



(51) International Patent Classification:
G06F 17/28 (2006.01)

(21) International Application Number:
PCT/US2010/051120

(22) International Filing Date:
1 October 2010 (01.10.2010)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
12/572,021 1 October 2009 (01.10.2009) US

(71) Applicant (for all designated States except US): **LANGUAGE WEAVER** [US/US]; Suite 150, 6060 Center Drive, Los Angeles, California 90045 (US).

(72) Inventors; and

(75) Inventors/Applicants (for US only): **SORICUT, Radu** [RO/US]; Suite 150, 6060 Center Drive, Los Angeles, California 90045 (US). **VISWANATHAN, Narayanaswamy** [US/US]; Suite 150, 6060 Center Drive, Los Angeles, California 90045 (US). **MARCU, Daniel** [CA/US]; Suite 150, 6060 Center Drive, Los Angeles, California 90045 (US).

(74) Agents: **BACHMANN, Stephen** et al.; Carr & Ferrell LLP, 120 Constitution Drive, Menlo Park, California 94025 (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, 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, KM, 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, PE, PG, PH, PL, PT, RO, RS, RU, 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, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, 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, ML, MR, NE, SN, TD, TG).

Published:

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

(54) Title: PROVIDING MACHINE-GENERATED TRANSLATIONS AND CORRESPONDING TRUST LEVELS

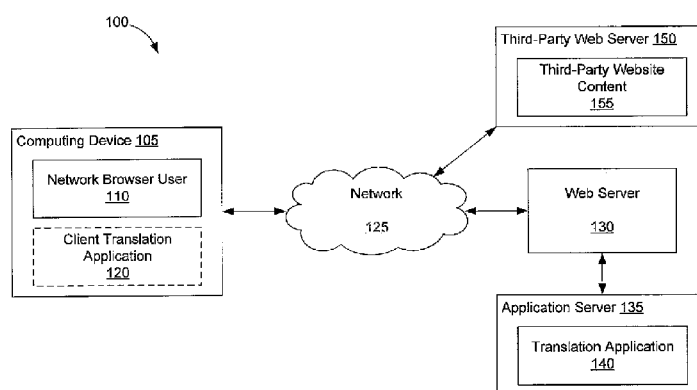


FIGURE 1

(57) Abstract: A quality-prediction engine predicts a trust level associated with translational accuracy of a machine-generated translation. Training a quality-prediction may include translating a document in a source language to a target language by executing a machine-translation engine stored in memory to obtain a machine-generated translation. The training may further include comparing the machine-generated translation with a human-generated translation of the document. Additionally, the training may include generating a mapping between features of the machine-generated translation and features of the human-generated translation based on the comparison. Machine-generated translations may then be credibly provided by translating a document from a source language to a target language by executing a machine-translation engine stored in memory to obtain a machine-generated translation predicting a trust level of the machine-generated translation by executing a quality-prediction engine stored in memory, and outputting the machine-generated translation and the trust level.



**PROVIDING MACHINE-GENERATED TRANSLATIONS
AND CORRESPONDING TRUST LEVELS**

By: Radu Soricut, Narayanaswamy Viswanathan, and Daniel Marcu

BACKGROUND OF THE INVENTION

Field of the Invention

[0001] The present invention relates generally to natural language translation. More specifically, the present invention relates to providing machine-generated translations and corresponding trust levels.

Related Art

[0002] Machine translation involves use of computer systems to translate text or speech from one natural language to another. Using corpus techniques, more complex translations can be achieved relative to simple word substitution approaches. Parallel corpora or other training datasets may be used to train, or effectively 'teach,' a machine translation engine to translate between two languages, thus allowing for better handling of differences in linguistic typology, phrase recognition, translation of idioms, and isolation of anomalies.

[0003] Presently, machine-generated translations are provided without any quantified assurance of translational accuracy. Without any assurance, machine translation users may unknowingly risk sending and receiving misinformation to contacts, clients, customers, colleagues, and so forth. In order for a consumer to obtain such assurance of translation accuracy for a given machine-generated translation, the users must either possess some degree of familiarity with the source and target languages, rely on another individual with that familiarity, or obtain a human-generated translation for

comparison with the machine-generated translation. In all of these cases, human expertise is necessitated. Counter to the objective of machine translation, limited supply of human expertise therefore still can hamper efficient and effective dissemination of information across language barriers.

[0004] In some machine translation systems, feedback associated with translational accuracy can be provided for improving those systems, but that feedback is not useful for machine translation users that need an indication of translational accuracy before sending or when receiving a translation. Such feedback is requested and provided subsequent to translations being provided and is often on a voluntary basis rendering availability of this feedback undependable. In addition, multiple individuals with varying levels of fluency in the pertinent languages may provide the feedback. As such, an accuracy metric or rating scale determined by multiple individuals is nearly impossible to standardize. Furthermore, feedback may not be available for some translated information due, for example, to sensitivity of that information. Therefore, there is a need for machine-generated translations to be provided concurrently with an indication of translational accuracy, without human involvement.

SUMMARY OF THE INVENTION

[0005] Embodiments of the present technology allow a machine-generated translation to be provided in conjunction with a corresponding trust level that is associated with translational accuracy of the machine-generated translation.

[0006] In one claimed embodiment, a method for training a quality-prediction engine is disclosed. The method may include translating a document in a source language to a target language by executing a machine-translation engine stored in memory to obtain a machine-generated translation. The method can further include comparing the machine-generated translation with a human-generated translation of the document. The human-generated translation is in the target language. Additionally, the method may include generating a mapping between features of the machine-generated translation and features of the human-generated translation based on the comparison. The mapping may allow determination of trust levels associated with translational accuracy of future machine-generated translations that lack corresponding human-generated translations.

[0007] Another claimed embodiment discloses a system for training a quality-prediction engine. The system may include a machine-translation engine, a feature-comparison module, and a mapping module, all of which may be stored in memory and executed by a processor to effectuate the respective functionalities attributed thereto. The machine-translation engine may be executed to translate a document in a source language to a target language to obtain a machine-generated translation. The feature-comparison module may be executed to compare the machine-generated translation with a human-generated translation of the document. The human-generated translation is in the target language. The mapping module can be executed to generate a mapping between features of the machine-generated translation and features of the human-generated translation based on the comparison.

