

(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2006-276915

(P2006-276915A)

(43) 公開日 平成18年10月12日(2006.10.12)

(51) Int.CI.

GO6F 17/27

(2006.01)

F 1

GO6F 17/27

テーマコード(参考)

D

5B091

審査請求 未請求 請求項の数 8 O L (全 11 頁)

(21) 出願番号	特願2005-90203 (P2005-90203)	(71) 出願人	000005496 富士ゼロックス株式会社 東京都港区赤坂二丁目17番22号
(22) 出願日	平成17年3月25日 (2005.3.25)	(74) 代理人	100098084 弁理士 川▲崎▼ 研二
		(72) 発明者	長尾 隆 神奈川県海老名市本郷2274番地 富士 ゼロックス株式会社内
		(72) 発明者	館野 昌一 神奈川県足柄上郡中井町境430 グリー ンテクなかい 富士ゼロックス株式会社内
		(72) 発明者	田中 圭 神奈川県足柄上郡中井町境430 グリー ンテクなかい 富士ゼロックス株式会社内

最終頁に続く

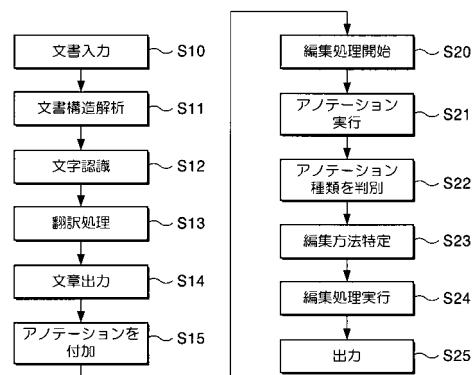
(54) 【発明の名称】翻訳処理方法、文書翻訳装置およびプログラム

(57) 【要約】

【課題】 高品質の翻訳文を取得する。

【解決手段】 ユーザは翻訳処理後の文書において誤訳や不適切な翻訳処理がなされている箇所がないかをチェックし、所望する編集方法に応じたアノテーションを当該翻訳後の文章に付加する。所定の指示を入力して編集対象箇所とアノテーションを確定させると、アノテーションが付加された状態の文書に対応する画像データが生成され、この画像データに対して編集処理（再翻訳処理）が開始される。文章構造解析が行われ文字情報とアノテーションとが分離され、各アノテーションに対し、そのアノテーションが付加されている対象の訳語の箇所と当該アノテーションの種類とが判別される。続いて、翻訳規則テーブルTrを参照し、判別されたアノテーションの種類に対応する編集方法を特定し、当該編集方法に従って編集処理（再翻訳処理）を行い、所定の方法で出力する。

【選択図】 図2



【特許請求の範囲】**【請求項 1】**

アノテーションの種類と翻訳方法とを対応付けてテーブルに登録する登録ステップと、文書を入力する入力ステップと、
前記入力ステップにて入力された文書から、アノテーションと文字情報とを抽出する抽出ステップと、
前記抽出ステップにて抽出されたアノテーションの種類と、当該アノテーションの付加対象となっている文書要素とを特定するアノテーション特定ステップと、
前記テーブルを参照し前記種類に対応する翻訳方法を決定する翻訳方法決定ステップと、
、
前記アノテーション特定ステップにて特定された文書要素に対し、前記編集方法決定ステップにて決定された翻訳方法を適用して翻訳処理を行う翻訳実行ステップと、
を有する翻訳処理方法。

【請求項 2】

アノテーションの種類と編集方法とを対応付けてテーブルに登録する登録ステップと、文書を入力する文書入力ステップと、
前記文書入力ステップにて入力された文書に対し翻訳処理を行う翻訳ステップと、
前記翻訳ステップにて翻訳された文章を提示して、アノテーションを付加する指示を受け付ける指示入力ステップと、
前記指示入力ステップにて入力されたアノテーションの種類と、当該アノテーションの付加対象である文書要素とを特定するアノテーション特定ステップと、
前記テーブルを参照し前記アノテーション特定ステップにて特定された種類に対応する編集方法を決定する編集方法決定ステップと、
前記アノテーション特定ステップにて特定された文書要素に対し、前記編集方法決定ステップにて決定された編集方法を適用して編集処理を行う編集実行ステップと、
を有する翻訳処理方法。

