# 1 Introduction

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

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

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

*1.1 E-mail as a Data Collection Method*

a researcher communicate with a great amount of people on a personalized level, as easy as possible by e-mail.

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

## E-mail as a Data Collection Method

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

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

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

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

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

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

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

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

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

## Problem Statement

To date, e-mail, as a popular medium for communication, is utilized for many purposes such as reaching groups of people to explore, describe, and explain things. However, as the group’s size gets larger, it becomes harder on the researchers’ part to maintain

* 1. *Problem Sstatement*

the consistency and effectiveness of the flow of the exchange of e-mails as compared to that of small groups. Therefore, the researchers tend to write generic e-mails—ignoring or using inadequate recipient-specific information with the help of a third-party software or an online tool. This results into low response rates, since recipients do realize that because they are a part of a large group of people being responded to means that you’ll feel less important and less valued, and the chance of getting a reply is less likely to happen. On the other hand, if researchers individually tailor those e-mails according to their recipients, it will require a huge additional effort at an increased cost, hence reducing the advantage of using e-mail as the primary communication medium.

Even though there are many products available in the market supporting e-mail communications, there’s just no available product allowing anyone to reach larger groups via e-mail, requiring minimum efforts while keeping the communication personalized at the same time.

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

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

This study also contributes on the following areas:

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

## 1.3 Outline

Outline goes here

*1.3 Outline*

# Foundation and Related Work

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

## Surveys and Data Collection

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

### E-mail 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). E-mail surveys offer more rapid surveying than other methods including regular mail and telephone surveys. In addition to that, e-mail surveys are inexpensive since it removes postage, paper and printing, and interview costs (Schaefer and Dill- man, 1998).

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

The result of the study indicated that the average duration of the data collection process for the e-mail version was less than a week, which is half of the duration of the written version. While the response rate of the e-mail 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 e-mail version. There were no differences in the nature of answers in the e-mail version compared with the written questionnaire.

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

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

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Author** | | | **Response Sample** | **Survey Topic** | | **Sample**  **Size** | **Usable**  **Sample** | **Method** | **Response**  **Rate** | **Time**  **(days)** |
| Kiesler & Sproull  (1986) | | | Employees of a  Fortune 500 | Corporate  Communication | | 115 | 77 | Mail | 67% | 10.8 |
| 115 | 86 | E-mail | 75% | 9.6 |
| Parker (1992) | | | Employees of AT&T | Internal  Communication | | 70 | 27 | Mail | 38% | NA |
| 70 | 48 | E-mail | 68% | NA |
| Schuldt & Totten  (1994) | | | Marketing & MIS  Professors (US) | Shareware Copying | | 200 | 113 | Mail | 56.5% | NA |
| 218 | 42 | E-mail | 19.3% | NA |
| Mehta & Sivadas  (1995) | | | Usenet Users | Internet  Communication | | 309 | 173 | Mail | 56.5%\* | NA |
| 182 | 99 | E-mail | 54.3%\* | NA |
| Tse, et al (1995) | | | University  Population (HK) | Business Ethics | | 200 | 54 | Mail | 27% | 9.79 |
| 200 | 12 | E-mail | 6% | 8.09 |
| Bachman, Elfrink &  Vazzana (1996) | | | Business School  Deans | TQM | | 224 | 147 | Mail | 65.6% | 11.18 |
| 224 | 117 | E-mail | 52.5% | 4.68 |
| Sheehan (1997) | & | Hoy | University Popu- lation (Southeast  US) | Privacy and Technology | New | 580 | 274 | E-mail | 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 | E-mail sur-  vey | 8% | NA |
| 150 | 42 | E-mail so-  licit | 11.3% | NA |
| Schillewaert, Langerak and Duhamel (1998) | Web users in Belgium | Attitudes toward the Web | 430 | 125 | E-mail | 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 | E-mail | 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 | E-mail | 58.0%\* | 9.16 |
| \*Differences not significant | | | | | | | |

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

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

Moreover, Paolo et al. (2000) reported that their respondents made longer comments to open-ended questions for the e-mail version of the survey, compared to the regular mail version. While the average number of words per comment was 58.33% in the mail version, the average for the e-mail version was 75.40%. Bachmann et al. (1999) had similar findings on their studies conducted on 1995 and 1998, where open-ended questions were more likely to be responded on by e-mail recipients than by mail recipients. In

the latter study conducted in 1998, the researchers also found out that e-mail respondents were more likely to expand their answers, even if it was not suggested on the survey, resulting in a more candid set of responses compared to the set of responses on mail surveys. Responses to open-ended questions are one of most the important measures on determining the quality of the returned surveys.

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

### Survey Errors

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

**Sampling Error** The greater number of people surveyed, the larger degree of precision can be achieved. Therefore, limitations on the number of people surveyed are considered under the sampling error. For example, while public opinion of 100 people results *±*$10% of the true percent, 2,200 people results higher confidence with the percent of *±*$2% (Dillman, 2006, page 9). Surveys relying on a predefined list of recipients considers that the list is randomly generated or with a systematic sampling. Hence, it has got little research to reduce sampling errors compared with face-to-face interviews in which multistage cluster designs1 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 elements of the population, coverage error happens (Dillman, 2006, page 9). Coverage error is consid- 1Cluster sampling selects preexisting groups of population elements instead of a single element of the pop- ulation (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 as one of the biggest issues of surveys ever since while surveying the general public (Dillman, 1991).

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

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

## Response Rate Influences

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

Low response rate was even considered a shortfall of the e-mail methodology despite to its advantages (Bachmann et al., 1999). In table 2.1, there are nine studies, where both postal mail and e-mail are compared side by side. Out of those nine studies, four of them showed a high response rate on the postal mail, three of them got a higher response on e-mail and two studies did not show any significant differences. Parker’s (1992) study of AT&T employees was the only study that was able to get an acceptable high response rate by e-mail. Schaefer and Dillman (1998) attributed this fact to the novelty of e-mail and that sent e-mails were carefully examined instead of considered as a company junk e-mail. Mavis and Brocato (1998) stated that studies cited by the others in support of e-mail surveys, as also shown in table 2.1, was not able to

* 1. *Response Rate Influences*

compare e-mail data collection with the more traditional methods, and their study design and analyses varied greatly. Sheehan and Hoy (2006) also focused the attention on many of these studies’ small and homogeneous population; therefore, it may not be applicable to represent larger population groups’ response tendencies.

