(19)日本国特許庁(JP)　(12)特　許　公　報(B2)　(11)特許番号

特許第4285272号
(P4285272)

(45)発行日　平成21年6月24日(2009.6.24)　(24)登録日　平成21年4月3日(2009.4.3)

(51)Int.Cl.　　F I
G06F 17/22 (2006.01)　　G06F 17/22　522L
G06F 17/27 (2006.01)　　G06F 17/27　E

請求項の数 3　(全 17 頁)

(21)出願番号　特願2004-53456(P2004-53456)
(22)出願日　平成16年2月27日(2004.2.27)
(65)公開番号　特開2005-242809(P2005-242809A)
(43)公開日　平成17年9月8日(2005.9.8)
審査請求日　平成18年10月31日(2006.10.31)

前置審査

(73)特許権者　000002945
オムロン株式会社
京都市下京区塩小路通堀川東入南不動堂町８０１番地
(74)代理人　100078916
弁理士　鈴木　由充
(74)代理人　100142114
弁理士　小石川　由紀乃
(72)発明者　伴　哲也
京都府京都市下京区西洞院木津屋橋通東入ル　オムロンソフトウェア株式会社内
(72)発明者　山田　堂弘
京都府京都市下京区西洞院木津屋橋通東入ル　オムロンソフトウェア株式会社内

最終頁に続く

(54)【発明の名称】形態素解析方法、この方法を用いたプログラムおよび情報処理装置

(57)【特許請求の範囲】
【請求項１】
所定数の変換後文字列が登録された変換用標準辞書および新規の変換後文字列を登録するための学習用辞書、ならびに所定数の形態素が登録された形態素解析用辞書が格納されたメモリを具備し、操作部よりかな文字列の入力を受け付けた後に辞書検索に基づく変換処理を指示する操作を受け付けたことに応じて、入力されたかな文字列により変換用標準辞書および学習用辞書を検索して、変換後文字列の候補を抽出する機能と、前記操作部よりかな文字列の入力を受け付けた後に、カタカナ、ひらがな、アルファベット、数字のいずれか一の文字種を選択して変換を指示する後変換操作を受け付けたことに応じて、前記入力されたかな文字列を選択された文字種による文字列に変換する機能と、前記後変換操作に応じて変換され、前記変換用標準辞書および学習用辞書に登録されていない変換後文字列を前記学習用辞書に登録する機能とを具備する文字変換手段として動作するコンピュータにおいて、漢字およびかな混じりの文字列を含むテキストデータの入力を受け付けてそのテキストデータを形態素に分解する形態素解析処理を実行する方法であって、
前記形態素解析処理の対象となるテキストデータの入力を受け付けたとき、当該テキストデータから所定長さの文字列を抽出するステップＡと、ステップＡで抽出された文字列により前記形態素解析用辞書を検索して形態素の候補を抽出するステップＢと、前記ステップＡで抽出された文字列を前記文字変換手段に渡して当該文字列による学習用辞書の検索を実行させ、その検索により抽出された文字列を形態素の候補として前記文字変換手段より受け付けるステップＣとを複数サイクル実行した後に、各サイクルのステップＢおよ

びステップＣにより得た候補の中の所定数を選択して出力するステップＤを実行し、
前記ステップＣでは、前記ステップＡで抽出された文字列がひらがな文字列である場合には、この文字列を前記文字変換手段に渡さずに、形態素の候補を得られなかったという処理結果を設定し、
前記ステップＤでは、前記ステップＣにより得たカタカナ文字列、アルファベット文字列、および数字による文字列を、ステップＢにより得た候補より優先して選択する、
ことを特徴とする形態素解析方法。
【請求項２】
所定数の変換後文字列が登録された変換用標準辞書および新規の変換後文字列を登録するための学習用辞書が格納されたメモリを具備し、操作部よりかな文字列の入力を受け付けた後に辞書検索に基づく変換処理を指示する操作を受け付けたことに応じて、入力されたかな文字列により変換用標準辞書および学習用辞書を検索して、変換後文字列の候補を抽出する機能と、前記操作部よりかな文字列の入力を受け付けた後に、カタカナ、ひらがな、アルファベット、数字のいずれか一の文字種を選択して変換を指示する後変換操作を受け付けたことに応じて、前記入力されたかな文字列を選択された文字種による文字列に変換する機能と、前記後変換操作に応じて変換され、前記変換用標準辞書および学習用辞書に登録されていない変換後文字列を前記学習用辞書に登録する機能とを具備する文字変換手段として動作するコンピュータに、形態素解析処理の機能を付与するためのプログラムであって、
漢字およびかな混じりの文字列を含むテキストデータの入力を受け付ける第１のステップ；および入力されたテキストデータを形態素に分解して各形態素の組み合わせを出力する第２のステップを前記コンピュータに実行させるためのプログラムと、所定数の形態素が登録された形態素解析用辞書を構成する電子データとを含み、
前記第２のステップでは、前記第１のステップで受け付けたテキストデータから所定長さの文字列を抽出するステップＡと、ステップＡで抽出された文字列により前記形態素解析用辞書を検索して形態素の候補を抽出するステップＢと、前記ステップＡで抽出された文字列を前記文字変換手段に渡して当該文字列による学習用辞書の検索を実行させ、その検索により抽出された文字列を形態素の候補として前記文字変換手段より受け付けるステップＣとを複数サイクル実行した後に、各サイクルのステップＢおよびステップＣにより得た候補の中の所定数を選択して出力するステップＤを実行し、
前記ステップＣでは、前記ステップＡで抽出された文字列がひらがな文字列である場合には、この文字列を前記文字変換手段に渡さずに、形態素の候補を得られなかったという処理結果を設定し、
前記ステップＤでは、前記ステップＣにより得たカタカナ文字列、アルファベット文字列、および数字による文字列を、ステップＢにより得た候補より優先して選択する、
ことを特徴とする形態素解析用のプログラム。
【請求項３】
操作部と、この操作部よりかな文字列の入力および変換指示操作を受け付けて、入力されたかな文字列を他の形態の文字列に変換する文字変換手段と、漢字およびかな混じりの文字列を含むテキストデータの入力を受け付けて、そのテキストデータを形態素に分解する形態素解析手段とを具備する装置であって、
前記文字変換手段は、
所定数の変換後文字列が登録された変換用標準辞書と、
新規の変換後文字列を登録するための学習用辞書と、
前記操作部よりかな文字列の入力を受け付けた後に辞書検索に基づく変換処理を指示する操作を受け付けたことに応じて、入力されたかな文字列により変換用標準辞書および学習用辞書を検索して、変換後文字列の候補を抽出する候補検索手段と、
前記操作部よりかな文字列の入力を受け付けた後に、カタカナ、ひらがな、アルファベット、数字のうちのいずれか一の文字種を選択して変換を指示する後変換操作を受け付けたことに応じて、前記入力されたかな文字列を指示された文字種による文字列に変換する

後変換操作時処理手段と、

前記後変換操作時処理手段により変換され、前記変換用標準辞書および学習用辞書に登録されていない変換後文字列を前記学習用辞書に登録するとともに、前記形態素解析手段から検索対象の文字列を受け付けたとき、この文字列により学習用辞書を検索して、その検索により抽出した変換後文字列を形態素解析手段に渡す学習用辞書処理手段とを具備し、

