Technical Disclosure Commons

Defensive Publications Series

April 03, 2019

Transliteration of Contact Names and/or Other Data Using an Automated Assistant

Diego Accame

Abraham Lee

Feng Ji

Ka Hei Cheung

Transliteration of Contact Names and/or Other Data Using an Automated Assistant

Abstract

Techniques are disclosed herein for transliterating contact names and/or other data from a source language to one or more target languages. The contact names and/or other data are received from a client device (*e.g.*, a feature phone) having an automated assistant. The data in the target language(s) can then be utilized in determining that further input, received from the client device in one of the target language(s), corresponds to certain data in the source language. In one aspect, a method includes receiving, over a network and at one or more remote servers, contact names that are in a source language and that are transmitted from a client device. The method can further include transliterating each of the contact names from the source language to one or more corresponding transliterated contact names in one or more target languages, and mapping each transliterated contact name in the target language(s) to a corresponding contact name in the source language and/or to a corresponding contact entry. The method can further include storing the mapping, along with an identifier of the client device, such as a temporary identifier of the client device. Thereafter, in response to further input received from the client device (*e.g.*, spoken input) in one of the target language(s), the spoken input can be matched to a transliterated contact name, and the mapping utilized to identify a corresponding contact entry and/or contact name in the source language. Thus, techniques disclosed herein can enable a user to create, at a client device, a contact name in a first alphabet (*e.g.*, Latin). The contact name can then be provided to remote server(s), which transliterates the contact name to one or more additional alphabets (*e.g.*, Devanāgarī). Subsequently, the user can provide, at the client device, spoken input that indicates a desire to contact (*e.g.*, call, message, *etc.*) the contact name, but that references the contact name in the Hindi language (which corresponds to the Devanāgarī alphabet). The spoken input can be provided to one or more of the servers, and the server(s) can match the contact name of the spoken input to a transliteration, of the contact name, that is in the Devanāgarī alphabet. Thus, the contact name can be contacted responsive to the spoken input.

Background

Electronic documents are typically written in many different languages. Each language is normally expressed in a particular writing system (*i.e.*, a script), which is usually characterized by a particular alphabet. For example, the English language is expressed using the Latin alphabet while the Hindi language is normally expressed using the Devanāgarī alphabet. The scripts used by some languages include a particular alphabet that has been extended to include additional marks or characters. However, it may be more practical for a user of a client device to store certain documents and/or data using the Latin alphabet, compared to, for example, the Devanāgarī alphabet, because the Devanāgarī alphabet has been extended to include additional marks or characters that result in certain hardware constraints (*i.e.*, keyboard constraints) and software constraints (*e.g.*, processing spoken input in different languages).

In transliteration, the script of one language is used to represent words normally written in the script of another language. For example, a transliterated term can be a term that has been converted from one script to another script or a phonetic representation in one script of a term in another script, and stored in one or more databases for subsequent use by the client device. Accordingly, the techniques disclosed herein enable transliteration and processing of data received from the client device.

Description

The techniques of this disclosure enable users to maintain an up-to-date database of transliterated data from a client device, and to subsequently access the database of transliterated data to perform one or more actions associated with the transliterated data.

In some implementations, components that contribute to implementations of maintaining a database of transliterated data from a client device described herein may be operated exclusively on the client device, may be operated using cloud-based automated assistant component(s), and/or may be operated using any combination thereof. In some of those implementations, a machine learning model may be trained and/or stored on the client device, *e.g.*, behind an Internet firewall, so that training data and other information generated by or associated with the machine learning model may be maintained in privacy. And in some such implementations, the cloud-based text-to-speech (TTS, cloud-based speech-to-text (STT) module, cloud-based aspects of natural language processing engine, and/or cloud-based transliteration engine may not be involved in implementing the machine learning model. In some of those implementations, the machine learning model may be trained and/or stored on one or more remote servers accessible by the off-line client device and cloud-based automated assistant component(s) over a network.