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

### Length

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

### Multiple Contacts

The researchers found out that the number of attempts in contacting people increases the response rates (Heberlein and Baumgartner, 1978; Schaefer and Dillman, 1998). The scenarios for multiple contacts include pre-notification contact, which is a brief notice for the main request, and follow-up contacts, aimed for the people do did not respond upon the initial contact. Heberlein and Baumgertner (1978) showed that follow-up mailing has a mean return rate of 19.9% at the initial contact, and continued on with 11.9% and 10.0% for the second and third contacts, respectively (Heberlein and Baumgartner, 1978). Schaefer and

Dillman (1998) also stated that the same conclusion applies for the multiple contacts for e-mail in their literature research. According to this, the average response rate for e-mail surveys with a single contact was 28.5% while 41% and 57% for two and more than two contacts, respectively (Schaefer and Dillman, 1998).

### Personalization

Personalization has been addressed as an important factor in increasing response rates by many researchers (Dillman, 1991; Schaefer and Dillman, 1998). It builds a connection between the respondent and researcher, by making the respondent feel important and drawing the respondent from out of the group (Dillman et al., 2009, page 272). Dillman and Frey (1974) conducted a study to see the effects of personalization, where they reached half of a university’s alumni sample via personalized cover letters, while the other half got impersonalized letters. The personalization treatment included personal salutations and real signatures affixed on the letters. They’ve 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 e-mails (Schaefer and Dillman, 1998). In the next section, we will continue with the application of personalization in e-mails and give the results of some studies.

## Personalization of E-mails

Studies on mail surveys showed that personalization helps increases the response rates (Dill- man, 1991; Schaefer and Dillman, 1998). Personalization is also important for e-mail 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 theory2 of the personalization of the e-mail.

On the other hand, Barron and Yechiam (2002) stressed on the socio-psychological

2Social 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 a volunteer’s dilemma. With a volunteer’s dilemma, one player is needed to volunteer in order to reach the outcome preferred by everyone else in the game. However, each person might be inclined on hoping that someone else will volunteer, resulting to a scenario of a higher instance of not volunteering, rather than volunteering. According to this, the greater the number of people in the group size, the lesser probability of volunteering will result, which will then produce the diffusion of responsibility effect. In order to experiment on the effect of diffusion of responsibility in the context of e-mail requests, they sent several e-mails asking for help either to a single address, or to a list of five addresses. In the e-mail body (see Appendix A), a fictitious graduate student asked if the university has a biology faculty, whose answer is actually a given to anyone familiar with the institute. The result of the study showed that the number of replies sent to a single e-mail address per e-mail got a 20% higher response rate than the number of replies sent to a group of e-mail addresses per e-mail. In addition to this, the study classified the given responses according to its level of helpfulness, and the rate of "very helpful" replies retrieved from the e-mails sent to a single e-mail address per instance was 187% higher compared to the responses retrieved from e-mails sent to a group of e-mail addresses per instance.

Another outcome regarding the use of multiple e-mail addresses in the "To" field resulted concerns from respondents in the study of Selm and Jankowski (2006). An introductory e-mail including a link to a web-based questionnaire was sent to recipients to explore the opinions of elderly Internet users about an electronic political debate. One of the respondents raised his privacy and confidentiality concerns when the header of the e-mail contained all the e-mail addresses of all of the 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 e-mail addresses of those you’ve sent this message to. Do you folks call that confidentiality !? I ’ve decided not to participate in this ’ carefully composed’ study, although I do have an opinion on the subject matter."

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

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

Dear Valued Customer,

Earlier today the e-mail 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 E-mail Message Showing the Importance of Confidentiality

In another study by Heerwegh (2005), personalization is applied to salutations in e-mails. The randomly drawn 2,540 samples from the student database of Katholieke Universiteit Leuven, Belgium were separated into two (2) equally sized groups. For the non-personalized group, the salutation of "Dear student" was used, while in the personalized group "Dear [First name] [Last name]" was used. The e-mail content was an invitation to a web survey which was about adolescent attitudes towards marriage and divorce.

The result of the study showed that the personalization-applied group got a 6.9% higher survey login rate than the un-personalized group. Therefore, they concluded that increased response rates were in line with the social exchange theory and the diffusion of responsibility theory.

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

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

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

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

Dear Don Dillman,

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

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

Kind regards, XYZ

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

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

## Conclusion

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

*2.4 Conculusion*

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

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

# Evaluation of Existing Applications

After building a foundation by giving out an overview about the related work on personalized mass e-mail communications, this section will evaluate existing systems and third-party applications available in the market.

## Application Categories and Their Relation with the Thesis

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

### Customer Relationship Management (CRM)

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

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

* 1. *Application Categories and Their Relation with the Thesis*

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

### Help Desk

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

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

### E-mail Marketing

Organizations and marketers use e-mail on marketing for several reasons. Some of those purposes are for brand and customer loyalty building, acquiring or converting customers,

advertising the brand or the product, solicit sales or donation, communicating for pro- motional offers and even educational purposes. At the end, these approaches can be grouped under the following categories according to Eley and Tilley (2009):

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

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

## Methodology

The analysis examined two products from each of the categories—CRM, help desk, and e-mail marketing. The selection of the products depends on several product comparison websites, including Toptenreviews.com1, Softwareshortlist.com2, as well as the suggestions of Stanford HCI group members3. In addition to those websites and suggestions, their demo or trial version availability was also considered, since some of the products actually require a certain fee before using them. After the products were shortlisted, the last

1http://{e-mail-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.com4, Alexa5, and Google Trends6. Finally, the trial accounts of those applications were created, and a scenario was simulated to get the full insight from them.

