

Politician – An Imitation Game

David Kuboň¹, Eleni Metheniti², and Barbora Hladká^{1(⊠)}

 ¹ Charles University, Prague, Czech Republic {dkubon,hladka}@mff.cuni.cz
 ² Saarland University, Saarbrücken, Germany elenim@coli.uni-saarland.de

Abstract. We present the question-answering system *Politician*, which is a chatbot designed to imitate a fictional politician. The chatbot accepts questions on political issues and answers them accordingly. The questions are analyzed using natural language processing techniques, mainly using a custom scenario built in the Treex system, and no complex knowledge base is involved. Once morphological and syntactic annotations added by language tools are available for the question, an appropriate answer template is selected from the manually created set of answer templates based on nouns, verbs, and named entities occurring in the question. Then the answer template is transformed into a grammatically correct reply. So far, two working versions of *Politician*, Czech and English, have been created. We conducted a Turing-like test to test Politician's intelligence. We also briefly investigated the differences between the two languages and potential generalization of the approach to other topics. Apparently, morphological and syntactic information provides enough data for a very basic understanding of questions on a specific topic.

Keywords: Chatbot \cdot Question answering Natural language processing

1 Introduction

Chatbots are programs designed to conduct a conversation with a person in both written and oral forms. A chatbot, as a conversational agent, interacts with a human usually by taking turns in answering questions posed by the human. Its main goals are to be able to carry a conversation successfully, to be as indistinguishable from a human as possible and to fulfill its purpose as effectively as possible (e.g. entertainment, customer service etc.).

Joseph Weizenbaum created the very first chatbot ELIZA that used pattern matching and substitution methodology in order to answer questions posed by humans, and given a rule-based scenario it could carry out a conversation with a human taking the role of a psychoanalyst [21].

Chatbots have been created in the rule-based style of ELIZA, by using heuristic conversation rules like ALICE [1], by using Artificial Intelligence methods, such as Watson by IBM [10], or in the recent years deep learning methods like the Neural conversational model chatbot by Google [19].

[©] Springer International Publishing AG, part of Springer Nature 2018

S. Diplaris et al. (Eds.): INSCI 2017 Workshops, LNCS 10750, pp. 201–212, 2018. https://doi.org/10.1007/978-3-319-77547-0_15

The *Politician* chatbot is a conversational agent where a user is interviewing a virtual fictional politician by posing questions and the system is returning an answer similar to what a real-life politician would answer. We named this fictional politician Humphrey Novotný/Smith.

Our approach uses a rule-based script that makes use of the Treex system (formerly TectoMT), a highly modular NLP software system implemented in the Perl programming language [14]. Our goal is to show that a chatbot can be effective without the use of a complex knowledge base or artificial intelligence methods, but applying the strategy similar to ELIZA. The first version of the chatbot was built for Czech to test it by Czech native speakers who were readily available. The second version was built for English without making any serious changes but translations of answer templates. The Treex system supports English morphological analysis.

2 Related Work

ELIZA, programmed with the 200 line code, answers humans' questions posing as a psychotherapist. Users were aware that ELIZA is a computer program, nevertheless they were very pleased to talk to ELIZA. After the success of ELIZA, many chatbots with a similar build have emerged, since building a rule-based system is rather simple. Chatbots through the years have mimicked a variety of different characters, from a patient with schizoaffective disorder (PARRY) to a virtual valet assisting with web searches (AskJeeves).

Restricting our research to conversational agents who do not use deep learning algorithms, the advances of natural language processing allow us to create more dynamic scenarios and more complex rules. Markup languages offer chatbot more in-depth knowledge. For example, the markup language AIML was used for pattern matching in A.L.I.C.E. [20], and Al-Zubaide [4] developed a chatbot which uses a mapping technique to transform ontologies and knowledge into relational database. Progress in the field of information retrieval enables building chatbots solely by using documents with question-answer sets (e.g., dialogue corpora [15]), or even from unstructured documents [7]. Parsing, named entity recognition, word vectors, and sentiment analysis present some of natural language processing tools that significantly help to build chatbots with more natural responses and adaptability (e.g., Wit.ai [22], Chatflow [12]).

