EXAMENSARBETE INOM TEKNIK,
GRUNDNIVÅ, 15 HP
STOCKHOLM, SVERIGE 2017

Chatbot as a potential tool for businesses

A study on chatbots made in collaboration with Bisnode

MAX KOTTORP

FILIP JÄDERBERG

KTH
SKOLAN FÖR INDUSTRIELL TEKNIK OCH MANAGEMENT

Abstract

The investigation aims to provide an answer to if a chatbot is a potential complement to an internal service desk of a company. The work has centered around developing a chatbot able to handle simple Q&A-interaction of the internal service desk of Bisnode, the company in question. The chatbot acted as an proof of concept, which then was tested by 15 individuals. The testing was done with pre-defined user scenarios, where the test person ultimately had to fill in a questionnaire with statements related to the overall experience. By summarizing the user evaluations from the questionnaires, combined with an SWOT analysis, the work concluded that a chatbot is indeed a potential complement to an internal service desk of a company, if it handles Q&A-interaction.

Table of Contents

1. Introduction

In 1950 Alan M. Turing published “Computing machinery and intelligence”, addressing the question, “Can machines think?”, which lay the ground upon which we until this day perceive and evaluate artificial intelligence (Turing, 1950). Ever since, computer scientists have tried to build intelligent software, and the first breakthrough was made May 11 1997 when IBM’s Deep Blue computer beat reigning world chess champion Garry Kasparov (Lewis and Writer, 2014). This event showed the true strength of AI to the public for the first time, and was the beginning of a new era of progress within the domain of machine learning and natural language processing.

In general, a chatbot is constructed to simulate a human counterpart in some sort of dialogue. This dialogue is often constructed to serve a greater purpose like information retrieval or service requests, but could also serve an entertaining purpose. The main problem that a chatbot is constructed to solve, is to understand the user's input and respond accordingly. The two main type of bots are retrieval based or generative based. A retrieval based bot uses a predefined set of responses and uses a heuristic to pick the most fitting response from a given input. In contrast, a generative model tries to translate from given input to an output by generating a fitting response without the help of a pre-defined set of responses. (Kojouharov, 2016)

Chatbots have been around since the 1960’s with the introduction of ELIZA, a simple implementation with today's standard but nevertheless gave the illusion of understanding the user. As the development of chatbots progressed, the first bot to allegedly complete the Turing test in 2014, was the bot named Eugene Goostman. Eugene managed to trick 33 % of a panel of judges into believing he was a real boy during a five minutes chat conversation (Aamoth, 2014). The AI community hasn’t settled whether the test was legit or not, but the event created a debate on the chatbots’ part of the near future.

Today, an increasing number of companies are trying to integrate chatbots as a part of their daily processes. Recent studies have shown that implementation of chatbots can be successful in order to automate certain processes related to education, information retrieval, business, e-commerce and amusement (Shawar and Atwell, 2007). Companies like Pizza Hut, Wholefoods and HBO have implemented chatbots as a way to interact and guide customers through their customer service (VentureBeat, 2016).

Because of the fact that chatbots are rapidly increasing in popularity, the spread of intelligent software has become substantial with a large variety of commercial/open-source application programming interfaces, hereinafter abbreviated to API. This development has enabled implementation of sophisticated chatbots made with only little effort and knowledge of the underlying natural language processing and learning algorithms. Some of the prime APIs with low knowledge threshold are Wit.ai owned by Facebook, LUIS developed by Microsoft, and Api.ai owned by Google. Other solutions with more complex features and more powerful potential is IBM’s Watson API and Google’s Tensorflow framework.

1.1 Investors interest

Bisnode was founded in 1989 as a project of the Bonnier company, with a budget of 15 MSEK. The project's goal was to create a profitable business model in the "digital business information solutions" segment. Today Bisnode has grown to have a yearly revenue of 3.5 BSEK with over 2000 employees and is present in over 15 countries.

The three main business areas of Bisnode are credit analysis, qualitative data, and market analysis, the core of all three areas is data and data analysis. Throughout the years Bisnode has developed a network of companies which they are collaborating with, where Bisnode gains access of the company's data and in exchange they keep the data updated and qualitative. This accumulated knowledge has enabled Bisnode to extract insights regarding customer segments and individual customers which every company by itself wouldn't be able to attain.

Within all their main business areas, Bisnode is one of the key actors, and they are at leading position of the qualitative data market in Sweden. However, the last years Bisnode has experienced a stagnant revenue and shrinking margins. This has resulted in an increased focus on business development where Bisnode tries to explore new solutions to create revenue streams.

1.1.1 Service Desk

Bisnode is currently investigating internal and external opportunities with the latest technologies and discoveries of artificial intelligence, with the overall goal to become more innovative. One of the current units of Bisnode being considered is the service desk. The service desk is the internal supporting unit within Bisnode which works with IT-related issues, system administration and order management.

The service desk handles a large amount of different internal errands, which is the primary workload. These kinds of errands are usually solved over email communication and could be put in various categories within service requests and incident support. The service desk's daily errands vary from more complex system administration issues, to Q&A-interaction. Q&A is the least time consuming, yet they handle this type of errands multiple times per day. The questions asked are recurring and often have basic solutions, these question concern e.g. password changes or guidance to specific persons. This, combined with a thorough documentation of handled errands, creates good conditions for an implementation of a chatbot.

1.2 Purpose

The main purpose of the study is to investigate if a chatbot is a suitable tool for a company to optimize and streamline their service desk. This investigation aims to discuss and evaluate the potential of the chatbot technology, and this will be conducted in collaboration with Bisnode. The final conclusions of the investigation can hopefully provide an answer to if the chatbot technology is mature enough to be implemented in an internal service desk of a company.

As AI and machine learning is a subject of discussion worldwide, this report won't only be of interest for Bisnode, but to numerous companies with similar functions. As this report hopefully will shed some light on the opportunities that the chatbot technology brings, the report might be relevant in further exploration on this subject.

1.3 Problem

This study will investigate how a chatbot can replace parts of the current labor associated with the internal service desk. The goal is not to succeed with a practical implementation that Bisnode will use to rationalize the internal service desk, but rather the goal is to investigate the potential of artificial intelligence. To clarify even further, the study is grounded in a solution, the chatbot, where the internal service desk will act as a potential area of implementation.

Overall, this segment (1.3) results in the following question:

"How suitable is a chatbot as a complement to the internal service desk of a company?"

1.4 Specified definition of problem

To be able to reach to a final conclusion, the main question is divided to three sub-questions with regards to relevant aspects of the problem. These three question will be answered later in the report, and will lay the ground on which the final conclusion will be based upon.

- Implementation: "How does the chatbot perform in a user test?"
- Organizational: "What are the organizational consequences of an implementation?"
- Relevancy: "Is the chatbot a relevant tool for the internal service desk?"

