

(51) International Patent Classification:  
*G06F 17/28* (2006.01)(21) International Application Number:  
PCT/CN2013/077161(22) International Filing Date:  
13 June 2013 (13.06.2013)

(25) Filing Language: English

(26) Publication Language: English

(71) Applicant (for all designated States except US):  
**GOOGLE INC.** [US/US]; 1600 Amphitheatre Parkway,  
Mountain View, California 94043 (US).

(72) Inventor; and

(71) Applicant (for US only): **SHEN, Hong** [CN/CN]; Room  
1002A, Block A, Zhongguancun E-Plaza, 11 Zhong-  
guancun Avenue, Haidian District, Beijing 100084 (CN).(74) Agent: **UNITALEN ATTORNEYS AT LAW**; 7th Floor,  
Scitech Place, No.22, Jian Guo Men Wai Ave., Chao Yang  
District, Beijing 100004 (CN).(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, DK, DM,  
DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT,  
HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KN, KP, KR,  
KZ, LA, LC, LK, LR, LS, LT, 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, 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, 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,  
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))

(54) Title: TECHNIQUES FOR USER IDENTIFICATION OF AND TRANSLATION OF MEDIA

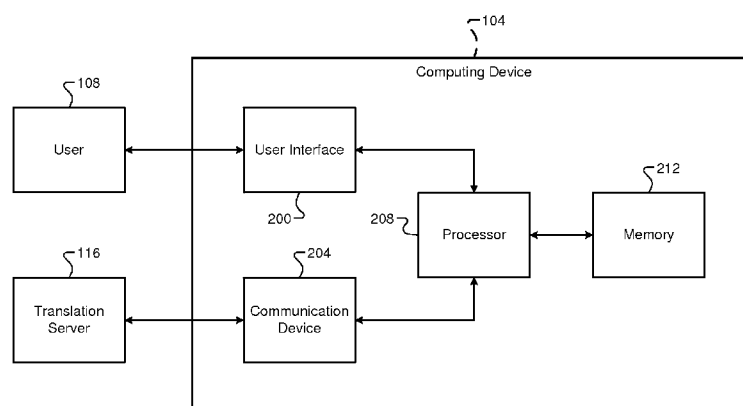


FIG. 2

(57) **Abstract:** A computer-implemented technique includes receiving, at a computing device including one or more processors, a user input (i) identifying a portion of a media stream being output from the computing device and (ii) indicating a request to translate the portion of the media stream from a source language to a target language. The technique includes transmitting, from the computing device, the portion of the media stream to a translation server via a network in response to receiving the user input. The technique includes receiving, at the computing device, a translated portion of the media stream from the translation server via the network, the translated portion of the media stream having been translated from the source language to the target language by the translation server. The technique also includes outputting, at the computing device, the translated portion of the media stream.

## TECHNIQUES FOR USER IDENTIFICATION OF AND TRANSLATION OF MEDIA

## FIELD

[0001] The present disclosure relates to language learning and, more particularly, to techniques for user identification of and translation of media.

## BACKGROUND

[0002] The background description provided herein is for the purpose of generally presenting the context of the disclosure. Work of the presently named inventors, to the extent it is described in this background section, as well as aspects of the description that may not otherwise qualify as prior art at the time of filing, are neither expressly nor impliedly admitted as prior art against the present disclosure.

[0003] A user may access a website from a computing device via a network such as the Internet. The website may display a web page to the user via a web browser executing on the computing device. The web page may include text, images, audio and/or video data, or a combination thereof, to be output to the user via one or more components associated with the computing device (a display, one or more speakers, etc.).

## SUMMARY

[0004] A computer-implemented technique is presented. The technique can include outputting, at a computing device including one or more processors, a media stream that includes at least one of audio data and video data. The technique can include during the outputting of the media stream: (i) receiving, at the computing device, a first user input (a) identifying a portion of the media stream and (b) indicating a request to

translate the portion of the media stream from a source language to a target language, (ii) determining, at the computing device, a text corresponding to the at least one of audio data and video data from the portion of the media stream, and (iii) transmitting, from the computing device, the text to a translation server via a network. The technique can include receiving, at the computing device, a translated text from the translation server via the network, the translated text having been translated from the source language to the target language by the translation server. The technique can also include outputting, at the computing device, the translated text after outputting of the media stream has completed.

[0005] Another computer-implemented technique is also presented. The technique can include receiving, at a computing device including one or more processors, a user input (i) identifying a portion of a media stream being output from the computing device and (ii) indicating a request to translate the portion of the media stream from a source language to a target language. The technique can include transmitting, from the computing device, the portion of the media stream to a translation server via a network in response to receiving the user input. The technique can include receiving, at the computing device, a translated portion of the media stream from the translation server via the network, the translated portion of the media stream having been translated from the source language to the target language by the translation server. The technique can also include outputting, at the computing device, the translated portion of the media stream.

