Crowdsourcing User Preferences and Query Judgements for Speech-Only Search

Johanne R. Trippas, Damiano Spina, Lawrence Cavedon, and Mark Sanderson (johanne.trippas, damiano.spina, lawrence.cavedon, mark.sanderson)@rmit.edu.au

Motivation

Increase usage of speech-only devices allowing us to use search engines

- How to present research results using audio so users can efficiently locate items, determine their relevance, provide feedback, and refine their query if needed?
- How to structure the conversation interaction in order to support the user in the information seeking processes with search engines?

Case Studies

Performance and effectiveness of snippets

- Relevance Judgments: Can users identify known items from inspecting summaries generated from automatic transcripts?
- Preference Judgments: Do users have any preference for different types of document summaries (e.g., generated from automatic/manual transcripts)?

Form of snippets

- How to better support processing of snippets lists
  - What is impact of length of results summaries?
  - Are shorter spoken summaries as effective and preferred as longer more informative summaries

Event Handlers

Gather information about whether:

- The audio was completely played
- The audio was paused
- The user moved or skipped the audio playback to a new position
- Ranges of audio have been played or skipped

Task Workload

Put humans in the loop to interact with speech-only search systems

- ParlAI Framework
  - Collect and evaluate conversations between agents and humans via Mechanical Turk
  - Talk to the bots to help train and evaluate them

End-to-End Evaluation

http://parl.ai/static/docs/mturk.html
