# Automatization of Customer Service: Could a Chatbot Answer Statistical Questions?

Katri Haverinen, Statistics Finland, katri.haverinen@stat.fi

**Abstract**

Statistics Finland has a customer service, where customers can ask for information about statistics. One of the channels used by the customer service is a web-based chat, which has been quite popular. In fact, so popular that it cannot be opened on the whole Statistics Finland website: if it was, the customer service would drown in questions. The service is good, but it can only be given during office hours, and it ties well-educated staff to answer also routine questions.

As automatization of customer service is being studied and tried out in different contexts, also Statistics Finland launched a project to discover whether it could be useful to use chatbot technology to aid customer service. The project gathered a sample of chat logs from the current human-based service and studied it. Topics of interest included among others, what topics the questions covered, whether there were easy or often repeated questions, and what kind of language the customers used.

The scope of the project was limited, as is the generalizability of the results. However, in the questions collected, it seems that, depending on the topic, about one half could potentially be answered by a chatbot based on a knowledge base. On the other hand, even though questions about the population dominated, the same questions did not seem to repeat very much. The customers used mostly plain language, which is potentially good for automatization but needs to be considered when designing answers for the chatbot.

Overall, the project group cautiously recommended exploring automatization, as there seems to be some potential in the questions, and other alternatives would not substantially improve the service in the chat. Potential automatization is dependent on future technology choices, and it will be considered based on choices made in the ongoing web service renewal.

**Keywords**: automatization, customer service, chatbots

## Background

Statistics Finland answers customers’ questions about statistics by phone, by email, and since 2013, also by chat. The web-based chat service has had its users since the beginning, and it has become increasingly popular over the years. In fact, the service is so popular that it needs to be “hidden” rather deep in the organization’s web service. On an occasion when a technical issue caused the chat to be visible on every page of the service, the resulting amount of questions overwhelmed the customer service completely. Although the service is good, it can only be given during office hours, and it requires well-educated staff to be available to answer also routine questions.

As part of its ongoing large-scale web service renewal, Statistics Finland launched a project to investigate whether it would be useful to partially automatize the chat service. The project analyzed chat discussions with customers to determine what kind of questions and topics are discussed in the chat and what kind of language the customers use. It also investigated the types of solutions being used in other Finnish organizations and considered the potential effects of automatization. The project cautiously recommended exploring automatization, as the analysis seemed to point towards some potential. However, issues such as technology choices should be considered carefully.

## Analysis of chat discussions

The project collected chat conversations between the customers and the customer service staff. The conversations were collected on four different days, with the chat open in different parts of the Statistics Finland web service on each of the days. The purpose was to see what kind of questions would be asked if the location of the chat differed from the usual. At the same time, the chat could not be opened very widely, to not overtax the customer service. There was some analysis available on a larger mass of conversations previously (Kuoppa 2017), but this analysis was less detailed and there was no variation in introduced into the chat locations.

Due to the timing of the project, the four days of collecting discussions took place in the late spring and early summer. This means that any time-based variation in the questions was not included. Naturally, four days is also a limited time to gather discussions. In total, the analysis included 186 discussions between customers and customer service staff. On these 186 discussions, several different kinds of analyses were performed.

### Topics of the questions

The first point of interest was the topic of each question. The questions asked by customers were divided into seven categories based on topics. The categories, in frequency order, were as follows:

* questions about a statistical number (such as unemployment rate or number of inhabitants)
* contacts with no clear need for information
* questions about classifications
* questions about products and services
* questions about concepts and definitions
* questions about data collection
* questions about information not collected by Statistics Finland

More than half of the questions, about 57 %, were about a single statistical number (see Figure 1). The second most common group was contacts where there was no clear need for information. These were for instance inappropriate messages, or, rarely, cases where customers opened the chat to inform the customer service that they do not need help. These customers seem to have interpreted the opening message of the chat window (“Hello, may I help you?”) as having received a message from the customer service.

Figure 1: Contacts by type. Questions about statistical figures were the most common type, followed by contacts that had no clear information need.

The questions were also classified into the 26 official topics in use at Statistics Finland. These topics include, among others, housing, living conditions, energy and education. When looking at the most common type of questions, concerning statistical numbers, the clearly most common topic was population, with 39 questions, while the second most common topic, enterprises, had 12 questions. See Figure 2 for details on the topics of the statistical figure question type.

Figure 2: Questions about statistical figures by topic. The most common topic was population, followed by enterprises.