[0006] In some embodiments, the media stream includes at least one of audio data and video data.

[0007] In other embodiments, the technique further includes determining, at the computing device, a text corresponding to the at least one of audio data and video data for the portion of the media stream.

[0008] In some embodiments, determining the text includes at least one of (i) optical character recognition of subtitles in the video data for the portion of the media stream and (ii) voice recognition of the audio data for the portion of the media stream.

[0009] In other embodiments, transmitting the portion of the media stream to the translation server via the network includes transmitting, from the computing device, the text to the translation server via the network during outputting of the media stream.

[0010] In some embodiments, receiving the translated portion of the media stream includes receiving a translated text, the translated text having been translated from the source language to the target language by the translation server.

[0011] In other embodiments, outputting the translated portion of the media stream includes outputting, at the computing device, the translated text.

[0012] In some embodiments, outputting the translated text includes outputting, at the computing device, the translated text after the outputting of the media stream has completed.

[0013] In other embodiments, the portion of the media stream is a predetermined period prior to when the user input was received.

[0014] In some embodiments, the user input includes one or more keystrokes of a keyboard.

[0015] A computing device is also presented. The computing device can include a user interface and a communication device. The user interface can be configured to

receive a user input (i) identifying a portion of a media stream being output from the computing device and (ii) indicating a request to translate the portion of the media stream from a source language to a target language. The communication device can be configured to: transmit the portion of the media stream to a translation server via a network in response to receiving the user input, and receive a translated portion of the media stream from the translation server via the network, the translated portion of the media stream having been translated from the source language to the target language by the translation server. The user interface can also be configured to output the translated portion of the media stream.

[0016] In some embodiments, the media stream includes at least one of audio data and video data.

[0017] In other embodiments, the computing device further includes one or more processors configured to determine a text corresponding to the at least one of audio data and video data for the portion of the media stream.

[0018] In some embodiments, the one or more processors are configured to determine the text by at least one of (i) optical character recognition of subtitles in the video data for the portion of the media stream and (ii) voice recognition of the audio data for the portion of the media stream.

[0019] In other embodiments, the communication device is configured to transmit the portion of the media stream to the translation server via the network by transmitting the text to the translation server via the network during outputting of the media stream.

[0020] In some embodiments, the communication device is configured to receive the translated portion of the media stream by receiving a translated text, the translated text

having been translated from the source language to the target language by the translation server.

[0021] In other embodiments, the user interface is configured to output the translated portion of the media stream by outputting the translated text.

[0022] In some embodiments, the user interface is configured to output the translated text by outputting the translated text after the outputting of the media stream has completed.

[0023] In other embodiments, the portion of the media stream is a predetermined period prior to when the user input was received.

[0024] Further areas of applicability of the present disclosure will become apparent from the detailed description provided hereinafter. It should be understood that the detailed description and specific examples are intended for purposes of illustration only and are not intended to limit the scope of the disclosure.

#### BRIEF DESCRIPTION OF THE DRAWINGS

[0025] The present disclosure will become more fully understood from the detailed description and the accompanying drawings, wherein:

[0026] FIG. 1 is a functional block diagram of a computing network including an example computing device according to some implementations of the present disclosure;

[0027] FIG. 2 is a functional block diagram of the example computing device of FIG. 1; and

[0028] FIG. 3 is a flow diagram of an example technique for user identification of and translation of media according to some techniques of the present disclosure.

## DETAILED DESCRIPTION

[0029] As previously mentioned, a user may access a website that displays a web page including text, images, audio and/or video data, or a combination thereof. Given the worldwide reach of the Internet, the web page may have been created in a language that the user cannot understand. For example, the user may be unable to read text or subtitles of a video stream or the user may be unable to understand an audio stream. The user, therefore, may request to translate a portion of the web page (text, video subtitles, audio, etc.) from its source language to a target language that the user can understand.

[0030] Accordingly, techniques are presented for user identification of and translation of media. The techniques can generally provide for a more efficient method of user identification of and translation of media, which can improve a user's overall experience. The techniques also provide for identification of and translation of only a portion of media item that the user cannot understand, instead of translating the entire media item. The techniques can receive, at a computing device including one or more processors, a user input (i) identifying a portion of a media item and (ii) indicating a request to translate the portion of the media item from a source language to a target language. The term "media item" as used herein can refer to a digital item capable of being output at the computing device. The term "media item," therefore, can include text, video, audio, or a combination thereof. A media item including at least one of audio data and video data can also be referred to as a "media stream" because the media item may be streamed for a period of time. In response to the user input, the computing device can transmit the portion of the media item, e.g., a text, to a translation server via a network. The computing device can then receive a translated portion of the media item, e.g., a

translated text, from the translation server via the network. In some cases, the media item may be a media stream, and the techniques can further include determining the text corresponding to a portion of the media stream, which can then be transmitted to the translation server via the network.