The mapping may allow determination of trust levels associated with translational accuracy of future machine-generated translations that lack corresponding human-generated translations.

[0008] A computer readable storage medium having a program embodied thereon is also disclosed as a claimed embodiment. The program is executable by a processor to perform a method for training a quality-prediction engine. The method may include translating a document in a source language to a target language using a machine-translation engine to obtain a machine-generated translation, comparing the machine-generated translation with a human-generated translation of the document, and generating a mapping between features of the machine-generated translation and features of the human-generated translation based on the comparison. The human-generated translation is in the target language. The mapping allows determination of trust levels associated with translational accuracy of future machine-generated translations that lack corresponding human-generated translations.

[0010] In yet another claimed embodiment, a method for credibly providing machine-generated translations is disclosed. The method can include translating a document from a source language to a target language by executing a machine-translation engine stored in memory to obtain a machine-generated translation, predicting a trust level of the machine-generated translation by executing a quality-prediction engine stored in memory, and outputting the machine-generated translation and the trust level. The trust level is associated with translational accuracy of the machine-generated translation.

[0011] A further claimed embodiment discloses a system for credibly providing machine-generated translations. The system may include a machine-translation engine, a communications engine, and a quality-prediction engine, all of which can be stored in memory and executed by a

processor to effectuate the respective functionalities attributed thereto. The machine-translation engine can be executed to translate a document from a source language to a target language to obtain a machine-generated translation. The quality-prediction engine can be executed to predict a trust level of the machine-generated translation. The trust level is associated with translational accuracy of the machine-generated translation. The communications engine may be executed to output the machine-generated translation and the trust level. The communications engine may be communicatively coupled with the machine-translation engine and the quality-prediction engine.

[0012] Still another claimed embodiment discloses a computer readable storage medium having a program embodied thereon. The program is executable by process to perform a method for credibly providing machine-generated translations. The method may include translating a document from a source language to a target language using a machine-translation engine to obtain a machine-generated translation, predicting a trust level of the machine-generated translation using a quality-prediction engine, and outputting the machine-generated translation and the trust level. The trust level is associated with translational accuracy of the machine-generated translation.

BRIEF DESCRIPTION OF THE DRAWINGS

[0013] FIG. 1 is a block diagram of an exemplary environment for practicing embodiments of the present technology.

[0014] FIG. 2 is a block diagram of an exemplary translation application invoked in the environment depicted in FIG. 1.

[0015] FIG. 3 is a block diagram of an exemplary quality-prediction engine included in the translation application.

[0016] FIG. 4 is a flowchart of an exemplary method for training a quality-prediction engine.

[0017] FIG. 5 is a flowchart of an exemplary method for credibly providing machine-generated translations.

[0018] FIG. 6 illustrates an exemplary computing system that may be used to implement an embodiment of the present technology.

DETAILED DESCRIPTION OF EXEMPLARY EMBODIMENTS

[0019] The present technology allows a trust level to be predicted for machine-generated translations. More specifically, given a machine-generated translation in a target language of a document in source language, a trust level associated with translational accuracy can be predicted and provided along with the machine-generated translation. Such a document can include any amount of text ranging, for example, from a few words to a batch of textual items such as websites, books, articles, or letters. The trust level may be presented in a number of manners such as on a scale between one and five, or a star rating scale. The trust level can be predicted without a human-generated translation of the document or any other human intervention. Since both the machine-generated translation and the corresponding trust-level prediction can be provided contemporaneously, it is immediately conveyed how much trust can be placed in the machine-generated translation without an understanding of the source language and/or the target language being necessary. Trust-level predictions can be provided in real-time and as a batch when several documents are processed together.

[0020] It is noteworthy that machine-generated translations obtained by way of statistical-translation techniques and non-statistical-translation techniques fall within the scope of the present technology. Furthermore, while the present technology is described herein in the context of textual translations, the principles disclosed can likewise be applied to speech translations such as when employed in conjunction with speech recognition technologies.

[0021] Referring now to FIG. 1, a block diagram of an exemplary environment 100 is shown in which embodiments of the present technology can be practiced. As depicted, the environment 100 includes a computing device 105 providing a network browser 110 and optionally a client

translation application 120, a web server 130, an application server 135 providing a translation application 140, and a third-party web server 150 providing third-party website content 155. Communication between the computing device 105, the web server 130, and the third-party web server 150 is provided by a network 125. Examples of the network 125 include a wide area network (WAN), local area network (LAN), the Internet, an intranet, a public network, a private network, a combination of these, or some other data transfer network. Examples of the computing device 105 include a desktop personal computer (PC), a laptop PC, a pocket PC, a personal digital assistant (PDA), a smart phone, a cellular phone, a portable translation device, and so on. The web server 130, the application server 135, and the third-party web server 150 may each be implemented as one or more servers. An exemplary computing system for implementing the computing device 105, the web server 130, the application server 135, and the third-party web server 150 is described in further detail in connection with FIG. 6. Additionally, other various components (not depicted) that are not necessary for describing the present technology may also be included in the environment 100, in accordance with exemplary embodiments.

[0022] As mentioned, the computing device 105 may include the network browser 110. The network browser 110 may retrieve, present, traverse, and otherwise process information located on a network, including content pages. For example, network browser 110 can be implemented as a web browser that can process a content page in the form of a web page. The network browser 110 may provide an interface as part of a content page or web page. The interface can be implemented from content page data received from the third-party web server 150 or the web server 130. Via the interface, the computing device 105 can receive an indication from a user to provide a translation from a source language to a target language along with a trust-level prediction of that translation. The user may provide the indication via the document itself,

location data for the document such as a link (*e.g.*, URL) associated with the document, or other information. The indication may convey a desire to obtain a highly accurate translation based on content included in or associated with the document. The indication may be forwarded either to the third-party website 155 or the web server 130 via the network 125.

