

(19) 日本国特許庁(JP)

(12) 公表特許公報(A)

(11) 特許出願公表番号

特表2016-507828
(P2016-507828A)

(43) 公表日 平成28年3月10日 (2016.3.10)

(51) Int.Cl.		F I			テーマコード (参考)
G06F 17/28	(2006.01)	G06F 17/28	609		5B091
G06F 17/27	(2006.01)	G06F 17/27	650		

審査請求 有 予備審査請求 未請求 (全 14 頁)

(21) 出願番号	特願2015-551995 (P2015-551995)	(71) 出願人	515191176 カタール・ファンデーション
(86) (22) 出願日	平成25年1月11日 (2013.1.11)		カタール ドーハ ビーオー ボックス 5825
(85) 翻訳文提出日	平成27年9月3日 (2015.9.3)	(74) 代理人	100093296 弁理士 小越 勇
(86) 国際出願番号	PCT/EP2013/050521	(74) 代理人	100173901 弁理士 小越 一輝
(87) 国際公開番号	W02014/108208	(72) 発明者	エレラ、フランシスコ ハビエル グスマン カタール ドーハ ビーオー ボックス 5825 カタール・ファンデーション
(87) 国際公開日	平成26年7月17日 (2014.7.17)	(72) 発明者	フォーゲル、ステファン カタール ドーハ ビーオー ボックス 5825 カタール・ファンデーション 最終頁に続く

(54) 【発明の名称】 機械翻訳システム及び方法

(57) 【要約】

機械翻訳システム(1)は、未知のテキスト(4)を受け取り、未知のテキスト(4)を小分けにして解析する言語解析モジュール(3)を備える。言語解析モジュール(3)は、未知のテキスト(4)における言語特徴を識別し、言語指紋を翻訳構成選択モジュール(5)に提供する。翻訳構成選択モジュール(5)は、識別された言語指紋に一致するメモリ(6)から翻訳構成(T₁~T₉)を選択し、選択された言語構成(T₁~T₉)を機械翻訳モジュール(7)に伝達する。機械翻訳モジュール(7)は、選択された翻訳構成(T₁~T₉)を用いて、未知のテキスト(4)を異なる言語に翻訳する。

【選択図】 図1

【特許請求の範囲】**【請求項 1】**

機械翻訳方法であって、

各々が少なくとも 1 つの言語指紋に対応する翻訳構成を複数提供すること、

テキストの一節を第 1 言語で受け取ること、

前記テキストの一節からテキストの第 1 部分における少なくとも 1 つの言語指紋を識別すること、

識別された言語指紋に基づいて、複数の翻訳構成から 1 つのグループの翻訳構成を選択すること、そして

選択されたグループの翻訳構成を用いて、前記テキストの一節の第 1 部分を第 2 言語に翻訳すること、

を特徴とする、機械翻訳方法。

【請求項 2】

前記翻訳構成は、最初に所定のグループに分類される、請求項 1 に記載の機械翻訳方法。

【請求項 3】

翻訳構成の各グループは、所定のテキスト・タイプに対応する、請求項 2 に記載の機械翻訳方法。

【請求項 4】

翻訳構成は、2 つ以上の前記所定のグループに属する、請求項 2 又は 3 に記載の機械翻訳方法。

【請求項 5】

前記方法は、更に、

前記テキストの一節からテキストの第 2 部分における少なくとも 1 つの言語指紋を識別すること、

テキストの第 2 部分における識別された各言語指紋に対応する複数の翻訳構成から第 2 グループの翻訳構成を選択すること、そして、

選択された第 2 グループの翻訳構成を用いて、前記テキストの一節の第 2 部分を第 2 言語に翻訳すること、

を特徴とする、請求項 1 から 4 のいずれか一項に記載の機械翻訳方法。

【請求項 6】

前記方法は、前記テキストの一節における各言語指紋に一致するよう翻訳構成のグループを動的に選択する、請求項 5 に記載の機械翻訳方法。

【請求項 7】

テキストの各部分は単独のワードである、請求項 1 から 6 のいずれか一項に記載の機械翻訳方法。

【請求項 8】

テキストの各部分は複数のワードである、請求項 1 から 6 のいずれか一項に記載の機械翻訳方法。

【請求項 9】

前記方法は、更に、

前記翻訳構成を生成し、前記翻訳構成をメモリに格納すること、

を特徴とする、請求項 1 から 8 のいずれか一項に記載の機械翻訳方法。

【請求項 10】

機械翻訳システムであって、

各々が少なくとも 1 つの言語指紋に対応する翻訳構成を複数格納するメモリと、

テキストの一節を第 1 言語で受け取り、前記テキストの一節からテキストの第 1 部分における少なくとも 1 つの言語指紋を識別することが可能な言語解析モジュールと、

