



(51) International Patent Classification:

G06F 40/216 (2020.01) G06F 40/35 (2020.01)

G06F 40/284 (2020.01) G06F 40/45 (2020.01)

G06F 40/289 (2020.01) G06F 40/56 (2020.01)

G06F 40/30 (2020.01)

(21) International Application Number:

PCT/US2020/024541

(22) International Filing Date:

25 March 2020 (25.03.2020)

(25) Filing Language:

English

(26) Publication Language:

English

(30) Priority Data:

201910447514.4 27 May 2019 (27.05.2019) CN

(71) Applicant: MICROSOFT TECHNOLOGY LICENSING, LLC [US/US]; One Microsoft Way, Redmond, Washington 98052-6399 (US).

(72) Inventors: LIANG, Yaobo; MICROSOFT TECHNOLOGY LICENSING, LLC, One Microsoft Way, Redmond, Washington 98052-6399 (US). DUAN, Nan; MICROSOFT TECHNOLOGY LICENSING, LLC, One Microsoft Way, Redmond, Washington 98052-6399 (US). GONG, Ming; MICROSOFT TECHNOLOGY LICENSING, LLC, One Microsoft Way, Redmond, Washington 98052-6399 (US). SHOU, Linjun; MICROSOFT TECHNOLOGY LICENSING, LLC, One Microsoft Way, Redmond, Washington 98052-6399 (US). JIANG, Daxin; MICROSOFT TECHNOLOGY LICENSING, LLC, One Microsoft Way, Redmond, Washington 98052-6399 (US). ZHOU, Ming; MICROSOFT TECHNOLOGY LICENSING, LLC, One Microsoft Way, Redmond, Washington 98052-6399 (US).

(54) Title: CROSS-LINGUAL TASK TRAINING

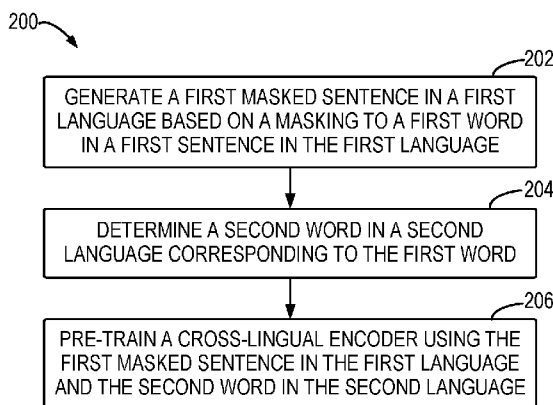


FIG. 2

(57) **Abstract:** In embodiments of the present disclosure, there is provided a cross-lingual task training method. After obtaining a training sentence in one language, a word in the training sentence is masked, and a corresponding word in another language corresponding to the masked word is obtained. The masked sentence in one language and the corresponding word in the other language are utilized to pre-train a cross-lingual model. The pre-trained cross-lingual model is capable of understanding and processing multiple languages, which may be further trained according to a specific task. In embodiments of the present disclosure, in a case that there is a large training corpus for one language while there is a small or even no training corpus for another language, the cross-lingual model may be pre-trained in a cross-lingual training manner, such that the cross-lingual model can be applied to the other language, thereby achieving model training for various languages.

(74) **Agent:** SWAIN, Cassandra T. et al.; MICROSOFT TECHNOLOGY LICENSING, LLC, One Microsoft Way, Redmond, Washington 98052-6399 (US).

(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, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, WS, 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, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

Declarations under Rule 4.17:

- *as to applicant's entitlement to apply for and be granted a patent (Rule 4.17(ii))*
- *as to the applicant's entitlement to claim the priority of the earlier application (Rule 4.17(iii))*

Published:

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

CROSS-LINGUAL TASK TRAINING

BACKGROUND

[0001] Natural language processing is a technology for processing natural human languages using a computer, which makes it possible for a computer to understand a human language. Through training by use of an artificially labeled corpus, a computer can generate semantic representations of natural languages. Natural language processing is now a trendy topic in the field of artificial intelligence, which is applicable to semantic analysis, information retrieval, machine translation, automatic answering, chatbots, and the like.

[0002] A language model is a probability distribution of a sequence of words, which is a foundation of natural language processing technology. Typically, a language model may be constructed through training with a large number of corpora. For example, a neural network-based language model may involve a three-layer feedforward neural network, which can optimize parameters in the neural network-based language model based on a back propagation technology during the training. The trained language model is capable of learning relationships between sentences or words and thus can be employed to predict the next word or sentence.

SUMMARY

[0003] In embodiments of the present disclosure, there is provided a cross-lingual task training method. After obtaining a training sentence in one language, a word in the training sentence is masked, and a corresponding word in another language corresponding to the masked word is obtained. The masked sentence in one language and the corresponding word in the other language are then utilized to pre-train a cross-lingual model. The pre-trained cross-lingual model is capable of understanding and processing multiple languages, which may be further trained according to a specific task. According to embodiments of the present disclosure, in a case that there is a large training corpus for one language while there is a small or even no training corpus for another language, the cross-lingual model may be pre-trained in a cross-lingual training manner, such that the cross-lingual model can be applied to the other language(s), thereby achieving model training for various languages.

[0004] The Summary is to introduce a selection of concepts in a simplified form that are further described below in the Detailed Description. This Summary is not intended to identify key features or essential features of the present disclosure, nor is it intended to be used to limit the scope of the present disclosure.

BRIEF DESCRIPTION OF THE DRAWINGS

[0005] Through the following detailed description with reference to the accompanying drawings, the above and other objectives, features, and advantages of example embodiments of the present disclosure will become more apparent, in which the same reference signs
5 generally refer to the same or similar elements.

[0006] Fig. 1 is a block diagram illustrating a computing device/server that may implement one or more embodiments of the present disclosure;

[0007] Fig. 2 is a flowchart illustrating a method for pre-training a cross-lingual model in accordance with embodiments of the present disclosure;

10 [0008] Fig. 3 is a schematic diagram illustrating a process of training a cross-lingual model for a specified task in accordance with embodiments of the present disclosure;

[0009] Fig. 4 is a schematic diagram illustrating an example for pre-training a cross-lingual model in accordance with embodiments of the present disclosure;

15 [0010] Fig. 5 is a schematic diagram illustrating another example for pre-training a cross-lingual model in accordance with embodiments of the present disclosure;

[0011] Fig. 6 is a schematic diagram illustrating a further example for pre-training a cross-lingual model in accordance with embodiments of the present disclosure;

[0012] Fig. 7 is a schematic diagram illustrating an example for further training the cross-lingual model in accordance with embodiments of the present disclosure;

20 [0013] Fig. 8 is a schematic diagram of an example for fine-tuning the cross-lingual model in accordance with embodiments of the present disclosure; and

[0014] Fig. 9 is a schematic diagram illustrating an example for providing a question and an answer in a search engine in accordance with embodiments of the present disclosure.

DETAILED DESCRIPTION OF EMBODIMENTS

25 [0015] Embodiments of the present disclosure will now be described in detail with reference to the drawings. Although some embodiments disclosed herein are illustrated in the drawings, it would be appreciated that the present disclosure may be implemented in various manners and should not be limited to the embodiments described herein. Rather, those embodiments are provided merely to describe thoroughly and completely the present
30 disclosure. It would be appreciated that the drawings and embodiments are provided only as examples and are not intended for limiting the scope of the present disclosure in any manner.

[0016] As used herein, the term “comprise” and its variants are to be read as open-ended terms that mean “comprise, but not limited to.” The term “based on” is to be read as “based

at least in part on.” The term “an embodiment” is to be read as “at least one example embodiment;” the term “another embodiment” is to be read as “at least one further embodiment;” and the term “some embodiments” is to be read as “at least some embodiments.” Relevant definitions of other terms will be given in the following description.

[0017] Traditionally, to train a language model for one language often requires a large training corpus for this language to complete the training of the language model. However, for some tasks, not every language is provided with a large corpus. In the presence of a small training corpus or even in the absence of a training corpus, language models of particular tasks for a lot of languages cannot be trained.

[0018] In view of this, embodiments of the present disclosure provide a cross-lingual task training method. In accordance with embodiments of the present disclosure, in a case that there is a large training corpus for one language while there is a small or even no training corpus for another language, a cross-lingual model may be pre-trained in a cross-lingual training manner, such that the cross-lingual model can be applied to other languages. As such, model training for various languages can be implemented. Thus, embodiments of the present disclosure propose training in the cross-lingual training manner, and thus can obtain a multilingual model applicable to multiple languages without requiring training corpora in multiple languages.

[0019] Basic principles and several example implementations of the present disclosure will now be described with reference to Figs. 1-9. Fig. 1 is a block diagram illustrating a computing device/server 100 that may implement therein one or more embodiments of the present disclosure. It would be appreciated that the computing device/server 100 as shown in Fig. 1 is merely provided as an example, rather than constituting any limitation to the functionalities and scope of implementations as described herein.