前記形態素解析手段は、

所定数の形態素が登録された形態素解析用辞書と、

処理対象のテキストデータの入力を受け付けたとき、当該テキストデータから所定長さの文字列を抽出するステップＡと、ステップＡで抽出された文字列により前記形態素解析用辞書を検索して形態素の候補を抽出するステップＢと、前記ステップＡで抽出された文字列を前記文字変換手段の学習用辞書処理手段に渡し、当該文字列による学習用辞書の検索を実行した学習用辞書処理手段から渡された文字列を形態素の候補として受け付けるステップＣとを複数サイクル実行する検索手段と、

前記検索手段による各サイクルのステップＢおよびステップＣにより得た候補の中の所定数を選択して出力する出力手段とを、具備し、

前記検索手段は、前記ステップＡで抽出された文字列がひらがな文字列である場合には、ステップＣにおいて、前記抽出された文字列を前記学習用辞書処理手段に渡さずに、形態素の候補を得られなかったという処理結果を設定し、

前記出力手段は、前記ステップＣにより得たカタカナ文字列、アルファベット文字列、および数字による文字列を、ステップＢにより得た候補より優先して選択する、情報処理装置。

【発明の詳細な説明】

【技術分野】

【０００１】

この発明は、コンピュータに所定容量のテキストデータを入力して、そのテキストデータに対する形態素解析を行う技術に関する。

【背景技術】

【０００２】

近年、インターネットの普及に伴い、ウェブページや電子メールなどに含まれるテキストデータに形態素解析を行って、抽出された単語を他の文書を作成する目的で使用することが提案されている。また、前記形態素解析の結果に基づき、テキストデータを音声データに変換して出力するようにしたソフトウェアなども開発されている。

【０００３】

一般的な形態素解析処理では、種々の形態素が登録された辞書を用いて処理対象のテキストデータに含まれる文字列を検索し、複数の候補を抽出する。さらに、形態素の組み合わせにかかる規則に基づいて最適な候補の組み合わせを決定し、その決定による形態素の組み合わせを出力する。

このような形態素解析において、解析の精度を向上するには、辞書に登録されていない未知語を抽出できるようにする必要がある。この点につき、下記の特許文献１のような先行技術が存在する。

【０００４】

【特許文献１】特開平６－１２４５３号　公報

【０００５】

上記の特許文献１では、未登録の単語やその前後の結合関係を種々のルールと照合することにより未知語を抽出した後、抽出された未知語をモニタに表示する。そして、ユーザーに登録すべき未知語を選択させた上で、読みや品詞情報などの入力を受け付け、これらを対応づけた新たな辞書データを作成してメモリに登録するようにしている。

【発明の開示】

【発明が解決しようとする課題】

【０００６】
形態素解析では、世間の情勢や流行により生まれた新たな名称や言い回し、ユーザーが仲間うちで使用する単語など（以下、これらを「新語」と総称する。）に速やかに対応できるようにする必要がある。特に、近年は、タレントの名前、ブランド名、略語などを、カタカナ文字列やアルファベット文字列で表す頻度が高くなっているから、形態素解析でもこれらの新語に速やかに対応できるように、簡単な方法で新語を登録できるようにするのが望ましい。
【０００７】
上述した特許文献１に記載の発明では、新語を含むテキストデータを形態素解析した場合、未知語として抽出された新語について、ユーザーが登録作業を行わなければならない。このため、ユーザーの負担が大きくなるという問題がある。また、ユーザーが登録すべき新語を見落としたり、登録作業に負担を感じて処理を中止した場合には、その新語は登録されないままとなるから、次回の形態素解析で同じ新語を含むテキストデータを処理した場合、その新語は、再び未知語として取り扱われることになる。
このように、特許文献１の発明は、未知語の登録のために特別な処理時間が必要である上、ユーザーが未知語の登録作業を行うことを前提にしており、登録処理を簡単に行うのは困難である。
【０００８】
一方、システム提供者が定期的に形態素解析用辞書に新語を追加して、これをデータ配信などの方法で各ユーザーに提供するようにすれば、ユーザーの負担を軽減でき、形態素解析の機能を高めることもできる。しかし、新規登録すべき新語が多数になるのに対し、実際に使用される単語はそのうちの一部になる可能性が高く、しかも、使用される新語はユーザーによって異なるものになる可能性がある。
このように、形態素解析用辞書を定期的に更新する方法では、効率が悪く、システム提供者の採算に見合わないという問題が生じる。
【０００９】
ところで、一般ユーザーは、電子メールなどの文書を作成する際に、自分が関心を持つ新語を入力する可能性が高い。また、他者からの電子メールやウェブページなどにアクセスする場合にも、自分が関心を持つ新語が含まれるデータを閲覧する可能性が高いと考えられる。
【００１０】
かな漢字変換処理用のソフトウェアには、一般に、内部辞書に登録されていない文字列への変換が行われたときに、その変換後文字列を自動的に登録する学習機能が設定されている。また、この種のソフトウェアでは、かな漢字以外の特定の文字種（アルファベットやカタカナなど）への変換を指示する機能を所定のキーに割り当て、そのキーを操作することによって、入力文字列を特定の文字種による文字列に変換するようにしている。この特定の文字種の文字列への変換を指示するための操作は、「後変換操作」と呼ばれている。
【００１１】
この明細書では、特定の文字種への変換を指示する操作からその操作に応じて変換された文字列を確定する操作までを含めて、「後変換操作」という。
前記したカタカナやアルファベットにより表記される新語は、この後変換操作により入力することができる。また、この後変換操作により確定された文字列は、前記した学習機能により登録することができるので、次回、同じ新語を入力する際には、変換前のかな文字列を入力して通常の変換操作を行うことによって、前回登録された文字列を呼び出すことが可能となる。このように、文字入力処理では、ユーザーが特別の登録作業を行わなくとも、通常の文字入力作業の過程で新語を辞書データに加えることが可能である。
【００１２】
この発明は、上記の点に着目し、文字変換処理で蓄積された辞書データを形態素解析処理でも使用できるようにすることにより、ユーザーが特別な登録作業を行ったり、形態素

