Capstone

Elicitation Plan

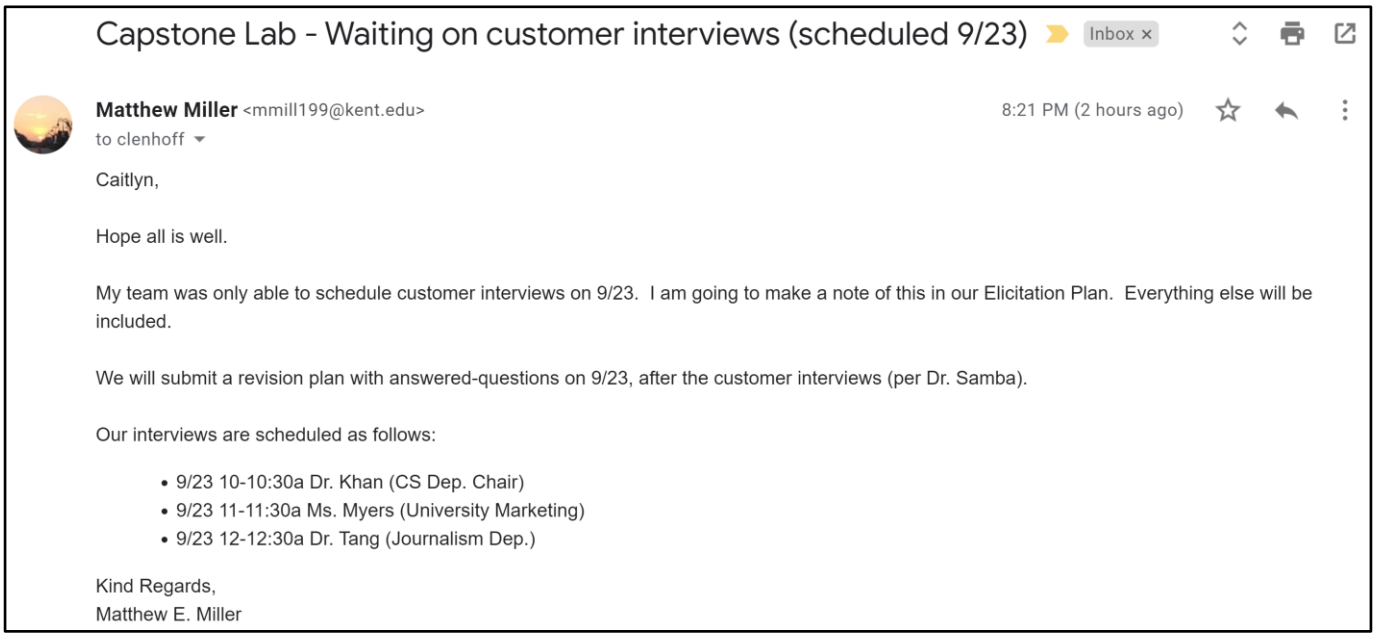
Sentiment Analysis and NLP

Matthew E. Miller, Manisha Kumari, Lucki Ratsavong
9-19-2019

*** Includes Problem Statement from 9/12 – Updates in RED ***

Table of Contents

Instructor Note:



Capstone Lab - Waiting on customer interviews (scheduled 9/23) Inbox ×

Matthew Miller <mmill199@kent.edu> 8:21 PM (2 hours ago)
to clenhoff

Caitlyn,

Hope all is well.

My team was only able to schedule customer interviews on 9/23. I am going to make a note of this in our Elicitation Plan. Everything else will be included.

We will submit a revision plan with answered-questions on 9/23, after the customer interviews (per Dr. Samba).

Our interviews are scheduled as follows:

- 9/23 10-10:30a Dr. Khan (CS Dep. Chair)
- 9/23 11-11:30a Ms. Myers (University Marketing)
- 9/23 12-12:30a Dr. Tang (Journalism Dep.)

