SearchBot: Supporting Voice Conversations with Proactive Search

Salvatore Andolina
Aalto University
salvatore.andolina@aalto.fi

Valeria Orso
University of Padova
valeria.orso@unipd.it

Hendrik Schneider
University of Göttingen
hendrik.schneider@stud.uni-goettingen.de

Khalil Klouche
University of Helsinki
khalil.klouche@helsinki.fi

Tuukka Ruotsalo
University of Helsinki
tuukka.ruotsalo@helsinki.fi

Luciano Gamberini
University of Padova
luciano.gamberini@unipd.it

Giulio Jacucci
University of Helsinki
giulio.jacucci@helsinki.fi

CSCW '18 Companion, November 3–7, 2018, Jersey City, NJ, USA
ACM 978-1-4503-6018-0/18/11.
https://doi.org/10.1145/3272973.3272990

Abstract

Searching during conversations and social interactions is becoming increasingly common. Although searching could be helpful for solving arguments, building common ground, and reinforcing mutual assumptions, it can also cause interactional problems. Proactive search approaches can enrich conversations with additional information without neglecting the shared and established social norms of being attentive to ongoing interaction. This demo showcases SearchBot, a tool that minimizes the issues associated with the practice of searching during conversations. It accomplishes this by tracking conversational background speech and then providing continuous recommendations of related documents and entities in a non-intrusive way [3].

Author Keywords

Spoken conversation support; proactive search; voice user interface; background speech.

ACM Classification Keywords

H.5.m [Information interfaces and presentation (e.g., HCI)]: Miscellaneous.

Introduction

Searching during a conversation allows people to check facts and to expand their knowledge about topics of interest. However, the typical explicit interactions between the user and the search engine may be laborious and cause disruptions in the flow of the conversation, interruptions of eye contact, and other social interaction problems [1, 8]. Although some approaches have been introduced to improve the social search experience by minimizing the need for textual input [4], here, we take things a step further by proposing a solution that uses proactive search directly from spoken conversational input. Proactive search can leverage information from users' contexts to retrieve information in an easily accessible and non-intrusive manner [9]. Despite the current limitations of automatic speech recognition, recent research on voice-based interaction [2, 5, 6] has shown that relevant contextual information can be extracted even from partial recognition. These findings have inspired our investigation into how proactive search from spoken conversational input could support conversations between individuals.

In [3], we introduce SearchBot, a proactive search agent that listens to conversations, detects the entities mentioned in the conversations, and proactively retrieves and presents information related to the topics of conversation. To demonstrate the viability of the SearchBot approach, we conducted a comparative study with 12 pairs of participants engaging in informal conversations on building travel or movie lists. The conversations were supported by either SearchBot or a more traditional search engine used as the control condition. Our findings showed that information retrieved proactively by an agent listening to the conversation had the potential to effectively support the conversation with facts and ideas without causing much interruption to the conversation's flow but at the cost of participants feeling less in control of the search process. The findings also showed that the proactive search approach retrieved the same number of useful resources supporting the conversation yet without the participants having to formulate explicit queries. In this demo we introduce the SearchBot system to the CSCW community.

Demonstration Description

During this demonstration, conference participants will be invited to engage in conversations on building travel or movie lists with the demo presenter. The conversations will be enriched with additional information automatically retrieved by the SearchBot system, which will be customized for the "movie" and "travel" topics to improve the relevance of the results displayed.

SearchBot Components

In this section we describe the main components of the system demo.

Spoken Conversation Analysis

SearchBot listens to conversations through a microphone. Speech recognition is performed by using Google's implementation of the HTML5 Web Speech API[1]. The speech API takes the audio recording as an input and outputs a transcript in natural language. The speech recognizer listens continuously to the conversation. The voice activity is automatically detected based on the audio input, and the system begins to build a sentence from the input. After the activity stops, the system returns the recognized sentence. As soon as the system recognizes and returns the sentence transcript, it triggers the entity detection and recommendation component.

