

Speech Search: Techniques and Tools for Spoken Content Retrieval

Gareth J. F. Jones
CNGL Centre for Global Intelligent Content
School of Computing, Dublin City University, Dublin 9, Ireland
gjones@computing.dcu.ie

1. OVERVIEW

Speech search is concerned with the retrieval of spoken content from archives of speech or multimedia materials. Importantly speech search does not reduce to application of text information retrieval to speech recognition transcripts. Although text information retrieval techniques form a core technology for speech search, speech search generally involves confronting issues less likely to arise in the text domain, such as high levels of noise in the data and the absence of clearly defined units of retrieval. This tutorial will provide researchers in information retrieval with an introduction to the challenges and technologies of spoken content search.

While speech search has been an active area of research for around twenty years, during much of this time activities have focused on retrieval from collections of well defined and clearly structured content such as radio and television news. The tutorial will review this established work based on the detailed descriptions contained in [4]. The current rapid growth in recordings of more informal audio content, e.g. in educational and enterprise settings, is creating new opportunities and demands for speech search technologies. Despite the developments in existing speech search technologies, these types of data pose challenges for which the solutions proposed to date are either incomplete or require extension or new methodologies. The tutorial will draw on current research to demonstrate the retrieval challenges posed by this content, and introduce work which aims to develop effective solutions, e.g. for search of internet video archives [3], meetings [2] and lectures and presentations [1].

2. TUTORIAL STRUCTURE

The tutorial will be structured as follows:

- Introduction to the scope of the topic of speech search, its relationship to text information retrieval, and potential application areas.
- History of spoken content retrieval research, including: early studies using private datasets, the TREC Spoken

Document Retrieval task for broadcast news, and interview retrieval in the CLEF CL-SR task.

- Overview of speech technologies for spoken content indexing, including methods for standard large vocabulary transcriptions which have been widely used in speech retrieval, and methods developed specifically for speech search including keyword spotting and phone and word lattices.
- Exploiting additional resources and techniques for spoken content retrieval, including: text metadata, collateral texts, and query and document expansion.
- Searching informal or spontaneous conversational speech where the content of the documents can be insufficient to actually support effective search. This will include advanced indexing and structuring methods such as use of external metadata and content segmentation to form effective retrieval units.
- Interactive access to spoken content within specific items, designed to address the inefficiencies of temporal browsing.
- Evaluation benchmarks for spoken content retrieval, including a summary of the conclusions of these campaigns and their technical strengths and limitations.
- Thoughts on research directions and open challenges.

3. REFERENCES

- [1] T. Akiba, H. Nizhizaki, K. Aikawa, X. Hu, Y. Itoh, T. Kawahara, S. Nakagawa, H. Nanjo, and Y. Yamashita. Overview of the NTCIR-10 SpokenDoc-2 Task. In *Proceedings of NTCIR-10 Workshop Meeting*, Tokyo, Japan, 2013.
- [2] M.Eskevich and G.J.F.Jones. Exploring speech retrieval from meetings using the AMI corpus. *Computer Speech and Language (Special Issue on Information Extraction and Retrieval)*, 2014.
- [3] M.Eskevich, G.J.F. Jones, M.Larson, C.Wartena, R.Aly, T.Verschoor, and R.Ordelman. Comparing retrieval effectiveness for alternative content segmentation methods for internet video. In *Proceedings of the 10th Workshop on Content-Based Multimedia Indexing (CBMI 2012)*, 2012.
- [4] M.Larson and G.J.F.Jones. Spoken content retrieval: A survey of techniques and technologies. *Foundations and Trends in Information Retrieval*, 5(4-5):235–422, 2011.

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage, and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s). Copyright is held by the author/owner(s).

SIGIR'14, July 6–11, 2014, Gold Coast, Queensland, Australia.

ACM 978-1-4503-2257-7/14/07.

<http://dx.doi.org/10.1145/2600428.2602293>.