A Knowledge Graph Based Speech Interface for Question Answering Systems

Ashwini Jaya Kumar, Christoph Schmidt, Joachim Koehler

PII: S0167-6393(16)30144-3
DOI: 10.1016/j.specom.2017.05.001
Reference: SPECOM 2459

To appear in: *Speech Communication*

Received date: 3 June 2016
Revised date: 1 January 2017
Accepted date: 17 May 2017

Please cite this article as: Ashwini Jaya Kumar, Christoph Schmidt, Joachim Koehler, A Knowledge Graph Based Speech Interface for Question Answering Systems, *Speech Communication* (2017), doi: 10.1016/j.specom.2017.05.001

A Knowledge Graph Based Speech Interface for Question Answering Systems

Ashwini Jaya Kumar, Christoph Schmidt and Joachim Koehler

***Abstract*—Speech interfaces to conversational systems have been a focus in academia and industry for over a decade due to its applicability as a natural interface. Speech recognition and speech synthesis constitute the important input and output modules respectively for such spoken interface systems. In this paper, the speech recognition interface for question answering applications is reviewed, and existing limitations are discussed. The existing spoken question answering (QA) systems use an automatic speech recogniser by adapting acoustic and language models for the speech interface and off-the-shelf language processing systems for question interpretation. In the process, the impact of recognition errors and language processing inaccuracies is neglected. It is illustrated in the paper how a semantically rich knowledge graph can be used to solve automatic speech recognition and language processing specific problems. A simple concatenation of a speech recogniser and a natural language processing system is a shallow method for a speech interface. An effort beyond merely concatenating these two units is required to develop a successful spoken question answering system. It is illustrated in this paper how a knowledge graph based structured data can be used to build a unified system combining speech recognition and language understanding. This facilitates the use of a semantically rich data model for speech interface. We conclude by discussing the limitations of existing spoken QA and a knowledge graph based speech interface.**

***Keywords*—Spoken Question Answering, Knowledge Graphs, Automatic Speech Recognition, Spoken Language Understanding, Spoken Interface, Linked Data**

I. Introduction

A QUESTION answering system can be defined as a system which searches for a suitable answer in a knowledge base for a given question by the user. The answer may be one word, a sentence snippet, a well constructed and meaningful sentence or a collection of sentences with a logical coherence. The answer type depends on the application for which a question answering system is developed. A question answering (QA) system can be developed on different paradigms: database querying, information retrieval and knowledge graph based. The database querying method typically involves developing a question-answer pair database for a specific domain and then fetching the answer depending on the user question. Information retrieval is usually finding unstructured data, i.e. text, that satisfies the user query from large collections of data, usually on the Web. The knowledge graph based approach involves a semantic analysis of the query and then accessing a structured database [1]. The database can be a fully relational database or simple structured databases like sets of Resource Description Framework (RDF)[1] triples. A speech interface to the above discussed QA systems is generally called spoken question answering. A general pipeline of spoken question answering is illustrated in Figure 1.

A speech interface to a question answering system makes the interaction more natural. To achieve this, the most common approach is to combine an automatic speech recognition and a speech synthesis unit with a question answering system. The development of an automatic speech recognition (ASR) and a speech synthesis (SS) modules are straight forward given a database of question-answer pairs. The complexity increases if a QA system is based on information retrieval or on a knowledge graph, since it involves a semantic analysis of the question. The challenges involved in developing a spoken QA is reviewed in [2]. The ASR and SS systems have their own dependencies and unsolved research problems, adding to the complexities of QA system development. In this paper, we are addressing only the speech recognition part of the interface, which is the first step of a spoken QA process and also more complicated than speech synthesis because of the necessity to recognise multiple speakers. In contrast, a speech synthesiser can be developed based on a single speaker. The core research challenges (e.g. multiple speakers, ambient noise, different dialects, etc.) in ASR are not the focus of this paper. The main focus is on the interface between a speech recogniser and a QA system. It has to be noted that many question answering systems which integrate speech recognition without speech synthesis in the overall architecture are also called "speech interfaced question answering" or "spoken question answering". We are following the same terminology, as our focus is only on the speech recognition interface. It should be mentioned that spoken question answering can also be developed for spoken documents, but the focus of this paper is only on text documents.

A review of existing spoken question answering systems is presented in this paper, and the limitations are discussed. The existing spoken QA systems use ASR by adapting the acoustic and the language model to a pre-determined set of QA data. After the recognition step, language processing is carried out to interpret the recognised questions. In most of the methods discussed in the review section, the focus is on developing a spoken QA system rather than to address

Ashwini Jaya Kumar is with Netmedia, Fraunhofer IAIS, Germany and Enterprise Information Systems, University of Bonn, Germany. e-mail: ashwini.jaya.kumar@iais.fraunhofer.de

Christoph Schmidt and Joachim Koehler are with Netmedia, Fraunhofer IAIS, Germany. e-mail: christoph.andreas.schmidt@iais.fraunhofer.de and joachim.koehler@iais.fraunhofer.de

Manuscript received June 2, 2016; revised .....

[1] https://www.w3.org/RDF/