[0020] As shown in Fig. 1, the computing device/server 100 is in the form of a general computing device. Components of the computing device/server 100 may include, but are not limited to, one or more processors or processing units 110, a memory 120, a storage device 130, one or more communication units 140, one or more input devices 150, and one or more output devices 160. The processing unit 110 may be any physical or virtual processor and can perform various processing based on programs stored in the memory 120. In a multi-processor system, multiple processing units execute computer-executable instructions in parallel to improve the parallel processing capacity of the computing device/server 100.

[0021] The computing device/server 100 typically includes a plurality of computer storage media, which may be any available media accessible by the computing device/server 100, including, but not limited to, volatile and non-volatile media, and removable and non-removable media. The memory 120 may be a volatile memory (for example, a register, cache, Random Access Memory (RAM)), non-volatile memory (for example, a Read-Only Memory (ROM), Electrically Erasable Programmable Read-Only Memory (EEPROM), flash memory), or any combination thereof. The storage device 130 may be any removable or non-removable medium, and may include a machine-readable medium, such as a flash drive, disk or any other medium, which can be used for storing information and/or data (for example, training data for training) and accessed in the computing device/server 100.

[0022] The computing device/server 100 may further include additional removable/non-removable, volatile/non-volatile memory media. Although not shown in Fig. 1, a disk drive may be provided for reading or writing a removable and non-volatile disk (for example, a floppy disk) and a disc drive may be provided for reading or writing a removable non-volatile disc. In these cases, each drive may be connected to the bus (not shown) via one or more data medium interfaces. The memory 120 may include a computer program product 125 having one or more program modules which are configured to perform the method or acts of various embodiments of the present disclosure.

[0023] The communication unit 140 communicates with a further computing device via communication media. In addition, functionalities of components in the computing device/server 100 may be implemented by a single computing cluster or multiple computing machines connected communicatively for communication. Therefore, the computing device/server 100 may be operated in a networking environment using a logical link with one or more other servers, network personal computers (PCs) or another general network node.

[0024] The input device 150 may include one or more input devices, such as a mouse, keyboard, tracking ball, and the like. The output device 160 may include one or more output devices, such as a display, loudspeaker, printer and the like. As required, the computing device/server 100 may also communicate via the communication unit 140 with one or more external devices (not shown) such as a storage device, display device and the like, one or more devices that enable users to interact with the computing device/server 100, or any devices that enable the computing device/server 100 to communicate with one or more other computing devices (for example, a network card, modem, and the like). Such communication may be performed via an input/output (I/O) interface (not shown).

[0025] As shown in Fig. 1, the storage device 130 may be provided with the training corpus for training a cross-lingual model. The program product 125 may implement cross-lingual model pre-training and fine-tuning based on the training corpus stored in the storage device. The cross-lingual model trained in accordance with embodiments of the present disclosure is applicable to a plurality of multi-lingual tasks. For example, a Chinese question answering model may be trained in the presence of a quite small Chinese question answering corpus. It should be appreciated that, although English is used as the first language while Chinese is used as the second language in the following embodiments, the second language may be other languages, such as French, German, Japanese, and the like. In addition, the first language is a language with a great amount of training data, which may be a language other than English.

[0026] Fig. 2 is a flowchart illustrating a method 200 for pre-training a cross-lingual model in accordance with embodiments of the present disclosure. It should be appreciated that the method 200 may be executed by the computing device/server 100 as described with reference to Fig. 1. Reference will be made with reference to the example 400 for pre-training a cross-lingual model as shown in Fig. 4 to describe the method 200 of Fig. 2.

[0027] At 202, a first masked sentence in a first language is generated based on a masking to a first word in a first sentence in the first language. In some embodiments, the first language, for example, may be English. Referring to Fig. 4, the first sentence may be “This is an example”. English sentences may act as training sentences sourced from an English encyclopedia, English books, and the like. Continuing to refer to Fig. 4, the word “example” in the above sentence is masked and turned into “[MASK],” so as to obtain a masked sentence 420 “This is a [MASK]”, where the mask conveys an occluded or sheltered meaning. In the example of Fig. 4, it is assumed that there is a great amount of training data in English (the first language) while few or even no training data are provided in Chinese (the second language).

[0028] At 204, a second word in the second language corresponding to the first word is determined, and the first language and the second language are different from each other. In the example of Fig. 4, the second language, for example, may be Chinese, and the Chinese word 450 “例子” corresponds to the masked English word “example.” In some embodiments, the Chinese word 450 may be a word that has been labeled manually. Alternatively, the word 450 may be obtained by translating the masked English word “example” using a translation system. It should be appreciated that various translation

systems, platforms, programs, interfaces, and the like, currently existing or to be developed in the future may be used in combination with embodiments of the present disclosure.

[0029] At 206, the first masked sentence in the first language and the second word in the second language are used to pre-train a cross-lingual encoder, and the cross-lingual encoder will be further trained based on a specific task. As shown in Fig. 4, the English masked sentence 420 “This is a [MASK]” may be used as an input to the cross-lingual model 410, and the Chinese word 450 “例子” may act as an output of the cross-lingual model 410, for pre-training the cross-lingual model, wherein the cross-lingual model includes a cross-lingual encoder, or a cross-lingual encoder and decoder. Therefore, the cross-lingual model trained in accordance with embodiments of the present disclosure is capable of understanding and processing multiple languages.

[0030] It should be appreciated that, in embodiments of the present disclosure, the term “pre-train” means that there are at least two stages of training processes. At the first stage which is the pre-training process, general training data are utilized to train the cross-lingual model, such that the cross-lingual model can have a multi-lingual understanding and processing capability. At the second stage which is a further training and/or fine-tuning process, the cross-lingual model is further trained for a particular task such that the trained cross-lingual model can process the particular task, such as a question answering task, a natural language inference task, and the like.

[0031] Fig. 3 is a schematic diagram of a process 300 of training a cross-lingual model for a particular task in accordance with embodiments of the present disclosure. As shown in Fig. 3, to train a cross-lingual model 340 for a particular task (such as a question answering task), a training process including three stages 310-330 may be performed. Any machine learning and/or neural network technology currently known or to be developed may be used in combination with embodiments of the present disclosure to implement the cross-lingual model 340. For example, the cross-lingual model 340 may be implemented through a Transformer Model, Recurrent Neural Network (RNN), or Long Short-Term Memory (LSTM) Network.

[0032] At the cross-lingual pre-training stage 310, pre-training may be performed based on general training data 305 which is not related to any particular task, and the cross-lingual model may be trained through at least two languages such that the cross-lingual model can understand and process at least two languages. Several different example training modes of the cross-lingual pre-training stage 310 are described with reference to Figs. 4-6.

[0033] At the cross-lingual task-related training stage 320, the cross-lingual training may be further performed based on the task-related training data 325 such that the trained cross-lingual model is capable of processing the cross-lingual task, and the task-related training data 325, for example, may include a question answering corpus in a certain language.

5 Reference will be made to Fig. 7 to describe an example training mode of the cross-lingual task-related training stage 320.

[0034] At the single-language task-related fine-tuning stage 330, the single-language fine-tuning may be performed based on task-related training data 325 such that the trained cross-lingual model 340 for the particular task has a task processing capability for the target language. By the fine-tuning of the single language, the task processing capability of the trained model for the targeted language can be further improved. For example, if the target language is Chinese, the cross-lingual model may be fine-tuned based on Chinese questions and answers. Reference will be made to Fig. 8 to describe below an example training mode of the single-language task-related fine-tuning stage 330.

15 **[0035]** Fig. 4 is a schematic diagram of an example 400 for pre-training a cross-lingual model in accordance with embodiments of the present disclosure. Fig. 4 shows a cross-lingual word alignment prediction process, in which the English word “example” in the sentence “This is an example” is masked, so as to generate an English masked sentence 420 “This is an [MASK].” Then, the English masked sentence 420 as token embedding, position information 430 of each word as position embedding, and language information 440 as language embedding, are input together into the cross-lingual model 410, and the Chinese word 450 “例子” is used as an output, for implementing the pre-training of the cross-lingual model 410.

25 **[0036]** In some embodiments, the cross-lingual model 410 may be of attention mechanism-based encoder-decoder architecture. As shown in Fig. 4, a large quantity of English masked sentences and respective Chinese words in pair may be used to train the cross-lingual model 410 such that the cross-lingual model can understand and process both English and Chinese. Moreover, the cross-lingual model 410 of Fig. 4 may be a bi-directional encoder representation. In some embodiments, if the subsequent model task is related to French, French may be used as the output 450 to pre-train the cross-lingual model 410.