[0031] It should be appreciated that the techniques of the present disclosure can be executed a plurality of times during a particular period, e.g., during the media stream. In other words, the user may provide the user input a plurality of different times to identify different portions of the media item for translation. The techniques can then transmit these different portions of the media item to the translation server for translation. For example, the techniques can transmit the different portions sequentially or in one or more groups. Similarly, the techniques can then receive different translated portions of the media item from the translation server. For example, the techniques can receive the different translated portions sequentially or in the one or more groups. Each portion may also include a unique identifier for associating the portion of the media item (pre-translation) to its corresponding translated portion. The computing device can then output the different translated portions in any suitable manner (sequentially, in the one or more groups, etc.). For example only, the user may identify a plurality of portions of a media stream and, after the media stream has completed, the user can review the translated text for each of the plurality of portions.

[0032] Referring now to FIG. 1, a computing network 100 including an example computing device 104 according to some implementations of the present disclosure is illustrated. It should be appreciated that the term “computing device” as used herein can refer to any suitable computing device including one or more processors (a desktop

computer, a laptop computer, a tablet computer, a mobile phone, etc.). The computing device 104 can receive input from and/or provide output to a user 108. The computing device 104 can communicate with other computing devices via a network 112. The network 112 can include a local area network (LAN), a wide area network (WAN), e.g., the Internet, or a combination thereof.

[0033] In particular, the computing device 104 can communicate with a server 116 via the network 112. It should be appreciated that the term “server” as used herein can refer to either a single server or two or more servers operating in a parallel or distributed architecture. The server 116 can be configured to perform translation of a text from a source language to a target language. The server 116, therefore, can be referred to hereinafter as translation server 116. The translation server 116 can perform translation of the text from the source language to the target language using any suitable language translation technique. For example only, the translation server 116 can translate the text from the source language to the target language using statistical machine translation (SMT).

[0034] According to the techniques of the present disclosure, the user 108 can provide a user input at the computing device 104. The user input can (i) identify a portion of a media item and (ii) indicate a request to translate the portion of the media item from a source language to a target language, which is described in further detail below. In response to the user input, the computing device 104 can transmit the portion of the media item, e.g., a text, to the translation server 116 via the network 112. The computing device 104 can then receive a translated portion of the media item, e.g., a translated text, from the translation server 116 via the network 112.

[0035] Referring now to FIG. 2, a functional block diagram of the example computing device 104 of FIG. 1 is illustrated. The computing device 104 can include a user interface 200, a communication device 204, a processor 208, and a memory 212

[0036] The user interface 200 can receive input to, e.g., from the user 108 and/or provide output from the computing device 104. It should be appreciated that the term “user interface” as used herein can refer to any suitable components for receiving input and/or providing output. For example, the user interface 200 can include a mouse, a keyboard, and/or a display.

[0037] The communication device 204 can control communication between the computing device 104 and other devices via the network 112. The communication device 204 can include any suitable components for communication via the network 112, e.g., a transceiver. In particular, the communication device 204 can control communication between the computing device 104 and the translation server 116 via the network 112. More specifically, the communication device 204 can transmit a text to the translation server 116 via the network 112 and can receive a translated text from the translation server 116 via the network 112.

[0038] The processor 208 can control operation of the computing device 104. It should be appreciated that the term “processor” as used herein can refer to either a single processor or two or more processors operating in a parallel or distributed architecture. For example, the processor 208 can perform functions including, but not limited to loading and executing an operating system of the computing device 104, controlling information input and/or output via the user interface 200, controlling communication over the network 112 via the communication device 204, and/or

controlling read/write operations at the memory 212. The memory 212 can be any suitable storage medium (flash, hard disk, etc.) configured to store information at the computing device 104.

[0039] The processor 208 can also execute the techniques according to the present disclosure. The processor 208 can execute the media item to provide output (text, audio, video, etc.) to the user 108. During the outputting of the media item, the user 108 can provide a user input to the computing device 104 that (i) identifies a portion of the media item and (ii) indicates a request to translate the portion of the media item. The user input can include one or more mouse clicks, one or more keyboard keystrokes, or some combination thereof. Other suitable user input configurations can also be implemented. The specific user input can be predefined or specified by the user 108, and may be stored at the memory 212. The user input can also depend on the type of media item being output. For example, when the media item includes only text, e.g., a text-only web page, the user input may be one or more mouse clicks at or near a portion of text. Alternatively, for example, when the media item is a media stream including audio data and/or video data, the user input may be one or more keyboard keystrokes.

