



Technische Universität Berlin



DOCUMENT BUILD DATE: 21st February 2018

DOCUMENT STATUS: Beta

Development and Evaluation of a Service Bot in the e-Government Sector

Bachelor Thesis

am Fachgebiet Agententechnologien in betrieblichen Anwendungen und der
Telekommunikation (AOT)

Prof. Dr.-Ing. habil. Şahin Albayrak
Fakultät IV Elektrotechnik und Informatik
Technische Universität Berlin

vorgelegt von
Mohamed Megahed

Betreuer: **Dr. Andreas Lommatzsch**
Gutachter: Prof. Prof. Dr. Dr. h.c. Şahin Albayrak
Prof. Dr. Odej Kao

Matrikelnummer: 342655

Abstract

Though not a recent phenomenon, chatbots and voice assistants are increasingly gaining unprecedented attention as a successor for mobile and web apps. While still emerging with no defined standards or set protocols, with a hype on the rise, tensions between industry giants with products like Amazon's Alexa, Apple's Siri or the Google Assistant unveil new examples in favour of providing an enriched user experience on consumer and business level. The surrounding ecosystem also plays a major role in widening the platforms available while exploring new horizons with alternative approaches and business models. Today voice assistance are already present around indoor spaces, in the car or on the go but are still a new terrain to discover and great potential to unleash.

One such scenarios involves the public sector. In this work, we explore Amazon's Alexa and respective platforms to develop a voice assistant for the local city council extending the current chatbot's functionality available on <http://service.berlin.de>. We touch on the technical challenges and possibilities in implementing a system for e-Government inquiries and analyse its usability as well as effectiveness in replacing a traditional lookup service. We then examine the goals we define for our use case to what we achieve with the available APIs and SDKs. With respect to those, we also report on the limitations developers could face in the process.

Finally, we study the current state of voice assistants and service bots in the market and the future of this trend from a technical and a social point of view. **[SUPERVISOR NOTE] CONCLUSION: DAS ERGEBNIS DER ARBEIT SOLLTE MAN ZUM SCHLUSS NOCH IM ABSTRACT ERGÄNZEN.]**

👉 **NOTE:** Im Abstract sollte man aim/will/going to meiden. Besser: Einfach Präsens oder sogar Perfekt

👉 **NOTE:** Ich finde es besser, im Präsens zu schreiben: Also besser: We analyze oder We study (anstatt von going to)

Zusammenfassung

👉 **NOTE:** translate to German to English or vice-versa.

[possibility to make inline notes] **[CITATION]** when a citation is missing

✍️ **To-DO:** todos

[TO CITE]

**[SUPERVISOR NOTE] BITTE BETREUERNOTIZEN AUCH IM FLIESSTEXT
MIT \SN{TEXT}]**

Contents

List of Figures	v
List of Tables	vi
1 Introduction	1
1.1 Motivation	3
1.2 The Analyzed Scenario	5
1.3 Approach and Goals	5
1.4 Structure of the Thesis	6
Bibliography	7
Appendices	8
Appendix A: Abbreviations	9
Appendix B: Glossary	10

List of Figures

List of Tables

Chapter 1

Introduction

With over a third of the world’s population projected to own a smartphone in 2018 [13] and a substantial fraction thereof using smarthome gadgets and appliances on a daily basis, AI’s role has become more interesting than ever in many disciplines including but not limited to productivity and entertainment. Many technologies we take for granted today, such as dictation and word prediction, recommender systems or other digital analytics depend on Machine Learning and Natural Language Processing techniques that were only made possible thanks to the high computational power shipped in most devices gradually overtaking the consumer market. This transition also facilitated the introduction of a new form of interaction through conversation with the hardware, paving the way to an aspiration modern societies have been striving globally [9]. Whether in blockbuster 60s drama as seen in “Breakfast at Tiffany’s” (1961) or in Sci-Fi romance in the movie “Her” (2013), an obsession with voice technologies is featured throughout from an answering machine on tape to a fully personalized but mass-produced voice-based operating system to even become a protagonist.

Conversational bots were already prevalent since the 80s in the form of Question/Answer systems based on query programming languages like PROLOG and SQL. ELIZA, considered as the world’s first chatbot and though quite superficial as an NLP-based programme for psychoanalysis, already at its early stages demonstrated how humans can become emotionally attached to machines, transcending over the anomaly of making conversation *not* with a human [15]. Today, combining ML with retrieval-based approaches allows a more advanced interaction with the system and yields smarter and more personalized conversations between man and machine. Consequently, it is no longer a surprise that chatbots acquire social skills to make Xiaoice, the empathetic bot from China, possibly a new kind of friend made of silicon revoking the fiction element from “Her”.