Kind Regards,
Matthew E. Miller

*** Includes Problem Statement from 9/12 – Updates in RED ***

Elicitation Plan Description:

Our elicitation plan is focused on exploring our customers' high-level needs. We are primarily concerned with finding out how they will use and benefit from our product. We don't spend much time asking them detail-oriented questions about system architecture, resource usage, and other technical aspects of the software product because many of these things are already known or independent of the user/customer. This is likely a consequence of our product being more like a SaaS product where the .

Our plan includes asking many high-level questions, having brain-storming sessions with customers/clients, and imagining better ways to accommodate and prioritize their needs.

Interview Questions:

Lindsey M Myers, M.A., Director, Marketing Strategy and Research (Destination Kent State)

High Level Understanding:

1. Could you please tell us more about your role?
2. What types of data do you use to make decisions?
3. How impactful is this data? Is it the primary driver of decisions?
4. Are they used to determine marketing efforts and campaigns?
5. How do you analyze this data? Current limitations?
6. What could be improved about the current analysis of this data? Likes and dislikes.
7. What are your thoughts about sentiment analysis?
8. Have you previously encountered obstacles with sentiment analysis?
9. How do you think sentiment analysis could benefit you in your role and your department?

Detail Oriented:

1. For both input and output, what should be the format of the data?
2. Must any data be retained for any period of time?
3. Are there constraints on size of the system (Handheld/Server/PC etc)?
4. Are there any COTS or other constraints on programming language, OS because of existing software components?
5. Is input coming from one or more other systems ("upstream")?
6. Is output going to one or more other systems ("downstream")?
7. What is the protocol for the upstream and downstream systems?
8. Who will use the system?
9. Will there be several types of users?
10. What is the skill level of each user?
11. What kind of training will be required for each type of user?
12. How easy should it be for a user to understand and use the system?
13. How much data will flow through the system?
14. How often will data be received or sent?
15. When and in what ways might the system be changed in the future?
16. How easy should it be to add features to the system?

The interview questions that I have selected for Lindsey Myers are with the goals in mind of understanding her role, how she utilizes data in her role, current efforts, improvements, and how our project could improve the overall process.

Next Page, Customer 2…

*** Includes Problem Statement from 9/12 – Updates in RED ***

Dr. Khan, Computer Science Department Chair (Course Evaluations)

High Level Understanding:

1. How significant or how much weight do student survey of instruction (SSI) carry?
2. What is the impact of SSI?
3. Do they inform decisions on promotions or salary increase? Whether a course is offered more or less in the future? If a Professor is removed from a course or not? Gaining tenure?
4. Informing instructors on their teaching skills?
5. How long does it take from when the SSI are submitted to what results are given?
6. What does the current analysis of SSI data look like? How could it be improved?
7. What differs from the new online SSI versus the paper ones? Limitations of the current process?

Detail Oriented:

1. For both input and output, what should be the format of the data?
2. Must any data be retained for any period of time?
3. Are there constraints on size of the system (Handheld/Server/PC etc)?
4. Are there any COTS or other constraints on programming language, OS because of existing software components?
5. Is input coming from one or more other systems ("upstream")?
6. Is output going to one or more other systems ("downstream")?
7. What is the protocol for the upstream and downstream systems?
8. Who will use the system?
9. Will there be several types of users?
10. What is the skill level of each user?
11. What kind of training will be required for each type of user?
12. How easy should it be for a user to understand and use the system?
13. How much data will flow through the system?
14. How often will data be received or sent?
15. When and in what ways might the system be changed in the future?
16. How easy should it be to add features to the system?

The interview questions that I have selected for Dr. Khan are with the goals in mind of understanding the significance of SSI, current efforts, improvements, and how our project could improve the overall process.