In recent years, rule-based approaches have fallen out of style in research communities, especially after the widespread appeal of neural networks. However, some chatbot engines that let users create a chatbot and train it by writing possible answers to prompts are quite popular (e.g., Rebot Me [9]) or setup prompts for users' inputs (e.g., ManyChat [13]). Other engines, such as Twyla [18], integrated rule-based systems along AI methods.

Shawar and Atwell [16] defined four possible uses for a chatbot in the modern era: (i) as a tool of entertainment, (ii) as a tool to learn and practice a language, (iii) as an Information Retrieval tool, and (iv) as an assistant in e-commerce, business, and other domains. In these contexts, while users found the chatbots' shortcoming funny when the only intention was fun, the communication problems (inability to understand, repetition, nonsensical or irrelevant answers) frustrated users when they wanted to learn information or accomplish a task. One important finding was that linguistic knowledge was crucial in teaching applications, as users would benefit from seeing grammatically correct sentences.

Regarding political issues, there exist some chatbots that can imitate politicians using simple techniques to mimic a real politician's speech with simple pattern matching. For example, the *Ask the Candidates* chatbot by Zabaware [5] which searches through the database of 2012 US presidential candidates' quotes to find the answer best fitting the question according to the politician selected. In the most recent 2016 US elections, more conversational bots of the presidential candidates have emerged, operating on the same principle [6].

Twitter hosts many bots that behave like politicians but they lack the conversational aspect, as they only reproduce ("retweet") the content from real politicians and cannot produce any original output [11].

3 Methods

3.1 Conversational Topic

The first step in designing the *Politician* chatbot was to determine how the system would behave in the conversation, mainly what topics will be discussed and how to engage humans to converse in a natural way, i.e., how to discourage users to follow a specific topic using several questions without changing the topic too quickly. Finally, the idea of having a politician as a persona of the system seemed ideal for our case. Politicians are characters who engage socially with people, [8] can converse about a variety of topics and sub-topics and can receive heterogeneous questions, so that the system can be tested on a variety of subjects; and anyone could find questions to ask a politician.

3.2 High-Level Concept

To build *Politician*, it was necessary to find a proper NLP framework to build upon. The **Treex** system offers a variety of language-specific tools, of which we used a morphological analyzer, a tagger, a parser, and a named-entity recognizer [14]. **Treex** is running on Perl and can be downloaded directly from the command line or from the Github repository.³

The *Politican* opens a conversation with a standard greeting (*Hello.* or *Greetings.*) and a prompt to pose a question (e.g., *I believe you have a question for me.* or *What would you like to discuss about?*). Then the user has to pose a question, and *Politician* replies to that question. The conversation is closed when the user does not pose any other question. The user input is analyzed in three basic steps visualized in Fig. 1:

³ https://github.com/ufal/treex.

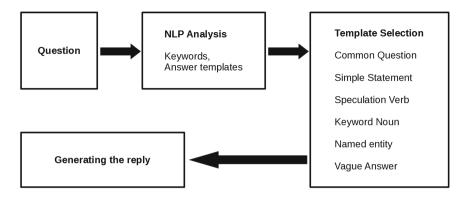


Fig. 1. Workflow of Politician

- 1. Analyzing user question the tools perform morphological analysis, partof-speech tagging and named-entity recognition analysis on the input string.
- 2. Selecting answer template we manually designed a list of answer templates. Thus our task is to select an answer template that fits the question best. More details are provided in Sect. 3.3 below.
- 3. Generating *Politician's* reply we transform the answer from the step 2 into a grammatically correct reply. More details are provided in Sect. 3.3 below.

The *Politician* chatbot keeps both the questions posed in the conversation and the keywords used for the answer template selection. When users fail to pose a question, they are prompted by the system to do so by receiving the question *Is there something else you would be interested in?* or a similar one. If a keyword is not detected in the question, the system returns a generic answer from a list of strings. Every time a keyword is detected in the question, *Politician* checks whether it has been used before or not. In the first case, it selects an answer template in accordance with the question sentiment. If the same keyword is used more than once, then the answer comes from a set of strings tagged with the same sentiment. However, we do not perform any sentiment analysis; we merely identify the answer templates with a given sentiment at design time, based on the human intuition that repetition can agitate a conversational partner.

This functionality also serves to ensure that two answers given after detecting the same keyword do not contradict each other. For example, in the English version, if the keyword war is identified, the system replies I have no right to comment on wars. This falls past my jurisdiction. If the keyword war appears again, there will be no chance of replying e.g., I do not think that war is one of the most burning issues.