【請求項 3】

前記編集方法決定ステップにて決定される編集方法は、前記編集実行ステップにて行われる編集処理として再翻訳処理と当該翻訳の際に使用する辞書とを規定したものであることを特徴とする請求項 2 に記載の翻訳処理方法。

【請求項 4】

前記編集方法決定ステップにて決定される編集方法は、前記辞書の使用の優先度を規定したものであることを特徴とする請求項 3 に記載の翻訳処理方法。

【請求項 5】

アノテーションの種類と編集方法とを対応付けてテーブルに記憶する記憶手段と、文書を入力する入力手段と、
前記入力手段にて入力された文書から、アノテーションと文字情報とを抽出する抽出手段と、
前記抽出手段にて抽出されたアノテーションの種類と当該アノテーションの付加対象である文書要素とを特定するアノテーション特定手段と、
前記テーブルを参照し前記種類に対応する翻訳方法を決定する翻訳方法決定手段と
前記アノテーション特定手段にて特定された文書要素に対し、前記翻訳方法決定手段にて決定された翻訳方法を適用して翻訳処理を行う翻訳実行手段と、
を有する文書翻訳装置。

【請求項 6】

アノテーションの種類と編集方法とを対応付けてテーブルに記憶する記憶手段と、文書を入力する文書入力手段と、
前記文書入力手段にて入力された文書を翻訳する翻訳実行手段と、
前記翻訳実行手段にて翻訳された文章を提示して、アノテーションを付加する指示を受

10

20

30

40

50

け付ける指示入力手段と、

前記指示入力手段にて入力されたアノテーションの種類と当該アノテーションの付加対象である文書要素とを特定するアノテーション特定手段と、

前記テーブルを参照し前記アノテーション特定手段にて特定された種類に対応する編集方法を決定する編集方法決定手段と

前記アノテーション特定手段にて特定された文書要素に対し、前記編集方法決定手段にて決定された編集方法を適用して編集処理を行う編集実行手段と、

を有する文書翻訳装置。

【請求項 7】

コンピュータを、

アノテーションの種類と編集方法とを対応付けて記憶する記憶手段と、

文書を入力する入力手段と、

前記入力手段にて入力された文書から文字情報とアノテーションとを抽出する抽出手段と、

前記抽出手段にて抽出されたアノテーションの種類と当該アノテーションの付加対象である文書要素とを特定するアノテーション特定手段と、

前記テーブルを参照し前記種類に対応する翻訳方法を決定する翻訳方法決定手段と、

前記アノテーション特定手段にて特定された文書要素に対し、前記翻訳方法決定手段にて決定された翻訳方法を適用して翻訳処理を行う翻訳実行手段と、

して機能させるコンピュータ読み取り可能なプログラム。

10

20

【請求項 8】

コンピュータを、

アノテーションの種類と編集方法と対応付けてをテーブルに記憶する記憶手段と、

文書を入力する文書入力手段と、

前記文書入力手段にて入力された文書を翻訳する翻訳手段と、

前記翻訳手段にて翻訳された文章を提示して、アノテーションを付加する指示を受け付ける指示入力手段と、

前記指示入力手段にて入力されたアノテーションの種類と当該アノテーションの付加対象である文字情報を特定するアノテーション特定手段と、

前記テーブルを参照し前記アノテーション特定手段にて特定された種類に対応する編集方法を決定する編集方法決定手段と

30

前記アノテーション特定手段にて特定された文字情報を対し、前記編集方法決定手段にて決定された編集方法を適用して編集処理を行う編集実行手段と、

して機能させるコンピュータ読み取り可能なプログラム。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、翻訳の品質を向上させるための技術に関する。

【背景技術】

【0002】

グローバルコミュニケーション時代の到来とともに、コンピュータを用いて、辞書データや所定のアルゴリズムを用いて文書構造を解析するなどして文字(語句)を他の文字(語句)に置換することにより、ある言語の文章を他の言語の文章に翻訳するという、いわゆる機械翻訳が盛んに行われるようになった。なお、文書が電子化されたものではない(すなわちJISコード等の文字情報がない)場合は、翻訳処理を行う前に、印刷された原稿をスキャナ装置等で読み取り文字認識処理を行って文字情報を抽出するOCR処理が行われることになる。