[0040] In response to receiving the user input, the processor 208 can obtain or capture the portion of the media item identified by the user input. The portion of the media item can also depend on the type of media item being output. For example, when the media item includes only text, the user input can cause the processor 208 to capture the portion of text within a predefined distance or area from the one or more mouse clicks. Alternatively, for example, when the media item is a media stream including audio data and/or video data, the user input can cause the processor 208 to

capture the portion of the media stream (its audio data and/or video data) for a predetermined period prior to when the user input was received. Other suitable rules specifying how to capture the portion of the media item can also be implemented.

[0041] As previously mentioned, when the media item is a media stream including at least one of audio data and video data, the processor 208 can also determine a text corresponding to the audio data and/or video data for the portion of the media item. For example, the processor 208 can perform voice/speech recognition to determine the text from the audio data for the portion of the media item. Any suitable voice/speech recognition techniques can be used. Additionally, in some cases the video data can include subtitles for corresponding audio data. In these cases, the processor 208 may also determine the text from the subtitles in the video data. For example only, the processor 208 can perform optical character recognition to determine the text from the subtitles in the video data. Other suitable image-to-text conversion techniques can also be used.

[0042] Additionally, after the communication device 204 has received the translated text from the translation server 116, the processor 208 can control the manner in which the translated text is presented to the user 108 via the user interface 200. Specifically, when the media item is a media stream, the processor 208 can capture the portion of the media item and receive the translated text from the translation server 116 before the outputting of the media stream has completed. In these cases, the processor 208 can wait until the media stream has completed before outputting the translated text via the user interface 200. Alternatively, the processor 208 can output the translated text immediately upon receipt from the translation server 116 or a predetermined period

after receipt. Further, the processor 208 can control additional information that is presented to the user 108 via the user interface 200, such as the text for the portion of the media item (pre-translation), a definition, related words/phrases, hyperlinks to Internet-based language resources, and the like.

[0043] Referring now to FIG. 3, a flow diagram of an example technique 300 for user identification of and translation of media is illustrated. At 304, the computing device 104 can receive a user input (i) identifying a portion of a media stream being output at the computing device 104 and (ii) indicating a request to translate the portion of the media stream from a source language to a target language. At 308, the computing device 104 can transmit the portion of the media stream to the translation server 116 via the network 112 in response to receiving the user input. Specifically, the computing device 104 can transmit a text corresponding to the portion of the media stream to the translation server 116 via the network 112. In some implementations, the computing device 104 can determine the text based on the techniques previously discussed herein (optical character recognition, voice/speech recognition, etc).

[0044] At 312, the computing device 104 can receive a translated portion of the media stream from the translation server 116 via the network 112, the translated portion of the media stream having been translated from the source language to the target language by the translation server 116. Specifically, the computing device 104 can receive a translated text from the translation server 116 via the network 112, the translated text having been translated from the source language to the target language by the translation server 116. At 316, the computing device 104 can output the translated text. In some implementations, the computing device 104 may wait until the

media stream has completed before outputting the translated text. The computing device 104 can also output additional information as previously discussed herein. The technique 300 can then end or return to 304 for one or more additional cycles.

[0045] Example embodiments are provided so that this disclosure will be thorough, and will fully convey the scope to those who are skilled in the art. Numerous specific details are set forth such as examples of specific components, devices, and methods, to provide a thorough understanding of embodiments of the present disclosure. It will be apparent to those skilled in the art that specific details need not be employed, that example embodiments may be embodied in many different forms and that neither should be construed to limit the scope of the disclosure. In some example embodiments, well-known procedures, well-known device structures, and well-known technologies are not described in detail.

[0046] The terminology used herein is for the purpose of describing particular example embodiments only and is not intended to be limiting. As used herein, the singular forms "a," "an," and "the" may be intended to include the plural forms as well, unless the context clearly indicates otherwise. The term "and/or" includes any and all combinations of one or more of the associated listed items. The terms "comprises," "comprising," "including," and "having," are inclusive and therefore specify the presence of stated features, integers, steps, operations, elements, and/or components, but do not preclude the presence or addition of one or more other features, integers, steps, operations, elements, components, and/or groups thereof. The method steps, processes, and operations described herein are not to be construed as necessarily requiring their performance in the particular order discussed or illustrated, unless

specifically identified as an order of performance. It is also to be understood that additional or alternative steps may be employed.

[0047] Although the terms first, second, third, etc. may be used herein to describe various elements, components, regions, layers and/or sections, these elements, components, regions, layers and/or sections should not be limited by these terms. These terms may be only used to distinguish one element, component, region, layer or section from another region, layer or section. Terms such as "first," "second," and other numerical terms when used herein do not imply a sequence or order unless clearly indicated by the context. Thus, a first element, component, region, layer or section discussed below could be termed a second element, component, region, layer or section without departing from the teachings of the example embodiments.

