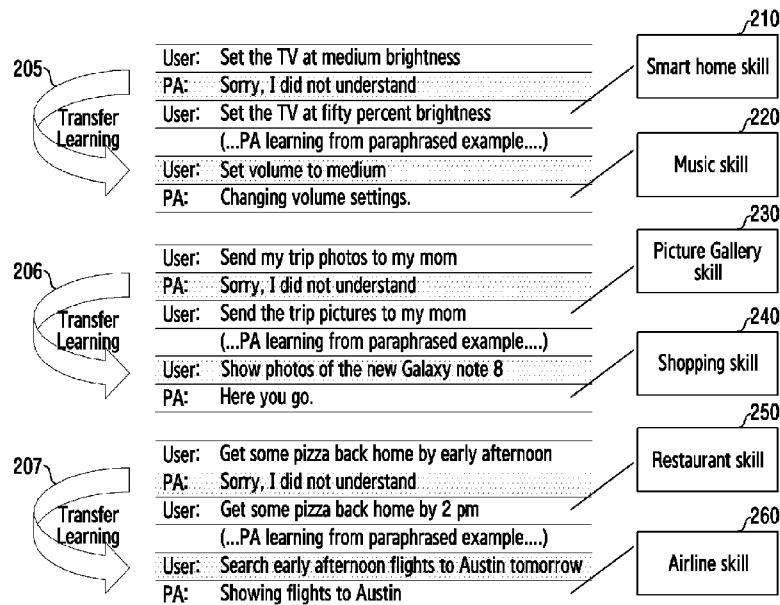




- (51) International Patent Classification:
G06F 17/27 (2006.01) G10L 15/22 (2006.01)
- (21) International Application Number:
PCT/KR2019/002406
- (22) International Filing Date:
28 February 2019 (28.02.2019)
- (25) Filing Language: English
- (26) Publication Language: English
- (30) Priority Data:
15/986,633 22 May 2018 (22.05.2018) US
- (71) Applicant: SAMSUNG ELECTRONICS CO., LTD.
[KR/KR]; 129, Samsung-ro, Yeongtong-gu, Suwon-si, Gyeonggi-do 16677 (KR).
- (72) Inventors: RAY, Avik; 665, Clyde Ave., Mountain View, California 94043 (US). SHEN, Yilin; 665, Clyde Ave., Mountain View, California 94043 (US). JIN, Hongxia; 665, Clyde Ave., Mountain View, California 94043 (US).

- (74) Agent: KWON, Hyuk-Rok et al.; 2F, 28, Gyeonghui-gung-gil, Jongro-gu, Seoul 03175 (KR).
- (81) Designated States (unless otherwise indicated, for every kind of national protection available): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.
- (84) Designated States (unless otherwise indicated, for every kind of regional protection available): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM,

(54) Title: METHOD FOR LEARNING CROSS DOMAIN PERSONALIZED VOCABULARY AND ELECTRONIC DEVICE THEREOF



(57) Abstract: A method for operating an electronic device includes detecting a first natural language (NL) input, determining a skill from the first NL input, if the NL input includes at least one unique skill word, transmitting the first NL input to a custom skill parser for determination of a skill intent, wherein the custom skill parser is trained based on data including at least a custom training data set, and if the NL input does not includes the at least one unique skill word, transmitting the first NL input to a generic parser for determination of a general intent of the first NL input.

WO 2019/225837 A1

TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW,
KM, ML, MR, NE, SN, TD, TG).

Published:

— *with international search report (Art. 21(3))*

Description

Title of Invention: METHOD FOR LEARNING CROSS DOMAIN PERSONALIZED VOCABULARY AND ELECTRONIC DEVICE THEREOF

Technical Field

- [1] One or more embodiments generally relate to personal assistants and, in particular, to learning cross domain personalized vocabulary and automatically generate structure variational paraphrases in smart personal assistants.

Background Art

- [2] Natural language understanding (NLU) is crucial in personal assistants. NLU helps understand a user's utterance and carry out the intent of the user. Each intent can be expressed in many different ways using very different utterances. One important variability arises from the appearance of partial content (e.g., adjuncts, noun phrases, etc.) at different portions in an utterance, as well as the different expressions of such content. For example, for an utterance of "get me a pizza to home by this evening," it can also be stated as follows: "by this evening get me a pizza to," "get me a pizza for home before this evening," "to home, buy me a pizza in this evening," etc. Such variability of one utterance is crucial to train a NLU engine. Otherwise, the NLU engine can only understand a limited number of fixed formats and fails to carry out a user's intent in many scenarios.
- [3] For each specific content, however, it cannot be arbitrarily placed into the sentence. The results of doing so will either change the semantic meaning of the whole utterance, or leave the new utterance completely meaningless. For example, consider the example "get me a pizza to home by this evening." There could be many invalid utterances "get to home me a pizza by this evening," "get to home by this evening a pizza," etc. Therefore, it is non-trivial to generate paraphrases that enable the appearance of partial contents at their correct places.
- [4] State-of-the-art approaches/tools (e.g., API.ai, etc.) or state-of-the-practice digital assistant products heavily rely on a user's manual inputs of utterances to cover the structural variational paraphrases for each intent. Existing paraphrase generation tools may use several simple syntax based sentence fragment rearrangement algorithms, resulting in incorrect paraphrases. Unfortunately, these approaches cannot scalably support a large number of correct variations of an utterance to generate a good training data set for an NLU engine.
- [5] Intelligent personal assistants (PA) are automated artificial intelligence (AI) systems which can take natural language input from a user and perform an action as desired by

the user. In order to enable a PA to perform more actions in a variety of domains, some PAs also provide a special "skills framework" for developers. Using this framework, a developer can add any domain specific skill to the PA, enabling it to understand natural language instructions and perform useful actions in that domain. However, the extent to which the PA is able to understand domain specific utterances are limited by the example utterances provided to the skills framework by the skill developer. Such examples provided by the developer are often limited and does not cover the huge variety of natural language utterances and paraphrases the users can provide to the PA, often using some preferred personal vocabulary. More importantly, the skills framework in current systems cannot leverage the examples provided for one particular skill to enhance the PA's vocabulary or natural language understanding capabilities for some other skill.

Disclosure of Invention

Technical Problem

- [6] One or more embodiments generally relate to learning cross domain personalized vocabulary and generation of structure variational paraphrases in intelligent personal assistants.

Solution to Problem

- [7] In one embodiment, a method for operating an electronic device includes detecting a first natural language (NL) input, determining a skill from the first NL input, if the NL input includes at least one unique skill word, transmitting the first NL input to a custom skill parser for determination of a skill intent, wherein the custom skill parser is trained based on data including at least a custom training data set, and if the NL input does not includes the at least one unique skill word, transmitting the first NL input to a generic parser for determination of a general intent of the first NL input.
- [8] In one embodiment, a method for operating an electronic device includes detecting a first utterance including a first word that is unable to be parsed based on a first skill, detecting a second utterance including a second word instead of the first word, identifying an intent corresponding to the second utterance by parsing the second utterance based on the first skill, adding the first utterance along with the intent into training data associated with the first skill, and training the first skill by using the training data. In one embodiment, a method includes determining, by an electronic device, a skill from a first natural language (NL) input. Upon successful determination of the skill, the first NL input is transmitted to a custom skill parser for determination of a skill intent. The custom skill parser is trained based on data including at least a custom training data set. Upon unsuccessful determination of the skill, the first NL input is transmitted to a generic parser for determination of a general intent of the first

NL input.

- [9] In one embodiment, an electronic device includes a memory storing instructions. At least one processor executes the instructions including a process configured to: determine a skill from a first NL input, upon successful determination of the skill, transmit the first NL input to a custom skill parser for determination of a skill intent, wherein the custom skill parser is trained based on data including at least a custom training data set, and upon unsuccessful determination of the skill, transmit the first NL input to a generic parser for determination of a general intent of the first NL input.
- [10] In one embodiment a non-transitory processor-readable medium that includes a program that when executed by a processor performing a method that includes determining, by an electronic device, a skill from a first NL input. Upon successful determination of the skill, the first NL input is transmitted to a custom skill parser for determination of a skill intent. The custom skill parser is trained based on data including at least a custom training data set. Upon unsuccessful determination of the skill, the first NL input is transmitted to a generic parser for determination of a general intent of the first NL input.
- [11] In yet another embodiment, a method for intelligent assistance that includes generating one or more segments for an NL input, where each segment is based on its dependency on an action term of the NL input. A leading segment is identified from the one or more segments. A first replacement phrase is generated based on concatenation of the leading segment with each non-leading segment of the one or more segments. An output is provided that includes at least a portion of the input and the first replacement phrase.
- [12] These and other aspects and advantages of one or more embodiments will become apparent from the following detailed description, which, when taken in conjunction with the drawings, illustrate by way of example the principles of the one or more embodiments.

Advantageous Effects of Invention

- [13] Various embodiments of the present disclosure provide a cross domain personalized vocabulary learning scheme.

Brief Description of Drawings

- [14] For a fuller understanding of the nature and advantages of the embodiments, as well as a preferred mode of use, reference should be made to the following detailed description read in conjunction with the accompanying drawings, in which:
- [15] FIG. 1 shows a block diagram of architecture for a system including an electronic device including a personal assistant app, according to some embodiments;
- [16] FIG. 2 shows example vocabulary and paraphrase transfer learning, according to

- some embodiments;
- [17] FIG. 3 shows a block diagram of processing for personalized skill intent parsing, according to some embodiments;
- [18] FIG. 4 shows a block diagram of processing for updating a personalized skill parser, according to some embodiments;
- [19] FIG. 5A shows a block diagram of processing for transferring personalized vocabulary, according to some embodiments;
- [20] FIG. 5B shows another block diagram of processing for transferring personalized vocabulary, according to some embodiments;
- [21] FIG. 6 shows a block diagram for segmentation rephrasing and swapping, according to some embodiments;
- [22] FIG. 7 shows a block diagram for utterance segmentation processing, according to some embodiments;
- [23] FIG. 8 shows an example of segmentation of an utterance, according to some embodiments;
- [24] FIG. 9 shows a block diagram of processing for segmentation rephrasing and swapping, according to some embodiments;
- [25] FIG. 10 shows an example of rephrased segments, according to some embodiments;
- [26] FIG. 11 shows an example of rephrased segments and combined paraphrases from rephrased segments, according to some embodiments;
- [27] FIG. 12 shows a block diagram for swapped paraphrase ranking, according to some embodiments;
- [28] FIG. 13 shows an example of ranked paraphrases, according to some embodiments;
- [29] FIGS. 14A and 14B show an example of rephrased and swapped utterances and an output of ranked paraphrases, according to some embodiments;
- [30] FIG. 15 shows a block diagram of a process for generating structural variational paraphrases, according to some embodiments; and
- [31] FIG. 16 is a high-level block diagram showing an information processing system comprising a computing system implementing one or more embodiments.

Best Mode for Carrying out the Invention

- [32] The following description is made for the purpose of illustrating the general principles of one or more embodiments and is not meant to limit the inventive concepts claimed herein. Further, particular features described herein can be used in combination with other described features in each of the various possible combinations and permutations. Unless otherwise specifically defined herein, all terms are to be given their broadest possible interpretation including meanings implied from the specification as well as meanings understood by those skilled in the art and/or as defined in

dictionaries, treatises, etc.

