

(19) 日本国特許庁(JP)

(12) 特 許 公 報(B2)

(11) 特許番号

特許第4318223号
(P4318223)

(45) 発行日 平成21年8月19日(2009.8.19)

(24) 登録日 平成21年6月5日(2009.6.5)

(51) Int.Cl. F 1
G 0 6 F 1 7 / 2 1 (2 0 0 6 . 0 1) G 0 6 F 1 7 / 2 1 5 5 0 J

請求項の数 2 (全 11 頁)

(21) 出願番号	特願2005-357442 (P2005-357442)	(73) 特許権者	000005223 富士通株式会社
(22) 出願日	平成17年12月12日(2005.12.12)		神奈川県川崎市中原区上小田中4丁目1番 1号
(62) 分割の表示	特願平9-84951の分割	(74) 代理人	100119161 弁理士 重久 啓子
原出願日	平成9年4月3日(1997.4.3)	(74) 代理人	100111822 弁理士 渡部 章彦
(65) 公開番号	特開2006-85739 (P2006-85739A)	(74) 代理人	100083297 弁理士 山谷 皓榮
(43) 公開日	平成18年3月30日(2006.3.30)	(72) 発明者	伊吹 潤
審査請求日	平成17年12月12日(2005.12.12)		神奈川県川崎市中原区上小田中4丁目1番 1号 富士通株式会社内
審査番号	不服2006-20977 (P2006-20977/J1)		
審査請求日	平成18年9月21日(2006.9.21)		
(31) 優先権主張番号	特願平8-83449		
(32) 優先日	平成8年4月5日(1996.4.5)		
(33) 優先権主張国	日本国(JP)		

最終頁に続く

(54) 【発明の名称】 文書校正装置およびプログラム記憶媒体

(57) 【特許請求の範囲】

【請求項1】

テキスト文書を校正する文書校正装置において、

テキストの各構成文字または文字列に対して所定の誤り確率計算手法により計算した誤り確率値によって誤り候補を検出する誤り候補検出手段と、

前記誤り候補検出手段により誤り候補とされた文字列について、所定の単語辞書に登録されている単語との類似度を計算し、前記文字列が所定値以上の類似度を持つ単語列から構成される場合に一致と判断する単語比較手段と、

前記単語比較手段により誤り候補の文字列と一致するとされた単語列が、前記誤り候補とされた文字列内の平均単語長を評価値とする評価を含む誤り候補としての評価を改善させるものである場合に、その単語列を誤り候補の文字列に対する正解単語を含む単語列と認定する拡張形態素解析手段と、

誤り候補の文字列を含むテキストとともに前記拡張形態素解析手段が認定した正解単語を表示する誤り表示処理手段とを備える

ことを特徴とする文書校正装置。

【請求項2】

テキスト文書を校正する文書校正装置を計算機によって実現するためのプログラムが記憶される計算機読み取り可能なプログラム記憶媒体であって、

前記計算機を、

テキストの各構成文字または文字列に対して所定の誤り確率計算手法により計算した誤

り確率値によって誤り候補を検出する誤り候補検出手段と、

前記誤り候補検出手段により誤り候補とされた文字列について、所定の単語辞書に登録されている単語との類似度を計算し、前記文字列が所定値以上の類似度を持つ単語列から構成される場合に一致と判断する単語比較手段と、

前記単語比較手段により誤り候補の文字列と一致するとされた単語列が、前記誤り候補とされた文字列内の平均単語長を評価値とする評価を含む誤り候補としての評価を改善させるものである場合に、その単語列を誤り候補の文字列に対する正解単語を含む単語列と認定する拡張形態素解析手段と、

誤り候補の文字列を含むテキストとともに前記拡張形態素解析手段が認定した正解単語を表示する誤り表示処理手段として、

機能させるための文書校正プログラムを格納したプログラム記憶媒体。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、文書処理装置においてユーザが入力もしくは電子的な媒体として獲得した文書データに対して、誤った部分を自動的に指摘することにより、ユーザが文書を校正する作業を軽減し、文書校正に係る作業効率を大幅に向上させることができるようにした文書校正装置およびその文書校正装置を実現するためのプログラム記憶媒体に関する。