1.5 Ethical implication

The emerge of Big Data analysis has raised concerns on its impact on some of our core values in society, as privacy, integrity and confidentiality. As institutions and companies mine larger and larger data sets to predict the future, they also achieve a greater knowledge and awareness of the people (Richards and King, 2014).

The data that has been received from Bisnode contains email conversations between employees at Bisnode and the internal service desk. These conversations treat IT-issues such as registrations, software issues, and orders. To minimize the risk of any integrity loss, all data received from Bisnode has been anonymized, and all information regarding sender and receiver has been erased.

1.6 Social implication

Artificial Intelligence has highlighted the question on whether or not machines will replace humans as workers. McKinsey and Company released an article in May 2016 which stated that 78 % of the predictable physical work could be replaced by a machine in the near future. This mainly entails food preparation, assembly line and packaging objects (Chui et al, 2016). As this report will work as a potential basis for Bisnode when deciding on automatization of internal processes, it might have a social impact in the long term.

2. Theory

2.1 Chatbot: models and problems

2.1.1 Retrieval based model

A chatbot can make use of a retrieval based model, when prompted to reply by a user's input. This model works by matching the input, to a specific pre-defined response. The heuristic to accomplish the mapping from input to response, can be either be rule-based or by using machine learning classifiers. The strength of this model is that the response will be of correct grammar, and most likely relevant to the user. The weakness is that an open domain make collecting responses an impossible task.

2.1.2 Generative model

This model focuses on generating output, without any pre-defined response. These kinds of systems are believed to solve many of the most complex tasks of artificial intelligence (Salakhutdinov Ruslan, 2015). The typical approach is the same as of machine translation, using deep learning techniques on large quantities of data in order to achieve the ability to "translate" from an input to a response. In contrast to the retrieval based model, this approach makes collecting responses on an open domain easier and with the ability to refer to certain entities in the input. The major drawback is that grammatical mistakes in the output is plausible, as well as it requires large quantities of training data. (Kojouharov Stefan, 2016)

2.2 Introduction to machine learning

To get a better understanding of what machine learning is, hereinafter abbreviated to ML, this segment aims to define the concept. There is still no widely accepted definition of what ML is, even though many have tried defining the concept. The starting point is often to divide the concept into two sub-questions, "What is a machine?" and "What is learning?".

In this case machine refers to a computer of some sort, which can process input and produce an output or result. Learning is a less clear concept, the Oxford dictionary defines learning as "*The acquisition of knowledge or skills through study, experience, or being taught*" (Oxford dictionary, 2012). This definition could be viewed as a bit too steep for the ML concept, and should be complemented with Marsland's thoughts on learning "*Learning is what gives us flexibility in our life; the fact that we can adapt to new circumstances and learn new tricks.*" (Marsland, 2015). Together these two definitions of learning, combined with what this report defines as a machine, gives a more understandable picture of ML as a concept.

One could summarize the above stated in a more formal way, "*A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P if its performance at tasks in T, as measured by P, improves with experience E.*" (Mitchell, 1997), which is Mitchell's definition of ML.

In other words, ML is the concept of a computer's or computer program's ability to adapt and change when presented new data, based on already seen data. The areas of usage are many today, and the future possibilities seem endless. ML is used for predicting upcoming events, such as demand shifts of a good or stock price changes. ML is also used for uncovering hidden patterns in large data sets which can be used for finding correlations of variables in a system.

As stated, the areas of usage are many, but all ML is mostly based on analyzing data and requires large quantity of data. For example, to predict a future shift in demand, the computer needs to learn the historical behavior of the market. When presented with new data, the earlier learning is what acts as a foundation for the prediction.

2.3 Introduction to natural language processing

Natural language processing, hereinafter abbreviated to NLP, is a field within computer science that explores human language and how it can be used by a computer to complete certain tasks. Research associated with NLP can be applied to several sub-genres such as natural language understanding, natural language generation and dialogue systems. All of the above sub-genres are based on a foundation of several areas of knowledge including mathematics, linguistics, psychology and computer science. (Gobinda G. Chowdhury, 2003)

The problems associated with natural language varies greatly depending on the situation and context, hence the problems should be dealt with accordingly. To understand why NLP is needed, one must first understand the complexity of the human language. The human language is constantly evolving and contains a comprehensive set of properties, making the input hard for a computer to understand and process. Stated below are some of the main properties in human language, that makes language processing difficult. (José D. Lopes, 2016)

- Lexical ambiguity - some words have different meanings depending on the context. An example could be the word *bank,* which can be interpreted as *"river bank"* or "*money bank".*
- Structural ambiguity - a sentence can be interpreted in multiple ways depending on its structure. Consider the meaning *"They ate pasta with forks.* Where is the fork, in the dish as an ingredient or in the eater's hand?
- Semantic ambiguity - a sentence can be interpreted in multiple ways if the sentence contains an ambiguous phrase or word without an agreed upon definition. The sentence "He is cold" does not say the actual temperature, hence leaving room for interpretations.
- Pragmatic ambiguity - a sentence can be interpreted differently by two agents in a conversation, where the agents does not share the same set of preferences or context. An example of this is the sentence *"Can you pass the salt?",* which can be interpreted as the question of ability, or as a request.

When building a chatbot, the above ambiguities must be taken into consideration in order for the chatbot to perfectly understand the user. What might be of even more importance, is to build a model for identifying the intents of the given input. Intents is a fundamental part of interaction, and is required to be identified in order to give a relevant response.

The identification process of intents can be approached in many different ways, but most of them share the same general idea. The identifier needs to break down the given input into tokens with internal representation. This representation is usually a word in its original form, or a lemmatized version of it, together with a word class tag derived from grammatical rules which is agreed upon before using a lexicon. This is followed by a parsing of the tokens to obtain the different syntactic relationships between the words, usually represented by one or several parse trees depending on the ambiguity of the input. Lastly, a semantical analysis, is done in order to derive the actual meaning of each word and ultimately the intent of the sentence. This semantical analysis is generally based on statistical analyses of large corpuses using statistical models, in order to gather information of the contexts in which specific words appear, e.g. machine learning. (Elizabeth D. Liddy, 2001)

2.4 Evaluation methods

Before a system can be deployed it needs to be tested and evaluated for accuracy on data it was not trained on (Marsland, 2015). The most common evaluation metrics used when evaluating a NLP system is precision and recall. These metrics are suitable for information retrieval based NLP systems, such as search engines. When evaluating a chatbot's performance one also should consider its ability to be perceived as a human, and not only its precision. This is what Turing discussed in his article, and in 1990 the Loebner prize competition was launched as the first formal instantiation of a Turing Test. The Loebner prize is a yearly contest which donates money to the most human-like chatbot. However, in Shawar and Atwell's article published in April 2007, they concluded that:

"*Our general conclusion is that we should NOT adopt an evaluation methodology just because a standard has been established, such as the Loebner Prize evaluation methodology adopted by most chatbot developers. Instead, evaluation should be adapted to the application and to user needs. If the chatbot is meant to be adapted to provide a specific service for users, then the best evaluation is based on whether it achieves that service or task.*" (Shawar and Atwell, 2007)

2.4.1 Qualitative evaluation and observations

Qualitative data differs from quantitative data as it is not represented in numbers, instead qualitative data can take the form of descriptions, quotes from interviewees, and observations. Qualitative analysis and evaluation focus on the nature of somethings and can be represented by themes and patterns. (Preece, Sharp and Rogers, 2015)

One data gathering method which often is used to gather qualitative data is observations. Users are observed while they are performing the activity the product aim to do, which can be done either directly by the investigator, or indirectly through records. These observations can take place in different environments, either in the field or in a controlled environment. Observation could both be used during the design process, helping designers to correct any flaws in the product, and in the final evaluation as a metric for how well the product fulfills its goals. (Preece, Sharp and Rogers, 2015)

2.4.2 Questionnaire

Questionnaire is a well-established method to collect data regarding user's opinions. One can formulate the questions to be both open and closed, similar to interviews, however questionnaires are easier to distribute and enables a more extensive data collection to be made. One important aspect to consider when developing a questionnaire is how the questions are formulated, as they need to be clearly worded and the collected data should be easy to analyze. To achieve this, one should try to ask closed questions and a range of answers should be provided to the person conducting the questionnaire. (Preece, Sharp and Rogers, 2015)

2.5 SWOT analysis

SWOT analysis is a method to identify Strengths, Weaknesses, Opportunities and Weaknesses, and it is often used in project planning and when formulating a business strategy. The SWOT analysis can help to develop a detailed decision basis when evaluating a matter, since it covers critical success and fail factors. The strengths and weaknesses are identified as internal aspects of the matter analyzed, compared to opportunities and threats which are external factors, e.g. market trends or politics. (Dyson, 2002)

3. Method

To be able to answer the question defined in section "1.3 Problem", *"How suitable is a chatbot as a complement to the internal service desk of a company?",* the question was broken down into three sub-questions. Two of these questions, "Is the chatbot a relevant tool for the internal service desk?" and "How does the chatbot perform in a user test?", directly requires a chatbot to be answered, therefore a chatbot was developed and acted as a proof of concept. The questions also require evaluation based on experience achieved through interaction with a chatbot, hence user evaluations were conducted to gather data using the developed chatbot. In order to achieve a more detailed picture of the current state of the chatbot technology, considering the business and organizational aspect, a SWOT analysis was made. Below follows a more detailed description of the chatbot development process and how the user evaluations were conducted.

3.1 Choice of technical solution

The theoretical study and an interview with Konrad Roziewski, IT Director at Bisnode Poland who recently developed a chatbot for the polish office, resulted in the choice of developing a chatbot using an API. A minor study was made with the sole purpose to determine which APIs seemed most suitable given the domain of the internal service desk's workflow. The main actors as of today showed to be: IBM Watson, Api.ai, Wit.ai and LUIS. All of them share the same general concept of extracting intents and entities of given input, calculate a probability and responding with predefined response.

Concurrently, semi-structured qualitative discussions and evaluations with Joachim Karlsson, Group Director of Innovation, Joakim Skog, Innovation Specialist, and Konrad Roziewski led to the choice of Wit.ai. The main reason why this framework was chosen over the others, is that it has support for Swedish natural language processing. As of today, Swedish is the main language of the internal service desk at Bisnode, making Wit.ai the most suitable choice.

3.2 Data gathering and evaluation of data

Data from the service desk was already gathered by documentation of past errands passing through their issue tracking system. However, it was not checked for quality or relevance, so the following step was to investigate whether it was of any use to this project. The conclusion that could be drawn was that the data would be used as support and guidance for training sentences, rather than being used as explicit training data. The main reason behind this, is that the data is in the shape of email conversations and not chat conversations. If the training data is in the form of email, the chatbot will only be able to recognize input with the structure of an email, which is not satisfactory due to the fact that the user input will have the structure of a chat message.

3.3 Development process

The workflow can be described as an iterative process with three main steps. The first step was to identify errands which the chatbot potentially could solve. This was done by conducting a series of interviews with Jonas Fagerlund, technician of the internal service desk at Bisnode. These interviews were done with the purpose to derive trustworthy user scenarios, which then could be translated into issue trees.

Once an issue tree was defined, it was translated into Wit's story function. The story function is basically a way to define how the bot should respond in a certain scenario of the conversation. In addition, every potential user input is marked with its entities and intents which is necessary for the bot to make future predictions.