- [33] It should be noted that the terms "at least one of" refers to one or more than one of the elements that follow. For example, "at least one of a, b, c, or a combination thereof" may be interpreted as "a," "b," or "c" individually; or as "a" and "b" together in combination, as "b" and "c" together in combination; as "a" and "c" together in combination; or as "a," "b" and "c" together in combination.
- [34] A portion of the disclosure of this patent document may contain material that is subject to copyright protection. The copyright owner has no objection to the facsimile reproduction by anyone of the patent document or the patent disclosure as it appears in the patent and trademark office patent file or records, but otherwise reserves all copyright rights whatsoever.
- [35] One or more embodiments provide for learning cross domain personalized vocabulary and generation of structure variational paraphrases in intelligent personal assistants. Some embodiments include a method that includes determining, by an electronic device, a skill from a first NL input. Upon successful determination of the skill, the first NL input is transmitted to a custom skill parser for determination of a skill intent. The custom skill parser is trained based on data including at least a custom training data set. Upon unsuccessful determination of the skill, the first NL input is transmitted to a generic parser for determination of a general intent of the first NL input.
- [36] In some embodiments, in order to enable a personal assistant (PA) (e.g., PA 129, FIG. 1) to learn personalized vocabulary used by an end user of the PA, a "personalized skill parser" 360 (FIGS. 3 and 6) may be implemented for each user which can parse skill specific utterances for the set of skills being used by the user. Some embodiments automatically generate structure variational paraphrases, and use them to train a NLU engine. The generated structure variational paraphrases may largely enrich the variability of a training set in order to train an improved NLU engine for the PA 129.
- [37] In some embodiments, the structure variational paraphrases that are generated allow appearances of partial content at different correct places in an utterance or a sentence. One or more embodiments automate user manual generation of such structure variational utterances for each intent, therefore reducing the cost of skill development for a personal assistant. Some embodiments provide an interface for inputting user-specified parts of content in an utterance which a user desires to generate its variability of expressions. In one embodiment, a new structure of a paraphrase of an utterance is generated by swapping a part of contents with other parts of the utterance. A different expression of each part of contents may be generated, and rephrasing the details in utterances is provided to generate more paraphrases. In one embodiment, the ranking of

generated structure variational paraphrases is also provided. Outputs from processing may be directly used to train existing NLU engines.

[38] FIG. 1 shows a functional block diagram of an architecture system 100 that may be used for PAs to enhance natural language understanding capabilities and personalization for PAs such as BIXBY® using an electronic device 120 (e.g., mobile telephone devices, television (TV) systems, cameras, camcorders, a device with audio video capabilities, tablets, pad devices, wearable devices, smart appliances, smart picture frames, smart lighting, etc.). In one embodiment, the electronic device 120 may comprise a display 121, a microphone 122, an audio output 123, an input mechanism 124, communications circuitry 125, control circuitry 126, a camera 128, a personal assistant (or PA) 129 (including, at least, processing for personalized skill intent parsing 300, FIG. 3, a personal skill parser 360 (FIGS. 3 and 6), and/or segmentation rephrasing and swapping 700 processing, FIG. 7, and/or communicating with the communications circuitry 125 to obtain/provide information thereof with the cloud or server 130, and may include any of the processing for, but not limited to, the examples and embodiments as described below), and/or any other suitable components. In one embodiment, applications 1-N 127 are provided and may be obtained from a cloud or server 130, a communications network 110, etc., where N is a positive integer equal to or greater than 1.

[39] In one embodiment, all of the applications employed by the audio output 123, the display 121, input mechanism 124, communications circuitry 125, and the microphone 122 may be interconnected and managed by control circuitry 126. In one example, a handheld music player capable of transmitting music to other tuning devices may be incorporated into the electronics device 120.

[40] In one embodiment, the audio output 123 may include any suitable audio component for providing audio to the user of electronics device 120. For example, audio output 123 may include one or more speakers (e.g., mono or stereo speakers) built into the electronics device 120. In some embodiments, the audio output 123 may include an audio component that is remotely coupled to the electronics device 120. For example, the audio output 123 may include a headset, headphones, or earbuds that may be coupled to communications device with a wire (e.g., coupled to electronics device 120 with a jack) or wirelessly (e.g., BLUETOOTH® headphones or a BLUETOOTH® headset).

[41] In one embodiment, the display 121 may include any suitable screen or projection system for providing a display visible to the user. For example, display 121 may include a screen (e.g., an liquid crystal display (LCD) screen, a light emitting diode (LED) screen, an organic LED (OLED) screen, etc.) that is incorporated in the electronics device 120. As another example, display 121 may include a movable

display or a projecting system for providing a display of content on a surface remote from electronics device 120 (e.g., a video projector). Display 121 may be operative to display content (e.g., information regarding communications operations or information regarding available media selections) under the direction of control circuitry 126.

[42] In one embodiment, input mechanism 124 may be any suitable mechanism or user interface for providing user inputs or instructions to electronics device 120. Input mechanism 124 may take a variety of forms, such as a button, keypad, dial, a click wheel, mouse, visual pointer, remote control, one or more sensors (e.g., a camera or visual sensor, a light sensor, a proximity sensor, etc., or a touch screen. The input mechanism 124 may include a multi-touch screen.

[43] In one embodiment, communications circuitry 125 may be any suitable communications circuitry operative to connect to a communications network and to transmit communications operations and media from the electronics device 120 to other devices within the communications network. Communications circuitry 125 may be operative to interface with the communications network using any suitable communications protocol such as, for example, Wi-Fi (e.g., an IEEE 802.11 protocol), Bluetooth®, high frequency systems (e.g., 900 MHz, 2.4 GHz, and 5.6 GHz communication systems), infrared, global system for mobile communication (GSM), GSM plus enhanced data GSM environment (EDGE), code division multiple access (CDMA), quadband, and other cellular protocols, voice over Internet protocol (VOIP), transmission control protocol (TCP)-IP, or any other suitable protocol.

[44] In some embodiments, communications circuitry 125 may be operative to create a communications network using any suitable communications protocol. For example, communications circuitry 125 may create a short-range communications network using a short-range communications protocol to connect to other communications devices. For example, communications circuitry 125 may be operative to create a local communications network using the BLUETOOTH® protocol to couple the electronics device 120 with a BLUETOOTH® headset.

[45] In one embodiment, control circuitry 126 may be operative to control the operations and performance of the electronics device 120. Control circuitry 126 may include, for example, a processor, a bus (e.g., for sending instructions to the other components of the electronics device 120), memory, storage, or any other suitable component for controlling the operations of the electronics device 120. In some embodiments, a processor may drive the display and process inputs received from the user interface. The memory and storage may include, for example, cache, flash memory, read only memory (ROM), and/or random access memory (RAM)/DRAM. In some embodiments, memory may be specifically dedicated to storing firmware (e.g., for device applications such as an operating system, user interface functions, and processor

functions). In some embodiments, memory may be operative to store information related to other devices with which the electronics device 120 performs communications operations (e.g., saving contact information related to communications operations or storing information related to different media types and media items selected by the user).

[46] In one embodiment, the control circuitry 126 may be operative to perform the operations of one or more applications implemented on the electronics device 120. Any suitable number or type of applications may be implemented. Although the following discussion will enumerate different applications, it will be understood that some or all of the applications may be combined into one or more applications. For example, the electronics device 120 may include applications 1-N 127 including, but not limited to: an automatic speech recognition (ASR) application, optical character reader (OCR) application, a dialog application, a map application, a media application (e.g., QuickTime, MobileMusic.app, or MobileVideo.app), social networking applications (e.g., FACEBOOK®, INSTAGRAM®, TWITTER®, etc.), a calendaring application (e.g., a calendar for managing events, appointments, etc.), an Internet browsing application, etc. In some embodiments, the electronics device 120 may include one or multiple applications operative to perform communications operations. For example, the electronics device 120 may include a messaging application, an e-mail application, a voicemail application, an instant messaging application (e.g., for chatting), a video-conferencing application, a fax application, or any other suitable application for performing any suitable communications operation.

[47] In some embodiments, the electronics device 120 may include a microphone 122. For example, electronics device 120 may include microphone 122 to allow the user to transmit audio (e.g., voice audio) for speech control and navigation of applications 1-N 127, during a communications operation or as a means of establishing a communications operation or as an alternative to using a physical user interface. The microphone 122 may be incorporated in the electronics device 120, or may be remotely coupled to the electronics device 120. For example, the microphone 122 may be incorporated in wired headphones, the microphone 122 may be incorporated in a wireless headset, the microphone 122 may be incorporated in a remote control device, etc.

[48] In one embodiment, the camera module 128 comprises one or more camera devices that include functionality for capturing still and video images, editing functionality, communication interoperability for sending, sharing, etc. photos/videos, etc.

[49] In one embodiment, the electronics device 120 may include any other component suitable for performing a communications operation. For example, the electronics device 120 may include a power supply, ports, or interfaces for coupling to a host

device, a secondary input mechanism (e.g., an ON/OFF switch), or any other suitable component.

[50] In the present disclosure, "skill" refers to a specialized domain where the PA can understand/interpret/parse commands related to that domain. For example, in "picture gallery skill", the PA can understand commands like "show my latest trip photos." The skill may be also referred as a 'domain', a 'field', a 'category' and another term having equivalent technical meanings. In some embodiments, a skill may correspond to an application. An "intent" refers to the specific goal the user wants to accomplish within the domain or the skill. For example, in "picture gallery skill", the command "show my latest trip photos" refers to an intent "DISPLAY-PICTURE," whereas the command "send my latest trip photos to my mom" may correspond to a different intent "SHARE-PICTURE." The PA may determine the skill being used by special skill invocation command, for example "open Samsung shopping," which precedes the main command; or by string matching when a unique "skill invocation phrase/identifier" is mentioned within the command itself e.g. phrase "Samsung shopping" in utterance "Open Samsung shopping, and search Galaxy S8." In some cases, the "skill" can also be determined by a special parser called "domain classifier" trained using machine learning algorithms.

[51] FIG. 2 shows example vocabulary and paraphrase transfer learning, according to some embodiments. In some embodiments, different domains, for example smart home skill 210, music skill 220, picture gallery skills 230, shopping skill 240, restaurant skill 250 and airline skill 260 may be present for a PA (e.g., PA 129, FIG. 1). In one example, a user utters "set the TV at medium brightness." The PA (using a smart home skill domain 210) may not understand the utterance and responds with "sorry, I did not understand." The user then rephrases the utterance as "set the TV at 50 percent brightness." The PA then learns this new vocabulary that "medium" equates to "50 percent", so that the next time the PA directly understands the intent when the user reissues the same command "set the TV at medium brightness". In some embodiments, transfer learning 205 shares or transfers the learned paraphrase of fifty percent for medium from smart home skill domain 210 to the music skill domain 220. As a result, when the user utters "set volume to medium," the PA (using music skill domain 220) changes the volume setting to fifty percent and replies with "changing volume settings." That is, the word "medium" is interpreted as a meaning of the word "50 percent" or is replaced with the word "50 percent", in case of parsing the utterance "set volume to medium" in the music skill domain 220.