40

【0003】

機械翻訳を用いると大量の文書を非常に速く翻訳処理することができるという利点がある反面、一般的に、翻訳後の文書の質があまり高くないという欠点がある。その理由とし

50

ては、文書の内容（ビジネス文書であるか技術文書であるかなど）に応じて翻訳方法（例えば使用する辞書データや翻訳処理のアルゴリズム）を柔軟に替えることができず、結果として、原文の文意にそぐわない語句に置換されてしまうといったことが挙げられる。よって、翻訳された文章の品質を向上を追求するには、翻訳処理された文章を人間（ユーザ）がチェックして、不適切な翻訳後の語句を正しい語句に置き換えるといった、最終的に人間の手による何らかの修正作業を行う必要がある。翻訳処理に係るユーザの作業を支援する技術としては、幾つかのものが存在する。例えば、特許文献1には、原文の行間に、当該原文中の所定の語の訳文を表示する技術が開示されている。また、特許文献2には、原文内の所定の語句とその訳文とを一覧表示する技術が開示されている。

【特許文献1】特開平5-2606号公報 10

【特許文献2】特開平5-54072号公報

【発明の開示】

【発明が解決しようとする課題】

【0004】

しかしながら、特許文献1および2の技術においては、原文と機械翻訳による翻訳文とを対比して表示させることができるので、作業のし易さ(表示画面の見易さ)といった点では効果があるが、自らの手で一つ一つ不適切な訳語に対し、正しい訳語を入力するという煩雑な作業をユーザに強いることには変わりがない。これでは、機械翻訳を行うことによる処理速度のメリットが失われてしまう。

【0005】 20

本発明は上述した背景に鑑みてなされたものであり、ユーザに負担を掛けずに、且つ処理速度を犠牲にせずに、高品質の翻訳文書を生成することができる文書処理装置を提供することを目的とする。

【課題を解決するための手段】

【0006】

上記課題を解決するため、本発明は、アノテーションの種類と翻訳方法とを対応付けてテーブルに登録する登録ステップと、文書を入力する入力ステップと、前記入力ステップにて入力された文書から文字情報とアノテーションとを抽出する抽出ステップと、前記抽出ステップにて抽出されたアノテーションの種類と当該アノテーションの付加対象である文書要素とを特定するアノテーション特定ステップと、前記テーブルを参照し前記種類に対応する翻訳方法を決定する翻訳方法決定ステップと、前記アノテーション特定ステップにて特定された文書要素に対し、前記翻訳方法決定ステップにて決定された翻訳方法を適用して翻訳処理を行う翻訳実行ステップと、を有する翻訳処理方法を提供する。本発明によれば、ユーザが編集対象となる箇所（文書要素）を指定してアノテーションを付加することにより、当該箇所に対して翻訳処理の際に所望の翻訳方法が適用されるので、翻訳の品質を向上させることができる。 30

【0007】

本発明の翻訳処理方法は、他の態様において、アノテーションの種類と編集方法とを対応付けてテーブルに登録する登録ステップと、文書を入力する文書入力ステップと、前記文書入力ステップにて入力された文書を翻訳する翻訳ステップと、前記翻訳ステップにて翻訳された文章を提示して、アノテーションを付加する指示を受け付ける指示入力ステップと、前記指示入力ステップにて入力されたアノテーションの種類と当該アノテーションの付加対象である文書要素とを特定するアノテーション特定ステップと、前記テーブルを参照し前記アノテーション特定ステップにて特定された種類に対応する編集方法を決定する編集方法決定ステップと前記アノテーション特定ステップにて特定された文書要素に対し、前記編集方法決定ステップにて決定された編集方法を適用して編集処理を行う編集実行ステップと、を有する。 40

【0008】

好みの態様において、前記編集方法決定ステップにて決定される編集方法は、前記実行ステップにて行われる編集処理として再翻訳処理の際に使用する辞書を規定したもの、 50

