

(19) 日本国特許庁(JP)

## (12) 公開特許公報(A)

(11) 特許出願公開番号

特開2015-5011

(P2015-5011A)

(43) 公開日 平成27年1月8日(2015.1.8)

(51) Int.Cl.

G06F 17/30 (2006.01)  
G06F 17/28 (2006.01)

F 1

G06F 17/30 240C  
G06F 17/28 Z  
G06F 17/30 170A

テーマコード(参考)

5B091

審査請求 有 請求項の数 5 O L (全 15 頁)

(21) 出願番号 特願2013-128180 (P2013-128180)  
(22) 出願日 平成25年6月19日 (2013.6.19)

(特許庁注：以下のものは登録商標)

1. フェイスブック

(71) 出願人 500257300  
ヤフー株式会社  
東京都港区赤坂9丁目7番1号  
(74) 代理人 100115129  
弁理士 清水 昇  
(72) 発明者 石川 裕貴  
東京都港区赤坂9丁目7番1号 ヤフー株式会社内  
(72) 発明者 鳩々野 学  
東京都港区赤坂9丁目7番1号 ヤフー株式会社内  
F ターム(参考) 5B091 AA03 BA02 CC15 CD03 EA02

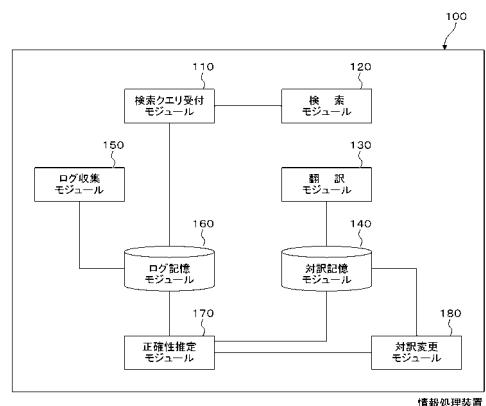
(54) 【発明の名称】情報処理装置、情報処理方法及び情報処理プログラム

## (57) 【要約】

【課題】検索サービスにおいて翻訳品質が求められる検索クエリ内の単語の翻訳の正確性を推定するようにした情報処理装置を提供する。

【解決手段】情報処理装置のクエリ受付手段は、第1の言語で記載された原We bページを機械翻訳システムによって第2の言語に翻訳した翻訳We bページに対して検索を行うための、第2の言語の単語を含む検索クエリを受け付け、対訳記憶手段は、機械翻訳システムが原We bページから翻訳We bページの翻訳に用いた第1の言語の単語と対訳である第2の言語の単語を記憶しており、ログ記憶手段は、クエリ受付手段によって受け付けられた検索クエリによって検索された翻訳We bページに対する操作者の操作のログを記憶しており、推定手段は、ログ記憶手段に記憶されているログを用いて、クエリ受付手段によって受け付けられた検索クエリ内の単語に対応する対訳記憶手段に記憶されている対訳の正確性を推定する。

【選択図】図1



**【特許請求の範囲】****【請求項 1】**

第1の言語で記載された原Webページを機械翻訳システムによって第2の言語に翻訳した翻訳Webページに対して検索を行うための、該第2の言語の単語を含む検索クエリを受け付けるクエリ受付手段と、

前記機械翻訳システムが前記原Webページから前記翻訳Webページの翻訳に用いた第1の言語の単語と対訳である第2の言語の単語を記憶している対訳記憶手段と、

前記クエリ受付手段によって受け付けられた検索クエリによって検索された翻訳Webページに対する操作者の操作のログを記憶しているログ記憶手段と、

前記ログ記憶手段に記憶されているログを用いて、前記クエリ受付手段によって受け付けられた検索クエリ内の単語に対応する前記対訳記憶手段に記憶されている対訳の正確性を推定する推定手段

を具備することを特徴とする情報処理装置。

**【請求項 2】**

前記推定手段によって推定された正確性に基づいて、前記対訳記憶手段に記憶されている対訳を変更する対訳変更手段

をさらに具備することを特徴とする請求項1に記載の情報処理装置。

**【請求項 3】**

前記ログ記憶手段は、前記対訳が正しいか否かを示す正解情報を付加して記憶しており、

前記ログ記憶手段に記憶された正解情報を教師データとして、該ログ記憶手段に記憶されているログを用いて、前記推定手段における推定処理を学習する学習手段をさらに具備し、

前記推定手段は、前記学習手段によって学習された推定処理にしたがって、対訳の正確性を推定する

ことを特徴とする請求項1又は2に記載の情報処理装置。

**【請求項 4】**

対訳記憶手段とログ記憶手段を具備する情報処理装置が行う情報処理方法であって、

前記対訳記憶手段は、機械翻訳システムが原Webページから翻訳Webページの翻訳に用いた第1の言語の単語と対訳である第2の言語の単語を記憶しており、

第1の言語で記載された原Webページを機械翻訳システムによって第2の言語に翻訳した翻訳Webページに対して検索を行うための、該第2の言語の単語を含む検索クエリを受け付けるクエリ受付ステップと、

前記クエリ受付ステップによって受け付けられた検索クエリによって検索された翻訳Webページに対する操作者の操作のログを前記ログ記憶手段に記憶させるステップと、

前記ログ記憶手段に記憶されているログを用いて、前記クエリ受付ステップによって受け付けられた検索クエリ内の単語に対応する前記対訳記憶手段に記憶されている対訳の正確性を推定する推定ステップ

を具備することを特徴とする情報処理方法。

**【請求項 5】**

コンピュータを、

第1の言語で記載された原Webページを機械翻訳システムによって第2の言語に翻訳した翻訳Webページに対して検索を行うための、該第2の言語の単語を含む検索クエリを受け付けるクエリ受付手段と、

