A personal dialog assistant-based approach to assist non-sighted people in web search

Md Javedul Ferdous
Old Dominion University, USA
mferdous@odu.edu

Abstract

With the ubiquitous use of the internet and improvement in internet security, online business goes to another level. It is easier for sighted people to buy a product from an online site. However, it is difficult for non-sighted people to figure out each item from a web page through a screen reader. Nowadays, personal dialog assistant (PDA) is the most popular among us. We introduce a different architectural solution for the development of an online shopping dialog system that helps non-sighted people complete a variety of purchase tasks, for example, searching for products and answering natural questions in a normal language conversation. Our main goal is to assist non-slighted people to have the same online experience as what we have. In addition, this work will illustrate how a knowledge-based system for the construction of language modeling combines with speech recognition. In this report, I show how a query can be made by using a microphone as well as how to rank keywords based on their context. Finally, I concluded the work with a glimpse of a product search result.

1 Introduction

Nowadays, a large number of industries focus on personal dialog assistants (Apple Siri, Microsoft Cortana, Google Assistant, Amazon Echo, Baidu Duer, and Facebook) and it gets popular among a lot of people. Web search preferences of users are evolving. They plan to search engines to answer questions, to be more conversational, And to provide means on their behalf to complete tasks. At the same time, PDA needs to better use information about the environment, a large amount of which is available in the knowledge bases and responses designed for search engines, to expand the scope of tasks. Improving natural language understanding (NLU) techniques for

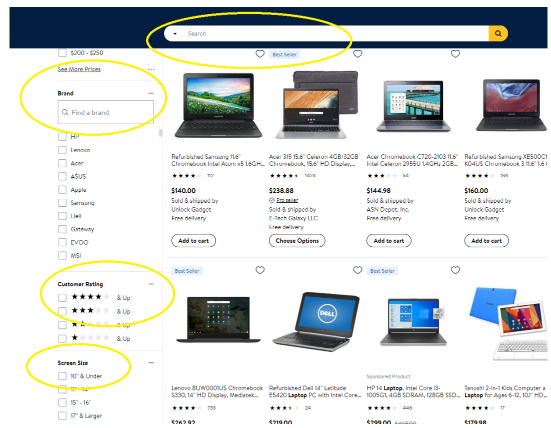

Figure 1: Search in Walmart website

parsing statements from users into predefined contextual positions gives a boost to the dialog system. However, Developing constructing NLU can be very difficult in the initial stage where it needs expertise in defining define the positions and possible outcomes for a specific domain.

For online shopping, one of the main obstacles is that an e-commerce shop typically contains hundreds of more types of items. In addition, based on semantic positions in different groups vary tremendously. As for normal people, can easily search for a product online. We can go to sites like Walmart, Amazon to buy anything that we want (figure 1). We can customize, set our own filter, or change any option just by hovering the cursor or typing any information using a keyboard. Sighted people have highly unlikely case where it is possible to face trouble in finding a product. But this scenario is quite opposite for the people who have disabilities, more specifically for those who have a problem with their vision. But it'll quite difficult for non-sighted people to search or customized any choice. There are a few reasons for it. Among them, one is it is difficult to go through each element of content by using voice-assisted technology such as screen reader. Only a screen reader reads the text on the screen, typically you do not have the opportunity to know the right wording, particularly if it is not so popular, such as medical terminology, etc. after hearing a word that they don't know to spell, they can make a reader read character by character, but it is rather time-consuming. Sometimes, some users feel bored using a screen reader. Some companies do their utmost to build speech synthesizers that can replicate the way people read a phrase; like the right intonation, but in recent years we have seen some great changes, they are yet far from achieving their targets.

Our work focuses on the issue we have listed usage of three common phases of the pipeline (i) a product knowledge base which is responsible for the storage and update of a user query, and (ii) a closely-connected search query Items, the generation of natural languages and semantic analysis; (iii) interpreting the search results and making a two-way conversation between user and device in natural language. In the mid-term report, the overall pipeline for this work is propose described in depth. In this final report, I included some recommendations, for example, some test case experiments of speech recognition and product search engine. Although, it is private now, when the project will be completed all Source code will be provided on GitHub.[1]

To show the efficiency of our proposed system, a dialog system is proposed to built as a personal assistant. All communication will be made via PDA with natural language and it will try to either help the consumers to complete their purchase-related tasks based on a session-level understanding of user utterances, or conversation with them. In future work, it'll be interesting to see how human-dialog assistant conversations involve making decisions on purchasing a product.