【背景技術】

【0002】

従来の文書処理装置において、誤り候補を指摘する方法として、形態素解析を行い、(1)その結果から未登録語部分を指摘するもの、(2)同音異義語のある単語を指摘するものがある。また、(3)特定の単語列が検出された場合に誤りと認定して指摘するものがある。これは、例えば現在では使用されない仮名遣いや表記等を予め登録した誤り検出用辞書を用いて、一致するものを誤りとするとか、名詞+動詞という普通にはあり得ない品詞列を検出するとか、あるいは、一文字の漢字単語があった場合に誤りとするなどの方法である。他の方法として、(4)テキスト中の片仮名または漢字文字列を辞書順に並べ、同じ単語であって僅かな表記の揺れのある単語が連続して並ぶようにして、表記の揺れを検出しやすくしたものがある。例えば、「ウィンドウ」と「ウインドウ」と「ウインドウ」などの場合に、表記を統一するために用いる方法である。

【発明の開示】

【発明が解決しようとする課題】

【0003】

上記(1)の未登録語部分を指摘する方法は、未登録語部分の周辺に誤った綴りの単語がある可能性が高いため、その部分を指摘する方法である。しかし、未登録語の生まれる原因としては、綴りの誤り以外にも、例えば固有名詞などのように、間違いではないが辞書に登録されていないというような本来の未登録語の存在も挙げられる。したがって、指摘された部分が誤りと断定できるわけではなく、ユーザが一つ一つ正誤を判断しなければならない。

【0004】

上記(2)の同音異義語の存在する単語箇所を指摘する方法は、仮名漢字変換のときに操作を誤りやすい箇所であるために確認を目的として指摘する方法である。この場合、一つでも同音異義語が存在すれば、本来は誤りでない単語であっても指摘されるため、ユーザが自ら正誤の判断を下す必要がある部分が非常に多くなってしまいう問題がある。

【0005】

一方、上記(3)の特定の単語列が検出された場合に誤りとして指摘する方法では、予め誤りとして検出しておくべき品詞列などを登録する必要があり、誤りの対象が非常に限定されたものとなるため、実際には、文章中の誤りの多くは検出することができないという問題がある。

【0006】

10

20

30

40

50

また、上記(4)の片仮名語句や漢字語句をソートしてユーザに示す方法では、校正時にユーザ自身ですべき作業量が多い割には作業能率があまり改善されないという問題がある。

【0007】

以上のように、従来の方法を用いると、誤り候補の指摘箇所が本来の誤りではないことが多くあり、過剰指摘の傾向になるという問題点があった。結果としてユーザが再チェックしなければならない量が多くなるため、校正作業の効率を低下させることがあった。また、誤り箇所の適切な選択が不十分なことが多いという問題点があった。

【0008】

また、例えば原テキストのそれぞれの単語が未登録語ではないため、誤りを指摘できないような場合などに、誤り候補の選出ができないという問題点があった。

【課題を解決するための手段】

【0009】

上記問題点を解決するため、本発明は、例えば図1ないし図5に示す各手段を備える。

【0010】

図1は、誤り候補検出部のブロック構成例を示す。図1において、1は誤り候補検出部、2a、2b、...は誤り確率計算部、3は誤り確率保存部、4は誤り確率閾値、5は誤り候補抽出部、6は誤り表示処理部を表す。

【0011】

誤り確率計算部2a、2b、...は、それぞれ異なる方法によりテキストの各構成文字または文字列に対して誤り確率値を計算し付与する手段である。各誤り確率計算部2a、2b、...は、後述する図2～図4に示すような手段を単独でまたは組み合わせて用いてもよく、また従来技術として説明したような誤り候補を検出する手段を組み合わせて用いることにより、それぞれ所定の誤り確率を付与するようにしてもよい。

【0012】

誤り確率保存部3は、誤り確率計算部2a、2b、...から得た誤り確率値を集計し、各構成文字等に対する最終的な誤り確率値を定める手段である。

【0013】

誤り候補抽出部5は、所定の誤り確率閾値4をもとに、これより高い誤り確率値を持つ文字または文字列を誤り候補として抽出する手段である。誤り確率閾値4は、外部入力としてまたは予め記憶された電子的データとして与えられる。