## Results

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

### CRM Applications

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

4[https://www.compete.com/](http://www.compete.com/) 5<http://www.alexa.com/>

6<http://www.google.com/trends/>

Table 3.1: Comparison Matrix for CRM Applications

|  |  |  |
| --- | --- | --- |
|  | **SugarCRM** | **Highrise** |
| **Versions** | On-premise and SaaS | SaaS |
| **Pricing** | $35 – $100 user/month, and a free community edition | $24 – $99/month, and a free  plan with limitations |
| **Task Manage-**  **ment** | Calendar based, no additional view | Individual module |
| **Syncronization** | Plugins available for Out-  look, Lotus Notes | Requires additional module  installation |
| **E-mail Client** | Built-in, allows e-mail marketing, with variable insertions | No |
| **Contact Import-**  **ing** | Via forwarding e-mails or plug-ins for Outlook, Lotus Notes | Outlook, Excel, vCard, or via  forwarding e-mails |
| **Mobile Support** | Yes | No |
| **Analytics** | Marketing Analytics, sales  forecasting and trends | No |

**SugarCRM** SugerCRM comes in three different deployment versions: On-premise, Software as a Service (SaaS) and the free community edition. It has a clean User Interface (UI) with a single navigation menu. Its calendar view can be synchronized with Outlook’s calendar or any other platform’s, which supports iCalendar7. It has a built-in e-mail management feature, as well as integrations with several platforms like Outlook and Gmail, or an Internet Message Access Protocol (IMAP) based e-mail server. Users can archive e-mails in the SugarCRM application by adding a unique e-mail address into the TO, Carbon Copy (CC) or Blind Carbon Copy (BCC) fields. This address can also be used to link an e-mail recipients’ information, including e-mail attachments with SugarCRM by simply forwarding the e-mails. Therefore, it removes the additional effort on manually importing them into the SugarCRM application and reduces dependency on a platform. The SugarCRM also comes with a built-in

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

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

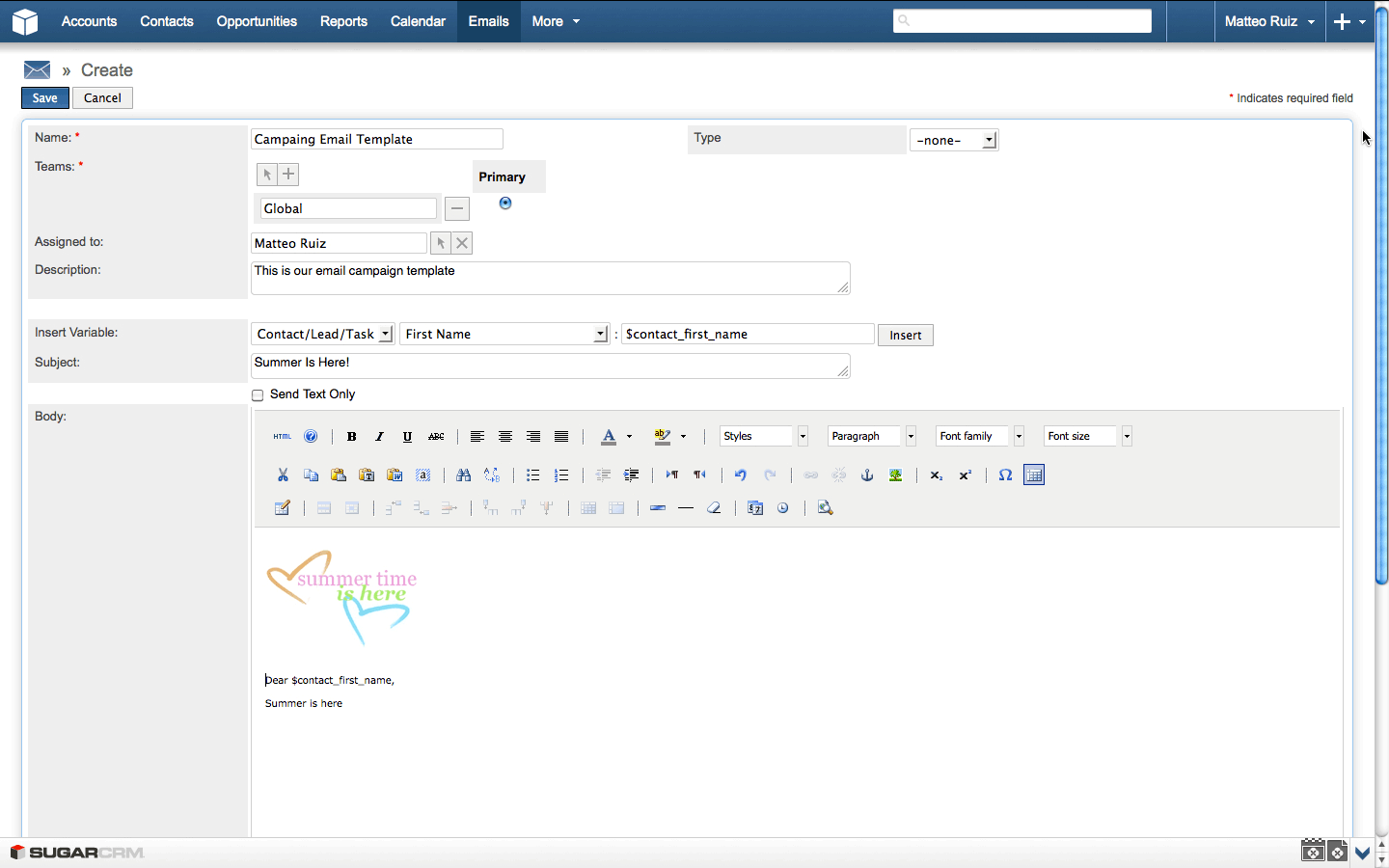


Figure 3.1: SugarCRM E-mail Composer with Embedded Variables (SugarCRM Inc., 2013)

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

**Highrise** Another SaaS application available on the marker is Highrise8. It offers several purchasing plans, with a 30-day free trial period. It has a simple User Interface (UI) like the SugarCRM, but it also has quick access buttons for adding a task or a contact. The Task management feature on the Highrise application makes it different from the SugarCRM, since unlike the latter, instead of

8<http://highrisehq.com/>

a calendar view, it offers a task view, which can be synchronized with iCalendar as well. In addition, users can create tasks from e-mails by using one the unique e-mail addresses for several time slots provided by Highrise, and adding them into BCC, CC, or simply by forwarding an existing e-mail, creating a task in Highrise. Contact information can be imported from Outlook or by uploading a vCard9 file. It provides all the basic contact information fields, including social accounts; however it does not offer custom field creation on those profiles. An e-mail, including its attachments, can also be linked to a contact profile just by simply forwarding it to the provided unique e-mail address. If a user does not exist in the Highrise contact database when an e-mail from him/her was forwarded to the link, a contact profile is created using any available information in that e-mail. Adding tags to contact profiles also makes it easier to organize contacts and browsing within them. However, the Highrise application does not offer an e-mail composer to do an e-mail marketing, as seen in the SugarCRM application. Therefore, users will have to depend on a different third-party application to do simple campaigns. The provided activity view helps users on keeping track of their own or other users’ recent actions within the Highrise application. Lastly, it offers options on customizing the look and feel of the application with the use of the system provided color schemes, depending on the user’s preference (37signals LLC, 2013).