2 Related Work

While several works have been published on the building of a domain-specific semantic search, we concentrate on three types of work and list the most known works in this section. The first one is based on dialog system and speech recognition, the second one is based on semantic web search and finally, the last one is finding those works which is about how both of those technologies is merger give benefit to the user. Web search is increasing in popularity and voice recognition technology is allowing users to open the door instead of typing their queries. While this form of voice search is still in its infancy, it is increasingly popular. We categorize all relevant work in eight types based on their activity.

[1] `https://github.com/javedulferdous/SpeechAssistant`

2.1 Voice Search

The advancement of voice search enables users to enter queries in the spoken language and afterward, is a distinctive feature of search progress retrieve the corresponding entries based on system-generated voice query transcripts[Jiang, 2013]. However the system, sometimes, did not perform traditional comparison search experiments, raise some error. Recent advances in language recognition, with high bandwidth coverage and the acquisition of high-quality speech signals, make for better voice search[Chelba, 2013]. Google has published a case study in 2010, which notes that its aim is to provide voice search with a ubiquitous range of services and to achieve a level of success that increases use[Schalkwyk, 2010]. Improvements in the personalized model have been noted for automatic speech recognition (ASR) for web research[Zweig, 2011]. However, Issue with fast normalization and unigram yet to resolve. Some improvement comes under specific language-based such as mandarin[Shan, 2010]. The proposed system is still adapting, did not reach the saturated point yet.

2.2 Personal Assistant

the personal dialog assistants became more popular since humans are likely to communicate with machines using voice. Instead of typing, people are likely to use voice because the world of the user is becoming increasingly complex, cooperative work is expanding; the management of information is increasingly spreading[Enembreck, 2004]. Next-Generation of Virtual Personal Assistants that can interact with users and the computers by using the Multi-modal dialogue system with techniques but proposed idea found a flaw in implementation and their fund comes out short[Kepuska, 2018]. A large-scale semantic search of customer behavior data is can be used to retrieve matching products in response to a query [Nigam, 2019]. However, their model failed to filter irrelevant results online.

2.3 Semantic Web Search

To give a boost to online shopping, task-oriented dialogue systems are developed for Chinese online shopping [Yan, 2017]. Although it was a good example of a dialogue system for online shopping, the system only conducts experiments on community sites. Yet to experiment on search engines and social networks. It is interesting to see an agent extracts background information from a simulated linked data network targeted search to make custom recommendations in the field Value domain [Zlatareva, 2018] and Information retrieval and question answering focused work can be seen where a sufficiently high-quality language model can be helped without domain data [Misu, 2006].

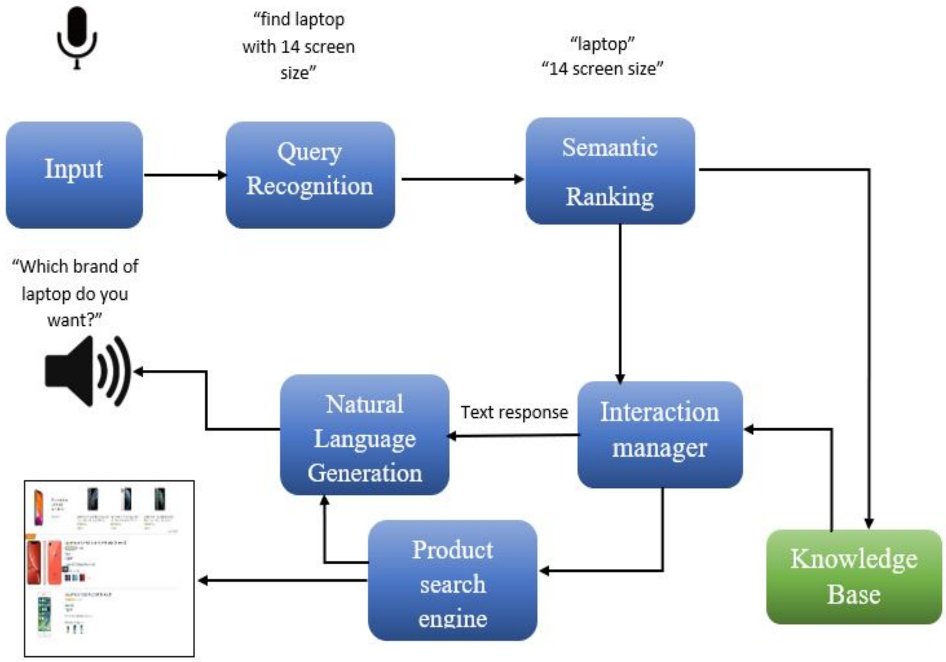

Figure 2: Architecture of proposed system.