3.3 Reply Generation

To further elaborate generating replies, we demonstrate it for the Czech question *Co si myslíte o dnešní demonstraci za zvýšení mezd v Olomouci?* (What do you think about today's demonstration for salary increase in Olomouc?)

- 1. We focus on nouns and verbs occurring in the question. In the sample question, the tagger identified the nouns demonstration, increase, salary, Olomouc and the verb think. The nouns are checked against the list of keywords that were identified as words typical for present-day political conversations, like důchod (pension), ekologický (ecologic), emancipace (emancipation), emigrant (emigrant), etc. Currently, the list consists of a few hundreds of words. The keyword list was built based on human intuition, but as a future step we plan to employ more advanced data mining techniques for removing words from the list that have a strong correlation with politics but mostly come with another topic-related keywords, which are more relevant to the question. For example, see the word strana (party) in the question Je postup vaší strany v kauze podplácení soudců ústavně v pořádku? (Is the practice of your party in the case of bribing judges in accordance with the constitution?), where bribing judges is clearly more important than party.
- 2. Some basic criteria are applied to filter out common questions such as What is your name?, Which political party are you a member of? or simple statements such as Hello, I'm pleased to meet you. We built the lists of words typical for such questions, e.g., the list of how, be, you for detecting the question How are you? Each list is associated with a reply that is posed to the user if the words from the list are detected in the question and no keywords occur in the question. For a simple statement, we use neither keywords nor lists, but we check if the sentence is a question by examining its word order and the presence of a question mark. If so, the user receives a generic answer from a list of strings, e.g., for the English version, This is a very pressing issue. or We will focus on this topic in the next election campaign. The sample Czech question is detected as neither common question nor simple statement.
- 3. The question is checked whether it consists of verbs of speculation such as to think, to suppose, to believe. They provide strong evidence for a particular topic that can be rather simply identified by checking the nouns in the question against the keyword list. In our example, the verb of speculation myslit (to think) was detected, meaning that the answer template for a question containing a verb of speculation is selected:
 - (1) [spekulativní sloveso], že vše je v souladu se [verb of speculation] that everything is in accordance with stranickou linií the party lines
- 4. The nouns detected in the question are checked against the keyword list. Unlike the previous step, the chance that a word which truly states the topic is found decreases.
 - First, it is mainly due to the fact that several keywords of different importance and relevance may occur in questions, e.g., *freedom of speech* and *democracy* in the sentence

(2) Je svoboda projevu skutečně tak významná pro Is the freedom of speech really so important for demokracii ? democracy ?

If one of the detected keywords is the subject of the sentence or the subject attribute, then it is chosen to select an appropriate answer template. If both are not dependent on the subject, the choice of the keyword is made randomly. There were detected two keywords in the sample question, namely *demonstration* and *salary*. Since none of them is dependent on the subject (in fact, there is a null subject in the question), we randomly select one of them, for example *salary*. Then we select the template

- Second, the actual topic may not be presented by a keyword, although there is a keyword in the question (see *prime minister* below). It could mean that the keyword was selected wrongly, but in most cases, it is a drawback of applying a simple approach that does not address the importance of the keywords, e.g.,
 - (4) Jak by se, dle vás, pan premiér How would , according to you , the prime minister respond vyjádřil k tomuto do očí bijícimu tunelu ? to this striking tunneling ?

Unfortunately there is no easy way to recognize the topic in this case, because our simple approach aims for the most common case. Therefore it is pure chance whether the selected template will fit or not.

- 5. The named entities recognized in the question are analyzed. While place names typically do not present a main topic, person names, especially accompanied by opinion words such as *to think, to consider, opinion*, often represent the substance of the question, see the following questions:
 - (5) Co si myslíte o dnešní demonstraci za zvýšení What do you think about today demonstration for salary mezd v Olomouci ? increase in Olomouc ?
 - (6) Myslíte si , že rozhodnutí pana Novotného je dobré z Do you think that the decision of Mr. Novotny is good from dlouhodobého hlediska ?
 a long-term angle ?

If people are called by name in the question, agreement or disagreement is mostly sufficient for keeping the conversation natural, even in cases when a person is not the main topic. For illustration, consider the question: (7) Pan Svoboda se nedávno vyjadřoval k několika Mr. Svoboda has recently expressed his views on several tématům, co si o těch problémech myslíte vy? issues, what do you think about those problems?