[0023] The computing device 105, as depicted in FIG. 1, can include the client translation application 120. The client translation application 120 may be a stand-alone executable application residing and executing, at least in part, on the computing device 105. The client translation application 120 may also provide an interface for selecting content to have translated. The client translation application 120 may communicate directly with the web server 130, the application server 135, or the third-party web server 150. In the description herein, it is intended that any functionality performed translation application 140, including providing an interface for implementing various functionality, can also be implanted by the client translation application 120. In some embodiments, client translation application 120 may be implemented in place of translation application 140, which is indicated by the dashed lines comprising the client translation application 120 in FIG. 1.

[0024] The web server 130 may communicate both with the application server 135 and over the network 125, for example to provide content page data to the computing device 105 for rendering in the network browser 110. The content page data may be used by the network browser 110 to provide an interface for selecting an indication of a document to translate, whether stored over a network or locally to the computing device 105. The web server 130 can also receive data associated with an indication from the computing device 105. The web server 130 may process the received indication and/or provide the indication, and optionally any document data, to the application server 135 for processing by translation application 140.

[0025] The application server 135 communicates with web server 130 and other applications, for example the client translation applications 120, and includes the translation application 140. The translation application 140 can generate a translated version of a document and a trust-level prediction associated therewith, as discussed in further detail herein. The translated document and the trust-level prediction may be transmitted to a user over the network 125 by the application server 135 and the web server 130, for example, through the computing device 105.

[0026] The translation application 140 may be part of a translation system that translates documents and predicts a trust level corresponding to the translated documents. A trust level may be presented on a numeric scale (*e.g.*, 1 through 5), a term-based scale (*e.g.*, poor through excellent), a star-rating scale (*e.g.*, one star through five stars), an analog scale (*e.g.*, a dial or meter), and so forth. Generally speaking, the translation application 140 receives an indication, such as via the network browser 110, to translate a document from a source language to a target language and to provide a quality prediction. The translation application 140 then accesses the document and translates the document by way of executing a machine-translation engine. Based on the translated document and other information discussed herein, the translation engine 140 predicts a trust level associated with the translational accuracy of the translated document. The translation and the trust level returned to the user such as via the computing device 105. The translation application 140 is described in further detail in connection with FIG. 2. Furthermore, although the translation application 140 is depicted as being a single component of the environment 100, it is noteworthy that the translation application 140 and constituent elements thereof may be distributed across several computing devices that operate in concert via the network 125.

[0027] In some embodiments, a content page for allowing a user to configure translation parameters can be provided through the network

browser 110. The translation configuration content page data can be provided to the network browser 110 by the web server 130 and/or by the third-party web server 150. When provided by the third-party web server 150, the web server 150 may access and retrieve information from the translation system (*i.e.*, the web server 130 and/or the application server 135) to provide a content page having an interface for configuring. In exemplary embodiments, the translation application 140 is accessed by the web server 150. A graphical user interface (GUI) may be implemented within a content page by the web server 150, rendered in the network browser 110, and accessed by a user via the network browser 110 of the computing device 105. According to exemplary embodiments, the GUI can enable a user to identify a document to be translated and select various options related to translating the documents. [0028] According to some exemplary embodiments, the web server 150 may not necessarily provide a translation configuration content page but, instead, may provide content pages containing text. As such, a content page provided by web server 150 may itself comprise a document to be translated. That is, a user may view a webpage in a source language (*e.g.*, English or French) through the network browser 110 from a content page received from the web server 150. The user may provide input to subsequently view the webpage in a different language (*e.g.*, Spanish). The translation application 140 may access and translate the text provided within the content page, predict a trust level of the translation, and return a translated version and trust-level prediction to the network browser 110 or the web server 150 in accordance with embodiments of the present technology.

[0029] FIG. 2 is a block diagram of an exemplary translation application 140 invoked in the environment 100. The translation application 140, as depicted, includes a communications module 205, an interface module 210, a translation engine 215, and a quality-prediction engine 220. Although FIG. 2 depicts one translation engine 215, the translation application 140 may comprise any

number of translation engines and may be in communication with other translation engines via the network 125. The translation engine 215 is associated with the training dataset 225. The training dataset 225 may or may not be included in the translation application 140. Programs comprising engines and modules of the translation application 140 may be stored in memory of a computing system such as the computing device 105, the web server 130, the application server 135, the third-party web server 150, or any computing device that includes the translation application 140. Additionally, the constituent engines and modules can be executed by a processor of a computing system to effectuate respective functionalities attributed thereto. It is noteworthy that the translation application 140 can be composed of more or fewer modules and engines (or combinations of the same) and still fall within the scope of the present technology. For example, the functionalities of the communications module 205 and the functionalities of the interface module 210 may be combined into a single module or engine.

[0030] When executed, the communications module 205 allows an indication to be received via a user interface to provide a translation of a document from a source language to a target language, as well as a prediction of a trust level of the translation. Such a user interface may include the network browser 110 or a GUI provided by the third-party website content 155. The communications module 205 may also facilitate accessing the document to be translated such as in response to an indication by a user. The document can be accessed based on location information associated with the document. Additionally, the document can be downloaded from the computing device 105, third-party website server 150, or any other site or device accessible via the network 125. Furthermore, the communications module 205 can be executed such that a translated document and an associated trust level is outputted from the translation application 140 to devices accessible via the network 125 (e.g., the computing device 105).

[0031] The interface module 210 can be executed to provide a graphical user interface through network browser 110, for example as a content page, that enables a user to request the translation and corresponding trust-level prediction. The graphical user interface may also provide various options to a user relating to, for example, pricing or translation domain. According to various embodiments, the graphical user interface may be presented to a user as a content page for network browser 110 via the third-party web server 150 or directly by client translation application 120 at the computing device 105.

[0032] The translation engine 215 comprises a machine translation engine capable of translating from a source language to a target language. Such translation capability may result from training the translation engine 215 on various training data. Higher translation accuracy may be achieved for domain-specific translations when a machine translation engine is trained using a training dataset associated with the same domain or similar subject matter as documents being translated. For example, a translation of a car-repair manual may be of higher quality if the machine translation engine employed was trained using a car-repair-domain-specific training dataset compared to, say, a general training dataset or an unrelated-domain-specific training dataset. In some embodiments, the translation application 140 may include more than one translation engine 215. Additionally, the translation engine 215 may be based on statistical-translation techniques, non-statistical-translation techniques, or a combination thereof.