[52] In another example, the PA (using the picture gallery skills domain 230) may not understand an input of "send my trip photos to my mom," and responds with "sorry, I did not understand." The input is then rephrased as "send the trip pictures to my mom,"

which the PA may understand. In some embodiments, transfer learning 206 shares or transfers the learned paraphrase of pictures for photos from the picture gallery skills domain 230 to the shopping skill domain 240. When the user utters "show photos of the new Galaxy note 8," the PA (using the shopping skill domain 240) presents the photos based on the transferred learning 206 and replies "here you go." In still another example, the PA (using the restaurant skills domain 250) may not understand an input of "get some pizza back home by early afternoon," and responds with "sorry, I did not understand." The input is then rephrased as "get some pizza back home by 2PM," which the PA may understand. In some embodiments, transfer learning 207 shares or transfers the learned paraphrase of early afternoon for 2PM from the restaurant skills domain 250 to the airline skill domain 260.

[53] An out-of-the box PA can understand utterances within a fixed set of domains/skills called "default/general skills," added by the manufacturer. In addition, many new skills (typically developed by a third party) can be added to the PA on the fly by user after deployment. These third party skills are referred as "special skills." FIG. 3 shows a block diagram of processing for personalized skill intent parsing 300, according to some embodiments. It can be understood that operations described below can be performed by the electronic device 120 or at least one processor (e.g., the control circuitry 126, the PA 129, etc.) of the electronic device 120. The processing for personalized skill intent parsing 300 includes the following.

[54] Referring to FIG. 3, in block 310, a natural language input, for example, an utterance/input X' is received (e.g., using a PA, such as PA 129, FIG. 1). In block 320, it is determined whether the utterance X' corresponds to a "special skill" (e.g., by string matching a skill invocation phrase stored within the device or a cloud server). If it is determined that the utterance X' corresponds to a special skill (i.e., if the utterance X' includes at least one unique skill word associated with the special skill), the processing for personalized skill intent parsing 300 proceeds to block 335. In block 335, it is determined whether a skill for the utterance X' is already added to a user's personalized parser of the PA or not. That is, the electronic device 120 determines whether the personalized parser is available to determine an intent and perform an action corresponding to the skill.

[55] If it is determined that the skill corresponding to the utterance X' is already added to a user's personalized parser of the PA, the processing for personalized skill intent parsing 300 proceeds to block 360 (of the personal skill processing block 350) where the personalized skill parser is invoked. If it is determined that the skill corresponding to the utterance X' is not already added to a personalized parser of the PA, the processing for personalized skill intent parsing 300 proceeds to block 355 (of the personal skill processing block 350) where the new skill invocation keyword is added

using an update personal skill parser algorithm and then processing proceeds to block 360.

[56] In some embodiments, from block 360 the processing for personalized skill intent parsing 300 proceeds to block 365 where it is determined whether the skill intent is found or not from the personal skill parser 360. If it is determined that the skill intent is found from the personal skill parser 360, the processing for personalized skill intent parsing 300 proceeds to block 380 where an action corresponding to the determined skill intent is performed by an electronic device (e.g., electronic device 120, FIG. 1), and processing then stops at block 390. If it is determined that the skill intent is not found from the personal skill parser 360, the processing for personalized skill intent parsing 300 proceeds to block 370 where transfer vocabulary processing is performed invoking a learn personalized vocabulary algorithm to add a new user-specific vocabulary/training data to the personalized skill parser. Accordingly, the electronic device may define a mapping between the new user-specific vocabulary and a skill intent based on a subsequent natural language input. Then the processing stops at block 390.

[57] In some embodiments, if it is determined that the utterance X' does not include a special skill invocation phrase, the processing for personalized skill intent parsing 300 proceeds to block 325 where an intent (general skill that is not associated with any individual user) parser is invoked. Processing for personalized skill intent parsing 300 then proceeds to block 330 where it is determined whether the intent is found or not. If the intent is found, the processing for personalized skill intent parsing 300 proceeds to block 340 where a corresponding action is performed by an electronic device (e.g., electronic device 120, FIG. 1). If the intent is not found in block 330, the processing for personalized skill intent parsing 300 stops at block 390.

[58] Each "skill" in a PA has an associated training data set, which we refer as "skill training data." FIG. 4 shows a block diagram of processing for updating a personalized skill parser block 355, according to some embodiments. It can be understood that operations described below can be performed by the electronic device 120 or at least one processor (e.g., the control circuitry 126, the PA 129, etc.) of the electronic device 120. In block 410, a new custom skill (the special skill determined in 320) is input and added in block 420 to the set of skills S being used by the user. In block 430, the processing for updating a personalized skill parser constructs a "combined skill training set" T by merging the "skill training set" for every skill in set S (merging simply refers to constructing a combined training data set containing utterances from skill training set of every skill in set S). In block 440, "user-specific training data" U (which were added by the user using the learn personalized vocabulary algorithm) is constructed. For example, in FIG. 2, when the user provides a paraphrased utterance

"set the TV at 50 percent brightness" 205, this utterance is added to the "user-specific training data" U. In block 450 a "personalized skill parser" is trained using the "combined skill training data" T and the "user-specific training data" U.

[59] FIG. 5A shows a block diagram for transferring personalized vocabulary processing block 370, according to some embodiments. It can be understood that operations described below can be performed by the electronic device 120 or at least one processor (e.g., the control circuitry 126, the PA 129, etc.) of the electronic device 120. In block 510, the intent parse of utterance X failed in the personal skill parser 360 (FIG. 3) (i.e., the personal skill parser 360 was unable to parse the utterance X). In block 520, it is determined whether to learn a new vocabulary or not (e.g., a user is prompted/queried if he/she wants the PA to pick up new personal vocabulary). If it is determined that a new vocabulary is not desired to be learned, then the transferring personalized vocabulary processing block 370 proceeds to block 570 and stops (e.g., a user replies "no" to the prompt/query). If it is determined that a new vocabulary is desired to be learned (e.g., a user replies "yes" to the prompt/query), the transferring personalized vocabulary processing block 370 proceeds to block 530 where a paraphrased utterance Y of X is received. In FIG. 2, for example, transfer learning 205 shows that the original utterance X is "set the TV at medium brightness," and the paraphrased utterance Y provided by the user is "set the TV at fifty percent brightness." A user is prompted to provide paraphrased utterance Y to any identified out-of-vocabulary words (e.g., an unknown word) in X. For example, the word "medium" is highlighted as out-of-vocabulary when a user is prompted to input a paraphrase. In block 535, processing checks if Y can be parsed correctly with the personal skill parser 360 to an intent I(Y). If not, then processing declares a failure and proceeds to block 570 and stops. Otherwise, processing proceeds to the personal skill parser retraining processing 540 to block 550 and adds X along with the intent I(Y) to the "user-specific training data" U. The personal skill parser 360 is retrained in block 560 using both "combined skill training data" T and "user-specific training data" U. The transferring personalized vocabulary processing in block 560 proceeds to transfer learning of the new personalized vocabulary across all personalized skills S being used by the user. That is, the new personalized vocabulary is applied to the all personalized skills S. In an embodiment, training using the new personalized vocabulary may be performed on the all personalized skills S. In another embodiment, information on the new personalized vocabulary is added to the all personalized skills S. Processing then proceeds to block 570 and stops.

[60] FIG. 5B shows another block diagram for transferring personalized vocabulary processing block 370, according to some embodiments. It can be understood that operations described below can be performed by the electronic device 120 or at least one

processor (e.g., the control circuitry 126, the PA 129, etc.) of the electronic device 120. In block 545 processing determines the skill s that corresponds to the utterance Y . In block 546, an extended set of paraphrases $P(X)$ of X within the skill s is generated, also containing the user provided paraphrase Y . In block 551, for each paraphrase Z in $P(X)$ processing adds $\{Z, \text{intent}(Y)\}$ to a user-specific training set U . The transferring personalized vocabulary processing block 370 proceeds to block 560 where the personal skill parser is retrained, and learning of the new personalized vocabulary is transferred across all personalized skills S being used by the user. Processing then proceeds to block 570 and stops.

[61] FIG. 6 shows a block diagram for segmentation rephrasing and swapping processing 600, according to some embodiments. Some embodiments generate structural variational paraphrases, meaning that the generated paraphrases have different variability in terms of the structure of details in the input utterance. A natural language utterance 610 is obtained/received (e.g., by a PA, such as PA 129, FIG. 1) and input into block 620 for utterance segmentation. In block 620, utterance segmentation generates the segments of the input utterance 610 and identifies the dependence of every segment in the utterance. The dependency here stands for a relation between a leading segment and a dependent segment, where the leading segment of a constituent was the central organizing word of a larger constituent (e.g., the primary noun in a noun phrase, or verb in a verb phrase) and the remaining words in the constituent are either direct, or indirect, dependents of their leading segment. In block 630 segments rephrasing and swapping includes generating the rephrasing of each segment and swapping the segments based on their dependencies. In block 640 generated paraphrase ranking includes ranking of the rephrased and swapped paraphrases based on a language model. The results of block 640 are output in block 650 as the ranked paraphrases. These ranked paraphrases can then be added to user-specific training set as shown in block 551 during transferring personalized vocabulary processing.

[62] FIG. 7 shows a block diagram for utterance segmentation processing of block 620 (FIG. 6), according to some embodiments. In some embodiments, the utterance segmentation processing of block 620 has the following inputs: a user utterance 610 and a probabilistic context-free grammar (PCFG) parser 710. In one embodiment, the PCFG parser 710 works out the grammatical structure of sentences, including which groups of words go together as segments and the dependency between segments. The utterance segmentation processing of block 620 generates the parser tree of the input utterance 610 using the PCFG parser 710 and identifies the dependency relations between every part of the utterance 610. For example, it identifies the verb 720, the object argument 730 and the adjuncts 740. From the identified verb 720 and the object argument 730, the utterance 610 is segmented to generate the leading segment

(including verb phrase) 750 of the input utterance 610; and the adjuncts 740 are segmented to generate the paraphrases 760 (paraphrase 1-N) of the input utterance 610.

[63] FIG. 8 shows an example of segmentation of an utterance 800, according to some embodiments. In the example, there are three segments including the leading segment 810 ("deliver some pizza") and two adjuncts 820 ("to my home" 840 and "by noon" 850). The utterance 800 includes a verb 835 of "deliver" 830 and object argument 836 of "some pizza." The utterance 800 is used to describe how some embodiments provide use of segmentation and rephrasing as shown in FIGS. 10 and 11.