あるいは前記辞書の使用の優先度を規定したものである。

【0009】

本発明は、他の観点において、アノテーションの種類と翻訳方法とを対応付けてテーブルに記憶する記憶手段と、文書を入力する入力手段と、前記入力手段にて入力された文書から文字情報とアノテーションとを抽出する抽出手段と、前記抽出手段にて抽出されたアノテーションの種類と当該アノテーションの付加対象である文書要素とを特定するアノテーション特定手段と、前記テーブルを参照し前記種類に対応する翻訳方法を決定する翻訳方法決定手段と、前記アノテーション特定手段にて特定された文書要素に対し、前記翻訳方法決定手段にて決定された翻訳方法を適用して翻訳処理を行う翻訳実行手段と、を有する文書翻訳装置を提供する。

10

【0010】

好みしい態様において、本発明の文書翻訳装置は、アノテーションの種類と編集方法とを対応付けてテーブルに記憶する記憶手段と、文書を入力する文書入力手段と、前記文書入力手段にて入力された文書を翻訳する翻訳実行手段と、前記翻訳実行手段にて翻訳された文章を提示して、アノテーションを付加する指示を受け付ける指示入力手段と、前記指示入力手段にて入力されたアノテーションの種類と当該アノテーションの付加対象である文字情報を特定するアノテーション特定手段と、前記テーブルを参照し前記アノテーション特定手段にて特定された種類に対応する編集方法を決定する編集方法決定手段と、前記アノテーション特定手段にて特定された文字情報に対し、前記編集方法決定手段にて決定された編集方法を適用して編集処理を行う編集実行手段と、を有する。

20

【0011】

本発明は、更に他の観点において、コンピュータに上記翻訳処理を実行させるコンピュータ読み取り可能なプログラムを提供する。

【発明を実施するための最良の形態】

【0012】

<実施例>

以下、図面を参照して本発明の好適な実施例を説明する。図1は、本発明の一実施例に係る文書翻訳装置1の機能構成を表した図である。同図に示すように、文書翻訳装置1は、制御部10、記憶部11、入力部12、操作部13、表示部14、および出力部15からなる。制御部10は、CPU等の制御用プロセッサを備え、文書翻訳装置1の各部を制御する。また、制御部10は、文書構造解析部101、アノテーション認識部102と、文字情報認識部103と、翻訳処理部104とを有する。文書構造解析部101は、入力部12にて取り込まれた画像データとしての文書に対し所定のアルゴリズムを用いてレイアウト解析等を行い、文書のレイアウト構造を決定する。具体的には、文書に文字と文字以外のもの（挿絵、罫線や注記等の付加情報（以下、アノテーションという）など）が含まれているか否かを判定し、文字以外のものがあった場合は、文字部分の領域とその他の部分の領域とを分離する。

30

【0013】

アノテーション認識部102は、分離抽出された文字以外の領域の画像データに対し、所定の解析処理を行って、アノテーションの種類およびアノテーションが付加されている対象となっている箇所（語や句といった文書要素）を決定する。抽出されるアノテーション種類とは、例えば、付箋、囲み線、下線、マーカ処理（ハイライト処理）、引き出し線、注記・注釈（原文の行間等に挿入された文字）などである。アノテーションの種類とアノテーションの付加対象の箇所に関する情報は、記憶部11に記憶される。文字情報認識部103は、文書構造解析部101にて分離抽出された領域に対して所定の文字認識処理を行って文字情報（字句）を抽出し、記憶部11に記憶する。翻訳処理部104は、文字情報認識部103にて抽出された文字情報に対し、記憶部11に格納される辞書データや所定のアルゴリズムを用いて置換処理を行うことにより、その文書の言語をユーザから指定された他の言語へ翻訳する翻訳処理を行う。翻訳後の文章データおよび原文中の語句と翻訳文中の語句との関係は、記憶部11に記憶される。

40

50

【0014】

文書構造解析部101、アノテーション認識部102、文字情報認識部103および翻訳処理部104により、アノテーションが付加された文書の画像データから、アノテーションと文字部分に対しては翻訳処理が行われ、各アノテーションに対してその種類とアノテーションの付加対象となっている原文中の語句とその翻訳後の語句に関する情報を抽出する機能が実現される。制御部10にて行われるこれらの処理の詳細については後述する。なお、制御部10にて実現されるこれら各部の機能は、各々独立したプロセッサによって実現されてもよいし、例えば一つのプロセッサが複数のソフトウェアを実行することにより実現されてもよい。