2.4 Knowledge-based systems

Knowledge-based approach shows that it is helpful to have previous information on specific user so that it can interact with user and can answer question [Wang, 2019]. However, an issue identified on agent profile. Different answer is getting more complex question. A question answering system by using speech interface (speech recognition) and language understanding can leads to less attentive that paid to identification errors and inaccuracies in language treatment [Kumar, 2017].

2.5 Web Mining

A unique approach for text analysis to a specific language is web mining. In this work, author developed an algorithm for Slavic language by using the content-monitoring method to correlate the content-analysis of text information of keywords [Lytvyn, 2017]. Based on analysis, a system that can give exact and precise answers to the user's question. But to do that, it needs with help of Speech Processing, Audio and Visual Capability [Prajapati, 2018]. By using the company database, web scraped companies categorize with their web text and perform multivariate analyzes based on company characteristics and a creative product likelihood predictor for deep learning [Kinne, 2019]. Based on sentiment extraction, a framework can be build for extracting reviews and deep sentiment of a product [Trupthi, 2019]. However, the overall sentiment of the product is difficult take decisions.

3 PDA for Non-sighted people

In this section, we give basic steps for building dialog system with the help of using natural language which is backbone of this project in understand context of the search pattern. The architecture of this work shown in Figure 2. Firstly, basic definitions are given for this work. Secondly, the method for building the system is listed. Thirdly, how the system is interact with each other is proposed. Finally, the ultimate outcome after going through each level and the search query is enable into the web search.

3.1 Query recognition

The Query recognition model functions in real time the input voice into microphone with the voice detection model Cloud Servers to know what a user is saying and convert it to a text, then sends the Text for Cloud Service systems for data processing and return the result. For example, if a user looking for an laptop with 14 inch screen size, it is likely they will use this voice command "Find laptop with 14 screen size" (figure: 3). The main goal is this section is to identify each word that users make. In this case I'm using Speech recognition library from python. This library provides some speech API such Google Speech Recognition, Microsoft Bing Voice Recognition, IBM Speech to Text, CMU Sphinx etc. In this case, I am using google speech recognition API. For microphone purpose, I used PyAudio library. Once I use this library, it will get the voice as audio, using google

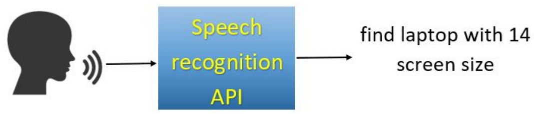

Figure 3: Query recognition

voice recognition API it'll return that audio into string.

3.2 Semantic Ranking

From high level perspective, a product search with words works by requiring that all products that contain words are found in the query, and that match is passed to a lexical content. The match set passes through classification stages. In the previous stage, the highest results are re-ranking prior to finally displaying the most relevant items. Searching a product as the queries tend to be shorter and the purchases are sparser than clicks. Each click and search may lead to incorrect results in favor of accessories (like a rice search) over the main product (like a rice cooker). It usually works in two stages: matching and ranking. The main candidates are products containing words in the query. Also included in the candidate set are products bought or clicked following a query. The ranking step is for those applicants to order and optimize customer satisfaction and business metrics through a machine-learning based classification function. To break a string into a sequence of smaller components such as words, phrases, sub-words, or characters and mapped to close, we'll tokenized the query in text and learn embedding for these tokens. we use a neural network to learn these embedding both the query and title are tokenized and the tokens are embedded using a shared [Nigam, 2019]. Since we are not bothering about complexity or building our own algorithm from scratch, it is possible to use other API to expedite the progress. For that, I'm using a pre-trained model for my project. This model is being trained by multi-task CNN with a Large Training Corpus that is used for context-specific token vectors, POS tags, dependency parse and named entities. Spacy a python-based library using here for making a pipeline with a pre-trained model. For text ranking, I am using pytextrank[Rada, 2004], which provides fast, effective phrase extraction from texts, along with extractive summarization. It returns with a probability score with the specific keyword. I tested with different search queries; it gives a very convincing level result. The author of this paper stated that this works well because it does not only rely on the local context of a text unit but rather it takes into account information recursively drawn from the entire text. .