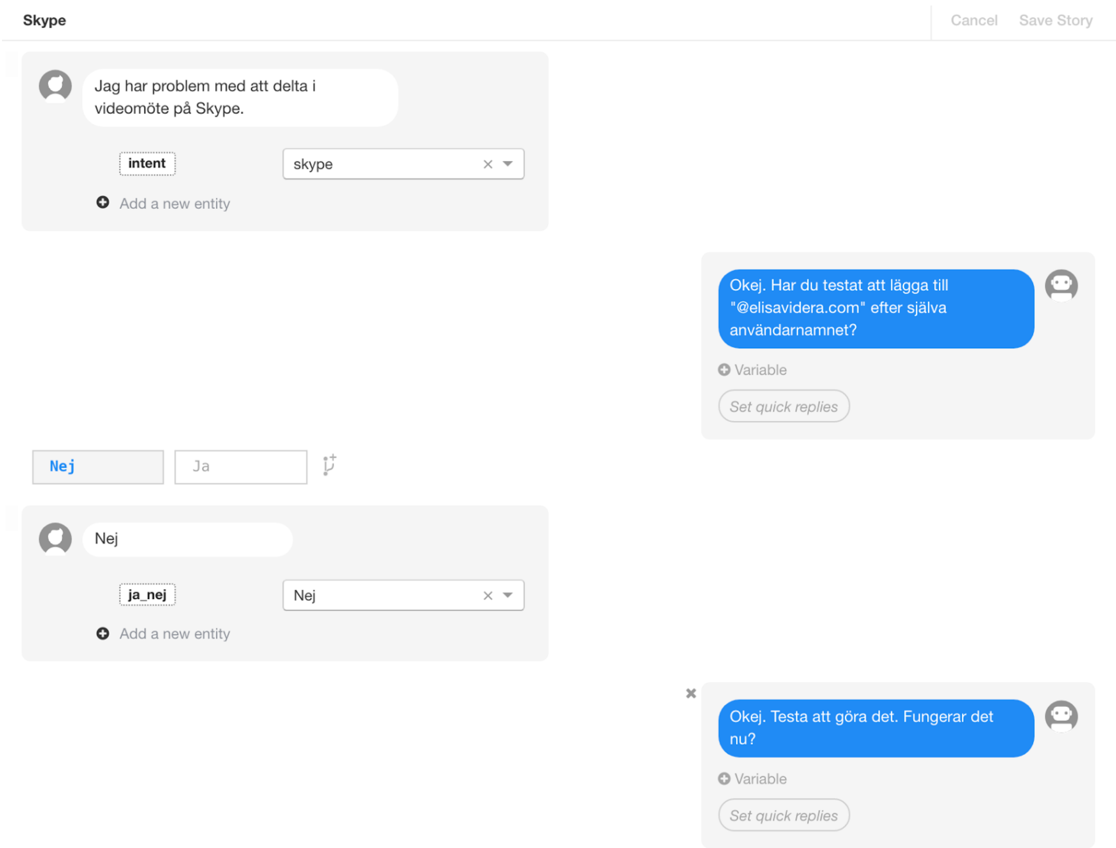

Figure 1: Screenshot from Wit's story function

The final step in the process is to feed the bot with input in order to train it. In 7.3, *"Example of training data sharing the same intent"*, an actual example of training data can be found. This process has been done with fictional input from the derived user scenarios. Once fed, the bot has been tested with the training data once more, in order ensure that it is able to make the correct interpretation. As the bot is used and faced with new unforeseen input by different users, it uses this data to train and become even more accurate over time. (Wit.ai, 2016)

3.4 Evaluation

Shawar and Atwell concluded in their report that the best evaluation of a chatbot is not to use quantitative metrics, as precision and recall, instead one should use qualitative metrics (Shawar and Atwell, 2007). During the interview held with Konrad Roziewski he stated that he shared this opinion and recommended an evaluation focusing on qualitative metrics of the chatbot. Based upon this, the evaluations of the chatbot was conducted focusing on qualitative properties.

The data was gathered through observations in a controlled environment where 15 different people, working at companies with a similar internal service desk, where given a set of predefined tasks to complete using the chatbot. Once the tasks were completed, each person had to answer a short survey of the experience, which altogether resulted in the final evaluation of the chatbot. This result is shown in section "4.2 User evaluation".

Continuous evaluations were also performed during the development of the chatbot, where Jonas Fagerlund tested its performance and got to comment it. From these evaluations the training of the chatbot could easily be modified to be more satisfying to the end user.

4 Results

4.1 Presentation of the bot

The final version of the bot is capable of handling simple QA-conversations in Swedish with guiding troubleshooting within the domain of the internal service desk at Bisnode. The bot is built with an underlying NLP method close to "Probabilistic Context Free Grammar" (Duckling Wit.ai, 2017) and is able to parse text and extract vital information given by the user. Every input is given a probability in order to give the most fitting predefined response, with respect to the context.

Figure 2: Screenshot of an actual conversation with the bot

4.1.1 Topics and issue tree

The bot can handle five main topics of conversations. These topics are:

- VPN
- Skype
- Outlook
- Mail in personal mobile device
- Order fulfillment

Every main topic has its own issue tree with different underlying structures. The issue tree is a fictional representation of how the user is believed to interact with the bot, and has been derived from real user scenarios in collaboration with Jonas Fagerlund. The bot is not forced to follow the issue tree scheme, since it can choose to skip parts of a conversation if enough information is given from the start. In addition, the bot is also able to respond to various greetings, uncertainties and farewells.

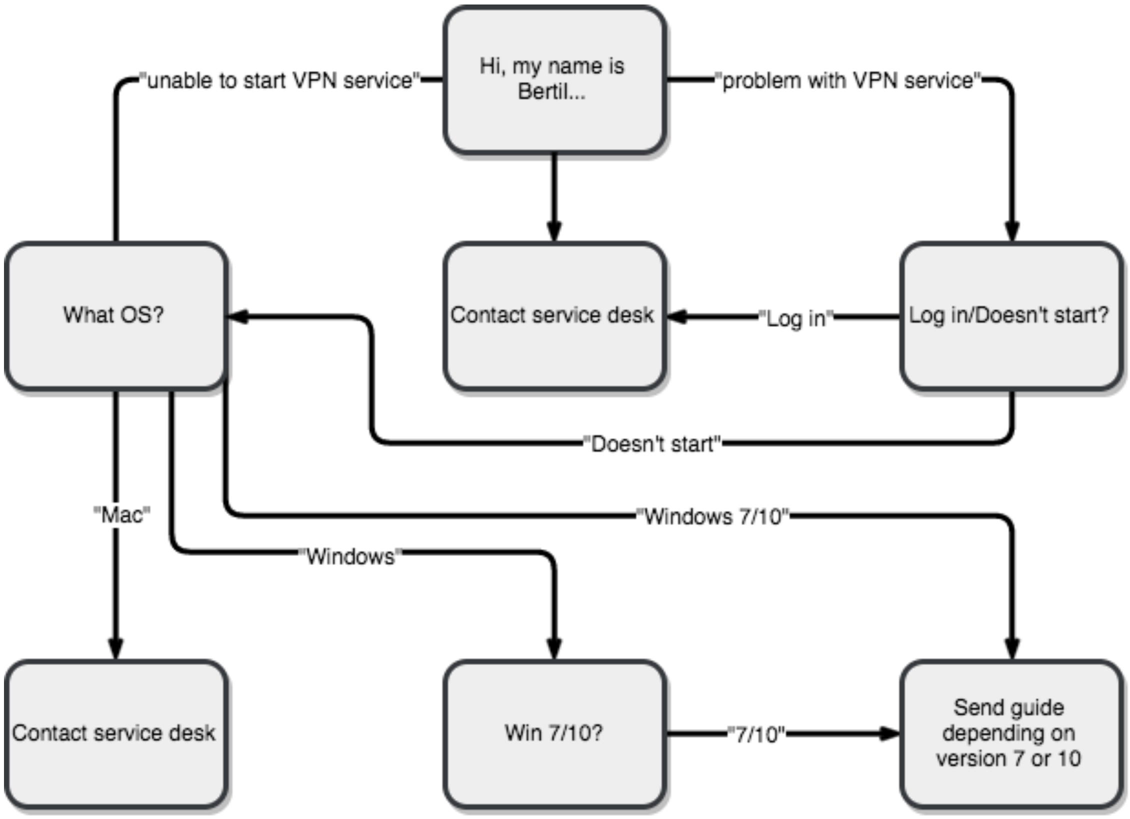

Figure 3: Conversation structure of VPN issue

4.1.2 Email/Order

The internal service desk at Bisnode acts as a proxy when co-workers want to make an hardware order. The usual scenario is that co-workers send a mail with order details to the service desk, which then is processed and granted permission. The issue is that these orders are frequently missing essential information and requires the service desk to contact the person in order to gain all essential information. This information includes: name, product, cost center, company and permission from manager.