解析用の辞書を定期的に更新しなくとも、形態素解析用の機能を自然に向上できるようにすることを目的とする。

また、この発明は、形態素解析において、個々のユーザーの関心に応じた新語を抽出できるようにすることによって、情報処理における利便性を高めることを目的とする。

【課題を解決するための手段】

【００１３】

この発明にかかる形態素解析方法は、所定数の変換後文字列が登録された変換用標準辞書および新規の変換後文字列を登録するための学習用辞書、ならびに所定数の形態素が登録された形態素解析用辞書が格納されたメモリを具備し、操作部よりかな文字列の入力を受け付けた後に辞書検索に基づく変換処理を指示する操作を受け付けたことに応じて、入力されたかな文字列により変換用標準辞書および学習用辞書を検索して、変換後文字列の候補を抽出する機能と、前記操作部よりかな文字列の入力を受け付けた後に、カタカナ、ひらがな、アルファベット、数字のいずれか一の文字種を選択して変換を指示する後変換操作を受け付けたことに応じて、前記入力されたかな文字列を選択された文字種による文字列に変換する機能と、後変換操作に応じて変換され、前記変換用標準辞書および学習用辞書に登録されていない変換後文字列を前記学習用辞書に登録する機能とを具備する文字変換手段として動作するコンピュータにおいて実行されるもので、処理対象のテキストデータ（漢字およびかな混じりの文字列を含むテキストデータである。以下においても同じ。）の入力を受け付けたとき、当該テキストデータから所定長さの文字列を抽出するステップＡと、ステップＡで抽出された文字列により前記形態素解析用辞書を検索して形態素の候補を抽出するステップＢと、ステップＡで抽出された文字列を前記文字変換手段に渡して当該文字列による学習用辞書の検索を実行させ、その検索により抽出された文字列を形態素の候補として前記文字変換手段より受け付けるステップＣとを複数サイクル実行した後に、各サイクルのステップＢおよびＣにより得た候補の中の所定数を選択して出力するステップＤを実行する。また、ステップＣでは、ステップＡで抽出された文字列がひらがな文字列である場合には、この文字列を文字変換手段に渡さずに、形態素の候補を得られなかったという処理結果を設定する。またステップＤでは、ステップＣにより得たカタカナ文字列、アルファベット文字列、および数字による文字列を、ステップＢにより得た候補より優先して選択する。

【００１４】

上記において、文字変換手段により実行される変換処理では、入力されたかな文字列（ひらがな文字列である場合が多い。）を、漢字文字列のほか、アルファベット、数字、カタカナ、ひらがななどの特定の文字種による文字列に変換することができる。なお、この明細書でいうところの漢字文字列には、漢字のみから成る文字列のほか、漢字とかな（主としてひらがな）とを組み合わせた文字列も含まれるものとする。

【００１５】

変換用標準辞書には、変換後文字列に読み／品詞／使用頻度などを対応づけたものを、辞書データとして格納することができる。学習用辞書にも同様の構成の辞書データを格納することができるが、この学習用辞書は最初は空の状態でよい。また、これら２種類の辞書に加え、ユーザーの設定操作によって作成された辞書データを格納する辞書（ユーザー辞書）を設けてもよい。

【００１８】

形態素解析用辞書には、形態素に読み／品詞／他の形態素との係り受けの関係などを対応づけた辞書データを格納することができる。形態素解析処理では、入力されたテキストデータを所定位置で区切ることにより検索キーワードとなる文字列を抽出し、その検索キーワードにより前記形態素解析用辞書および学習辞書を検索して、候補を抽出することができる。

【００１９】

上記の形態素解析方法によれば、通常のかな文字列の変換処理を実行する際に行われた後変換操作に応じて学習用辞書に蓄積された変換後文字列を形態素解析処理でも使用する

ことができるので、特別の登録処理を行ったり、形態素解析用の辞書を更新しなくとも、形態素の解析機能を向上することができる。また、後変換操作によりかな文字列が名称や略語などの新語に変換され、その新語が学習用辞書に登録されると、形態素解析においても、登録された新語を含むテキストデータが入力された場合には、前記学習用辞書を用いた検索により、その新語を抽出することが可能となる。

また、この方法では、学習用辞書から抽出されたカタカナ文字列、アルファベット文字列、および数字による文字列を形態素解析用辞書から抽出された候補より優先的に選択するので、形態素解析用辞書に登録されていないカタカナ文字列、アルファベット文字列、および数字による文字列を、形態素解析処理で容易に抽出することが可能になる。

【００２０】

つぎに、形態素解析処理のために受け付けたテキストデータから抽出された文字列がひらがな文字列である場合には、ステップＣにおいて、この文字列を文字変換手段に渡さずに形態素の候補を得られなかったという処理結果を設定するのは、ひらがな文字列は、『送りがな』などを特定する用途で入力されることが多く、重要な意味のある単語を表す可能性が低いためである。このようにすれば、後変換操作により学習用辞書に登録された文字列のうち、名称や略語を表す可能性が高いカタカナ文字列、アルファベット文字列、数字による文字列のみが形態素解析処理で抽出されるようにすることができ、検索にかかる時間を短縮することができる。

【００２２】

つぎに、この発明にかかる形態素解析用のプログラムは、所定数の変換後文字列が登録された変換用標準辞書および新規の変換後文字列を登録するための学習用辞書が格納されたメモリを具備し、操作部よりかな文字列の入力を受け付けた後に辞書検索に基づく変換処理を指示する操作を受け付けたことに応じて、入力されたかな文字列により変換用標準辞書および学習用辞書を検索して、変換後文字列の候補を抽出する機能と、前記操作部よりかな文字列の入力を受け付けた後に、カタカナ、ひらがな、アルファベット、数字のいずれか一の文字種を選択して変換を指示する後変換操作を受け付けたことに応じて、前記入力されたかな文字列を選択された文字種による文字列に変換する機能と、後変換操作に応じて変換され、前記変換用標準辞書および学習用辞書に登録されていない変換後文字列を前記学習用辞書に登録する機能とを具備する文字変換手段として動作するコンピュータに導入される。

【００２３】

上記のプログラムは、処理対象のテキストデータの入力を受け付ける第１のステップ；入力されたテキストデータを形態素に分解して各形態素の組み合わせを出力する第２のステップ；の各ステップをコンピュータに実行させるためのプログラムと、前記した形態素解析用辞書を構成する電子データとを含む。このうちの第２のステップでは、前記第１のステップで受け付けたテキストデータから所定長さの文字列を抽出するステップＡと、ステップＡで抽出された文字列により形態素解析辞書を検索して形態素の候補を抽出するステップＢと、ステップＡで抽出された文字列を文字変換手段に渡して当該文字列による学習用辞書の検索を実行させ、その検索により抽出された文字列を形態素の候補として前記文字変換手段より受け付けるステップＣとを複数サイクル実行した後に、各サイクルのステップＢ及びステップＣにより得た候補の中の所定数を選択して出力するステップＤを実行する。

【００２４】

上記の形態素解析用のプログラムでは、ステップＣにおいて、ステップＡで抽出された文字列がひらがな文字列である場合には、この文字列を文字変換手段に渡さずに、形態素の候補を得られなかったという処理結果を設定する。またステップＤでは、ステップＣにより得たカタカナ文字列、アルファベット文字列、および数字による文字列を、ステップＢにより得た候補より優先して選択する。

【００２６】

上記のプログラムによれば、後変換操作により入力された単語を抽出可能な形態素解析

処理を実行することができる。また、このプログラムを一度インストールすれば、形態素解析用辞書の更新処理を行わなくとも、後変換操作に応じて学習用辞書に蓄積された変換後文字列を形態素解析で使用することができる。すなわち、ユーザーは、形態素解析のために特別な登録作業を行う必要も、形態素解析用辞書の更新データをインストールする必要もなしに、通常の文字変換処理を行うだけで、形態素解析の能力を向上させることができる。また、システム開発者も、形態素解析用辞書を定期的に更新することなく、各ユーザーの要望に応えることができる。

