May 2022

# Automated Detection and Organization of Objects of Interest in Human Conversations

Jaime Guajardo

Gabi Lanning

Dmitrii Boiarshinov

# Automated Detection and Organization of Objects of Interest in Human Conversations

ABSTRACT

Human conversations often involve discussion of various objects of interest, e.g., recommended restaurants or movies, items to buy, books to read, etc. However, to remember objects mentioned in a conversation, the participants need to write them down immediately. This disclosure describes the use of a virtual assistant to help users memorize objects of interest mentioned in a conversation. A participant can activate a virtual assistant and instruct it to identify and store objects of interest mentioned in a conversation. Upon receipt of the user command and permission to activate a microphone to listen to the conversation, the virtual assistant utilizes natural language processing to detect the mention of objects of interest and stores those in one or more categorized lists. At a later time, the user can ask the virtual assistant to retrieve the identified objects and cross-validate them with their historical location data for prioritization and personalization.

KEYWORDS

- Object of Interest (OOI)
- Point of Interest (POI)
- Restaurant recommendation
- Shopping list
- Virtual assistant
- Natural Language Processing (NLP)
- Personalization
- Contextual suggestion

## BACKGROUND

Conversations among people often involve references to various objects that might be of interest to the conversing parties. For instance, a person may ask a friend for suggestions for restaurants in a neighborhood. In such a case, the various restaurants mentioned in the response constitute objects of interest to the person who initiated the conversation. Currently, users

capture any objects of interest mentioned in a conversation by relying mostly on their memories or by creating a written list. Both of these approaches are suboptimal since it is hard to remember conversations completely and accurately, while engaging in writing during a conversation disrupts the natural flow of conversation.

Voice-based virtual assistants offer assistance for tasks related to various objects of interest. For example, a user can ask a voice-based virtual assistant how far a given restaurant is located from the user's current location and whether a reservation can be made for the user's desired date and time. Currently, users must explicitly invoke a voice-based virtual assistant with a designated wake word before their speech is captured and analyzed to determine whether it includes a command related to an object of interest. Such an operation is unable to capture objects of interest that occur in a natural conversation among users.

## DESCRIPTION

This disclosure describes techniques, implemented with specific user permission, to enable a voice-based virtual assistant to automatically detect and record objects of interest that are mentioned in a natural conversation among users. Users can invoke the functionality by issuing an explicit command for the voice assistant to capture the speech that occurs in a conversation and analyze it via natural language processing (NLP) to detect the mention of objects of interest of specific types, such as restaurants, shops, neighborhoods, brands, etc.

The various objects of interest detected during the conversation can be used to create corresponding lists. For instance, the lists can be constructed by grouping objects based on various criteria such as category (e.g., type of cuisine served at a restaurant), suggested intent (e.g., places to avoid), relevant parameters (e.g., cost), etc. Each object of interest detected in the

conversation can be shown on the user's device, such as a smartphone, along with relevant metadata such as location, ratings, etc.

If an object of interest mentioned in the conversation is unclear or matches with several possibilities, the virtual assistant can optionally seek clarification from the users at an opportune moment during the conversation. For instance, users can be asked to clarify which object they referenced or to select the desired one from a list of possible matches. Users can provide the clarification via spoken answers, on-screen choice selection, etc.