30 **[0037]** Fig. 5 is a schematic diagram of another example 500 for pre-training a cross-lingual model in accordance with embodiments of the present disclosure. Fig. 5 illustrates

a cross-lingual interpretation classification process, in which the input of the model is a combination of the English sentence “This is an example” and the Chinese sentence “这是一个例子”, and its output is a token, for example, “TRUE” or “FALSE,” which indicates whether the English sentence and the Chinese sentence convey the same meaning. As shown in Fig. 5, the combination 520 of the English sentence “This is an example” and the Chinese sentence “这是一个例子” as token embedding, position information 530 of each word in the respective sentence as position embedding, and language information 540 as language embedding are input together into the cross-lingual model 410, and a token 550 (for example, “TRUE”) is used as an output, for implementing the pre-training of the cross-lingual model 410. It would be appreciated that the pre-training method of the example 500 is different from the traditional training method for a translation system, because the sentences in two languages in the example 500 are both inputs as the model inputs into the cross-lingual model, rather than generating an output in one language based on an input in another language.

[0038] Fig. 6 is a schematic diagram illustrating a further example 600 for pre-training a cross-lingual model in accordance with embodiments of the present disclosure. Fig. 6 illustrates a cross-lingual subsequent sentence prediction process, in which the model input is a combination of an English sentence “This is an example” and a Chinese sentence “明天出发去纽约,” and its output is a token, for example, “TRUE” or “FALSE,” which indicates whether the subsequent sentence is a next one of the preceding sentence in semantics. As shown in Fig. 6, the combination 620 of the English sentence “This is an example” and the Chinese sentence “明天出发去纽约” as token embedding, position information 630 of each word in the respective sentence as position embedding, and language information 640 as language embedding, are input together into the cross-lingual model 410, and the token 650 (for example “TRUE”) is used as an output, for implementing the pre-training of the cross-lingual model 410. It would be appreciated that, although Figs. 4-6 only show a respective example in each training mode, training in each training mode involves a great number of training data.

[0039] Fig. 7 is a schematic diagram of an example 700 for further training the cross-lingual model in accordance with embodiments of the present disclosure. As shown in Fig. 7, after completing the pre-training of the cross-lingual model 410, a further task-related

training may be performed according to the specific task. For example, Fig. 7 illustrates an example 700 for training a question answering task. Since there is only a large English question answering corpus, and there is lacking a sufficient Chinese question answering corpus, a pair of question and answer is acquired first from the English question answering corpus, for example, an English question “Why is the sky blue” and an English answer “The atmosphere of the Earth scatters blue light most.” Then, a translation system may be utilized to translate the English answer “The atmosphere of the Earth scatters blue light most” into a Chinese answer “地球大气散射蓝光最多” automatically. As shown in Fig. 7, the English question “Why is the sky blue” and the Chinese answer “地球大气散射蓝光最多” may be used as an input 720, and then the cross-lingual model 410 generates input embedding 730 for each token and further generates a context representation 740 of each token. In the example 700, the output 750 is a “TRUE” or “FALSE” token which indicates whether the English question matches the Chinese answer.

[0040] Alternatively, or in addition, the English question may be translated into a Chinese question when keeping the English answer unchanged, and then the Chinese question and the English answer may be used together to train the cross-lingual model 410, for use in the question answering in Chinese.

[0041] Fig. 8 is a schematic diagram illustrating an example 800 for fine-tuning the cross-lingual model in accordance with embodiments of the present disclosure. As shown in Fig. 8, after completing the task-related training of the cross-lingual model 410 in Fig. 7, task-related fine-tuning may be performed according to the specific task. For example, Fig. 8 illustrates an example 800 for training a question answering task. Since there is a large English question answering corpus only, and there is lacking a sufficient Chinese question answering corpus, a pair of question and answer is obtained first from the English question answering corpus, for example, an English question “Why is the sky blue” and an English answer “The atmosphere of the Earth scatters blue light most.” Then, a translation system may be utilized to translate the English question and the English answer into the Chinese question “为什么天空是蓝的” and the Chinese answer “地球大气散射蓝光最多,” respectively. The Chinese question and the Chinese answer are used as an input 820, and then the cross-lingual model 410 generates input embedding 830 of each token and further generates a context representation 840 of each token. In the example 700, the output 850 is a token of “TRUE” or “FALSE,” which indicates whether the Chinese question matches

the Chinese answer. In other words, at the fine-tuning stage, a question and an answer in a target language may be used to fine-tune the cross-lingual model for question answering.

[0042] Fig. 9 is a schematic diagram of an example 900 for providing a question and an answer in a search engine in accordance with embodiments of the present disclosure. For example, the above cross-lingual model trained with reference to Fig. 8 may be used to process Chinese questions and answers. After the search engine has received a Chinese query 910, the cross-lingual model determines a Chinese answer 920 and presents the Chinese answer 920 at a search result page in the search engine. In this way, even in the absence of the Chinese question answering corpus, a question answering model for Chinese can still be trained and provided. It would be appreciated that the Chinese is merely provided herein as an example of the target language, and any other language may be used as the target language in combination with embodiments of the present disclosure.

[0043] In some embodiments, a pair of sentences may be obtained from an English natural language inference library, and the pair of sentences includes an English premise sentence and an English hypothetical sentence, wherein a relation between the English premise sentence and an English hypothetical sentence is marked with entailment, contradiction or neutrality. For example, the premise sentence “You don’t have to stay there” and the hypothetical sentence “You can leave” are in an entailment relation. Optionally, an English hypothetical sentence may be translated from English into Chinese, and an English premise sentence, a Chinese hypothetical sentence and a sentence relation may be used to train a cross-lingual model, for use in natural language inference in Chinese. Alternatively, the English premise sentence may be translated into Chinese. For example, the English premise sentence “You don’t have to stay there” is translated into Chinese “你不需要留在那里,” and then the Chinese premise sentence “你不需要留在那里,” the English

hypothetical sentence “You can leave,” and the sentence relation (such as entailment) may be used to train the cross-lingual model, for use in natural language interference in Chinese.

[0044] Therefore, in accordance with embodiments of the present disclosure, in a case that there is a training corpus for one language while there is no training corpus for another language, a cross-lingual model may be pre-trained in a cross-lingual training manner such that the cross-lingual model can be applied to the other language(s). In this way, model training for various languages can be implemented.

[0045] The methods and functionalities described herein can be performed, at least in part, by one or more hardware logic components. For example, and without limitation,

illustrative types of hardware logic components that can be used include Field-Programmable Gate Arrays (FPGAs), Application-specific Integrated Circuits (ASICs), Application-specific Standard Products (ASSPs), System-on-a-chip systems (SOCs), Complex Programmable Logic Devices (CPLDs), and the like.

5 **[0046]** Program code for carrying out methods of the present disclosure may be written in any combination of one or more programming languages. These program codes may be provided to a processor or controller of a general purpose computer, special purpose computer, or other programmable data processing devices, such that the program codes, when executed by the processor or controller, cause the functions/operations specified in the
10 flowcharts and/or block diagrams to be implemented. The program code may execute entirely on a machine, partly on the machine, as a stand-alone software package, partly on the machine and partly on a remote machine or entirely on the remote machine or server.

[0047] In the context of this disclosure, a machine-readable medium may be any tangible medium that may contain, or store a program for use by or in connection with an instruction
15 execution system, apparatus, or device. The machine-readable medium may be a machine-readable signal medium or a machine-readable storage medium. A machine-readable medium may include, but not limited to, an electronic, magnetic, optical, electromagnetic, infrared, or semiconductor system, apparatus, or device, or any suitable combination of the foregoing. More specific examples of the machine-readable storage medium would
20 include an electrical connection having one or more wires, a portable computer diskette, a hard disk, a random access memory (RAM), a read-only memory (ROM), an erasable programmable read-only memory (EPROM or Flash memory), an optical fiber, a portable compact disc read-only memory (CD-ROM), an optical storage device, a magnetic storage device, or any suitable combination of the foregoing.

25 **[0048]** Further, while operations are depicted in a particular order, this should not be understood as requiring that such operations be performed in the particular order shown or in sequential order, or that all illustrated operations be performed, to achieve desirable results. In certain circumstances, multitasking and parallel processing may be advantageous. Likewise, while several specific implementation details are contained in
30 the above discussions, these should not be construed as limitations on the scope of the present disclosure, but rather as descriptions of features that may be specific to particular implementations. Certain features that are described in the context of separate implementations may also be implemented in combination in a single implementation. Conversely, various features that are described in the context of a single implementation

may also be implemented in multiple implementations separately or in any suitable sub-combination.

[0049] Some example implementations of the present disclosure will be given below.

[0050] In one aspect, there is provided a computer-implemented method. The method comprises generating a first masked sentence in a first language based on a masking to a first word in a first sentence in the first language. The method further comprises determining a second word in a second language corresponding to the first word, wherein the first language is different from the second language. The method further comprises pre-training a cross-lingual encoder using the first masked sentence in the first language and a second word in the second language, wherein the cross-lingual encoder will be further trained based on a specific task.