【００２７】

この形態素解析用のプログラムは、ウェブブラウザやメールリーダなどのアプリケーション（以下、「上位アプリケーション」という。）の稼働時に、ユーザーの操作などに応じて起動させることができる。この場合のステップＤでは、前記決定した形態素の組み合わせを、前記上位アプリケーションに出力することができる。また、これら上位アプリケーションとは別のアプリケーション（たとえば音声出力システム）に、決定した形態素の組み合わせを渡すこともできる。

【００２８】

上記のプログラムは、パーソナルコンピュータのほか、携帯電話、ＰＤＡなどの携帯端末の制御部を構成するコンピュータに組み込むことができる。また、このプログラムは、ＣＤ－ＲＯＭなどの記憶媒体に格納する方法や、電気通信回線により伝送する方法によって、ユーザーに提供することができる。また、この形態素解析用のプログラムと前記文字変換処理用のプログラムとを組み合わせたものを、１つのパッケージソフト（上記２種類のプログラムが格納された１または複数の記憶媒体から成り、各プログラムをコンピュータに同時または選択的にインストールできるようにしたもの）として提供したり、電気通信回線を介して所定のコンピュータに提供することができる。

【００２９】

つぎに、この発明にかかる情報処理装置は、操作部と、この操作部よりかな文字列の入力および変換指示操作を受け付けて、入力されたかな文字列を他の形態の文字列に変換する文字変換手段と、処理対象のテキストデータの入力を受け付けて、そのテキストデータを形態素に分解する形態素解析手段とを具備するものである。前記文字変換手段は、所定数の変換後文字列が登録された変換用標準辞書と、新規の変換後文字列を登録するための学習用辞書と、前記操作部よりかな文字列の入力を受け付けた後に辞書検索に基づく変換処理を指示する操作を受け付けたことに応じて、入力されたかな文字列により変換用標準辞書および学習用辞書を検索して、変換後文字列の候補を抽出する候補検索手段と、前記操作部よりかな文字列の入力を受け付けた後に、カタカナ、ひらがな、アルファベット、数字のうちのいずれか一の文字種を選択して変換を指示する後変換操作を受け付けたことに応じて、入力されたかな文字列を指示された文字種による文字列に変換する後変換操作時処理手段と、この後変換操作時処理手段により変換され、前記変換用標準辞書および学習用辞書に登録されていない変換後文字列を学習用辞書に登録するとともに、前記形態素解析手段から検索対象の文字列を受け付けたとき、この文字列により学習用辞書を検索して、その検索により抽出した変換後文字列を形態素解析手段に渡す学習用辞書処理手段とを具備する。一方、前記形態素解析手段は、所定数の形態素が登録された形態素解析用辞書と、処理対象のテキストデータの入力を受け付けたとき、当該テキストデータから所定長さの文字列を抽出するステップＡと、ステップＡで抽出された文字列により形態素解析用辞書を検索して形態素の候補を抽出するステップＢと、ステップＡで抽出された文字列を文字変換手段の学習用辞書処理手段に渡し、当該文字列による学習用辞書の検索を実行した学習用辞書処理手段から渡された文字列を形態素の候補として受け付けるステップＣとを複数サイクル実行する検索手段と、前記検索手段による各サイクルのステップＢおよびステップＣにより得た候補の中の所定数を選択して出力する出力手段とを具備する。

【００３０】

さらに検索手段は、ステップＡで抽出された文字列がひらがな文字列である場合には、ステップＣにおいて、前記抽出された文字列を学習用辞書処理手段に渡さずに、形態素の

候補を得られなかったという処理結果を設定する。また出力手段は、ステップＣにより得たカタカナ文字列、アルファベット文字列、および数字による文字列を、ステップＢにより得た候補より優先して選択する。

【００３１】

上記の情報処理装置において、文字変換手段と形態素解析手段とは、いずれも、プログラムによって、情報処理装置の制御用コンピュータに設定されるものである。文字変換手段は、ワードプロセッサ、電子メールエディタなどのアプリケーションに連動して動作することができる。また、形態素解析手段は、ウェブブラウザやメールリーダが動いているときに、これらからのコマンドやユーザーの呼び出し操作に応じて起動するものとして構成することができる。

【発明の効果】

【００３４】

この発明によれば、ユーザの後変換操作に応じて自動的に学習された変換後文字列を、形態素解析処理用の辞書データとして使用することが可能になるから、形態素解析のために特別な登録作業を行ったり、形態素解析用辞書の更新データを取り込んだりしなくとも、形態素解析の機能を自然に向上することができる。

【００３５】

また、この発明では、後変換操作に応じて学習用辞書に登録された変換後文字列のうちの重要な意味を持たない可能性が高いひらがな文字列が形態素の候補として抽出されないようにすることにより、学習用辞書から抽出される形態素の候補を、カタカナ文字列、アルファベット文字列、および数字による文字列に限定することができる。さらに学習用辞書から抽出された候補を形態素解析用辞書から抽出された候補より優先して選択するので、後変換操作により入力された特定の文字種（カタカナ、アルファベット、数字）による文字列を優先的に形態素の解析結果に含めることが可能になる。よって、たとえば、携帯電話において、受信メールを形態素解析し、その処理で抽出された単語を用いて返信メールを作成する際に、ユーザーが使用する可能性の高い単語を抽出するなど、ユーザーの関心に応じた新語を精度良く抽出することができる。また、システム開発者により形態素解析用辞書を更新するサービスを行わなくとも、各ユーザーは、自身の関心に応じた新語を抽出できるようになるなど、情報処理における利便性を大いに向上することができる。

【発明を実施するための最良の形態】

【００３６】

図１は、この発明が適用された情報処理装置の構成例を示す。

この情報処理装置は携帯電話（図示せず。）に組み込まれるものであって、かな漢字変換処理部１と形態素解析処理部２とを含む。これらの処理部１，２は、いずれも、プログラムによって、前記携帯電話の制御部（ＣＰＵ）に設定されるものである。また、かな漢字変換処理部１は電子メールなどの文書作成用のアプリケーションとともに動作する。一方、形態素解析処理部２は、電子メールやウェブページの閲覧用のアプリケーションなどの上位アプリケーションが動いているときに、形態素解析処理を指定する操作に応じて起動する。

【００３７】

かな漢字変換処理部１には、標準辞書１６、ユーザー辞書１７、自動学習辞書１８の３種類の辞書や、入出力部１１、変換制御部１２が設けられるほか、各辞書毎に、その辞書に対する検索や登録のための処理部１３，１４，１５が設けられる。いずれの辞書１３，１４，１５にも、単語の表記（変換後文字列）に読みや使用頻度などを対応づけた辞書データが格納される。

【００３８】

入出力部１１は、前記携帯電話の表示部に展開されるヒューマンインターフェース（図示せず。）と情報をやりとりしつつ、操作キーの入力を受け付けるものである。たとえば、かな漢字変換処理のためにかな文字列（変換前文字列）が入力されている間は、操作されたキーの種類や操作回数に基づき、入力文字を認識し、その認識した文字をヒューマン

インターフェースに出力する。また、かな文字列の入力後に変換操作が行われると、その時点での未確定の変換後文字列をヒューマンインターフェースに出力する。さらに、この入出力部１１は、未確定の変換後文字列が確定された場合には、文字入力対象のアプリケーション（メールエディタなど）に確定された文字列を出力する。