In the sample question, the named entity *Olomouc* was detected. Since it is a place name, it is not likely to be a topic of the conversation.

- 6. If all the preceding steps are considered failed, then a vague reply is provided. It is selected randomly from the set of replies prepared for such situation. For example,
 - (8) Tuto problematiku jsem důkladně rozebral ve svém I analyzed these issues in detail in my včerejším tiskovém prohlášení a press release yesterday and již se k ní nemíním vracet.
 I do not intend to get back to them any longer.
 - (9) Nevěřím , že si na tuto otázku nedovedete odpovědět sám . I do not believe that you cannot answer this question yourself .
- 7. When the answer template is selected, a grammatically correct reply is generated so that *Politician* inserts the keywords detected in the question into a reply in a proper grammatical form. For example, we can generate the two replies (11) and (12) from the template (10).
 - (10) [keyword] [být] [plod] našeho mnohaletého úsilí [keyword] [to be] [fruit] of our longterm efforts
 - (11) Reformy jsou plody našeho mnohaletého úsilí. Reforms are the fruit of our longterm efforts.
 - (12) Vzdělání je plodem našeho mnohaletého úsilí. Education is the fruit of our longterm efforts.

For the verb of speculation answer template (see 3. above), the singular form of the first person for the verb myslit (to think) is generated. Then the final reply is

(13) Myslím, že vše je v souladu se stranickou linií. I think that everything is in accordance with the party lines.

The answer templates are designed to make the conversation as captivating as possible. We found out that a user who is enjoying the conversation tends to ask more relevant questions rather than trying to catch out the computer. We included replies from real political interviews into the answer templates. They make the conversation more authentic. Some of them are well-known phrases of Czech politicians (14) Jsem přesvědčen, že jsem toho nevykonal málo a že to I am convinced that I have done enough and that it bude ve střízlivější chvíli, než je tato, doceněno. will be appreciated in a moment more sober than this one.

We included non-real replies as well, some of them are rather ironic, e.g.,

(15) Novotného si vážíme nejvíce když mlčí.We appreciate Novotny the most when he is silent.

We provide a sample conversation selected from the data collected at public promotion events. We denote *Politician* (P) and the user (U). We use italics to emphasize the words, which were selected as the keywords for the answer template selection.

- (P): Hello! What are you interested in?
- (U): Good afternoon. How much is the Czech state debt per capita?
- (P): This is a very good question. We have already voted to create a special committee that will deal with the debt. Let us wait for its conclusion.
- (U): What impact will *inflation* have on Czech market?
- (P): This topic will be emphasized in the next election campaign.
- (U): What is your opinion on amnesty granted by President Václav Klaus?
- (P): I do not like this person at all.
- (U): Do not you think that MPs' salaries are too high?
- (P): Unfortunately, judging this topic is not within my competence. I am sorry, but I have to go now, I have an important meeting.

3.4 Non-political Scenarios

It is necessary to realize that this simple approach can be applied in a limited number of scenarios. For comparison, we tried to create a football player behaving in the same style.

While *Politician* requires a single list of keywords for political topics, where each and every can be, with a good rate of success, inserted into any of the answer templates, football player topics are more varied and a single keyword list is not sufficient at all. For instance, it is necessary to differentiate whether the question relates to a game situation, a player, a strategy, etc. That can be achieved by having multiple keyword lists, each for one category. Obviously, even then the coverage is not perfect. However, this solution is clearly less convincing for the interviewer than that of a single keyword list, which is a problem that increases with more and more complex topics.

On the other hand, having multiple, clearly differentiable keyword lists is a viable approach that proposes a practical use to such a chatbot. By matching questions to a single sub-topic, the AI of some customer service support could provide conversation openers and cluster questions to be handed over to respective human agents. The big advantage we see is the simplicity of designing such a system.

4 Evaluation

4.1 Testing the Politician

To investigate how convincing *Politician* could be as a conversational partner, we have been presenting it mainly to students e.g., attending a university open house, and we were conducting a test loosely inspired by the Turing Test [17] during these events. There are many variations to the Turing test. We worked with the following settings:

- a user was faced with a single terminal and chatted with either *Politician* or a human,
- the human was instructed to pose only vague and non-specific replies and the users were told to ask only political questions. If they did not follow these instructions, we did not include their conversations into a final evaluation,
- each conversation contained exactly four user questions,
- each user did two interviews, one with the human and one with *Politician*. Then he judged in which of the two interviews he was interviewing the computer.