Due to this issue, the bot has been given the ability to ensure that all required information is present, if a co-worker wants to make an order. Throughout the conversation, the bot stores the necessary information in a dictionary using Python, which is then used to summarize an email which is passed on using the smtplib module. The service desk ultimately receives a full order description, where the bot has acted as a quality assurer.

4.1.3 Implementation

The bot is mainly developed using WIT and Python, which gives the bot its fundamental ability to understand and make actions. In order to deploy it to a platform, the source code of the bot is connected to the cloud platform Heroku which provides the ability to communicate with Facebook Messenger using HTTP webhooks.

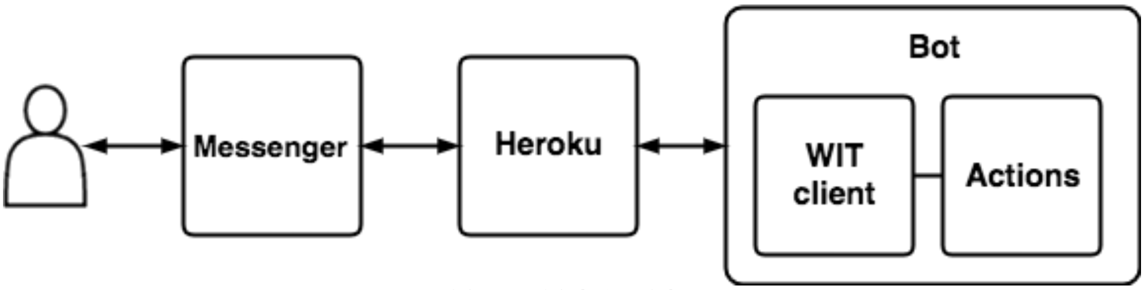

Figure 4: Schematic of the bot's underlying structure

4.2 User evaluation

The following results are the user evaluations performed on 15 different users. Every user completed the same predefined tasks and had the same scenarios to relate to. The scenarios can be found in section "7.1 Scenarios". Once completed, the user filled out a questionnaire with four statements which are presented below in figure 5. The user gave an integer between 1-5 to each statement, where 1 represents "I fully disagree" and 5 represents "I fully agree".

Statement	User average rating
"I consider the quality of the conversation to be good"	4,7
"I understood what the bot replied to me"	4,1
"I consider that the bot understood me"	4,1
"I would like to use a chatbot in a similar scenario"	4,8

Figure 5: Evaluation chart

5 Analysis

5.1 SWOT - using an API

In this section follows a SWOT analysis of the chatbot technology, and more specific the SWOT analysis will treat the aspect of using an "Application Programming Interface" (API), since it is the method used in this report. The SWOT analysis aims to provide a detailed picture of what strengths, weaknesses, opportunities and threats are connected to the chatbot technology for a company considering developing and/or implement a chatbot into their business. The analysis is based on experience using Wit.ai as an API when developing the chatbot, and the findings of this report.

Strengths	**Weaknesses**
Automatization of standardized conversation Low knowledge barriers Low Costs Improves over time Implementation	Only in Beta version for Swedish Correct data Robustness
Opportunities	**Threats**
Ride the wave of development Numerous functionalities Enhanced work climate	Costs Integrity Availability Acceptability

Figure 6: SWOT analysis

5.1.1 Strengths

Automatization of standardized conversation

This study has shown that the chatbot technology has reached the point where it is able to converse and lead a user through a predefined issue/conversation tree. These conversations are suitable to derive from simple problems, where there is only a few possible answers from the users. This creates the possibility to automate the issue handling by creating standardized conversations where the bot guides the user, step by step, through an issue tree.

Low knowledge barriers

The most complex part of the chatbot system is the natural language processing (NLP) and its ability to predict the context of a user input. To create such a system one needs not only high skill in coding, one also needs to grasp the fundamentals of NLP and machine learning (ML) to its fullest. However, the emerge of API solutions has eliminated that part of the development as they provide state of the art systems, and the focus can shift to tailor the chatbot's functionalities instead of the underlying code.

Low Costs
Most APIs provide their services for free, or to a low price. The cost of a chatbot is therefore in many cases limited to its implementation-, development-, and maintenance costs. However, as the API solutions already has developed the fundamentals of the chatbot, the development/tailoring often do not demand a large amount of time.

Improves over time
The chatbot's prediction algorithm is based on ML, which means that the more data it trains on, the more accurate predictions it is able to perform. This property entails that the more user input the chatbot gets, the more data it bases its predictions on. This results in that it will not only be able to better distinguish the context of an input, it will also dynamically change its predictions if the user inputs changes.

Implementation
One strength is the flexibility regarding implementation and where to deploy the chatbot. In this study, Facebook messenger was chosen as platform for the chatbot, but most API supports many more platforms e.g. Skype, Slack and own websites. This enables the developer to tailor the deployment according to end user preferences. In the case of Bisnode, Skype is used for internal communication, and therefore may be suitable for implementation of the chatbot.

5.1.2 Weaknesses

Only in Beta version for Swedish
The standard language for all API solutions is English, and those who support other languages only provide a Beta-version in the available languages. This limits the API solutions available if one would like to implement the chatbot in a different language. Furthermore, it also affects the quality of the chatbot. The NLP system is lacking much of the training and its information retrieval and prediction will therefore not perform with equal quality as for other languages, e.g. Swedish.

Correct data
As mentioned, the prediction algorithm is based on ML, which requires training to recognize and retrieve key information from user input. To develop a robust chatbot, which can handle a variety of input, the system therefore needs a data set to train on, which should be equivalent of the user input. To obtain this data, the developers either need to guess how a user would converse with the bot or they need to have documented previous conversations with users. However, as the transformation often goes from email (or equivalent) to chatbot, the documented conversation is not very useful as users write on a different form when conversing over email compared to a chat window. Therefore, the data set is difficult to obtain without any interaction with the users, and one often needs to deploy a soft launch to gather correct data.

Robustness
Depending on the scope of the bot, the range of possible questions from users will vary. To create a robust chatbot all possible inputs of the users must be considered when developed, otherwise the chatbot will fail to understand the user's intents and answer incorrectly. The developer has two options, either set a clear scope of the chatbot and try to lead the users through the conversations. This limits the dynamic in the conversations and the bot risks to be perceived as unintelligent. The other option is to widen the scope and let the users steer the conversations, which will increase the complexity of the chatbot as it has to understand a wider range of inputs. However, this can interfere with the core function of the bot and the prediction model will be less accurate. Furthermore, one need to consider another aspect of how the user input will be perceived by the chatbot, regardless of whether the scope is narrow or wide. There are several ways for a user to express the same intent, and the complexity of the system increases with larger vocabularies. The larger the vocabulary is, the longer the processing times will be as well as a substantial increase of perplexity errors (Brennan, 1998). Hence, to create a robust chatbot the developer need to cover a wide range of expressions of the same intents to ensure the chatbot's precision. This phenomenon could also create another issue,

that the chatbot understands the user, but answers with a predefined answer using a different vocabulary than the user.