【００３９】

この情報処理装置が導入される携帯電話の操作部では、前記変換操作や変換後文字列を確定するための操作のほか、後記する後変換操作のために、それぞれ特定のキーが設定される。前記入出力部１１は、後変換用のキーが操作されると、前記ヒューマンインターフェースに文字種の選択画面を表示させ、この画面上で、カタカナ、ひらがな、アルファベット、数字のいずれかの文字種を選択させるようにしている。また、カタカナ、アルファベット、数字については、全角文字または半角文字を選択することもできる。

なお、以下で、「後変換操作」という場合には、前記後変換用のキーの操作から最終の文字種を選択する操作までを指すものとする。

【００４０】

変換制御部１２は、入出力部１１を介してかな文字列の入力を受け付け、これを各辞書の処理部１３，１４，１５に渡して検索処理を実行させる。そして、各処理部１３，１４，１５から返された検索結果に基づき、前記かな文字列を所定の変換後文字列に変換し、その変換後文字列を入出力部１１に出力する。

【００４１】

３種類の辞書のうち、標準辞書１６には、システム設計者によりあらかじめ選択された単語の辞書データが格納される。この標準辞書１６には、新たな辞書データを書き込むことはできないが、使用頻度については、辞書データの文字列が採用される都度、更新することができる。

【００４２】

ユーザー辞書１７は、ユーザーが指定した単語を登録するためのものである。変換制御部１２は、登録対象の単語について、変換後文字列、読み、品詞情報などの入力を受け付けると、これをユーザー辞書処理部１４に渡し、ユーザー辞書１７に登録させる。

【００４３】

自動学習辞書１５は、漢字文字列以外の文字種への変換を指示する後変換操作が行われたときに、その操作による変換後文字列を登録するためのものである。この自動学習辞書１８への登録処理も、変換制御部１２により制御されるが、直接の登録処理は、自動学習辞書処理部１５により行われる。

なお、ユーザー辞書１７、自動学習辞書１８についても、登録されている辞書データが使用される都度、その使用頻度を更新することができる。

【００４４】

つぎに、形態素解析処理部２には、入出力部２１、解析制御部２２、辞書検索部２３、形態素解析用辞書２４、候補制御部２５などが含められる。形態素解析用辞書２４には、品詞情報、他の形態素との係り受けの関係などの属性データを形態素の表記に対応づけた辞書データが格納される。この形態素解析用辞書２４も、前記かな漢字変換処理部１の標準辞書１６と同様に、システム設計者により作成されたもので、新規の辞書データを登録するようには設定されていない。

【００４５】

入出力部２１は、前記した上位アプリケーションと連絡するもので、処理対象のテキストデータを入力し、そのテキストデータに対する形態素解析の結果（抽出した形態素を属性データとともに順に並べたもの）を出力するように、設定される。

【００４６】

解析制御部２２は、入出力部２１から前記テキストデータの提供を受けて、このテキストデータから検索キーワードを切り出す。そして、この文字列と表記が一致する形態素を検索するために、辞書検索部２３、候補制御部２５、およびかな漢字変換処理部１のユーザー辞書処理部１４に、前記検索キーワードを出力する。

【００４７】
辞書検索部２３は、解析制御部２２から検索キーワードを渡されると、形態素解析用辞書２４を検索し、形態素の表記が前記検索キーワードに一致する辞書データを抽出する。ユーザー辞書処理部１４も、同様に、解析制御部２２から渡された検索キーワードを用いてユーザー辞書１７を検索し、変換後文字列の表記が前記検索キーワードに一致する辞書データを抽出する。これらの検索で抽出された辞書データは、いずれも解析制御部２２に渡される。
【００４８】
一方、候補制御部２５は、かな漢字変換処理部１の自動学習辞書処理部１５に前記検索キーワードを渡して、自動学習辞書１８を用いた検索を実行させる。ただし、候補制御部２５では、検索キーワードがひらがな文字列である場合には、これを自動学習辞書処理部１５に渡さずに、解析制御部２２に「候補なし」との検索結果を返す。ひらがな文字列以外の検索キーワードは、自動学習辞書処理部１５に与えられるので、自動学習辞書処理部１５は、与えられた検索キーワードにより自動学習辞書１８を検索し、変換後文字列の表記が前記検索キーワードに一致する辞書データを抽出する。抽出された辞書データは、候補制御部２５を介して解析制御部２２に渡される。
なお、形態素解析処理における自動学習辞書１８やユーザー辞書１７の使用は、データの読み出しのみであって、データの書込みは行わないのが望ましい。
【００４９】
解析制御部２２は、詳細は後記するが、複数の検索キーワードを設定し、これらの検索キーワード毎に、上記３通りの検索により形態素の候補を求める。さらに、解析制御部２２は、検索キーワード毎に得た候補の中から、形態素の組み合わせとして最適なものを決定し、その決定に基づいて前記テキストデータの構造を表したデータ（以下、「解析データ」という。）を出力する。なお、解析データは、形態素の区切り位置のほか、各形態素の品詞情報などを含むものとなる。
【００５０】
以下、前記かな漢字変換処理部１および形態素解析処理部２で実行される処理のうち、この発明に関連する処理について、詳細に説明する。なお、図２以下のフローチャートでは、各ステップを「ＳＴ」と略記し、各ステップを３桁の番号で示す。
【００５１】
図２は、かな漢字変換処理部１による文字入力処理の手順を示す。なお、この実施例の手順では、複数文節分の文字列を入力して一括変換した後、最終的な確定操作が行われるまで、必要な文節を修正できるようにしている。
【００５２】
この手順は、かな文字の入力操作が開始されることによってスタートする。まず、最初のＳＴ１０１では、ユーザーによる操作の内容を判断する。ユーザーは、文字入力のために、所定のかな文字が割り当てられたキーを操作するが、その操作の都度、ＳＴ１０１からＳＴ１１５に進み、その操作に応じた処理（たとえば、入力文字をメモリ内の一時保存領域に蓄積する処理、蓄積された文字列を表示する処理など）を実行する。
【００５３】
所定数のかな文字が入力された時点で変換操作が行われると、ＳＴ１０１からＳＴ１０２に進み、蓄積された入力文字列（かな文字列）を取得する。つぎのＳＴ１０３では、この入力文字列を変換する処理（かな漢字変換処理）を実行する。このかな漢字変換処理は、公知の技術であるので、詳細は省略するが、入力文字列を所定数の文節に切り分け、標準辞書処理部１６、ユーザー辞書処理部１７、自動学習辞書処理部１８に検索を実行させることによって、文節毎に最適な候補を抽出する。この実施例では、抽出された文節毎の候補を組み合わせた変換後文字列は、未確定の文字列として表示部に表示されるものとする。このとき、先頭の文節が注目文節として初期設定され、カーソルによる識別表示が行われる。
【００５４】

この実施例では、前記未確定の変換後文字列に含まれる文節毎に、後変換フラグを設定している。この後変換フラグは、後変換操作による変換処理を受けた旨を記憶するためのものであり、かな漢字変換処理の直後のＳＴ１０４では、すべての文節の後変換処理をリセットする処理を実行する。