### Repeating questions

Another interesting point was whether the same questions tended to repeat. Seeing that in natural language, the same question can be asked in numerous unpredictable ways, a definition was needed for what it is to be a repeating question. Two questions were considered repetitions of each other if they could be answered with the same (automatized) textual answer, possibly including a link. It should be noted that this is quite a broad definition. For instance, questions with quite different information needs could be answered by a link to the same listing of database tables and some general instructions on making selections in the database.

With even this broad definition, repetition in the questions was not very common. 57 questions were considered instances of repeating questions, that is, approximately 31%. These repeating questions were instances of 12 different “questions”, meaning that 12 different answers could cover them. If about one third of the questions were instances of repeating questions, that also means that about two thirds were unique.

### Difficulty of the questions

Perhaps the most interesting, and at the same time, the most difficult question regarding the customer conversations, was “can they be automatized?” What does it mean for a question to be automatable? Many technologies rely on pre-created answers hand-crafted by humans. Thus, the project used this as a criterion: a question can be automatized if it is possible to write an answer to it beforehand, possibly linking to a web page in it.

Two things are relevant here that the project did not know. The project did not know what exact technology Statistics Finland might adopt for automatization and what the exact capabilities of that technology might be. In addition, Statistics Finland was, and currently still is, undergoing a major website renewal. This means that the project did not know what the website would be like in the future, that is, when the chat automatization would be taken into use. For example, requests for a specific employee’s contact information could be automatized by linking to a page containing a phonebook of employees – but only if such a page exists.

Both unknown matters cause predictions to be overly optimistic. The project was not able to consider whether a technology could interpret a customer’s question correctly, so it is assumed that this can be done. Similarly, as the project did not know, what the future website will be like, it is assumed to be the best possible from the point of view of automatization.

With this definition and these caveats, approximately 55% of all questions in the data were found automatable. There was some variation by question type: for instance, the contacts with no clear information need were not considered automatable at all, while questions about classifications almost always fell into the automatable class. See Figure 3 for the assessment of automatization degree per question type.

Figure 3: The degree of automatization per question type. Altogether about 55% of all questions were automatable.

### Language used by customers

The project also examined the conversations from the point of view of the language used by customers. The analysis considered syntax, vocabulary and the complexity of the conversations. In the analysis of syntax and vocabulary, only the lines of the customer were considered, not the lines of the staff member.

The syntactic complexity of the customer questions was measured in running words, sentences and clauses contained in each question. A **sentence** is the unit of written language that starts with a capital letter and ends with a period, question mark or exclamation mark. A **clause** is a syntactic unit usually centered around a main verb. A sentence can have one or more clauses.

One customer line had on average 7.8 words. The lines included a large number of lines with one or two words only: 93 lines with one word and 57 with two. Many customers start by writing a separate message with a greeting only, and often they also thank the staff member with one or two words. Additionally, the inappropriate messages are often one or two words long. Disregarding the one- and two-word lines, the rest of the customer lines had on average 10.7 words.

One customer line had on average 1.4 sentences, and most lines (356 out of 474) had exactly one sentence. The longest customer lines had seven sentences. Similarly, in the clause counts, short lines dominated: a customer line had on average 2.0 clauses. As a curiosity, the longest line had 95 words, seven sentences and 22 lines. This consisted mostly of a quote from the Statistics Finland website, as the customer wanted clarification on it. The lengthiness and the considerable clause count of the line thus better describe the language used by Statistics Finland than by its customers.

The analysis also considered the vocabulary of the customer lines, especially from a machine point of view. An automatic analysis tool, Voikko,[[1]](#footnote-1) was used to check the customer lines: if the tool recognizes a word, it can tell the user what base form the word has and how it has been inflected. For instance, the Finnish word “talossa” (“in the house”) is a form of the word “talo” (“house”) and it has the inessive ending “-ssa” (“in”). This kind of analysis is particularly important (and demanding) for automatic analysis of languages that have rich inflection, such as Finnish.

For the analysis of this project, the interesting point was the degree of difficulty of automatic analysis. Thus, the work considered those words that Voikko was **not** able to analyze, that is, words **unknown** to the tool. On average, a customer line had 0.5 unknown words. The most common case was with no unknown words at all, with 337 occurrences. This case was followed by lines with one unknown word, and at most, a line had nine unknown words (three cases).

