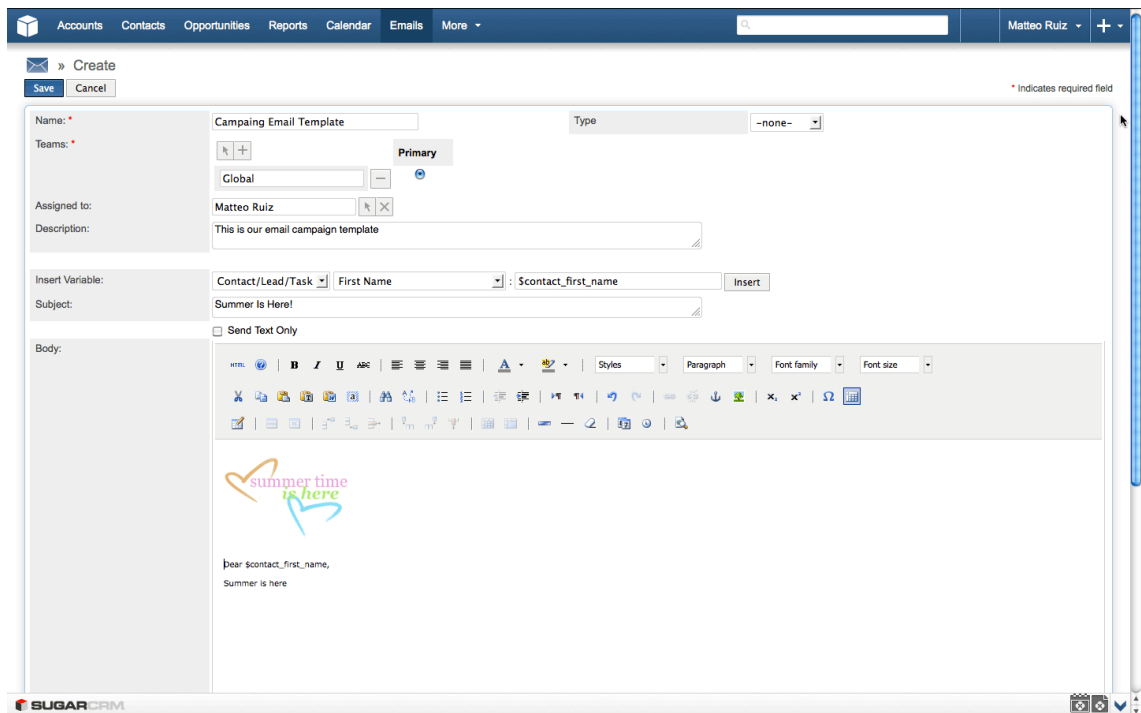

Figure 3.1: SugarCRM Email Composer with Embeded Variables (SugarCRM Inc., 2013)

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

Highrise   Another SaaS is Highrise[8] offering several purchasing plans with 30 days trial period. It has a simple UI like SugarCRM, but also provides quick access buttons to add a task or a contact. Task management differentiates it from SugarCRM since there is

---

[8]http://highrisehq.com/

no calendar view, but a task view in Highrise. These tasks 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 forwarding an existing email creates 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 the 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 by simply forwarding it to the provided unique email address. If a user does not exist in Highrise when an email from him/her forwarded to link, a contact profile is created using available information in that email. Adding tags to the contact profiles also makes it easier to organize contacts and browsing within them. Highrise does not offer any email composer to do email marketing as in SugarCRM. Therefore, users will depend on another product to do simple campaigns. The provided activity view helps users to keep track of their own or other users' recent actions within Highrise. Lastly, it offers options to customize the look and feel of the application by provided color schemes (37signals LLC, 2013).

3.3.2 Help Desk Applications

The two help desk applications investigated in this thesis are Zendesk and Kayako. Table 3.2 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, and free trial	$29 – $49 user/month, and free trial
Channels	Website, email, phone, and social platforms	Website, email, and only 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 via many different communication channels including website, email, phone, and social platforms like Facebook and Twitter. Hence, support requests coming from those platforms can be turned in to a support ticket. Those support tickets can be grouped under categories, and further classification can be done via tags for each ticket. Those feature also help to find 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 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 to create a knowledge base. Customer ticket history and basic personal information are kept in the system. However, it does not support to add additional fields to customers' profiles. In addition to the desktop version, Zendesk has a mobile version for smartphones and tablet devices. Therefore, support teams have no dependency on a device. Lastly, provided analytics view by reports give an overview of customer satisfaction and performance of the support team (Zendesk Inc., 2013b,a).

---

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

Kayako   Kayako's[11] complete solution for customer support is named as Kayako Fusion. It comes as software and SaaS. Comparing with Zendesk, its UI seems more complicated. Kayako got more than 30,000 clients within ten years. Kayako does not have social platforms integration like Zendesk, therefore support tickets are generated over website, 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 a click. Kayako also keeps basic information of customers, if they are registered to the system. Registered customers can also support to build a knowledge base in a forum-like environment by contributing others questions along with the support team. Kayako does not have any native app for mobile platforms like Zendesk. Finally, it has a analytics view to keep track of ticket reports, measuring customer satisfaction and support team performance (Kayako Inc., 2013b,a).

---

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

3.3.3 Email Marketing Applications

MailChimp and Constant Contact are the selected two email marketing applications to analyze in this thesis. Table 3.3 shows their features side by side to give an overview. The details are provided in the upcoming 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 fixed cost monthly plans or pay as you go plan. Along with its intuitive UI, it offers a drag and drop functionality on the email content creation. It supports email marketing processes starting from designing the sign up form, so users can add all the desired fields, and apply brandings to it. The recipients' list can be filtered out with several conditions like campaign name, location, or ratings assigned by application's user. There are different levels of access privileges to MailChimp. Hence, a person who has an "Admin" account can grant permissions different type of permissions allowing different access levels to MailChimp. This allows a separation of tasks in mail marketing. For instance, while managers manage recipients

---

[12]http://mailchimp.com/

list, an author team can focus on emails' content and design (The Rocket Science Group LLC, 2013e). To design email content, users can pick one of the available templates provided by MailChimp or create their own HTML templates with its drag and drop editor (See figure 3.2).
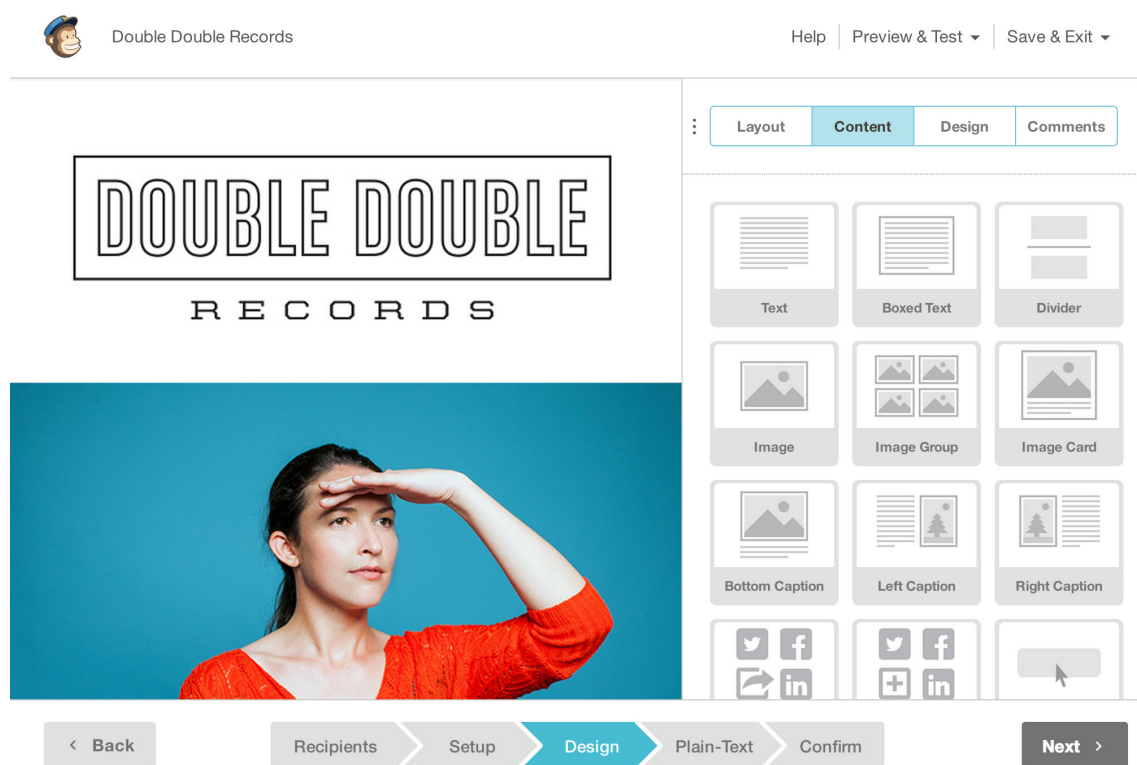
Figure 3.2: MailChimp Drag and Drop Content Editor (The Rocket Science Group LLC, 2013b)