[0048] As used herein, the term module may refer to, be part of, or include: an Application Specific Integrated Circuit (ASIC); an electronic circuit; a combinational logic circuit; a field programmable gate array (FPGA); a processor or a distributed network of processors (shared, dedicated, or grouped) and storage in networked clusters or datacenters that executes code or a process; other suitable components that provide the described functionality; or a combination of some or all of the above, such as in a system-on-chip. The term module may also include memory (shared, dedicated, or grouped) that stores code executed by the one or more processors.

[0049] The term code, as used above, may include software, firmware, byte-code and/or microcode, and may refer to programs, routines, functions, classes, and/or objects. The term shared, as used above, means that some or all code from multiple modules may be executed using a single (shared) processor. In addition, some or all

code from multiple modules may be stored by a single (shared) memory. The term group, as used above, means that some or all code from a single module may be executed using a group of processors. In addition, some or all code from a single module may be stored using a group of memories.

[0050] The techniques described herein may be implemented by one or more computer programs executed by one or more processors. The computer programs include processor-executable instructions that are stored on a non-transitory tangible computer readable medium. The computer programs may also include stored data. Non-limiting examples of the non-transitory tangible computer readable medium are nonvolatile memory, magnetic storage, and optical storage.

[0051] Some portions of the above description present the techniques described herein in terms of algorithms and symbolic representations of operations on information. These algorithmic descriptions and representations are the means used by those skilled in the data processing arts to most effectively convey the substance of their work to others skilled in the art. These operations, while described functionally or logically, are understood to be implemented by computer programs. Furthermore, it has also proven convenient at times to refer to these arrangements of operations as modules or by functional names, without loss of generality.

[0052] Unless specifically stated otherwise as apparent from the above discussion, it is appreciated that throughout the description, discussions utilizing terms such as "processing" or "computing" or "calculating" or "determining" or "displaying" or the like, refer to the action and processes of a computer system, or similar electronic computing device, that manipulates and transforms data represented as physical (electronic)

quantities within the computer system memories or registers or other such information storage, transmission or display devices.

[0053] Certain aspects of the described techniques include process steps and instructions described herein in the form of an algorithm. It should be noted that the described process steps and instructions could be embodied in software, firmware or hardware, and when embodied in software, could be downloaded to reside on and be operated from different platforms used by real time network operating systems.

[0054] The present disclosure also relates to an apparatus for performing the operations herein. This apparatus may be specially constructed for the required purposes, or it may comprise a general-purpose computer selectively activated or reconfigured by a computer program stored on a computer readable medium that can be accessed by the computer. Such a computer program may be stored in a tangible computer readable storage medium, such as, but is not limited to, any type of disk including floppy disks, optical disks, CD-ROMs, magnetic-optical disks, read-only memories (ROMs), random access memories (RAMs), EPROMs, EEPROMs, magnetic or optical cards, application specific integrated circuits (ASICs), or any type of media suitable for storing electronic instructions, and each coupled to a computer system bus. Furthermore, the computers referred to in the specification may include a single processor or may be architectures employing multiple processor designs for increased computing capability.

[0055] The algorithms and operations presented herein are not inherently related to any particular computer or other apparatus. Various general-purpose systems may also be used with programs in accordance with the teachings herein, or it may prove

convenient to construct more specialized apparatuses to perform the required method steps. The required structure for a variety of these systems will be apparent to those of skill in the art, along with equivalent variations. In addition, the present disclosure is not described with reference to any particular programming language. It is appreciated that a variety of programming languages may be used to implement the teachings of the present disclosure as described herein, and any references to specific languages are provided for disclosure of enablement and best mode of the present invention.

[0056] The present disclosure is well suited to a wide variety of computer network systems over numerous topologies. Within this field, the configuration and management of large networks comprise storage devices and computers that are communicatively coupled to dissimilar computers and storage devices over a network, such as the Internet.

[0057] The foregoing description of the embodiments has been provided for purposes of illustration and description. It is not intended to be exhaustive or to limit the disclosure. Individual elements or features of a particular embodiment are generally not limited to that particular embodiment, but, where applicable, are interchangeable and can be used in a selected embodiment, even if not specifically shown or described. The same may also be varied in many ways. Such variations are not to be regarded as a departure from the disclosure, and all such modifications are intended to be included within the scope of the disclosure.

## CLAIMS

What is claimed is:

1. A computer-implemented method, comprising:
  - outputting, at a computing device including one or more processors, a media stream that includes at least one of audio data and video data;
  - during the outputting of the media stream:
    - (i) receiving, at the computing device, a first user input (a) identifying a portion of the media stream and (b) indicating a request to translate the portion of the media stream from a source language to a target language,
    - (ii) determining, at the computing device, a text corresponding to the at least one of audio data and video data from the portion of the media stream, and
    - (iii) transmitting, from the computing device, the text to a translation server via a network;
  - receiving, at the computing device, a translated text from the translation server via the network, the translated text having been translated from the source language to the target language by the translation server; and
  - outputting, at the computing device, the translated text after outputting of the media stream has completed.