[0033] As depicted in FIG. 2, the translation engines 215 is associated with the training dataset 225. According to other exemplary embodiments, the translation engine 215 can be associated with any number of training datasets. The training dataset 225 may comprise documents in source languages and corresponding translations of those documents in target languages (*i.e.*, parallel corpora). The translated documents may be human-generated or machine-generated. The training dataset 225 may be domain-specific or

generic. Accordingly, the translation engine 215 may be associated with specific subject matter. For example, the translation engine 215 may be associated with consumer electronics or with agriculture.

[0034] According to exemplary embodiments, the quality-prediction engine 220 is executable to predict a trust level of a translation provided by the translation engine 215. The trust-level prediction is indicative of translational accuracy of translations generated by the translation engine 215. The trust level is predicted independent of a human-generated translation or other human intervention. The quality-prediction engine 220 is described in further detail in connection with FIG. 3.

[0035] FIG. 3 is a block diagram of an exemplary quality-prediction engine 220 included in the translation application 140. The quality-prediction engine 220 provides a trust-level prediction of a translation generated through execution of the translation engine 215. The depicted quality-prediction engine 220 includes a feature-comparison module 305, a mapping module 310, an analysis module 315, a trust-level-determination module 320, and a calibration module 325, all of which may be stored in memory and executed by a processor to effectuate the functionalities attributed thereto.

Furthermore, the quality-prediction engine 220 can be composed of more or fewer modules (or combinations of the same) and still fall within the scope of the present technology. For example, the functionalities of the feature-comparison module 305 and the functionalities of the mapping module 310 may be combined into a single module or engine.

[0036] Some modules included in the quality-prediction engine 220 may be used primarily during training of the quality-prediction engine 220, while other modules may be used primarily when the quality-prediction engine 220 is utilized to provide credibility for machine-generated translations. During training, the quality-prediction engine 220, in effect, learns to predict translational accuracy. In general terms, this learning process can be achieved

using several parallel corpora, namely a source-language corpus, a human-generated translation of the source-language corpus (*i.e.*, a human-generated target-language corpus), and a machine-generated translation of the source-language corpus (*i.e.*, a machine-generated target-language corpus). These corpora may reside in the training dataset 225. By forming a comparison, relative to the source-language corpus, between the human-generated target-language corpus and the machine-generated target-language corpus, the quality-prediction engine 220 can learn aspects of accurate and inaccurate translation by the translation engine 215.

[0037] According to exemplary embodiments, two or more machine-generated translations of the source-language corpus may be utilized. These two or more machine-generated translations may be obtained from two or more corresponding translation engines 215, which may each be based on different translation techniques or similar translation techniques using different translation algorithms. One of these translation engines 215 is a primary engine that generates final translations outputted to a user. One or more other translation engines 215 are secondary engines used for during training and trust-level prediction.

[0038] The feature-comparison module 305 can be executed to compare a machine-generated target-language corpus with a human-generated target-language corpus, relative to a corresponding source-language corpus. This comparison can be used in mapping features such as similarities and differences between the machine-generated target-language corpus and the human-generated target-language corpus. Another feature may be the length of input text segments. A metric referred to herein as perplexity may also be invoked as a feature. Perplexity describes the likelihood that an output string would occur in a target language. For example, if English were the target language, the string "red car" would have a lower perplexity than the string "car red," wherein lower perplexity is more desirable than higher perplexity.

In embodiments having more than one translation engine **215**, an exemplary feature may be a correlation between translations generated by a primary engine and a secondary engine relative to a correlation between translations generated by a primary engine and a human.

[0039] Execution of the mapping module **310** allows generation of a mapping between features of the machine-generated target-language corpus and features of the human-generated target-language corpus based on the comparison obtained through execution of the feature-comparison module **305**. This mapping allows determination of trust levels associated with translational accuracy of future machine-generated translations that lack corresponding human-generated translations. More specifically, the quality-prediction engine **220** can use this mapping learned from the parallel corpora comparison, when human-generated translations are not available, to make a prediction as to how close a machine-generated translation is to what a human translator might generate. This prediction can be expressed in many manners such as a percentile or scaled value.

[0040] When the quality-prediction engine **220** is utilized to provide credibility for machine-generated translations, the analysis module **315** is executed to analyze discrete units of a machine-generated translation obtained from the translation engine **215** to determine a trust level or a trust value for each discrete unit. The discrete units may comprise words, phrases, sentences, paragraphs, pages, and so forth. The trust value may be associated with translation accuracy, which in turn may represent an alignment-estimation between the machine-generated translation and a prospective human-generated translation for the same source-language document. Trust values for discrete units can allow weak points in machine-generated translations to be identified. In some cases, such as when a trust value for a discrete unit falls below a threshold value, a human translator or another

translation engine 215 may be invoked to retranslate or otherwise improve that discrete unit to elevate the associated trust value.

[0041] In embodiments having more than one translation engine 215, the analysis module 315 may also consider a correlation between a translation generated by a primary engine and a translation generated by a secondary engine in determining a trust level or a trust value for each discrete unit. To illustrate, say there is a high correlation between translations generated by a primary engine and a secondary engine during training when there is also a high correlation between translations generated by the primary engine and a human. While providing machine-generated translations, when no human-generated translations are available, a high correlation between translations generated by a primary engine and a secondary engine may be assumed to indicate a high trust level or trust value.

[0042] The trust-level-determination module 320 may be executed to obtain the trust level for the machine-generated translation by combining the trust values of each discrete unit analyzed by the analysis module 315. The trust values can be combined using a weighted average, for example. According to exemplary embodiments, a contribution of each trust value to the weighted average is associated with the length of the respective discrete unit. When the machine-generated translation includes a batch of documents, an aggregate trust level indicative of the overall translational accuracy of the batch may also be obtained by way of execution of the trust-level-determination module 320.