5.1.3 Opportunities

Ride the wave of development

The chatbot technology is currently in an early stage of development, but there are many who see a great potential in the future. A study by McKinsey argued that most repetitive work can be replaced by AI in the near future, among those jobs customer service is included (Michael Chuy, James Manyika, and Mehdi Miremadi, 2016). Thus, it is valuable to obtain knowledge of the technology in an early stage when it is still free and to do so in a controlled environment as in the case of this study. This would enable a rapid adoption of the technology, e.g. in the area of customer service, when it is mature. Furthermore, it could lead to an advantage towards competitors.

Numerous functionalities

Many of the API's available online support integration of additional code to the chatbot. The code can be triggered by actions from the chatbot which reacts to certain input from the users. This enables developers to write own code with tailored functions to specific actions, e.g. the chatbot retrieves information from a website or sends an email. Thus, the chatbot can execute any given function from the developer, which creates flexibility.

Enhanced work climate

Today the chatbot's strength lies in handling issues with a clear step-by-step solution. These issues can be categorized as repetitive and easy to perform for a human, and it could be argued that they are less stimulating than other tasks. Nevertheless, in the case of Bisnode, 10-15% of the daily errands for the service desk consist of this type of issues. If a chatbot could partly or completely manage these issues, the employees could instead focus on more value creating tasks, resulting in a more stimulating working environment.

5.1.4 Threats

Costs

At this point in time it is completely cost free to use an API implementation for the chatbot. This is due to the fact that the API gets access to all data that the user's input which then can be used to train the system, and over time the API will become more accurate. The owners of the API therefore have incitement to offer their API for free, but only as long as they believe their system needs to be trained. This causes concern for a lock-in effect which might get costly in the future, where the owners of the API decide to charge for the service and the company has become dependent on the chatbot.

Integrity

As stated in section Costs, all data that users input to the chatbot are accessible of the API owners. This results in an integrity loss for the users which can be problematic depending on the function of the chatbot. In the case of customer service the users might not want information regarding their customer to be leaked as this could be sold to a competitor.

Availability

Even though the chatbot is deployed on a website or a platform, the requests will always go via Wit.ai. This causes concern regarding availability as Wit has the authorization to shut down the service for e.g. maintenance. Thus, the company cannot ensure availability of the service even though their website is accessible.

Acceptability

The introduction of the chatbot will affect the user behavior, and it is therefore important to consider who the end user is and if he/she has a high acceptability towards change. If the users have low acceptability the chatbot may be unused and in the case of Bisnode this might result in a continued email contact with the service desk.

6. Discussion

6.1 Development and implementation of a chatbot

The following segment aims to discuss the second break down question defined in segment "1.4 Specified definition of problem", "What are the organizational consequences of an implementation?". The discussion will detail how we believe the development and ownership should be orchestrated in an organization. Furthermore, it is based on the findings of this report and the experiences achieved when working with Bisnode.

One key aspect to consider when initiating a project is who should own and lead it. Furthermore, it's important to consider who will be affected by the project and how this might affect the progression and result of the project. This of course varies from case to case, but when developing a chatbot, this study has shown that it is important that the developer understands the problems and conversations that the bot will handle. In the case of Bisnode the most suitable owner of the chatbot is the service desk, as they have the best understanding of how the end user behaves and what problems the chatbot could handle. Furthermore, this configuration also addresses the second aspect, of important stakeholders to the project. By handing the ownership to the service desk, it becomes easier to anchor the development as an opportunity instead of a threat. The chatbot's function doesn't necessarily mean that workforce will be rationalized, rather that more time can be allocated to the more complex issues that they handle.

Another aspect to consider is how the group should acquire enough knowledge to be able to develop the chatbot. Firstly, one need to bear in mind that there is no need for coding if the chatbot only should be able to converse. To be able to create a chatbot without extra functions, one does not need any previous training in chatbot development either, the API's provides documentation on how the chatbot works and how to develop them. However, to be able to further develop the chatbot as adding other functionalities it requires basic knowledge of supported coding languages such as JavaScript or Python (Wit.ai, 2016). There are two ways to acquire this knowledge, either by hiring or by educate all or some parts of the team. In the case of Bisnode, the recommendation is to invest in existing personnel and educate a part of the team in the basic of coding related to chatbots. This due to the fact that is stated in the paragraph above, they carry most knowledge of the end user and are therefore most suitable to develop the chatbot.

The project of developing a chatbot stretches beyond deployment. The chatbot will need maintenance in form of tailoring the conversation to be more pedagogical for the end user, and further development in form of new conversations. This part is crucial for a successful implementation as the chatbot most certainly won't work as desired in the beginning, see segment "5.1.2 Weaknesses - Correct Data", and the developers need to gather input from users and amend the chatbot to handle their way of interacting with the chatbot. The risk if the bot is not amended to fit the users is that the users goes back to the old way of handling their problem, which in Bisnode's case would be to email the service desk.

Furthermore, one need to consider how to change the user behavior to start interacting with the bot instead of emailing the service desk. As mentioned in the paragraph above, one important aspect is to tailor the bot in an early stage when users has given input on the bot. Furthermore, it is important to consider where to implement the chatbot. It should be easy and natural for the users to interact with the chatbot and therefore one needs to implement the chatbot in a channel which the users already use in some way. In the case of Bisnode Skype for Business is a suitable option as a substantial part of the internal communication is in that channel.

6.2 Final conclusions

The user evaluation clearly indicates that the user acceptability is high for the chatbot technology, with a 4,8/5 score on the question "I would like to use a chatbot in a similar scenario". Furthermore, the evaluation shows that the chatbot technology has reached the point where the users find the performance satisfactory, and that the chatbot could understand the input as well as answering in an intelligent manner (Scored above 4,0/5 on both questions). This finding answers the question "How does the chatbot perform in a user test?", which is one of the three break down questions of the problem definition *"How suitable is a chatbot as a complement to the internal service desk of a company?".* In the light of the user evaluation we argue that the performance of the chatbot is more than sufficiently satisfying for a company to implement it as a complement to the internal desk.

As mentioned, all questions and problems the chatbot was trained to handle was derived from real user scenarios and formulated in collaboration with Jonas Fagerlund. This property enables the user evaluations to not only reflect users' opinions regarding the chatbot technology in general, but also to indicate the users' experiences of how well the chatbot can handle the same questions as the service desk does. As a result, this can help answering the third break down question "Is the chatbot a relevant tool for the internal service desk?". All questionnaire questions scored an average over 4,0/5, which indicates that the chatbot, in a satisfying way, can handle the same questions as the service desk does. Although one need to consider that the chatbot only has been tested on a limited amount of issues, and all of these issues has a low complexity and are highly repetitive. Therefore, we cannot conclude that the chatbot can handle all of the service desk problems, but it has shown to be able to handle Q&A-interaction.