2. A computer-implemented method, comprising:

receiving, at a computing device including one or more processors, a user input (i) identifying a portion of a media stream being output from the computing device and (ii) indicating a request to translate the portion of the media stream from a source language to a target language;

transmitting, from the computing device, the portion of the media stream to a translation server via a network in response to receiving the user input;

receiving, at the computing device, a translated portion of the media stream from the translation server via the network, the translated portion of the media stream having been translated from the source language to the target language by the translation server; and

outputting, at the computing device, the translated portion of the media stream.

3. The computer-implemented method of claim 2, wherein the media stream includes at least one of audio data and video data.

4. The computer-implemented method of claim 3, further comprising determining, at the computing device, a text corresponding to the at least one of audio data and video data for the portion of the media stream.

5. The computer-implemented method of claim 4, wherein determining the text includes at least one of (i) optical character recognition of subtitles in the video data for the portion of the media stream and (ii) voice recognition of the audio data for the portion of the media stream.

6. The computer-implemented method of claim 4, wherein transmitting the portion of the media stream to the translation server via the network includes transmitting, from the computing device, the text to the translation server via the network during outputting of the media stream.

7. The computer-implemented method of claim 6, wherein receiving the translated portion of the media stream includes receiving a translated text, the translated text having been translated from the source language to the target language by the translation server.

8. The computer-implemented method of claim 7, wherein outputting the translated portion of the media stream includes outputting, at the computing device, the translated text.

9. The computer-implemented method of claim 8, wherein outputting the translated text includes outputting, at the computing device, the translated text after the outputting of the media stream has completed.

10. The computer-implemented method of claim 2, wherein the portion of the media stream is a predetermined period prior to when the user input was received.

11. The computer-implemented method of claim 2, wherein the user input includes one or more keystrokes of a keyboard.

12. A computing device, comprising:

a user interface configured to receive a user input (i) identifying a portion of a media stream being output from the computing device and (ii) indicating a request to translate the portion of the media stream from a source language to a target language; and

a communication device configured to:

transmit the portion of the media stream to a translation server via a network in response to receiving the user input, and

receive a translated portion of the media stream from the translation server via the network, the translated portion of the media stream having been translated from the source language to the target language by the translation server,

wherein the user interface is also configured to output the translated portion of the media stream.

13. The computing device of claim 12, wherein the media stream includes at least one of audio data and video data.

14. The computing device of claim 13, further comprising one or more processors configured to determine a text corresponding to the at least one of audio data and video data for the portion of the media stream.

15. The computing device of claim 14, wherein the one or more processors are configured to determine the text by at least one of (i) optical character recognition of subtitles in the video data for the portion of the media stream and (ii) voice recognition of the audio data for the portion of the media stream.