So far, voice assistants represent an additional layer of abstraction from software

beyond the graphical user interface (GUI) and are hence the closest we have come towards human communication. They de-construct another barrier between the user and the hardware as voice communication generally does not require profound computer literacy and the conversational models rely on our inherent ways of expression. As they simulate the human aspect and imitate its behaviour for instance with small-talk abilities [12], voice assistants are regarded as a convenience for daily tasks and are on their way to becoming a de-facto replacement to sophisticated actions we perform on the screen. In fact voice searches now compose a large marget segment 「 [check quality of statistics](#) 」 and are predicted to make up about 30% [6] [14] of all searches by 2020 with currently highest rates coming from the youngest generations based on a survey [16] by Global Web Index. According to the Alpine.ai 2017 Voice Labs report there were 33 million voice-first devices in circulation in 2017 worldwide [1] with tremendous shifts in number of units sold between 2015 and 2017. This and more statistics hint at a revolutionary change towards our use as much as the introduction of the GUI and mouse were to the command line interface. Increasingly, we recognize voice as a *new* user interface, also known as Voice User Interface (VUI), and analyse good practices and design guidelines for it.

Since sound and voice are primal stimuli in the human brain [8], using them becomes more instinctive, making us in turn process our ideas starting with an “inner voice” that translates easiest into words when we speak it before “funneling” it to actions [3]. Arguably, on a market scale, this gives voice assistants today a competitive innovation advantage (CIA) for they are hence more accessible to further demographic groups with a growing wider acceptance.

Meanwhile, statistics [10] show that time spent on messaging apps already surpassed average uptime on social media , which indicates how the former is more desirable as a communication format on mobile platforms and how having a conversation is an easier way to interact with a device instead of downloading an app for every task. 「 [link ok?](#) 」 Further, the speech-to-text/text-to-speech domain has become more powerful with steadily increasing processing power, an effect of Moore’s law we only get to understand lately in addition to the gradual lessening of the dominance of Chomskyan theories of linguistics (e.g. transformational grammar) [4]. With an integration in most modern operating systems, speaking to a device has become no longer a absurd novelty. Inadvertently, Google Now, the voice assistant built into the OS currently with the highest usage shares among computer and smartphones [5], supports understanding of multiple languages in the same sentence for multilingual interaction. Moreover, the Echo Dot recorded the peak for best-sellers for 2017 on Amazon with unprecedented numbers showing a high customer retention and satisfaction rate [11]

As we constantly challenge our expectations towards technology, our not too far-fetched imagination makes us question the ability of AI to make a machine able to react to everything we say. Although we are still far from this step, at least for the consumer level, we dedicate a lot of effort to make it happen with examples like IBM's Watson. And as long as we still do not exactly understand human intelligence yet, it is hard to fathom AI as a holistic field. As such, it is therefore more realistic to consider the current works in the field as *intelligence amplification* [3] empowering human take better decisions beyond their normal brain processing power.

1.1 Motivation

With the aforementioned, we try to think how voice assistants could come handy and why we want to invest in such technology, juxtaposing it to available alternatives. If we consider human workforce (i.e. customer service agents) on one hand, it is commonplace that these are more expensive, less available due to restrictive working hours and not always aware of the full circumstances related to an issue they are supposed to fix or a question they are to answer.

Sometimes knowing more about a person could almost become a dangerous tool since it gives room to manipulate them. A client in a shop for instance can have their decision influenced by the seller and eventually get tricked into buying a product based on wrong advice. Although there is practically little the client can normally do to circumvent misinformation if that seller is replaced by an algorithm, having an automated system like a voice assistant step in gives at least a more neutral impression since it are not directly expected to act with malicious intentions.

Since the point of availing voice assistant is to act in a person's interest, we also want an information system not to confuse us or to limit our cognitive abilities. 「 talk about cognitive behaviour 」 Besides, we can at least ensure that a voice assistant will not become moody and intentionally want to make our lives harder for this reason as opposed to a human. And so though a voice assistant or a chat bot may not potentially answer every question we throw at it, we want to at least presume that it would give us no information and not partial truths or lies. This is why most credible companies elaborate explicitly on their terms, conditions and privacy statements making them accountable on the products they produce and the services they offer. A consumer therefore feels more empowered to assert any faults originating from an automated system than from a human and conditionally has an assurance that they can prevent any violations more systematically.