前記言語解析モジュールによって識別された各言語指紋に対応する複数の翻訳構成から、1 つのグループの翻訳構成を選択することが可能な構成選択モジュールと、

10

20

30

40

50

選択されたグループの翻訳構成を用いて、前記テキストの一節の第 1 部分を第 2 言語に翻訳することが可能な機械翻訳モジュールと、
を備える、機械翻訳システム。

【請求項 1 1】

前記翻訳構成は、最初に所定のグループに分類される、請求項 1 0 に記載の機械翻訳システム。

【請求項 1 2】

翻訳構成の各グループは、所定のテキスト・タイプに対応する、請求項 1 1 に記載の機械翻訳システム。

【請求項 1 3】

翻訳構成は、2 つ以上の前記所定のグループに属する、請求項 1 1 又は 1 2 に記載の機械翻訳システム。

【請求項 1 4】

前記言語解析モジュールは、前記テキストの一節からテキストの第 2 部分における少なくとも 1 つの言語指紋を識別することが可能であり、前記構成選択モジュールは、テキストの第 2 部分における識別された各言語指紋に対応する複数の翻訳構成から、第 2 グループの翻訳構成を選択することが可能であり、そして、前記機械翻訳モジュールは、選択された第 2 グループの翻訳構成を用いて、前記テキストの一節の第 2 部分を第 2 言語に翻訳することが可能である、請求項 1 0 から 1 3 のいずれか一項に記載の機械翻訳システム。

【請求項 1 5】

前記構成選択モジュールは、前記テキストの一節における各言語指紋に一致するよう翻訳構成のグループを動的に選択することが可能である、請求項 1 4 に記載の機械翻訳システム。

【請求項 1 6】

テキストの各部分は単独のワードである、請求項 1 0 から 1 5 のいずれか一項に記載の機械翻訳システム。

【請求項 1 7】

テキストの各部分は複数のワードである、請求項 1 0 から 1 5 のいずれか一項に記載の機械翻訳システム。

【請求項 1 8】

前記システムは、更に、

翻訳構成を生成し、前記翻訳構成をメモリに格納することが可能である構成生成器を備える、請求項 1 0 から 1 7 のいずれか一項に記載の機械翻訳システム。

【発明の詳細な説明】

【技術分野】

【0 0 0 1】

本発明は機械翻訳システム及び方法に関し、特に、異なる文体で書かれたテキストを第 1 言語から第 2 言語に翻訳することが可能である機械翻訳システム及び方法に関する。

【背景技術】

【0 0 0 2】

従来の統計的機械翻訳システムは、テキストを第 1 言語から第 2 言語へ翻訳するために、翻訳用モデル及び調整パラメータを含む翻訳構成又は翻訳パラメータを使用する。翻訳モデルは、第 1 及び第 2 言語の両言語におけるテキストの節（テキスト・パッセージ）を含む訓練データを解析することによって導き出される。調整パラメータは、各翻訳モデルの強度（又は重み）を設定することによって導き出されるが、これは、調整データセットと呼ばれる特定のデータセットに従って、最良の翻訳を得るためである。

【0 0 0 3】

統計的機械翻訳システムは、特定のドメイン（領域）に対する新しい訓練データを用いて、適切な翻訳構成を作成すること（モデル又は調整パラメータを再訓練すること）によって改善され得る。例えば、機械翻訳システムは、異なる文体で書かれたテキストの節（

10

20

30

40

50

テキスト・パッセージ)を含む訓練データを使用することで、再構成できる。従って、機械翻訳システムは、法律用語又は俗語を含むテキストのような、異なる文体で書かれたテキストを翻訳するように構成できる。

【0004】

従来の機械翻訳システムに関する問題は、特定のタイプのテキストに対して機械翻訳システムを再訓練することが難しいことと、時間がかかることの両方であるが、その理由は、2つの異なる言語で正確に書かれている長いテキストの一節に、起源を求める必要があるためである。