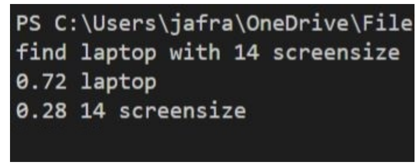

Figure 4: Semantic Ranking

3.3 Interaction Modeling

This is a major part of how the data from the input model will be transmitted and received and the data will be analyzed based on its tasks, and then the result will then return used to make the final decision. It will interact with the users and the system. It will communicate at both ends of a whole system simultaneously. With the Interaction Manager, the result will be improved through the chain of conditions and derivatives. All the facts and rules are analyzed and then sorted before a solution is concluded. The purpose of using a knowledge base is to store previous query and current query so that it can be used in future. Also, when you create another query or update information, it needs to change its status and keep update to maintain a coherent to each text response that made.

3.4 Natural Language Generation

The Natural Language Generation must provide a natural language utterance of the system in the light of a dialog system response. This was often done with handcrafted rules in commercial dialog systems. A further way to select a natural language answer is to learn a discriminatory model.[Serban, 2015] In this case, a number of so-called phrases of surface form can be defined in the output space. The classification model has to select a suitable surface shape in view of the system response. The selected surface form then adequately replaces the placeholder values.

3.5 Product Search Engine

Although this is a non-NLP part, it is one of the important parts of the whole research. Our search query should produce some results based on the input the user given on the system. Suppose, if we get "laptop", "14 screensize" from interaction manager, it should have the ability to make a search query on the web those words and able to return it by their context. In this case,I saved a search query for a laptop from the Walmart website and save it as HTML format. With the help of beautifulsoup4, I extracted all the information from DOM and filter it out by a specific keyword. In this case, I'm getting two keywords "laptop" with 0.72 probability, "14 screensize" with 0.28 (figure: 4). It is very likely possible that the first keyword should be a search query, followed by the rest of the keyword will be used as filtering on

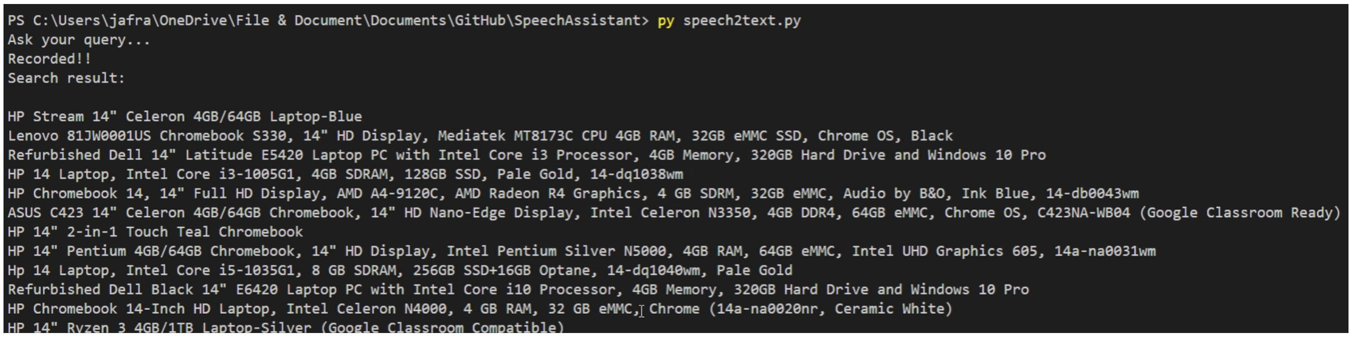

Figure 5: Product search result

the web page. Having this rule, I filter out information and get a set of lists of laptops that each laptop has 14-inch screen size(figure: 5). Since we are working for non-sighted people, so the result should be interpreted by the speaker, so that they understand they understand the request they made vocally for a specific product is successfully return its result.

3.6 Dataset

We introduce a custom dataset in our project. We collected over 200 web pages and label them with custom attributes. We focus on four specific sections webpage: filter, search, sort, page. We added data-attribute="section name" on each element on webpage for further implementation. For example, we added a "data-atttribute= filter" for those elements, who used filter product on the webpage. Our next task is to validate dataset characteristics and this result will be added to the final report.