Entity Detection and Recommendation

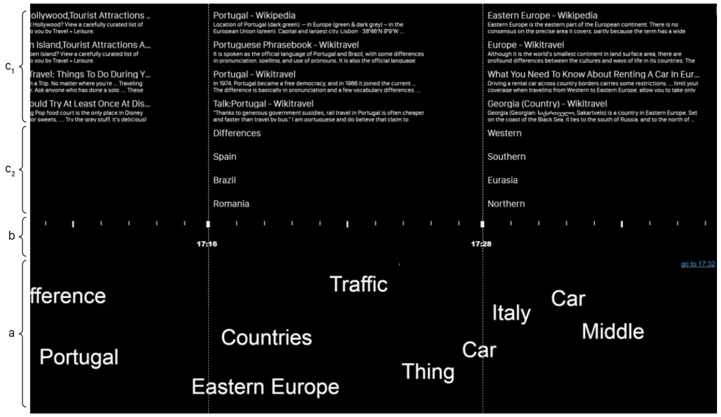
Each transcription is processed by Google's Cloud Natural Language API[2], which is used to extract recognized entities from the transcripts. The API returns entities along with

[1] https://w3c.github.io/speech-api/speechapi.html

[2] https://cloud.google.com/natural-language/

information about their named entity types. For example, people, locations, and organizations are separately typed.

To recommend new entities based on the detected ones, we model them using a vector space model [10]. To recommend new entities, we train an entity embedding model using Word2Vec [7] on a complete English Wikipedia. Each of the detected entities from the current transcript is represented as a vector in the embedding space. The embedding model is used by first combining the vectors of the words in the recognized entities and then retrieving new entities by ranking other entities using their cosine similarity in the embedding space. Altogether, the four highest-ranking entities are retrieved in response to each transcript (Figure 1c$_2$). For example, for the input "Bordeaux", "France", and "wines", the system computes a cosine distance for an input vector "Bordeaux" + "France" + "wines" and retrieves the entities "Bandol," "sauternes," "wines," and "Marseille," which have the smallest cosine distance to that vector.



ollywood,Tourist Attractions ..
t Hollywood? View a carefully curated list of
o you by Travel + Leisure.

n Island,Tourist Attractions A...
pen Island? View a carefully curated list of
o you by Travel + Leisure.

Travel: Things To Do During Y...
n a Trip. No matter where you're ... Traveling
. Ask anyone who has done a solo .... These

ould Try At Least Once At Dis...
g Pop food court is the only place in Disney
or sweets. .... Try the grey stuff, it's delicious!

Portugal - Wikipedia
Location of Portugal (dark green). – in Europe (green & dark grey) – in the European Union (green). Capital and largest city. Lisbon · 38°46′N 9°9′W ...

Portuguese Phrasebook - Wikitravel
It is spoken as the official language of Portugal and Brazil, with some differences in pronunciation, spelling, and use of pronouns. It is also the official language

Portugal - Wikitravel
In 1974, Portugal became a free democracy, and in 1986 it joined the current ... The difference is basically in pronunciation and a few vocabulary differences. ...

Talk:Portugal - Wikitravel
"Thanks to generous government susidies, rail travel in Portugal is often cheaper and faster than travel by bus." I am portuguese and do believe that claim to

Eastern Europe - Wikipedia
Eastern Europe is the eastern part of the European continent. There is no consensus on the precise area it covers, partly because the term has a wide

Europe - Wikitravel
Although it is the world's smallest continent in land surface area, there are profound differences between the cultures and ways of life in its countries. The

What You Need To Know About Renting A Car In Eur...
Driving a rental car across country borders carries some restrictions. ... limit your coverage when traveling from Western to Eastern Europe, allow you to take only

Georgia (Country) - Wikitravel
Georgia (Georgian: საქართველო, Sakartvelo) is a country in Eastern Europe. Set on the coast of the Black Sea, it lies to the south of Russia, and to the north of ...

Figure 1: The user interface of the SearchBot system. The system monitors a conversation and provides continuous recommendations of related documents and entities in a non-intrusive way. a) Stream of recognized entities; b) timescale with timecodes; c$_1$) recommended documents; c$_2$) recommended entities.

