

WikiBot: Spoken Information Retrieval Dialog System for Wikipedia

Ayushi Aggarwal

Department of Linguistics
University of Washington
ayushiag@uw.edu@uw.edu

Wenxi Lu

Department of Linguistics
University of Washington
wenxil@uw.edu

1 Introduction

Spoken dialog systems often scrape the web and parse out information to create a structured database (Gruenstein et al., 2006) whereas question-answering systems (such as Schofield et al., 2003) rely on a highly semantically structured knowledge base. WikiBot aims to quickly fetch information and respond to keyword-driven queries by mimicking the web-based and site-constrained search experience of Wikipedia through speech-only input. It provides a hands-off search experience which is applicable in public information systems and poverty-stricken neighborhoods where computing systems are sparsely available.

This spoken dialog system was designed to provide the user with a simple and satisfying user experience. The user can search Wikipedia by mentioning the title of the article they want to peruse, switch to Chinese and search the Chinese Wikipedia similarly, or control the volume or pitch of the system. If the system is already in Chinese, the user can switch back to English in addition to all the other features that were available in English. Detailed instructions for using the system is available here¹. The user can access WikiBot through Chrome browser². User manual can be found on the same web page as well.

2 Related Work

Wilcock et al. (Wilcock et al., 2015) built a multilingual spoken information access system that utilizes the Wikipedia database and language-specific speech components. Further, the user can switch between English, Japanese and Finish, and

change topics smoothly, following the hyperlinks in Wikipedia by simply mentioning the name of that link. VoicePedia (Sherwani, et al., 2007) developed a voice-based Wikipedia navigation system that closely mirrors our vision for WikiBot.

3 Resources and Framework

For the purpose of this application, we built a framework from the ground-up using some of the open-source and readily available Google and Wikipedia APIs. Primarily, this voice-driven information access system is powered by the Javascript Web Speech API³ which provides various voice recognition features such real-time speech recognition, speech to text, event handling, and language and speech settings control. To provide visual feedback to the user, WikiBot possesses a front-end that was designed using Javascript. However, the user will not be able to interact with the web-application in any way other than by providing speech input. Speech Synthesis API⁴ facilitates text to speech and allows language, pitch, volume and speaking rate control. The search string provided by the user is used to perform a title-based search of Wikipedia topics using the Wikipedia API⁵. API calls to the English and Chinese Wikipedias returns a JSON object with the content of the entire Wikipedia page for the user-given search string. For the purpose of this project, we extract and return a snippet of the first search result matching the user's query. Further work in this direction has been discussed in the Future Work section.

¹<http://students.washington.edu/wenxil/575/README.pdf>

²<http://students.washington.edu/wenxil/575/WikiBot.html>

³<https://developers.google.com/web/updates/2013/01/Voice-Driven-Web-Apps-Introduction-to-the-Web-Speech-API>

⁴<https://developer.mozilla.org/en-US/docs/Web/API/SpeechSynthesisUtterance>

⁵https://www.mediawiki.org/wiki/API:Main_page

4 Evaluation and Results

We evaluated our system informally to roughly gauge the success of user queries and the current and future scope of functionality.

1. **Basic functionality** - The app has realized basic bilingual wikipedia search along with some possible setting changes such as volume, pitch.
2. **Search speed** - Speech-based search was found to be faster than web-based search by comparing which result was generated faster for five queries. Currently, the system performs explicit confirmation. Information access speech can further be enhanced with implicit confirmation of the user's search query.
3. **Disambiguation** - In its current state, the system cannot disambiguate between keywords with different Wikipedia article entries. It returns the first matching article from the result of the API call to the Wikipedia API. While some of this is due to the existence of topics with different meanings - such as 'Julius Caesar' which could represent the historical figure or the play by Shakespeare. Additionally, lack of disambiguation sometimes leads to results that are not the most accurate. For instance, a search for 'dumpling' does not result in a snippet from the 'Dumpling' wikipedia article(as checked by web-based Wikipedia search).
4. **Grammar** - The system is highly keyword-driven and does not possess a flexible grammar. For example - If you the user says 'search' instead of 'search for something' when prompted with the main menu prompt, system will not recognize the input and will re-prompt the user.

5 Challenges

One of the foremost challenges with this system was asynchronous communication between the user and the Web Speech API. Coordination of user input with the system's responses and prompts was a major challenge that could only be overcome through a menu-based dialog design and turning off the continuous attribute of the speech recognizer. While we were able to perform confirmation of user input, user barge-in remains to

be implemented. Given the length and descriptive unstructured nature of the text in Wikipedia articles, the user should be allowed to stop the system and request a different query.

We contemplated using VoiceXML for our system, but opted for a web-based application in the interest of learning new up-to-date technologies. This decision also seems prudent since a speech-based Wikipedia application is mostly keyword driven and does not require a lot of grammar variation. Moreover, the Wikipedia API can be plugged in seamlessly with Javascript.

6 Caveats and Future Work

WikiBot is currently heavily keyword-driven and has a limited grammar. It is also mostly a system initiative application that does not allow barge-in by the user.

While our system successfully searches Wikipedia based on a keyword and extracts snippets from Wikipedia given that an article exists for it, there is clearly some disparity between the voice and web-based Wikipedia search experience. To bridge this gap we could further implement search result navigation to access specific sections of the Wikipedia article. Also, to interrupt the system while it is speaking requires activating speaking and listening simultaneously. Another future work could be navigation across articles by clicking on hyper-links. Disambiguation strategies could also be performed on the user input to further enhance the success rate.

The system can currently only be run in Chrome browser since it uses the browser's built in Speech Recognition and Speech Synthesis APIs. Since this Web Speech API support is also available in Firefox and Edge, we predict that the system would work in these browsers too. However, system compatibility testing is intended for the future.

Lastly, we envisage addition of more languages to the system, however, we would require access to language-based speech components to make multi-lingual search more robust.

7 Conclusion

WikiBot performs a faster keyword-based search for a voice interface than the GUI-based alternative. However, search result navigation and page browsing are significantly slower for speech-based search applications. With the addition of navigation across hyper-links and to different sections of

the Wikipedia articles, WikiBot could become a useful information access application deployable on public information systems targeted towards audiences that do not have access to personal computers or smartphones.

8 References

Graham Wilcock and Kristiina Jokinen. (2015) Multilingual WikiTalk: Wikipedia-based talking robots that switch languages. SIGDial 2015.

Sherwani, J., Dong Yu, Tim Paek, Mary Czerwinski, Yun-Cheng Ju, and Alex Acero. "Voicepedia: towards speech-based access to unstructured information." In INTERSPEECH, vol. 7, pp. 146-149. 2007.

Gruenstein, A., Seneff, S., and Wang, C., Scalable and Portable Web-Based Multimodal Dialogue Interaction with Geographical Databases. Proc. Interspeech 2006.

Schofield, E., and Zheng, Z., A speech interface for open-domain question-answering, in Proc. ACL 2003.