【0014】

誤り表示処理部6は、誤り候補抽出部5により抽出された誤り箇所を表示する手段である。

【0015】

以上のような構成によって、誤り確率保存部3は、各誤り確率計算部2a、2b、...から得た複数の誤り確率値を集計し、各文字または文字列ごとに、それぞれの誤り確率値を定めるため、誤っている確率の大きい候補を従来より正確に抽出することができる。

【0016】

また、誤り表示処理部6は、誤り候補抽出部5によって所定の誤り確率閾値4を超えたものだけを抽出して表示するため、過剰な誤り指摘を防止することができる。

【0017】

さらに、誤り確率閾値4を変化させることにより、誤り候補として抽出し表示する単語等の量を調節できるため、校正作業の目的等により、また原テキストの種類や内容により、誤り指摘の精度を保持するか、または作業効率を向上させるかを選択することができる。

【0018】

図2は、誤り確率計算部の構成例を示す。

【0019】

誤り確率計算部20は、形態素解析部21、未登録語検出部22、固有名詞パターン検

10

20

30

40

50

出部 2 3 を備える。

【 0 0 2 0 】

形態素解析部 2 1 は、テキストを形態素解析し単語列に分割し、単語群を未登録語検出部 2 2 および固有名詞パターン検出部 2 3 へ渡す手段である。形態素解析についての具体的方法については、従来からよく知られているので、ここでの詳細な説明は省略する。

【 0 0 2 1 】

未登録語検出部 2 2 は、単語群から未登録語とされた単語領域を検出してその単語領域に所定の誤り確率値を付与し、図 1 の誤り確率保存部 3 へ渡す。

【 0 0 2 2 】

また、固有名詞パターン検出部 2 3 は、形態素解析結果の単語列を受けて、統語的な特徴を利用することにより固有名詞が出現しやすい領域を認識し、その固有名詞が出現しやすい領域に対して負の誤り確率値を付与し、図 1 の誤り確率保存部 3 へ渡す。統語的な特徴とは、前後の既知の単語の意味や品詞などから類推できる特徴であり、「社長」や「部長」というような肩書の前は人名であることが多いとか、「市」や「州」という字の前にくる漢字列または片仮名は地名であることが多いという単語が並ぶパターンの特徴である。固有名詞に対して負の誤り確率値を与えるのは、固有名詞は未登録語であっても誤りではないことが多いからである。したがって、未登録語でない固有名詞には、負の誤り確率値を付与する必要はない。

【 0 0 2 3 】

このような固有名詞の検出では、例えば「未登録語 + 肩書」のように、予め固有名詞と特定の品詞または単語の組み合わせを解析用データとして登録しておき、固有名詞パターン検出部 2 3 で、形態素解析の結果をこれらの固有名詞パターンと照合すればよい。

【 0 0 2 4 】

これにより、従来の方法では必ず誤り候補として指摘されるような固有名詞に関する単語領域を、誤り指摘箇所から除くことが可能となる。

【 0 0 2 5 】

図 3 は、誤り確率計算部の構成例を示す。

【 0 0 2 6 】

誤り確率計算部 3 0 は、語彙抽出部 3 1、カテゴリ化部 3 2、確率付与部 3 3 を備える。語彙抽出部 3 1 は、対象テキストから単語を抽出し、カテゴリ化部 3 2 へ送る手段である。カテゴリ化部 3 2 は、語彙抽出部 3 1 から得た単語群について、単語間の文字構成の類似度を計算し、計算した類似度に基づきカテゴリ化を行う手段である。確率付与部 3 3 は、カテゴリ化部 3 2 によりカテゴリ化された単語相互の類似度および単語の出現頻度等に基づいて誤り確率値を付与する手段である。

【 0 0 2 7 】

図 3 に示す誤り確率計算部 3 0 によれば、例えば表記の揺れを生じている単語のように、単語相互の類似度の高い単語を同一カテゴリに分類し、その単語の類似度および出現頻度により、どの表記の単語を正当または誤りとするかを誤り確率値として付与することができるため、最終的に表示する誤り候補に表記の揺れによる誤りの傾向を反映することができる。