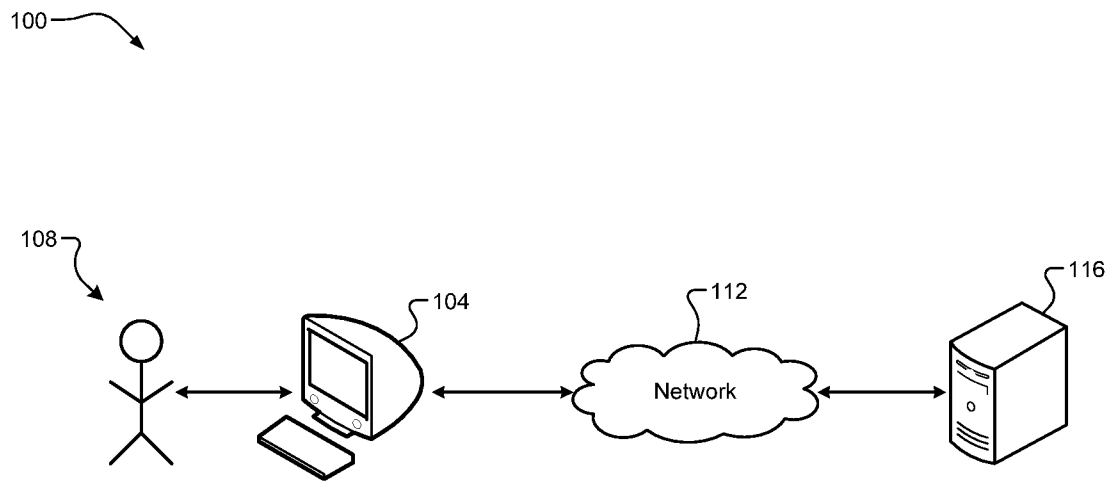
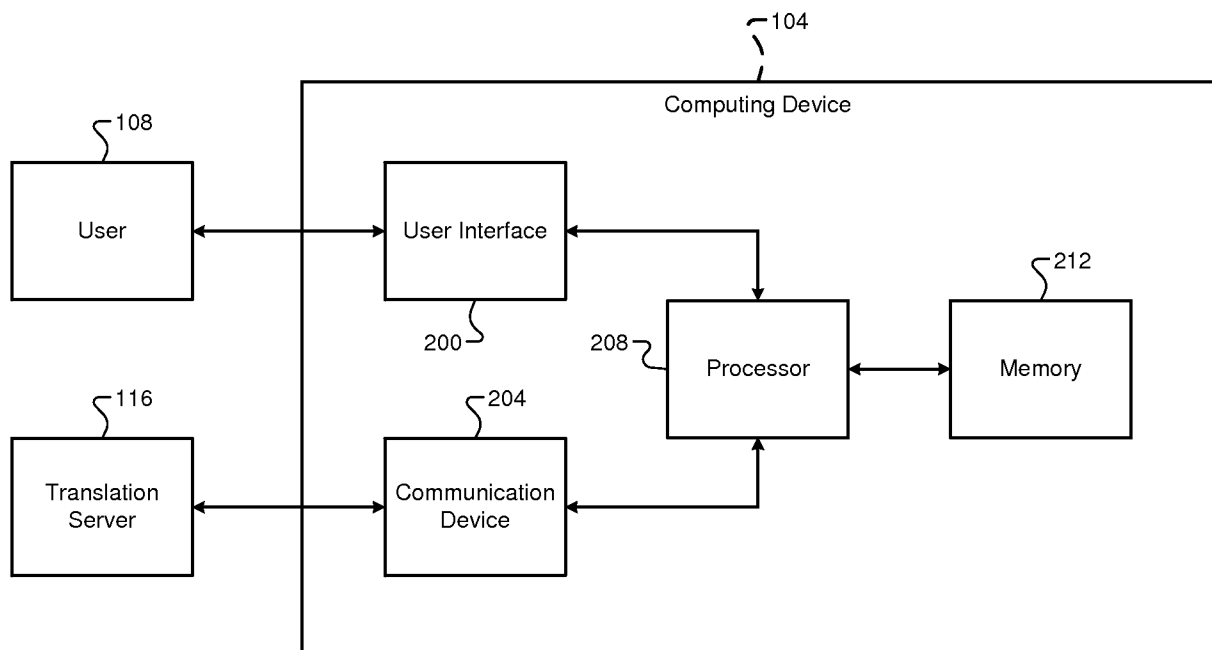
16. The computing device of claim 14, wherein the communication device is configured to transmit the portion of the media stream to the translation server via the network by transmitting the text to the translation server via the network during outputting of the media stream.

