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, July 26, 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

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

SaaS Software as a Service

UI User Interface

KVP Key-Value Pair

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 Sstatement

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 on 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 focusing on email communication directly or indirectly. Followings section will give a brief description of those are these product types, 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 customer (Reinartz et al., 2004).

One of the reasons why this thesis considers CRM applications to evaluate is its communication aspects 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 digital world make the true authentic personalization more rare. To achieve such a level of personalization requires getting know each recipient very well by considering not only the recent conversation, but also earlier conversations, and all the information that might be extracted from those conversations helping to build a relation with 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 adequate amount of personalized information to a email conversation.

3.1.2 Help Desk

Another application that focuses on a company and its relation with their clients is help desk applications. It's main purpose to provide information and support related to a company's products and services to their customers. As a part of knowledge acquisition, help desks supports 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 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 initiate an open ended questionnaire, extracting information from the coming replies, and organize them according to the answers that researcher 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 require great amount of effort from a researcher's side, where 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

group under following categories Eley and Tilley (2009):

- **Educational Communication:** An educational message is given in the form of newsletter, avoiding sale push, but it might still consist some content indirectly by encouraging recipients. 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, 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 marketing ads, and clear message 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 a long with the registration confirmation email.

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

3.2 Methodology

The analysis examined two products from the categories of CRM, help desk, and email marketing. Selection of the products depends on the several product comparison websites including Toptenreviews.com[1], Softwareshortlist.com[2], as well as from the suggestions of Stanford HCI group members[3]. In addition to those websites and suggestions, their demo or trial version availability were also considered, since some of the products required fee before using them. After the products were shortlisted, the last filtering was done by getting their web traffic rankings from Compete.com[4], Alexa[5], and Google

---

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

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

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

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

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

Trends[6]. Finally, trial accounts were created on those application, and a scenario is 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 an in-depth exploration for these products.

---

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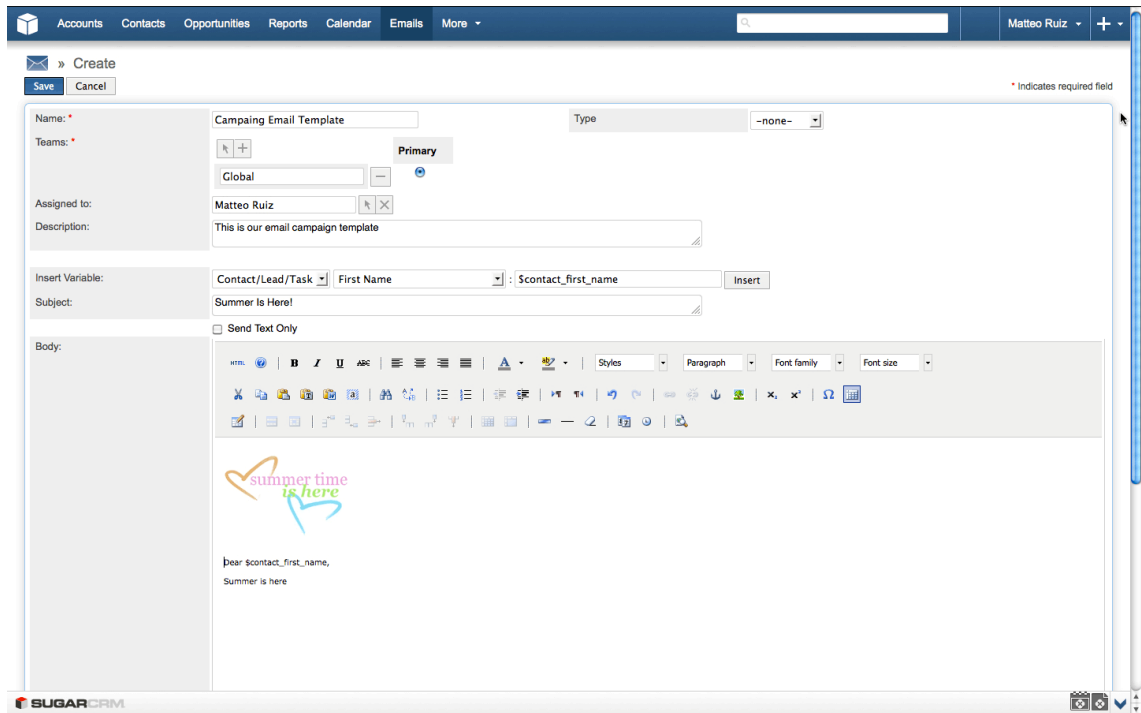