【 0 0 2 8 】

図 4 は、誤り確率計算部の構成例を示す。

【 0 0 2 9 】

誤り確率計算部 4 0 は、形態素解析部 4 1、領域分割処理部 4 2、確率付与部 4 3 を備える。形態素解析部 4 1 は、対象テキストを形態素解析し単語列に分割する手段である。領域分割処理部 4 2 は、対象テキストを、例えば漢字、平仮名、片仮名、英字、... というような同種の文字で構成される領域に分割する手段である。

【 0 0 3 0 】

確率付与部 4 3 は、記憶している各文字種ごとの平均単語長を用いて、同一文字種からなる単語領域ごとに算出したその領域内での予測単語数と、形態素解析による単語群から

10

20

30

40

50

求めた実際の単語数とを比べ、その差に基づいた誤り確率値を各文字または文字列ごとに付与する手段である。すなわち、確率付与部43は、予測単語数と実際の単語数との差が大きい領域には、綴り誤り等により1単語として認識されなかった単語が含まれている確率が高いとみなし、この領域に相対的に高い誤り確率値を付与する。

【0031】

上述したような誤り候補検出部1によって、ユーザに対し誤りであるかどうかの手掛かりを一つ一つ直接提示するのではなく、複数の手掛かりをもとに総合的に判断した結果により誤りの可能性が高い部分のみを提示することが可能となる。また、それぞれの手掛かりの確からしさを誤り確率値によって定量化するので、ユーザは、個々の手掛かりごとの調整に煩わされることなく、誤り確率閾値の調整のみによって、誤り候補を広く抽出して誤り指摘箇所の洩れをなくすか、明白な誤り箇所のみを抽出して校正作業の能率を向上させるかについてのバランスの調整を任意にできるようになる。

10

【0032】

以上の図1～図4に示す各手段は、計算機を動作させるプログラムを用いて実現される。そのプログラムは適当な記憶媒体に格納することができる。

【0033】

図5は、本発明のブロック構成例を示す図である。図5において、51は誤り候補検出部、52は拡張形態素解析部、53は単語比較部、54は単語辞書、55は誤り表示処理部を表す。

【0034】

誤り候補検出部51は、例えば図1に示す誤り候補検出部1と同様な手段によって、入力テキストから誤り確率の大きい文字列を検出する手段である。誤り確率の計算では、例えば図4などに示すような手段を用いる。単語比較部53は、誤り候補検出部51から得た誤り候補とされた文字列と、単語辞書54に登録されている単語との類似度を比較して、前記文字列が所定値以上の類似度を持つ単語列から構成される場合に一致と判断する手段である。拡張形態素解析部52は、誤り候補検出部51から得た誤り候補とされた文字列に対して拡張形態素解析を行い、拡張形態素解析の結果、その単語領域の評価値が誤り候補検出部51から得た評価値よりよい場合、すなわち全体としての誤り確率が小さくなるような場合には、拡張形態素解析により得た単語(単語群)を正解単語(正解単語群)と認定する。

20

30

拡張形態素解析部52における誤り候補としての評価は、単語比較部53により誤り候補の文字列と一致するとされた単語列が、前記誤り候補とされた文字列内の平均単語長を評価値とする評価を含む。

【0035】

ここで拡張形態素解析とは、厳密に単語辞書54に登録されている単語と一致する場合の情報だけを用いて形態素解析を行うのではなく、厳密には一致しないがかなり似ている単語の情報も用いて形態素解析を行うことを意味する。

【0036】

誤り表示処理部55は、元の対象テキストとともに拡張形態素解析部52から得た正解単語群を表示する。

40

【0037】

以上の図5に示す各手段は、計算機を動作させるプログラムを用いて実現される。そのプログラムは適当な記憶媒体に格納することができる。

【発明の効果】

【0038】

本発明によれば、例えば原テキストのそれぞれの単語が未登録語ではないため誤りを指摘できないような場合や、付与される誤り確率値が所定の誤り確率閾値よりも小さいため実際の誤り候補として抽出されないような場合にも、誤り候補とその正解単語の選出が可能になる。