Next Page, Customer 3…

*** Includes Problem Statement from 9/12 – Updates in RED ***

Dr. Tang, Professor - School of Journalism and Design

High Level Understanding:

1. How are the Tweets selected?
2. Do they have a specific hashtag?
3. Include a specific user?
4. Posted on a particular page?
5. What does the process of current analysis look like (i.e. manually)?
6. How long does it take to complete?
7. What results does it offer?
8. Tools used?
9. Limitations?
10. How would sentiment analysis improve upon this process?

Detail Oriented:

1. For both input and output, what should be the format of the data?
2. Must any data be retained for any period of time?
3. Are there constraints on size of the system (Handheld/Server/PC etc)?
4. Are there any COTS or other constraints on programming language, OS because of existing software components?
5. Is input coming from one or more other systems (“upstream”)?
6. Is output going to one or more other systems (“downstream”)?
7. What is the protocol for the upstream and downstream systems?
8. Who will use the system?
9. Will there be several types of users?
10. What is the skill level of each user?
11. What kind of training will be required for each type of user?
12. How easy should it be for a user to understand and use the system?
13. How much data will flow through the system?
14. How often will data be received or sent?
15. When and in what ways might the system be changed in the future?
16. How easy should it be to add features to the system?

The interview questions that I have selected for Dr. Tang are with the goals in mind of understanding the business value resulting from sentiment analysis, journalism school has a client that wants to know acceptance of this e-sport event (impression) a bunch of tweets, go through and do sentiment analysis 10% manual paper for her client started with Dr. Guan, work with him, or pass work onto us.

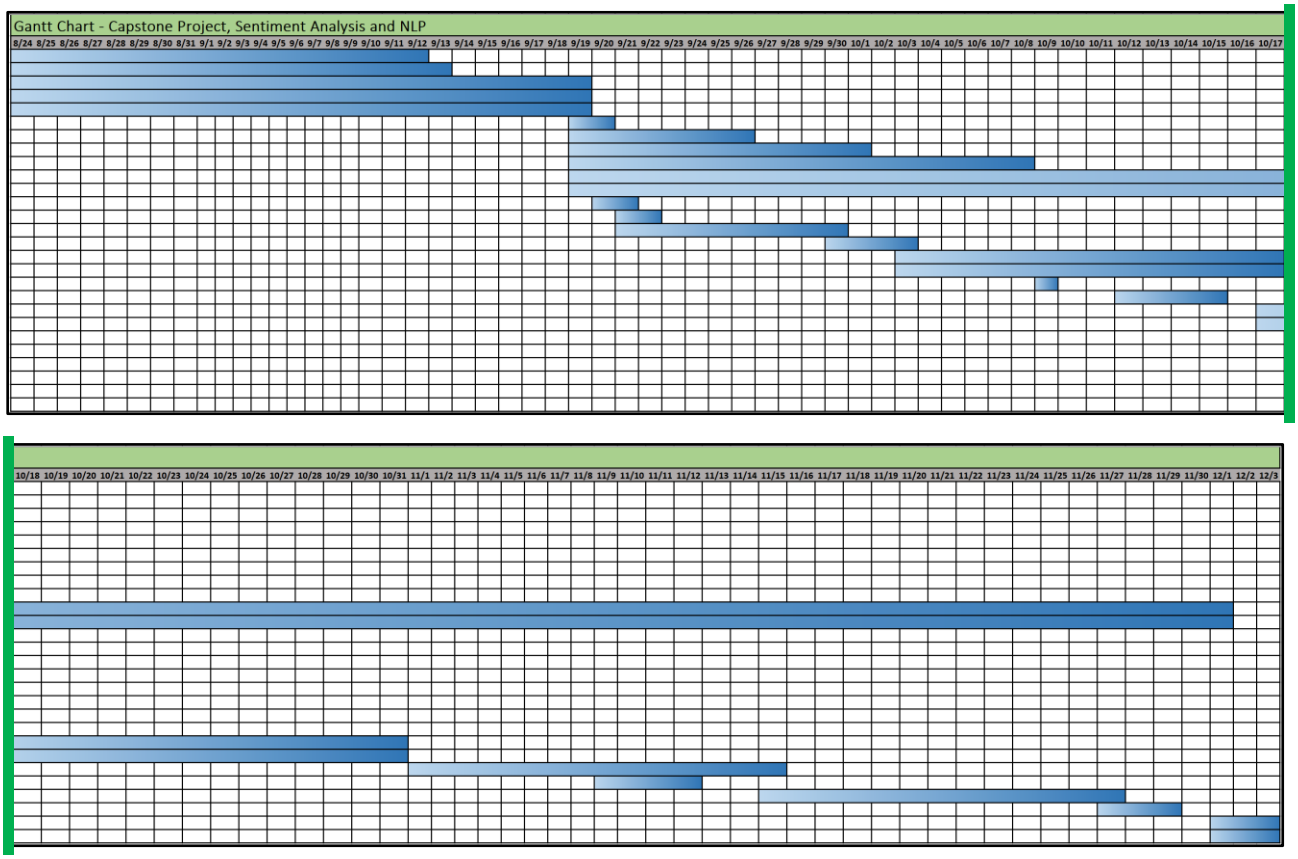
End of Customers.