前記機械翻訳システムが前記原Webページから前記翻訳Webページの翻訳に用いた第1の言語の単語と対訳である第2の言語の単語を記憶している対訳記憶手段と、

前記クエリ受付手段によって受け付けられた検索クエリによって検索された翻訳Webページに対する操作者の操作のログを記憶しているログ記憶手段と、

前記ログ記憶手段に記憶されているログを用いて、前記クエリ受付手段によって受け付けられた検索クエリ内の単語に対応する前記対訳記憶手段に記憶されている対訳の正確性

10

20

30

40

50

**を推定する推定手段**

として機能させるための情報処理プログラム。

**【発明の詳細な説明】****【技術分野】****【0001】**

本発明は、機械翻訳で翻訳された文に含まれる単語の対訳の正しさを推定する情報処理装置、情報処理方法及び情報処理プログラムに関する。 10

**【背景技術】****【0002】**

インターネットにおけるWebページにおいて、ユーザが用いる言語（母国語）とは異なる言語（外国語）で作成されたWebページを翻訳して、その翻訳したWebページを検索できるようにすることが行われている。その翻訳は、一般的に機械翻訳システムが用いられている。機械翻訳システムを用いた翻訳の場合、翻訳者が翻訳を行った場合と比べると、翻訳品質の劣ることが多い。

機械翻訳システムによる翻訳の品質を向上するために、対訳コーパスから統計的に対訳の正しさを推定することが行われている。又は、人手によって作成された対訳辞書を利用すること等が行われている。

また、関連する技術として、特許文献1等がある。

**【0003】**

特許文献1には、表音文字からなる言語の単語間で文字の表記に基づいて単語意味関係を正確に抽出する単語意味関係抽出装置を提供することを目的とし、データに含まれる単語から、二つの単語によって構成される単語ペアを抽出し、抽出した単語ペアの単語意味関係を判定する単語意味関係抽出装置において、単語意味関係辞書に登録された単語意味関係単語ペアの単語から複数の文字からなる意味素を抽出し、前記単語意味関係単語ペアの単語の意味素間の類似度を算出し、意味素間の類似度に基づいてデータから抽出された単語ペアの類似度を算出し、単語ペアの類似度に基づいて単語ペアの単語意味関係を判定することが開示されている。 20

**【先行技術文献】****【特許文献】****【0004】****【特許文献1】特開2012-108570号公報 30****【発明の概要】****【発明が解決しようとする課題】****【0005】**

人手による対訳辞書、対訳コーパスとともに、それを用意するのにコストが掛かる。また、対訳辞書を用いる場合、訳語に曖昧性がある単語は誤って翻訳される可能性がある。そして、対訳コーパスを用いる場合、翻訳確率は基本的には出現頻度に依存するため頻度の低い語は誤って訳される可能性がある。また、検索サービスにおいては、検索クエリ内の単語についての翻訳の正確性が求められている。

本発明は、このような背景技術の状況の中でなされたもので、検索サービスにおいて翻訳品質が求められる検索クエリ内の単語の翻訳の正確性を推定するようにした情報処理装置、情報処理方法及び情報処理プログラムを提供することを目的としている。 40

**【課題を解決するための手段】****【0006】**

かかる目的を達成するための本発明の要旨とするところは、次の各項の発明に存する。

[1] の発明は、第1の言語で記載された原Webページを機械翻訳システムによって第2の言語に翻訳した翻訳Webページに対して検索を行うための、該第2の言語の単語を含む検索クエリを受け付けるクエリ受付手段と、前記機械翻訳システムが前記原Webページから前記翻訳Webページの翻訳に用いた第1の言語の単語と対訳である第2の言語の単語を記憶している対訳記憶手段と、前記クエリ受付手段によって受け付けられた検 50

索クエリによって検索された翻訳Webページに対する操作者の操作のログを記憶しているログ記憶手段と、前記ログ記憶手段に記憶されているログを用いて、前記クエリ受付手段によって受け付けられた検索クエリ内の単語に対応する前記対訳記憶手段に記憶されている対訳の正確性を推定する推定手段を具備することを特徴とする情報処理装置である。

#### 【0007】

[2]の発明は、前記推定手段によって推定された正確性に基づいて、前記対訳記憶手段に記憶されている対訳を変更する対訳変更手段をさらに具備することを特徴とする[1]に記載の情報処理装置である。

#### 【0008】

[3]の発明は、前記ログ記憶手段は、前記対訳が正しいか否かを示す正解情報を付加して記憶しており、前記ログ記憶手段に記憶された正解情報を教師データとして、該ログ記憶手段に記憶されているログを用いて、前記推定手段における推定処理を学習する学習手段をさらに具備し、前記推定手段は、前記学習手段によって学習された推定処理にしたがって、対訳の正確性を推定することを特徴とする[1]又は[2]に記載の情報処理装置である。

10

#### 【0009】

[4]の発明は、対訳記憶手段とログ記憶手段を具備する情報処理装置が行う情報処理方法であって、前記対訳記憶手段は、機械翻訳システムが原Webページから翻訳Webページの翻訳に用いた第1の言語の単語と対訳である第2の言語の単語を記憶しており、第1の言語で記載された原Webページを機械翻訳システムによって第2の言語に翻訳した翻訳Webページに対して検索を行うための、該第2の言語の単語を含む検索クエリを受け付けるクエリ受付ステップと、前記クエリ受付ステップによって受け付けられた検索クエリによって検索された翻訳Webページに対する操作者の操作のログを前記ログ記憶手段に記憶させるステップと、前記ログ記憶手段に記憶されているログを用いて、前記クエリ受付ステップによって受け付けられた検索クエリ内の単語に対応する前記対訳記憶手段に記憶されている対訳の正確性を推定する推定ステップを具備することを特徴とする情報処理方法である。

20

#### 【0010】

[5]の発明は、コンピュータを、第1の言語で記載された原Webページを機械翻訳システムによって第2の言語に翻訳した翻訳Webページに対して検索を行うための、該第2の言語の単語を含む検索クエリを受け付けるクエリ受付手段と、前記機械翻訳システムが前記原Webページから前記翻訳Webページの翻訳に用いた第1の言語の単語と対訳である第2の言語の単語を記憶している対訳記憶手段と、前記クエリ受付手段によって受け付けられた検索クエリによって検索された翻訳Webページに対する操作者の操作のログを記憶しているログ記憶手段と、前記ログ記憶手段に記憶されているログを用いて、前記クエリ受付手段によって受け付けられた検索クエリ内の単語に対応する前記対訳記憶手段に記憶されている対訳の正確性を推定する推定手段として機能させるための情報処理プログラムである。