The template editor also provides a photo editor, and authors can add comments to give feedback on the content and design of the templates. Created templates can be previewed as if they are viewed in a software or an online email client, or even a mobile browser (The Rocket Science Group LLC, 2013b). As in SugarCRM, MailChimp allows you to use dynamic variables, called merge tags, in the email content. Therefore, send out emails can be personalized with information specific to the recipients. However, it provides more different type 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 the US state of recipients.

```
* | 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 offers auto-response based on a triggering event. These events can be a link clicked in the email, being on a specific date like birthday of a contact, or scheduled dates. Finally, there exists an analytics dashboard 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 Constant Contact Email Marketing[13], whose purchase plan depends on the amount of contacts you have, but there is a free 60 days trial period as well. It offers drag and drop content creation like MailChimp with a clean UI. It has quite many templates to pick and start customizing it. However, users can embed sign up forms into websites or Facebook. Recipients list can be imported from several sources including Excel, Outlook, and Gmail. In addition, recipients can be grouped under sublists which can also be merged into each other easily. An option is available to remove duplicate contacts from the lists or delete recipients who unsubscribed from the list. Users can track opens, clicks, forwards, and social platform shares of their email campaign. On the other hand, it does not offer any sophisticated email variables to be replaced by actual content from the application (Constant Contact Inc., 2013b,a, 2011).

---

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

3.4 Conclusion

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

- Contact User profiles kept in CRM applications can help a researcher to get to know their respondents better and to identify basic attributes like name, gender, address, and phone numbers. However, those fields were limited with fixed fields in those applications.

- Importing contacts information from other popular software, e.g. Outlook, can ease the time to create recipient lists for email communication. Again as SugarCRM and Highrise supports, importing an email into the system by just forwarding it to a specific email address can make researchers life easier in the same way. Providing such flexibility will reduce the dependency on a platform, therefore while researchers continue to use their email clients that they are familiar with, they can switch to another platform when it is necessary.

- Both CRM applications provides a module to create tasks, so this can be helpful to remind researchers what is the next high priority thing to do about an email campaign. This can be a task showing what is next to do in an email campaign initiated by the researcher. For instance replaying the email in which the respondent has asked a clarifying question at our initial campaign.

- Both CRM applications support archiving of emails by simply forwarding them to a provided unique email address, and linking those emails to the users' profile. This can be helpful to see important conversations happened with respondents earlier time, so it can provide content or an opinion about 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 not to write them again. As we have seen, SugarCRM also allows saving emails as a template to reuse them again. However, there was no filtering mechanism or similar functionality, but just remembering the given name of the template to help users find the corresponding template to reuse.

- It is not always the case that a researcher initiate an email communication. It might be the case that a high amount of email can be dropped into the researchers inbox. For example, students attending a course may ask questions regarding their home-

work. In that scenario, the same question might be asked several times. Help desk applications provide a ticketing system for customer related issues, which is also applicable to the mentioned homework 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 to identify conversations belonging to each other in a situation where a researcher initiated more than one campaign. However, there was no visual representation of the state of the communication of a support ticket, but just status labels like "resolved" or "assigned".

- Another feature of help desk software is a support ticket that can be shared or assign anyone by anyone from the support team. Hence, this will decrease the answering time of those tickets. This can be also useful in a mass email communication to share the responsibility to reply or extract information from incoming emails.

- The email marketing application MailChimp provides dynamic variables that users can add into email content and its variable will be replaced with actual values. Such a feature can be helpful in personalized mass email communication, where it is difficult to add recipient specific personalized information into emails. However, there was no attached information regarding those variables to show users in which state of the communication they are extracted, and again they were created separately in an additional view where users are away from the actual emails where they can extract information.

- MailChimp also provides different type of permissions to leverage in an email marketing task. For instance, while an author can create the email content, a viewer can just follow the reports to see what is the success rate of an initiated campaign. Such functionality can also be helpful in mass email communication, where some users can extract the information from emails, others can reply the emails.

- Both email marketing and helpdesk applications provide analytic reports to keep track of the success of a campaign or a support team. That is a useful function in a mass email communication as well to get a quick overview of the 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 doing all the men-

tioned 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.

3.4 Conclusion

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 its drawback will be discussed.

4.1 Mass Email Communication Concept

When a researcher initiates a mass email communication, there will be many cases that cannot be predicted beforehand, but they affect the flow of the communication. Respondents may come up with clarifying questions to an initial question, email addresses that are considered active might not exist anymore, an auto responder might have been set since the respondent is not available during a temporary period, and many more cases.

4.1.1 A Mass Email Communication Scenario

In section 1.1, the online learning platform scenario was used to illustrate the possible reasons for email data collection. If we will consider it in here again, assuming that we are exploring our users to improve our platform. Therefore, we start sending out emails to our registered users to get their permission 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 to improve our platform.

Kind regards,
your online course team.

Listing 4.1: Email Invitation to Questionnaire for Online Learning Platform

4.1 Mass Email Communication Concept

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

1. Yes, I would like to involve

2. No, I am quite busy

3. I could involve, but I am busy till the end of this month

4. An automated message from recipients to indicate his or her vacation period

Figure 4.1 shows a simplified version of this scenario. The people who accepted our invitation will get a second email including the questionnaire. As soon as they respond back with the answers of the questions, we will return a thank you email for the participation. Those who did not accept the invitation will get a motivation email to encourage them for the next time. Next, we will group the people who returned a conditional acceptance under a "maybe" case. Again, there might be several reason for them for instance they might involve in the the questionnaire but they are busy during that week. Hence, we could reply with an email telling an opinion for their situation like if people attend other exams, we can wish good luck for it, and afterwards we can mention about a reminder email that will be send out after a time period that the participants mentioned in their response. After the reminder email, we expect to get a acceptance of the questionnaire email. Then conversation will continue as with the people who initially accepted to involve our research. The communication with an "auto-respond" case is not clear, therefore the communication will not go further with those people until they reply us in the near future. However, if the auto respond provides some information, we could also set a reminder for an upcoming date to send a reminder email.