【発明を実施するための最良の形態】

50

【 0 0 3 9 】

以下、本発明の実施の形態を図を用いて説明する。

【 0 0 4 0 】

図6は、図2に示す誤り確率計算部20による誤り確率計算の例を説明する図である。

【 0 0 4 1 】

校正対象の原テキストが「ジュペ外相はその意見にうなづいた」であるとき、図2に示す形態素解析部21は、形態素解析を行い、その結果として「/ジュペ/外相/は/その/意見/に/うなづ/いた/」の単語列を、未登録語検出部22へ渡すとともに、固有名詞パターン検出部23へ品詞等の属性情報を含めて渡す。

【 0 0 4 2 】

未登録語検出部22では、「ジュペ」、「うなづ」が未登録語として検出される。検出された未登録語の文字列に対して、それぞれ誤り確率〔1〕として、0.5を付与する。この0.5という値は、予め統計的なデータによって決められた値である。

【 0 0 4 3 】

また、固有名詞パターン検出部23では、予め登録されている固有名詞パターンの一つである「未登録語+肩書」に一致するものとして、「/ジュペ/外相/」が検出される。ここで、「肩書」の前は固有名詞であることが一般的であり、通常、固有名詞は未登録語であるから、「/ジュペ/外相/」の単語列は、誤りである確率は低くなる。しがたって、「/ジュペ/外相/」の「ジュペ」に、負の誤り確率〔2〕として、-0.4を付与する。

【 0 0 4 4 】

誤り確率計算部20からの誤り確率値により、誤り確率保存部3は、文字列「ジュペ」の誤り確率値を0.1(=0.5-0.4)とし、文字列「うなづ」の誤り確率値を0.5とする。

【 0 0 4 5 】

以上の結果から、図1に示す誤り候補抽出部5と同様な誤り候補の抽出を行う。ここで、誤り確率閾値4が例えば0.2に設定されているとすると、「ジュペ」の部分は、『0.1 誤り確率閾値(0.2)』であるため誤り候補から外され、「うなづ」の部分が、『0.5 > 誤り確率閾値(0.2)』であるため誤り候補として抽出される。抽出された誤り候補は、誤り表示処理部55へ送られる。誤り表示処理部55は、誤り指摘箇所として「うなづ」の部分をも、下線付き表示または色違い表示等により他の部分と区別して、校正用のテキストとして表示装置等へ出力する。

【 0 0 4 6 】

このように、本来は誤りではない固有名詞である未登録語の誤り確率値を低く抑えることで、誤り候補として「ジュペ」が表示されることを防止している。なお、図5の拡張形態素解析部52の結果を、図1に示す誤り確率保存部3に渡すような構成にしてもよい。

【 0 0 4 7 】

図7は、図3に示す誤り確率計算部30による誤り確率計算の例を説明する図である。

【 0 0 4 8 】

校正対象の原テキスト中に、同一の意味を示す単語が「インパーテッド」、「インヴァーテッド」、「インパーテット」の3種の異なる表記で表されているとする。語彙抽出部31は、原テキストの単語ごとの出現頻度を計算し、単語と出現頻度の情報をカテゴリ化部32へ渡す。

【 0 0 4 9 】

カテゴリ化部32は、単語群の中で文字構成の類似度の高いもの同士を同一のカテゴリに分類する。このカテゴリ化の処理では、例えば「バ」と「ヴァ」は類似度が非常に大きい、「ズ」と「ヅ」も類似度が非常に大きい、長音記号「ー」がある場合とない場合の類似度は大きい、濁点付きと濁点なしの同字は類似度がやや大きい、というような情報に基づいて、特定の文字の関係に対して予め定められた類似度の値を用いて、単語の類似度が算出される。これにより、「インパーテッド」、「インヴァーテッド」、「インパーテッ