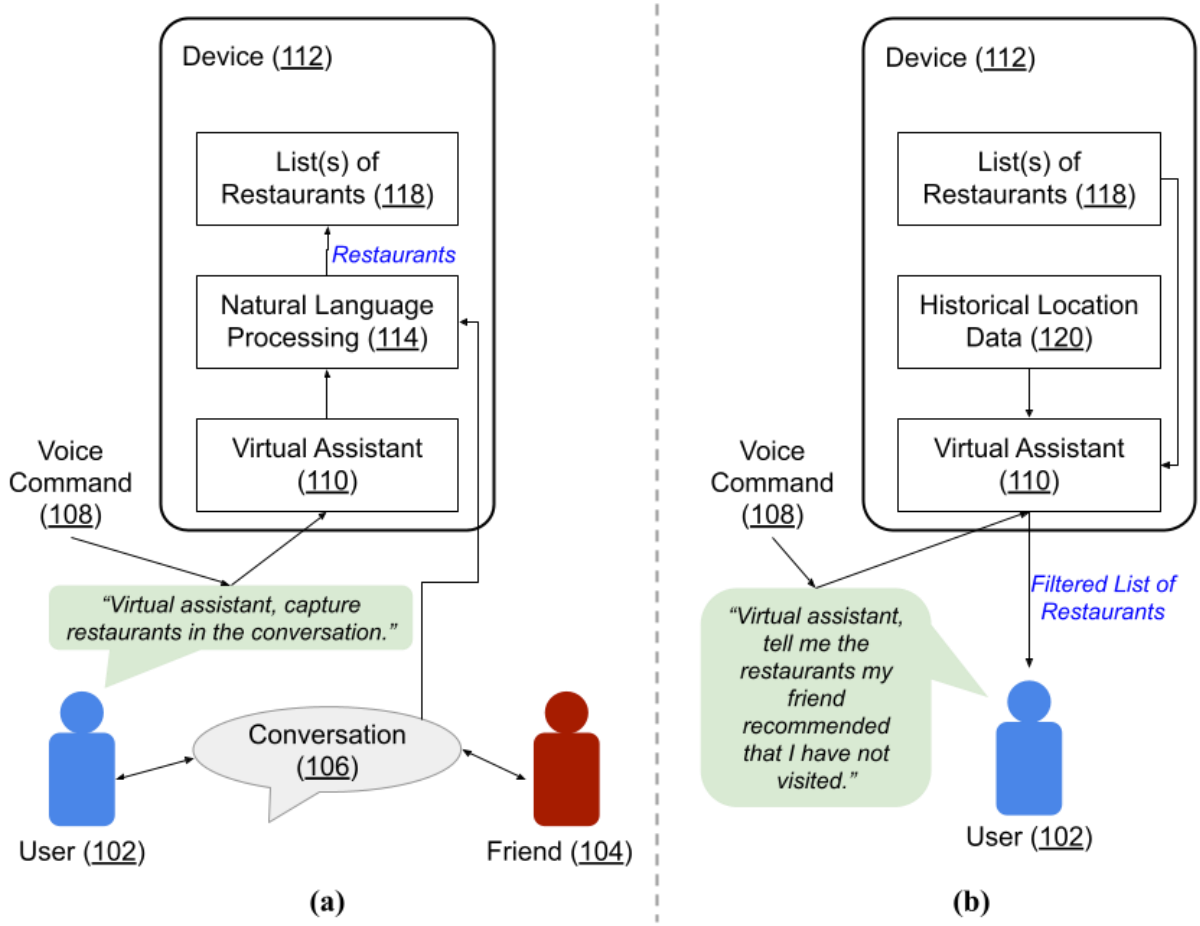


**Fig. 1: Employing a virtual assistant to automatically capture lists of objects of interest**

Fig. 1 shows an example of operational implementation of the techniques described in this disclosure. As shown in Fig. 1(a), a user (102) initiates a conversation (106) with a friend

(104) seeking suggestions for restaurants. The user issues a voice command (108) to a virtual assistant (110) available via the user's device (112) to capture objects of interest from the conversation.

Upon receiving the command, the virtual assistant can activate a device microphone to capture the conversation content and perform natural language processing analysis (114) on the conversation content locally. Based on the analysis, the restaurants (118) mentioned in the conversation are identified and compiled in the form of one or more semantically meaningful lists. The virtual assistant discards captured content as soon as identification is complete; no speech content is stored locally or transmitted to a remote device.

The user can issue another command to instruct the virtual assistant to stop the capture and analysis, or the listening mode is disabled automatically after a short period of time. Upon such a command, all objects of interest captured during the conversation and their associated metadata are presented to the user. The user can make edits to the presented list(s), such as confirming objects detected with low certainty, deleting incorrectly detected objects, moving objects between lists, etc.

After making edits, users can export the list(s) of objects of interest to one or more relevant applications, such as maps, to do lists, files, chat, email, etc. In addition, as depicted in Fig. 1(b), users can review the collection of objects of interest captured in the conversation by issuing commands to the virtual assistant. For instance, a user can ask the virtual assistant, "remind me of the restaurants that my friend recommended." Further, the user can optionally grant the virtual assistant access to historical data about the user's location (120) for cross-validating with the locations of the various objects of interest. Such cross-validation can be used

to filter the objects of interest, such as selecting objects of interest that the user has not yet visited, prioritizing objects of interest near locations frequented by the user, etc.

In addition to listing the objects of interest that fit the user's query, the virtual assistant can provide personalized suggestions to help the user prioritize among the listed options. For example, after listing restaurants that fit the user's query, the virtual assistant can recommend a specific restaurant based on the user's interests and context obtained with the user's permission.

The techniques described in this disclosure can be incorporated with any application or device that includes voice-based virtual assistant capabilities. Implementation of the techniques can eliminate the need for users to memorize and/or write down lists of objects of interest mentioned in everyday natural conversations. The techniques can enable a virtual assistant to automate the storage, retrieval, organization, prioritization, and personalization of objects of interest, thus enhancing their user experience (UX).

Further to the descriptions above, a user may be provided with controls allowing the user to make an election as to both if and when systems, programs or features described herein may enable collection of user information (e.g., information about a user's voice commands, audio conversations, social network, social actions or activities, profession, a user's preferences, or a user's current location), and if the user is sent content or communications from a server. In addition, certain data may be treated in one or more ways before it is stored or used, so that personally identifiable information is removed. For example, a user's identity may be treated so that no personally identifiable information can be determined for the user, or a user's geographic location may be generalized where location information is obtained (such as to a city, ZIP code, or state level), so that a particular location of a user cannot be determined. Thus, the user may

have control over what information is collected about the user, how that information is used, and what information is provided to the user.

## CONCLUSION

This disclosure describes the use of a virtual assistant to help users memorize objects of interest mentioned in a conversation. A participant can activate a virtual assistant and instruct it to identify and store objects of interest mentioned in a conversation. Upon receipt of the user command and permission to activate a microphone to listen to the conversation, the virtual assistant utilizes natural language processing to detect the mention of objects of interest and stores those in one or more categorized lists. At a later time, the user can ask the virtual assistant to retrieve the identified objects and cross-validate them with their historical location data for prioritization and personalization.

## REFERENCES

1. Massai, Lorenzo, Paolo Nesi, and Gianni Pantaleo. "PAVAL: A location-aware virtual personal assistant for retrieving geolocated points of interest and location-based services." Engineering Applications of Artificial Intelligence 77 (2019): 70-85.