### Help Desk Applications

The two help desk applications reviewed in this study are Zendesk and Kayako. Table 3.2 provides a comparison matrix of their features, and the details are described in the following paragraphs.

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

Table 3.2: Comparison Matrix for Help Desk Applications

|  |  |  |
| --- | --- | --- |
|  | **Zendesk** | **Kayako** |
| **Versions** | SaaS | Software and SaaS |
| **Pricing** | $24 – $119 agent/month, with a limited free trial version | $29 – $49 user/month, with a limited free trial version |
| **Channels** | Website, e-mail, phone, and  social platforms | Website, e-mail, and only the Fusion version supports phone |
| **Macros** | Yes, basic | Yes, advanced |
| **Ticket Manage-**  **ment** | 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 User Interface (UI). Zendesk has more than 30,000 businesses from a wide variety of industries. Zendesk offers one-on-one support through different communication channels including their website, e-mail, phone, and social media platforms like Facebook and Twitter. Hence, support requests coming from those platforms can be turned in to a support ticket, and those support tickets can be grouped under categories, and further classification can be done via tags for each ticket. This feature also helps in finding related archived resolved tickets, so they can be reused for new tickets. Thanks to the automated process coming with macros, a combination of actions can be done with a one-click like setting status, priority, type of a ticket, and assign it to another person with a predefined comment for the ticket. A ticket can be merged with another one, or copied to the forum to make it available to the public, which helps creating a reliable a knowledge base. Customer ticket histories and basic personal information are kept in the system. However, it does not allow adding additional fields on the customers’ profiles. In addition to the desktop version, Zendesk has a mobile version for smartphones and tablet devices. Therefore, the support team does not have to depend on the desktop platform, as long as they have it on their mobile devices as well. Lastly, the provided analytics view by reports gives an overview of customer satisfaction, and performance of the support team (Zendesk Inc., 2013b,a).

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

**Kayako** Kayako’s11 complete solution for customer support is named as Kayako Fu- sion. It comes as software and SaaS. Kayako has more than 30,000 clients within the last ten years. Unlike Zendesk, using its User Interface (UI) seemed to be a little more complicated than the latter, and does not have a social media integration, therefore support tickets are generated only over their website, or via e-mail and phone. Tickets can have customized types, statuses, priorities, and tags. Similar to Zendesk, it also supports macros to assign tickets into a department, owner, type, priority and provide canned responses for tickets with just a single click. Kayako also keeps basic information of customers, as long as they are registered in the system. Registered customers can also participate in building knowledge base in a forum-like environment by contributing other questions to support team. However, Kayako does not have a native app for smartphones and tablet devices like Zendesk. Finally, it has an analytic view to keep track of ticket reports, measuring customer satisfaction and support team’s performance (Kayako Inc., 2013b,a).

11<http://www.kayako.com/>

### E-mail Marketing Applications

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

Table 3.3: Comparison Matrix for E-mail Marketing Applications

|  |  |  |
| --- | --- | --- |
|  | **MailChimp** | **Constant Contact** |
| **Versions** | SaaS | SaaS |
| **Pricing** | $10/month with max of 500 subscribers – $240/month with max of 50,000 subscribers.  Pay as you go available | $15/month with max of 500 subscribers – $75/month with max of 10,000 subscribers. |
| **Template Editor** | Drag-and-drop including ad-  vanced photo editor | Drag-and-drop including ba-  sic 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** MailChimp12 comes as SaaS, and offers either a fixed monthly plan or a pay as you go plan. Along with its intuitive User Interface (UI), it offers a drag-and- drop function on the e-mail content creation. It supports e-mail marketing processes ranging from designing the sign-up form so that users can add their desired fields and apply brandings on it to applying personalization on e-mails using dynamic variables. The recipients’ list can also be filtered out according to several conditions like their campaign name, location or ratings, as assigned by the user. There are different user account types with different levels of privileges in accessing MailChimp. A person who has an "Admin" account is the only one capable of granting permissions to other users, as well as determining one’s access limitations on using MailChimp. This allows an efficient way for the distribution of mail marketing tasks. For an instance, while the assigned manager manages the recipients

12<http://mailchimp.com/>

list, the author team can focus on the e-mails’ content and design (The Rocket Science Group LLC, 2013e). To design an e-mail content, users can either decide on picking an available template from a collection provided the MailChimp application or to create their own HTML templates with its drag-and-drop editor (See figure 3.2).

