SELECTING OPTIONS IN LISTS PRESENTED BY VIRTUAL ASSISTANTS

ABSTRACT

An interactive assistant, referred to herein as “an interactive assistant,” “a virtual assistant,” or simply an “assistant,” is described that allows a user to interrupt presentation of a list of options and select one of the options without having to wait until the virtual assistant presents all of the options in the list. The user may, in other words, stop enumeration of the options in the list by the virtual assistant and select one of the options from the options presented by the virtual assistant up to the point of being stopped by the user. The user may select the option from the list using a reference to the numerical order of the option in the list.

DESCRIPTION

Virtual, intelligent, or computational assistants (e.g., also referred to simply “assistants”) execute on counter-top devices, mobile phones, automobiles, and many other type of computing devices. Assistants output useful information, responds to users’ needs, or otherwise performs certain operations to help users complete real-world and/or virtual tasks. Some assistants are configured to record reminders, notes, lists, and other information as persistent records that a user can later retrieve or modify during a subsequent interaction with the assistant. Some assistants may output lists of options, such as when there are multiple options that satisfy a query by a user. The user selects one of the options from the list only after the assistant has presented the list in its entirety.

A virtual assistant is described that allows a user to interrupt presentation of a list of options and select one of the options without having to wait until the virtual assistant presents all of the options in the list. The user may, in other words, stop enumeration of the options in the
list by the virtual assistant and select one of the options from the options presented by the virtual assistant up to the point of being stopped by the user. The user may select the option from the list using a reference to the numerical order of the option in the list.

In operation, a device may provide an interface with which to interact with the virtual assistant. The device may include a cellular phone (such as a so-called “smart phone”), a tablet computer, a laptop computer, a desktop computer, a workstation, a home audio device (such as a so-called “smart speaker”), a gaming console, a portable gaming console, an audio/visual (AV) receiver, a digital disc player, and the like.

The device may include, in some examples, a display by which to present the interface for interacting with the digital assistant. The device may further include one or more sensors by which to receive data input by a user of the device, the data specifying interactions with the virtual assistant. The sensors may include capacitive touch sensors integrated with the display to capture text interactions, one or more transducers (e.g., a microphone) for capturing audio interactions, cameras for capturing image and/or video interactions, a global positioning system (GPS) sensor for capturing location-based interactions, a compass for capturing positional-based interactions, etc.

In some examples, a server or dedicated device (which may be separate from the device providing the interface by which to interact with the digital assistant - which itself may be referred to as a “client device”) may provide some or all of the functionality of the digital assistant. The server may receive the data indicative of the interaction (e.g., any combination of the above text, audio, image, video, location-based, or position-based interactions) from the client device.
As a user interacts with the assistant, the assistant may obtain personal information about the user. Examples of personal information include: habits, routines, preferences, notes, lists, contacts, communications, interests, location histories, and other types of user information. After receiving explicit permission from the user, the assistant may store, the personal information at user information data stores and, in the course of providing assistant services, make use of the personal information stored at the user information data stores.

A user may be provided with controls allowing the user to make an election as to both if and when the assistant, the computing device, or the computing systems described herein can collect or make use of supplemental data (e.g., user information or contextual information about a user’s social network, social actions or activities, profession, a user’s preferences, or a user’s current location), and if and when 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 supplemental data is collected about the user, how that supplemental data is used, and what supplemental data is provided to the user.

Furthermore, the external systems, the computing devices, and the assistant treat the information stored at the information stores so that the information is protected, encrypted, or otherwise not susceptible to unauthorized use. The information stored at the information data
stores may be stored locally at the computing devices and/or remotely (e.g., in a cloud computing environment provided by the external systems and which is accessible via a network).

In any event, the server may include one or more processors configured to process the data indicative of the interactions in accordance with the digital assistant to generate a result. The server may provide the result to the client device responsive to the data indicative of the interactions.

The digital assistant may represent a software agent trained through, at least in some examples, machine learning to process data indicative of natural language text and audio, and possibly images and video to perform tasks or services for individuals. The digital assistant may, to provide a few examples, answer queries, provide directions to destinations, schedule meetings and/or appointments, generate electronic mail (e-mail) and/or text messages, initiate a telephone call, and the like.

In some instances, the digital assistant may be unable to provide a single result that is responsive to the interactions (which often take the form of a query) due to ambiguity in the query or other reasons. The digital assistant may, as a result of the inability to provide the single result, seek clarification by presenting enumerated lists of options.

To illustrate, consider that a user may enter an interaction requesting that an e-mail be drafted to “David,” the digital assistant may process contact information associated with the user to identify the e-mail address associated with “David.” The user, however, may have a number of different contacts associated with “David,” such as “David A.,” “David B.,” “David C.,” etc. As such, the digital assistant may generate a response with the list of “David” contacts, which the client device may begin to present, e.g., audibly (possibly along with a text version of the list of “David” contacts).
TRANSMIT QUERY TO SERVER

CLIENT

RECEIVE INTERACTION (E.G., QUERY) FROM USER

TRANSMIT QUERY TO SERVER

PROCESS QUERY

SERVER

CLARIFICATION?

PRESENT RESULT TO USER

RETURN RESULT TO CLIENT

PRESENT LIST OF OPTIONS TO USER

GENERATE LIST OF OPTIONS

INTERRUPTION?

NO

YES

HALT PRESENTATION OF LIST OF OPTIONS TO USER

RECEIVE INDICATION OF ONE OF OPTIONS

TRANSMIT INDICATION TO SERVER

RETURN RESULT TO CLIENT

PRESENT RESULT TO USER
As shown in the above flowchart, the client device may receive an interaction from the user (e.g., in the form of a query). The client device may transmit the query to the server. The server may process the query. When a single result is determined, the server may return the result to the client, as no further clarification is required.

However, when multiple results are returned, the server may request additional clarification by generating a list of options that potentially satisfy the query. The server may then transmit the list of options to the client device.

The client device presents the list of options to the user. When audibly presented, the client device may attempt to determine when the user is interrupting the audible presentation in an attempt to better adhere to social norms, and thereby better emulate human assistants. When no interruption is detected, the client device may continue to present the list of options. When an interruption is detected (e.g., possibly via touch or proximity gestures or other touch interactions, audio interactions, physical button interactions – including keyboard interactions, etc., including combinations of interactions), the client device may halt the presentation of the list of options to the user, and process the interruption, allowing the user to specify an indication of one of the options of the list as the option intended in the query.

The user may, when providing audio interactions, specify the intended option by the number of the option in the list. For example, assuming the “David” list above, where “David B.” was the intended option, the user provides data indicative of audio, saying “the second one,” “the second David,” “number two,” etc. The user may also provide data indicative of audio, saying “David B.” Regardless of how the intended option is selected, the client device may transmit the indication to the server, which may process the indication and return a result to the client device.
The client device may then present the result to the user (where the result in the example above would involve opening an e-mail application and generating an e-mail to David. B). That is, the presentation of the result may not necessarily involve actual presentation of some item to the user, but that the server or the client device may perform some operation related to the query, such as texting a contact, emailing a contact, opening an application, scheduling a meeting or other event, initiating a phone call, pushing information to an attached wearable device (such as a head mounted display or a so-called smart watch), etc.

Although described as providing the indication to the server, the client device may locally process the indication of the intended option and provide the result to the user. That is, the client device need not interface with the server hosting the digital assistant to return the result to the client when clarification is required. Furthermore, in some examples, the client device may host the digital assistant itself, thereby avoiding network expense, power consumption and the like associated with interfacing with the remotely hosted virtual assistant.