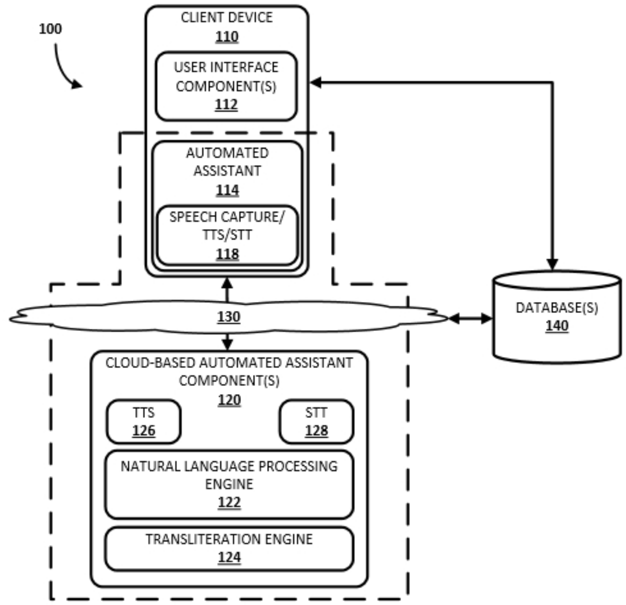

FIG. 1

FIG. 1 illustrates a system 100 that includes a client device 110. The client device 110 may be, for example, a desktop computer, a tablet computer, a laptop computer, a smart phone, a feature phone, and/or a wearable device. The client device 110 may include user interface component(s) 112. In some implementations, the user interface component(s) 112 may include user interface input components and user interface output components. In some of those implementations, the user interface input components may include a keyboard, pointing devices such as a mouse or graphics tablet, a touchscreen of the display, audio input devices such as microphones, or the like. In some of those implementations, the user interface output components may include a display subsystem, a printer, a fax machine, or non-visual displays such as audio output devices. The display subsystem may include a cathode ray tube (CRT), a liquid crystal display (LCD), or the like for creating a visible image.

The client device 110 may further include an automated assistant 114. In some implementations, the automated assistant 114 may include a corresponding speech capture/ TTS/STT module 118. In some of those implementations, the speech capture/TTS/STT module 118 may be implemented separately from the automated assistant 114. Each speech capture/TTS/STT module 118 may be configured to perform one or more functions: capture a user’s speech, *e.g.*, via a microphone; convert that captured speech to text; and/or covert text to speech. For example, in some implementations, because the client device 110 may be relatively constrained in terms of computing resources (*e.g.*, processor cycles, memory, battery, *etc.*), the speech capture/TTS/STT module 118 that is local to the client device 110 may be configured to convert a finite number of different spoken phrases to text. Other speech input may be sent over a network 130 to cloud-based automated assistant components 120, which may include a cloud-based TTS module 126 and/or a cloud-based STT module 128.

Cloud-based STT module 128 may be configured to leverage the resources of the cloud to convert audio data captured by speech capture/TTS/STT module 118 into text (which may then be provided to natural language processing engine 122). Cloud-based TTS module 126 may be configured to leverage the resources of the cloud to convert textual data (*e.g.*, natural language responses formulated by the automated assistant 114) into computer-generated speech output. In some implementations, TTS module 126 may provide the computer-generated speech output to the off-line client device 110 to be output directly, *e.g.*, using one or more speakers. In other implementations, textual data (*e.g.*, natural language responses) generated by automated assistant 114 may be provided to speech capture/TTS/STT module 118, which may then convert the textual data into computer-generated speech that is output locally.

The automated assistant 114 (and in particular, cloud-based automated assistant component(s) 120) may include a natural language processing engine 122, a transliteration engine 124, the aforementioned TTS module 126, the aforementioned STT module 128, and/or other components. In some implementations, one or more of the engines and/or modules of automated assistant 1114 may be omitted, combined, and/or implemented in a component that is separate from the automated assistant 114. And as noted above, to protect privacy, one or more of the components of the automated assistant 114 may be implemented at least in part on the client device 110 (*e.g.*, to the exclusion of the cloud). In some such implementations, the speech capture/TTS/STT module 118 may be configured to perform selected aspects of the present disclosure, while in some cases leaving other aspects to cloud-based components when suitable.

In some implementations, the automated assistant 114 generates responsive content in response to various inputs generated by a user of the client device 110 during a dialog session with the automated assistant 114. The automated assistant 114 may provide the responsive content (*e.g.*, over one or more networks 130 when separate from a client device of a user) for presentation to the user as part of a dialog session. For example, automated assistant 114 may generate responsive content in response to free-form natural language input provided via the client device 110. As used herein, free-form input is input that is formulated by a user, typed or spoke, and that is not constrained to a group of options presented for selection by the user.