[0043] Execution of the calibration module 325 allows calibration of the quality-prediction engine 220. Calibration of the quality-prediction engine 220 may be desirable for a number of reasons. For example, in some domains such as legal writing, precision is of great importance so the quality-prediction engine 220 may overestimate the trust level compared to what a human translator might suggest. Conversely, in domains such as internet

chat where form is secondary to conveying concepts, the quality-prediction engine 220 may underestimate the trust level compared to what a human translator might suggest. In exemplary embodiments, the calibration module 325 uses ratings provided by third-party sources of samples of translations generated by the translation engine 215 to adjust or tune the scale used for the trust levels determined by the trust-level-determination module 320. Such third-party sources may include human translators. Additionally, adjustments to a trust-level scale can be linear or non-linear.

[0044] In some instances, content can shift such as in documents that are continuously updated. To ensure that translations and corresponding trust levels remain dependable, feedback may be provided by users. This feedback may be obtained from the interface module 210. In one example, a user can provide feedback through a widget presented on a website or elsewhere. By obtaining feedback from users, the calibration module 325 can be automatically triggered to perform further calibrations ensuring that the calibration of the quality-prediction engine 220 remains consistent with user feedback.

[0045] FIG. 4 is a flowchart of an exemplary method 400 for training a quality-prediction engine (*e.g.*, the quality-prediction engine 220). The steps of the method 400 may be performed in varying orders. Additionally, steps may be added or subtracted from the method 400 and still fall within the scope of the present technology.

[0046] In step 405, a document in a source language is translated to a target language to obtain a machine-generated translation. Step 405 may be performed by executing a machine-translation engine (*e.g.*, the translation engine 215) stored in memory.

[0047] In step 410, the machine-generated translation is compared with a human-generated translation of the document. The human-generated translation is also in the target language.

[0048] In step 415, a mapping between features of the machine-generated translation and features of the human-generated translation is generated. The mapping can be generated based on comparison of step 410. In addition, the mapping allows determination of trust levels, which as associated with translational accuracy, of future machine-generated translation that lack corresponding human-generated translations.

[0049] FIG. 5 is a flowchart of an exemplary method 500 for credibly providing machine-generated translations. The steps of the method 500 may be performed in varying orders. Additionally, steps may be added or subtracted from the method 500 and still fall within the scope of the present technology.

[0050] In step 505, a document is translated from a source language to a target language to obtain a machine-generated translation. Step 505 may be performed by executing a machine-translation engine (*e.g.*, the translation engine 215) stored in memory.

[0051] In step 510, a trust level of the machine-generated translation is predicted. Such a trust-level prediction is associated with translational accuracy of the machine-generated translation. The trust-level prediction may be obtained through execution of a quality-prediction engine (*e.g.*, the quality-prediction engine 220) stored in memory.

[0052] In step 515, the machine-generated translation and the trust level are outputted. According to exemplary embodiments, the machine-generated translation and the trust level are outputted to a user via a computing device such as computing device 105.

[0053] FIG. 6 illustrates an exemplary computing system 600 that may be used to implement an embodiment of the present technology. The computing system 600 may be implemented in the contexts of the likes of the computing device 105, a server implementing the third-party website 155, and a server implementing the translation application 140. The computing system 600

includes one or more processors 610 and main memory 620. Main memory 620 stores, in part, instructions and data for execution by processor 610. Main memory 620 can store the executable code when in operation. The computing system 600 further includes a mass storage device 630, portable storage medium drive(s) 640, output devices 650, user input devices 660, a graphics display 670, and peripheral device(s) 680.

[0054] The components shown in FIG. 6 are depicted as being connected via a single bus 690. The components may be connected through one or more data transport means. The processor 610 and the main memory 620 may be connected via a local microprocessor bus, and the mass storage device 630, the peripheral devices 680, the portable storage medium drive(s) 640, and display system 670 may be connected via one or more input/output (I/O) buses.

[0055] The mass storage device 630, which may be implemented with a magnetic disk drive or an optical disk drive, is a non-volatile storage device for storing data and instructions for use by the processor 610. The mass storage device 630 can store the system software for implementing embodiments of the present invention for purposes of loading that software into the main memory 620.

[0056] The portable storage device 640 operates in conjunction with a portable non-volatile storage medium, such as a floppy disk, compact disk, digital video disc, or USB storage device, to input and output data and code to and from the computer system 600 of FIG. 6. The system software for implementing embodiments of the present invention may be stored on such a portable medium and input to the computer system 600 via the portable storage device 640.

[0057] The input devices 660 provide a portion of a user interface. The input devices 660 may include an alpha-numeric keypad, such as a keyboard, for inputting alpha-numeric and other information, or a pointing device, such as a mouse, a trackball, stylus, or cursor direction keys. Additionally, the

computing system 600 as shown in FIG. 6 includes the output devices 650.

Suitable output devices include speakers, printers, network interfaces, and monitors.

[0058] The display system 670 may include a liquid crystal display (LCD) or other suitable display device. The display system 670 receives textual and graphical information, and processes the information for output to the display device.

[0059] The peripheral device(s) 680 may include any type of computer support device to add additional functionality to the computer system. The peripheral device(s) 680 may include a modem or a router.

[0060] The components contained in the computer system 600 of FIG. 6 are those typically found in computer systems that may be suitable for use with embodiments of the present invention and are intended to represent a broad category of such computer components that are well known in the art. Thus, the computer system 600 of FIG. 6 can be a personal computer, hand held computing device, telephone, mobile computing device, workstation, server, minicomputer, mainframe computer, or any other computing device. The computer can also include different bus configurations, networked platforms, multi-processor platforms, etc. Various operating systems can be used including Unix, Linux, Windows, Macintosh OS, Palm OS, webOS, Android, iPhone OS and other suitable operating systems.

[0061] It is noteworthy that any hardware platform suitable for performing the processing described herein is suitable for use with the technology.

Computer-readable storage media refer to any medium or media that participate in providing instructions to a central processing unit (CPU), a processor, a microcontroller, or the like. Such media can take forms including, but not limited to, non-volatile and volatile media such as optical or magnetic disks and dynamic memory, respectively. Common forms of computer-readable storage media include a floppy disk, a flexible disk, a hard

disk, magnetic tape, any other magnetic storage medium, a CD-ROM disk, digital video disk (DVD), any other optical storage medium, RAM, PROM, EPROM, a FLASH EPROM, any other memory chip or cartridge.