【0015】

記憶部11は、RAM、ROM、ハードディスク等の記憶装置であって、制御部10にて上述した処理を行う際に必要となる辞書データベースDBやその他の参照データが格納される。図1に示すように、データベースDBには、翻訳処理の際に使用され得る各種辞書データ111～115が格納される。加えて、アノテーションの種類と編集方法とを対応付けて格納した翻訳規則テーブルTr(詳細は後述)を格納する。さらに、注記文字と翻訳の際に使用する辞書の優先順位とを対応付けて格納した辞書テーブルTp(詳細は後述)とを格納する。

【0016】

入力部12は、スキャナ装置等であり、紙等に印刷された原稿文書をデジタル画像データとして読み込み、制御部10および記憶部11に供給する装置である。操作部13はキーボードやマウス等の入力デバイスであって、文書翻訳装置1のユーザが、翻訳対象の文書の指定、辞書テーブルTpおよび翻訳規則テーブルTrへの情報の書き込み、編集対象箇所の指定(詳細は後述)、その他の必要な情報の入力の際に使用されるものである。入力された指示や情報は制御部10へ供給される。表示部14は、描画用プロセッサ(図示せず)および液晶ディスプレイ等の表示装置(図示せず)から構成され、制御部10の指示の下、原文書やおよび翻訳処理中の文書やユーザへの各種メッセージを画面に表示する。ユーザは、表示部14の表示画面を見ながら入力部12から各種指示を入力することによって各種処理を文書翻訳装置1に実行させる。出力部15は、編集処理後の原稿を紙等に印刷するためのプリンタや付加情報編集処理を行って得られた文書データを印刷装置に供給するための通信インターフェースや、あるいは文書データをフラッシュメモリやCD-ROM等の記憶媒体に記憶するための記憶装置である。

【0017】

以下、図2～5を用いて、文書翻訳装置1の動作の一例を説明する。なお、図4に示す翻訳規則テーブルTrおよび図5に示す辞書テーブルTpには、予め必要な情報が登録されているものとする。

【0018】

図2は、特徴情報の登録処理の流れを示したものである。同図に示すように、まず、ユーザは所定指示を入力して翻訳元の言語および翻訳先の言語を指定するとともに、翻訳したい文書(以下、翻訳対象文書という)をスキャナ装置にセットし、その文書を取り込んで画像データを取得する(ステップS10)。ここでは英語の文章を日本語に翻訳する場合について説明する。図3の(a)は、翻訳対象文書の例を示したものである。図2に戻り、取得した画像データに対し文書構造を解析して文字部分の領域を特定し(ステップS11)、文字認識処理を行って文字情報を抽出する(ステップS12)。続いて、抽出した文字情報に対して翻訳処理を行って(ステップS13)、翻訳結果を表示部14に出力する(ステップS14)。なお、この翻訳処理の際に用いられる辞書データは予め決められたものが使用される。例えば、汎用的な辞書である英和辞書111が選択される。翻訳された文章の一例を図3(b)に示す。さらに、制御部10は、表示部14の表示画面に「翻訳が完了しました。編集対象箇所がある場合は指定してください。」といったメッセージを表示させ、ユーザに確認を促す。

【0019】

10

20

30

40

50

図2に戻り、ユーザは表示画面を見ながら、誤訳や不適切な翻訳処理がなされている箇所がないかをチェックする。該当箇所を見つけると、ユーザが所望する編集方法に応じたアノテーションを当該翻訳後の文章に付加する(ステップS15)。この処理を図3(c)を参照して具体的に示す。同図において、ユーザは「big-endian(訳語なし)」、「little-endian(訳語なし)」、「骨形成タンパク質」、「武勇伝勲章」、「通訳」の計5箇所において不適切な翻訳処理がなされたことを確認した場合の例が示されている。「big-endian」および「little-endian」はコンピュータ専門用語であるため、翻訳処理の際に使用された英和辞書111には訳語が存在せず、このため「訳語なし」の語が原稿に付加されてしまっている。「骨形成タンパク質」および「武勇伝勲章」、「通訳」については、それぞれ「BMP」、「CGM」および「interpreter」の訳語として選択されたものであるが、これは誤訳である。ユーザこれらの箇所を見つけると、編集対象箇所としてマウスやキーボードを用いて所定のアノテーションを付加する。

