Two-stream indexing for Spoken Web search

Date Reviewed: 06/29/11

The Spoken Web affects many people in rural India and other developing areas that have just recently adopted the widespread use of voice-only cell phones (as opposed to smartphones). The Spoken Web, like the World Wide Web (WWW), enables users to browse and surf for information. The presentation, however, is spoken language instead of text and graphics, so users can access content using a cell phone with no display. Since there is no textual information, though, how does one search this Web? This is exactly the topic of this paper.

This paper presents a technique that combines metadata and speech recognition into an indexing approach for Spoken Web search. The paper describes different variants of algorithms, and then presents an evaluation of precision and recall based on more than 20,000 voice documents. It provides evidence that the combination of the metadata and the speech recognition stream performs better than comparable algorithms on a single modality.

This paper is for speech and natural language processing researchers. I definitely recommend it to those working in such fields, even if they are not working on the Spoken Web. Though the authors performed the experiments within this domain, as a speech and multimedia researcher, I see no reason why the techniques could not also be applied to "found data" from the WWW, such as consumer-produced videos on social networking sites. The characteristics of the data match reasonably.

Reviewer:  Gerald Friedland
Review #: CR139192