【発明の概要】

【発明が解決しようとする課題】

10

【0005】

本発明は、改善された機械翻訳システム及び方法を提供することに努める。

【課題を解決するための手段】

【0006】

本発明の実施形態では機械翻訳方法が提供され、この機械翻訳方法は、各々が少なくとも1つの言語指紋に対応する翻訳構成を複数提供すること、テキストの一節を第1言語で受け取ること、その一節からテキストの第1部分における少なくとも1つの言語指紋を識別すること、識別された言語指紋に基づいて、複数の翻訳構成から1つのグループの翻訳構成を選択すること、そして、選択されたグループの翻訳構成を用いて、その一節の第1部分を第2言語に翻訳することを特徴とする。

20

【0007】

好ましくは、翻訳構成は、最初に所定のグループに分類される。

【0008】

有利なことには、翻訳構成の各グループは、所定のテキスト・タイプに対応する。

【0009】

好ましくは、翻訳構成は、2つ以上の所定のグループに属する。

【0010】

都合のよいことには、本機械翻訳方法は、更に、テキストの一節からテキストの第2部分における少なくとも1つの言語指紋を識別すること、テキストの第2部分における識別された各言語指紋に対応する複数の翻訳構成から第2グループの翻訳構成を選択すること、そして、選択された第2グループの翻訳構成を用いて、テキストの一節の第2部分を第2言語に翻訳することを特徴とする。

30

【0011】

有利なことには、本方法は、テキストの一節における各言語指紋に一致するよう翻訳構成のグループを動的に選択する。

【0012】

都合のよいことには、テキストの各部分は単独のワードである。

【0013】

好ましくは、テキストの各部分は複数のワードである。

【0014】

40

有利なことには、本方法は、更に、翻訳構成を生成し、メモリに翻訳構成を格納することを特徴とする。

【0015】

本発明の別の態様では、機械翻訳システムを提供し、この機械翻訳システムは、各々が少なくとも1つの言語指紋に対応する翻訳構成を複数格納するメモリと、テキストの一節を第1言語で受け取り、その一節からテキストの第1部分における少なくとも1つの言語指紋を識別することが可能な言語解析モジュールと、該言語解析モジュールによって識別された各言語指紋に対応する複数の翻訳構成から、1つのグループの翻訳構成を選択することが可能な構成選択モジュールと、選択されたグループの翻訳構成を用いて、テキストの一節の第1部分を第2言語に翻訳することが可能な機械翻訳モジュールと、を備える。

50

【0016】

都合のよいことには、翻訳構成は、最初に所定のグループに分類される。

【0017】

好ましくは、翻訳構成の各グループは、所定のテキスト・タイプに対応する。

【0018】

有利なことには、翻訳構成は、2つ以上の所定のグループに属する。

【0019】

都合のよいことには、言語解析モジュールは、テキストの一節からテキストの第2部分における少なくとも1つの言語指紋を識別することが可能であり、構成選択モジュールは、テキストの第2部分における識別された各言語指紋に対応する複数の翻訳構成から、第2グループの翻訳構成の選択することが可能であり、そして、機械翻訳モジュールは、選択された第2グループの翻訳構成を用いて、テキストの一節の第2部分を第2言語に翻訳することが可能である。

10

【0020】

有利なことには、構成選択モジュールは、テキストの一節において各言語指紋に一致するよう翻訳構成のグループを動的に選択することが可能である。

【0021】

好ましくは、テキストの各部分は単独のワードである。

【0022】

都合のよいことには、テキストの各部分は複数のワードである。

20

【0023】

有利なことには、本システムは、翻訳構成を生成し、該翻訳構成をメモリに格納することが可能である構成生成器(コンフィグレーション・ジェネレーター)を更に備える。

【0024】

本発明がより容易に理解され、且つ、それによって、本発明の更なる特徴が正しく認識されるために、本発明の実施形態が、添付図面を参照しながら、例として、これから説明する。

【図面の簡単な説明】

【0025】

【図1】本発明の一実施形態の機械翻訳システムを示す概略図である。

30

【図2】翻訳構成の4つの所定のグループを示す概略図である。

【図3】本発明の、ある実施形態の一要素を形成する翻訳構成生成器のシステムを示す概略図である。

【発明を実施するための形態】

【0026】

初めに、添付の図1を参照すると、機械翻訳システム1は、処理配列2を含む。該処理配列2は言語解析モジュール3を含み、この言語解析モジュールは、未知のテキストの一節4を受け取ることが可能である。未知のテキストの一節4は、より短い未知のテキストの節を複数備えてもよい。