[0062] Various forms of transmission media may be involved in carrying one or more sequences of one or more instructions to a CPU for execution. A bus carries the data to system RAM, from which a CPU retrieves and executes the instructions. The instructions received by system RAM can optionally be stored on a fixed disk either before or after execution by a CPU.

[0063] While various embodiments have been described above, it should be understood that they have been presented by way of example only, and not limitation. The descriptions are not intended to limit the scope of the technology to the particular forms set forth herein. Thus, the breadth and scope of a preferred embodiment should not be limited by any of the above-described exemplary embodiments. It should be understood that the above description is illustrative and not restrictive. To the contrary, the present descriptions are intended to cover such alternatives, modifications, and equivalents as may be included within the spirit and scope of the technology as defined by the appended claims and otherwise appreciated by one of ordinary skill in the art. The scope of the technology should, therefore, be determined not with reference to the above description, but instead should be determined with reference to the appended claims along with their full scope of equivalents.

CLAIMS

What is claimed is:

1. A method for credibly providing machine-generated translations, the method comprising:
 - translating a document from a source language to a target language by executing a machine-translation engine stored in memory to obtain a machine-generated translation;
 - predicting a trust level of the machine-generated translation by executing a quality-prediction engine stored in memory, the trust level associated with translational accuracy of the machine-generated translation;
 - and
 - outputting the machine-generated translation and the trust level.
2. The method of claim 1, wherein the trust level is predicted independent of a human-generated translation of the document.
3. The method of claim 1, wherein the translational accuracy represents an alignment-estimation between the machine-generated translation and a prospective human-generated translation.
4. The method of claim 1, wherein the trust level is indicated as one of a numerical scale, a term-based scale, a star-rating scale, or an analog scale.
5. The method of claim 1, wherein predicting the trust level includes analyzing discrete units of the document to determine a trust value for each discrete unit.

6. The method of claim 3, wherein predicting the trust level further includes combining the trust values using a weighted average to obtain the trust level.
7. The method of claim 4, wherein a contribution of each trust value to the weighted average is associated with the length of the respective discrete unit.
8. The method of claim 3, wherein a discrete unit includes a sentence.
9. A system for credibly providing machine-generated translations, the system comprising:
 - a machine-translation engine stored in memory and executable by a processor to translate a document from a source language to a target language to obtain a machine-generated translation;
 - a quality-prediction engine stored in memory and executable by a processor to predict a trust level of the machine-generated translation, the trust level associated with translational accuracy of the machine-generated translation; and
 - a communications engine stored in memory and executable by a processor to output the machine-generated translation and the trust level, the communications engine being communicatively coupled with the machine-translation engine and the quality-prediction engine.
10. The system of claim 9, wherein the trust level is predicted independent of a human-generated translation of the document.

11. The system of claim 9, wherein the quality-prediction engine includes an analysis module stored in memory and executable by a processor to analyze discrete units of the machine-generated translation to determine a trust value for each discrete unit.

12. The system of claim 11, wherein the quality-prediction engine includes a trust-level-determination module stored in memory and executable by a processor to obtain the trust level by combining the trust values using a weighted average.

13. A computer readable storage medium having a program embodied thereon, the program executable by process to perform a method for credibly providing machine-generated translations, the method comprising:

translating a document from a source language to a target language using a machine-translation engine to obtain a machine-generated translation;

predicting a trust level of the machine-generated translation using a quality-prediction engine, the trust level associated with translational accuracy of the machine-generated translation; and

outputting the machine-generated translation and the trust level.

14. A method for training a quality-prediction engine, the method comprising:

translating a document in a source language to a target language by executing a machine-translation engine stored in memory to obtain a machine-generated translation;

comparing the machine-generated translation with a human-generated translation of the document, the human-generated translation in the target language; and

generating a mapping between features of the machine-generated translation and features of the human-generated translation based on the comparison, the mapping allowing determination of trust levels associated with translational accuracy of future machine-generated translations that lack corresponding human-generated translations.

15. The method of claim 14, further comprising calibrating the quality-prediction engine.

16. The method of claim 15, wherein calibrating the quality-prediction engine includes:

obtaining a plurality of opinions for a plurality of sample translations generated by execution of the machine-translation engine, each of the opinions from a human and indicating a perceived trust level of corresponding sample translations;

using the quality-prediction engine to determine a trust level of each of the plurality of sample translations;

determining a relationship between the plurality of opinions and the trust levels; and

tuning the mapping to minimize any difference between the plurality of opinions and the trust levels.

17. The method of claim 16, wherein calibrating the quality-prediction engine is automatically triggered to ensure that determined trust levels are continually consistent with user feedback.

18. A system for training a quality-prediction engine, the system comprising:
- a machine-translation engine stored in memory and executable by a processor to translate a document in a source language to a target language to obtain a machine-generated translation;
 - a feature-comparison module stored in memory and executable by a processor to compare the machine-generated translation with a human-generated translation of the document, the human-generated translation in the target language; and
 - a mapping module stored in memory and executable by a processor to generate a mapping between features of the machine-generated translation and features of the human-generated translation based on the comparison, the mapping allowing determination of trust levels associated with translational accuracy of future machine-generated translations that lack corresponding human-generated translations.
19. The system of claim 18, further comprising a calibration module stored in memory and executable by a processor to calibrate the quality-prediction engine.

20. A computer readable storage medium having a program embodied thereon, the program executable by a processor to perform a method for training a quality-prediction engine, the method comprising:

translating a document in a source language to a target language using a machine-translation engine to obtain a machine-generated translation;

comparing the machine-generated translation with a human-generated translation of the document, the human-generated translation in the target language; and

generating a mapping between features of the machine-generated translation and features of the human-generated translation based on the comparison, the mapping allowing determination of trust levels associated with translational accuracy of future machine-generated translations that lack corresponding human-generated translations.