30

#### 【発明の効果】

#### 【0011】

本発明にかかる情報処理装置、情報処理方法及び情報処理プログラムによれば、検索サービスにおいて翻訳品質が求められる検索クエリ内の単語の翻訳の正確性を推定することができる。

40

#### 【図面の簡単な説明】

#### 【0012】

【図1】第1の実施の形態の構成例についての概念的なモジュール構成図である。

【図2】本実施の形態における主にデータの流れの例を示す説明図である。

【図3】第1の実施の形態を実現するためのシステム構成例を示す説明図である。

【図4】第1の実施の形態における処理例を示すフローチャートである。

【図5】対訳テーブルのデータ構造例を示す説明図である。

50

【図6】ログのデータ構造例を示す説明図である。

【図7】第2の実施の形態の構成例についての概念的なモジュール構成図である。

【図8】第2の実施の形態における学習処理例を示すフローチャートである。

【図9】ログのデータ構造例を示す説明図である。

【図10】本実施の形態を実現するコンピュータのハードウェア構成例を示すブロック図である。

【発明を実施するための形態】

【0013】

以下、図面に基づき本発明を実現するにあたっての好適な各種の実施の形態の例を説明する。

図1は、第1の実施の形態の構成例についての概念的なモジュール構成図を示している。

なお、モジュールとは、一般的に論理的に分離可能なソフトウェア（コンピュータ・プログラム）、ハードウェア等の部品を指す。したがって、本実施の形態におけるモジュールはコンピュータ・プログラムにおけるモジュールのことだけでなく、ハードウェア構成におけるモジュールも指す。それゆえ、本実施の形態は、それらのモジュールとして機能させるためのコンピュータ・プログラム（コンピュータにそれぞれの手順を実行させるためのプログラム、コンピュータをそれぞれの手段として機能させるためのプログラム、コンピュータにそれぞれの機能を実現させるためのプログラム）、システム及び方法の説明をも兼ねている。ただし、説明の都合上、「記憶する」、「記憶させる」、これらと同等の文言を用いるが、これらの文言は、実施の形態がコンピュータ・プログラムの場合は、記憶装置に記憶させる、又は記憶装置に記憶させるように制御するの意である。また、モジュールは機能に一対一に対応していてもよいが、実装においては、1モジュールを1プログラムで構成してもよいし、複数モジュールを1プログラムで構成してもよく、逆に1モジュールを複数プログラムで構成してもよい。また、複数モジュールは1コンピュータによって実行されてもよいし、分散又は並列環境におけるコンピュータによって1モジュールが複数コンピュータで実行されてもよい。なお、1つのモジュールに他のモジュールが含まれていてもよい。また、以下、「接続」とは物理的な接続の他、論理的な接続（データの授受、指示、データ間の参照関係等）の場合にも用いる。「予め定められた」とは、対象としている処理の前に定まっていることをいい、本実施の形態による処理が始まる前はもちろんのこと、本実施の形態による処理が始まった後であっても、対象としている処理の前であれば、そのときの状況・状態に応じて、又はそれまでの状況・状態に応じて定まることの意を含めて用いる。「予め定められた値」が複数ある場合は、それぞれ異なる値であってもよいし、2以上の値（もちろんのことながら、全ての値も含む）が同じであってもよい。

また、システム又は装置とは、複数のコンピュータ、ハードウェア、装置等がネットワーク（一対一対応の通信接続を含む）等の通信手段で接続されて構成されるほか、1つのコンピュータ、ハードウェア、装置等によって実現される場合も含まれる。「装置」と「システム」とは、互いに同義の用語として用いる。もちろんのことながら、「システム」には、人為的な取り決めである社会的な「仕組み」（社会システム）にすぎないものは含まれない。

また、各モジュールによる処理毎に又はモジュール内で複数の処理を行う場合はその処理毎に、対象となる情報を記憶装置から読み込み、その処理を行った後に、処理結果を記憶装置に書き出すものである。したがって、処理前の記憶装置からの読み込み、処理後の記憶装置への書き出しについては、説明を省略する場合がある。なお、ここでの記憶装置としては、ハードディスク、RAM(Random Access Memory)、外部記憶媒体、通信回線を介した記憶装置、CPU(Central Processing Unit)内のレジスタ等を含んでいてもよい。

【0014】

第1の実施の形態である情報処理装置100は、検索クエリ受付モジュール110、検

10

20

30

40

50

索モジュール120、翻訳モジュール130、対訳記憶モジュール140、ログ収集モジュール150、ログ記憶モジュール160、正確性推定モジュール170、対訳変更モジュール180を有している。

#### 【0015】

検索クエリ受付モジュール110は、検索モジュール120、ログ記憶モジュール160と接続されている。検索クエリ受付モジュール110は、第1の言語で記載された原Webページを機械翻訳システムによって第2の言語に翻訳した翻訳Webページに対して検索を行うための、その第2の言語の単語を含む検索クエリを受け付ける。ここで第1の言語は、原Webページの記載に用いられている言語であり、例えば、英語等がある。第2の言語は、原Webページにアクセスしたいユーザが読むことができる言語であり、例えば、日本語等がある。そのユーザにとっては、一般的に、第1の言語は外国語であり、第2の言語は母国語となる。第1のWebページは、機械翻訳システムによって翻訳されたWebページである。また、検索クエリ内には、第2の言語の単語が含まれ、複数あってもよい。例えば、複数の単語のAND検索、OR検索等である。一般的な例として、検索クエリ受付モジュール110は、Webブラウザを通じてユーザが入力した検索クエリを受け付ける。例えば、ショップサイトの検索等がある。

検索モジュール120は、検索クエリ受付モジュール110と接続されている。検索モジュール120は、検索クエリ受付モジュール110が受け付けた検索クエリにしたがって、翻訳Webページを検索する。従来からある検索システムを用いればよい。