The second break down question, "What are the organizational consequences of an implementation?" is discussed in segment "6.1 Development and implementation of a chatbot". As stated in the first paragraph, the discussion is based on the findings of this report, and the experience achieved working with Bisnode. Therefore, we cannot present any conclusions on the subject, rather to present a basis for discussion for further investigation and research.

The overall conclusion from the discussion and the result is that the chatbot technology is suitable as a complement to the internal service desk of a company. This is supported by the strengths and opportunities presented in the SWOT analysis. However, we believe that to present and definite answer and a complete roadmap on the implementation process, further research needs to be conducted in regards of the organizational consequences of an implementation. Furthermore, an investigation on how to manage the weaknesses and threats presented in the SWOT analysis needs to be conducted.

7. List of references

Computing machinery and intelligence 1950, Alan Mathison Turing [23 Februari 2017]
http://www.loebner.net/Prizef/TuringArticle.html

Livescience 2014, History of AI, Tanya Lewis, Staff Writer [23 Februari 2017]
http://www.livescience.com/49007-history-of-artificial-intelligence.html

Time 2014, Interview with Eugene Goostman, the Fake Kid Who Passed the Turing Test, Doug Aamoth [25 Februari 2017]
http://time.com/2847900/eugene-goostman-turing-test/

JLCL 2007, Chatbots: Are they Really Useful?, Bayan Abu Shawar, Eric Atwell [3 Mars 2017]
http://www.jlcl.org/2007_Heft1/Bayan_Abu-Shawar_and_Eric_Atwell.pdf

Venturebeat 2016, Bot boom will blow up B2B,B2C, Venturebeat Staff [3 Mars 2017]
http://venturebeat.com/2016/07/24/better-believe-the-bot-boom-is-blowing-up-big-for-b2b-b2c-businesses-vb-live/

Chatbotslife 2016, Ultimate Guide to Leveraging NLP & Machine Learning for your Chatbot, Stefan Kojouharov [23 Februari 2017]
https://chatbotslife.com/ultimate-guide-to-leveraging-nlp-machine-learning-for-you-chatbot-531ff2dd870c

Wake Forest Law Review 2014, Big Data Ethics, Richards, Neil M. and King, Jonathan H. [3 Mars 2017]
https://poseidon01.ssrn.com/delivery.php?ID=358086124121107117102125125115082066005053028051000082065087071119118115101070068106013057098023052127028119127088001083102079109016014016004002025112114005030096073066040001112020030009026127110089093122089125117086093104067095086008123085003069104&EXT=pdf

McKinsey insights 2015, An executive's guide to machine learning, Michael Chui, James Manyika, and Mehdi Miremadi [3 Mars 2017]
http://www.mckinsey.com/industries/high-tech/our-insights/an-executives-guide-to-machine-learning

The Annual Review of Statistics and Its Application 2015, Learning Deep Generative Models, Ruslan Salakhutdinov [3 Mars 2017]
http://www.cs.cmu.edu/~rsalakhu/papers/annrev.pdf

Stephen Marsland, Machine Learning An Algorithmtic Perspective Second Edition - Taylor & Francis Group 2015, pp. 4.

Tom Mitchell, Machine Learning, McGaw-Hill 1997, pp. 2-5

Annual Review of Information Science and Technology 2003, Natural language processing, Chowdhury G. pp. 51-89

Introduction to natural language processing, Jose D. Lopes 2016
http://www.csc.kth.se/~jdlopes/teaching/sprakteknologi1617/Introduction.pdf

Encyclopedia of Library and Information Science 2001, Liddy, E.D. Natural Language Processing, pp 5-10 [3 Mars 2017]
http://surface.syr.edu/cgi/viewcontent.cgi?article=1043&context=istpub

Jenny Preece, Helen Sharp, Yvonne Rogers, Interaction Design: Beyond Human-Computer Interaction, WILEY 2015, pp. 362-396

Robert G. Dyson, 2002, Strategic Development and SWOT analysis at the University of Warwick, European Journal of Operational Research, pp. 632

Wit.ai 2016, Recipes [5 Maj 2017]
https://wit.ai/docs/recipes

Duckling Wit.ai 2017, Introduction [5 Maj 2017]
https://duckling.wit.ai/

Susan E. Brennan, The Vocabulary Problem in Spoken Dialogue Systems, Stony Brook 1998

McKinsey insights 2016, Where machines could replace humans—and where they can't (yet), Michael Chui, James Manyika, and Mehdi Miremadi [5 Maj 2017]
http://www.mckinsey.com/business-functions/digital-mckinsey/our-insights/where-machines-could-replace-humans-and-where-they-cant-yet

8. Appendix

8.1 Scenarios

1. Your mail does not work on your iPhone/Android cellphone. It is the first time you are using it. Once you are provided with a guide, the problem is solved.
2. You cannot connect to a video conference using Skype. You have tried to add "@elisavidera.com" after your username, and you are not connected to the network using a cable. If asked to connect using a cable, the problem is solved.
3. You want to order X amount of HDMI-cables. You work at "Bisnode Sverige, Avdelning B". If asked, you have not asked for permission from your manager to do this order. Your name is Filip Jäderberg, and your manager's name is Max Kottorp.

8.2 Screenshots of chatbot

Figure 8: Example of typical conversation

Figure 9: Example where the bot's understanding is limited, leading to redundant questions