Eventually a person is more likely to develop a certain kind of trust in a machine

more than in a human once the technology is established and widespread. Cars, email, and other gadgets or services we take for granted today are living proof of how this trust can grow is inevitable, for the better or worse and depending on the degree of affinity, aversion or ignorance in a business sector. Trust in a system can grow once it is certified to have little to minimal exceptions. Besides enriching the value chain, it is a key in setting a technology to become an industry standard. Therefore, if a voice assistant is shown to deliver reproducible results disregarding a person's profile, this definitely contributes towards the credibility of the system. This is however not an easy case, since an advanced voice assistant is not expected to be deterministic in most situations, otherwise it becomes boring! We elaborate later on this thesis how utterances [] handle this problem.

On the other hand, information systems range from web pages (e.g. frequently asked questions section (FAQ), forums). These are in most cases even less effective than contacting a human as getting the proper information takes a lot of time, or the level of trustworthiness or participation particularly in a forum is too low, the problem stated is too broad or too specific compared to the answer we are seeking. A user also could come purposefully across false positives in a search and rely on irrelevant information without knowing. Furthermore, some case-related information might be required to have a proper understanding of a situation or a scenario and provide adequate answers. For Example, if a user would like to know if a certain accessory is compatible with their mobile device, they might need to give a model number, which they may or may not know. Consequently, it is of high interest to maintain a system that could determine all these factors autonomously or with the least possible human interference such that a system supersedes the abilities of the classical Q&A approach.

Internationalization is also another factor to take into account. Since languages differ not only in their vocabulary but also grammar, word and sentence structure, developing a voice-first device requires a flexible infrastructure and software stack both able to accommodate these deviations. In that respect, not only is region-specificity important, but also being able to cater for people in a region who do not speak its official language or are residing there temporarily. Especially in businesses where it's difficult to hire skilled foreign-language speaking personnel, a voice assistant can overcome this challenge as it would communicate more accurately and will not have language problems. A customer is then given the option to avoid an inconvenient experience with a call centre representatives where neither of both parties understands the other. For those who have minimal understanding of the language, there are possibly also options to provide help in the native language if the user is given the options he/she can answer a prompt with. At the very least, a VUI could still give feedback to the user of whether it understands the language or not, since algorithms for detecting a language are not as

complex as answering the question once the language is understood. [approaches to localization? (Translators via Wordnet, Stammsprache,...refer to IRS lecture notes)]

In the following, we take all these factors into consideration and narrow down a tailored scenario in the e-Government sector. For this we choose to present our solution using Alexa, since among Apple and Google it has the voice-first devices best equipped for its platform, a user base larger than its competition and provides the most mature API and SDKs.

1.2 The Analyzed Scenario

Berlin.de is an online one-stop-shop for approx. 3,7 million residents of the German capital [2] with [hundreds/thousands] of visitors daily for information lookup, appointment bookings and even access to local news. As part of a federal modernization procedure with the help of the ministry of interior, D115 was launched in 2009 [7] as a phone service to help residents find relevant information about a public service or municipality, something that can be tricky if a person has no overview of the local government structure and still not always easy even with the help of search engines nowadays. To promote information accessibility, D115 continuously aims at expanding its reach and services. It is therefore worth exploring, how to offer D115 services in a fashion that takes advantage of conversational abilities beyond its human personnel.


For now, although local authorities rely heavily on their websites to communicate information to the public, the challenge is mainly finding the right service. In a metropolis with a high influx of incomers, it is also very likely that certain services are frequently pursued, meaning that helping find the right public service or authority is a repetitive task. In this context, thinking of a voice assistant as a public service could have several advantages, like offloading some traffic from the phone service, getting over the language barrier in the case of non-german speakers, expatriates or simply helping native customers formulate the right wording for a query in a more intuitive way than using a search box.

1.3 Approach and Goals

To-do: Aufgabenstellung:

-AL: Anschließend soll das Ziel der Arbeit formuliert werden: Entwicklung und Evaluation eines Prototypen für den Anwendungsfall.

- 1- es sollen die Stärken und Schwächen eines solchen System zu analysieren.
- 2- Es sollte zunächst eine Dienstleitung aus dem Berliner Service-Katalog mit dem Chatbot beaufschlagt werden können.
- 3- Nuancen beachten (e.g. 10243 / FHain)
- 4- Smalltalk Fähigkeiten

 **To-do:** - ..it would speak as an advantage for bots if they can determine these things automatically..

- currently most tasks revolve around performing tasks like setting an alarm,
- answer suggestions functionality in chatbot equivalent - next step is to get around the user's frustration by making the bot at least more human.
- Alexa Skill will work in Germany in english and german -> add english after german