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

- Figure out the courses that participants attended for our email list, and add them to the content of the email for each person.

- Write a generic answer for those who accepted to involve in the questionnaire, and write the questions.

- Write a generic motivation email to support those who did not accept to involve to
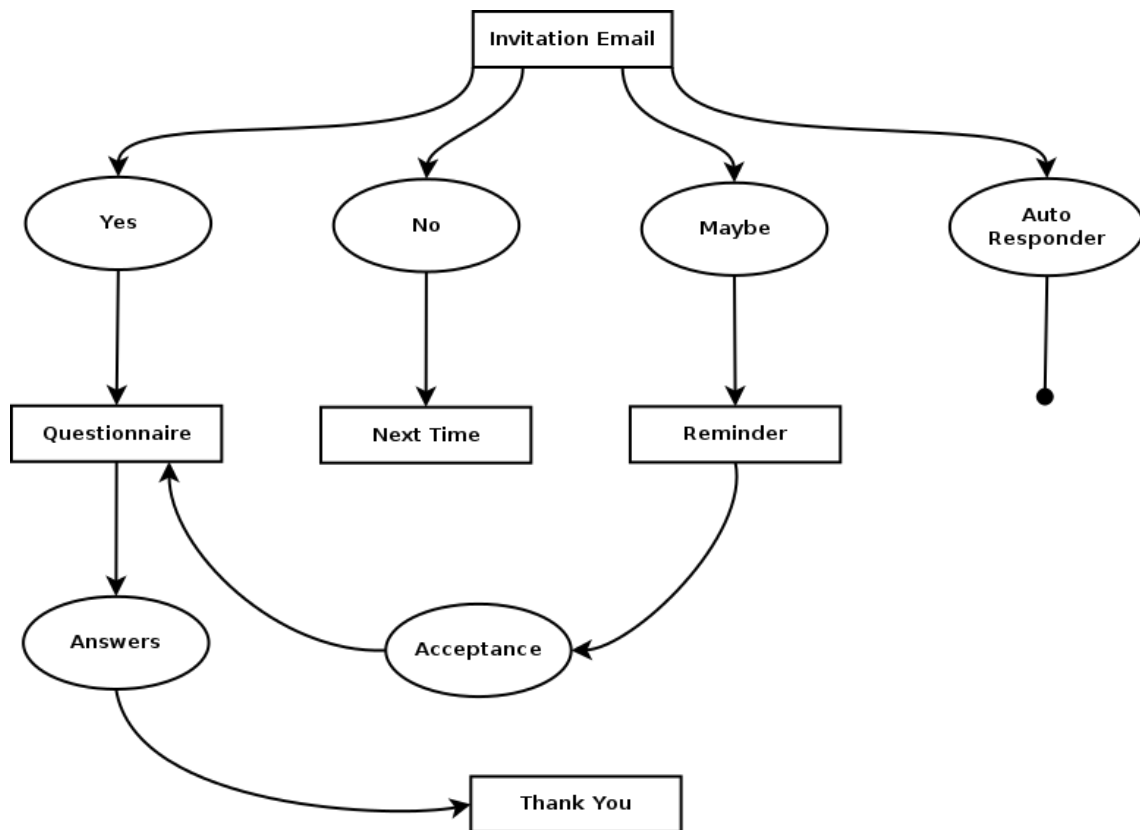
Figure 4.1: An Email Campaign Scenario

the questionnaire.

• Go over all the "maybe" cases, and record the date information if they mention a time period that they will be willing to participate, but busy at the moment.

• Write a generic email to those "maybe" cases mention about the reminder that we will send in the time period that they mentioned. This email can involve a couple sentences to give a feedback and share the feelings about the respondent's situation that make them not to involve to the questionnaire at the moment.

• Go over for all the "auto-respond" cases to figure out if there is any condition that we need to ignore those people at all for the upcoming questionnaires in case they change their email addresses, or any date that can be recorded to setup a reminder email later on.

• Send a thank you email to those who answered the questionnaire.

However, when the amount of participants is great, doing this manual work will get harder, and the personalization that is added into the emails will get less. For example,

for the "maybe cases", we will not be able to give a feedback or share the feelings about respondent's situation, but just write an email to mention that we will send them a reminder later on. More over, these are simplified possible cases. There might be much more cases in real life situations, which might require to ignore some respondents totally or to write a generic email by again ignoring the context of their email since individual efforts will not be enough to write custom emails for those cases. In such a case, we might even terminate the communication forever, or results in the side effects of diffusion of responsibility theory as mentioned in section 2.3.

In the figure 4.2, you see that the personalization of emails is directly proportional to the effort of composing those emails. The shaded are on the graph represent the situation, when researchers use an application to create mass email communication, while the area marked with a star represents the gold standard, really hard to reach with the available applications.
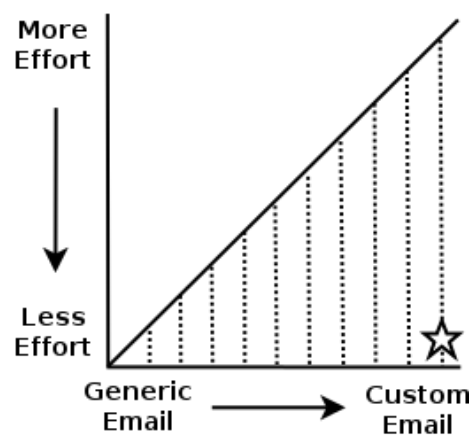
Figure 4.2: Effort vs Personalization on Composing Email

In the next section, we will see that how can we help researchers to reduce the efforts, and increase the personalization of the 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 in this section to reduce the efforts on the researchers'

side, as well as increasing the personalization in the content in some degree.

Contact Information   The first step would be to import contacts, and their relevant information. These could the information that we asked for during the sign up form when they registered to our learning platform e.g. first and last name, email address, and gender. Therefore, we could easily export those information into a file format like Excel or CSV[1] to make it ready to use to initiate our mass email conversation. The application that was discussed in section 3.3 offers several options to import data into those systems. Hence, researcher will not need to enter all those information manually into those systems.

Dynamic Variables   The invitation email included user specific data, which are the name of the recipients at the salutation, and the courses that he or she attended. As it is hard to write those fields individually into all the emails, we could apply dynamic variables insertion approach that is used by both email marketing applications (MailChimp and Constant Contact) and one of the CRM application (SugarCRM) analyzed in section 3.3. However, the limitation of SugarCRM was that it only allows us to record a limited amount of contact profile fields. This could help us only to dynamic generation of the salutations in the emails not the course list of the recipients. On the other hand, Constant Contact and MailChimp have the feature to add custom fields into a contact profile, and dynamically use them in the emails' content. So, this could also remove the manual effort on the initial invitation email with an email template who has some dynamic fields to be replaced by the application while sending the emails. Those dynamic fields also add some degree of personalization into the emails.

Information Extraction   When we start getting answers from respondents, we might need to extract some information from those emails. For example, this could be a date that we need to return to the recipients with a reminder as it is discussed in section 4.1.1. Storing those information as we stored 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.

---

[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).

Reusability of Existing Emails   After the initial invitation email, there will be many cases to reuse different versions of an email for further states of the conversations. We mentioned the term "template" as an email consisting some dynamic variables in it. Later on, those variables will get their actual values, and it will be the actual message that we want to send. Here, we will add another meaning to the term "template" to emphasize a message's reusability. Let us consider the "maybe" case that was mentioned in section 4.1.1 to represent respondents who were too busy to participate in the questionnaire at that moment, but maybe it is possible for them to participate in later (See figure 4.1). In those responds, recipients will come up with a date that is possible for them to participate, or an unclear answer about the availability of themself. It is definite that we cannot ignore if they provide us a date, and send them a reminder before that date to participate in the questionnaire in a generic email. This will make them realize that the messages that we send are automated, and it will result the drawbacks that we discussed in chapter 2. However, if we create two email templates: one with a dynamic variable in it to mention when we will send a reminder, and another one with a more generic content including a time far enough to encourage them to participate in the questionnaire. Therefore, when we encounter different "maybe" conditions, we have now two ready to use email templates to reply them.

Visualization of Conversation State   At this point, we begin to have several email templates for different conditions to ease a researcher's life by allowing reusability on the replies. However, it has also started to be difficult to pick a template to use since the state of conversation get advanced, and we have had many email templates for different conditions. This will result in additional effort to choose the proper template. Besides, we will not need some of the templates after conversation's state get further, and use the latest ones more frequently than the older ones. Therefore, we need a view to show us what the current state of the conversation is, as well as 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 wrote are collected in a hierarchical way starting with the first invitation email at the root node. After the researcher started to get messages from re-

spondents, all the following answers that we can write are at the second level of the tree right after the invitation email node. When ever we get an answer for the questionnaire, we write an email to say thank you. This is also represented in the third level of the tree, right after the questionnaire node.