10

20

30

40

【0020】

具体的には、図4に示すように、ユーザが所望する編集方法に対応する種類のアノテーションを付加する。例えば、「big-endian」および「little-endian」については、コンピュータ専門用語であり一般には原語のまま用いられるので原文のまま残したい(すなわち、「big-endian(訳語なし)」を「big-endian」と、「little-endian(訳語なし)」を「little-endian」と編集したい)、とユーザが考えた場合、それらの語にアノテーションとして囲み線を付加する。「骨形成タンパク質」については、原文中では「BMP」に対応するものであり、原文をそのまま当てはめる(すなわち「骨形成タンパク質」を「BMP」と編集する)のが最適であると考えた場合、「骨形成タンパク質」に対して下線を引くというアノテーション処理を行う。「通訳」については、対応する原文中の語句(この場合は「interpreter」)の訳語の複数の選択肢のうち、「通訳」とは異なる他の1つの語句であって当該英和辞書111において次に高い優先順位がつけられている語(例えば「解釈」)を当てはめることを希望する場合は、翻訳後の「通訳」部分にマーカ処理を施す。また、「武勇伝勲章」については、本文書の分野の翻訳に適した辞書を選択し、当該辞書に登録されている訳語(例えば「CGM(Computer Graphic Metafile)」)を当てはめたいと考えた場合は、アノテーションとして「引き出し線と、文書の分野を指定する文字(この例では「画像処理」)を付加する。ユーザがこの対応関係を頭に入れておくてもいいように、表示画面の図3(c)に示した翻訳後の文章の周辺に表示するようにしてもよい。ユーザは図4に示した対応関係を確認しながら、所望する編集方法に対応するアノテーションの種類を容易に特定することができる。

30

40

【0021】

図2に戻り、ユーザが所望の編集対象箇所に所望のアノテーションを付加する作業が完了すると、所定の指示を入力して編集対象箇所とアノテーションとを確定させると、図3(c)に示したアノテーションが付加された状態の文書に対応する画像データが生成され、この画像データに対して編集処理(再翻訳処理)が開始する(ステップS20)。文書構造解析部101にてこの画像データに対し文章構造解析が行われ、文字情報とアノテーションとが分離抽出される(ステップS21)。続いて、アノテーション認識部102において、各アノテーションに対し、そのアノテーションが付加されている対象の訳語の箇所と当該アノテーションの種類とが判別される(ステップS22)。なお、アノテーションとして注記(図3(b)の例における「画像処理」)が付加されている場合は文字認識処理を行ってその文字を特定する。

50

【0022】

続いて、翻訳規則テーブルTrを参照し、判別されたアノテーションの種類に対応する編集方法を特定する(ステップS23)。ここで、アノテーションとして注記があった場合、辞書テーブルTpを参照し、注記に含まれる文字に対応する辞書および各辞書の使用に係る優先順位を特定する。図5に辞書テーブルTpの記憶内容の一例を示す。同図に示すように、辞書テーブルTpには、指定文字に対応付けて、使用可能な辞書とその優先順位とが登録されている。例えば、「画像処理」という注記があった場合、辞書テーブルT

50

pに登録されている指定文字「画像」を含んでいるから、辞書として英和辞書111、和英辞書112、および画像処理用語辞書113が、この順番で使用され得ることが決定される。すなわち、当該注記の対象となっている語句(図3(c)の例における「武勇伝勲章」；原文はCGM)に対しては、まず既に使用されている英和辞書111は使用候補から除外される。次に優先順位の高い「和英辞書112」については和英翻訳の際にのみ使用されるものであるから当然に使用候補から除外される。この結果、次に優先順位の高い辞書である画像処理用語辞書113を編集対象となっている語(CGM)に適用して翻訳処理を行うことが決定される。この結果、例えば画像処理用語辞書113に登録されている「CGM」の訳語として、「CGM(Computer Graphic Metafile)」が選択される。