Figure 3.1: SugarCRM Email Composer with Embeded Variables (SugarCRM Inc., 2013)

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 email composer. Finally, it offers a mobile version to let access 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 no

---

[8]http://highrisehq.com/

24

calendar view, but a task view in Highrise. These tasks can be synchronized with iCalendar as well. In addition, user 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 simple forwarding an existing email creates a task in Highrise. Contact information can be imported from Outlook or 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 in them. Highrise does not offer any email composer to do email marketing as in SugarCRM. Therefore, user will depend another product to do simple campaigns. Provided activity view helps users to keep track of their 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 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 channel including website, email, phone, and social platforms like Facebook and Twitter. Hence, support request coming from those platforms can be turned in to a support ticket. Those support tickets can be group 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 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 or tablet devices. Therefore, support team has 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 has been had more than 30,000 clients since ten years. Kayako does not have social platforms integration like Zendesk, therefore support tickets are generated over website, email, and phone. Tickets can have custom 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 customer information 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 a long with support team. Kayako does not have any native app for mobile platforms like Zendesk. Finally, it has a analytics view to keep track of tickets report 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 a overview, and 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. A long with its intuitive UI, it offers a drag and drop functionality on the email content creation. It supports email marketing process from designing the sign up form, so users can add all the desired fields, and apply branding 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 "Admin" account can grant permissions different type of permissions allowing different access levels to MailChimp. This allows separation of tasks in mail marketing. For instance, while managers manage recipients list, author team can

---

[12]http://mailchimp.com/

focus on emails' content, and design (The Rocket Science Group LLC, 2013e). To design a 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).
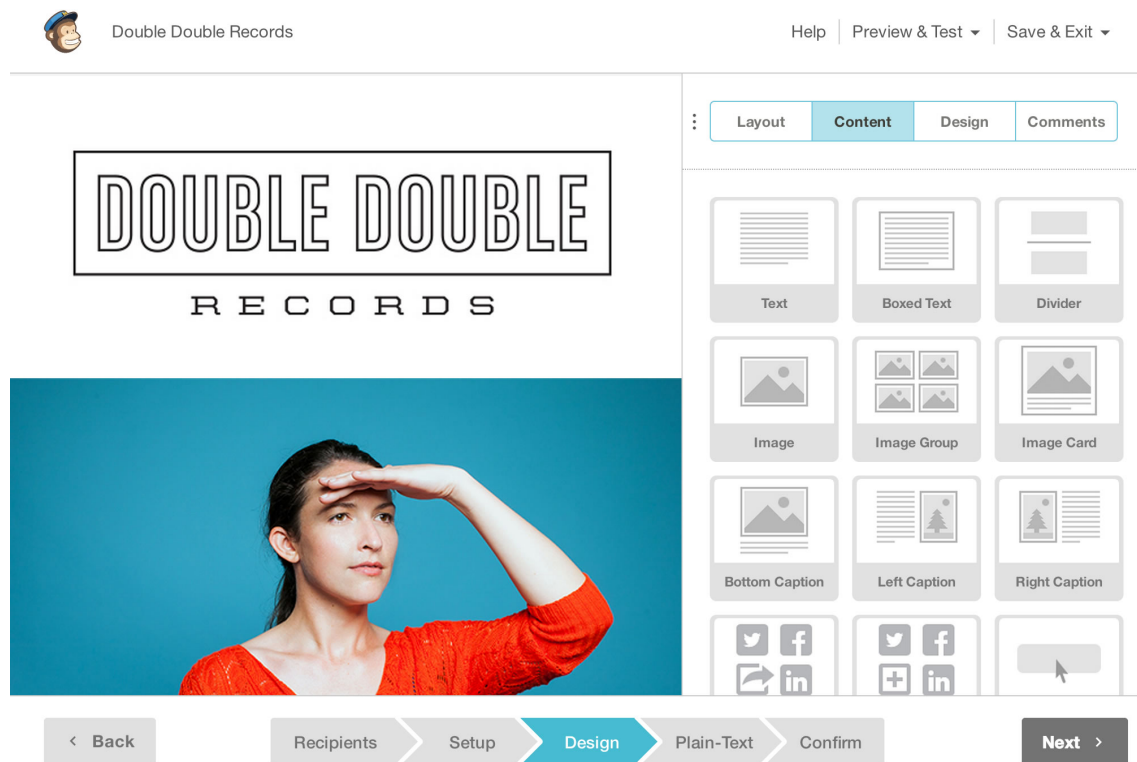

Figure 3.2: MailChimp Drag and Drop Content Editor (The Rocket Science Group LLC, 2013b)

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 software or an online email client, and a mobile browser (The Rocket Science Group LLC, 2013b). As in the SugarCRM, MailChimp allows you to use placeholder, 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 placeholders then SugarCRM, and it is possible to add a conditional logic to it. 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 a 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 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 amount 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 group under sublists which can also be merged into each other easily. An option is available to remove duplicate contacts from the lists or clean 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 researcher to know their respondents better 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 recipients list 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 to 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 other 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 a email campaign. This can be a task showing what is next to do in an email campaign initiated by researcher. For instance replaying the email in which respondent has asked clarifying question at our initial campaign.