【0027】

言語解析モジュール3は、未知のテキスト4からテキスト部分の言語を解析し、且つテキストにおいて少なくとも1つの言語特徴を識別するように構成される。各言語特徴は、以下に限定されるものではないが、例えば、言語タイプ、文体、言語特性又は複雑さのような、テキスト特性を表す。言語特徴やテキスト特性の集合体を、ここでは、テキストの言語指紋と呼ぶ。正しく認識されるべきことであるが、言語解析モジュールは、そのようなテキストに対する言語指紋を生成するために、未知のテキスト4における他の可能性のある言語特徴を識別するように構成される。

40

【0028】

言語解析モジュール3は、未知のテキスト4における言語の結合された特徴を表す言語指紋として、識別された各言語特徴を格納する。

50

【 0 0 2 9 】

言語解析モジュール3は、翻訳構成選択モジュール5に接続される。言語解析モジュール3は、識別された言語指紋を、構成選択モジュール5に伝達することが可能である。

【 0 0 3 0 】

翻訳構成選択モジュール5は、複数の所定の翻訳構成 $T_1 \sim T_9$ を格納する構成メモリ6と通信している。他の実施形態において、翻訳構成 $T_1 \sim T_9$ は、別のモデル及び別の調整パラメータを含む。翻訳構成 $T_1 \sim T_9$ の各々は、特定の言語特性を機械翻訳するためのパラメータである。正しく認識されるべきことであるが、図1には9個の翻訳構成だけが示されているが、本発明の実施形態では、翻訳構成が幾つあってもよい。実際に、本発明のある実施形態では、無数の構成があり、構成の各々は、一意的に識別できる言語指紋に対して生成される。一実施形態において、翻訳構成は統計的に選択され、短文/長文、固有表現、固有ブランド、センテンス中での動詞/名詞/形容詞の位置決め、句読点などのような、いかなる言語特性をもカバーする。

10

【 0 0 3 1 】

各翻訳構成 $T_1 \sim T_9$ 又は各翻訳構成 $T_1 \sim T_9$ のグループは、1つ以上の言語指紋に対応する。構成選択モジュール5は、翻訳構成 $T_1 \sim T_9$ を識別するために、構成メモリ6を検索するように機能し、該翻訳構成 $T_1 \sim T_9$ は、未知のテキスト4において言語解析モジュールによって識別される各言語指紋に対応している。構成選択モジュール5は、構成メモリ6に格納された複数の翻訳構成 $T_1 \sim T_9$ から1つのグループの翻訳構成を選択するように機能し、該構成メモリ6は、言語解析モジュール3によって識別された言語指紋に対応している。選択されたグループの翻訳構成は、ただ1つの翻訳構成を備えてもよく、又は複数の翻訳構成を備えてもよい。構成選択モジュール5は、言語解析モジュール3によって生成される各言語指紋に対して、翻訳構成を選択することが可能である。

20

【 0 0 3 2 】

いったん翻訳構成選択モジュール5が1つのグループの翻訳構成を選択すると、該構成選択モジュール5は、そのグループの翻訳構成を機械翻訳モジュール7に伝達する。機械翻訳モジュール7は、その機械翻訳アルゴリズムの中で、選択されたグループの翻訳構成を使用して未知のテキスト4を異なる言語に翻訳する。機械翻訳システム1は、それ故に、テキストの言語指紋に従って、未知のテキストを解析し、且つ翻訳構成を動的に選択することが可能である。

30

【 0 0 3 3 】

更なる実施形態において、言語解析モジュール3は、未知のテキスト4からテキストの第1部分を解析することが可能である。構成選択モジュール5は、その後、テキストの第1部分の言語指紋に対応する翻訳構成を選択し、そして、機械翻訳モジュール7は、前述のとおり、選択された翻訳構成を使用してテキストの第1部分を翻訳する。しかしながら、この更なる実施形態では、機械翻訳システム1は、その後、未知のテキスト4からテキストの第2部分を解析し、翻訳することが可能である。未知のテキストからのテキストの第2部分は、第1部分とは別に解析され、テキストの第2部分では、テキストの第1部分とは異なる言語指紋が識別されることもある。テキストの第2部分に対する翻訳構成は、それ故に、テキストの第1部分に対する翻訳構成とは独立して選択される。従って、機械翻訳システム1は、未知のテキストにおけるテキストのさまざまな部分に対して、翻訳構成を動的に選択することが可能である。