In some implementations, the client device 110 can be a signed-out client device. As used herein, a "signed-out" client device may include a client device that is not associated with an account for a particular service, connected to an account for the particular service, and/or logged into an account for the particular service. Accordingly, very little user information (*i.e.*, an identity of the user, a username of the user, a unique identifier associated with the user, *etc.*), if any, is known about a user of the signed-out client device. Therefore, a corresponding unique signed-out identifier (*e.g.*, ephemeral token) may be generated for the user of the signed-out client device. In some implementations, a single corresponding unique signed-out identifier may be generated for the signed-out client device the first time the automated assistant 114 of the signed-out client device communicates with the cloud-based automated assistant component(s) 120. In other implementations, a new corresponding unique signed-out identifier may be generated for the signed-out client device each time the automated assistant 114 of the signed-out client device communicates with the cloud-based automated assistant component(s) 120.

The natural language processing engine 122 of the cloud-based automated assistant component(s) 120 processes natural language input generated by users via the client device 110, and may generate annotated output for use by one or more other components of cloud-based automated assistant component(s) 120. For example, the natural language processing engine 122 may process natural language free-form input that is generated by a user via one or more user interface component(s) 112 of the client device 110. The generated annotated output includes one or more annotations of the natural language input and optionally one or more (*e.g.*, all) of the terms of the natural language input.

In some implementations, the natural language processing engine 122 is configured to identify and annotate various types of grammatical information in natural language input. For example, the natural language processing engine 122 may include a part of speech tagger configured to annotate terms with their grammatical roles (such as “noun,” “verb,” “adjective,” “pronoun,” *etc.*). Also, for example, in some implementations the natural language processing engine 122 may additionally and/or alternatively include a dependency parser (not depicted) configured to determine syntactic relationships between terms in the input. For example, the dependency parser may determine which terms modify other terms, subjects and verbs of sentences, and so forth (*e.g.*, a parse tree) — and may make annotations of such dependencies.

In some implementations, the natural language processing engine 122 may additionally and/or alternatively include an entity tagger (not depicted) configured to annotate entity references in one or more segments such as references to people (including, for instance, contacts, literary characters, celebrities, public figures, *etc.*), organizations, locations (real and imaginary), and so forth. In some implementations, data about entities may be stored in one or more databases 140, such as in a knowledge graph (not depicted). In some implementations, the knowledge graph may include nodes that represent known entities (and/or entity attributes), as well as edges that connect the nodes and represent relationships between the entities. For example, a "chakrabarty" node may be connected (*e.g.*, as a child) to a "contact name" node, which in turn may be connected (*e.g.*, as a child) to "person" and/or "human" nodes.

The transliteration engine 124 of the cloud-based automated assistant component(s) 120 processes the annotated output generated by the natural language processing module 122. Transliteration converts data, or a term included in the data, in a source language to a transliterated term in a target language, or vice versa. After conversion, the letters or characters of the term in the source language are represented by letters or characters of the target language. In some implementations, a machine learning system may be trained and utilized for transliteration. In some implementations, the transliteration engine 124 may generate a mapping from one or more terms in an original language to one or more terms in a target language. In some implementations, the transliteration engine 124 may generate a mapping from one or more terms in the target language to one or more terms in the source language. The techniques disclosed herein use the transliteration engine 124 to determine what terms in a target language (*e.g.*, English) are transliterated from the same term in a source language (*e.g.*, Hindi).

In some implementations, the mapping of the transliteration 124 may score and rank a list of candidate transliterated terms using an associated confidence value for each transliterated term. Each of the candidate transliterated terms may be scored and ranked with respect to one another. In some of these implementations, if a particular transliterated candidate term has a confidence value with respect to the first candidate transliterated candidate term that is above a specified threshold, the particular candidate transliterated term is a second candidate transliterated term identified as a candidate transliterated term of the first transliterated term. If the mapping does not produce a transliteration score for each transliteration, the confidence value for a given second candidate transliterated term can be a function of the number of terms in the source language that are mapped from both the first candidate transliterated term and the given second candidate transliterated term. If the mapping produces a transliteration score for each candidate transliteration, the confidence value for the given second candidate transliterated term can be a function of the transliteration scores of the first candidate transliterated term and of the given second candidate transliterated term. The transliteration engine 124 may choose the candidate transliteration term with the highest confidence value as the transliterated term.