```
Invitation Email
├──Questionnaire – Graduate Version (GV)
│  └──Thank You GV
├──Questionnaire
│  └──Thank You
├──Questionnaire After Reminder
├──Next Time
├──Next Time with Motivation
├──Reminder with Date
├──Reminder Generic
├── ⋮
```
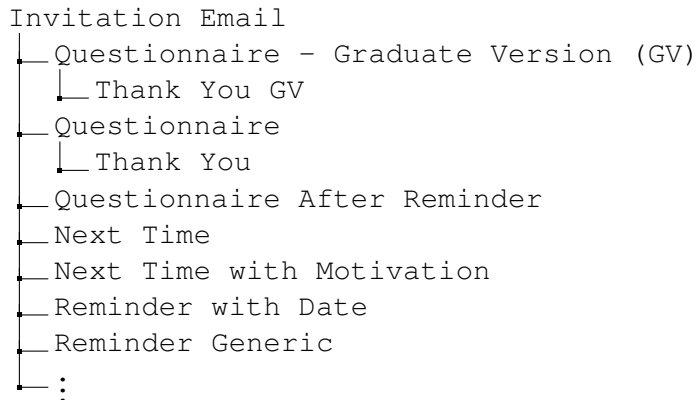
Figure 4.3: A Sample Conversation State and Template View

A view like in figure 4.3 also help researchers to let them pick a template to reuse according to the email of the respondents. Therefore, they can easily see the previous answers that they gave, and the level of the tree provides an idea on the overall state of the conversation. When we consider existing applications in the market in section 3.3, none of them provides such a 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 gathering to figure out how can we ease the efforts of the researchers, while keeping the personalization in the emails high, the next section will show how this initial idea was applied into an application.

4.2 Prototype

A software prototype provides developers a better understanding of the requirements, what could be developed with existing technology, and by allowing client interaction, it demonstrates them what is functionally feasible, and revise their imagination accordingly, which results in better inputs from them and more forward looking systems (Bern-

stein, 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 to improve a researcher's personalized mass email communication. 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 recipients' first names, last names, and email addresses. The application must store those information under a contact list permanently.

2. Application user can use variables in the email message for a recipient's first name and last name. Therefore, application must allow user to add those dynamic variables into email's content to personalize the salutation of the emails.

3. Application user can extract information from an email message into a KVP[2] while reading responds. Therefore, the application must offer an option to get KVPs from users. At the end, each contact's record will have enough information saved in KVP to describe all details that we know about them.

4. The application must allow users to use those extracted KVPs as dynamic variables in the email content. This will allow us to personalize the email with the information specific to each recipients.

5. User can reuse a sent email as a template to reply to respondents or initiate a conversation for a group of people. Therefore, application must offer a view to pick the templates from a list.

6. Application must show the whole state 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 that the user created, and have a given template name by the user.

7. Application must offer the template selection and conversation state representation under one view. Therefore, user should be able to pick the templates to reuse by

---

[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.

looking at nodes in the tree to figure out the state of the conversation .

The next 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 in the evaluation of the prototype.

4.2.2 System

Initial system was built on top of an existing project named as EmailValet[3] also known as GmailValet in Human-Computer Interaction (HCI) group[4] at 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 on the initial requirements as discussed in section 4.2.1. From the technology point of view, we were also limited with the existing EmailValet's architecture, which was developed in a Web framework Ruby on Rails (RoR)[5], and a JavaScript framework jQuery[6].

EmailValet requires a Gmail account to register and sign in[7] to the system, therefore 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[8]. On the left hand side, there is the email list pane, and a selected email can be read on the reading pane, at right.

---

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

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

[5]http://rubyonrails.org/

[6]http://jquery.com/

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

[8]In this thesis, the word campaign is used as a shorhand for a mass email communication campaign, where a researcher initiates a conversation with groups of people related to one specific topic.
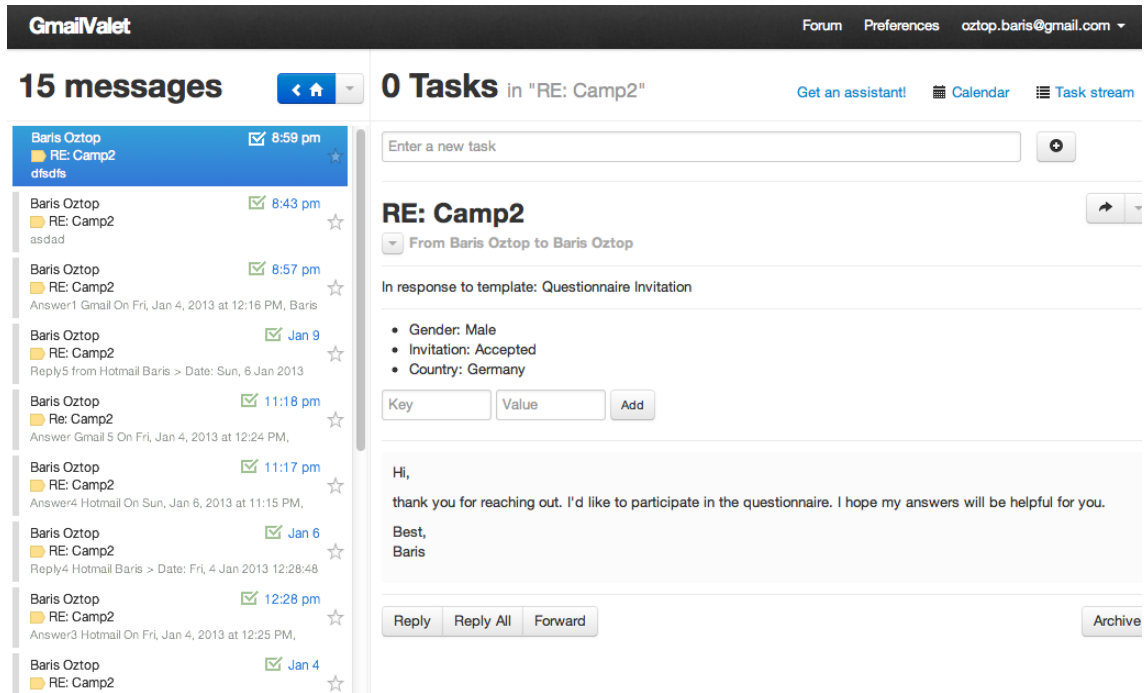
Figure 4.4: EmailValet Inbox View and Message Reading Pane

Starting a campaign is done exactly the same way as in composing an email. However, the corresponding view has additional two input fields to get the campaign's and the template's names from the user. A campaign name will help to identify the groups of conversations from other campaigns, or from regular email conversations in the inbox. A template name will help us to identify which respondents's answers correspond to which emails that we sent. For example, in figure 4.4, the recipient's answer is response to the questionnaire invitation that we sent earlier. A template name will also help us to represent the template in a tree structure to pick from to reuse, and show the state of the conversation in the same tree structure.

Researcher can add their recipients list into the "Merge"[9] input field as in the figure 4.5, which corresponds to the "To" field in a regular email client. However, the format should be as in the figure, that is the first name, the last name, and the 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 that recipients list

---

[9]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).

is entered, and the email is sent, those contacts will be recorded in the Gmail's contacts book.
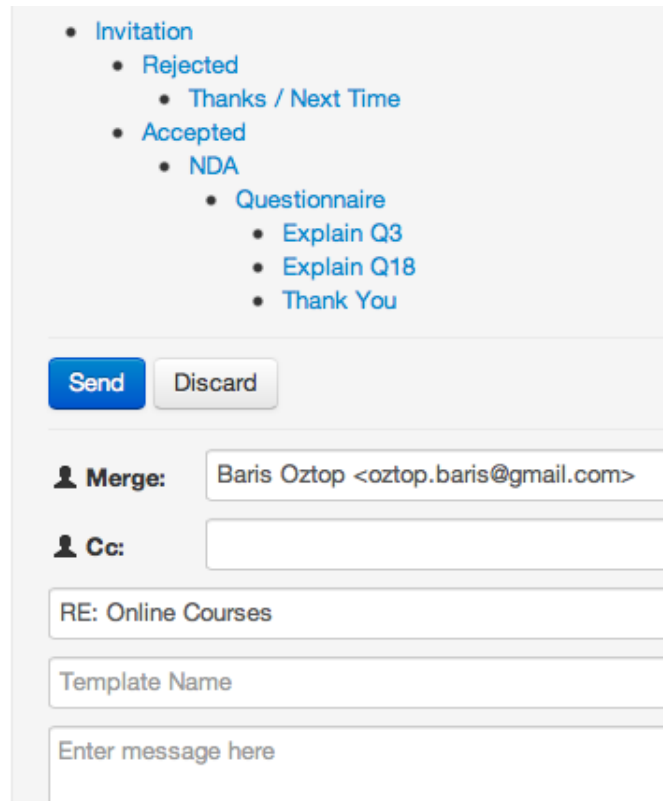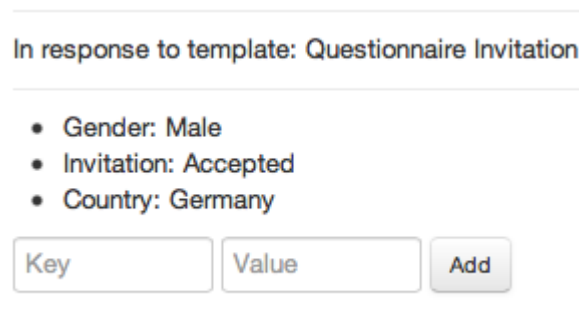
Figure 4.5: Email Template Selection to Reuse Earlier Replies

Figure 4.5 also shows the template tree to represent the state of the conversation, and allow users to select a template from. The indentation at the nodes of the tree helps to identify which templates are used after which one, and therefore it gives an idea about the current state of the conversation. According to this sample scenario in the figure, right after we sent the invitation email, and started to get responds, there were two possible answers that we gave to the recipients. For those who rejected the invitation, we can select the "Rejected" node from the tree to reuse that answer that we gave before. This can be a motivation email as discussed in section 4.1.2. At the parallel, the conversation continued with the people who accepted the invitation, and we continued to the conversation by sending Non-Disclosure Agreement (NDA), and afterwards the questionnaire. After we sent the questionnaire, there were cases, where respondents returned back with an email to ask some questions to clarify the questions in the questionnaire.

