ReMap: Multimodal Help-Seeking
C. Ailie Fraser, Julia M. Markel, N. James Basa, Mira Dontcheva, Scott Klemmer

THE CHALLENGES OF WEB SEARCH

People often seek help via online resources. However, searching the web requires switching context away from the task at hand and coming up with the right keywords for an effective search query.

When asking for help in person, people leverage language, gestures, and shared context. But most online search engines only take text as input.

How can we help people search for learning resources in a more natural way?

LEVERAGING ACCESSIBILITY APIs

OS Accessibility APIs provide a useful entry point for gathering contextual information at a system-wide level.

ReMap uses Accessibility APIs to gather context about the user’s activity for improving search results, and to allow users to make “deictic” references like one might in person (i.e. referencing elements with the words “this” and “that”).

VIDEO SEARCH USING SPEECH, POINTING, AND CONTEXT

ReMap allows users to:
- search for videos using speech (to avoid context-switching),
- point at software elements to include their names in the search query, and
- use voice commands to navigate the video results while working.

QUALITATIVE STUDY

ReMap has been used by 13 participants in a creative task experiment:
- 125 total queries made (118 via speech)
- Only 24 attempts at deictic resolution, 6 of which were successful, likely because:
  - Not all objects had accessibility labels
  - Interface of the study application was relatively simple
- Users tended to multitask while navigating the videos with speech

The graph below summarizes user feedback from this study.

Helpfulness of Speech Features [1-5]

SERVER AND WEB SPEECH API

ReMap leverages the Web Speech API for voice recognition by opening a webpage in the background.

Each video result is displayed as a webpage with a YouTube video player. These webpages render clickable markers on the video timelines to highlight relevant moments.

The server manages all communication between ReMap and the external webpages.

FUTURE WORK

How might we implement more robust and context-aware deictic resolution?

How might we balance the tradeoffs of voice assistants listening too much vs. not enough?

ACKNOWLEDGEMENTS

We would like to thank Professor Christine Alvarado and Vignesh Gokul from the ERSP program for their support.