[64] FIG. 9 shows a block diagram of processing for block 630 (FIG. 6) segmentation rephrasing and swapping, according to some embodiments. In some embodiments, the segmentation and rephrasing and swapping processing includes block 910 segment concatenation and block 920 generation of paraphrases for concatenated segments. For block 910 segment concatenation, the segments of an input utterance are input, including one or more verb phrase and adjuncts. Block 910 segment concatenation generates the rephrase forms, the non-leading segments, by concatenating each of verb phrase with one of the adjuncts. For block 920 generation of paraphrases for concatenated segments, some embodiments may use any paraphrase generation algorithm to rephrase the non-leading segments, and use any paraphrase generation algorithm to rephrase the leading segment. For adjunct swapping, some embodiments permute the segments generated in block 910 to generate the list of swapped paraphrases. If a set of segments are dependent on the same segment, they are swapped between each other to generate the list of swapped paraphrases.

[65] FIG. 10 shows an example of rephrased segments, according to some embodiments. For this example, the input utterance 1010 is "deliver some pizza to my home by noon." The example non-leading segments 1020 are: "deliver some pizza to my home," and "deliver some pizza by noon," where the underlined words refer to the non-leading segments. The leading segment 1030 is: "deliver some pizza." The example result of rephrasing processing for the non-leading segments 1025 results in the following rephrases: "deliver some pizza for home," "deliver some pizza back home," "deliver some pizza to bring home...," "deliver some pizza before noon," "deliver some pizza at noon," and "deliver some pizza around noon...," where the underlined portions are the rephrased non-leading segment. For the rephrasing of the leading segment 1030, the example rephrased leading segments 1035 are: "get me some pizza," "buy a pizza," and "I want a piece of pizza..." The rephrasing of a segment is generated by any existing paraphrasing algorithm.

[66] FIG. 11 shows an example of rephrased segments 1110 and combined paraphrases from rephrased segments 1120, according to some embodiments. The rephrased segments are used to substitute the original segment in the utterance to generate a list

of paraphrases. In this example, the rephrased segments 1110 are: "get me some pizza," "buy a pizza," and "I want some pizza...," "for home," "back home," and "to bring home;" and "before noon," "at noon," and "around noon...". The combined paraphrases from rephrased segments 1120 are: "get me some pizza back home by noon," "by noon get me some pizza and home," "I want a pizza before noon back home," "before noon I want a pizza back home," "to home get me some pizza via noon," "buy a pizza for home before noon," "for home buy some pizza before noon," "around noon buy a pizza to home," "buy a pizza by noon back home," and "get me some pizza by noon to home..."

[67] FIG. 12 shows a block diagram for swapped paraphrase ranking for block 640 (FIG. 6), according to some embodiments. In some embodiments, block 640 includes combined paraphrase ranking that takes the generated swapped and rephrased paraphrases 1220 and a language model 1210 as input. The language model 1210 is used to score each generated swapped and rephrased paraphrases based on the likelihood from the language model 1210. In some embodiments, in block 1230 the processing extracts each tuple of segments in the dictionary of rephrased segments, uses the language model 1210 to score each of the combination of permuted tuples, selects the lowest score among all combinations, marks this score as the score of the selected tuple, and derives the overall ranking of all swapped paraphrases (ranked paraphrases 650) by ranking the score on each combined tuple.

[68] FIG. 13 shows an example of ranked paraphrases, according to some embodiments. In the example 1310 of ranked paraphrases of "deliver some pizza to my home by noon," the high ranked paraphrases 1320 (from the highest) include: "get me some pizza back home by noon," "buy a pizza for home before noon," "I want a pizza before noon back home," "before noon I want a pizza back home," "around noon buy a pizza to home," and "get me some pizza by noon to home." The low ranked paraphrases 1330 include (from higher to lower): "by noon get me some pizza and home," "for home buy some pizza before noon," "to home get me some pizza via noon," and "buy a pizza by noon back home."

[69] FIGS. 14A and 14B show an example of rephrased and swapped utterances, and an output of ranked paraphrases, according to some embodiments. For the example input utterance 1410 of "deliver some pizza to my home by noon," the action term 1420 is "deliver," the three segments 1430 are: "deliver some pizza," "to my home," and "by noon," and the leading segment 1440 is "deliver some pizza." In some example embodiments, the first replacement of phrases 1450, which is the rephrasing of the non-leading segments, includes replacement of "to my home" with "for home," "back home," and "to bring home;" and replacement of "by noon" with "before noon," "at noon," and "around noon."

- [70] The second replacement phrase of the action term 1460 "deliver some pizza" includes rephrasing the leading segment with "get me some pizza," "buy a pizza," "I want some pizza," etc. In some embodiments, the example rephrased and swapped utterances 1470 (by combining first and second replacements) include: "get me some pizza back home by noon," "by noon get me some pizza and home," "I want a pizza before noon back home," "before noon I want a pizza back home," "to home get me some pizza via noon," "buy a pizza for home before noon," "for home buy some pizza before noon," "around noon buy a pizza to home," "buy a pizza by noon back home," and "get me some pizza by noon to home," etc.
- [71] The final output of ranked paraphrases 1480 may include high ranked results 1490 of "get me some pizza back home by noon," "buy a pizza for home before noon," "I want a pizza before noon back home," "before noon I want a pizza back home," "around noon buy a pizza to home," and "get me some pizza by noon to home." The final output of ranked paraphrases 1480 may include high ranked results 1495 of "by noon get me some pizza and home," "for home buy some pizza before noon," "to home get me some pizza via noon," and "buy a pizza by noon back home."
- [72] FIG. 15 shows a block diagram of a process 1500 for generating structural variational paraphrases, according to some embodiments. In block 1510, process 1500 determines, by an electronic device (e.g., electronic device 120, FIG. 1), a skill based on a keyword from a first NL input. In block 1520, upon successful determination of the skill, process 1500 transmits the first NL input to a custom skill parser (e.g., personal skill parser 360, FIG. 3) for determination of a skill intent. In some embodiments, in process 1500 the custom skill parser is trained based on a custom training data set and a combination of multiple skill training data sets, and each skill training data set of the multiple skill training data sets includes data specific to a distinctive skill of a set of skills. In block 1530, upon unsuccessful determination of the skill, process 1500 transmits the first NL input to a generic parser (e.g., Intent Parser 325, FIG. 3) for determination of a general intent of the first NL input. In block 1540, upon successful determination of the skill intent, process 1500 adds the skill to the set of skills. In block 1550, upon unsuccessful determination of the skill intent, process 1500 prompts for a second NL input. In process 1500, the second NL input is a paraphrase of the first NL input.
- [73] In some embodiments, process 1500 may include transmitting the second NL input to the custom skill parser for determination of the skill intent, updating the custom training data set based on the first NL input and the skill intent, and training the custom skill parser based on both the combination of the multiple custom skill training data sets and the custom training data set.
- [74] In some embodiments, process 1500 may include generating segments from the first

NL input. In process 1500, the segments may include a leading segment and one or more non-leading segments. Process 1500 may further include generating multiple rephrases from the segments, and swapping the segments based on dependencies of the segments to generate swapped rephrases. In some embodiments, process 1500 may include ranking the multiple rephrases and the swapped rephrases based on an input language model.

- [75] In some embodiments, in process 1500 generating the segments may include generating a parsing tree of the first NL input and a parser, identifying dependency relations between portions of the first NL input, identifying details of the first NL input, where the details include at least one verb, an object argument and one or more adjuncts, generating the leading segment of the first NL input based on the at least one verb and the object argument, and generating the multiple paraphrases based on the one or more adjuncts.
- [76] FIG. 16 is a high-level block diagram showing an information processing system comprising a computing system implementing one or more embodiments. The system 1600 includes one or more processors 1611 (e.g., ASIC, CPU, etc.), and may further include an electronic display device 1612 (for displaying graphics, text, and other data), a main memory 1613 (e.g., random access memory (RAM), cache devices, etc.), storage device 1614 (e.g., hard disk drive), removable storage device 1615 (e.g., removable storage drive, removable memory, a magnetic tape drive, optical disk drive, computer-readable medium having stored therein computer software and/or data), user interface device 1616 (e.g., keyboard, touch screen, keypad, pointing device), and a communication interface 1617 (e.g., modem, wireless transceiver (such as Wi-Fi, Cellular), a network interface (such as an Ethernet card), a communications port, or a PCMCIA slot and card).
- [77] The communication interface 1617 allows software and data to be transferred between the computer system and external devices through the Internet 1650, mobile electronic device 1651, a server 1652, a network 1653, etc. The system 1600 further includes a communications infrastructure 1618 (e.g., a communications bus, cross bar, or network) to which the aforementioned devices 1617 through 1617 are connected.
- [78] The information transferred via communications interface 1617 may be in the form of signals such as electronic, electromagnetic, optical, or other signals capable of being received by communications interface 1617, via a communication link that carries signals and may be implemented using wire or cable, fiber optics, a phone line, a cellular phone link, a radio frequency (RF) link, and/or other communication channels.
- [79] In one implementation of one or more embodiments in a mobile wireless device (e.g., a mobile phone, tablet, wearable device, etc.), the system 1600 further includes an image capture device 1620, such as a camera 128 (FIG. 1), and an audio capture device

1619, such as a microphone 122 (FIG. 1). The system 1600 may further include application processing or processors as MMS 1621, SMS 1622, email 1623, social network interface (SNI) 1624, audio/video (AV) player 1625, web browser 1626, image capture 1627, etc.

[80] In one embodiment, the system 1600 includes personal intent learning processing 1630 that may implement processing similar as described regarding personal intent learning app 129 (FIG. 1), and for processing described above in relation to FIGS. 5-9. In one embodiment, the personal intent learning processing 1630 along with an operating system 1629 may be implemented as executable code residing in memory of the system 1600. In another embodiment, the personal intent learning processing 1630 may be provided in hardware, firmware, etc.

[81] In one embodiment, the main memory 1613, storage device 1614 and removable storage device 1615, each by themselves or in any combination, may store instructions for the embodiments described above that may be executed by the one or more processors 1611.

[82] As is known to those skilled in the art, the aforementioned example architectures described above, according to said architectures, can be implemented in many ways, such as program instructions for execution by a processor, as software modules, microcode, as computer program product on computer readable media, as analog/logic circuits, as application specific integrated circuits, as firmware, as consumer electronic devices, AV devices, wireless/wired transmitters, wireless/wired receivers, networks, multi-media devices, etc. Further, embodiments of said Architecture can take the form of an entirely hardware embodiment, an entirely software embodiment or an embodiment containing both hardware and software elements.

[83] One or more embodiments have been described with reference to flowchart illustrations and/or block diagrams of methods, apparatus (systems) and computer program products according to one or more embodiments. Each block of such illustrations/diagrams, or combinations thereof, can be implemented by computer program instructions. The computer program instructions when provided to a processor produce a machine, such that the instructions, which execute via the processor create means for implementing the functions/operations specified in the flowchart and/or block diagram. Each block in the flowchart/block diagrams may represent a hardware and/or software module or logic, implementing one or more embodiments. In alternative implementations, the functions noted in the blocks may occur out of the order noted in the figures, concurrently, etc.