[0051] In some embodiments, wherein pre-training the cross-lingual encoder comprises: determining the first masked sentence in the first language as a first input; determining the second word in the second language as a first output; and pre-training the cross-lingual encoder using the first input and the first output.

[0052] In some embodiments, the method further comprises: obtaining a second sentence in the second language; combining the first sentence in the first language and the second sentence in the second language to obtain a second input; and pre-training the cross-lingual encoder using the second input and a second output, wherein the second output comprises a token indicating whether the first sentence and the second sentence have a same meaning.

[0053] In some embodiments, the method further comprises: obtaining a third sentence in the second language; combining the first sentence in the first language and the third sentence in the second language to obtain a third input; and pre-training the cross-lingual encoder using the third input and a third output, wherein the third output comprises a token indicating whether the third sentence is a next sentence of the first sentence in semantics.

[0054] In some embodiments, the method further comprises: after completing the pre-training of the cross-lingual encoder, obtaining a first pair of question and answer from a question answering corpus in the first language, wherein the first pair of question and answer comprises a first question and a first answer in the first language; and performing at least one of the following: translating the first answer from the first language into the second language, and training the cross-lingual encoder using the first question in the first language and the first answer in the second language, for use in question answering in the second language, and translating the first question from the first language into the second language, and training the cross-lingual encoder using the first question in the second language and

the first answer in the first language, for use in question answering in the second language.

[0055] In some embodiments, the method further comprises: obtaining a second pair of question and answer from the question answering corpus, wherein the second pair of question and answer comprises a second question and a second answer in the first language;
5 translating the second question and the second answer from the first language into the second language; and fine-tuning the cross-lingual encoder using the second question in the second language and the second answer in the second language.

[0056] In some embodiments, the method further comprises: in response to receiving, in a search engine, a user query in the second language: determining an answer in the second
10 language to the user query using the cross-lingual encoder; and presenting the answer in the second language in a search result page of the search engine.

[0057] In some embodiments, the method comprises: obtaining a first pair of sentences from a natural language inference library in the first language, wherein the first pair of sentences comprises a first premise sentence and a first hypothetical sentence in the first
15 language, the first premise sentence and the first hypothetical sentence are labeled with one of the following relations: entailment, contradiction and neutrality; and performing at least one of the following: translating the first hypothetical sentence from the first language into the second language, and training the cross-lingual encoder using the first premise sentence in the first language, the first hypothetical sentence in the second language and the relation,
20 for use in inference for the second language, and translating the first premise sentence from the first language into the second language, and training the cross-lingual encoder using the first premise sentence in the second language, the first hypothetical sentence in the first language and the relation, for use in inference for the second language.

[0058] In some embodiments, the method further comprises: determining a third word in
25 a third language corresponding to the first word; and pre-training the cross-lingual encoder using the first masked sentence in the first language and the third word in the third language, for use in a task associated with the third language.

[0059] In a further aspect, there is provided an electronic device. The electronic device comprises a processing unit and a memory coupled to the processing unit and storing
30 instructions. The instructions, when executed by the processing unit, perform acts comprising: generating a first masked sentence in a first language based on a masking to a first word in a first sentence in the first language; determining a second word in a second language corresponding to the first word, wherein the first language is different from the second language; and pre-training a cross-lingual encoder using the first masked sentence

in the first language and a second word in the second language, wherein the cross-lingual encoder will be further trained based on a specific task.

[0060] In some embodiments, wherein pre-training the cross-lingual encoder comprises: determining the first masked sentence in the first language as a first input; determining the
5 second word in the second language as a first output; and pre-training the cross-lingual encoder using the first input and the first output.

[0061] In some embodiments, the acts further comprise: obtaining a second sentence in the second language; combining the first sentence in the first language and the second sentence in the second language to obtain a second input; and pre-training the cross-lingual
10 encoder using the second input and a second output, wherein the second output comprises a token indicating whether the first sentence and the second sentence have a same meaning.

[0062] In some embodiments, the acts further comprise: obtaining a third sentence in the second language; combining the first sentence in the first language and the third sentence in the second language to obtain a third input; and pre-training the cross-lingual encoder using
15 the third input and a third output, wherein the third output comprises a token indicating whether the third sentence is a next sentence of the first sentence in semantics.

[0063] In some embodiments, the acts further comprise: after completing the pre-training of the cross-lingual encoder, obtaining a first pair of question and answer from a question answering corpus in the first language, wherein the first pair of question and answer
20 comprises a first question and a first answer in the first language; and performing at least one of the following: translating the first answer from the first language into the second language, and training the cross-lingual encoder using the first question in the first language and the first answer in the second language, for use in question answering in the second language, and translating the first question from the first language into the second language,
25 and training the cross-lingual encoder using the first question in the second language and the first answer in the first language, for use in question answering in the second language.

[0064] In some embodiments, the acts further comprise: obtaining a second pair of question and answer from the question answering corpus, wherein the second pair of question and answer comprises a second question and a second answer in the first language;
30 translating the second question and the second answer from the first language into the second language; and fine-tuning the cross-lingual encoder using the second question in the second language and the second answer in the second language.

[0065] In some embodiments, the acts further comprise: in response to receiving, in a search engine, a user query in the second language: determining an answer in the second

language to the user query using the cross-lingual encoder; and presenting the answer in the second language in a search result page of the search engine.

[0066] In some embodiments, the acts further comprise: obtaining a first pair of sentences from a natural language inference library in the first language, wherein the first pair of sentences comprise a first premise sentence and a first hypothetical sentence in the first language, and the first premise sentence and the first hypothetical sentence are labeled with one of the following relations: entailment, contradiction and neutrality; and performing at least one of the following: translating the first hypothetical sentence from the first language into the second language, and training the cross-lingual encoder using the first premise sentence in the first language, the first hypothetical sentence in the second language and the relation, for use in inference for the second language, and translating the first premise sentence from the first language to the second language, and training the cross-lingual encoder using the first premise sentence in the second language, the first hypothetical sentence in the first language and the relation, for use in inference for the second language.

[0067] In some embodiments, the acts further comprise: determining a third word in a third language corresponding to the first word; and pre-training the cross-lingual encoder using the first masked sentence in the first language and the third word in the third language, for use in a task associated with the third language.

[0068] In a further aspect, there is provided a computer program product. The computer program product being stored in a computer storage medium and including machine-executable instructions. The machine-executable instructions, when running in a device, cause the device to perform acts comprising: generating a first masked sentence in a first language based on a masking to a first word in a first sentence in the first language; determining a second word in a second language corresponding to the first word, wherein the first language is different from the second language; and pre-training a cross-lingual encoder using the first masked sentence in the first language and a second word in the second language, wherein the cross-lingual encoder will be further trained based on a specific task.

[0069] In some embodiments, wherein pre-training the cross-lingual encoder comprises: determining the first masked sentence in the first language as a first input; determining the second word in the second language as a first output; and pre-training the cross-lingual encoder using the first input and the first output.

[0070] In some embodiments, the machine-executable instructions further cause the device to perform acts comprising: obtaining a second sentence in the second language; combining the first sentence in the first language and the second sentence in the second

language to obtain a second input; and pre-training the cross-lingual encoder using the second input and a second output, wherein the second output comprises a token indicating whether the first sentence and the second sentence have a same meaning.

[0071] In some embodiments, the machine-executable instructions further cause the device to perform acts comprising: obtaining a third sentence in the second language; combining the first sentence in the first language and the third sentence in the second language to obtain a third input; and pre-training the cross-lingual encoder using the third input and a third output, wherein the third output comprises a token indicating whether the third sentence is a next sentence of the first sentence in semantics.

[0072] In some embodiments, the machine-executable instructions further cause the device to perform acts comprising: after completing the pre-training of the cross-lingual encoder, obtaining a first pair of question and answer from a question answering corpus in the first language, wherein the first pair of question and answer comprises a first question and a first answer in the first language; and performing at least one of the following: translating the first answer from the first language into the second language, and training the cross-lingual encoder using the first question in the first language and the first answer in the second language, for use in question answering in the second language, and translating the first question from the first language into the second language, and training the cross-lingual encoder using the first question in the second language and the first answer in the first language, for use in question answering in the second language.

[0073] In some embodiments, the machine-executable instructions further cause the device to perform acts comprising: obtaining a second pair of question and answer from the question answering corpus, wherein the second pair of question and answer comprises a second question and a second answer in the first language; translating the second question and the second answer from the first language into the second language; and fine-tuning the cross-lingual encoder using the second question in the second language and the second answer in the second language.

[0074] In some embodiments, the machine-executable instructions further cause the device to perform acts comprising: in response to receiving, in a search engine, a user query in the second language: determining an answer in the second language to the user query using the cross-lingual encoder; and presenting the answer in the second language in a search result page of the search engine.