In addition to counting the unknown words, the project also classified them, based on why they were unknown to the software. See Figure 4 for the types encountered, ordered by frequency. Clearly the most common case of unknown words, with 71 occurrences out of a total of 193 unknowns, was spelling errors. The second most common group was colloquial or dialectical words, followed by abbreviations. Notably absent were jargon and words of foreign derivation.

Figure 4: Unknown words by type. Words unrecognized by the automatic tool were most commonly spelling errors or colloquial or dialectical language.

Finally, the analysis considered the complexity of the conversations. For this, the project analysed all 186 conversations gathered, including both the customer lines and the staff member lines. On average, a conversation consisted of a total of 6.7 lines, including both customer and staff lines. The most common count of lines per conversation was one. However, many of these were cases where the customer’s question got directed to the customer service email, because the maximum number of simultaneous conversations allowed by the chat software were already underway. These conversations likely had more to them, but this was not included in the data collected. On the other hand, the one-line conversations also included inappropriate messages to which the staff members did not react.

The project also examined the conversations from the point of view of numbers of separate and clarifying questions asked. A customer asked on average 1.1 separate questions per conversation, and by far the most common number of questions in a conversation was one. In the most complex conversations (three cases), the customer asked three separate questions. As for clarifying questions, the results were similar for customers and the staff. For both, most questions did not need any clarifying questions. For both customers and staff, one conversation had on average 0.2 clarifying questions.

## Types of chatbots in use

To get an idea of what kind of solutions are used and what their capabilities are, the project examined some chatbots in use in Finnish organizations. These chatbots could be loosely divided into three groups.

**A button bot** is a solution where the customer can use buttons to select a topic to ask about. The conversation then goes on with button selection, narrowing down the need for information until a specific question is reached and can be answered. A button bot uses a knowledge base to give the customer a readymade answer.

**A keyword bot**, in turn, has a free text interface. The customer types a question, in either a single word or a full sentence, and the chatbot interprets it by looking for keywords or combinations thereof. Like a button bot, the keyword bot also uses a knowledge base with readymade answers.

**A guided conversation bot** initiates the conversation by telling the customer what kind of questions it can answer and how they should be formulated. The customer types a question, and the bot guides the conversation by asking more questions. The topics handled by guided conversation bots are typically quite limited.

It is also possible for a chatbot to be a hybrid of more than one type. While it is not possible to tell for sure what the internals of a specific chatbot are like from only seeing its user interface, most chatbots encountered seemed to either follow human-written rules or look for keywords in the user’s questions. This would mean that they do not contain artificial intelligence. This impression is consistent with the discussion in events such as Chatbot Day (arranged in Helsinki, late October 2018).

The analysis by Kuoppa (2017) seemed to have similar results. Examined in the summer of 2017, automatization solutions found in Finnish public sector organizations mainly consisted of different knowledge bases and lists of frequently asked questions, apart from Kela,[[2]](#footnote-2) who had a pilot chatbot known to utilize AI. Some solutions in use in foreign organizations seemed to include artificial intelligence or natural language processing. In the autumn of 2018, at the time of the project, some non-AI solutions had emerged in Finnish organizations, but overall the project concluded that true AI solutions were still very challenging and mostly only available with the resources of large commercial organizations.

## Potential effects of automatization

This Section considers the potential effects of automatization, including benefits, risks and other effects on the work of the customer service.

### Benefits

Naturally, automatization has some benefits that Statistics Finland hopes to achieve. First, it is hoped that automatization would reduce the workload of the skilled staff members, freeing time from answering routine enquiries and enabling better concentration on complicated information needs.

In addition, it is hoped that the chat service could be made more available. Currently the chat is only open on a small number of pages on the Statistics Finland website, and a staff member is available for about five hours each day to answer questions. If automatization were used, it is hoped that the service could be opened more widely on the website. Naturally, a chatbot could also serve customers around the clock.

### Risks

The project considered that the biggest risks in automatization have to do with attempting to automatize too much or trying to use too demanding a technology.

One risk is that the chatbot might be used to give very precise answers, which could be problematic if the bot has misunderstood the question. If a customer asked for a specific statistical figure and the chatbot responded simply by giving the customer the desired number, it would be very difficult for the customer to know if the chatbot misunderstood the question. In the worst case, this could cause financial harm to the customer, and Statistics Finland could be held accountable for the erroneous information. This could be avoided if the bot was to give written answers with more context and links to the relevant part of the website. This way, the customer would be able to check the appropriateness of the answer, and the final responsibility of accepting it would rest with them.