*** Includes Problem Statement from 9/12 – Updates in RED ***

Timeline:

Schedule				
Task Name	**Start Date**	**End Date**	**Duration**	**Percent Complete**
First Deliverable - Problem Statement	8/24/2019	9/12/2019	19	100%
Email and Visit Potential Customers and Users	8/24/2019	9/13/2019	20	100%
Identify Project Stakeholders	8/24/2019	9/19/2019	26	80%
Interview Customers and Users	9/12/2019	9/19/2019	7	10%
Second Deliverable - Elicitation Reqs. and Statement	9/12/2019	9/19/2019	7	20%
Setup Python Anywhere Server	9/19/2019	9/20/2019	1	100%
Third Deliverable - Functional Reqs. and Project timeline	9/19/2019	9/26/2019	7	0%
Fourth Deliverable - Present Progress Report - Systems Architecture and Software Design	9/19/2019	10/1/2019	12	0%
Fifth Deliverable - Systems Arch. and Software Design via BBLearn	9/19/2019	10/8/2019	19	0%
Develop NLP and Sentiment Analysis - AI Model	9/19/2019	12/1/2019	73	5%
Develop NLP and Sentiment Analysis - Syntax/Dictionary Style	9/19/2019	12/1/2019	73	5%
Setup Project's Dir. Structure	9/20/2019	9/21/2019	1	90%
Setup Remote Request Processing server	9/21/2019	9/22/2019	1	0%
Design ER Diagrams	9/21/2019	9/30/2019	9	60%
Setup Remote Databases - myDB	9/30/2019	10/3/2019	3	0%
Design Processing Server's Program logic	10/3/2019	10/17/2019	14	0%
Design User Interface - Web Application	10/3/2019	10/17/2019	14	0%
Sixth Deliverable - Mid-semester Peer Evaluations via BBLearn	10/9/2019	10/9/2019	0	0%
Seventh Deliverable - Present Progress Report	10/12/2019	10/15/2019	3	0%
Implement Processing Server's Design	10/17/2019	10/31/2019	14	0%
Implement User Interface	10/17/2019	10/31/2019	14	0%
Connect System Components	11/1/2019	11/15/2019	14	0%
Eigth Deliverable - Present Progress Report	11/9/2019	11/12/2019	3	0%
Test, Revise, Maintain	11/15/2019	11/27/2019	12	0%
Curate Supporting Docs.	11/27/2019	11/29/2019	2	0%
Final Deliverable - Group Presentations	12/1/2019	12/3/2019	2	0%
Create Project Presentation	12/1/2019	12/3/2019	2	0%

Gantt Chart (line up with timeline-table):



*** Includes Problem Statement from 9/12 – Updates in RED ***

Project Description:

The goal of this project is to create business value through the use of sentiment analysis and natural language processing (NLP). Sentiment analysis is the process of surmising context, tone, and meaning from written text. It's a sub-discipline of natural language processing (NLP) and it's a powerful tool in a world with increasingly digital text.