1 of 6

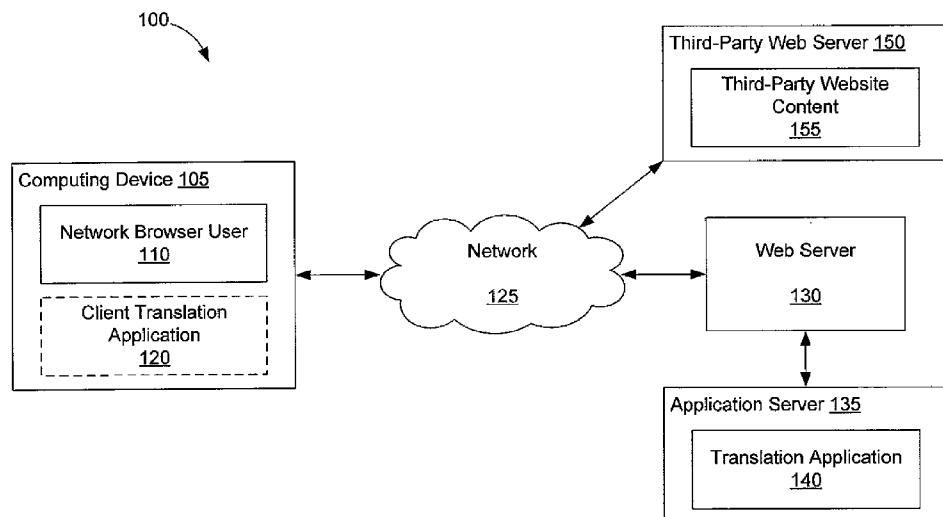


FIGURE 1

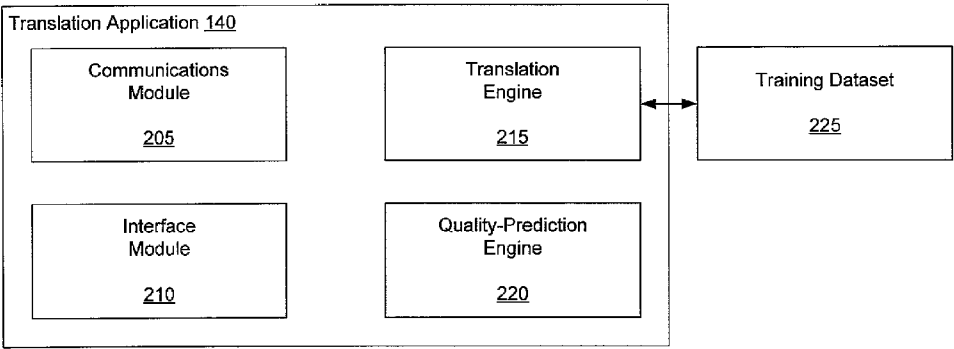


FIGURE 2

3 of 6

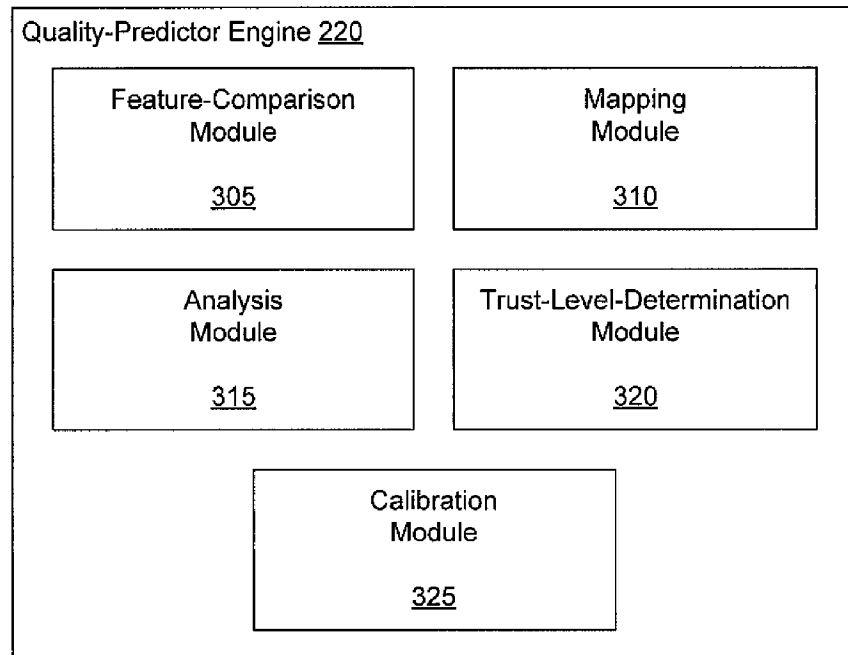


FIGURE 3

4 of 6

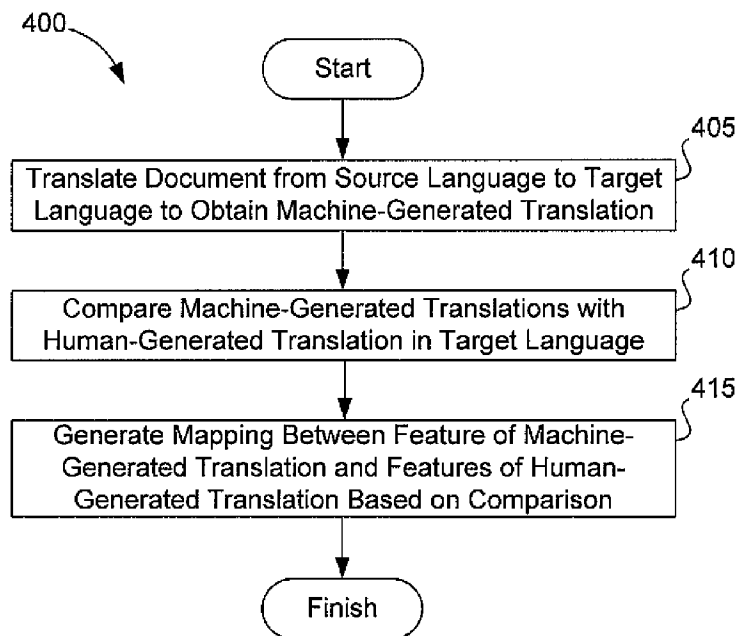


FIGURE 4

5 of 6

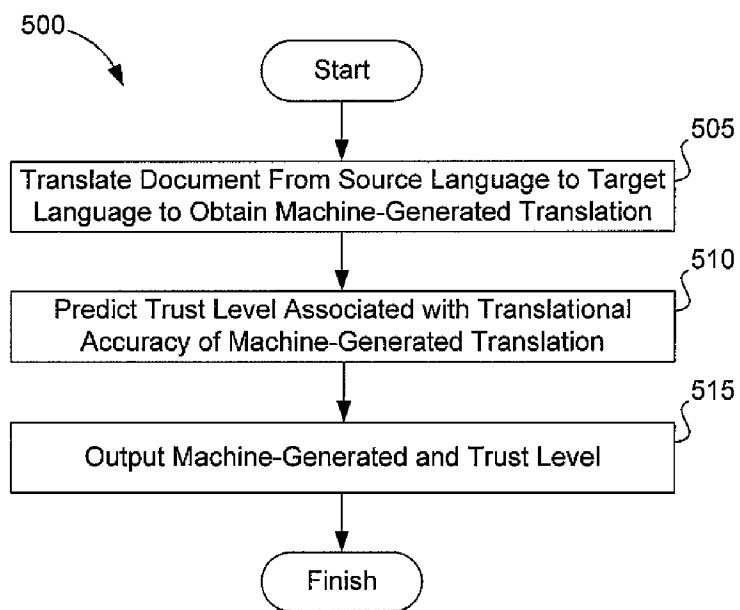


FIGURE 5

6 of 6

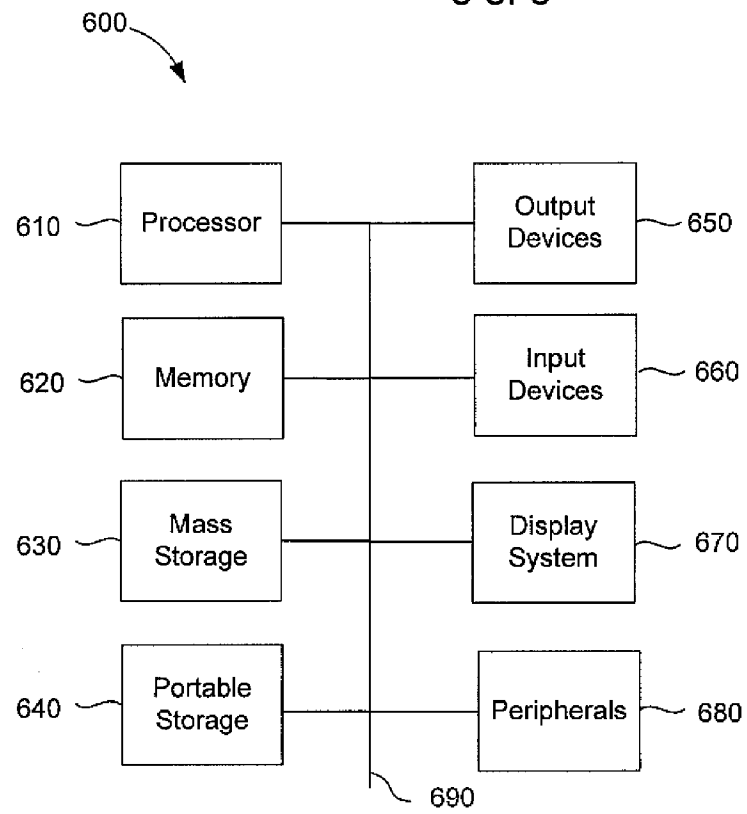


FIGURE 6

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US 10/51120

A. CLASSIFICATION OF SUBJECT MATTER

IPC(8) - G06F 17/28 (2010.01)

USPC - 704/2

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)

USPC: 704/2

IPC(8): G06F 17/28 (2010.01)

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

USPC: 704/1, 4, 5, 200, 270, 277; 709/223; 726/2, 3, 11, 13, 22; 714/1, 25, 37, 39, 48, 49, 50, 51, 100, 699, 799, 811, 815; 705/1.1, 7, 26; 706/12, 13 - search terms below.

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

WEST (PGPB, USPT, EPAB, JPAB) - language, linguistic, translate, translation, interpret, interpretation, interpreter, human, manual, quality, trust, accuracy, correctness, precision, closeness, credibility, trustworthiness, certainty, confidence, reliability, accurateness, weight, average, prefer, favor, preference.

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	US 2004/0260532 A1 (RICHARDSON et al.) 23 December 2004 (23.12.2004) entire document, especially Abstract; para [0002], [0005], [0018], [0025]-[0028], [0032], [0038]-[0040], [0042]-[0045], [0050], [0052], [0059], [0060], [0083]-[0086], [0088], [0089], [0092], [0095], [0096].	1-15, 18-20 -----