Each transliterated term may be stored in one or more database(s) 140, and if applicable, along with a corresponding unique signed-out identifier associated with the client device 110. The one or more of the databases(s) 140 may be accessed by the client device 110 and/or the cloud-based automated assistant component(s) 120 over the network 130. The one or more database(s) 140 may include a cloud-based database, a local cache stored on the client device 110, one or more servers at a remote location, and/or any combination thereof. Further, the one or more database(s) 140 may index each of the transliterated terms, list of candidate transliterated terms, and their respective confidence scores for each candidate transliterated term.

Turning now to an example implemented by the system 100 of FIG. 1, data on a client device 110 may include a list of contact names. Even though a user of the client device 110 may not speak English, or use a writing system that utilizes the Latin alphabet, the user may store each contact in the list of contact names using the Latin alphabet because of certain hardware and/or software constraints, such as constraints associated with feature phones. As an example, the Hindi surname, चक्रबर्ती, may be stored in the list of contact names as “chakrabarty”. However, if a user provides spoken input “call ‘चक्रबर्ती’” to an automated assistant 114, the automated assistant 120 may struggle to associate the received spoken input “call ‘चक्रबर्ती’” with the contact name “chakrabarty” stored in the list of contacts. Therefore, it may be advantageous to have the transliteration of चक्रबर्ती stored in one or more databases for subsequent recognition.

Initially, a list of contact names may be stored in one or more database(s) 140, one or more documents, one or more applications, and/or any combination thereof. Although a user may not speak English or typically use the Latin alphabet, each contact in the list of contact names may be stored using the Latin alphabet. Cloud-based automated assistant component(s) 120 may receive the list of contact names sent over a network 130 from the client device 110. However, when an automated assistant 114 receives spoken input related to a contact name in the list of contact names, the automated assistant 114 (or the cloud-based automated assistant component(s) 120) may struggle to process the spoken input because the user does not speak English or use the Latin alphabet.

After storing the list of contact names in one or more of the database(s) 140, the list of contact names may be updated. In some implementations, updating the list of contact names in one or more of the database(s) 140 may occur automatically after a threshold period of time. For example, the list of contact names may be updated every seven days or every fourteen days. In some implementations, updating the list of contact names in one or more of the database(s) 140 may occur when the list of contacts changes. For example, if the client device 110 detects a change in one or more contact names for the list of contact names, then the client device 110 may compare the list of contact names in one or more of the database(s) 140 with the list of contact name(s) in one or more of the documents or one or more of the applications to determine which contact names in the list of contact names has been added, omitted, changed, and/or any combination thereof. The comparison may occur locally on the client device 110 or by one or more servers connected to the client device 110 over the network 130. In some implementations, the automated assistant 120 may receive a request from the user to update the list of contact names in one or more of the database(s). In each of these implementations, the list of contact names may be updated using batch processing, stream processing, and/or a combination thereof.

Further, the list of contact names stored in one or more of the database(s) 140 may be associated with a corresponding unique signed-out identifier. In some implementations, the corresponding unique signed-out identifier may be associated with the client device 110 and/or a list of contact names, and remain the same corresponding unique signed-out identifier after the updating. In some implementations, the corresponding unique signed-out identifier may be associated with a particular update of the list of client names, and replace the previous corresponding unique signed-out identifier. In some implementations, a new corresponding unique signed-out identifier may be generated and associated with the client device 110 and/or list of contact names after the updating. Once the system 100 has been trained and the list of contacts names has been stored in one or more of the database(s) 140, the system 100 may receive voice input, transliterate the voice input, and determine one or more contact names from the list of contact names to perform an action associated with one or more of the contact names.