Document Retrieval
Related documents are retrieved via Google Custom Search by combining entities recognized in the present transcript to a query. Entities of type "location" or "person" are prioritized to improve the relevance of the results displayed. If an entity of such type is identified, a separate query is generated using that specific entity, and the other entities are combined to that query. Altogether, four search results are retrieved in response to each transcript (Figure 1c$_1$).

More specifically, anytime the recognizer detects pauses, silence, or non-speech audio, a new sentence is returned. From the sentence, a set of entities is extracted, and a type is determined for each entity. If some of these entities are named entities–in our case, "location" or "person" type–they are stored in a separate named entity query vector. All of the entities are also used to form another general query vector. The final set of search results shown to the user is then computed as the union of the highest-ranked results in response to both query vectors. In case none of the entities falls into the location or person categories, only the latter vector is used for retrieval.

An example of a sentence, the extracted entities and their types, and the query vectors is given below:

Sentence: Bordeaux is famous for its wines.
Entities: Bordeaux (type *location*),
wines (type *consumer good*)
Named entity query vector: Bordeaux
General query vector: Bordeaux + wines

User Interface Design

The user interface operates on a regular Web browser. It consists of a timeline that displays a stream of recognized entities in the lower area of the window (Figure 1a), a timescale with timecodes displayed in the center (Figure 1b), and successive sets of four retrieved documents (Figure 1c_1) and four recommended entities in the upper part of the window (Figure 1c_2). A new set extends the timeline every time a new transcription is available. The user can interact with the system in multiple ways. Clicking on recognized or recommended entities triggers a search and opens the most relevant article in a new tab. Clicking on a document will open its content in a new tab. Users can also move back and forth on the timeline by clicking on and dragging the central portion of the window.

Conclusions

In this demo we introduce SearchBot [3] to the CSCW community. SearchBot supports conversations by minimizing the need to perform explicit searches. With SearchBot searches are proactively performed in the background by using naturally occurring spoken conversational contexts between individuals while they can remain focused on the conversation and pay attention to their personal devices only when additional information is needed.

REFERENCES

1. Jesper Aagaard. 2016. Mobile devices, interaction, and distraction: a qualitative exploration of absent presence. *AI & SOCIETY* 31, 2 (2016), 223–231.
2. Salvatore Andolina et al. 2015. InspirationWall: Supporting Idea Generation Through Automatic Information Exploration. In *Proceedings of the 2015 ACM SIGCHI Conference on Creativity and Cognition (C&C '15)*. ACM, 103–106.
3. Salvatore Andolina et al. 2018a. Investigating Proactive Search Support in Conversations. In *Proceedings of the 2018 Designing Interactive Systems Conference (DIS '18)*. ACM, 1295–1307.
4. Salvatore Andolina et al. 2018b. Querytogether: Enabling entity-centric exploration in multi-device collaborative search. *Information Processing & Management* 54, 6 (2018), 1182–1202.
5. Barry Brown, Moira McGregor, and Donald McMillan. 2015. Searchable Objects: Search in Everyday Conversation. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing (CSCW '15)*. ACM, 508–517.
6. Donald McMillan, Antoine Loriette, and Barry Brown. 2015. Repurposing Conversation: Experiments with the Continuous Speech Stream. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (CHI '15)*. ACM, 3953–3962.
7. Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013. Efficient Estimation of Word Representations in Vector Space. *CoRR* abs/1301.3781 (2013).
8. Martin Porcheron, Joel E. Fischer, and Sarah Sharples. 2016. Using Mobile Phones in Pub Talk. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing (CSCW '16)*. ACM, 1649–1661.
9. B. J. Rhodes and P. Maes. 2000. Just-in-time Information Retrieval Agents. *IBM Syst. J.* 39, 3-4 (July 2000), 685–704.
10. G. Salton, A. Wong, and C. S. Yang. 1975. A Vector Space Model for Automatic Indexing. *Commun. ACM* 18, 11 (Nov. 1975), 613–620.