[84] The terms "computer program medium," "computer usable medium," "computer readable medium", and "computer program product," are used to generally refer to media such as main memory, secondary memory, removable storage drive, a hard disk

installed in hard disk drive. These computer program products are means for providing software to the computer system. The computer readable medium allows the computer system to read data, instructions, messages or message packets, and other computer readable information from the computer readable medium. The computer readable medium, for example, may include non-volatile memory, such as a floppy disk, ROM, flash memory, disk drive memory, a CD-ROM, and other permanent storage. It is useful, for example, for transporting information, such as data and computer instructions, between computer systems. Computer program instructions may be stored in a computer readable medium that can direct a computer, other programmable data processing apparatus, or other devices to function in a particular manner, such that the instructions stored in the computer readable medium produce an article of manufacture including instructions which implement the function/act specified in the flowchart and/or block diagram block or blocks.

[85] Computer program instructions representing the block diagram and/or flowcharts herein may be loaded onto a computer, programmable data processing apparatus, or processing devices to cause a series of operations performed thereon to produce a computer implemented process. Computer programs (i.e., computer control logic) are stored in main memory and/or secondary memory. Computer programs may also be received via a communications interface. Such computer programs, when executed, enable the computer system to perform the features of the embodiments as discussed herein. In particular, the computer programs, when executed, enable the processor and/or multi-core processor to perform the features of the computer system. Such computer programs represent controllers of the computer system. A computer program product comprises a tangible storage medium readable by a computer system and storing instructions for execution by the computer system for performing a method of one or more embodiments.

[86] Though the embodiments have been described with reference to certain versions thereof; however, other versions are possible. Therefore, the spirit and scope of the appended claims should not be limited to the description of the preferred versions contained herein.

Claims

- [Claim 1] A method for operating an electronic device, the method comprising:
detecting a first natural language (NL) input;
determining a skill from the first NL input;
if the NL input includes at least one unique skill word, transmitting the first NL input to a custom skill parser for determination of a skill intent, wherein the custom skill parser is trained based on data including at least a custom training data set; and
if the NL input does not includes the at least one unique skill word, transmitting the first NL input to a generic parser for determination of a general intent of the first NL input.
- [Claim 2] The method of claim 1, wherein the data includes a combination of a plurality of skill training data sets, and
wherein each skill training data set of the plurality of skill training data sets includes data specific to a distinctive skill of a set of skills.
- [Claim 3] The method of claim 2, further comprising:
if the NL input includes the at least one unique skill word, adding the skill to the set of skills.
- [Claim 4] The method of claim 2, further comprising:
if the NL input does not includes the at least one unique skill word, prompting for a second NL input, wherein the second NL input is a paraphrase of the first NL input.
- [Claim 5] The method of claim 4, further comprising:
transmitting the second NL input to the custom skill parser for determination of the skill intent;
updating the custom training data set based on the first NL input and the skill intent; and
training the custom skill parser based on both the combination of the plurality of custom skill training data sets and the custom training data set.
- [Claim 6] The method of claim 4, further comprising:
generating segments from the first NL input, wherein the segments comprise a leading segment and one or more non-leading segments;
generating a plurality of rephrases from the segments; and
swapping the segments based on dependencies of the segments to generate swapped rephrases.
- [Claim 7] The method of claim 6, further comprising:

- ranking the plurality of rephrases and the swapped rephrases based on an input language model.
- [Claim 8] The method of claim 6, wherein generating the segments comprises: generating a parsing tree using the first NL input and a parser; identifying dependency relations between portions of the first NL input; identifying details of the first NL input, wherein the details comprise at least one verb, an object argument and one or more adjuncts; generating the leading segment based on the at least one verb and the object argument; and generating the plurality of rephrases based on the one or more adjuncts.
- [Claim 9] A method for intelligent assistance, comprising: generating one or more segments for a natural language (NL) input; identifying a leading segment from the one or more segments; generating a first replacement phrase based on concatenation of the leading segment with one non-leading segment of the one or more segments; and providing an output comprising at least the first replacement phrase.
- [Claim 10] The method of claim 9, further comprising: swapping the one or more segments based on dependencies of the one or more segments to generate swapped rephrases.
- [Claim 11] The method of claim 9, further comprising: generating a second replacement phrase of an action term based on semantic of the NL input, wherein the semantic of the NL input is based on a language model.
- [Claim 12] The method of claim 9, wherein generating the one or more segments comprises: generating a parsing tree using the NL input and a parser; identifying dependency relations between portions of the NL input; and identifying details of the NL input, wherein the details comprise at least one verb, an object argument and one or more adjuncts.
- [Claim 13] A method for operating an electronic device, the method comprising: detecting a first utterance including a first word that is unable to be parsed based on a first skill; detecting a second utterance including a second word instead of the first word; identifying an intent corresponding to the second utterance by parsing the second utterance based on the first skill; adding the first utterance along with the intent into training data as-

sociated with the first skill; and
training the first skill by using the training data.

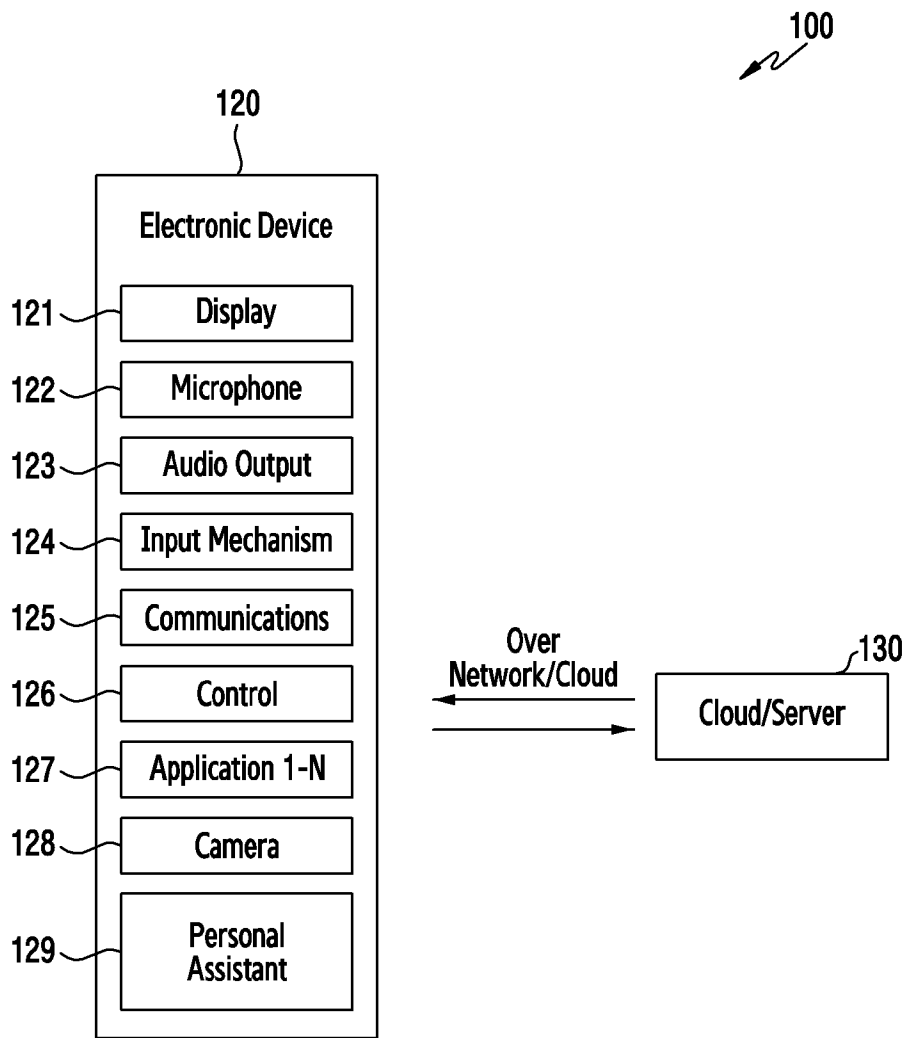
[Claim 14]

The method of claim 13, further comprising:
transferring learning of the first word to a second skill
detecting a third utterance including the first word; and
identifying an intent corresponding to the third utterance by parsing the
third utterance based on the second skill; and
performing at least one operation according to the intent corresponding
to the third utterance.

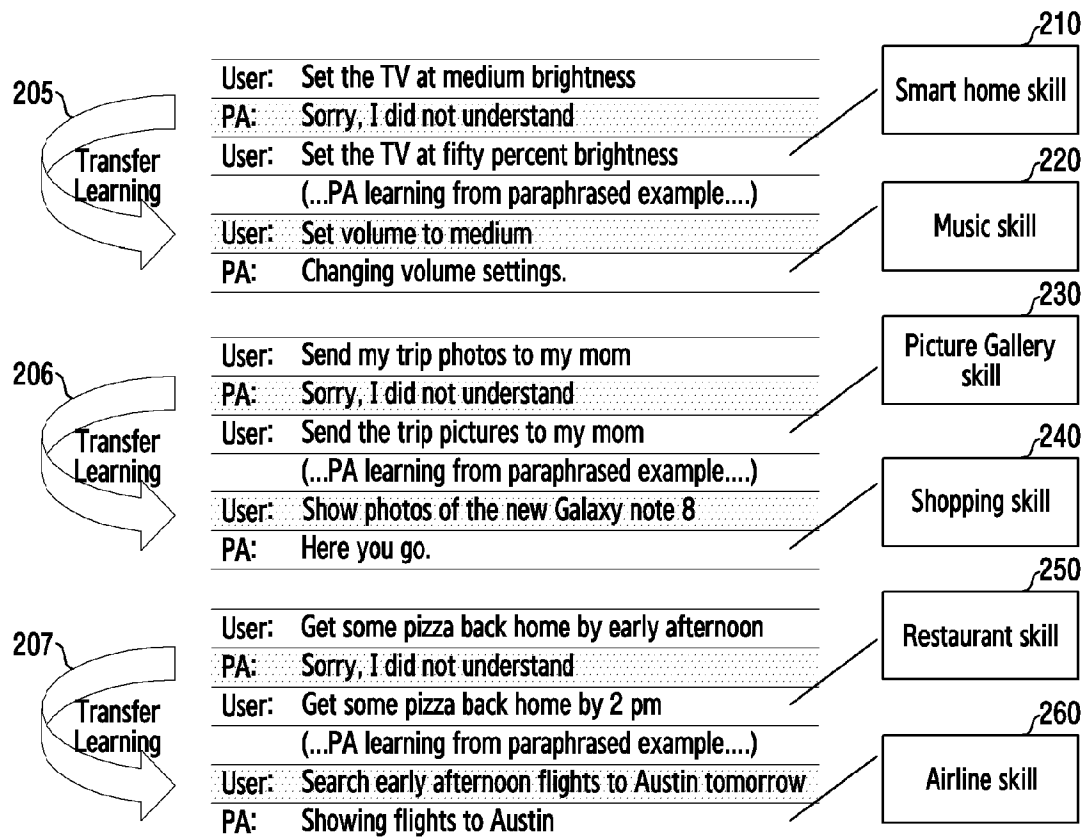
[Claim 15]

The method of claim 13, wherein the parsing the third utterance
comprises,
interpreting the first word in the third utterance as the meaning of the
second word, or
replacing the first word in the third utterance with the second word.