The Hindi surname, चक्रबर्ती, can be transliterated into English as “chakrabarti” or “chakrabarty”. However, if a given telephone keypad utilizes the ITU E 1.161 International Standard, it is not practical to type चक्रबर्ती as a surname for a contact. Thus, the transliterated term “chakrabarti” or “chakrabarty”, rather than चक्रबर्ती, may be stored in the list of contact names in one or more of the database(s) 140. Further, the transliterated term “chakrabarti” or “chakrabarty” may be stored with a mapping from “chakrabarti” or “chakrabarty” to चक्रबर्ती,

stored with a mapping from a mapping from चक्रबर्ती to "chakrabarti" or "chakrabarty", and stored with a corresponding unique signed-out identifier. The corresponding unique signed-out identifier may enable the client device 110 to access one or more databases and, during subsequent use, receive the transliterated term using the automated assistant 114.

For example, cloud-based automated assistant component(s) 120 may receive a list of contact names of a Hindi user stored using the Latin alphabet sent over a network 130 from a client device 110. The list of contact names, a transliterated term for each contact name in the list of contact names, and a corresponding unique signed-off identifier may be stored in one or more database(s) 140. An automated assistant 114 may receive a spoken input from the Hindi user to "call चक्रबर्ती". The natural language processing engine may parse and annotate the spoken input. For example, the natural language processing engine 122 may tag the term "call" as a verb, or action to be performed in connection with an entity. However, because the user speaks Hindi, and because the contact name चक्रबर्ती is stored in the Hindi user's list of contact names as either "chakrabarti" or "chakrabarty", the natural language processing engine 122 of the automated assistant 114 may struggle to process the Hindi user's spoken input and determine the entity reference associated with the action "call".

The cloud-based automated assistant component(s) 120 may receive the Hindi user's spoken input of "call चक्रबर्ती" and the corresponding unique off-line identifier over the network 130, and transliterate a portion of the Hindi user's spoken input by mapping the entity reference "चक्रबर्ती" to the entity reference "chakrabarti" or "chakrabarty". The system 100, at the client device 110 or at the cloud-based automated assistant component(s) 120, may use the corresponding unique signed-out identifier to determine a list of contact names associated with the Hindi user, where the list of contact names is stored in one or more database(s) 140. Further, the system 100 may compare the transliteration of the Hindi user's spoken input "chakrabarti" or "chakrabarty" to determine one or more contacts stored in the Hindi user's list of contacts. The automated assistant 114 may select one or more of the contact names from the list of contact names based on the confidence value (as described hereinabove) for each of the contact names, and provide the entity reference and the action to be performed in association with the selected contact name, such that the automated assistant 114 may provide either "call chakrabarti" or "call chakrabarty" in response to the spoken input. The automated assistant 114 may automatically perform the action associated with the entity reference or prompt the user to perform the action associated with the entity reference.

Although this example focuses on transliterating spoken input in the Hindi language that uses the Devanāgarī alphabet to the English language that uses the Latin alphabet, the system 100 may use the same model to transliterate spoken input the English language that uses the Latin alphabet to the Hindi language that uses the Devanāgarī alphabet.

A machine learning model may be trained and/or stored on the client device 110, or on one or more servers at a remote location accessible over the network 130. Accordingly, the trained model may be used to transliterate data received from the client device 110. Namely, the machine learning model may be trained to transliterate one or more terms received from the client device 110 as spoken input. For example, the trained machine learning model may be trained to transliterate a term in the Hindi language using the Devanāgarī alphabet to a term in the English language using the Latin alphabet. Conversely, the same trained machine learning model may transliterate a term in the English language using the Latin alphabet to a term in the Hindi language using the Devanāgarī alphabet. Although the examples focus on transliteration between the English and Hindi languages that is not meant to be limiting. A machine learning model may be trained for transliteration between any two or more languages.

Moreover, privacy of a user's data may be a concern to a user or a contact in the list of contact names. In some implementations, only a list of contact names may be stored in one or more of the database(s). For example, the stored list of contact names does not include a telephone number associated with a contact name in the list of contact names, or any other identifying information. Further, the system 100 may prompt a user to provide consent to storing the list of contact names each time the list of contact names is initially stored or updated. Thus, all that is stored is a list of contact names, not any data associated with the contact name.

Conclusion

This disclosure provides techniques for transliterating data received from an off-line client device running an automated assistant. Namely, these techniques may be used for transliterating a list of contact names form a source language to a target language, where the stored transliterated terms may be retrieved for subsequent use by the automated assistant. Further, these techniques may be implemented using a trained machine learning model.