40

【 0 0 3 4 】

この更なる実施形態の機械翻訳システム1は、未知のテキスト4における全てのテキスト部分に対して、解析及び機械翻訳を繰り返すことが可能である。機械翻訳システム1は、このように、テキストの各部分に対して翻訳構成が動的に選択された状態で、未知のテキスト4を小分けにして翻訳する。本発明の実施形態では、各部分は、単独のワードであってもよい。或いは、少なくとも一部分は、センテンス、段落、又はページのように複数のワードであってもよい。

【 0 0 3 5 】

50

さて、添付の図2を参照すると、本発明の更に別の実施形態において、翻訳構成 $T_1 \sim T_9$ は構成メモリ6の中で、所定のグループに分類される。図2に示された配列において、4つの所定のグループ8~11があるが、しかし翻訳構成の所定のグループが幾つあってもよいことは、正しく認識されるべきである。図2に示される第4グループ11によって表されるように、1つの翻訳構成が2つ以上のグループに属することもある。

【0036】

一実施形態において、所定のグループ8~11は、特定のテキスト・タイプと一致するように、各々が選択される。テキスト・タイプは、例えば、特定の言語指紋とマッチしてもよい。

【0037】

本発明の更に別の実施形態において、構成メモリ6に、翻訳構成の所定のグループ・セットが複数格納される。例えば、図2に示された配列において、第1のセットは、第1及び第2のグループ8、9を備え、且つ第2のセットは、第3から第4のグループ9~11を備える。セットは、特定のテキスト・パッセージのタイプにマッチするように選択される。例えば、セットは、法律専門用語を用いた文書に対して選択されてもよい。機械翻訳システム1は、未知のテキスト4におけるテキスト部分に一致するよう翻訳構成のグループ又は翻訳構成のグループ・セットを複数選択することが可能であるが、これは、未知のテキスト4の機械翻訳を動的に最適化するためである。

【0038】

翻訳構成選択の動的性質によって、本発明の実施形態は、未知のテキストにおける全てのテキスト部分に対する機械翻訳を最適化できるようになる。本発明の実施形態は、それ故に、そのテキストに対してシステムを再訓練しなくても、未知のテキストの個々の部分を翻訳できる。

【0039】

さて図3を参照すると、翻訳構成 $T_1 \sim T_9$ を生成するためのシステム12は、構成生成器14に接続される言語解析モジュール13を備える。言語解析モジュール13は、第1言語15及び第2言語16における対訳テキスト(パラレルテキスト)の部分を受け取ることが可能である。言語解析モジュール13は、第1言語15における入力テキストに対して言語指紋を生成し、そして、構成作成モジュール14は、第1言語15におけるこの入力テキストに対して最適な翻訳構成を作り出し、翻訳システム1が後で使用できるように、該構成を構成メモリ6に格納する。