Y	US 2007/0294076 A1 (SHORE et al.) 20 December 2007 (20.12.2007) entire document, especially Abstract; para [0005]-[0007], [0022], [0035], [0054], [0056].	16, 17
Y	US 2007/0294076 A1 (SHORE et al.) 20 December 2007 (20.12.2007) entire document, especially Abstract; para [0005]-[0007], [0022], [0035], [0054], [0056].	16, 17
A	US 2009/0119091 A1 (SARIG) 07 May 2009 (07.05.2009) entire document	1-20
A	US 2008/0281578 A1 (KUMARAN et al.) 13 November 2008 (13.11.2008) entire document	1-20
A	US 2008/0270112 A1 (SHIMOHATA) 30 October 2008 (30.10.2008) entire document	1-20
A	US 2007/0219774 A1 (QUIRK et al.) 20 September 2007 (20.09.2007) entire document	1-20
A	US 2002/0169592 A1 (AITYAN) 14 November 2002 (14.11.2002) entire document	1-20



Further documents are listed in the continuation of Box C.



* 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

14 November 2010 (14.11.2010)

Date of mailing of the international search report

02 DEC 2010

Name and mailing address of the ISA/US

Mail Stop PCT, Attn: ISA/US, Commissioner for Patents
P.O. Box 1450, Alexandria, Virginia 22313-1450

Facsimile No. 571-273-3201

Authorized officer:

Lee W. Young

PCT Helpdesk: 571-272-4300

PCT OSP: 571-272-7774