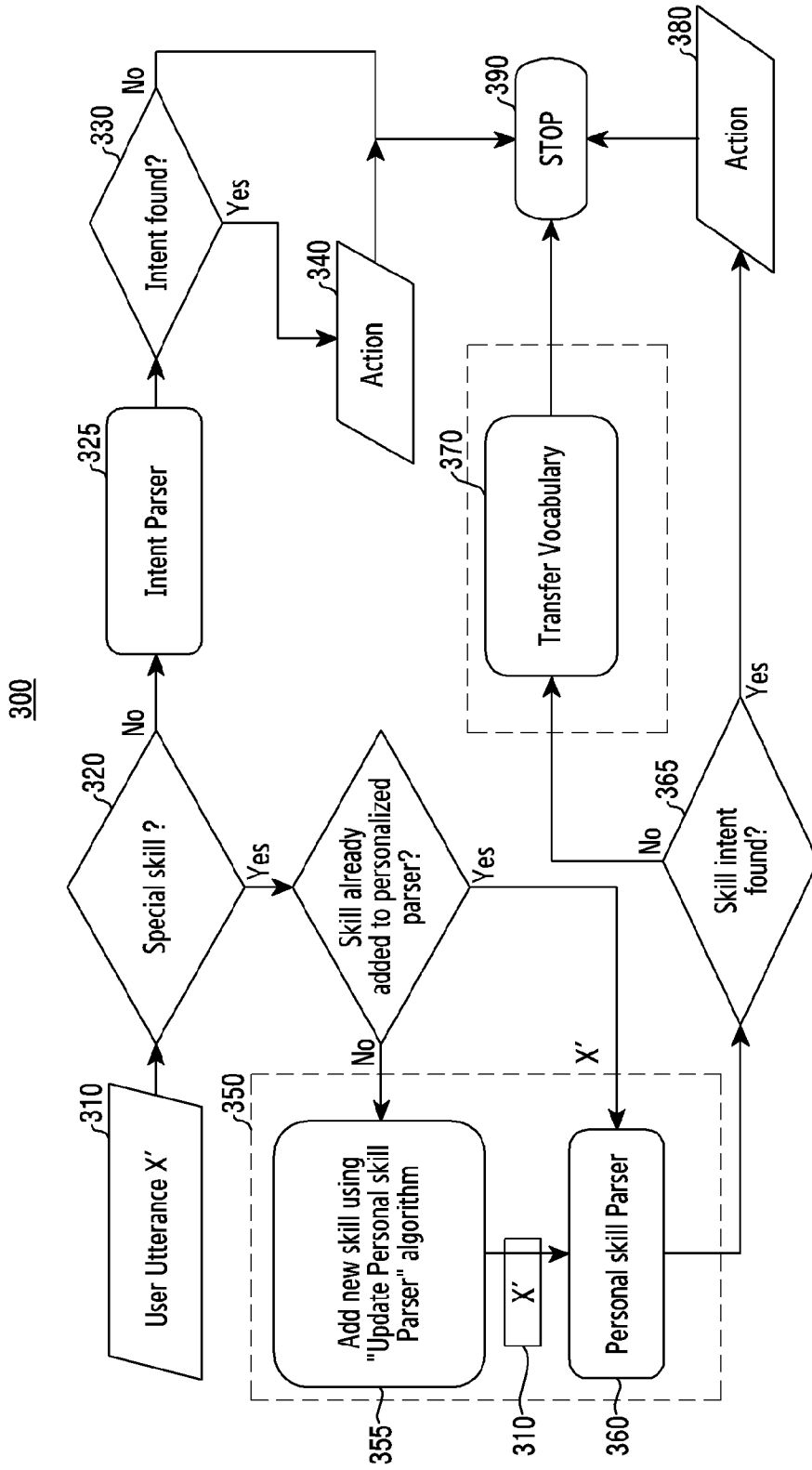
[Fig. 1]



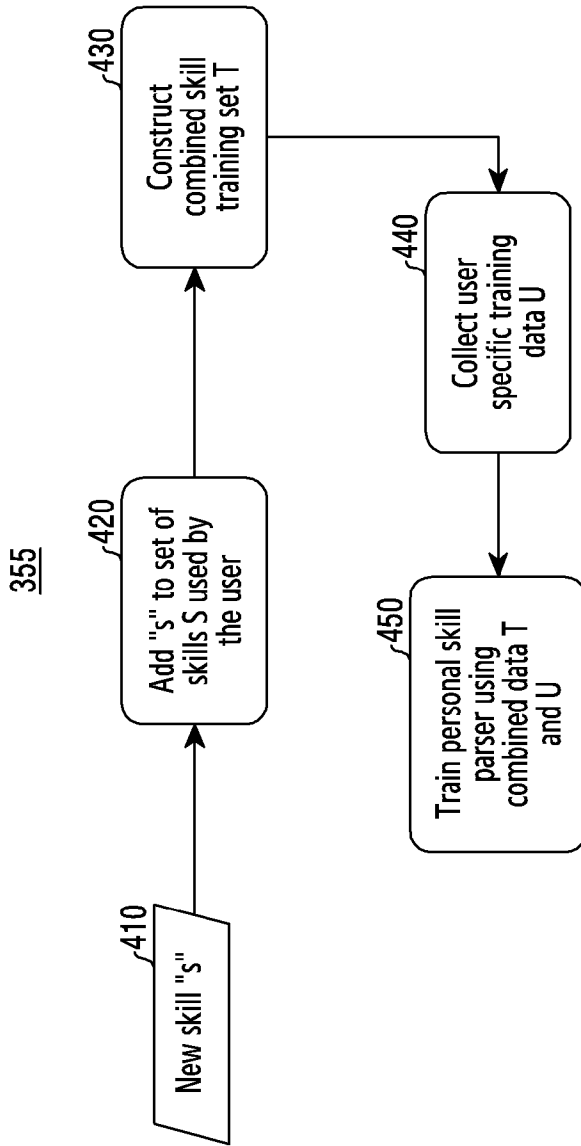
[Fig. 2]



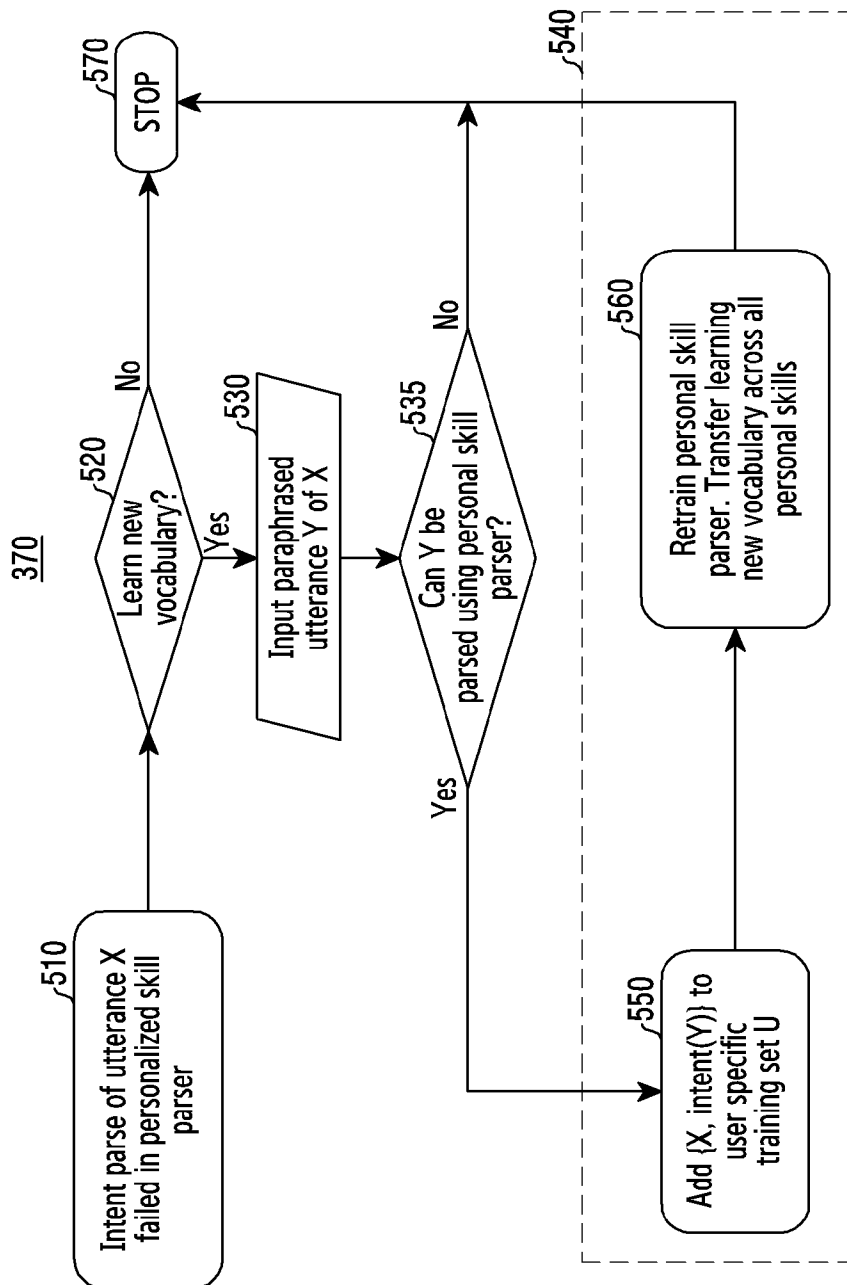
[Fig. 3]



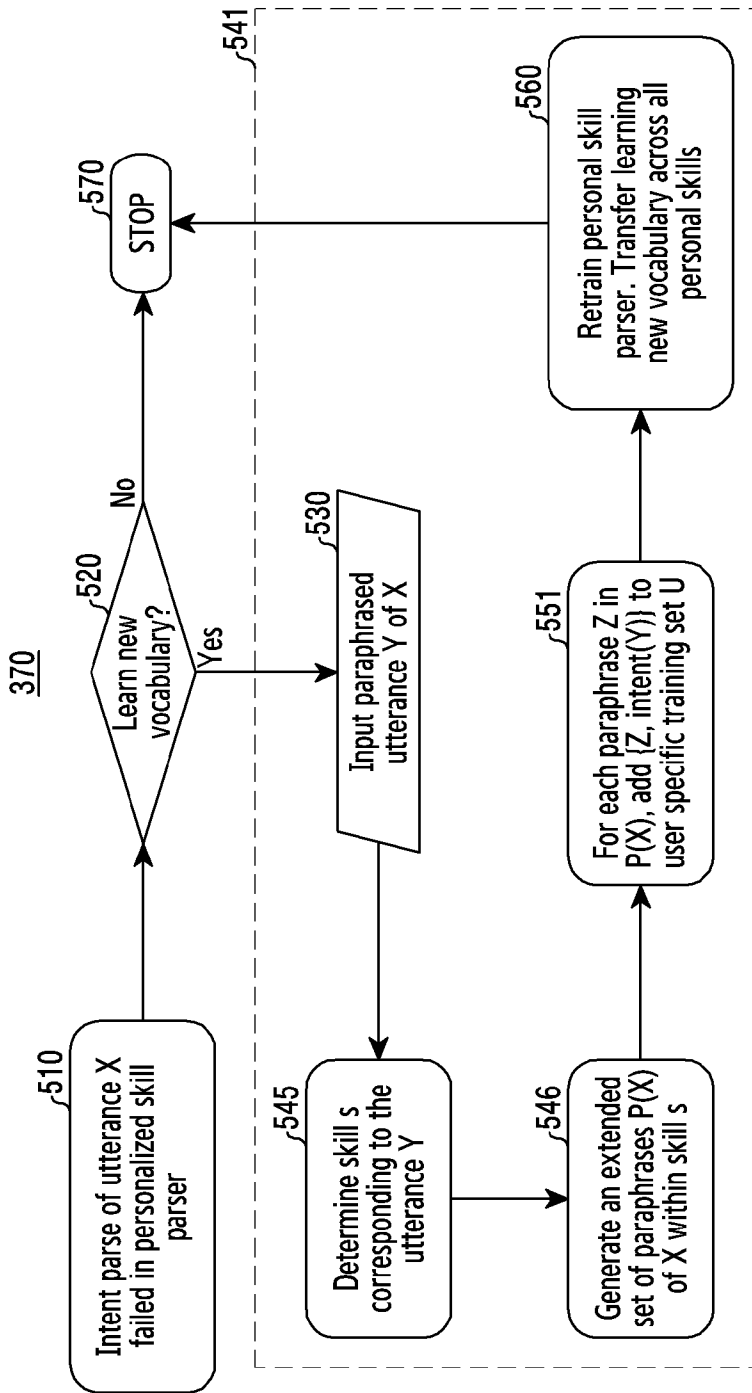
[Fig. 4]