#### 【0016】

翻訳モジュール130は、対訳記憶モジュール140と接続されている。翻訳モジュール130は、前述の機械翻訳システム内の翻訳処理を行うモジュールであって、対訳記憶モジュール140内の対訳テーブルを用いて、原Webページから翻訳を行って、翻訳Webページを作成する。また、翻訳モジュール130は、翻訳Webページを作成する際に用いた対訳の組をログ記憶モジュール160に記憶させてもよい。

対訳記憶モジュール140は、翻訳モジュール130、正確性推定モジュール170、対訳変更モジュール180と接続されている。対訳記憶モジュール140は、機械翻訳システムが原Webページから翻訳Webページの翻訳に用いた第1の言語の単語と対訳である第2の言語の単語を記憶している。例えば、対訳テーブル500を記憶している。図5は、対訳テーブル500のデータ構造例を示す説明図である。対訳テーブル500は、第1言語欄510、第2言語欄520、翻訳確率欄530を有している。第1言語欄510は、第1の言語である単語を記憶している。第2言語欄520は、その第1の言語の単語の対訳である第2の言語の単語を記憶している。翻訳確率欄530は、その対訳を用いる確率を記憶している。対訳である単語対応には複数の可能性がある。ここでの確率は、これらの単語対応を採用する可能性（確率、条件付き確率）を示している。一般的に、第1の言語の単語が多義語である場合、対訳は複数あり、その第1の言語の単語が用いられているWebページ内の文脈に応じて、対訳が選択される。

#### 【0017】

ログ収集モジュール150は、ログ記憶モジュール160と接続されている。ログの収集は、翻訳Webページを表示している端末が収集し、その収集したログをログ収集モジュール150に対して送信してもよいし、その翻訳Webページに対する指示を受け付けるWebサーバーがログを収集し、その収集したログをログ収集モジュール150に対して送信してもよい。収集するログとしては、(1)翻訳WebページのURL(Uniform Resource Locator)、(2)滞在時間(次のWebページへ移動するまでの期間(翻訳Webページの表示期間))、(3)検索結果を示す翻訳Webページがクリックされたか否かを示す情報(その翻訳Webページが表示されたか否かを示す情報)、(4)検索クエリ内で用いられた単語毎の対訳、等がある。また、これらに限られず、以下のものを収集してもよい。

##### (1) クリックログ

以下の操作の基本データとなるものであり、Webページ上でマウス操作(カーソル

10

20

30

40

50

の移動操作、右ボタン又は左ボタン等の選択操作)をログとして記憶する。また、Webページを表示する装置がタッチパネルである場合は、指等の操作(タップ、ドラッグ、フリック、ピンチイン、ピンチアウト、長押し、シェイク等)をログとして記憶する。このクリックログを解析して、以下の操作をログとして収集してもよい。

#### (2) スクロールに関する操作

具体的には、スクロールが行われた回数、スクロールによる移動距離、移動方向等がある。

#### (3) 原Webページ(原言語(第1の言語)のページ)に関する操作

「原Webページ」を「翻訳Webページ」とともに表示する。例えば、タグ等を用いて原Webページを表示できるようにしてよい。

具体的には、「原Webページ」への移動回数、「原Webページ」の滞在時間等がある。タグを用いた場合は、「原Webページ」のタグが選択された回数が「原Webページ」への移動回数となる。「原Webページ」のタグが選択されてから、他のタグが選択されるまでの期間又は次のWebページへ移動するまでの期間が、「原Webページ」の滞在時間となる。

#### (4) 「いいね！」ボタン等に関する操作

ソーシャルネットワーキングサービス(例えば、フェイスブック、ツイッター)等における「いいね！」ボタンに関する操作である。

具体的には、「いいね！」ボタン等が選択された回数等がある。

#### (5) 売上等に関する操作

対象としている翻訳Webページが商用目的である場合は、購入の回数、売上額、予約数等がある。

#### (6) コンバージョン(CV)に関する操作

コンバージョンとは、商用目的等のウェブサイト上で獲得できる最終的な成果をいい、商用目的の翻訳Webページならば商品購入、情報提供やコミュニティのための翻訳Webページならば会員登録等がコンバージョンにあたる。なお、資料請求、問い合わせ等の一次的な成果をコンバージョンとして含めてもよい。また、コンバージョン率(Webページへのアクセス数に対して、コンバージョンに結びついた件数の割合)を収集してもよい。

#### (7) 辞書検索への遷移に関する操作

具体的には、「原Webページ」、「翻訳Webページ」内の単語に対して、辞書検索を行った回数、辞書の種類、検索対象となった単語等がある。

#### (8) 検索結果ページからの流入トラフィック

つまり、インターネットの検索システムにおける検索による検索結果から対象としているWebページへの流入の回数(いわゆる検索エンジンを経由したWebページへの訪問回数)を操作ログとする。

この操作ログと翻訳品質とは一定の相関がある。以下に説明する。翻訳品質が悪いと、そもそもユーザの検索クエリと一致しない可能性がある。また、検索結果のページで表示される部分の翻訳品質が悪い場合、例えば、商品タイトルの翻訳品質が悪い場合、又は、検索結果で一部だけ表示する本文の翻訳品質が悪い場合、検索結果であるWebページに対するクリックの可能性が低くなる。

#### 【0018】

ログ記憶モジュール160は、検索クエリ受付モジュール110、ログ収集モジュール150、正確性推定モジュール170と接続されている。ログ記憶モジュール160は、検索クエリ受付モジュール110によって受け付けられた検索クエリによって検索された翻訳Webページに対する操作者の操作のログを記憶している。例えば、ログ600を記憶している。図6は、ログ600のデータ構造例を示す説明図である。ログ600は、クエリ欄610、ページ欄620、滞在時間欄630、クリック欄640、対訳1欄650、対訳2欄660等を有している。クエリ欄610は、検索クエリ受付モジュール110が収集した検索クエリを記憶している。ここでは、検索クエリに含まれている単語の列を