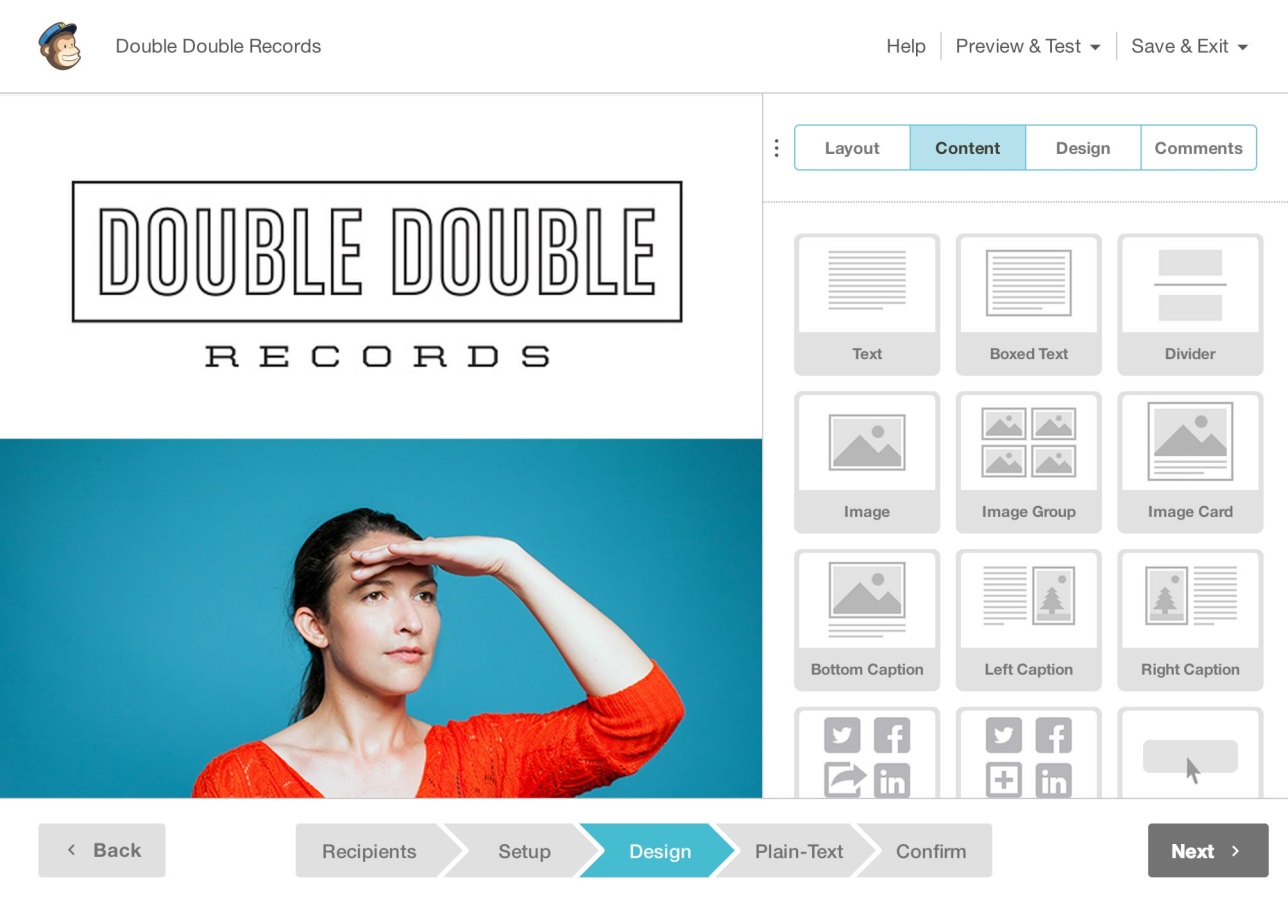


Figure 3.2: MailChimp Drag and Drop Content Editor (The Rocket Science Group LLC, 2013b)

The template editor has a photo editing feature, and authors can add comments to give a feedback on the content and design of the templates. Created templates can be previewed as if they are viewed on either a software, an online e-mail client, or even on a mobile browser (The Rocket Science Group LLC, 2013b). Similar to the SugarCRM application, MailChimp allows you to use dynamic variables—merge tags—in the e-mail content. Therefore, sending out e-mails can be personalized with recipient-specific information. However, it provides more types of dynamic variables than SugarCRM, and it is possible to add a conditional logic to them. For example, in listing 3.1, a custom discount message will be shown in the e-mail depending on what state in the US the recipient is from.

*∗*|IF:STATE=CA|*∗*

Save 20% on surf boards!

*∗*|END:IF|*∗*

*∗*|IF:STATE=GA|*∗*

Save 20% on Mountain Bikes!

*∗*|END:IF|*∗*

*∗*|IF:STATE=FL|*∗*

Save 40% on water skis!

*∗*|END:IF|*∗*

*∗*|IF:STATE=CO|*∗*

Save 50% on ski gear

*∗*|END:IF|*∗*

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

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

**Constant Contact** Another e-mail marketing SaaS is the Constant Contact E-mail Marketing13, whose purchase plan depends on the number of contacts you have, but a free 60-day trial period is available upfront. It offers a drag-and-drop content creation like MailChimp’s, but with a cleaner User Interface (UI). It offers quite a number of templates to choose from and customize. Users can embed sign-up forms into their websites or Facebook accounts. The recipients list can also be imported from different sources like Microsoft Excel, Outlook, and Gmail. In addition, recipients can be grouped into sub-lists, which can also be merged into each other easily. An option removing duplicate contacts from the lists or delete recipients who unsubscribed from the list is also available. Users can track how many opens, clicks, forwards, and social media platform shares are done for their e-mail campaign. On the other hand, it does not offer any sophisticated e-mail variables to be replaced by an actual content from the application (Constant Contact Inc., 2013b,a, 2011).

13[http://www.constantcontact.com/e-mail-marketing](http://www.constantcontact.com/email-marketing)

## Conclusion

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

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

Given the scenario, similar or identical questions might be asked several times. A Help desk application provides a ticketing system for customer related issues, which is also applicable to the above-mentioned scenario. Therefore, existing e-mail replies can be reused for further recipients.

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

As mentioned above, there are many useful features that can be helpful to ease a mass e-mail communication. However, there is no one specific application capable of doing all of the men-

tioned features, or doing them in a way to support their main purposes, which are CRM, help desk, and e-mail marketing.

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

# 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.

## Mass E-mail Communication Concept

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

### A Mass E-mail Communication Scenario

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

Dear John,

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

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

Your online course team.

Listing 4.1: E-mail Invitation to Questionnaire for Online Learning Platform

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

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

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

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

* Figure out the courses for the participants to attended in our e-mail list, and adds them to the e-mail content for each person.
* Write a generic answer for those who accepted the invitation for the questionnaire, and write the questions.
* Write a generic motivation e-mail to support those who did not accept the invitation