4 Conclusion

In this research, we present a general architecture of a dialogue system for online shopping, aiming to assist non-sighted people with the help of natural language conversation. We discussed each stage of its architecture at a very high level. In the future, we'll continue to add experiments to this project at each stage. Based on it, we will make a decision about whether we need to prove any particular section or need to explore different approaches for both dialog system and product search.

References

[Jiang, 2013] Harold Abelson, Gerald Jay Sussman, and Julie Sussman. *Structure and Interpretation of Computer Programs.* MIT Press, Cambridge, Massachusetts, 1985.

[Chelba, 2013] Ciprian, and Johan Schalkwyk. *Empirical exploration of language modeling for the google. com query stream as applied to mobile voice search..* Mobile Speech and Advanced Natural Language Solutions. Springer, New York, NY, 2013. 197-229.

[Schalkwyk, 2010] Beeferman, D., Beaufays, F., Byrne, B., Chelba, C., Cohen, M., Kamvar, M. and Strope, B. your word is my command. Advances in speech recognition. Springer, Boston, MA, 2010. 61-90.

[Zweig, 2011] Geoffrey, and Shuangyu Chang. *Personalizing model m for voice-search.* Twelfth Annual Conference of the International Speech Communication Association. 2011.

[Shan, 2010] Wu, G, Hu, Z, Tang, X, Jansche, M, and Moreno, P J. *Search by voice in mandarin chinese.* In Eleventh Annual Conference of the International Speech Communication Association.

[Enembreck, 2004] Fabrício, and Jean-Paul Barthès. *Dialog with a personal assistant.* Brazilian Symposium on Artificial Intelligence. Springer, Berlin, Heidelberg.

[Kepuska, 2018] Veton, and Gamal Bohouta. *Next-generation of virtual personal assistants (microsoft cortana, apple siri, amazon alexa and google home).* IEEE 8th Annual Computing and Communication Workshop and Conference (CCWC). IEEE.

[Yan, 2017] Duan, N., Chen, P., Zhou, M., Zhou, J., and Li, Z. *Building task-oriented dialogue systems for online shopping.* 25th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining.

[Nigam, 2019] Song Y, Mohan V, Lakshman V, Ding W, Shingavi A, Teo CH, Gu H, Yin B *Semantic product search.* In Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence (pp. 4618-4625).

[Zlatareva, 2018] *A Specialized Recommender Agent for the Semantic Web.* The 4th World Congress on Electrical Engineering and Computer Systems and Science.

[Misu, 2006] Teruhisa, and Tatsuya Kawahara. *A bootstrapping approach for developing language model of new spoken dialogue systems by selecting web texts.*

Ninth International Conference on Spoken Language Processing.

[Wang, 2019] Jiang, Jingfang Xu, and Qi Zhang. *Knowledge graph construction and applications for Web search and beyond.* Data Intelligence 1.4 (2019): 333-349.

[Kumar, 2017] Christoph Schmidt, and Joachim Köhler. *A knowledge graph based speech interface for question answering systems.* Speech Communication 92 (2017): 1-12.

[Lytvyn, 2017] Victoria Vysotska, Petro Pukach, Oksana Brodyak, and Dmytro Ugryn *Development of a method for determining the keywords in the slavic language texts based on the technology of web mining.* Computer Science Eastern-European Journal of Enterprise Technologies (2017): 14-23

[Prajapati, 2018] Atul Prakash, Ashish Chandiok, and D. K. Chaturvedi *Semantic Network Based Cognitive, NLP Powered Question Answering System for Teaching Electrical Motor Concepts.* International Conference on Intelligent Information Technologies. Springer, Singapore

[Kinne, 2019] David Lenz *Predicting innovative firms using web mining and deep learning.* ZEW-Centre for European Economic Research Discussion Paper 19-01 (2019).

[Trupthi, 2019] Mandhula, Suresh Pabboju, and Narsimha Gugulotu *Deep Sentiments Extraction for Consumer Products Using NLP-Based Technique.* Soft Computing and Signal Processing. Springer, Singapore, 2019. 191-201.

[Serban, 2015] Iulian Vlad, Ryan Lowe, Peter Henderson, Laurent Charlin, and Joelle Pineau *A survey of available corpora for building data-driven dialogue systems.* The journal version." Dialogue and Discourse 9.1 (2018): 1-49.

[Rada, 2004] Rada Mihalcea, Paul Tarau *Textrank: Bringing order into text.* Proceedings of the 2004 conference on empirical methods in natural language processing.