According to figure 4.5, those cases were related to the questions number 3 and 18, and we explained those two questions of the questionnaire. Again, since the system allow us to reuse the our previous answers as a template, we don't need to rewrite our explanation if those questions were not clear for upcoming prospective people. On the other hand, if researchers want to create a new template, he or she can simply add the name of the template into the corresponding input field, and send the email. The next time, the researcher will find it in the template tree under the corresponding level of the node. Also, it is possible to select a template, made some modification, and save it as a new template. Therefore, a slightly different versions of 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 offered to enter a key and its value corresponding to the extracted information from those emails while reading them. As in the figure 4.6, while a user can add new KVPs, they can also see the existing ones added earlier to that email message.

In response to template: Questionnaire Invitation

- Gender: Male
- Invitation: Accepted
- Country: Germany

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

In conclusion, while the prototype satisfied many of the initial requirements that was discussed in section 4.2.1, there were some drawbacks that will be mentioned in the section 4.3 to explain why it was discarded instead of making into the production code. But before that, we will go into some implementation details about EmailValet's mass email communication modifications in the next section.

4.2.3 Architecture

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

Identification of a Campaign Message   An email message consists 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 consists 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[10], 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.

EmailValet had already the data model to store an email message's related data when it is synchronized[11] with Gmail. However, creating a mass email campaign requires to identify if the synchronized email belongs to a mass email campaign that was created or if it is an answer to a campaign message that was sent before. Because, EmailValet's initial de-

---

[10]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>`

[11]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, 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).

sign were able to fetch emails from Gmail, then assign properties afterwards. The emails composed by EmailValet are become exist in EmailValet after recipients sent a reply to them, since then EmailValet fetches them from Gmail's inbox as a 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 in a Campaign data model as in figure 4.7. Hence, we were able to identify the messages during the synchronization with Gmail whether a message is an initial campaign message, or if it is an answer to a campaign message that was created earlier. Before, setting a message-ID was done by EmailValet's "mail" gem[12] during its execution time without allowing to get its value.
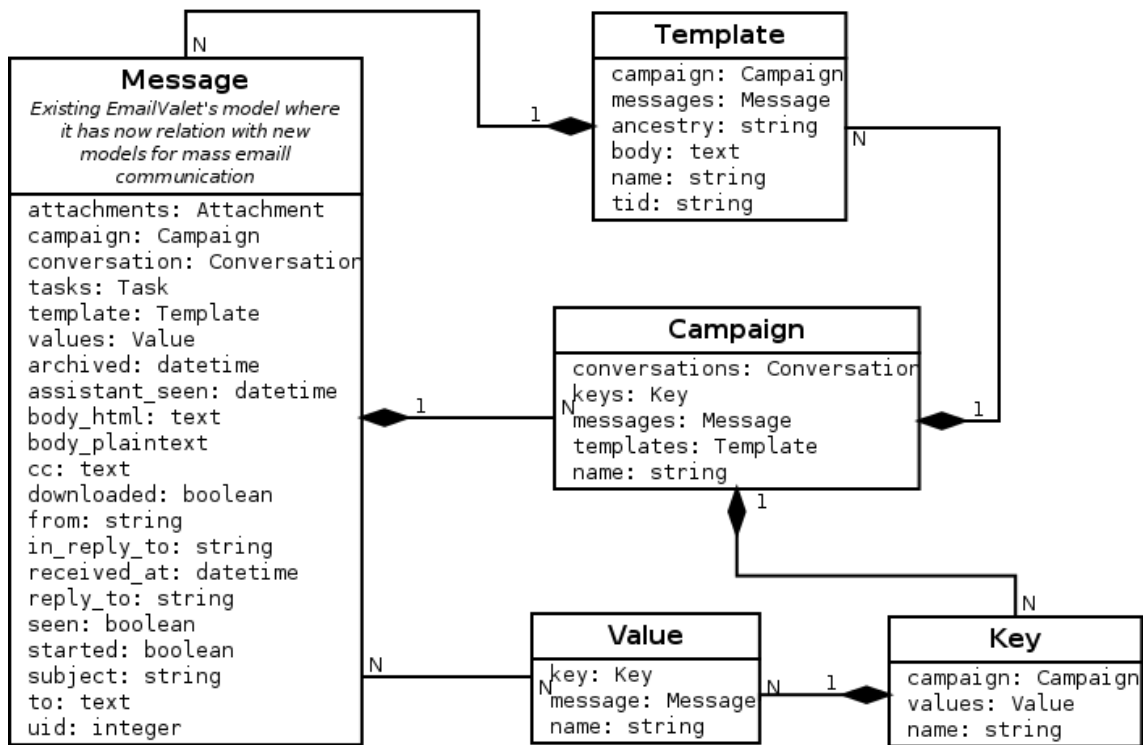
Figure 4.7: Model Dependency of the New Features of EmailValet

Template and Email Message Relation   Each campaign message that is composed in EmaiValet is actually a template with dynamic variables in it. Therefore, the campaign messages that are fetched from Gmail correspond to a template in EmailValet. This relation can be seen in figure 4.7. As it was mentioned in the previous section, there is a tree

---

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

view when a user replies a respondent's answer. This tree view's purpose was to choose earlier templates to reuse, and give a visualization 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 keeps for the each child nodes as well, hence each child node has a left and right value that represent another set of children, as a part of outer nested set. Figure 4.8 depicts a tree in a nested set model.
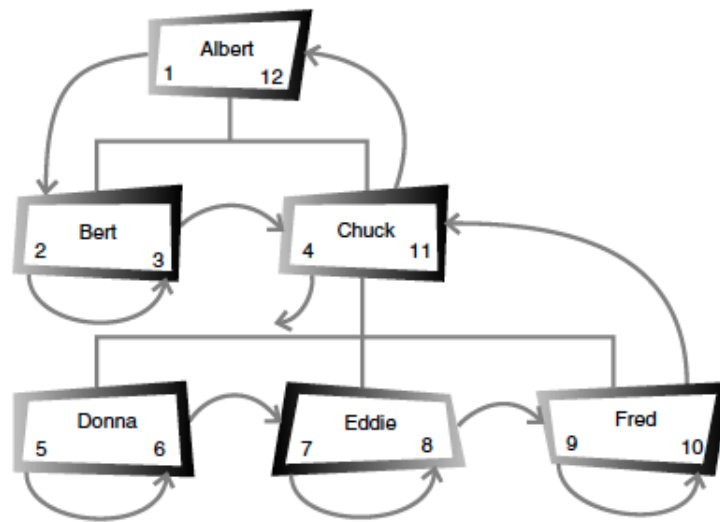
Figure 4.8: A Nested Set Model Tree Hiererchy (Celko, 2004, page 46)

Nested set model helps us to find 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, to find all the nodes of the root node 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 adjacency list, however, we first need to find all the nodes whose parent ids are equal to "Albert"s id, then for those nodes' ids, we need to find other nodes whose parent ids are equal to them, and it continues 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[13].

---

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

KVPs Relation with a Campaign and a Message   As in figure 4.7, each email message can have more than one values that correspond 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 the 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 what was missing on this initial design comparing with initial requirements described in section 4.2.1, the problems with it, and which new ideas that it was emerged.

4.3 Evaluation of the Prototype

In this section, we will analyze what are the drawbacks of the prototype, even though it fulfilled most of the requirements mentioned in 4.2.1. Some of these drawbacks were related with developing a prototype on an existing application whose main purpose was different than a regular email client. As one of the reason for prototyping is to make developers and users realize the new requirements or review the existing ones, we will also mention where we realized some new requirements, and as a result the new feature ideas.

4.3.1 Limitation of the Existing Application

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

Number of Message Limitation   One of the problem with EmailValet was its limit on the amount of messages that are synchronized from Gmail. It was limited to the latest 100 messages as a technique for EmailValet's privacy and accountability (Kokkalis et al., 2013). However, a mass email campaign will reach quickly to that limit, and it might ignore some of the older emails.