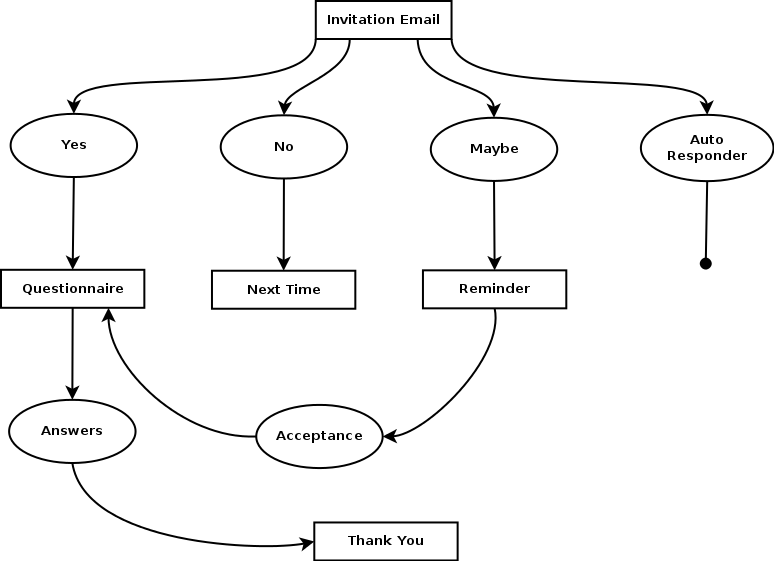


Figure 4.1: An E-mail Campaign Scenario

* Go over the "maybe" cases, and record the date and other important information they mentioned involving their participation.
* Write a generic e-mail to the "maybe" cases about the reminder that will be sent according to the time period they mentioned. This e-mail may include a couple sentences encouraging them to give a feedback and share the reason why they are unable to accept the invitation for the questionnaire at the moment.
* Go over the "auto-respond" cases to figure out if there is any condition that we need to ignore those people at all for the upcoming questionnaires in case they change their e-mail addresses, or any date that can be recorded to setup a reminder e-mail later on.
* Send a thank you e-mail to everyone who answered the questionnaire.

However, the larger the amount of participants there is, the harder it is to manually work on all of them, thus, the lesser personalization the e-mails will get. For example,

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

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



Figure 4.2: Effort vs Personalization on Composing E-mail

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

### Possible Improvements to Reduce Efforts

After a mass e-mail communication scenario was described in the previous section, we will see the possible improvements presented on this section to reduce the efforts on the researchers’

side, while increasing the personalization of the content on some degree.

**Contact Information** The first step would be to import contacts and their relevant information. These will include some of the basic information we asked when they filled-up the registration form to our learning platform e.g. first and last name, e-mail address and gender. Therefore, we can easily export these information into a different, easily-accessible file formats like Microsoft Excel or CSV1 to make it ready for use to initiate our mass e-mail conversation. The application that was discussed in section 3.3 offers several options in importing data onto the mentioned systems. Therefore, the researchers does not need to manually enter those information to the system.

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

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

1CSV 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 E-mails** After the initial invitation e-mail, there will be many instances calling the need on reusing different versions of an e-mail for further states of conversations. We mentioned the term "template" as an e-mail with some dynamic variables in it. Later on, these variables will get their actual values, and it will be the actual message that we want to send. From here on, we will add another meaning to the term "template", to emphasize a message’s reusability. Let us consider the "maybe" cases mentioned in section 4.1.1 to represent respondents who were too busy to participate on answering the questionnaire at that moment, but, maybe it is possible for them to participate on a later time (See figure 4.1). For instances like this, our recipients has to indicate a possible date for them to participate or provide an unclear answer about their availability. It is definite that we cannot ignore them if they provide us a date, and we should send them a reminder before that date to participate in the questionnaire in a generic e-mail. This might give them a hint that the messages we send are actually automated, and it may result to some drawbacks that we discussed in chapter 2. However, if we create two different e-mail templates; one with a dynamic variable in it to mention when we will send a reminder, and the other one with a more generic content, including a time far enough to encourage them to participate in the questionnaire. Therefore, the next time we encounter a "maybe" case, we have two ready-to-use e-mail templates to send them.

**Visualization of Conversation State** At this point, we have begun on creating different e-mail templates for different conditions, to ease a researcher’s life by allowing reusability on the replies. However, it is also difficult to pick out a template since a flow of a certain conversation is different from another one, one way or the other, and we have different e-mail templates for different conditions. This will result to an additional effort in choosing the appropriate template. Besides, we might not need some of the templates after conversation’s state gets further, and just use the latest ones more frequently compared to the older ones. Therefore, we need an overview to show us the current state of the conversation is, as well as to allow us to pick the proper template by leveraging this view.

Figure 4.3 represents a view in a directory tree structure for this purpose. All the possible e-mail messages that we have written are collected in a hierarchical way, starting with the first invitation e-mail at the root node. After the researcher began on obtaining messages from re-

spondents, all of the possible answers we can write are at the second level of the tree, right after the invitation e-mail node. Whenever we get an answer for the questionnaire, we’d write an e-mail to say thank you. This is also represented in the third level of the tree, right after the questionnaire node.

Invitation E-mail

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

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

## Prototype

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

stein, 1996). In this section, we will analyze the requirements and the actual prototype.

### Requirements Analysis

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

* + - 1. Contact information can come from different sources. Assuming we have atleast the recipients’ first names, last names, and e-mail addresses, the application must store these information under a contact list permanently.
      2. The application user can use variables in the e-mail message for a recipient’s first name and last name. Therefore, the application must allow a user to add different dynamic variables into an e-mail’s content to personalize the salutation of the e-mails.
      3. The application user can extract information from an e-mail message into a Key-Value Pair (KVP)2 while reading responses. Therefore, the application must offer an option on obtaining KVPs from users. At the end, each contact’s record will have enough information saved in KVP to describe all details we know about them.
      4. The application must allow users to use the extracted KVPs as dynamic variables in the e-mail content. This will allow us to personalize the e-mail with information specific to each recipient.
      5. The user can reuse a sent e-mail as a template to reply to respondents or initiate a conversation for a group of people. Therefore, the application must offer an overview for picking out templates from a list.
      6. The application must show the whole flow of the conversation. Therefore, the template list can be showed in a directory tree structure as in figure 4.3. Each node must represent the templates the user created, and let the user pick out a custom name for the given template.
      7. The application must offer a template selection and conversation flow representation under a specific single view. Therefore, the user should be allowed to pick which templates to reuse by

2A 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 the nodes in the tree to figure out the current flow of the conversation.

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

### System

The initial system was built on top of an existing project named as E-mailValet3 also known as GmailValet in the Human-Computer Interaction (HCI) group4 at the Stanford University. E-mailValet is a SaaS application that combines an e-mail client with a task manager. Remote assistants from an expert crowdsourcing marketplace extract tasks from incoming e-mails, and annotate them in a task list separated from the inbox (Kokkalis et al., 2013).

