Usable Transcriptions of Webcast Lectures

A Thesis Proposal

Cosmin Munteanu

Department of Computer Science, University of Toronto

May 2007

Abstract

Webcasts are one of the emerging technologies associated with the expansion of the World Wide Web. Despite voice being currently present in all webcasts, it is not used beyond playback. My goal is to exploit this information-rich resource and improve webcast users' experience in browsing and searching for specific information, by combining research in Human-Computer Interaction and Automatic Speech Recognition that would ultimately see text transcripts of lectures and presentations being integrated in webcast archives.

In this dissertation, I show that the usefulness and usability of automatically-generated transcripts of webcast lectures and presentations can be improved by integrating novel speech recognition techniques specifically addressed at increasing the accuracy of webcast transcriptions with the development of an interactive collaborative interface that facilitates

users' contribution to the improvement of machine-generated transcripts. For this, I first investigate the user needs for transcription accuracy in webcast archives and show that users' performance and transcript quality perception is linearly affected by the Word Error Rate (WER), with WER equal to or less than 25% being acceptable for use in webcast archives. As current Automatic Speech Recognition (ASR) systems can only deliver, in realistic lecture conditions, WERs of around 45%, I propose and evaluate an extension to the ePresence webcast system that engages users to collaborate in a wiki manner on editing the imperfect ASR transcripts. My research on ASR systems focuses on reducing the WER for lectures by making use of available external knowledge sources, such as the World Wide Web and the lecture slides, to better model the conversational and the topic-specific styles of lectures. Further ASR improvements are proposed that combine the research on language modelling with aspects of collaborative transcript editing by extracting information from the user-edited transcripts about the conversational style of a particular lecturer and about the topic of the presentation to further refine the language models. Finally, as WER does not always reflect the users' view of text quality, a user-motivated measure of transcript quality is also proposed and evaluated.