【００５５】

ここで変換後文字列がユーザーが目的とする文字列に一致していれば、確定操作が行われる。この操作により、ＳＴ１０５，１０６が「ＹＥＳ」となってＳＴ１１３に進み、後記する自動学習処理を実行する。さらに、つぎのＳＴ１１４では、前記未確定の変換後文字列を確定して出力する。これにより、入力文字列に対する処理は終了したことになる。

【００５６】

一方、前記未確定の変換後文字列の中に、ユーザーが意図しない文字列を含む文節がある場合には、ユーザーは、その文節にカーソルを移動させ、修正操作を実行する。これらの操作が行われる毎に、ＳＴ１０５が「ＹＥＳ」、ＳＴ１０６が「ＮＯ」となり、ＳＴ１０７以下の処理に進む。

【００５７】

まず、カーソルの移動操作が行われた場合には、ＳＴ１０７が「ＹＥＳ」となってＳＴ１０８に進み、前記操作に応じて注目文節を変更する。このとき、前記カーソルも、移動操作に応じて注目文節に移動することになる。

【００５８】

つぎに、注目文節に対し、後変換操作が行われると、ＳＴ１０９が「ＹＥＳ」となってＳＴ１１０に進み、注目文節を操作に応じた文字種に変更する。さらに、つぎのＳＴ１１１では、前記注目文節の後変換フラグをオンにする。

【００５９】

また、注目文節に対し、後変換操作以外の操作（他の変換候補を選択する操作、変換をキャンセルする操作など）が行われた場合には、ＳＴ１１２に進み、その操作に応じた処理を実行する。修正処理が終了し、確定操作が行われると、前記したＳＴ１１３、１１４を実行し、処理を終了する。

【００６０】

図３は、前記ＳＴ１１３の自動学習処理について、詳細な手順を示す。

この処理では、ＳＴ２０１において、先頭の文節を注目文節に設定した後、ＳＴ２０２～２０８のループを繰り返すことにより、各文節を順に処理する。ただし、後変換フラグがオフの場合には、つぎの文節に注目文節を移すだけであり、後変換フラグがセットされている文節に対してのみ、実質的な処理（ＳＴ２０３～２０６）を実行する。

【００６１】

注目文節に後変換フラグがセットされている場合、ＳＴ２０３では、その注目文節の読み（変換前のかな文字列）と変換後文字列とを対応づけた辞書データを設定する。つぎのＳＴ２０４では、自動学習辞書１８に空き領域があるか否かをチェックする。空き領域がある場合には、ＳＴ２０５をスキップしてＳＴ２０６に進み、その空き領域に新規の辞書データを登録する。なお、この登録の際に、辞書データには、使用頻度や登録順序が付加される。

【００６２】

自動学習辞書１８に空き領域がない場合には、ＳＴ２０４からＳＴ２０５に進み、前記した登録順序に基づき、辞書内の最も古い辞書データを抽出して、これを削除する。ただし、この辞書データの使用頻度が所定値以上である場合には、つぎに古い辞書データを抽出するものとする。これにより、登録後の経過時間が比較的長い辞書データの中から使用頻度が最も少ないものを削除することができる。この後は、ＳＴ２０６に進み、前記削除により空いた領域に新規の辞書データを登録する。

【００６３】

なお、自動学習辞書１８内の辞書データは、登録順序や使用頻度に基づき、適宜ソートするようにしてもよい。

また、文字入力処理の手順は、上記図２，３に限らず、たとえば、入力文字列を一括変換した後に、文節毎に確定できるようにしてもよい。この場合には、確定操作が行われる毎に、その操作にかかる文節の後変換フラグをチェックし、この後変換フラグがオンであれば、確定後の文字列を自動学習辞書１８に登録する処理を行うことになる。

【００６４】

図４は、前記自動学習辞書１８のデータ構成例を示す。図中の「読み」は、前記した入力文字列に、「表記」は変換後文字列に、それぞれ対応する。読みには、ひらがな文字列のほか、半角数字による文字列も含まれている。表記には、後変換操作における選択に応じて、カタカナ、ひらがな、アルファベット、数字のいずれかの文字種による文字列が格納される。

【００６５】

つぎに、図５を用いて、形態素解析処理の概要を説明する。

この実施例では、処理対象のテキストデータを文の単位に区切った上で、文毎に、形態素解析の一手法である最長一致法を実行するようにしている。すなわち、処理対象の文を構成する文字列（以下、「処理対象文字列」という。）内の各文字（図５の例の場合、「朝」「日」「が」「富」・・・の各文字）に順に注目して検索を実行する。この検索では、注目文字から文の末尾までの文字列を検索対象として、１番目の文字から所定位置の文字までの文字列に一致する文字列を抽出する。

【００６６】

たとえば、図５の処理対象文字列中の４番目の文字「富」を例にして説明すると、この「富」から図示しない末尾の文字までの文字列を検索対象文字列として、前記「富」から所定位置の文字までの文字列に一致する文字列を抽出する検索を実行する。この結果、図示例では、「富士山」が最も長い文字列として抽出され、以下、「富士」、「富」の２種類の文字列が抽出されている。

【００６７】

処理対象文字列内の各文字毎に抽出された文字列は、それぞれ、その文字から始まる形態素の候補としてメモリに格納される。以下では、文字毎に抽出された候補が格納されるメモリ領域を「検索結果リスト」といい、すべての文字の検索結果リストを含むメモリ領域を「候補リスト」ということにする。

【００６８】

このようにして候補リストが作成されると、処理対象文字列の文字の並びに沿って各検索結果リストをチェックし、所定の候補を形態素として選択する。この場合に、最長一致法では、文字列の長いものから優先して候補を選択し、先に選択した候補に含まれる文字について、候補の選択をスキップするようにしている。ただし、候補の選択は１組に限らず、複数とおりの候補の組み合わせを設定し、その中から形態素の区切りが最適なものを選択するようにしている。

【００６９】

図６は、前記形態素解析処理部２における詳細な処理の手順を示す。なお、この手順は、テキストデータ中の１つの文に対するものである。テキストデータ中に複数の文が含まれる場合には、この図６の手順が文毎に実行されることになる。

【００７０】

まず、最初のＳＴ３０１では、処理対象文字列の先頭の文字を注目文字として初期設定する。つぎのＳＴ３０２では、前記候補リストを設定するためのメモリ領域をクリアし、しかる後にＳＴ３０３に進む。

【００７１】

ＳＴ３０３では、前記注目文字から処理対象文字列の末尾の文字までの文字列を切り出し、これを検索対象として設定する。続くＳＴ３０４では、この検索対象の文字列を用いた検索処理を実行する。この検索処理の詳細については後述する。

【００７２】

検索が終了すると、ＳＴ３０５において、候補が抽出されたかどうかを判断する。この

判断結果が「ＮＯ」の場合には、ＳＴ３０６に進み、注目文字のみから成る文字列を候補として抽出する。検索処理で候補が抽出された場合には、このＳＴ３０６をスキップしてＳＴ３０７に進む。ＳＴ３０７では、抽出された候補を、注目文字の検索結果リストに格納する。ＳＴ３０６を実行した場合にも、ＳＴ３０７において、注目文字から成る文字列を検索結果リストに格納する。