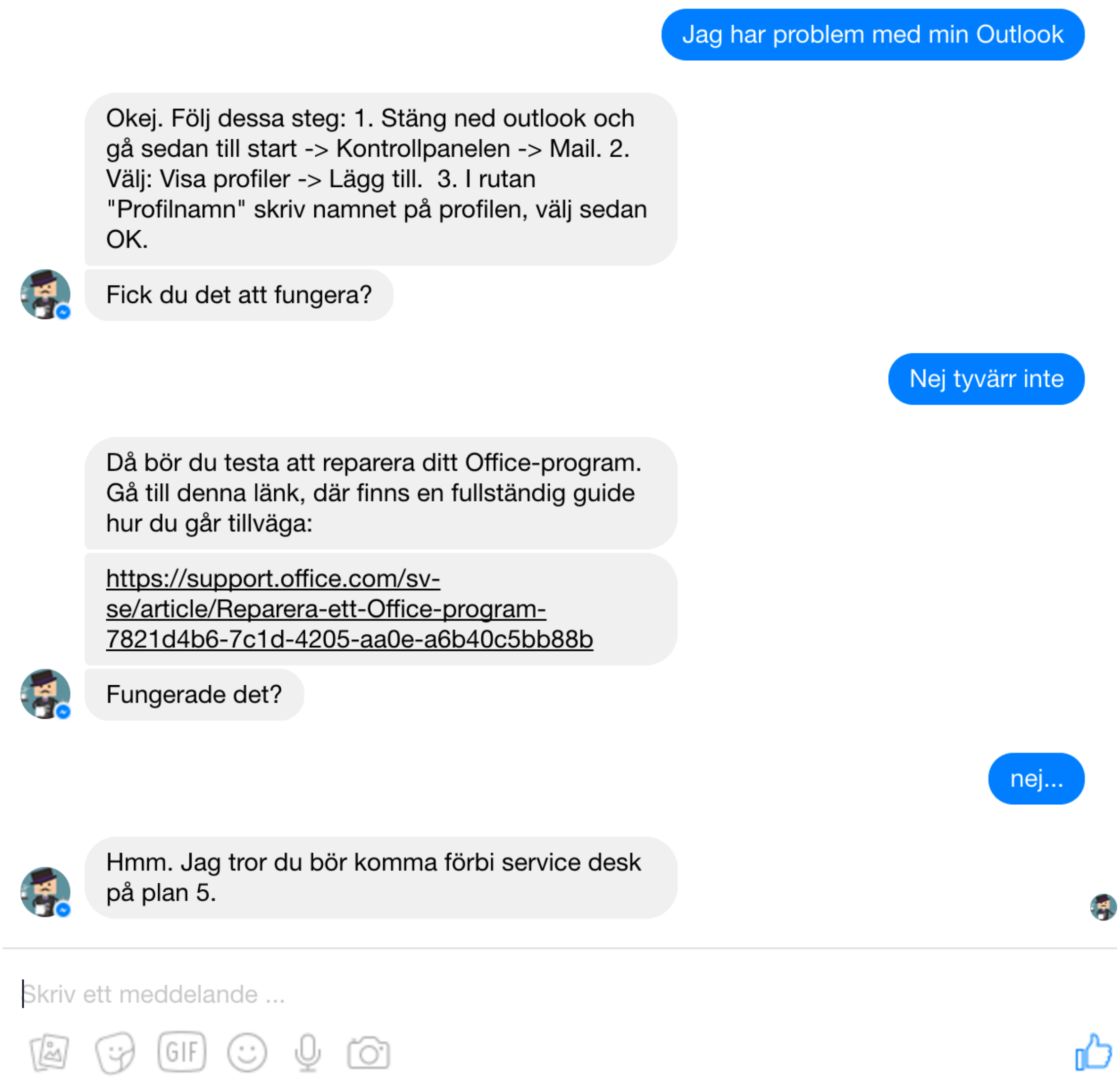

Figure 10: Example of conversation where the bot is unable to help further

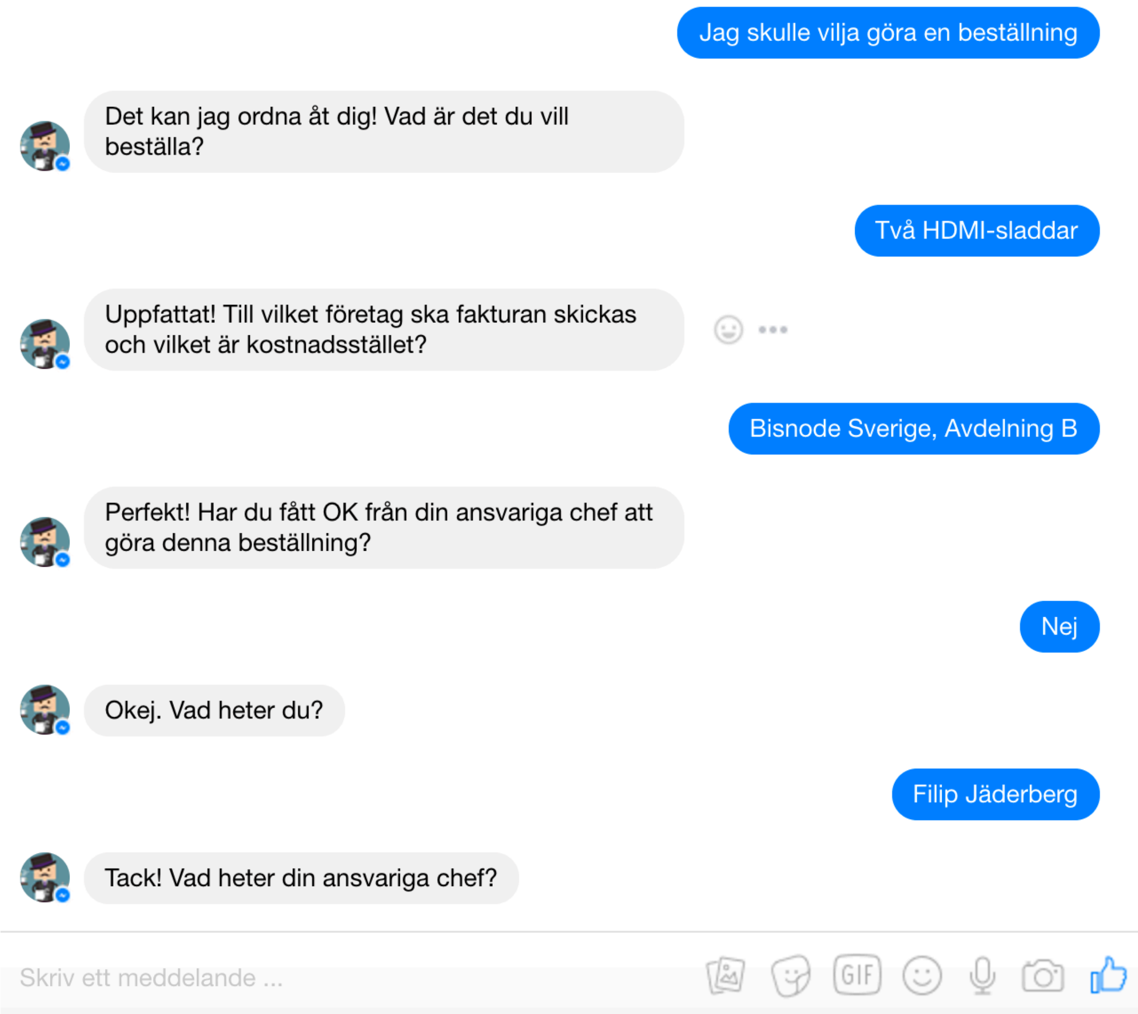

Figure 11: Example of order fulfillment through the bot

8.3 Example of training data sharing the same intent

Mailen strular
Min mail strular
Mailen funkar inte
Mailen fungerar inte
Min mail funkar inte
Mailen har slutat fungera
Min mail har slutat fungera
Jag har problem med mailen
Jag får inte min mail att funka
Jag har problem med min mail
Hur får jag min mail att funka?
Jag får inte min mail att fungera
Hur får jag min mail att fungera?
Hur kan jag få min mail att funka?
Hur kan jag få min mail att fungera?