1.4 Structure of the Thesis

This thesis endeavours to shed light on the following: In Chapter ??, we discuss related work as currently available. We first show a few use cases for chat bots and voice assistants and their implementation. For that we consider [the City of Vienna chat bot service "WienBot" for German and Singapore/ LA/... for English] and compare them to the current application available on service.berlin.de. We then introduce the Alexa platform and the idea behind Skills in chapter ?? and discuss how it can overcome a few struggles with respect to its counterparts [like redundant boilerplate code, smalltalk, but also limitations]. We also present the frameworks we use in our own implementation, the prerequisites and the artefacts provided initially. Chapter ?? and ?? represent a detailed analysis of our implementation. In particular, [SOME MORE SENTENCES. In Chapter ??, our solution is presented. This solution covers ... (SOME MORE SENTENCES).] Chapter ?? evaluates our implementation with respect to our target in the defined [4-5] use cases. [SOME MORE SENTENCES] In Chapter ??, we conclude with how the implementation of this e-Government solution is attainable through Alexa and where future works can be directed.

Bibliography

- [1] fka VoiceLabs Alpine.ai, jan 2017. accessed on 16.02.2018.
- [2] Amt für Statistik Berlin-Brandenburg. Statistischer Bericht - A I 5 hj 1. https://www.statistik-berlin-brandenburg.de/publikationen/stat_berichte/2017/SB_A01-05-00_2017h01_BE.pdf, August 2017. Accessed on 06.02.2018.
- [3] Brian Roemmele, Dave Isbitski. Episode 018 - A Voice First Future with Brian Roemmele. *Alexa Dev Chat Podcast*, 11 2017. Accessed on 13/02/2018.
- [4] Wikipedia contributors. History of natural language processing — wikipedia, the free encyclopedia, 2017. [Accessed 14 February 2018].
- [5] Wikipedia contributors. Usage share of operating systems — wikipedia, the free encyclopedia, 2018. [Accessed 14 February 2018].
- [6] Gartner. Gartner reveals top predictions for it organizations and users in 2017 and beyond. Oct 2017. accessed on 15.02.2018.
- [7] Geschäfts- und Koordinierungsstelle 115 im Bundesministerium des Innern. https://www.115.de/SharedDocs/Publikationen/Service_Publikationen/Infomaterialien/infobroschuere_%20BMI08324_screen_barrierefrei.pdf?__blob=publicationFile&v=8, November 2013.
- [8] Tobias Grossmann, Regine Oberecker, Stefan Koch, and Angela D Friederici. The developmental origins of voice processing in the human brain. *Elsevier*, 65:852–8, 03 2010.
- [9] Judy Goldsmith and Nicholas Mattei. Science fiction as an introduction to ai research, 2011.
- [10] M. Ballve. Messaging apps are now bigger than social networks, 2016.

- [11] NBC. Amazon's alexa had a breakout holiday — people even used echoes to buy more echoes. Accessed on 15/02/2018.
- [12] Shankar Vedantam. Radio Replay: I, Robot. <https://www.npr.org/podcasts/510308/hidden-brain>, 2018. Accessed on 25.01.2018.
- [13] Statista. Smartphone user penetration as percentage of total global population from 2014 to 2020, November 2016. Accessed on 06.02.2018.
- [14] Laurie Sullivan. Gartner predicts 30 12 2017. accessed on 12.2.2018.
- [15] Joseph Weizenbaum. *Computer Power and Human Reason: From Judgment to Calculation*. W. H. Freeman & Co., New York, 1976.
- [16] Katie Young, June 2016. accessed on 16.02.2018.

Appendices

Appendix A: Abbreviations

AWS	Amazon Web Services
ASK	Alexa Skills Kit
AVS	Alexa Voice Service
ARN	Amazon Resource Name
MVP	Minimum Viable Product
AI	Artificial Intelligence
NLP	Natural Language Processing
ML	Machine Learning
GUI	Graphical User Interface
VUI	Voice User Interface
ACID	Atomicity, Identity, .. criteria
appx	approximately

Appendix B: Glossary

Intent	erklärung
Slot	erklärung
Utterance	erklärung
Alexa	erklärung
Alexa Skill	erklärung
Lambda Function	erklärung
Alexa Skills Kit	erklärung
Amazon Developer Console	erklärung
AWS Lambda	erklärung
Amazon Lex	erklärung
Amazon Polly	erklärung
ElasticSearch	erklärung
node.js	Framework built on top of JavaScript
Interaction Model	erklärung
Service	bot, AWS, Berlin.de
https://docs.aws.amazon.com/general/latest/gr/glos-chap.html	
Application ID	erklärung
Skill ID	erklärung
Bot	Unless otherwise mentioned, yeb2a Cha
Hitlist	erklärung
Voice-first	an always-on, intelligent piece of hardwa