10

20

30

40

50

記憶している。ページ欄 620 は、その検索クエリによって検索された翻訳 Web ページ（検索クエリに該当すると表示された翻訳 Web ページ）の URL を記憶している。滞在時間欄 630 は、その翻訳 Web ページの滞在時間を記憶している。クリック欄 640 は、その翻訳 Web ページがクリックされたか否かを示す情報を記憶している。対訳 1 欄 650 は、検索クエリ内の第 1 の単語について、その翻訳 Web ページ内での対訳を記憶している。つまり、機械翻訳システムが、その翻訳 Web ページを翻訳した際に用いた対訳テーブル 500 の対訳の組を記憶している。対訳 2 欄 660 は、検索クエリ内の第 2 の単語について、対訳 1 欄 650 と同等の内容を記憶している。もちろんのことながら、検索クエリ内の単語の数が 3 以上の場合は、対訳 1 欄 650 と同等の欄が続き、検索クエリ内の単語の数が 1 である場合は、対訳 2 欄 660 は不要（対訳 2 欄 660 に NULL を記憶させてもよい）である。

10

#### 【0019】

正確性推定モジュール 170 は、対訳記憶モジュール 140、ログ記憶モジュール 160、対訳変更モジュール 180 と接続されている。正確性推定モジュール 170 は、ログ記憶モジュール 160 に記憶されているログを用いて、検索クエリ受付モジュール 110 によって受け付けられた検索クエリ内の単語に対応する対訳記憶モジュール 140 に記憶されている対訳の正確性を推定する。

20

例えば、検索モジュール 120 による出てきた検索結果には検索クエリの各語が含まれるが、正しい訳となっているページのクリック数は多く滞在時間も長くなり、誤った訳となっているページのクリック数は少なく滞在時間も短くなると考えられる。翻訳モジュール 130 は対訳記憶モジュール 140 を用いて翻訳しているので、検索クエリ内の各語の対訳は判明する。各対訳の正確性は、ログ記憶モジュール 160 に記憶されているログを手がかりとして用いて推定する。つまり、クリックされており（ログ 600 のクリック欄 640 を参照）、滞在時間（ログ 600 の滞在時間欄 630 参照）が予め定められた値よりも長い翻訳 Web ページでの対訳の正確性の値を大きく、逆の傾向がある対訳の正確性の値を小さくして、対訳記憶モジュール 140 に反映する。より具体的には、滞在時間が予め定められた値よりも長い翻訳 Web ページの場合は、その翻訳 Web ページにおける対訳の翻訳確率に予め定められた係数（1 以上の係数）を乗算して、正確性の値としてもよい。クリックされなかった翻訳 Web ページ又は滞在時間が予め定められた値以下である翻訳 Web ページの場合は、その翻訳 Web ページにおける対訳の翻訳確率に予め定められた係数（1 未満の係数）を乗算して、正確性の値としてもよい。また、正確性の値は、元の翻訳確率を用いずに、滞在時間に比例した値としてもよい。

30

なお、対訳の正確性を推定に用いるログとしては、クリックされたか否か、滞在時間の情報以外に、前述した翻訳 Web ページ上のクリックログ、スクロールに関する操作、原 Web ページに関する操作、「いいね！」ボタン等に関する操作、売上等に関する操作、コンバージョン（CV）に関する操作、辞書検索への遷移に関する操作等がある。例えば、翻訳 Web ページ上のクリックの回数、スクロール回数、「いいね！」ボタンの選択回数、売上、コンバージョン率が多い（高い）ほど、その対訳の正確性の値を大きくするようにしてよい。又は、原 Web ページへの移動が少ないほど、辞書検索への遷移回数が少ないほど、その対訳の正確性の値を大きくするようにしてもよい。それぞれ逆の場合は、その対訳の正確性の値を小さくするようにしてもよい。

40

#### 【0020】

対訳変更モジュール 180 は、対訳記憶モジュール 140、正確性推定モジュール 170 と接続されている。対訳変更モジュール 180 は、正確性推定モジュール 170 によって推定された正確性に基づいて、対訳記憶モジュール 140 に記憶されている対訳を変更する。ここでの対訳の変更は、対訳テーブル 500 内の第 2 言語欄 520 内の対訳の変更の他に、翻訳確率欄 530 内の翻訳確率の変更であってもよい。例えば、正確性の値が予め定められた値未満である場合は、その対訳を対訳記憶モジュール 140 から削除するようにしてもよい。

50

また、正確性の値を翻訳確率と同等に扱ってもよい。つまり、前述の正確性推定モジ

ール 170 が推定する正確性の値を翻訳確率と読み替えて、正確性推定モジュール 170 が算出した正確性の値を新たな翻訳確率として、対訳記憶モジュール 140 内の対応する翻訳確率を変更するようにしてもよい。

翻訳確率を調整して機械翻訳システムで再度、原 Web ページを翻訳（翻訳 Web ページを作成）することにより、その語を含む原 Web ページの翻訳結果が改善し、それまで異なる語に翻訳されていたために検索結果に現れなかつた翻訳 Web ページが現れるようになる。また、検索クエリとして現れやすい単語の翻訳精度が優先的に改善していくため、検索結果や売上の改善が見込まれる。

#### 【0021】

図 2 は、本実施の形態（情報処理装置 100、機械翻訳システム 220）における主にデータの流れの例を示す説明図である。10

機械翻訳システム 220 は、海外ショッピングサイト 210A を翻訳して翻訳ショッピングサイト 210B を作成する。そして、ユーザ 201、202 等による操作によって検索可能なようにインターネット上で公開される。ユーザ 201、202 等は、それぞれが所持している端末（ノート PC、スマートフォン等を含む携帯情報端末等）を用いて、Web ブラウザを用いて検索クエリを入力する。その検索クエリは、検索クエリ受付モジュール 110 によって受け付けられ、検索モジュール 120 が翻訳 Web ページに対して検索を行い、翻訳ショッピングサイト 210B を検索結果として提示する。ユーザ 201、202 等は、翻訳ショッピングサイト 210B 上で操作を行う。これらの操作の履歴等は、ログ収集モジュール 150 が収集し、ログ記憶モジュール 160 内にログ 230 として格納する。正確性推定モジュール 170、対訳変更モジュール 180 は、対訳の翻訳確率算出結果 240 を算出する。そして、対訳変更モジュール 180 は、対訳の翻訳確率算出結果 240 を用いて、機械翻訳システム 220 内の対訳を修正する。20