[0075] In some embodiments, the machine-executable instructions further cause the device to perform acts comprising: obtaining a first pair of sentences from a natural language

inference library in the first language, wherein the first pair of sentences comprises a first premise sentence and a first hypothetical sentence in the first language, and the first premise sentence and the first hypothetical sentence are labeled with one of the following relations: entailment, contradiction and neutrality; and performing at least one of the following:

5 translating the first hypothetical sentence from the first language into the second language, and training the cross-lingual encoder using the first premise sentence in the first language, the first hypothetical sentence in the second language and the relation, for use in inference for the second language, and translating the first premise sentence from the first language to the second language, and training the cross-lingual encoder using the first premise

10 sentence in the second language, the first hypothetical sentence in the first language and the relation, for use in inference for the second language.

[0076] In some embodiments, the machine-executable instructions further cause the device to perform acts comprising: determining a third word in a third language corresponding to the first word; and pre-training the cross-lingual encoder using the first

15 masked sentence in the first language and the third word in the third language, for use in a task associated with the third language.

[0077] Although the present disclosure has been described in language specific to structural features and/or methodological acts, it is to be understood that the subject matter specified in the appended claims is not necessarily limited to the specific features or acts

20 described above. Rather, the specific features and acts described above are disclosed as example forms of implementing the claims.

CLAIMS

1. A computer-implemented method, comprising:
generating a first masked sentence in a first language based on a masking to a first word in a first sentence in the first language;
determining a second word in a second language corresponding to the first word, the first language being different from the second language; and
pre-training a cross-lingual encoder using the first masked sentence in the first language and the second word in the second language, the cross-lingual encoder being further trained based on a specific task.
2. The method of claim 1, wherein pre-training the cross-lingual encoder comprises:
determining the first masked sentence in the first language as a first input;
determining the second word in the second language as a first output; and
pre-training the cross-lingual encoder using the first input and the first output.
3. The method of claim 1, further comprising:
obtaining a second sentence in the second language;
combining the first sentence in the first language and the second sentence in the second language to obtain a second input; and
pre-training the cross-lingual encoder using the second input and a second output, the second output comprising a token indicating whether the first sentence and the second sentence have a same meaning.
4. The method of claim 1, further comprising:
obtaining a third sentence in the second language;
combining the first sentence in the first language and the third sentence in the second language to obtain a third input; and
pre-training the cross-lingual encoder using the third input and a third output, the third output comprising a token indicating whether the third sentence is a next sentence of the first sentence in semantics.
5. The method of claim 1, further comprising:
after completing the pre-training of the cross-lingual encoder, obtaining a first pair of question and answer from a question answering corpus in the first language, the first pair of question and answer comprising a first question and a first answer in the first language; and
performing at least one of the following:
translating the first answer from the first language into the second language,

and training the cross-lingual encoder using the first question in the first language and the first answer in the second language, for use in question answering in the second language, and

translating the first question from the first language into the second language, and training the cross-lingual encoder using the first question in the second language and the first answer in the first language, for use in question answering in the second language.

6. The method of claim 5, further comprising:

obtaining a second pair of question and answer from the question answering corpus, the second pair of question and answer comprising a second question and a second answer in the first language;

translating the second question and the second answer from the first language into the second language; and

fine-tuning the cross-lingual encoder using the second question in the second language and the second answer in the second language.

7. The method of claim 6, further comprising:

in response to receiving, in a search engine, a user query in the second language:

determining an answer in the second language to the user query using the cross-lingual encoder; and

presenting the answer in the second language in a search result page of the search engine.

8. The method of claim 1, further comprising:

obtaining a first pair of sentences from a natural language inference library in the first language, the first pair of sentences comprising a first premise sentence and a first hypothetical sentence in the first language, the first premise sentence and the first hypothetical sentence being labeled with one of the following relations: entailment, contradiction and neutrality; and

performing at least one of the following:

translating the first hypothetical sentence from the first language into the second language, and training the cross-lingual encoder using the first premise sentence in the first language, the first hypothetical sentence in the second language and the relation, for use in inference in the second language, and

translating the first premise sentence from the first language into the second language, and training the cross-lingual encoder using the first premise sentence in

the second language, the first hypothetical sentence in the first language and the relation, for use in inference in the second language.

9. The method of claim 1, further comprising:

determining a third word in a third language corresponding to the first word; and
pre-training the cross-lingual encoder using the first masked sentence in the first language and the third word in the third language, for use in a task associated with the third language.

10. An electronic device, comprising:

a processing unit; and

a memory coupled to the processing unit and storing instructions, the instructions, when executed by the processing unit, perform acts comprising:

generating a first masked sentence in a first language based on a masking to a first word in a first sentence in the first language;

determining a second word in a second language corresponding to the first word, the first language being different from the second language; and

pre-training a cross-lingual encoder using the first masked sentence in the first language and the second word in the second language, the cross-lingual encoder being further trained based on a specific task.

11. The device of claim 10, wherein pre-training the cross-lingual encoder comprises:

determining the first masked sentence in the first language as a first input;

determining the second word in the second language as a first output; and

pre-training the cross-lingual encoder using the first input and the first output.

12. The device of claim 10, the acts further comprising:

obtaining a second sentence in the second language;

combining the first sentence in the first language and the second sentence in the second language to obtain a second input; and

pre-training the cross-lingual encoder using the second input and a second output, the second output comprising a token indicating whether the first sentence and the second sentence have a same meaning.

13. The device of claim 10, the acts further comprising:

obtaining a third sentence in the second language;

combining the first sentence in the first language and the third sentence in the second language to obtain a third input; and

pre-training the cross-lingual encoder using the third input and a third output, the

third output comprising a token indicating whether the third sentence is a next sentence of the first sentence in semantics.

14. The device of claim 10, the acts further comprising:

after completing the pre-training of the cross-lingual encoder, obtaining a first pair of question and answer from a question answering corpus in the first language, the first pair of question and answer comprising a first question and a first answer in the first language; and

performing at least one of the following:

translating the first answer from the first language into the second language, and training the cross-lingual encoder using the first question in the first language and the first answer in the second language, for use in question answering in the second language, and

translating the first question from the first language into the second language, and training the cross-lingual encoder using the first question in the second language and the first answer in the first language, for use in question answering in the second language.

15. A computer program product being stored in a computer storage medium and including machine-executable instructions which, when running in a device, cause the device to perform acts comprising:

generating a first masked sentence in a first language based on a masking to a first word in a first sentence in the first language;

determining a second word in a second language corresponding to the first word, the first language being different from the second language; and

pre-training a cross-lingual encoder using the first masked sentence in the first language and the second word in the second language, the cross-lingual encoder being further trained based on a specific task.