10

20

30

40

50

ト」の3つの単語は同一カテゴリに分類される。

【0050】

次に、同一カテゴリ内の単語間で、最高頻度の単語を基準に、文字構成の類似度により誤り確率を計算する。この例では、最高頻度の「インバーテッド」を正解単語と認定し、「バ」と「ヴァ」の表記の類似度が高いことから「インバーテッド」に対する「インヴァーテッド」の類似度が10、また、語尾の「ド」と「ト」の表記の類似度はそれほど高くはなく、「インバーテッド」に対する「インバーテット」の類似度が2となっている。続いて、この類似度に基づき誤り確率値を定める。「インバーテッド」は、正解単語と認定しているので誤り確率を0とし、正解単語に対する類似度の高い「インヴァーテッド」は誤り確率が0.2、類似度の低い「インバーテット」は、誤り確率が0.8となっている。

10

【0051】

ここで、図1に示す誤り確率閾値4を0.7と高く設定すると、誤り候補抽出部5は、文字列「インバーテット」のみを誤り候補として抽出する。一方、誤り確率閾値4を0.1と低くすれば、誤り候補抽出部5は、文字列「インヴァーテッド」、「インバーテット」の両方を抽出し、元の単語と別表記と認められる多くの単語が、誤り候補として校正テキストに表示される。なお、この誤り確率値は一例であり、テキストの種類によっては、同一カテゴリ内で類似度の低い単語の誤り確率を小さくしたほうがよい場合もある。

【0052】

図8は、図4に示す誤り確率計算部40による誤り確率計算と図5に示す拡張形態素解析部52等の処理の例を説明する図である。

20

【0053】

原テキストが「キーワードファイルの指定」であるとき、図4に示す形態素解析部41は、形態素解析の結果として「/キイ/ワード/ファイル/の/指定/」を出力する。ここでは、単語辞書中に「キイ」、「ワード」、「ファイル」という単語があったため、未登録語は検出されていない。

【0054】

また、領域分割処理部42は、同一文字種ごとに領域分割処理を行い、片仮名部分の「キーワードファイル」、平仮名部分の「の」、漢字部分の「指定」の3つの部分に分割する。

30

【0055】

確率付与部43は、「/キイ/ワード/ファイル/」について、形態素解析で求めた実際の単語数と、領域分割処理による予測単語数とを比較する。予測単語数は、文字種によって統計的に定められた平均単語長から計算される。例えば片仮名文字列の平均単語長が4であるとすると、「キーワードファイル」の予測単語数は、 $2(9 \div 2)$ となる。

【0056】

この部分の形態素解析から得られた単語数は3であり、「キーワードファイル」について、実際の単語数と予測した単語数に不一致が生じることになる。確率付与部43は、この不一致の生じた単語領域「キーワードファイル」に誤りが存在する可能性があると判断して、各文字に不一致の割合に応じた誤り確率(例えば0.1)を付与する。

40

【0057】

図1の例では、誤り確率計算部40からの誤り確率値と、他の誤り確率計算部からの誤り確率値とを、図1に示す誤り確率保存部3で集計し、誤り候補抽出部5で誤り候補を抽出し、誤り表示処理部6で誤り候補を指摘した校正テキストを表示する。

【0058】

図5の例では、さらに次のように拡張形態素解析の利用により正解単語列を求める処理を行う。

【0059】

拡張形態素解析部52は、誤り候補検出部51が検出した、ある一定値以上の誤り確率値を持つ文字列「キーワードファイル」を拡張形態素解析の対象とする。単語比較部53

50

は、「キーワードファイル」と類似度が大きい単語列として、単語辞書54に登録された単語から「キーワード」、「ファイル」を検出し、拡張形態素解析部52では、拡張形態素解析の結果として、「/キーワード/ファイル/」が抽出される。

【0060】

拡張形態素解析の結果として得られた「/キーワード/ファイル/」を正解単語列とみなしてよいかどうかを評価するために、ある評価値を計算する。ここでは、例えば誤り候補とされた領域内の平均単語長を評価値とする。形態素解析の結果による「/キイ/ワード/ファイル/」の平均単語長は3である。一方、拡張形態素解析の結果による「/キーワード/ファイル/」の平均単語長は4.5である。