機械翻訳システム 220 は、修正された対訳を用いて海外ショッピングサイト 210A を再度翻訳する。つまり、新しく作成された翻訳ショッピングサイト 210B は、以前の翻訳ショッピングサイト 210B よりも翻訳の正確性をより増した Web ページとなっている。

なお、海外ショッピングサイト 210A、翻訳ショッピングサイト 210B は、1 つの Web ページだけではなく、複数の Web ページからなり、ある Web ページに対するログ 230 から算出された対訳の翻訳確率算出結果 240 が他の Web ページの再翻訳に活用されることになる。30

#### 【0022】

図 3 は、第 1 の実施の形態を実現するためのシステム構成例を示す説明図である。Web ページ評価システム 300 は、情報処理装置 100、Web ページサーバー 310、ログ収集装置 320 を有している。Web ページ評価システム 300、機械翻訳システム 220、端末 340A、端末 340B、端末 340C、端末 340D、端末 340E は、通信回線 399 を介してそれぞれ接続されている。図 1 の例に示した翻訳モジュール 130、対訳記憶モジュール 140 は、機械翻訳システム 220 内に含まれていてもよく、ログ収集モジュール 150 はログ収集装置 320 内に含まれていてもよい。

例えば、原 Web ページを機械翻訳システム 220 が翻訳し、その翻訳結果である Web ページを Web ページサーバー 310 内に記憶させる。40

そして、端末 340A、端末 340B 等は、検索クエリを入力し、その検索結果である Web ページサーバー 310 内の Web ページにアクセスする。その際に、端末 340A、端末 340B 等が表示している Web ページに対する操作のログをログ収集装置 320 が収集する。例えば、端末 340A、端末 340B 等が操作を検知し、その検知結果を、通信回線 399 を介してログ収集装置 320 へ送信する。そして、ログ収集装置 320 が収集したログを情報処理装置 100 に渡し、情報処理装置 100 は検索クエリ内に用いられている単語の対訳の翻訳の正確性を推定し、対訳を調整する。機械翻訳システム 220 は、調整された対訳を用いて、原 Web ページを再度翻訳する。

#### 【0023】

50

20

30

40

50

図4は、第1の実施の形態における処理例を示すフローチャートである。

ステップS402では、翻訳モジュール130が、機械翻訳Webページを準備する。

ステップS404では、検索クエリ受付モジュール110が、検索クエリを受け付ける。

ステップS406では、ログ収集モジュール150が、機械翻訳Webページに対するログを収集する。

ステップS408では、正確性推定モジュール170が、翻訳確率を算出する。この例は、正確性の値として翻訳確率を用いたものである。

ステップS410では、対訳変更モジュール180が、翻訳確率が閾値以上か否かを判断し、閾値以上の場合は処理を終了し(ステップS499)、それ以外の場合はステップS412へ進む。

ステップS412では、対訳変更モジュール180が、検索クエリ内に用いられた単語の対訳データを修正する。

さらに、この後に、翻訳モジュール130が、修正された対訳データを用いて、機械翻訳Webページを作成し直してもよい。

#### 【0024】

図7は、第2の実施の形態の構成例についての概念的なモジュール構成図である。

情報処理装置100は、検索クエリ受付モジュール110、検索モジュール120、翻訳モジュール130、対訳記憶モジュール140、ログ収集モジュール150、ログ記憶モジュール160、正確性推定モジュール170、対訳変更モジュール180、学習モジュール750を有している。なお、第1の実施の形態と同種の部位には同一符号を付し重複した説明を省略する。

対訳記憶モジュール140は、翻訳モジュール130、正確性推定モジュール170、対訳変更モジュール180、学習モジュール750と接続されている。

ログ記憶モジュール160は、検索クエリ受付モジュール110、ログ収集モジュール150、正確性推定モジュール170、学習モジュール750と接続されている。ログ記憶モジュール160は、対訳が正しいか否かを示す正解情報を附加して記憶している。例えば、ログ900を記憶している。図9は、ログ900のデータ構造例を示す説明図である。ログ900は、クエリ欄910、ページ欄920、滞在時間欄930、クリック欄940、対訳1欄950、対訳2欄960、教師データ欄970を有している。ログ900は、ログ600に教師データ欄970を附加したものである。教師データ欄970は、その翻訳Webページにおいて、その対訳が正しいことを示す情報を記憶している。対訳が正しいか否かについては、人手で(翻訳者等が)判断した結果である。翻訳については機械翻訳システム220を用いることが前提であるため、全ての対訳について、対訳が正しいか否かを判断しているわけではない。つまり、正確性推定方法の学習のために教師データが必要であり、その教師データとして正しい対訳を用いる。そのために、ピックアップされた検索結果に対して判断を行うものである。

#### 【0025】

学習モジュール750は、対訳記憶モジュール140、ログ記憶モジュール160、正確性推定モジュール170と接続されている。学習モジュール750は、ログ記憶モジュール160に記憶された正解情報を教師データとして、ログ記憶モジュール160に記憶されているログを用いて、正確性推定モジュール170における推定処理を学習する。ここで学習は、正しい対訳における正確性の値が高くなるように、滞在時間(滞在時間欄930)、クリック(クリック欄940)のログから正確性の値を算出する式の係数、閾値である前述の予め定められた値を、学習によって求める。なお、前述したように、正確性推定方法のために、滞在時間、クリック以外のログを用いてもよい。したがって、学習においても、それらのログに基づいて行う。学習方法として、具体的には、ニューラルネットワークを用いた学習によって係数等を定めてよい。

正確性推定モジュール170は、対訳記憶モジュール140、ログ記憶モジュール160、対訳変更モジュール180、学習モジュール750と接続されている。正確性推定モ

10

20

30

40

50

ジユール 170 は、学習モジュール 750 によって学習された推定処理にしたがって、対訳の正確性を推定する。したがって、第 1 の実施の形態のように予め定められた係数等を用いて正確性を推定するよりも、実際のログを反映した正確な推定処理を行うことができるようになる。