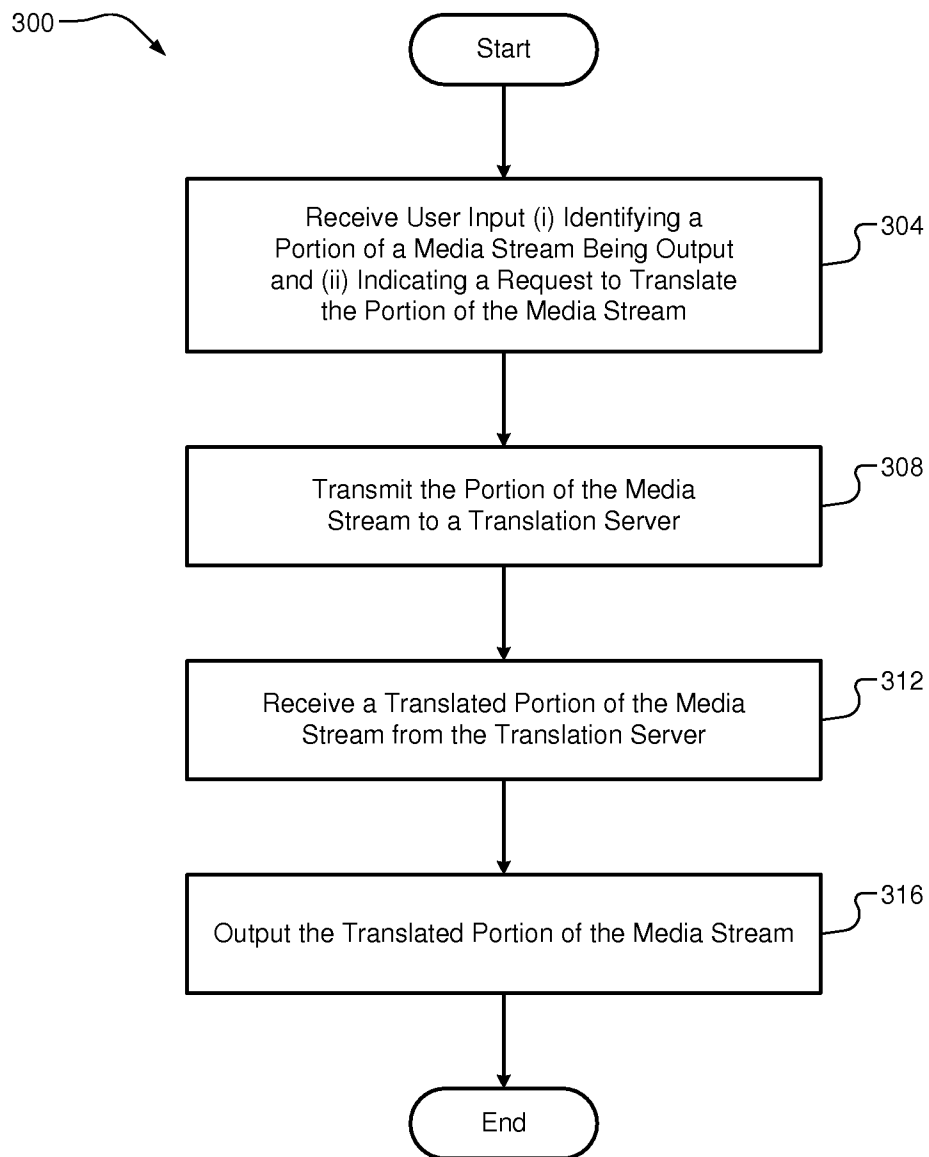
17. The computing device of claim 16, wherein the communication device is configured to receive the translated portion of the media stream by receiving a translated text, the translated text having been translated from the source language to the target language by the translation server.

18. The computing device of claim 17, wherein the user interface is configured to output the translated portion of the media stream by outputting the translated text.

19. The computing device of claim 18, wherein the user interface is configured to output the translated text by outputting the translated text after the outputting of the media stream has completed.

20. The computing device of claim 12, wherein the portion of the media stream is a predetermined period prior to when the user input was received.

**FIG. 1****FIG. 2**

**FIG. 3**

# INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2013/077161

## A. CLASSIFICATION OF SUBJECT MATTER

G06F 17/28 (2006.01) i

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

## B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC: G06F

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

CNPAT, CNKI, WPI, EPODOC, IEEE: web w (page or site), translat+, media, audio, video, flow or stream, network, server, text

## C. DOCUMENTS CONSIDERED TO BE RELEVANT

| Category* | Citation of document, with indication, where appropriate, of the relevant passages                                         | Relevant to claim No. |
|-----------|----------------------------------------------------------------------------------------------------------------------------|-----------------------|
| X         | CN 101309449 A (SHENYANG GE-SOFT CO., LTD.) 19 November 2008 (19.11.2008)<br>description, page 5, lines 10-18 and figure 1 | 2-3, 10-13, 20        |
| A         |                                                                                                                            | 1, 4-9, 14-19         |
| A         | CN 103136969 A (AGE OF LEARNING, INC.) 05 June 2013 (05.06.2013)<br>the whole document                                     | 1-20                  |
| A         | CN 102446169 A (ZHANG, Longbu) 09 May 2012 (09.05.2012)<br>the whole document                                              | 1-20                  |
| A         | US 8379072 B2 (EXHIBITONE CORPORATION) 19 February 2013 (19.02.2013)<br>the whole document                                 | 1-20                  |

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

|                                                                                                                                                                          |                                                                                                                                                                                                                                                  |
|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| * Special categories of cited documents:                                                                                                                                 | “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                                              |
| “A” document defining the general state of the art which is not considered to be of particular relevance                                                                 | “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                                                                     |
| “E” earlier application or patent but published on or after the international filing date                                                                                | “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 |
| “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) | “&” document member of the same patent family                                                                                                                                                                                                    |
| “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                                                                   |                                                                                                                                                                                                                                                  |

|                                                                                            |                                                                                        |
|--------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------|
| Date of the actual completion of the international search<br>25 February 2014 (25.02.2014) | Date of mailing of the international search report<br><b>20 Mar. 2014 (20.03.2014)</b> |
|--------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------|

Name and mailing address of the ISA/CN  
The State Intellectual Property Office, the P.R.China  
6 Xitucheng Rd., Jimen Bridge, Haidian District, Beijing, China  
100088  
Facsimile No. 86-10-62019451

Authorized officer

LIU, Tong

Telephone No. (86-10)82245501

International application No.  
PCT/CN2013/077161

Form PCT/ISA /210 (patent family annex) (July 2009)