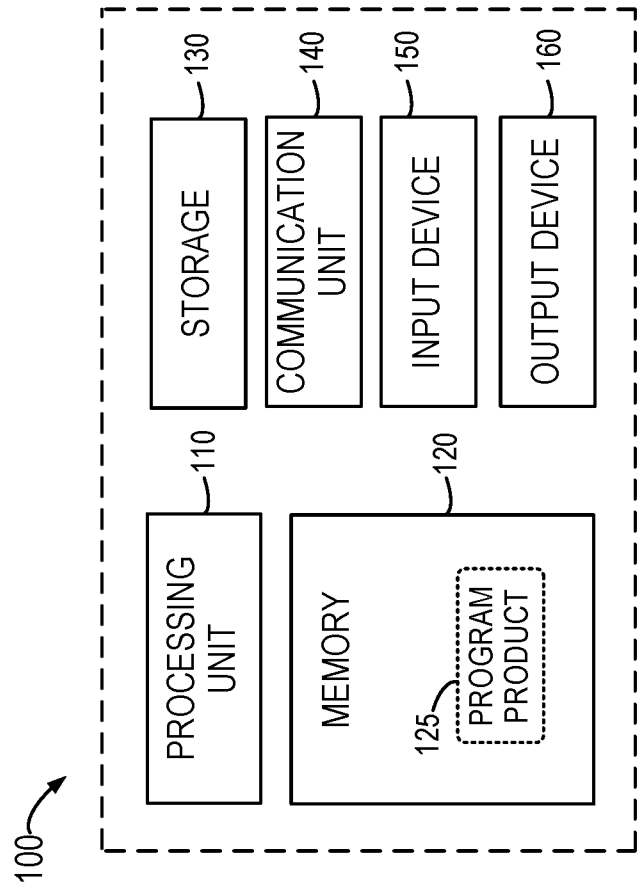


FIG. 1

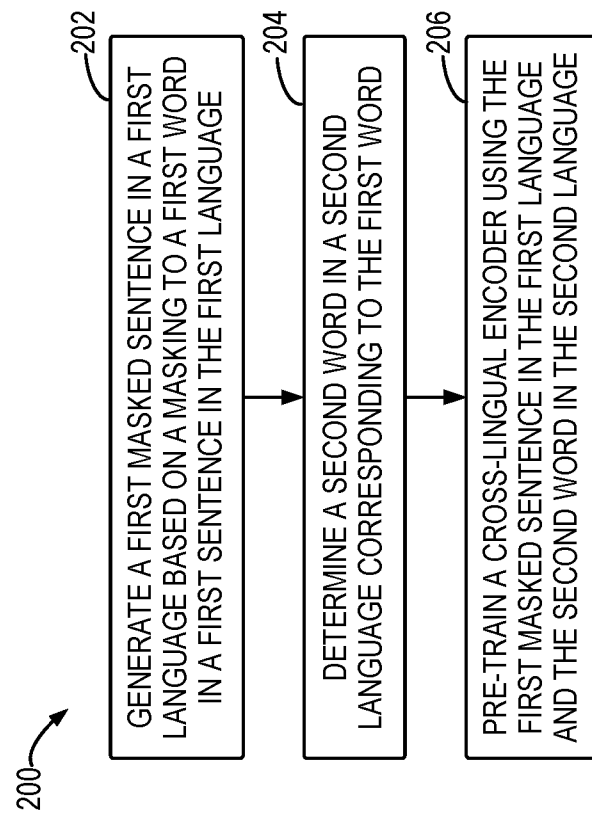


FIG. 2

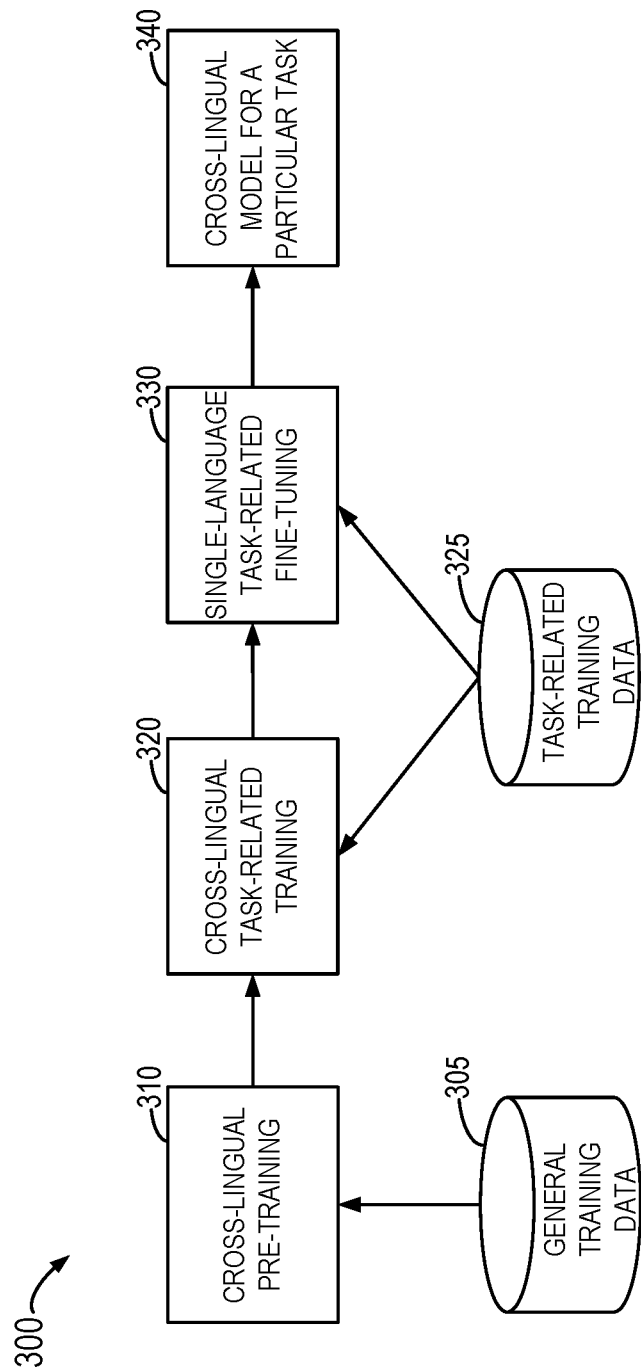


FIG. 3

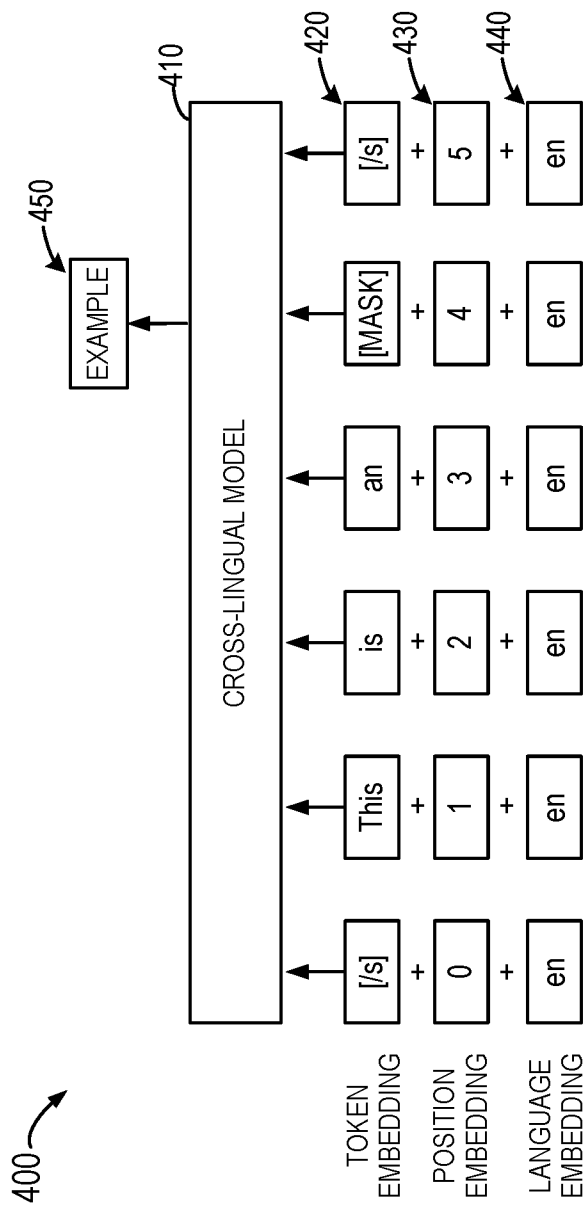


FIG. 4

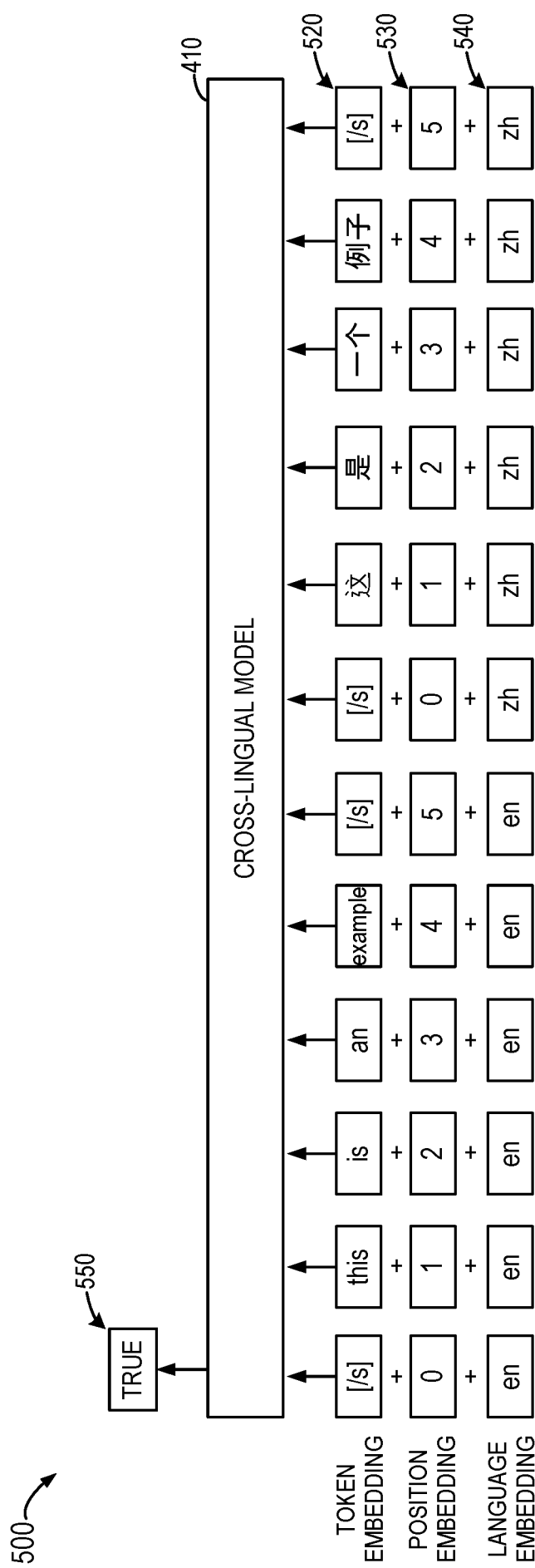


FIG. 5

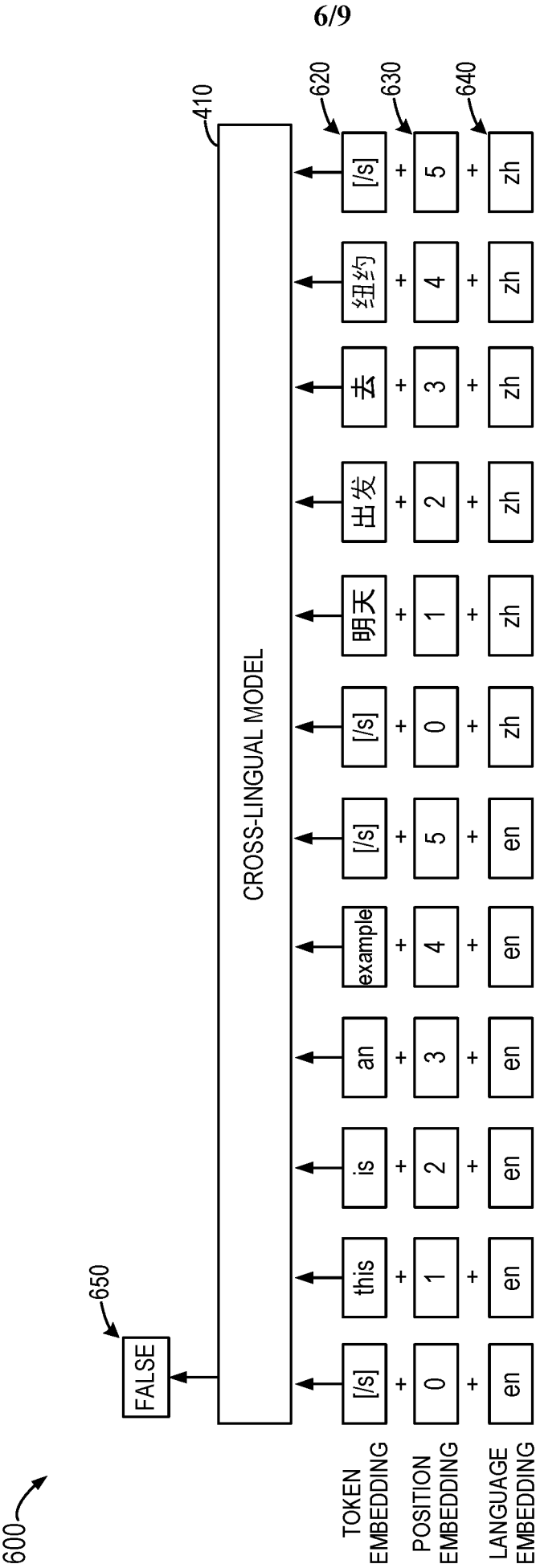


FIG. 6

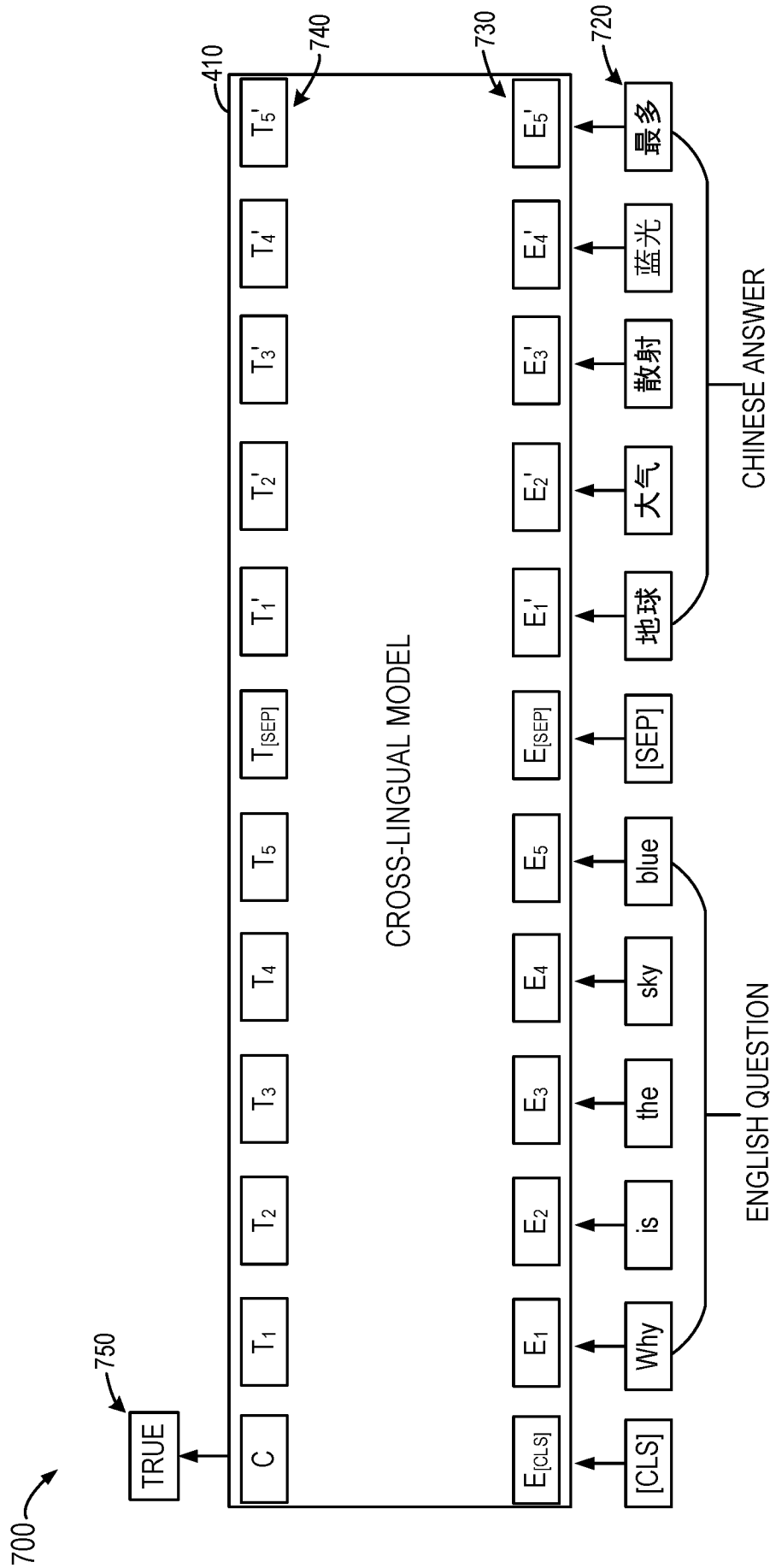


FIG. 7

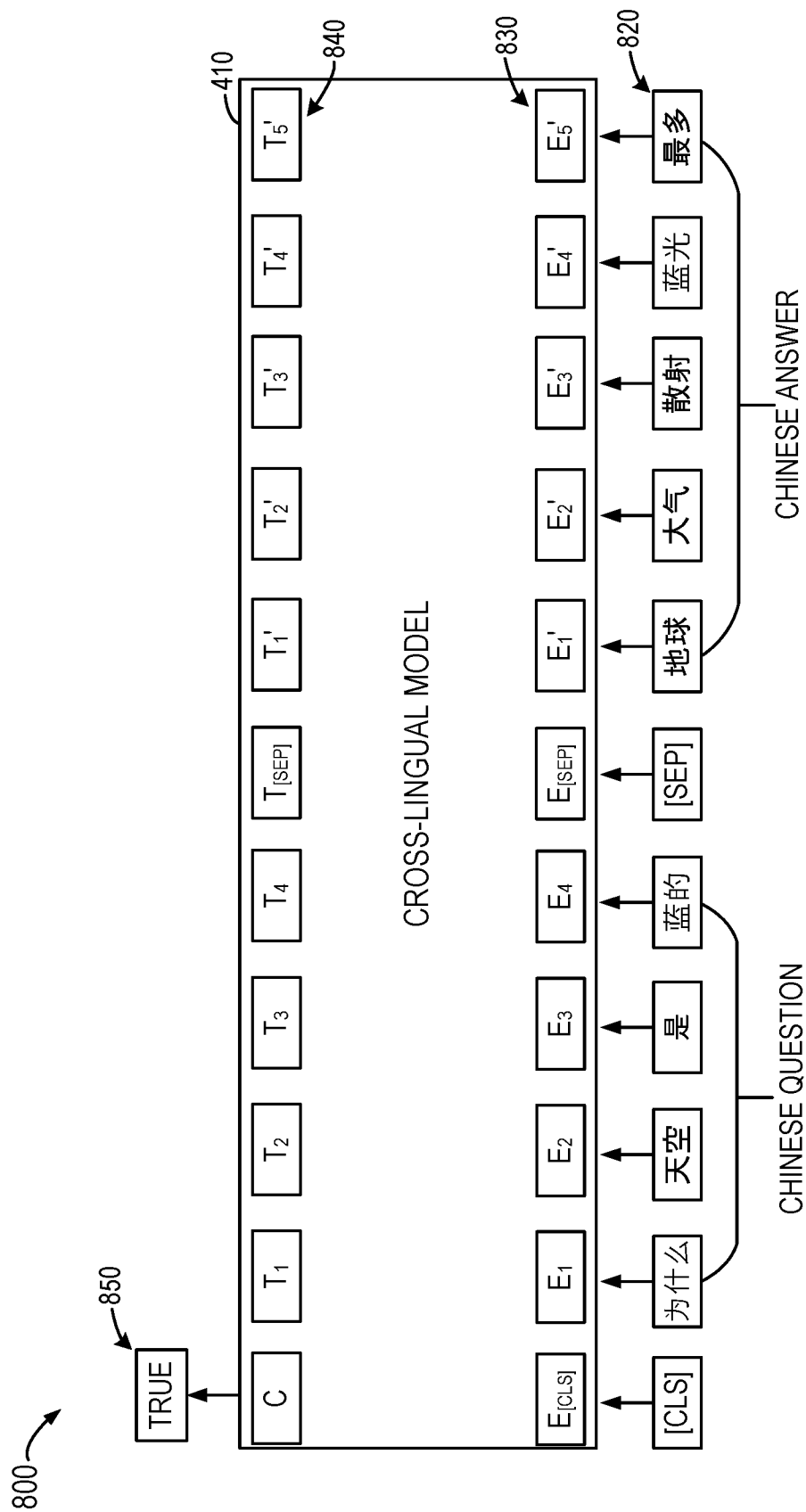


FIG. 8

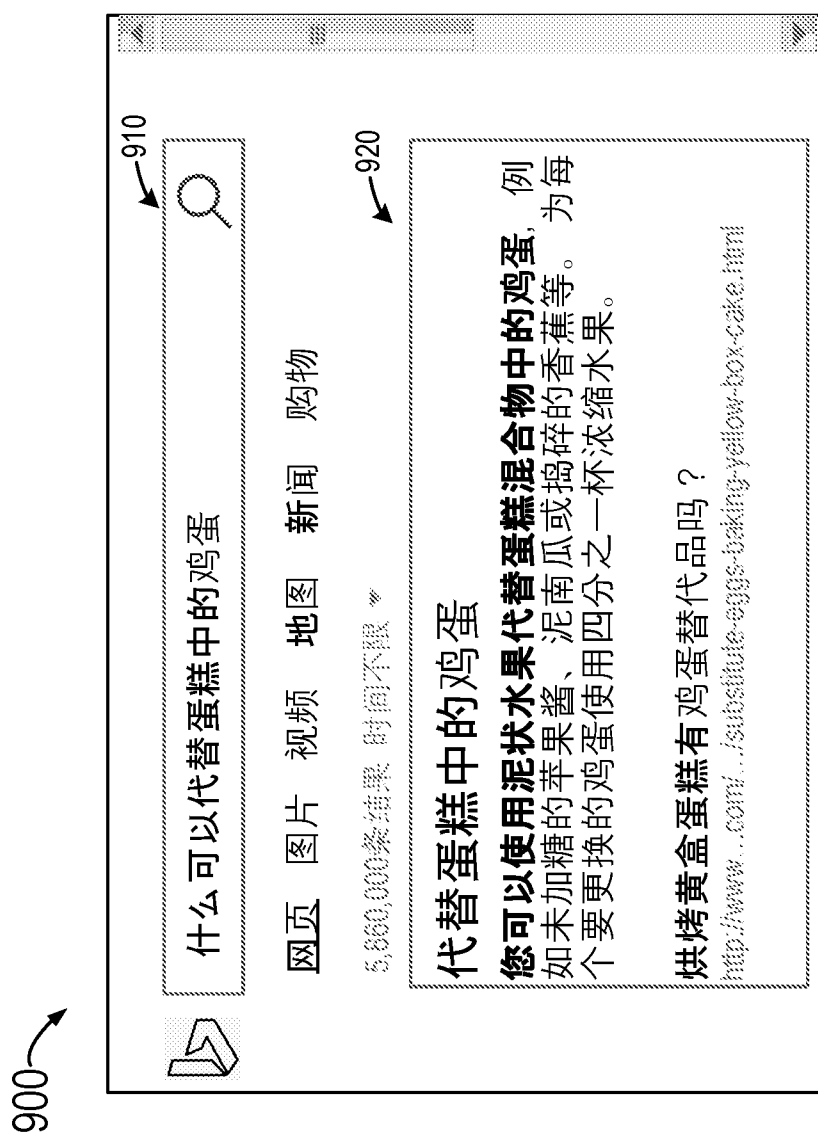


FIG. 9

INTERNATIONAL SEARCH REPORT

International application No
PCT/US2020/024541

A. CLASSIFICATION OF SUBJECT MATTER

INV. G06F40/216 G06F40/284 G06F40/289 G06F40/30 G06F40/35
G06F40/45 G06F40/56

ADD.

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)

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)

EPO-Internal, WPI Data

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X,P	HAOYANG HUANG ET AL: "Unicoder: A Universal Language Encoder by Pre-training with Multiple Cross-lingual Tasks", ARXIV.ORG, CORNELL UNIVERSITY LIBRARY, 201 OLIN LIBRARY CORNELL UNIVERSITY ITHACA, NY 14853, 3 September 2019 (2019-09-03), XP081472256, the whole document	1-15