【0061】

これらの結果を比較すると、拡張形態素解析の結果の単語列「/キーワード/ファイル/（平均単語長=4.5）」の方が、通常の形態素解析の結果の単語列「/キイ/ワード/ファイル/（平均単語長=3）」よりも、片仮名文字列の平均単語長=4に近いため、評価値（平均単語長）が改善されることが分かる。このため、原テキストの文字列と相違する部分を持つ「キーワード」を、元のテキストの「キーワード」の正解単語と認定して、誤り表示処理部55により、校正テキスト中に誤り候補と正解単語部分とを対比させて表示する。

【0062】

この拡張形態素解析を用いた方式によれば、例えば原テキストの「キーワードファイル」の部分について、「キイ」、「ワード」、「ファイル」のそれぞれの単語が未登録語ではないため、誤りを指摘できないような場合や、「キーワードファイル」の部分に付与される誤り確率値が所定の誤り確率閾値よりも小さいため、実際の誤り候補として抽出されないような場合にも、誤り候補とその正解単語の選出が可能になるという効果がある。

【図面の簡単な説明】

【0063】

【図1】誤り候補検出部のブロック構成例を示す図である。

【図2】誤り確率計算部の構成例を示す図である。

【図3】誤り確率計算部の構成例を示す図である。

【図4】誤り確率計算部の構成例を示す図である。

【図5】本発明のブロック構成例を示す図である。

【図6】誤り確率計算の例を説明する図である。

【図7】誤り確率計算の例を説明する図である。

【図8】誤り確率計算と拡張形態素解析の例を説明する図である。

【符号の説明】

【0064】

- 1 誤り候補検出部
- 2 a, 2 b, ... 誤り確率計算部
- 3 誤り確率保存部
- 4 誤り確率閾値
- 5 誤り候補抽出部
- 6 誤り表示処理部
- 20 誤り確率計算部
- 21 形態素解析部
- 22 未登録語検出部
- 23 固有名詞パターン検出部
- 30 誤り確率計算部
- 31 語彙抽出部
- 32 カテゴリ化部
- 33 確率付与部
- 40 誤り確率計算部

10

20

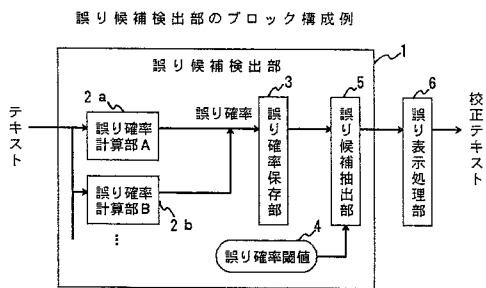
30

40

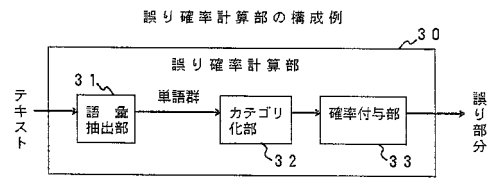
50

- 4 1 形態素解析部
- 4 2 領域分割処理部
- 4 3 確率付与部
- 5 1 誤り候補検出部
- 5 2 拡張形態素解析部
- 5 3 単語比較部
- 5 4 単語辞書
- 5 5 誤り表示処理部

【図1】

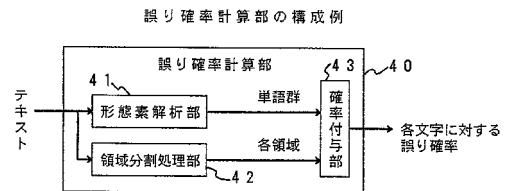
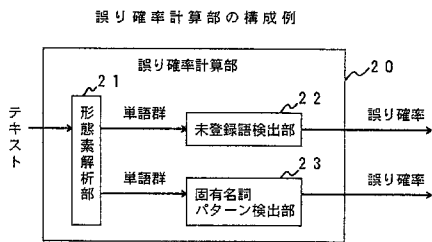