Different Inbox Purpose   EmailValet was offering an email client with a task manager feature as a replacement of people's regular email client. However, a mass email communication is not something that a person does on a daily basis like their 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 the emails except the ones related to an email campaign. Combining both inboxes also resulted drawbacks on the performance when it synchronized with Gmail to fetch new emails. Because, EmailValet was also performing some additional steps during the synchronization such as checking and comparing the message headers to find the changes in the last 100 emails.

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 all KVPs in an aggregated way when we browse to a message belonging to a specific person. Therefore, assigning KVPs to an email message actually means that assigning them to a contact in a mass email campaign. This requires to keep each individual contacts in a separate data model, and relating it with the existing EmailValet's design. Since we already spotted the limitations of EmailValet in the previous section, the decision was adding those enhancements into a new application.

Getting Contact List into the System   At Stanford HCI group, we realized that people who do such a mass email communication import their contact list from a spreadsheet or a similar data store. Since, the prototype requires to input a contact list in a special syntax that was mentioned in section 4.2.2, this makes the users change their existing contact list in a format that the prototype will be able to parse. This reduces flexibility, and adds an additional effort to the researchers who have already a contact list ready in a data file. Again, it was a feature where it would be better to start implementing it in a new project separated from EmailValet.

4.4 Conclusion

In this section, a mass email communication concept was introduced with an illustrative scenario. We pointed out how could be possible to reduce the efforts on the researchers' side, while to increase the personalization of the emails. Later, we collected those ideas as software requirements for the prospective prototype, where we introduced the features and their architecture. Finally, the evaluation of the prototype made the basis to the next chapter, where the mentioned limitations are removed or minimized on the final design, and the new requirements are created from the feedbacks of the users' 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. Firstly, the improvements on the requirements will be compared to the ones defined for the prototype from previous chapter. Later, the actual product's features, and its architecture will be discussed.

5.1 The Improved Requirements

Having seen the prototype in action made us to review the requirements, and bring the ones. Some of those requirements are shaped according to the feedbacks that we got at the Stanford HCI group, as well as the organizations who do mass email communication on weekly basis to reach their community. Following sections will discussed those new features on the final product.

5.1.1 Assistant Support

As it it discussed in section 4.1.1, the standard that we want to achieve within a mass email communication is the most adequately personalized emails according to recipients with minimum effort on a researcher's side. The initial idea and the prototype brought several features as discussed in chapter 4 for this purpose. However, if you consider the gold standard in the figure 4.2 we would like to achieve, the prototype still leaves effort on a researcher to accomplished a successful mass email campaign.

In order to minimize the efforts on the researcher's side, the additional assistants' involvement was considered. Therefore, a primary researcher will be able to share the tasks with permitted assistants. These tasks may be extracting information from the incoming answers or even writing replies to those answers. Hence, the primary researcher

will only need to interact with the flow of a mass email campaign when there is situation where it requires primary researcher attention. However, the system still needs to provide the necessary features that we will see in the next sections to support the work flow in a mass email campaign, hence assistants will only need to interact with this work flow to let the continuation of the email campaign by providing answers with the email templates, and extract information as KVPs.

5.1.2 Dynamic Variables and KVPs

We introduced dynamic variables at the initial prototype, however it was only limited at the salutation of the email. Since it allows the personalization of emails easier, and email marketing applications also supports this feature in the same purpose in section 3.3.3, the final product will also include this feature. As a result, application users can create KVPs and use those keys in the content of the email message to be replaced dynamically by its value according to the recipients. Therefore, the extracted information from the emails will not just help us to gather information in an organized, but also use them to personalize the emails

As a result, instead of keeping the KVPs in the system according to the responds, system should keep them according to the recipients itself. This is where the KVP-idea differs from the prototype as well. So, we have now profiles of contacts for a campaign having all KVPs of a recipients visible during the whole state of the conversation, not just at one message of the recipient. However, system should offer an option to hide individual KVPs to avoid cluttering on the view, and keep the KVP list with the ones that are actively used.

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

5.1.3 Importing and Exporting Contacts and Their Information

In the prototype, application user had to enter the basic information of the recipients such as first name, last name, and email address into the system manually. However, we realized while they were doing this, they use a spreadsheet and copy the contact information from there. This was also the case for their regular email client that they used for email campaigns. The applications that was investigated in section 3.3 had an option to import from spreadsheet to ease the process. Therefore, the system should offer an option to import contacts from the spreadsheet environment.