【００７３】

以下、ＳＴ３０８，３０９によって注目文字を処理対象文字列の末尾の文字まで動かしながら、上記ＳＴ３０３～３０７の処理を繰り返す。この後は、ＳＴ３０８からＳＴ３１０に進み、前記した候補リストから候補の組み合わせを設定し、その中から最適なものを選択する処理を実行する。そして、ＳＴ３１１では、選択した候補の組み合わせに基づいて解析データを作成し、これを上位のアプリケーションに出力する。

【００７４】

図７は、前記ＳＴ３０４の候補検索処理、すなわち、前記ＳＴ３０３で設定された検索対象の文字列を用いた検索処理の詳細な手順を示す。

まず、ＳＴ４０１では、現在の注目文字に対応する検索結果リストをクリアする。つぎに、ＳＴ４０２において、検索対象の文字列の長さをカウンタＬに格納した後、ＳＴ４０３において、前記検索対象の文字列を検索キーワードとして設定する。

【００７５】

以下では、設定された検索キーワードの末尾を一文字ずつ削除しながら、前記した各辞書に対する検索を実行する。まず、ＳＴ４０４では、候補制御部２５に検索キーワードを出力することによって、自動学習辞書１８を検索させ、つぎのＳＴ４０５で、その検索結果を取得する。なお、この検索結果に候補が含まれている場合には、ＳＴ４０５では、その候補を検索結果リストに格納する処理を実行する。

【００７６】

ＳＴ４０６では、ユーザー辞書処理部１４に検索キーワードを出力することによって、前記ユーザー辞書１７を検索させ、つぎのＳＴ４０７で、その検索結果を取得する。このＳＴ４０７でも、検索結果に候補が含まれている場合には、その候補を検索結果リストに格納する処理を実行する。

【００７７】

つぎのＳＴ４０８では、辞書検索部２３に検索キーワードを出力することによって、前記形態素解析用辞書２４を検索させ、続くＳＴ４０９で、その検索結果を取得する。このＳＴ４０９でも、検索結果に候補が含まれている場合には、その候補を検索結果リストに格納する処理を実行する。

【００７８】

このようにして、各辞書に対する検索を終了すると、前記Ｌの値を１つ小さくする。更新後のＬが０より大きければ、ＳＴ４１２に進み、検索キーワードの末尾文字を削除した後、ＳＴ４０４に戻る。Ｌの値が０となった場合は、処理を終了する。

【００７９】

上記の形態素解析処理において、先に後変換操作によって確定され、自動学習辞書１８に登録された文字列（以下、「自動登録文字列」という。）が処理対象文字列中に含まれている場合には、この自動登録文字列が検索キーワードとして設定されたときの前記ＳＴ４０４，４０５の処理によって、自動登録文字列を検索結果リストに格納することができる。

【００８０】

ただし、検索キーワードがひらがな文字列である場合には、前記候補制御部２５は、この文字列による検索処理をキャンセルするから、この検索キーワードが自動登録文字列であっても、その文字列は候補として抽出されない。すなわち、自動学習辞書１８から候補として抽出されるのは、カタカナ、アルファベット、数字の３種類の文字種による文字列に限定されることになる。また、重要な意味を持たない可能性が高いひらがな文字列について、自動学習辞書１８による検索をスキップすることができるので、検索にかかる時間

を短縮することができる。

【００８１】

なお、前記図６のＳＴ３１０で、各検索結果リストの候補を組み合わせる際には、自動学習辞書１８から抽出された候補を最優先し、つぎにユーザー辞書１７から抽出された候補を選択するようにしている。よって、文字入力処理においてユーザーが後変換操作により入力した単語や、ユーザーの登録処理によってユーザー辞書に登録された単語を、形態素解析処理でも優先的に抽出することができる。

【００８２】

また、この実施例では、自動学習辞書１８には、後変換操作により変換された文字列のみが登録されるものとしたが、これに限らず、標準辞書１６やユーザー辞書１７に登録されていない漢字文字列を登録することもできる。たとえば、１文節として変換された文字列が意図しない文字列であったために、ユーザーが入力文字列を２つに区切って文節毎に変換操作を行うことにより、所望の文字列を入力した場合、確定された漢字文字列を自動学習辞書１８に登録することができる。このように自動学習辞書１８に漢字文字列を登録した場合には、形態素解析処理でも、これらの漢字文字列まで抽出できるようにするとよい。

【００８３】

また、文字変換処理では、自動学習辞書１８に登録された文字列を通常の変換操作により呼び出して使用した場合に、使用頻度を更新することができる。よって、形態素解析処理でも、自動学習辞書１８に対する検索によって自動登録文字列を抽出した場合には、その文字列の使用頻度を考慮して候補の組み合わせ処理を行うようにしてもよい。

【図面の簡単な説明】

【００８４】

【図１】この発明が適用された情報処理装置の機能ブロック図である。

【図２】文字入力処理の手順を示すフローチャートである。

【図３】自動学習処理の手順を示すフローチャートである。

【図４】自動学習辞書のデータ構成を示す図である。

【図５】形態素解析処理の概要を説明する図である。

【図６】形態素解析処理の手順を示すフローチャートである。

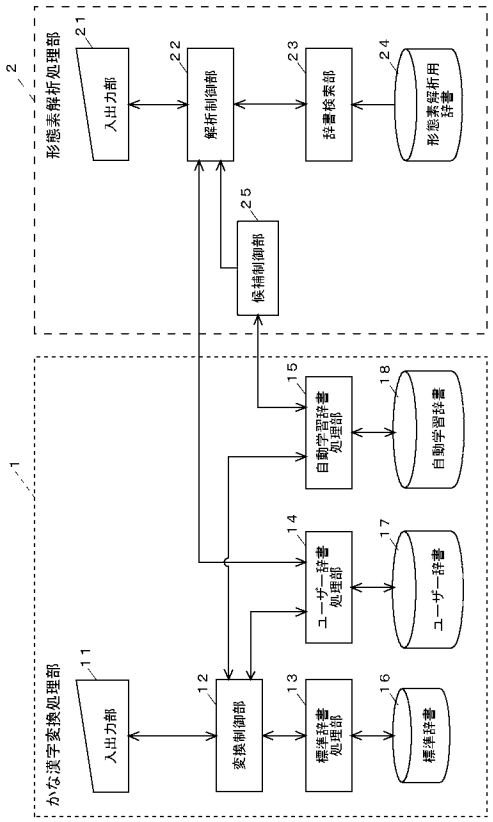
【図７】候補検索処理の手順を示すフローチャートである。

【符号の説明】

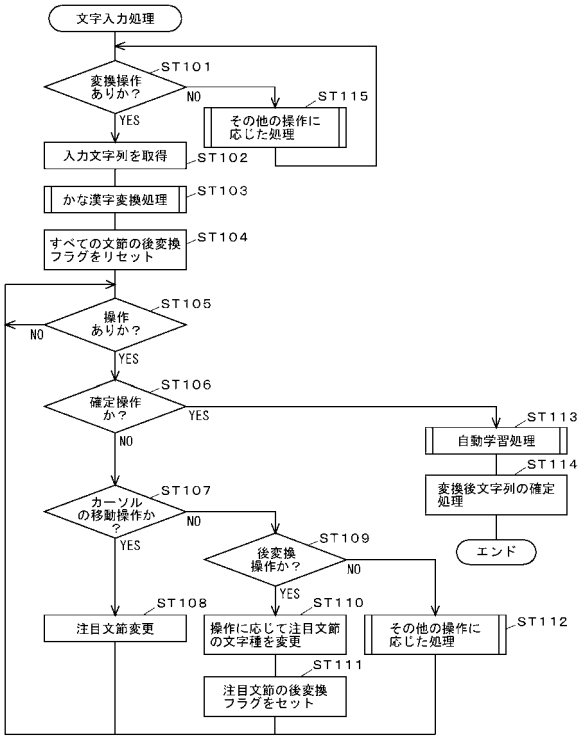
【００８５】

１　かな漢字変換処理部
２　形態素解析処理部
１１、２１　入出力部
１２　変換制御部
１５　自動学習辞書処理部
１６　標準辞書
１８　自動学習辞書
２２　解析制御部
２３　辞書検索部
２４　形態素解析用辞書
２５　候補制御部