【0023】

10

図2に戻り、編集方法が決定されると、当該編集方法に従って編集処理(再翻訳処理)を行う(ステップS24)。図3(d)には、前述した計5つの編集対象箇所が、対応する編集方法に従ってそれぞれ編集された文書を示したものである。続いて、制御部10は、表示部14の表示画面に「編集(再翻訳)処理が完了しました。編集対象箇所を追加したい場合は指定し直してください」といったメッセージを表示させ、ユーザに編集結果の確認を促す。ユーザは思ったように編集されなかつと判断した場合、または他の箇所に誤訳があることを新たに発見した場合等は、所定の指示を入力する。すると、処理は図2のステップS15へ戻り、再度編集対象箇所の指定を受け付ける。ユーザが編集内容に満足した場合は所定の指示を入力し、翻訳処理を確定させる。確定した翻訳文は所定の方法で出力される(ステップS25)。

【0024】

20

このように文書翻訳装置1によれば、一度翻訳処理がされた文書をユーザが確認し、編集が必要な箇所と編集方法とをアノテーションにより指定することにより当該箇所が適切に修正されるから、短時間でかつユーザに過度の負担を掛けることなく、高品質の翻訳文を取得することができる。

【0025】

30

<変形例>

本発明は上記実施例に限定されるものでなく、各種の変形を施すことが可能である。以下、変形例を示す。上記実施例においては、一度汎用の辞書(英和辞書111)を用いて文書翻訳装置1にて翻訳処理(仮翻訳処理)を行い、その結果をユーザがチェックして編集対象箇所を指定したが、原文にアノテーションを付加し、このアノテーションに基づいて翻訳処理を行ってもよい。すなわち、アノテーションつきの原文をスキヤナで読み込み、アノテーションの種類とアノテーションの対象箇所を特定し、翻訳規則テーブルTrおよび辞書テーブルTpを参照して翻訳の方法(原文のままにするのか否か、使用辞書、優先順位など)を決定してもよい。この場合、翻訳処理を一回省くことができ、例えば原文をユーザがチェックして誤訳が発生しそうな箇所が予想できるような場合に特に有効である。

【0026】

40

また、仮翻訳された文章にアノテーションを付加する際には、当該文書を紙等に印刷し、その紙等に手書きで行ってもよい。この場合は、アノテーション付きの文書を再度スキャンして画像データを取得することになる。

【0027】

また、上記実施例においては、全ての編集対象箇所を指定した後、編集(再翻訳)処理を行ったが、これに限らず、例えば一つアノテーションを付加するごとに当該箇所の編集処理を実行してもよい。

【0028】

文書の内容、登録するアノテーションの種類、注記の指定文字や使用辞書については、上述したものに限られないことはいうまでもない。

【図面の簡単な説明】

【0029】

50

【図1】本発明の実施例に係る文書翻訳装置1の機能構成を示す図である。

【図2】文書翻訳装置1において実行される処理の流れを説明するための図である。

【図3】(a)～(d)は、それぞれ翻訳対象となる原文、仮翻訳された文章、編集処理中の文章、編集後の文書の一例を示す図である。

【図4】アノテーションの種類と編集方法との対応関係を示す図である。

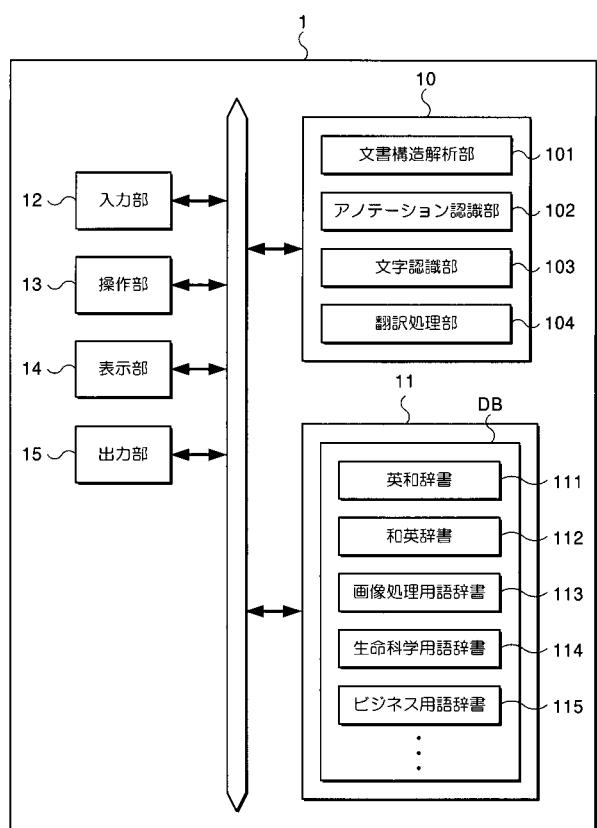
【図5】指定文字と使用辞書および優先順位の対応関係が記述されたテーブルを示す図である。