#### 【0026】

図 8 は、第 2 の実施の形態における学習処理例を示すフローチャートである。

ステップ S802 では、翻訳モジュール 130 が、機械翻訳 Web ページを準備する。

ステップ S804 では、検索クエリ受付モジュール 110 が、検索クエリを受け付ける。

ステップ S806 では、ログ収集モジュール 150 が、機械翻訳 Web ページに対するログを収集する。 10

ステップ S808 では、学習モジュール 750 が、教師データのログを用いて翻訳確率算出モデルを生成する。この例は、正確性の値として翻訳確率を用いたものである。

ステップ S810 では、正確性推定モジュール 170 が、翻訳確率算出モデルを組み込む。

この後、図 4 の例に示すフローチャートにしたがった処理を行って、対訳を変更する。なお、変更対象とする対訳は、教師データ以外の対訳としてもよい。

#### 【0027】

なお、本実施の形態としてのプログラムが実行されるコンピュータのハードウェア構成は、図 10 に例示するように、一般的なコンピュータであり、具体的にはパーソナルコンピュータ、サーバーとなり得るコンピュータ、携帯電話（スマートフォンを含む）等である。つまり、具体例として、処理部（演算部）として CPU1001 を用い、記憶装置として RAM1002、ROM1003、HD1004 を用いている。HD1004 として、例えばハードディスクを用いてもよい。検索クエリ受付モジュール 110、検索モジュール 120、翻訳モジュール 130、ログ収集モジュール 150、正確性推定モジュール 170、対訳変更モジュール 180、学習モジュール 750 等のプログラムを実行する CPU1001 と、そのプログラムやデータを記憶する RAM1002 と、本コンピュータを起動するためのプログラム等が格納されている ROM1003 と、補助記憶装置である HD1004 と、キーボード、マウス、タッチパネル等に対する利用者の操作に基づいてデータを受け付ける受付装置 1006 と、液晶ディスプレイ等の出力装置 1005 と、ネットワークインターフェースカード等の通信ネットワークと接続するための通信回線インターフェース 1007、そして、それらをつないでデータのやりとりをするためのバス 1008 により構成されている。これらのコンピュータが複数台互いにネットワークによって接続されていてもよい。

#### 【0028】

前述の実施の形態のうち、コンピュータ・プログラムによるものについては、本ハードウェア構成のシステムにソフトウェアであるコンピュータ・プログラムを読み込ませ、ソフトウェアとハードウェア資源とが協働して、前述の実施の形態が実現される。

なお、図 10 に示すハードウェア構成は、1 つの構成例を示すものであり、本実施の形態は、図 10 に示す構成に限らず、本実施の形態において説明したモジュールを実行可能な構成であればよい。例えば、一部のモジュールを専用のハードウェア（例えば ASIC 等）で構成してもよく、一部のモジュールは外部のシステム内にあり通信回線で接続しているような形態でもよく、さらに図 10 に示すシステムが複数互いに通信回線によって接続されていて互いに協調動作するようにしてもよい。

#### 【0029】

また、前述の実施の形態の説明において、予め定められた値との比較において、「以上」、「以下」、「より大きい」、「より小さい（未満）」としたものは、その組み合わせに矛盾が生じない限り、それぞれ「より大きい」、「より小さい（未満）」、「以上」、「以下」としてもよい。

#### 【0030】

10

20

30

40

50

なお、説明したプログラムについては、記録媒体に格納して提供してもよく、また、そのプログラムを通信手段によって提供してもよい。その場合、例えば、前記説明したプログラムについて、「プログラムを記録したコンピュータ読み取り可能な記録媒体」の発明として捉えてもよい。

「プログラムを記録したコンピュータ読み取り可能な記録媒体」とは、プログラムのインストール、実行、プログラムの流通などのために用いられる、プログラムが記録されたコンピュータで読み取り可能な記録媒体をいう。

なお、記録媒体としては、例えば、デジタル・バーサタイル・ディスク（DVD）であって、DVDフォーラムで策定された規格である「DVD-R、DVD-RW、DVD-RAM等」、DVD+RWで策定された規格である「DVD+R、DVD+RW等」、コンパクトディスク（CD）であって、読み出し専用メモリ（CD-ROM）、CDレコーダブル（CD-R）、CDリライタブル（CD-RW）等、ブルーレイ・ディスク（Blu-ray（登録商標）Disc）、光磁気ディスク（MO）、フレキシブルディスク（FD）、磁気テープ、ハードディスク、読み出し専用メモリ（ROM）、電気的消去及び書換可能な読み出し専用メモリ（EEPROM（登録商標））、フラッシュ・メモリ、ランダム・アクセス・メモリ（RAM）、SD（Secure Digital）メモリーカード等が含まれる。10

そして、前記のプログラム又はその一部は、前記記録媒体に記録して保存や流通等させてもよい。また、通信によって、例えば、ローカル・エリア・ネットワーク（LAN）、メトロポリタン・エリア・ネットワーク（MAN）、ワイド・エリア・ネットワーク（WAN）、インターネット、イントラネット、エクストラネット等に用いられる有線ネットワーク、あるいは無線通信ネットワーク、さらにこれらの組み合わせ等の伝送媒体を用いて伝送させてもよく、また、搬送波に乗せて搬送させてもよい。20

さらに、前記のプログラムは、他のプログラムの一部分であってもよく、あるいは別個のプログラムと共に記録媒体に記録されていてもよい。また、複数の記録媒体に分割して記録されていてもよい。また、圧縮や暗号化など、復元可能であればどのような様態で記録されていてもよい。