Having a ready-to-use e-mail client to implement the first version of the personalized mass e-mail communication system helped us save time and effort, and make us focus more on the initial requirements, as discussed in section 4.2.1. From the technology point of view, we were also limited by E-mailValet’s existing architecture, which was developed in a Web framework Ruby on Rails (RoR)5, and a JavaScript framework jQuery6.

E-mailValet requires a Gmail account to register and sign in7 to the system, therefore the application’s users have to grant access to E-mailValet before using it. Figure 4.4 shows E-mailValet’s inbox view after a user signed in and browse to a campaign8. On the left hand side, there is the e-mail list pane, and a selected e-mail can be read on the reading pane on the right.

3[https://www.gmailvalet.com/](http://www.gmailvalet.com/) 4https://hci.stanford.edu/

5<http://rubyonrails.org/>

6<http://jquery.com/>

7E-mailValet uses an open protocol named OAuth to provide credentials for authentication with Gmail.

OAuth’s community webpage is at <http://oauth.net/>

8In this study, the word campaign is used as a short-hand for a mass e-mail communication campaign, wherein a researcher initiates a conversation with groups of people related to one specific topic.

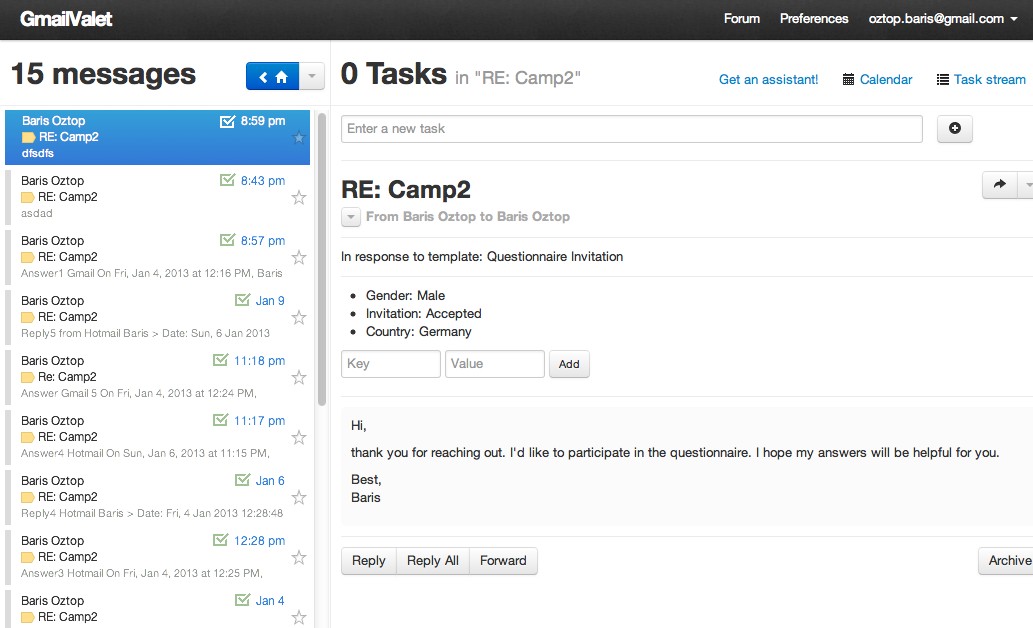


Figure 4.4: E-mailValet Inbox View and Message Reading Pane

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

The researchers can add their recipients list into the "Merge"9 input field as shown in figure 4.5, which corresponds to the "To" field in a regular e-mail client. However, the format should be similar to the shown figure—first name, the last name, and the e-mail address written in angle brackets. Hence, those fields can be used dynamically in the salutation of the e-mails by writing one of the variables of *{{first\_name}}* and *{{last\_name}}*. Once the recipients list 9The 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 e-mail is sent, these contacts will be recorded on Gmail’s contacts book.

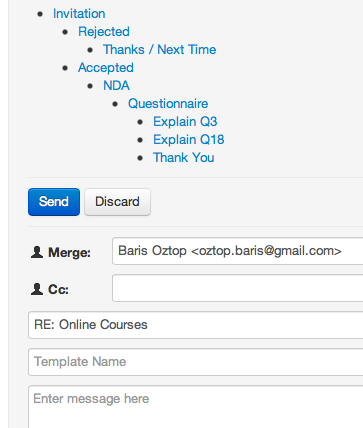


Figure 4.5: E-mail Template Selection to Reuse Earlier Replies

Figure 4.5 also shows the template tree used to represent the state of the conversation, and allow users to select a template from. The indentation at the nodes of the tree helps on identifying which templates are used after which one, therefore it gives an idea about the current state and flow of the conversation. According to this sample scenario in the figure, right after we sent the invitation e-mail and started on getting responses, there were two possible answers to give the recipients. For those who rejected the invitation, we can select the "Rejected" node from the tree to reuse an answer we gave before. This can be a motivation e-mail as discussed in section 4.1.2. On the same note, the people who accepted the invitation would receive a Non-Disclosure Agreement (NDA), and afterwards the questionnaire as the conversation with them carries on. After sending the questionnaire, there were cases where some respondents sent an e-mail clarifying some questions found in the questionnaire.

According to figure 4.5, these cases are related to the questions number 3 and 18, and we explained those questions, to clear off the confusion or issues they’re having. Again, since the system allow us to reuse our previous answers as a template; we don’t need to rewrite our explanation for those questions not clear for future querries. On the other hand, if the researchers wish to create a new template, he or she can simply do so by adding the name of the template into the corresponding input field, and send the e-mail. The next time the researcher fires up the application, he or she will find it in the template tree under the corresponding level of the node. Also, it is possible to select a template, made some modifications on it and save it as a new template. Therefore, a slightly different version of the templates can be reused during the communication as well.

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

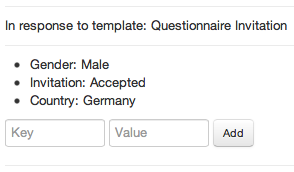


Figure 4.6: Extracted Key-Value Pairs and Input Fields to Add New Ones

In conclusion, while the prototype satisfied many of the initial requirements discussed in section 4.2.1, there were some drawbacks that will be discussed on the next section, section 4.3 to explain why it was discarded, instead of considering it for consumer production. But before that, we address some implementation details about E-mailValet’s mass e-mail communication modifications in the next section.

### Architecture

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

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

One of the fields in an e-mail message’s header is called Message-ID10—a unique message identifier, and set by e-mail 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.