X	----- GUILLAUME LAMPLE ET AL: "Cross-lingual Language Model Pretraining", ARXIV.ORG, CORNELL UNIVERSITY LIBRARY, 201 OLIN LIBRARY CORNELL UNIVERSITY ITHACA, NY 14853, 22 January 2019 (2019-01-22), XP081006616, the whole document ----- -/-	1-15



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 June 2020

Date of mailing of the international search report

09/07/2020

Name and mailing address of the ISA/

European Patent Office, P.B. 5818 Patentlaan 2
NL - 2280 HV Rijswijk
Tel. (+31-70) 340-2040,
Fax: (+31-70) 340-3016

Authorized officer

Stauch, Marc

INTERNATIONAL SEARCH REPORT

International application No
PCT/US2020/024541

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	JACOB DEVLIN ET AL: "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding", ARXIV.ORG, CORNELL UNIVERSITY LIBRARY, 201 OLIN LIBRARY CORNELL UNIVERSITY ITHACA, NY 14853, 11 October 2018 (2018-10-11), XP081064287, the whole document	1-15
A	----- Rani Horev: "XLM - Enhancing BERT for Cross-lingual Language Model Cross-lingual Language Model Pretraining", 12 February 2019 (2019-02-12), XP055709742, Retrieved from the Internet: URL:https://towardsdatascience.com/xlm-enhancing-bert-for-cross-lingual-language-model-5aeed9e6f14b [retrieved on 2020-06-29] the whole document	1,10,15
A	----- XIAODONG LIU ET AL: "Improving Multi-Task Deep Neural Networks via Knowledge Distillation for Natural Language Understanding", ARXIV.ORG, CORNELL UNIVERSITY LIBRARY, 201 OLIN LIBRARY CORNELL UNIVERSITY ITHACA, NY 14853, 20 April 2019 (2019-04-20), XP081171813, abstract	1,10,15
A	----- OLIVER ADAMS ET AL: "Cross-Lingual Word Embeddings for Low-Resource Language Modeling", PROCEEDINGS OF THE 15TH CONFERENCE OF THE EUROPEAN CHAPTER OF THE ASSOCIATION FOR COMPUTATIONAL LINGUISTICS: VOLUME 1, LONG PAPERS, 2016, pages 937-947, XP055709653, Stroudsburg, PA, USA DOI: 10.18653/v1/E17-1088 abstract	1,10,15
A	----- MIKEL ARTETXE ET AL: "Massively Multilingual Sentence Embeddings for Zero-Shot Cross-Lingual Transfer and Beyond", ARXIV.ORG, CORNELL UNIVERSITY LIBRARY, 201 OLIN LIBRARY CORNELL UNIVERSITY ITHACA, NY 14853, 26 December 2018 (2018-12-26), XP081493212, abstract	1,10,15