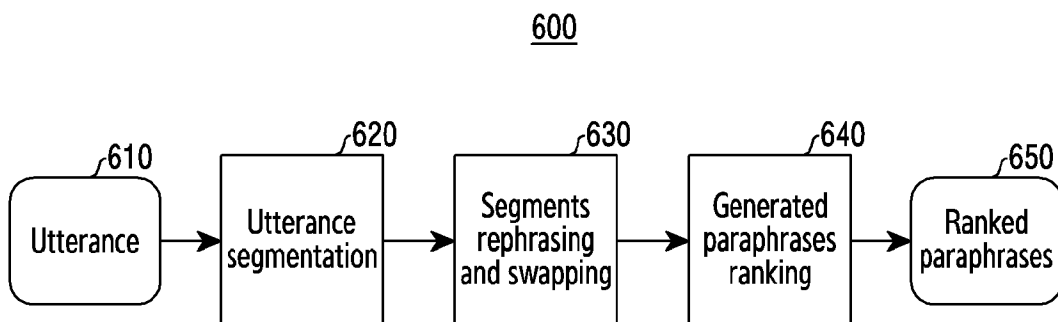
[Fig. 5A]



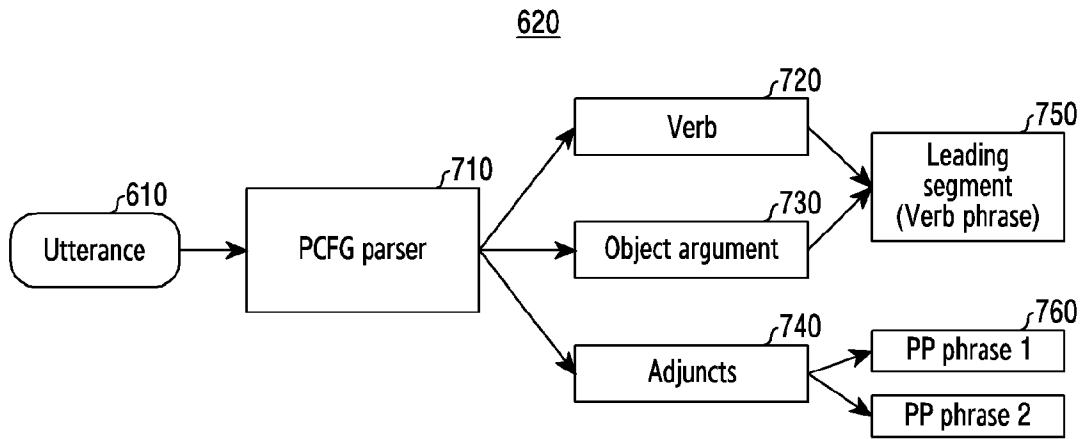
[Fig. 5B]



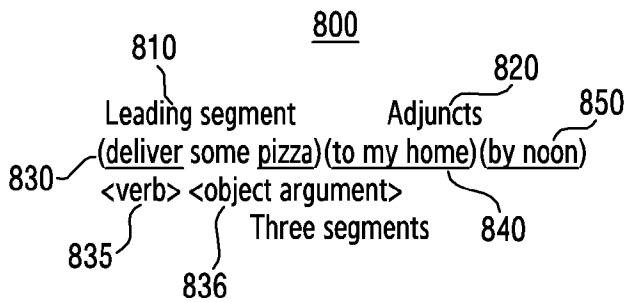
[Fig. 6]



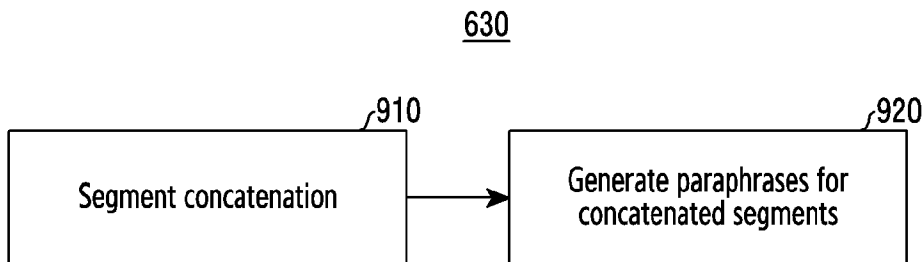
[Fig. 7]



[Fig. 8]



[Fig. 9]



[Fig. 10]

Input: deliver some pizza to my home by noon ~ 1010

Non-leading segments ~ 1020

deliver some pizza to my home

deliver some pizza by noon

Rephrase non-leading segment ~ 1025

deliver some pizza for home
 deliver some pizza back home
 deliver some pizza to bring home ...

deliver some pizza before noon
 deliver some pizza at noon
 deliver some pizza around noon ...

Leading segment ~ 1030

deliver some pizza

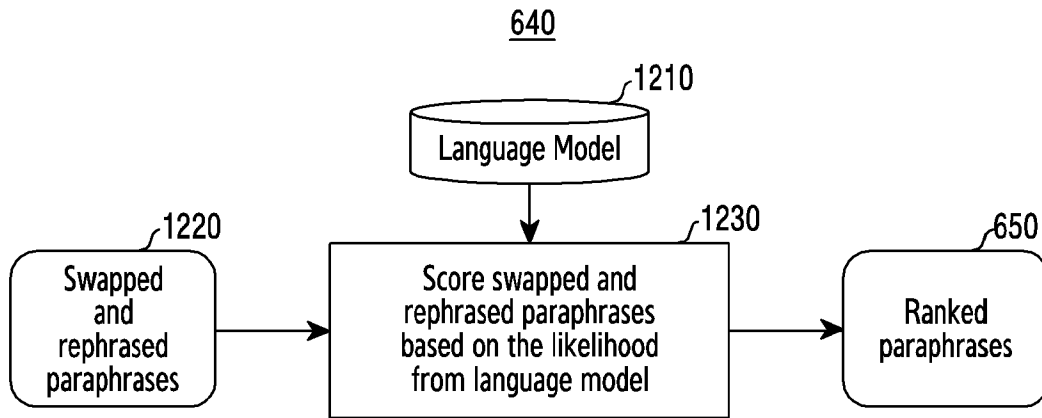
Rephrase leading segment ~ 1035

get me some pizza
 buy a pizza
 i want a piece of pizza ...

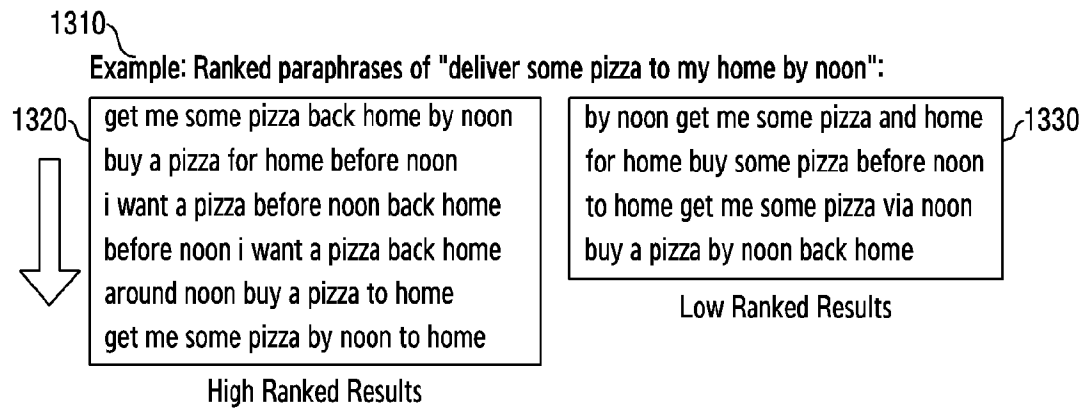
[Fig. 11]

| | |
|---|--|
| <p>Rephrased segment ~ 1110</p> <p>get me some pizza buy a pizza i want some pizza ...</p> <p>for home back home to bring home ...</p> <p>before noon at noon around noon ...</p> | <p>Combined paraphrases from rephrased segments ~ 1120</p> <p>get me some pizza back home by noon by noon get me some pizza and home i want a pizza before noon back home before noon i want a pizza back home to home get me some pizza via noon buy a pizza for home before noon for home buy some pizza before noon around noon buy a pizza to home buy a pizza by noon back home get me some pizza by noon to home</p> |
|---|--|

[Fig. 12]



[Fig. 13]



[Fig. 14A]

1410 ~ Input: deliver some pizza to my home by noon

1420 ~ Action term : deliver

1430 ~ Three segments: deliver some pizza; to my home; by noon
 ("pizza", "to my home", "by noon" depend on "deliver".)