During the test, *Politician* used several tricks to confuse the user. For example, it randomly generated spelling errors by swapping two random adjacent characters. Also, it provided the correct answer to the *ultimate question of life, the universe, and everything* [3], which confused some users. However, these conversations were not evaluated since this is not a political question.

In total, we conducted over 30 conversations on political topics. The *Politician* chatbot was able to fool users more than 33% of the time. It turned out that the user being fooled did not fully depend on the quality of the conversation. One third of the users which were fooled were confused because of the spelling errors, half of them considered *Politician's* replies more natural and the rest of them did not have any relevant reason.

Further, we evaluated the *Politician*'s replies only in the conversations that we considered relevant to the topic. Eight annotators of different ages, some having NLP or language background, marked the *Politician's* replies as *fitting* the question when the answer made full sense, *not fitting* when the reply was not answering the question and *nonsense* when it simply made little or no sense in the given context.

Unlike the Turing-inspired test, which was interesting in terms of human behaviour and the proneness to be deceived, this experiment gave us more solid data on the actual quality of the replies.

As seen in Table 1, of 140 annotated replies, 30.6% on average were considered *fitting*, 34.4% were *not wrong* and 35% were *nonsense*. There is only one reply

	A	В	\mathbf{C}	D	Е	F	G	Η	Avg.	Perc.
Fitting	17	48	34	18	61	33	63	69	42.9	30.6
Not fitting	71	59	35	56	24	44	48	48	48.1	34.4
Nonsense	52	33	71	66	55	63	29	23	49.0	35.0

Table 1. Summary of the evaluation of *Politician*'s replies. Annotators A to H markedthe replies as *fitting*, not *fitting*, and nonsense.

rated as *fitting* by all the annotators and three replies with all *fitting* except one *not fitting*.

It is worth noting that the annotation is subjective. It happened that seven annotators marked a reply as *fitting* and one as *nonsense*, or vice versa, or that the reply scored roughly equal number of *fitting* and *nonsense*. An example of the former follows:

Co si myslíte o globálním oteplování? (What is your opinion on global warming?)

To je dobrá otázka. Hodnocení tohoto tématu je bohužel mimo moji kompetenci. To by pro tuto chvíli stačilo.

(That is a good question. Further analysis of this topic is unfortunately not withing my competence. That would be all for now.)

4.2 Language Comparison

Czech and English are two languages somewhat different in structure. For instance, while English maintains a strict word order as far as the position of the subject and the verb is concerned, Czech allows more of a word play. Also, English is a morphologically poor language with very few inflectional morphemes for verbs, nouns and adjectives and agreement is rarely an issue (e.g., in third person singular). However, Czech is a morphologically rich language with a rich nominal declension, meaning that the agreement has more strict requirements (gender, case etc.). Therefore it would seem that the difficulty of implementing these two languages into *Politician* would also differ — but it does not.

Implementing Czech required extra steps, but the existence of the proper tools from **Treex** allowed us to handle these cases and produce grammatically correct sentences. Implementing the Czech model prior to the English model allowed us to pay attention to morphological details in English, and create a more believable model that can efficiently deal with many word types.

Obviously, there are obstacles that relate to only one of the languages. For example, figuring out the keyword in the loose word order of a Czech sentence is considerably harder than in the English one. On the other hand, using the correct tense in a reply is not really needed in Czech and can be partly avoided even in English with good design of reply patterns. Furthermore, detecting a question in the past tense can be done in English by simply looking at the tense of the verb *to do*, while in Czech it is necessary to find the correct full meaning verb in the whole sentence and check its tense.

Overall, it is hard to say if any of the languages is better for the chatbot. English is a little easier to handle, but based on user experience, people do not expect the chatbot to be able to generate all the correct forms of the verbs and declinations needed in Czech, and thus are more impressed by it.

5 Conclusion

The *Politician* chatbot and its extensions and derivatives provide insight into a very simple technique for question analysis. It shows that for a known topic, it is not difficult to cluster questions into predefined answer templates. The chatbot is user friendly and fun to talk to, so gathering data is much easier because users simply enjoy talking to a virtual politician. Our insight into other conversational topics has shown that with such a simple system, disallowing the user from going deeper on topics or asking for specific facts is vital for the conceivability of the system.