However, importing should not be limited only the basic information of the recipients. Since we already mentioned about importing KVPs from spreadsheet in the previous section, system should detect and import the contact information and KVPs related to that contact if it is available from a provided spreadsheet.

System should provide a bi-directional synchronization, not just importing data from the spreadsheet. Therefore, system should provide an option to export contact information and their created KVPs from the system to the spreadsheet. This will also gives a reporting functionality to the application users, where they can see all the contacts, and their extracted information from a campaign in one view.

5.1.4 Interoperability with Other Email Clients

Even though we provide a new system for the users to conduct their mass email communication, there might be some cases where a mass email communication initiated in the user's regular email client, and the system is not aware of that campaign since it was not created with it. Application should provide an option to import those emails messages created with another email client in to the system by recognizing those annotated messages by the user.

Enabling the system to import email conversations from other email clients reduce the dependency on the application, and while a researcher continue to use their own email

client, assigned assistants can take care of those imported emails by the researcher. We saw such a import feature in CRM applications in section 3.3.1 as well, where a user is able to forward an email into the CRM application, and system takes care of assigning the imported email to the corresponding recipient entry in the system.

5.1.5 Automated Decision-Making and Notifications

Even though the involvement of assistants make the initial researcher's life easier, the system still needs to provide an automated approach to answer the emails whose status is clear in the flow of a mass email communication. Therefore, a rule based decision-making mechanism should be used, where users sets the values of the keys of KVPs, and it triggers the action of sending an email to a respondent.

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

5.2 Final System

In this section, we will see that how the revised and improved requirements from section 5.1 reflected to the final product, named "Myriad".

5.2.1 Log-In and Campaigns Overview

As it was the same case for the prototype in section 4.2.2, Myriad also requires a Gmail account to work with. the reason behind is not just the popularity of the Gmail, but

also Stanford University uses Google Apps by default. Therefore, each member of the university has a Google account to use with Myriad. This also provides flexibility to use Myriad, because there is no registration form or sign in screen.

After user sign in to the system, the first screen that they will see is the campaigns overview screen, in where all the created campaigns including the ones that are shared with the user (as a assistant) are shown as in the figure 5.1. It has a simple and a clean UI, deferring from regular email clients to emphasize its focus on mass email communication.
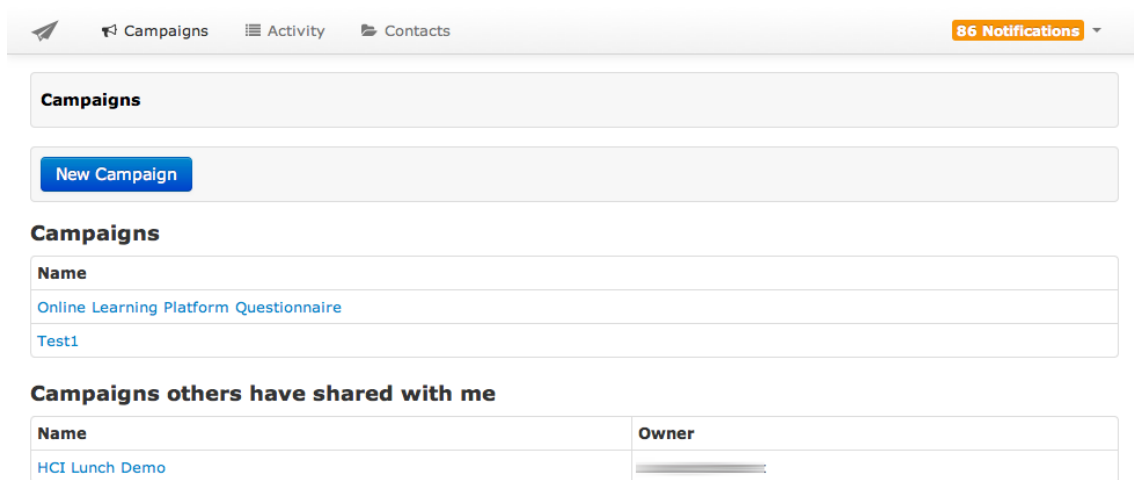
Figure 5.1: Myriad's Campaigns Overview Screen
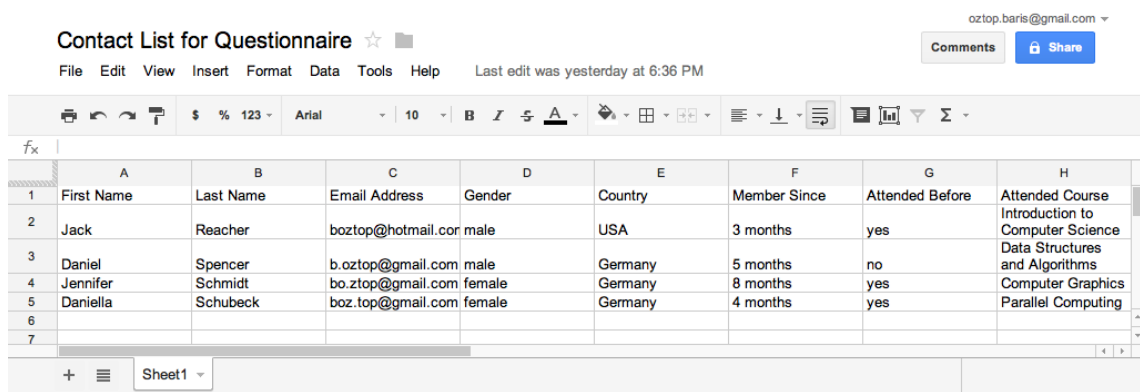
5.2.2 Synchronization with Other Sytems

Myriad is able to get the contacts' information and their KVPs from Google Spreadsheet. This is a convenient way to import contacts into the system since many people already keeps their information in a spreadsheet environment as we discussed in section 5.1.3. Therefore, Myriad offers a bi-directional syncing from and to the spreadsheet defined at the creation of a campaign. Besides, Myriad also offers an option to enter recipients first name, last name, and email address into the system directly.

The corresponding columns in the spreadsheet start with "first name", "last name", and "email address" as in the figure 5.2. The rest of the columns will be acted as a KVP, and

5.2 Final System

imported into the system as well.

Figure 5.2: A Google Spreadsheet to Import into Myriad

As it was discussed in section 5.1.4, importing existing email conversations as campaigns into the system is an important feature, since researchers may initiate conversations in their regular email client, and later on to make them more manageable, they can import them into Myriad. Myriad leverages Gmail's labeling feature which is equivalent to the IMAP protocol's folders for this purpose (Google Inc., 2013). Myriad creates a Gmail label in the user's account, and the only thing that user needs to do is the enable label syncing feature at the create campaign screen. Next, there will appear a label with the campaign name, and group under the root label of "myriad" in Gmail's inbox as in the figure 5.3. Hence, researcher will able to imports email messages, which were not considered to belong to a campaign for some reason, into Myriad.
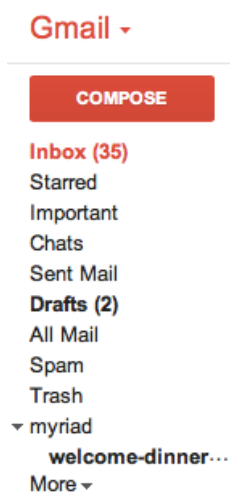
Figure 5.3: Gmail's Labels and the Myriad's Campaigns Under the Myriad Label

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 URL to synchronize with. Other two checkboxes for the spreadsheet are to manage the frequency of synchronization and disabling the warning in case of 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 under label synching.

Figure 5.4: Creating a Campaing in Myriad

After the campaign is created, all the contacts and their KVPs will imported if the user

provided a Google Spreadsheet as in the figure 5.5.

Figure 5.5: Contacts and Their KVPs After Synchronization in Myriad

5.2.4 Composing an Email Message

Users can send emails to all the contacts that they entered or imported into the system, or select a subset of it by using the filtering functionality as in figure 5.6. Provided filtering option are according to values of KVPs, conversation status such as unread, replied, or unreplied, or the last message's template name.

Figure 5.6: Filtering the Contact List in Myriad

Filtered recipients are ready to be composed of email at the compose screen after pressing the corresponding button. The compose email pane (see figure 5.7) contains a section that lists the previously sent emails to reuse. This is the same templating idea that was introduced at the prototype in chapter 4 section 4.2.2. It also shows the visualization of the state 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 C. The system also suggest an email template while repling a responded by considering the nodes at the same tree structure.