- Both CRM applications support archiving of emails by simply forwarding to a unique email address provided by them, and linking those emails to the users' profile. This can be helpful to see important conversation happened with respondents at earlier time, so it can provide content or opinion about how to initiate upcoming conversations with those people. However, forwarding an email is an additional step, which require additional effort and time.

- Reusability of earlier emails are important not to write them again. As we have seen, SugarCRM also allows saving emails as a template to reuse again. However, there were 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 with their homework. In that scenario, it is probable many repetition of the same question will happen. Help desk applications provides 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 supports tagging or grouping of incoming emails, which can be helpful two 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 softwares was a support ticket can be shared or assign 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.

- Email marketing application MailChimp provides variable placeholder where users can add into email content and its variable will be replaced with actual value. Such a feature can helpful in personalized mass email communication, where it is difficult to add recipient specific personalized information into emails. However, there were 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 provided different type of permissions to leverage in a 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 analytics report to keep track of the success of a campaign or 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 it is 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 mentioned features, or doing them in a way to support their main purposes, which are CRM, help desk, and email marketing.

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

3.4 Conclusion

4 Initial Idea and the Prototype

In this chapter, a mass communication schema, and possible exceptional cases will be introduces. Afterwards, the initial prototype will be reviewed, and its drawback will be discussed.

4.1 Mass Email Communication Concept

When a researcher initiate 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 initial question, email addresses that are considered active might not be exist anymore, an auto responder might 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 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 to get their permission for 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 responds back with the answers of the questions, we will return a thank you email for the participation. For those who did not accept the invitation will get a motivation email to encourage for the next time. Next, some people returned conditional acceptance, which we collected them 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 a good luck for it, and afterwards we can mention about a reminder email that will be send out after a time period where participants mentioned in their response. After the reminder email, we will 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 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 a 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 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 to the questionnaire, and write the questions

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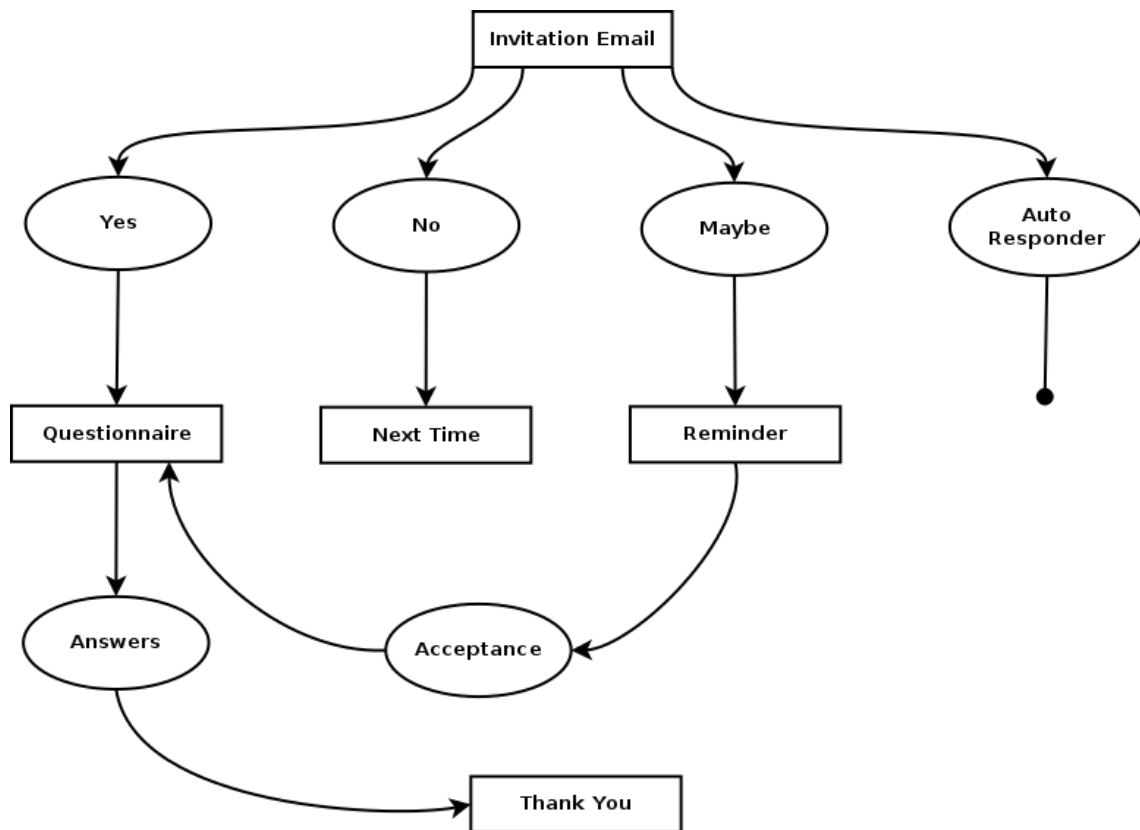
Figure 4.1: An Email Campaign Scenario