【0040】

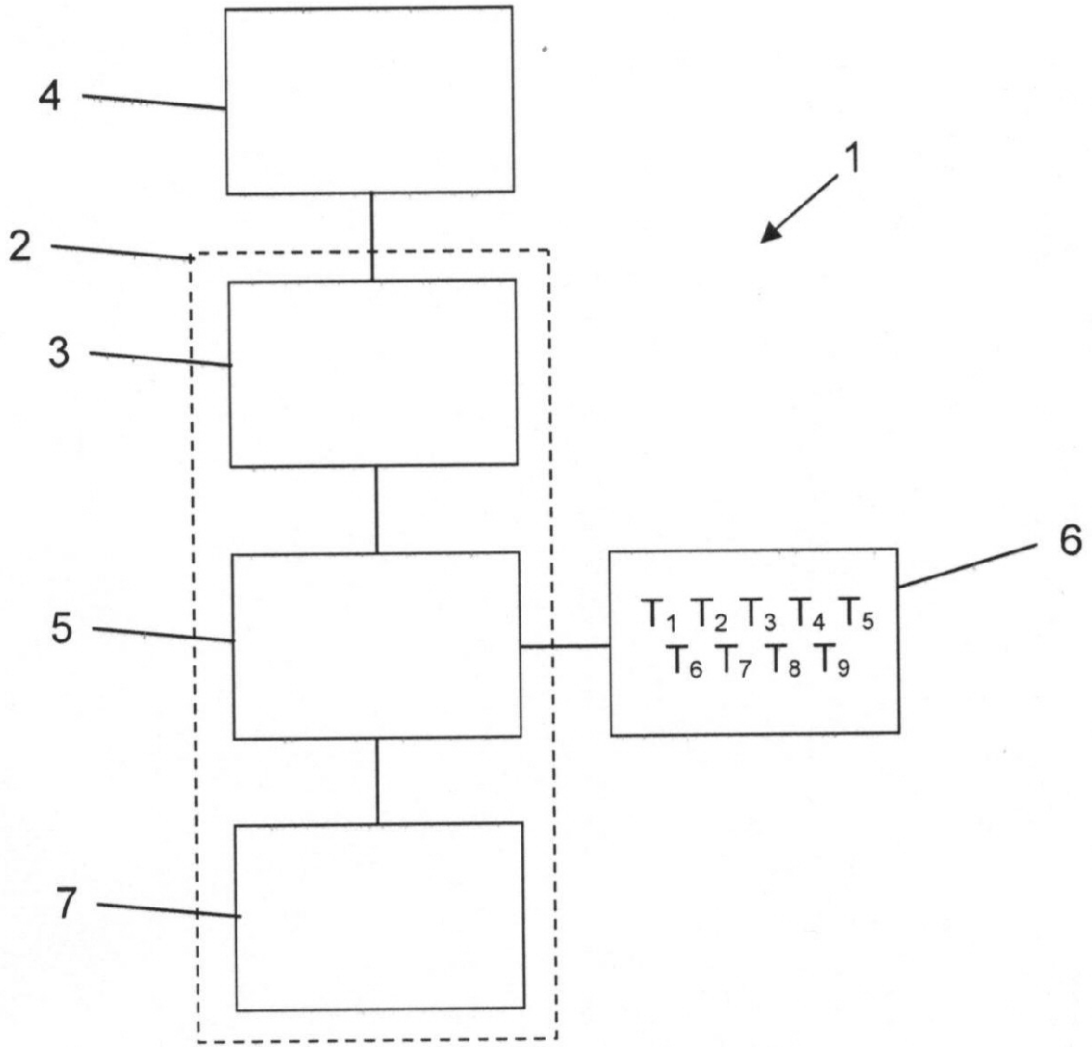
この明細書及び請求項で使用される場合、用語「備える(comprise)」及び「備えている(comprising)」及びこれらの変形は、明記された特徴、ステップ又は整数が含まれることを意味する。これらの用語は、他の特徴、ステップ又は要素の存在を排除すると解釈されるべきではない。