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

The compose pane allows users to dynamic variables into the content of messages to personalize them according to recipients. Those variables are not just limited to the first name and last name, but also can be any keys from the assigned KVPs in the campaign.
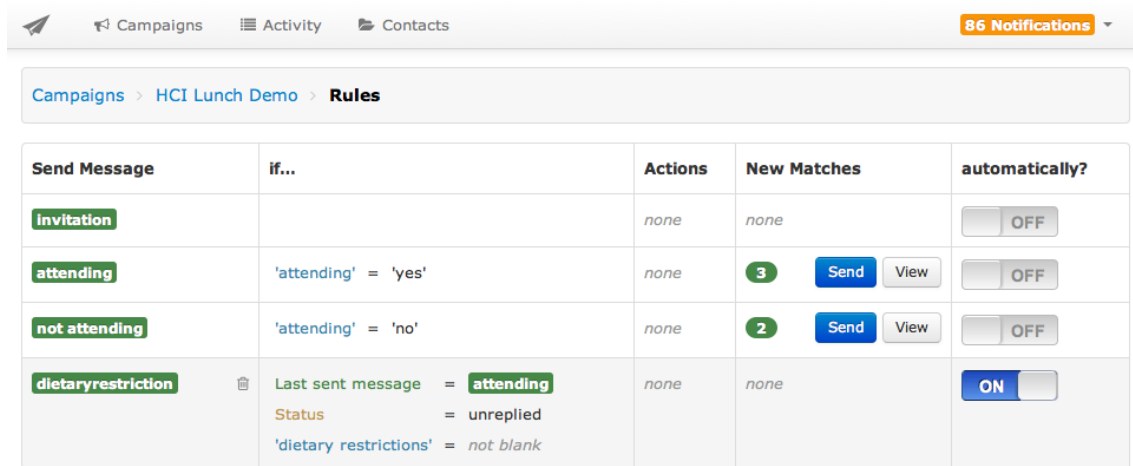
Each campaign has a link to the other campaigns if the recipient was involved for them as well as we will see in the next section. A researcher can access to the conversations and the KVPs from other campaigns by switching to the campaigns via that link. Providing an option to see the earlier campaigns that a recipients was involved gives broader

knowledge about a recipient that may help researchers to personalize the content of the emails more easily and properly. For example, if we had extracted information regarding which sports that the recipients do in an earlier campaign. We can use the same information to make a more frankly and friendly start in the new campaign by mentioning about the latest events of those sports areas in the country. Such a technique also supports the social exchange and the diffusion of responsibility theories 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 became obsolete during the flow of the campaign to avoid cluttering of KVPs in the application's view.

In the 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 another screen to change or set KVPs whose values depends on the email user has just sent.

At the beginning of this section it is mentioned that a user can filter the recipients list, and send an email to the subset of it according to filtered condition. Myriad saves those filtered condition under the "Rules" menu as in the figure 5.8 after the user sent the email. Hence, the next time if there are new emails satisfying the same conditions as before, user will notice them under the rules menu, and the earlier sent email message can be sent to those new matching recipients automatically if user chose to enable it automated option or a user can simple press the send button manually in the same screen. For example, it the figure 5.8, there are 3 recipients whose "attending" key was set to "yes", and the system already recognized that we already sent an email to those whose attending, and there are new matches for the same condition. In this case, the user can simply press the send button after, or automated 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 where all the messages from a recipient are visually grouped together. The advantage of threaded view is that it allows the researcher to get a quick overview of the whole state of a conversation with an individual, therefore a researcher can write a customized message more easily by focusing on the specific personality of the individual being respond to.

Each of the researcher's message is also annotated with the name of the message template that was used, hence it makes the latest state of the communication easily recognizable. In the figure 5.9, the emails that was send by the application user had a green highlighted labels writing the name of the templates that was used, which is "Accepted" and "Invitation".

While reading a recipient's answer, the extracted information can be recorded in KVPs at the right-hand side of the reading pane (see figure 5.9). Earlier recorded keys' values can be also updated at the same pane. Having KVPs along with the conversation thread gives a researcher necessary information about the person that is going to be replied. Under the KVP pane, a researcher can also see if the responded was involved in other campaigns with a link to those campaigns and the amount of emails that was exchanged next to it.

This is an helpful feature to remind the researcher about the existing of previous relation with the respondent, and if necessary the researcher can switch to other campaign to get an overview of extracted informations in KVPs.

Figure 5.9: Reading Pane and 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 email addresses into Myriad as in the figure 5.10. The task of a researcher can be extracting information from the email, writing answer to the recipients, or even proofreading of a researcher's emails before sending them.
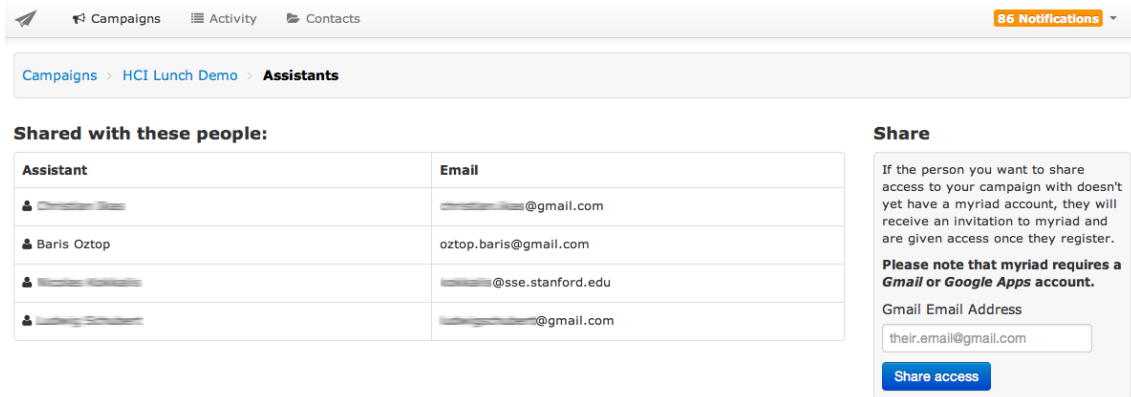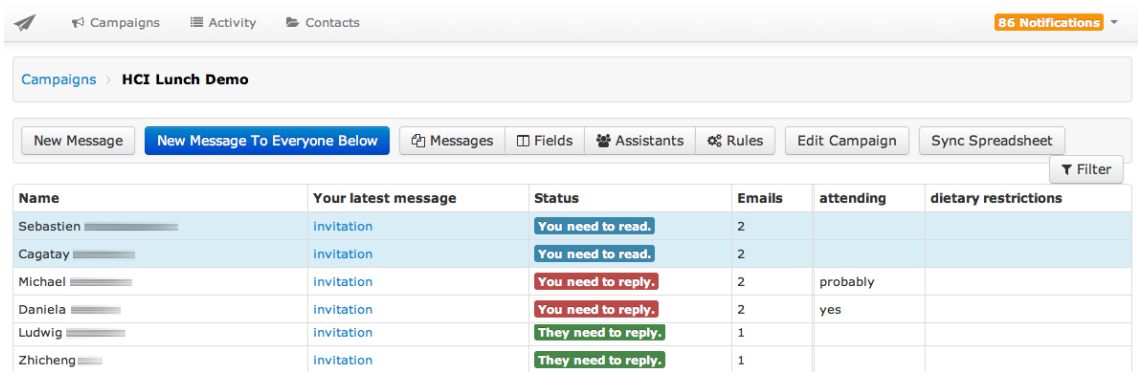
Figure 5.10: Rules to Automate the Sending Process of the Emails in Myriad

After an assistant is assigned to a campaign, he or she will get an notification email including the link to the campaign. Again, there will be a notification email for each email coming from respondents to the assistants' email address to let them know about the action required situation.

Myriad provides status labels for each received or sent email to give a hint about the next required actions as in the figure 5.11. These status labels give a hint about the next action that should be done in the state the conversation such as if the application user needs to read or reply a message, or the action is required from the recipient's side to continue the communication. There are also status labels related with Myriad's internal state regarding to an email message such as if Myriad sent the messages, or there was a failure with sending. The same view also provides a column to show what was the last message that is sent by an application user. Therefore, researchers and 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

5.4 Experiences

5.5 Conclusion

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.
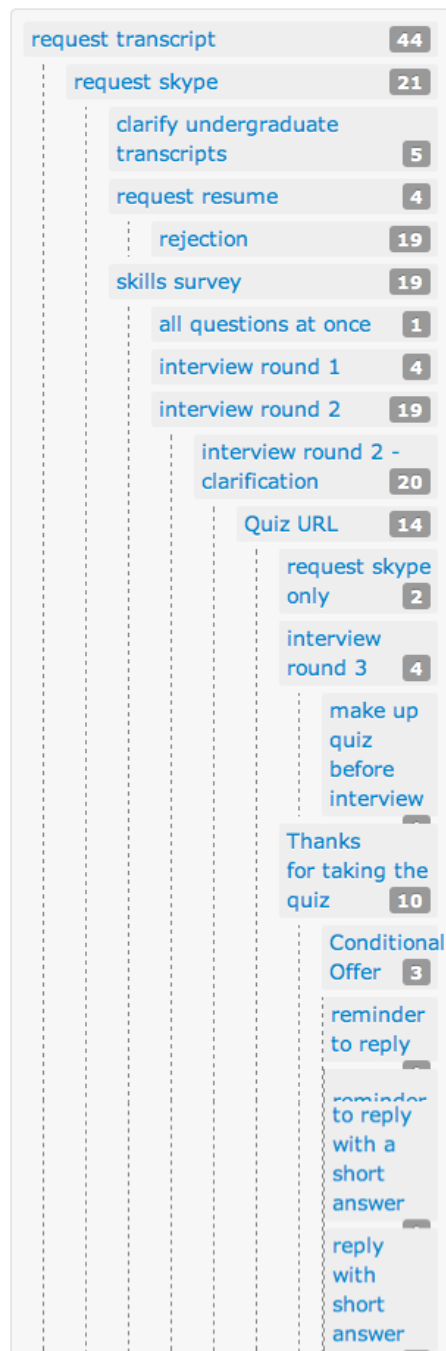
C Visualization of the Communication State

Figure C.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.

Bibliography

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 question-naire 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. (2013). 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 ques-tionnaires: 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.

Bibliography

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

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

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.

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.

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