Different languages, with no surprise, provide different challenges when designing the reply patterns. However, as far as we can tell for English and Czech, neither seems clearly easier. More research into this can be done with NLP tools for a chosen new language.

The *Politician* chatbot experiments proved valuable for verification of the used technology, most notably to test the limits of this simple method. The modularity of the system allows for a future expansion by changing the answer templates for a more complex approach.

Acknowledgement. This research was supported by the Charles University project No. SVV26033.

References

- Abu Shawar, B., Atwell, E.: Using corpora in machine-learning chatbot systems. Int. J. Corpus Linguist. 10(4), 489–516 (2005)
- Shawar, B.A., Atwell, E.: Different measurements metrics to evaluate a chatbot system. In: Proceedings of the Workshop on Bridging the Gap: Academic and Industrial Research in Dialog Technologies. Association for Computational Linguistics (2007)
- Adams, D.: The Hitchhiker's Guide to the Galaxy, 1st edn. Pan Books, London (1979)
- Al-Zubaide, H., Issa, A.A.: Ontbot: ontology based chatbot. In: 2011 Fourth International Symposium on Innovation in Information & Communication Technology (ISIICT), pp. 7–12. IEEE, November 2011
- 5. Ask the Candidates 2012 (n.d.). https://web.archive.org/web/20120218013142/, http://www.askthecandidates2012.com:80/. Accessed 11 Nov 2017

- Ask The Donald a Question (n.d.). https://web.archive.org/web/20161229070115, https://www.donalddrumpfbot.com/. Accessed 13 Nov 2017
- Bao, J., Chen, P., Duan, N., Li, Z., Yan, Z., Zhou, M., Zhou, J.: DocChat: an information retrieval approach for chatbot engines using unstructured documents. In: ACL (2016)
- Compton, K.: Little procedural people: playing politics with generators. In: Proceedings of the 12th International Conference on the Foundations of Digital Games (FDG 2017). ACM, New York, NY, USA, Article 75, 2 p. https://doi.org/10.1145/3102071.3110573 (2017)
- 9. Create Chatbot (n.d.). http://rebot.me/. Accessed 11 Nov 2017
- Ferrucci, D.A.: Introduction to "this is Watson". IBM J. Res. Develop. 56(3.4), 1 (2012)
- Forelle, M.C., Howard, P.N., Monroy-Hernandez, A., Savage, S.: Political bots and the manipulation of public opinion in Venezuela. In: Proceedings of the 12th International Conference on the Foundations of Digital Games (FDG 2017). ACM, New York, NY, USA, Article 75, 2 p. https://doi.org/10.1145/3102071.3110573 (2015)
- 12. KITT.AI (n.d.). http://kitt.ai/. Accessed 11 Nov 2017
- ManyChat The easiest way to create Facebook Messenger bot (n.d.). https://manychat.com/ Accessed 11 Nov 2017
- Popel, M., Žabokrtský, Z.: TectoMT: modular NLP framework. In: Proceedings of IceTAL, 7th International Conference on Natural Language Processing, pp. 293– 304, Reykjavík, Iceland (2010)
- Shawar, B.A., Atwell, E.: Using dialogue corpora to train a chatbot. In: Proceedings of the Corpus Linguistics 2003 Conference, pp. 681–690 (2003)
- Shawar, B.A., Atwell, E.: Chatbots: Are they really useful? In: LDV Forum, vol. 22, no. 1, pp. 29–49 (2007)
- 17. Turing, A.M.: Computing machinery and intelligence, pp. 433-460. Mind (1950)
- Twyla AI Customer Support Chat Bots That Learn From Human Agents (n.d.). https://www.twylahelps.com/. Accessed 11 Nov 2017
- Vinyals, O., Le, Q.: A neural conversational model. arXiv preprint arXiv:1506.05869 (2015)
- 20. Wallace, R.: The elements of AIML style, pp. 12–20. Alice AI Foundation (2003)
- 21. Weizenbaum, J.: ELIZA–a Computer program for the study of natural language communication between man and machine. Commun. ACM, pp. 36–45 (1966)
- 22. Wit.ai (n.d.). https://wit.ai/. Accessed 11 Nov 2017