On the flipside, a second risk is that customers may need more specific answers than the chatbot is able to give. This would result in customers first asking the chatbot and then, when the answer turned out to be insufficient, turning to the human customer service. This would increase the amount of work for the customer service staff. At the time of the project, it was not possible to estimate how many of the questions of those asked from the chatbot would in fact be returned to the human staff afterwards.

Third, the project considered it risky to use very advanced solutions, especially technology that learns more by itself. The issues of independently learning methods are well demonstrated by a Microsoft chatbot named Tay (see for instance Guardian, 2016). Tay discussed with human users on Twitter, and in less than a day, they had taught the bot to generate inappropriate, offensive content – at a rate only possible via automatization. This kind of issues could be avoided by careful technology choices, and if machine learning was to be used, by making sure to allow only Statistics Finland staff to teach the chatbot.

### Effects on customer service work

If Statistics Finland used chat automatization, one important thing to consider would be the changes that it would cause to the customer service work.

It is not to be expected that a chatbot would be able to answer all questions. The customer service should be prepared to handle the questions that the chatbot cannot answer; that is, the difficult questions would still require human attention. If at the same time Statistics Finland hopes to open the chat more widely than previously, this would mean more questions altogether – not just more easy questions that a chatbot could handle, but also more difficult questions.

Automatization would also cause additional work, work that has not been done previously at all. Regardless of the technology chosen, the chatbot needs to be maintained and developed. The knowledge base used by the chatbot also needs to be updated continuously, and this demands human resources. In addition, if the chatbot needs to provide service not only in Finnish but also in Swedish and English, this causes translation work.

## Possible choices

Based on the analysis presented above, the project presented four choices for consideration. Briefly, these choices were as follows:

* improving the findability of the Frequently Asked Questions,
* waiting for other parts of the website renewal to progress before working on automatization,
* starting to develop a knowledge base,
* or starting to work on automatization in the form of a chatbot.

At the moment, Statistics Finland has a simple page for Frequently Asked Questions. If the goal is to free some time for the customer service staff to concentrate on difficult questions instead of routine enquiries, the path of least work would be to improve this page and its findability. This could help customers to find answers to easy questions themselves, but the staff would still only be available during office hours and the chat likely could not be opened more widely than it currently is.

As Statistics Finland is currently undergoing a large website renewal, one possibility would be to wait for this renewal to progress further before attempting automatization. The project has examined the questions that arrive to the current chat service located on the current website, but it is possible that questions will be different in the renewed service. Statistics Finland has also been looking into a ticketing system for the various ticketing needs within the organization. Some ticketing systems include a knowledge base, meaning that it could be beneficial to develop chat automatization together with ticketing. This would avoid two, possibly incompatible knowledge bases. Naturally, in this option the service rate of the chat would not rise at all before either the website renewal or the ticketing system, or both, have progressed.

If one were to develop customer service but not directly leap into automatization via a chatbot, one way to do this would be begin with a knowledge base. This would limit the immediate costs to acquiring a technical solution for a knowledge base and the person hours spent on gathering the questions and their answers. The costs could further be adjusted by selecting the topic or topics covered by the knowledge base. The chat automatization could be developed later based on the existing knowledge base. However, due to the ongoing website renewal and ticketing work, it is possible that some work could later turn out redundant or would have to be redone.

The furthest-going option regarding automatization would be to start developing a chatbot directly. This would be the fastest route towards the desired benefits: the higher service rate in the chat and more time for the customer service to focus on difficult questions. Directly moving towards automatization would mean that the project’s analysis would be directly applicable. In contrast, if Statistics Finland was to wait for the new website to be published first, there might be a need to analyze the kind of questions that are often asked on the new website. Naturally, if the knowledge base for a chatbot was built using the current website, it would have to be updated after the new website is published. The work would not be wasted, as the amount of information offered by Statistics Finland is so large that no web service will be able to solve all information needs and findability issues. However, it is possible that some time would be spent on questions that will not be asked on the new website. Also, accidentally ending up with two different knowledge bases with possibly differing contents would be problematic for both the customer service and the customers.

## Discussion

It should be noted that the generalizability of the results is limited. Due to the limited time and resources of the project, only a rather small amount (186) of conversations could be collected and used for analysis. Due to the timing of the project, only conversations from the late spring and early summer could be collected. In addition, no information was available on what the future website for Statistics Finland will be like from the point of view of automatization, nor what the exact capabilities of the chosen technology may be.