1440 ~ Leading segment: deliver some pizza

1450 ~ First replacement of phrases (rephrase non-leading segments):

| | |
|----------------------------------|--------------------------------|
| deliver some pizza to my home | deliver some pizza by noon |
| deliver some pizza for home | deliver some pizza before noon |
| deliver some pizza back home | deliver some pizza at noon |
| deliver some pizza to bring home | deliver some pizza around noon |
| | |

1460 ~ Second replacement phrase of the action term:
 (rephrase the leading segment)

| |
|--------------------|
| deliver some pizza |
| get me some pizza |
| buy a pizza |
| i want some pizza |
| |

[Fig. 14B]

1410 ~ Input: deliver some pizza to my home by noon

1470 ~ Rephrased and swapped utterances (by combining first and second replacement):

| | |
|--------------------------------------|---|
| get me some pizza back home by noon | buy a pizza for home before noon |
| by noon get me some pizza and home | for home buy some pizza before noon |
| i want a pizza before noon back home | around noon buy a pizza to home |
| before noon i want a pizza back home | buy a pizza by noon back home |
| to home get me some pizza via noon | get me some pizza by noon to home |

1480 ~ Final output of ranked paraphrase:

1490 ~

get me some pizza back home by noon
 buy a pizza for home before noon
 i want a pizza before noon back home
 before noon i want a pizza back home
 around noon buy a pizza to home
 get me some pizza by noon to home

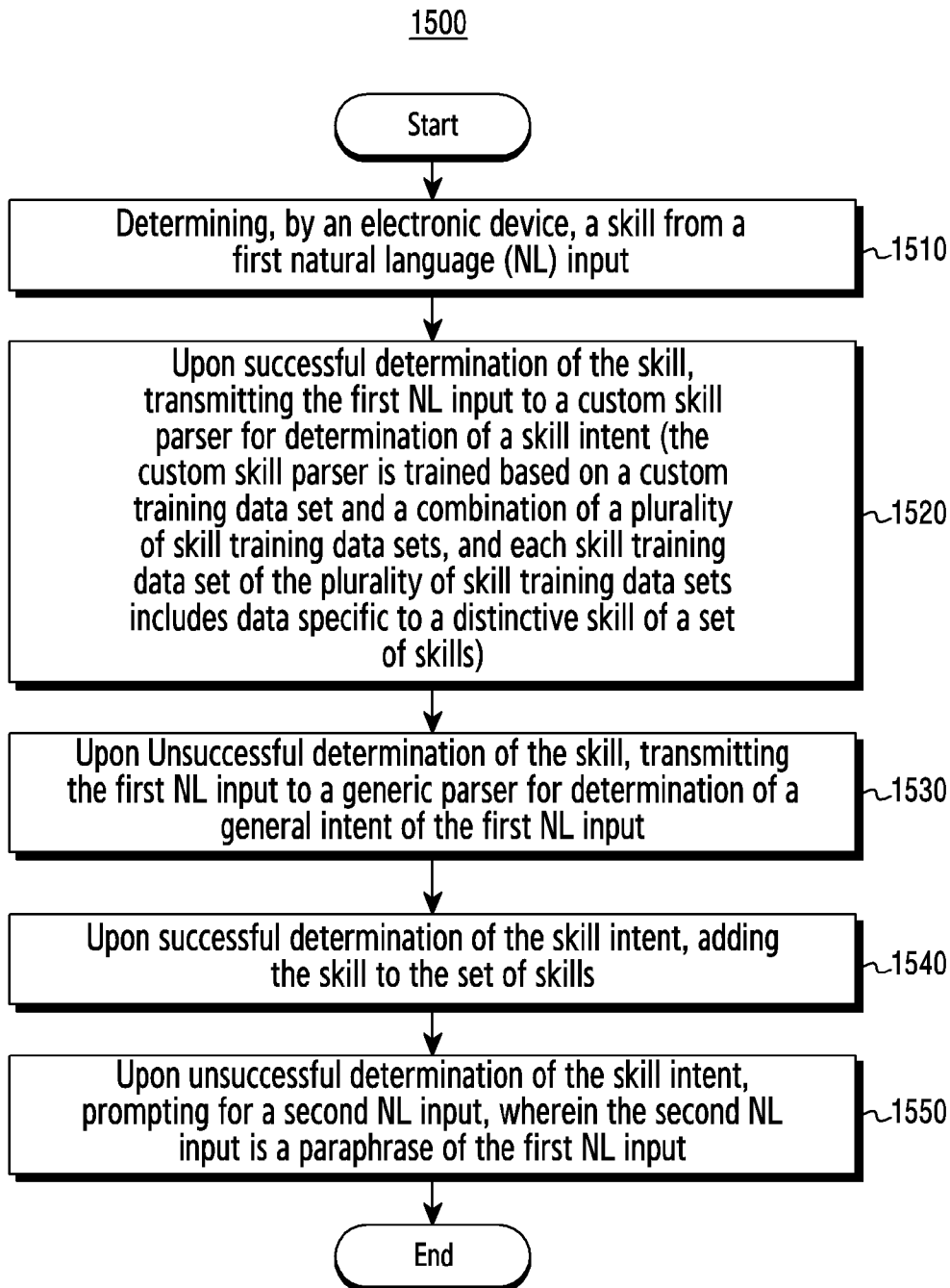
High Ranked Results

1495 ~

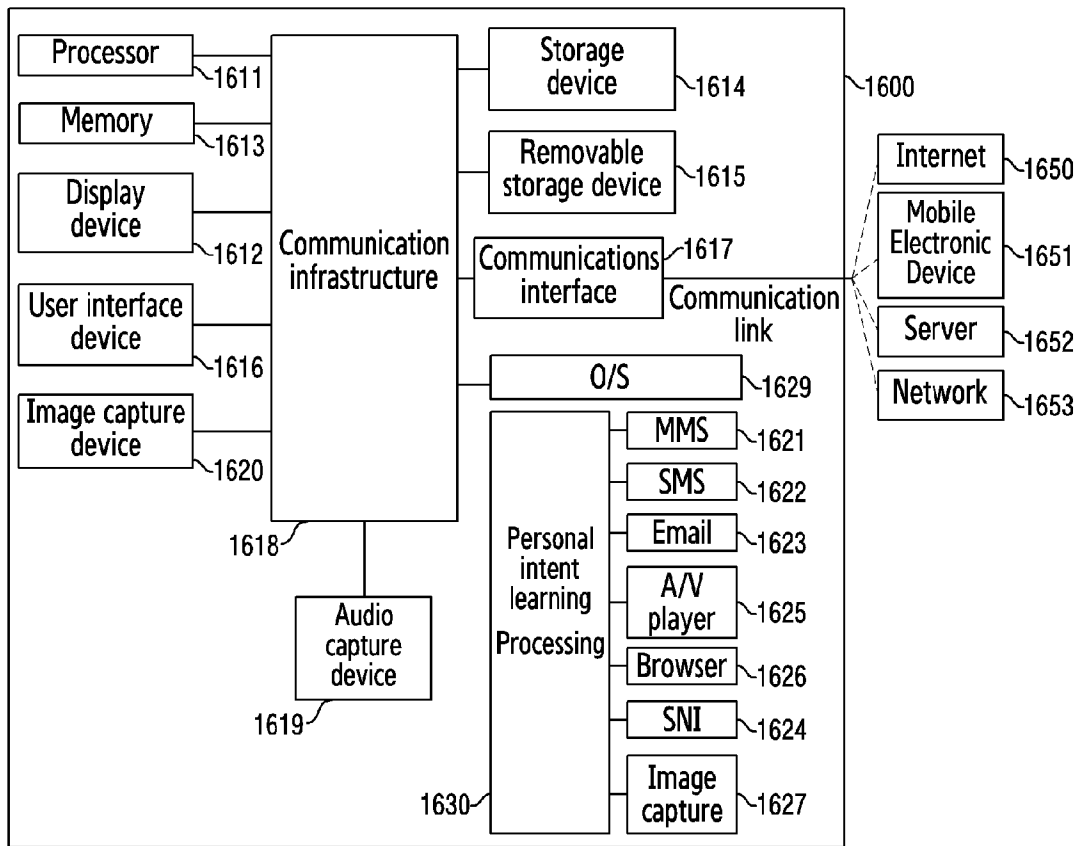
by noon get me some pizza and home
 for home buy some pizza before noon
 to home get me some pizza via noon
 buy a pizza by noon back home

Low Ranked Results

[Fig. 15]



[Fig. 16]



A. CLASSIFICATION OF SUBJECT MATTER**G06F 17/27(2006.01)i, G10L 15/22(2006.01)i**

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHEDMinimum documentation searched (classification system followed by classification symbols)
G06F 17/27; G06F 17/24; G06F 17/28; G06F 17/50; G10L 15/22Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched
Korean utility models and applications for utility models
Japanese utility models and applications for utility modelsElectronic data base consulted during the international search (name of data base and, where practicable, search terms used)
eKOMPASS(KIPO internal) & Keywords:personal assistant, learning cross domain, NL(Natural language), parser, training data, segments, transfer learning, voice**C. DOCUMENTS CONSIDERED TO BE RELEVANT**

| Category* | Citation of document, with indication, where appropriate, of the relevant passages | Relevant to claim No. |
|-----------|---|-----------------------|
| X | US 2018-0033435 A1 (DESPREZ, LLC) 01 February 2018 See paragraphs [0006], [0037]-[0038]; claim 1; and figures 2, 4. | 1 |
| Y | | 2-15 |
| Y | US 2016-0179785 A1 (INTERNATIONAL BUSINESS MACHINES CORPORATION) 23 June 2016 See paragraphs [0027]. | 2-8, 10-11, 14-15 |
| Y | US 2018-0011838 A1 (INTERNATIONAL BUSINESS MACHINES CORPORATION) 11 January 2018 See paragraphs [0019], [0038], [0041]. | 2-8 |
| Y | US 2016-0062982 A1 (FIDO LABS INC.) 03 March 2016 See paragraphs [0007], [0054]-[0055], [0058], [0069], [0095]; and figure 26. | 5-12 |
| Y | US 9922642 B2 (APPLE INC.) 20 March 2018 See column 1, lines 51-52; and column 20, lines 34-36. | 5-8, 13-15 |

 Further documents are listed in the continuation of Box C. See patent family annex.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search

30 May 2019 (30.05.2019)

Date of mailing of the international search report

01 July 2019 (01.07.2019)

Name and mailing address of the ISA/KR

International Application Division
Korean Intellectual Property Office
189 Cheongsa-ro, Seo-gu, Daejeon, 35208, Republic of Korea

Facsimile No. +82-42-481-8578

Authorized officer

KIM, Seong Woo

Telephone No. +82-42-481-3348



INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No.

PCT/KR2019/002406

| Patent document cited in search report | Publication date | Patent family member(s) | Publication date |
|--|------------------|--|--|
| US 2018-0033435 A1 | 01/02/2018 | US 10079016 B2 US 10229679 B1 US 9613020 B1 | 18/09/2018 12/03/2019 04/04/2017 |
| US 2016-0179785 A1 | 23/06/2016 | US 2016-0179940 A1 | 23/06/2016 |
| US 2018-0011838 A1 | 11/01/2018 | US 9910848 B2 | 06/03/2018 |
| US 2016-0062982 A1 | 03/03/2016 | EP 2915068 A2 US 2014-0136188 A1 US 9152623 B2 WO 2014-071330 A2 WO 2014-071330 A3 | 09/09/2015 15/05/2014 06/10/2015 08/05/2014 07/08/2014 |
| US 9922642 B2 | 20/03/2018 | AU 2014-233517 A1 AU 2014-233517 B2 CN 105027197 A CN 105027197 B KR 10-1759009 B1 KR 10-2015-0113127 A US 2014-0278413 A1 WO 2014-144949 A2 WO 2014-144949 A3 | 17/09/2015 25/05/2017 04/11/2015 14/12/2018 17/07/2017 07/10/2015 18/09/2014 18/09/2014 20/11/2014 |