#### 【符号の説明】

#### 【0031】

100...情報処理装置30

110...検索クエリ受付モジュール

120...検索モジュール

130...翻訳モジュール

140...対訳記憶モジュール

150...ログ収集モジュール

160...ログ記憶モジュール

170...正確性推定モジュール

180...対訳変更モジュール

300...Webページ評価システム

310...Webページサーバー40

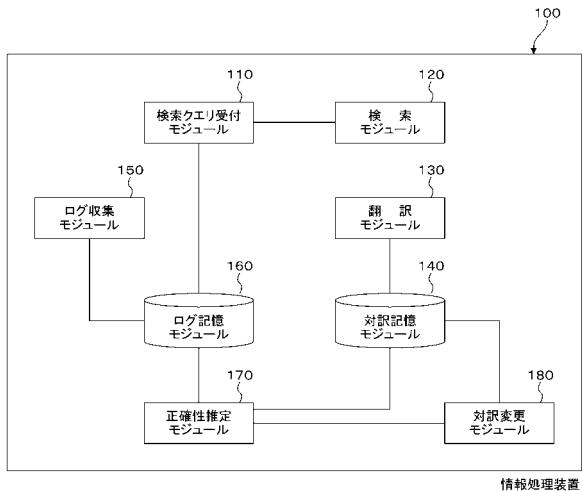
320...ログ収集装置

340...端末

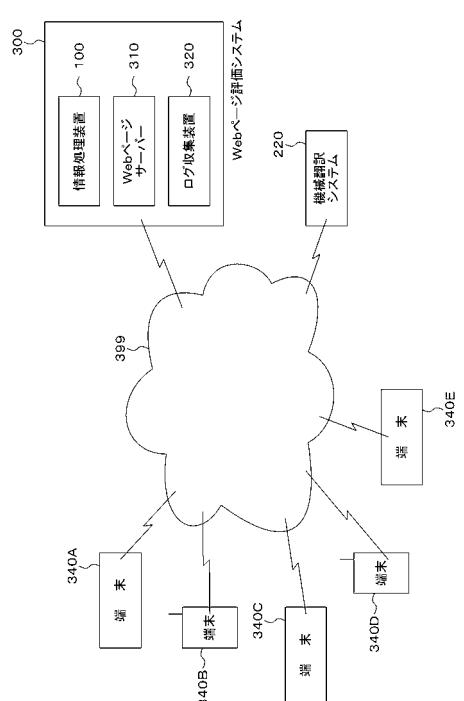
399...通信回線

750...学習モジュール

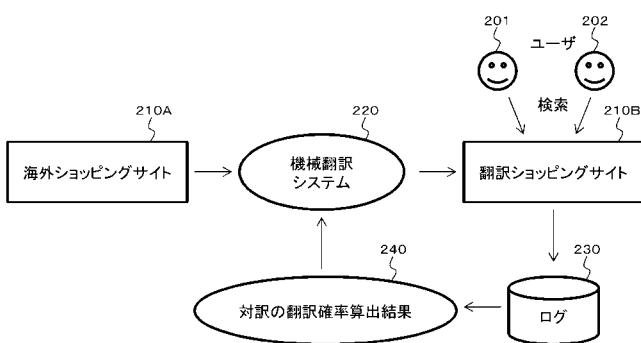
【図1】



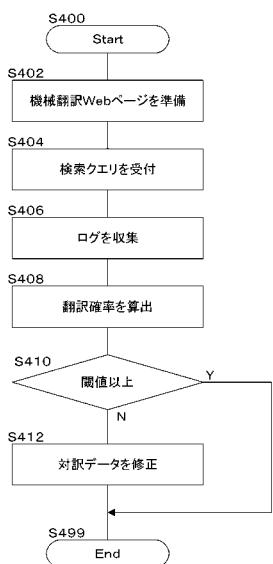
【図3】



【図2】



【図4】



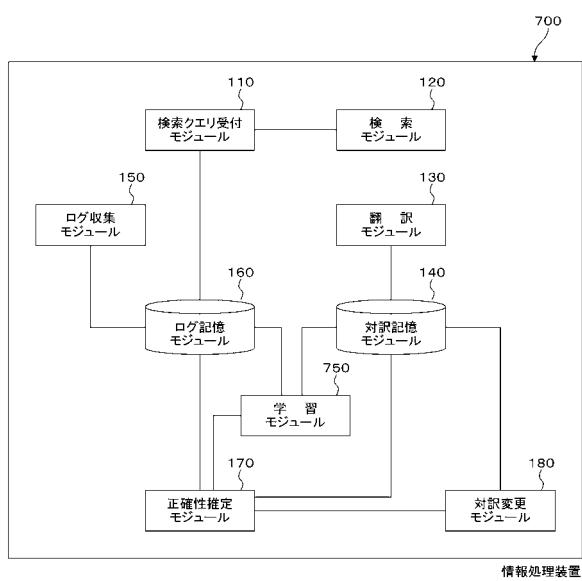
| 第1言語 | 第2言語 | 翻訳確率 |
|------|------|------|
| a    | 財布   | ...  |
| b    | 財布   | ...  |
| c    | 財布   | ...  |
| ⋮    | ⋮    | ⋮    |
| x    | 革    | ⋮    |
| y    | 革    | ⋮    |
| z    | 革    | ⋮    |
| ⋮    | ⋮    | ⋮    |

【図5】

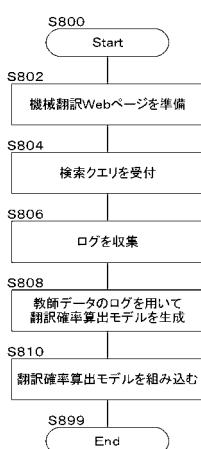
【図6】

| クエリ | ページ       | 滞在時間 | クリック | 対訳1    | 対訳2   | ... |
|-----|-----------|------|------|--------|-------|-----|
| 財布革 | www/test1 | 0:00 | ×    | 財布 ⇔ a | 革 ⇔ x | ... |
| 財布革 | www/test2 | 0:30 | ○    | 財布 ⇔ b | 革 ⇔ y | ... |
| 財布革 | www/test3 | 0:01 | ○    | 財布 ⇔ c | 革 ⇔ z | ... |
| ... | ...       | ...  | ...  | ...    | ...   | ... |

【図7】



【図8】



【図9】

| クエリ | ページ       | 滞在時間 | クリック | 対訳1    | 教師データ | ... |
|-----|-----------|------|------|--------|-------|-----|
| 財布革 | www/test1 | 0:00 | ×    | 財布 ⇔ a | 革 ⇔ x | 0   |
| 財布革 | www/test2 | 0:30 | ○    | 財布 ⇔ b | 革 ⇔ y | 1   |
| 財布革 | www/test3 | 0:01 | ○    | 財布 ⇔ c | 革 ⇔ z | 1   |
| ... | ...       | ...  | ...  | ...    | ...   | ... |

【図10】