Especially when it comes to the language analysis, it should be remembered that the analysis was performed on conversations between two humans, whereas customers might communicate differently when discussing with a chatbot. This was indeed the case in the work of Hill et al. (2015): people discussing with a computer used for instance shorter sentences than those discussing with other humans, but on the other hand, exchanged more messages per conversation.

Based on the analysis performed in the project, about 55% of the questions in the data collected were estimated to be automatable in principle. On the other hand, only about one third of the questions were an instance of a repeating question, meaning that about two thirds of all questions were unique. This could mean that while some questions may be possible to answer automatically, it may not always be sensible to do so if they are asked rarely.

Over one half of the questions had to do with a single statistical figure. Disregarding contacts with no need for information, the second most common type of questions was those having to do with classifications. When examining the questions about a statistical figure, the most common topic of those was population. If automatization is done gradually over time, questions about statistical figures on population seem a promising place to start.

Based on the language analysis, it seems that the language used by the customers is not very complex when it comes to syntax. Even disregarding the shortest lines, a customer line had on average less than 11 words. There were on average 1.4 sentences and 2 clauses per line. The relatively uncomplicated language is a good thing for automatization, even if actual natural language processing technology is not used. In a keyword bot, a simple question might be tagged with fewer keywords than a complex one, or it could be easier to direct the customer to formulate the question in a way that is understood by a directed conversation bot. In a button bot, a simple question may be easier to transform into buttons than a complex one, or it may be found behind fewer buttons.

In vocabulary, the questions seemed reasonably uncomplicated as well. Almost three out of four customer lines had no unrecognized words at all, and on average there were 0.5 unrecognized words per line. More than a third of the unrecognized words were spelling errors, likely in part due to fast communication in the chat. How unrecognized words affect automatization depends on the technology selected. If for instance a keyword-based approach is used, the effect depends on whether the customer misspells the critical keyword. Meanwhile natural language processing technology, which attempts to understand the whole message, may be more sensitive to misspellings.

It is also interesting to note that while some of the unrecognized words in customer lines were colloquial language or words from a regional dialect, no unrecognized words were classified as jargon, nor words of foreign derivation. The customers seem to speak largely general language, and this should be remembered when creating any automatization solution. The chatbot interface, regardless of type, should be designed to use general language to match the customer’s language, as should naturally also the questions and answers in the knowledge base.

Generally, customers only asked one question per conversation, and clarifying questions were rarely needed on either side. The conversations were, however, on average 6.7 lines long. This seems to be explained by the customer service often splitting their answers into several messages: the answers can contain multiple parts and several links to the website. The relative simplicity of conversations is a good thing for automatization, especially when it comes to clarifying questions. The possibly lengthy answers by the customer service may simply indicate that sometimes the answers in the knowledge base may be rather long.

## Conclusions

The project ended up cautiously recommending automatization. Both the work done in the project and the previous analysis by Kuoppa (2017) seem to indicate that there is some automatization potential in the questions received by the customer service. In addition, the other options presented above would not allow Statistics Finland to open the chat more widely on its website or to enhance the service hours.

If automatization is to be applied, care should be taken that the technology selections are compatible with both the choices made in the website renewal and with the possible ticketing solutions. For instance, two separate knowledge bases with possibly conflicting information should be avoided. Also, the preferences of customers are worth investigating carefully when making choices. As Statistics Finland is currently working on renewing several background systems necessary for the web service, automatization will be given due consideration when the supporting infrastructure has been built.

## References

The Guardian: *Tay, Microsoft’s AI chatbot, gets a crash course in racism from Twitter*, 2016. Available at <https://www.theguardian.com/technology/2016/mar/24/tay-microsofts-ai-chatbot-gets-a-crash-course-in-racism-from-twitter?CMP=twt_a-technology_b-gdntech> (Accessed: 9 May, 2019).

Hill, Jennifer, Ford, Randolf ja Farreras, Ingrid: *Real conversations with artificial intelligence: A comparison between human-human online conversations and human-chatbot conversations*, Computers in Human Behavior, 2015. DOI: 10.1016/j.chb.2015.02.026.

Kuoppa, Arttu: *Selvitys asiakaspalvelun automatisointiratkaisuista*. Statistics Finland (internal document), 2017.

1. <https://voikko.puimula.org/> [↑](#footnote-ref-1)
2. <https://www.kela.fi/web/en> [↑](#footnote-ref-2)