10

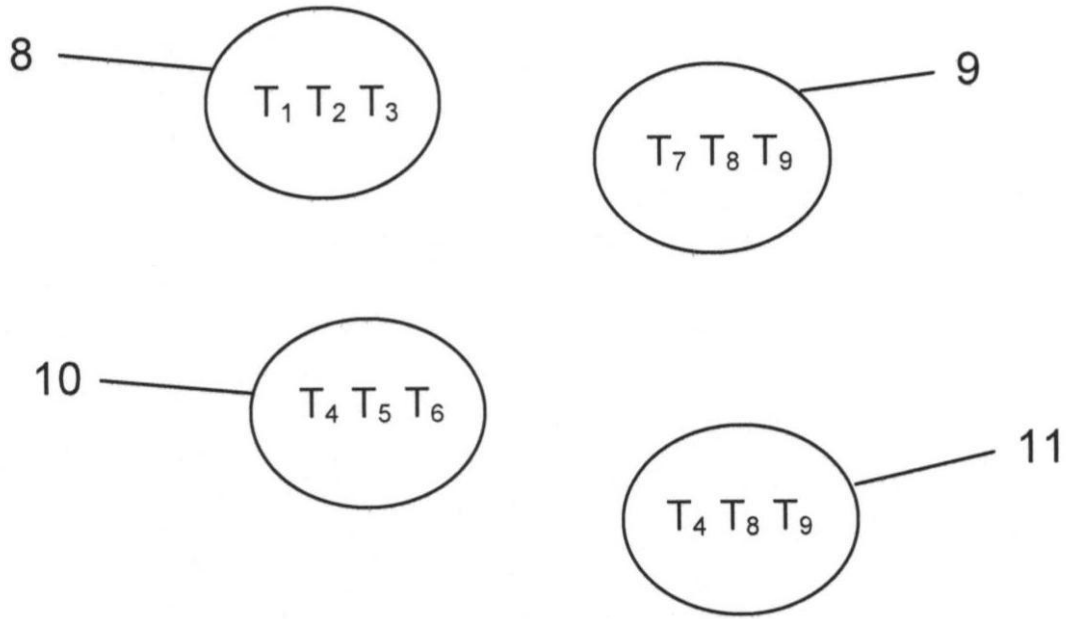
20

30

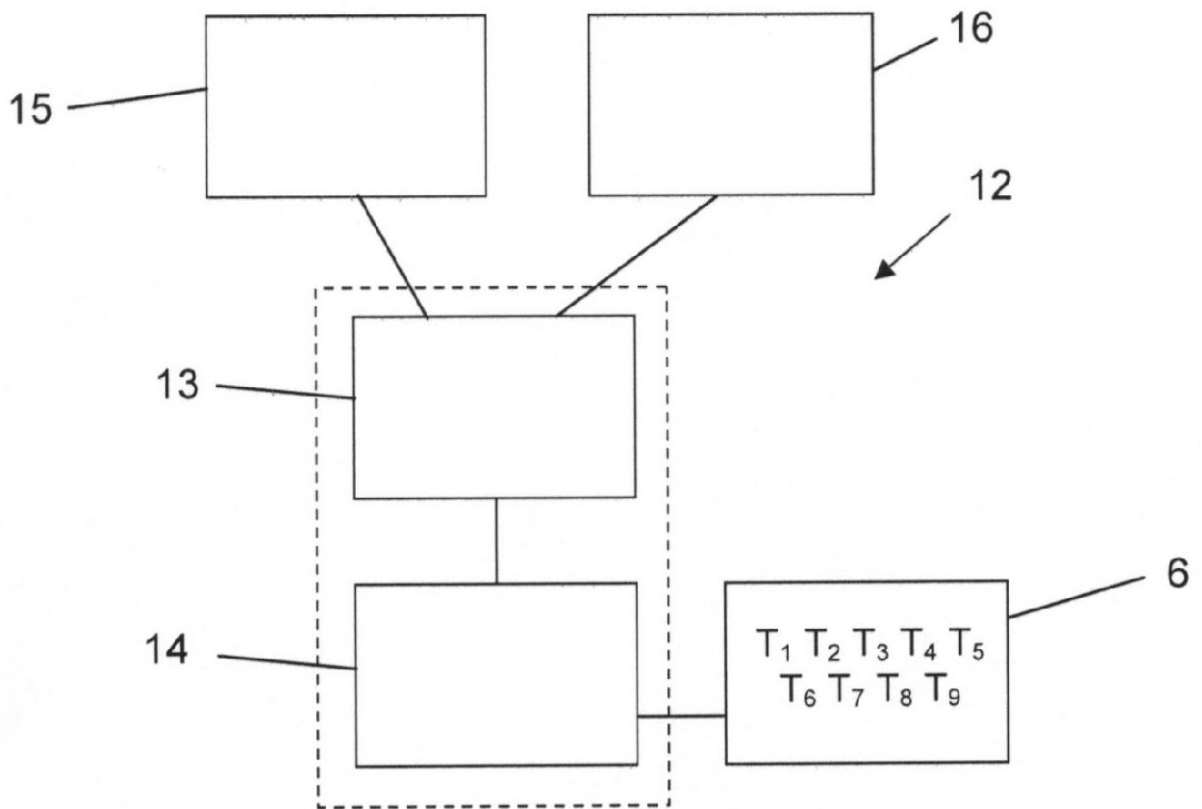
【 図 1 】



【 図 2 】



【 図 3 】



【 国際調査報告 】

INTERNATIONAL SEARCH REPORT

International application No PCT/EP2013/050521

A. CLASSIFICATION OF SUBJECT MATTER INV. G06F17/28 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		
C. DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	JIA XU ET AL: "Domain Dependent Statistical Machine Translation", MACHINE TRANSLATION SUMMIT XI, 10 September 2007 (2007-09-10), XP055093824, Copenhagen , Denmark abstract Sections 1-5 ----- -/--	1-6, 8-15,17, 18
<input checked="" type="checkbox"/> Further documents are listed in the continuation of Box C. <input type="checkbox"/> 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 17 December 2013		Date of mailing of the international search report 07/01/2014
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 Woods, Justin