【符号の説明】

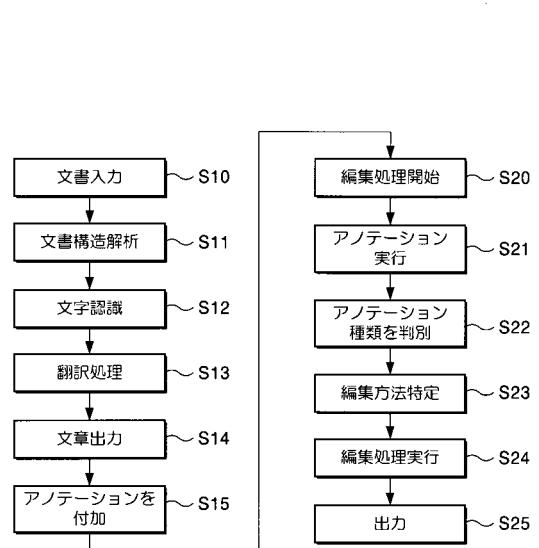
【0030】

1・・・文書翻訳装置、10・・・制御部、11・・・記憶部、12・・・入力部、13 10
・・・操作部、14・・・表示部、15・・・出力部、101・・・文書構造解析部、1
02・・・アノテーション認識部、103・・・文字情報認識部、104 10
・・・翻訳処理部。
02・・・アノテーション認識部、103・・・文字情報認識部、104 10
・・・翻訳処理部。

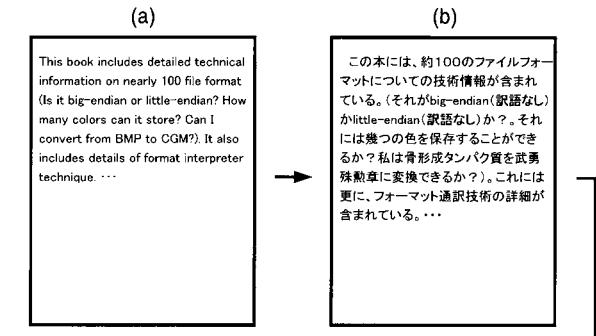
【図1】



【図2】



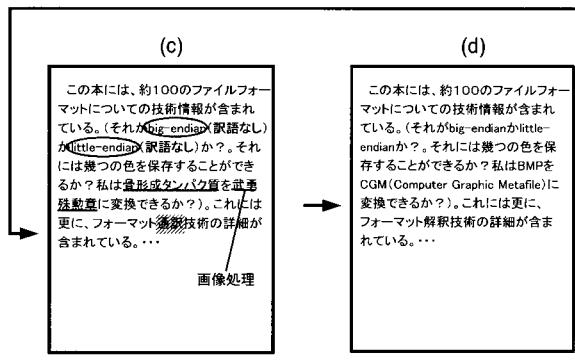
【図3】



【図4】

種類	編集方法
囲み線	原文のままにする
下線	訳文を削除して原文を使用
マークで塗る	他の意味を選択
注記	辞書の指定

【図5】



指定文字	優先順位				
	1	2	3	4	...
画像	英和辞書	和英辞書	画像処理用語辞書	—	...
ビジネス	英和辞書	和英辞書	—	—	...
バイオ	英和辞書	和英辞書	生命科学用語辞書	—	...
:	:	:	:	:	:

フロントページの続き

(72)発明者 中村 浩太郎
東京都港区赤坂二丁目17番22号 富士ゼロックス株式会社内

(72)発明者 横原 正義
神奈川県海老名市本郷2274番地 富士ゼロックス株式会社内

(72)発明者 彦 新宇
神奈川県海老名市本郷2274番地 富士ゼロックス株式会社内

(72)発明者 斎藤 照花
神奈川県海老名市本郷2274番地 富士ゼロックス株式会社内

(72)発明者 小山 俊哉
神奈川県海老名市本郷2274番地 富士ゼロックス株式会社内

F ターム(参考) 5B091 AA04 CB09 DA04 DA14 EA07