【図3】

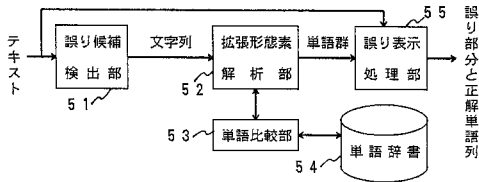


【図4】

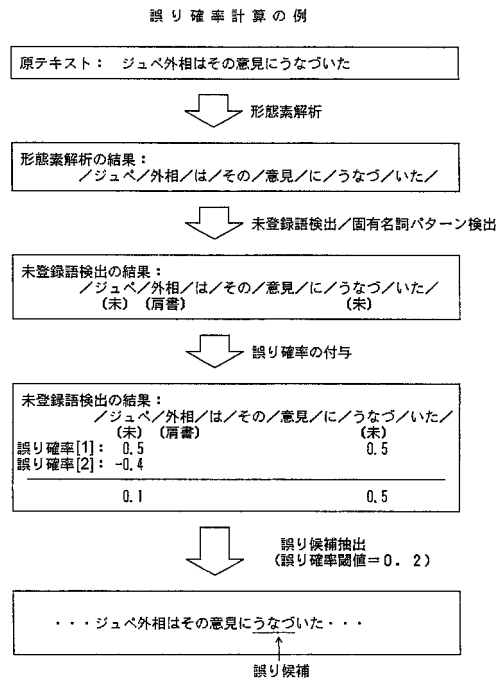
【図2】



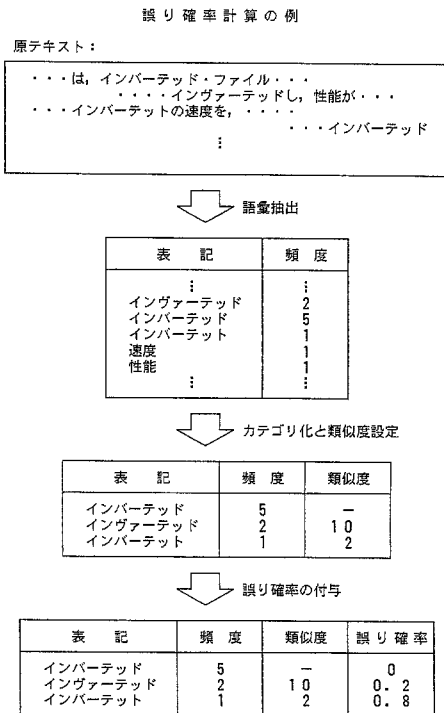
【図5】



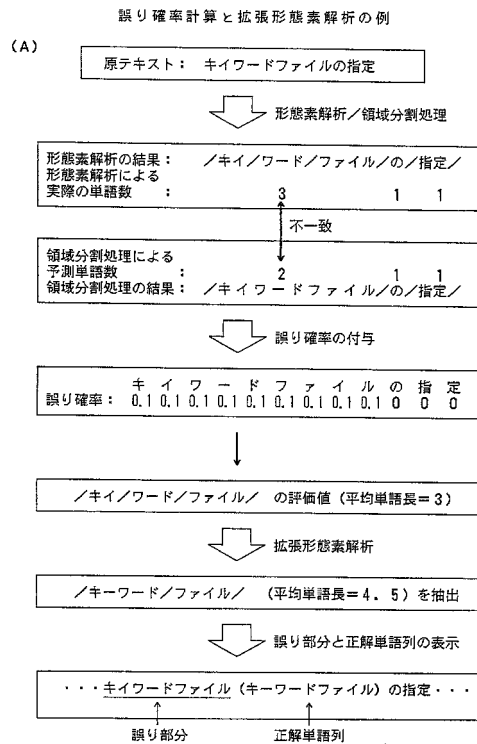
【図6】



【図7】



【図8】



フロントページの続き

(72)発明者 足立 顕
神奈川県川崎市中原区上小田中4丁目1番1号 富士通株式会社内

合議体

審判長 小曳 満昭

審判官 飯田 清司

審判官 長島 孝志

(56)参考文献 特開平4 - 213155 (JP, A)

特開平3 - 278168 (JP, A)

特開平2 - 39357 (JP, A)

(58)調査した分野(Int.Cl., DB名)

G06F17/21