1

INTERNATIONAL SEARCH REPORT

International application No

PCT/EP2013/050521

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	LEMAO LIU ET AL: "Locally Training the Log-Linear Model for SMT", PROCEEDINGS OF THE 2012 JOINT CONFERENCE ON EMPIRICAL METHODS IN NATURAL LANGUAGE PROCESSING AND COMPUTATIONAL NATURAL LANGUAGE LEARNING, 12 July 2012 (2012-07-12), pages 402-411, XP055093323, Jeju Island, Korea abstract Sections 1-3, 5 -----	1-6, 8-15,17, 18
X	CONG HUI ET AL: "An Empirical Study on Development Set Selection Strategy for Machine Translation Learning **", PROCEEDINGS OF THE JOINT 5TH WORKSHOP ON STATISTICAL MACHINE TRANSLATION AND METRICS, 15 July 2010 (2010-07-15), pages 67-71, XP055093827, Uppsala, Sweden abstract Sections 1-3 -----	1-6, 8-15,17, 18
T	PRESLAV NAKOV ET AL: "Parameter Optimization for Statistical Machine Translation: It Pays to Learn from Hard Examples", PROCEEDINGS OF RECENT ADVANCES IN NATURAL LANGUAGE PROCESSING, 7 September 2013 (2013-09-07), pages 504-510, XP055093296, Hissar, Bulgaria abstract Sections 1-4 -----	

INTERNATIONAL SEARCH REPORT

International application No.
PCT/EP2013/050521**Box No. II Observations where certain claims were found unsearchable (Continuation of Item 2 of first sheet)**

This international search report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:
2. Claims Nos.: 7, 16
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:
see FURTHER INFORMATION sheet PCT/ISA/210
3. Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box No. III Observations where unity of invention is lacking (Continuation of Item 3 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

1. As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.
2. As all searchable claims could be searched without effort justifying an additional fees, this Authority did not invite payment of additional fees.
3. As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:
4. No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest

- The additional search fees were accompanied by the applicant's protest and, where applicable, the payment of a protest fee.
- The additional search fees were accompanied by the applicant's protest but the applicable protest fee was not paid within the time limit specified in the invitation.
- No protest accompanied the payment of additional search fees.

International Application No. PCT/ EP2013/ 050521

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

Continuation of Box II.2

Claims Nos.: 7, 16

The term "linguistic fingerprint" is explicitly defined in the application (see description, page 5, lines 15-17) as "the ensemble of language features and characteristics of a text ", a "language feature" being "a feature representing a characteristic of the text, such as, but not limited to, the language type, writing style, linguistic characteristics or complexity" (see description, page 5, lines 13-15). The definitions have been used to interpret the terms for the purposes of search. In claim 1, linguistic fingerprints as defined in the passages of the description cited above are identified from portions of text. However, claim 7 and corresponding claim 16 limit the portion of text used to identify the linguistic fingerprint to a single word. This is completely inconsistent with the definition of the term, which requires "an ensemble of language features". It is not obvious to a person skilled in the art how such an ensemble of features should be derived from a single word, as the description merely repeats the wording of the claim, contrary to the requirements of Article 6 PCT, the subject matter of claims 7 and 16 are not clear, concise and supported in the description. As the description provides no further indication as to how the linguistic fingerprint should be established from a single word, the application fails to meet the requirements of Article 5 PCT that the description shall disclose the invention in a manner sufficiently clear and complete for the invention to be carried out by a person skilled in the art. As the application as a whole provides the person skilled in the art with no means of understanding and carrying out their subject matter, no meaningful search could be carried out with regard to claims 7 and 16.

The applicant's attention is drawn to the fact that claims relating to inventions in respect of which no international search report has been established need not be the subject of an international preliminary examination (Rule 66.1(e) PCT). The applicant is advised that the EPO policy when acting as an International Preliminary Examining Authority is normally not to carry out a preliminary examination on matter which has not been searched. This is the case irrespective of whether or not the claims are amended following receipt of the search report or during any Chapter II procedure. If the application proceeds into the regional phase before the EPO, the applicant is reminded that a search may be carried out during examination before the EPO (see EPO Guidelines C-IV, 7.2), should the problems which led to the Article 17(2) declaration be overcome.

フロントページの続き

(81)指定国 AP(BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, SZ, TZ, UG, ZM, ZW), EA(AM, AZ, BY, KG, KZ, RU, TJ, TM), EP(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), OA(BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG), AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, 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, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC

Fターム(参考) 5B091 AA01 AA15 BA11 CA01 CA21 CD03 EA17