E-mailValet had already the data model to store an e-mail message’s related data when it is synchronized11 with Gmail. However, creating a mass e-mail campaign requires identifying if the synchronized e-mail belongs to a mass e-mail campaign created or if it is an answer to a campaign message that was sent before. Because E-mailValet’s initial de-

10The 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>](mailto:CAF2E4bfH4%2BGAYHcJFZJ6dTJJ%2Bpux4mTjff2neCS_VR_zVCUY9g@mail.gmail.com)

11Fetching the e-mails 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, E-mailValet uses Simple Mail Transfer Protocol (SMTP) protocol to connect to the Google’s SMTP servers to send a composed e-mail in E-mailValet. SMTP is a mail transport and delivery protocol defined by RFC5321 (Klensin, 2008).

sign was able to fetch e-mails from Gmail, then assigning of different properties should follow. The e-mails composed by E-mailValet comes to an existence in the E-mailValet database after recipients send a reply to them, since E-mailValet fetches them from Gmail’s inbox as part of the thread of messages. Therefore, Message-ID and In-Reply-To fields of an e-mail message are leveraged to keep track of a campaign message by setting a Message-ID into it, and store it in a Campaign data model as illustrated on figure 4.7. With this, we were able to identify the type of messages during the synchronization with Gmail, whether the message is an initial campaign message, or if it is an answer to a campaign message that was created earlier. Before this, setting a Message-ID was done by E-mailValet’s "mail" gem12 during its execution time without allowing to get its value.

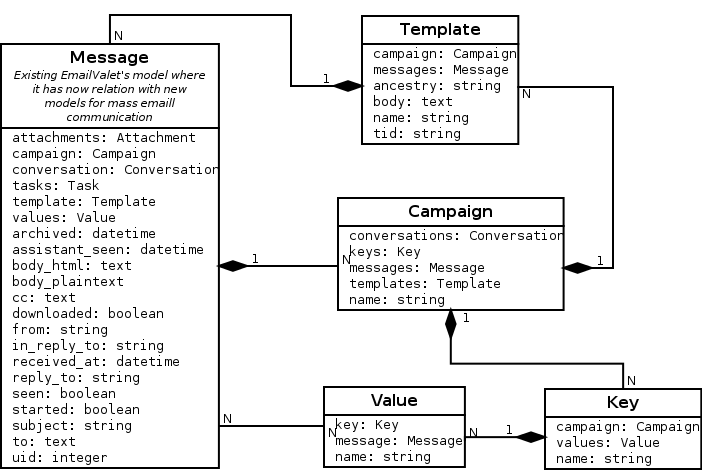


Figure 4.7: Model Dependency of the New Features of E-mailValet

**Template and E-mail Message Relation** Each campaign message composed in EmaiValet is actually a template with dynamic variables in it. Therefore, the campaign messages fetched from Gmail corresponds to a template in E-mailValet. This interaction can be seen in figure 4.7. As mentioned in the previous section, there is an available tree

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

view when a user replies to a respondent’s answer. This tree view’s purpose is to let users choose earlier templates to reuse, and give an overview of the latest state of the conversation. The hierarchy between the nodes of the tree is a nested set model, in which each node has a left and a right integer values. The left and right values in a node represent a set of child nodes. The same concept applies for the each child nodes as well, hence, each child node has a left and right value that represents another set of children, as a part of outer nested set. Figure 4.8 depicts a tree in a nested set model.

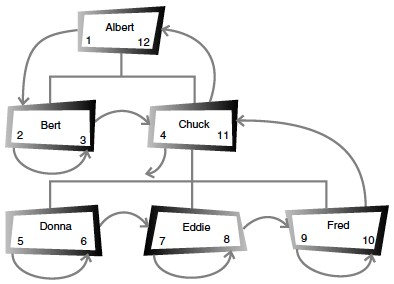


Figure 4.8: A Nested Set Model Tree Hiererchy (Celko, 2004, page 46)

Nested set model helps us on finding a node’s children faster than a usual adjacency set model, where each node only keeps information about its parent node, which is the ID (Celko, 2004, pages 45–47). For example, in order to find all the nodes of the root node named "Albert" in figure 4.8, we just need to find all nodes whose left value is greater than 1, and right value is less than 12. In the adjacency list, however, we first need to find all the nodes whose parent IDs are equal to "Albert’s”, then for those nodes’ IDs, we need to find other nodes whose parent IDs are equal to them, and it continues on until all nodes are finished in the tree in the same way.

The template data model in 4.7 keeps those left and right number information in one column named "ancestry", which is populated by a gem named Ancestry13.

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

*4.3 Evaluation of the Prototype*

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

After getting an overview of the technical implementation of the prototype, the next section will focus on what was missing from this initial design comparing with the initial requirements, as described in section 4.2.1, the problems with it, and what new ideas can emerge with its development.

## Evaluation of the Prototype

In this section, we will analyze the drawbacks of the prototype, despite the fact that it was able to fulfill most of the requirements mentioned in 4.2.1. Some of these drawbacks are related with the development of a prototype from an existing application, whose main purpose is different than that of a regular e-mail client. One of the reasons for making a prototype is to let developers and users conceptualize different requirements or review the existing ones. In here is where we will also mention new and different requirements we realized as we went through the whole process, as well as our future ideas.

### Limitation of the Existing Application

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

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

**Different Inbox Purpose** E-mailValet offers an e-mail client with a task manager feature, as a replacement to everyone’s regular e-mail client. However, a mass e-mail communication is not something a regular person needs to resort to for personal e-mail communication. Moreover, a mass e-mail communication application should be separated from a person’s regular inbox, since it does not need to show all their e-mails, except those related to an e-mail campaign. Combining both inboxes also resulted to some drawbacks on the performance while it was synchronizing with Gmail to fetch new e-mails. It is because E-mailValet was also performing some additional steps during the synchronization on the side, such as checking and comparing message headers, checking for changes in the last 100 e-mails.

### Limitation of the Prototype

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

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

*4.4 Conclusion*

## 4.4 Conclusion

In this section, a mass e-mail communication concept was introduced with an illustrative scenario. We pointed out how it could be possible to reduce the efforts on the researchers’ side while increasing the personalization of the e-mails at the same time. Later on, we converted those ideas into 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 user feedbacks during the final application’s iterative design cycles.