Sentiment analysis processing can be performed on any set of digital text. For this reason, we can reasonably extend its reach by using auxiliary technologies to bridge the gap between non-digital text and the digital world by translating non-digital data into a digital form. For example, optical character recognition (OCR) can be used to process images of printed, hand-written text, into digital text. This data can then be processed with sentiment analysis and NLP tools, allowing us to quickly quantify historical documents/data.

Another possibility is the transcription of audio from real-time telephone conversations. First, audio transcription technologies must be used to translate the audio into digital text. Then, this digital text can be processed with sentiment analysis and NLP tools. For example, this might be used to evaluate the performance of a call-center employee.

As you can image, there are many forms of data which can be analyzed with NLP and sentiment analysis. This project is about processing these data sets and providing meaningful feedback to our customers and users.

Problem Description:

Currently, our customers and users are manually interpreting written text from various sources like course evaluations, customer surveys, twitter mentions, etc. The problem is that manual interpretation is slow, unquantified, and prone to discrepancies in understanding between readers. Our goal is provide automatic, programmatic, consistent, and quantifiable interpretations of written text. The outcome is that our customers/users gain better understanding, stronger analytical tools, and an improved ability to provide high-quality service in their professional roles.

We want to make better, more responsive, data-driven organizations and professionals.

Project Objectives:

- Build a functional sentiment analysis and NLP web application.
- Provide meaningful and accurate sentiment analysis and NLP results for customers and users.
- Improve upon existing sentiment analysis and NLP technologies.

Customer Identification:

Customer 1	**Customer 2**
Name: Lindsey M. Myers (Uni. Marketing) **Objectives:** Gain meaningful insight from customer feedback. **Stakeholders:** Employer (KSU), Clients	**Name:** Dr. Khan (CS Department Chair) **Objectives:** Improve quality of course offerings and the department. **Stakeholders:** Employer (KSU), Faculty
Customer 3	**Customer 4**
Name: Dr. Tang (Journalism and Comm.) **Objectives:** Unspecified. **Stakeholders:** Unspecified.	**Name:** Kent State University **Objectives:** Improve students' experience, course offerings, and quality of course offerings. **Stakeholders:** The State, Board of Directors, Customers (students), Employees

*** Includes Problem Statement from 9/12 – Updates in RED ***

User Identification:

Our users are researchers, managers, and other people that want to quantify, previously, unquantified digital text. This pool of users is large, but we have chosen to focus our project on a few target users (our customers). These users typically have a background that includes higher education, management work, critical thinking, and sophisticated use of technology. These users likely work in areas that receive customer feedback which would normally be manually interpreted and possibly quantified using many manual man-hours.

Customers' Constraints:

In some cases, the customer's data is sensitive and cannot be released to aid in the software product's development. The result is that mock data must be manufactured to accurately reflect the characteristics of an authentic dataset. This is a major constraint on development.

The other customer-constraints are schedule and resource based. The project must be completed within three months and programming work must be done within two. Furthermore, our customers are not paying for machine resources, so we must abide by the free-tier-restrictions and limit CPU time to 100s per day, store only 512MB on the remote server, and adapt to a low bandwidth environment.

Risks and Assumptions:

Assumptions	Risks
They care to improve services. They want to quantify written comments, reviews, etc. Quantifiable data is more valuable than unquantifiable, human-interpreted, information.	Low accuracy of the sentiment analysis. Poor AI modeling. Product may not be valuable, useful, usable for customers/users.

Conclusion:

We are building a web-based sentiment analysis and NLP tool to enable our customers and users to quantify previously unquantified text.