the questionnaire

- Go over for 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 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 replied with the answers of the questionnaire

However, when the amount of participants is great, the possibility of doing this manual work will get harder, and the personalization that is added into 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 situation, where it might require to ignore some respondents totally or writing 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 break up 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 to compose 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 star represent the standard where where it is really hard to reach with the available applications.
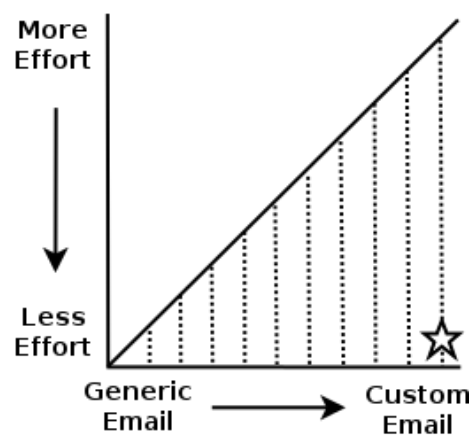

Figure 4.2: Effort vs Personalization on Composing Email

In the next section, we will see that how can we help to 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 during the sign up form when they registered in 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 require to enter all those information manually into those systems.

Dynamic Variables   The invitation email had 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 with 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 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 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 of 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 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 them. 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 have began 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 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 find the proper template to choose from. 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 researcher started to get messages from respon-

dents, 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 a 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
├─ ⋮
```
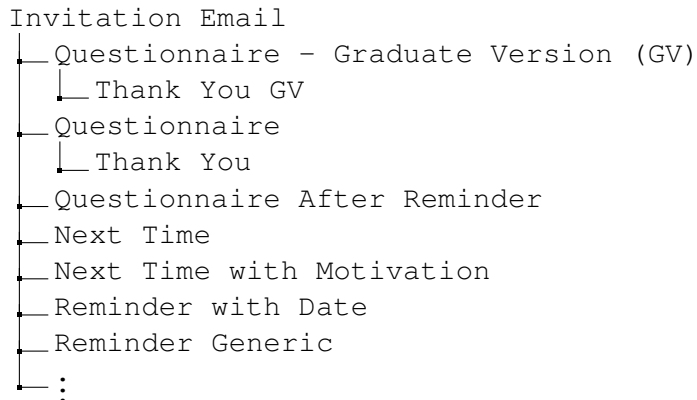
Figure 4.3: A Sample Conversation State and Template View

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 provided such a conversation view, and available templates had to be chosen out of mixed list which includes messages belonging 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, 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 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.

- Contact information can come from different sources, and assuming we have at least recipients' first names, last names, and email addresses. Application must store those information under a contact list permanently.

- Application user can use place holders for first and last names, therefore application must allow user to add those dynamic variables into email's content to personalize the salutation of the emails.

- Application user can extract information while reading responds into a Key-Value Pair (KVP)[2]. Therefore, application must offer an option to get KVPs from user. At the end, each contact's record will have enough information saved in KVP to describe all findings that we know about them.

- Application must allow user 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.

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

- 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 user created, and have a given template name by the

---

[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 represents its class. For example, "first name" can be a key, and each recipients first names will be a value for the "first name" key.

user.

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

Next, the prototype implemented to satisfy initial idea of a personalized mass email communication system, and the aforementioned requirements.

4.2.2 System

Initial system was built on the 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 email client that extracts tasks from incoming emails by the remote assistants from an expert crowdsourcing marketplace (Kokkalis et al., 2013).

---

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

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

4.2 Prototype

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)

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.

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.

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.

Bibliography

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

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

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

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

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

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

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

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

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

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

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

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

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.

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

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).

Bibliography

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.