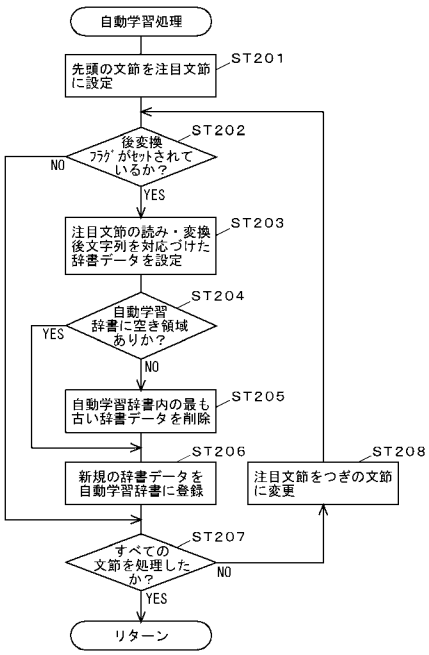
【図１】


形態素解析処理部
2
入出力部
21
解析制御部
22
辞書検索部
23
形態素解析用辞書
24
候補制御部
25
かな漢字変換処理部
1
自動学習辞書処理部
15
自動学習辞書
18
ユーザー辞書処理部
14
ユーザー辞書
17
入出力部
11
変換制御部
12
標準辞書処理部
13
標準辞書
16

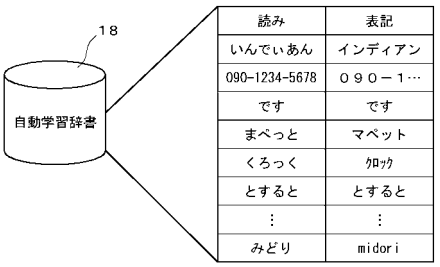
【図２】


文字入力処理
変換操作ありか？
ST101
NO
YES
ST115
その他の操作に応じた処理
入力文字列を取得
ST102
かな漢字変換処理
ST103
すべての文節の後変換フラグをリセット
ST104
操作ありか？
ST105
NO
YES
確定操作か？
ST106
YES
NO
ST113
自動学習処理
ST114
変換後文字列の確定処理
エンド
カーソルの移動操作か？
ST107
NO
YES
後変換操作か？
ST109
NO
YES
ST108
注目文節変更
ST110
操作に応じて注目文節の文字種を変更
ST112
その他の操作に応じた処理
ST111
注目文節の後変換フラグをセット

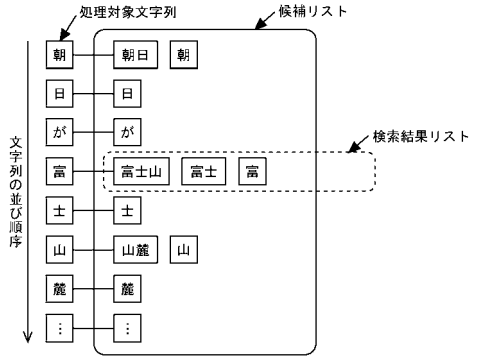
【図３】


自動学習処理
先頭の文節を注目文節に設定
ST201
後変換ﾌﾗｸﾞがｾｯﾄされているか？
ST202
NO
YES
注目文節の読み・変換後文字列を対応づけた辞書データを設定
ST203
自動学習辞書に空き領域ありか？
ST204
YES
NO
自動学習辞書内の最も古い辞書データを削除
ST205
新規の辞書データを自動学習辞書に登録
ST206
注目文節をつぎの文節に変更
ST208
すべての文節を処理したか？
ST207
NO
YES
リターン

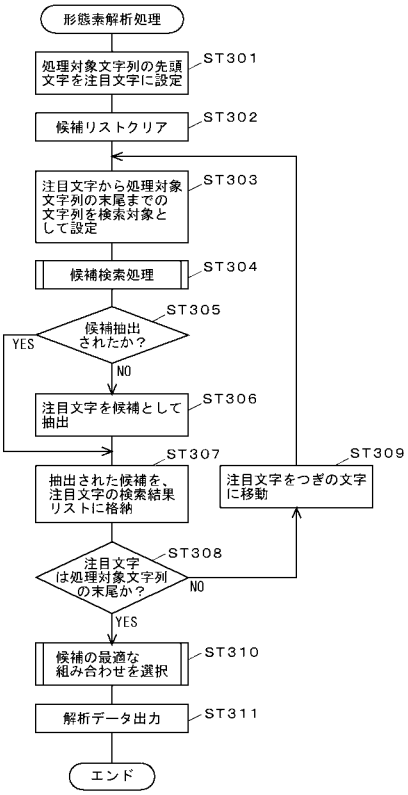
【図４】


18
自動学習辞書
読み
表記
いんでぃあん
インディアン
090-1234-5678
０９０－１…
です
です
まぺっと
マペット
くろっく
ｸﾛｯｸ
とすると
とすると
⋮
⋮
みどり
midori

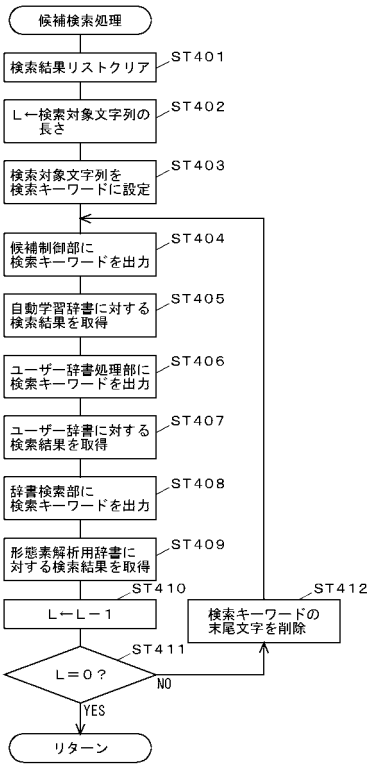
【図５】


処理対象文字列
候補リスト
検索結果リスト
文字列の並び順序
朝
朝日
朝
日
日
が
が
富
富士山
富士
富
士
士
山
山麓
山
麓
麓
⋮
⋮

【図６】


形態素解析処理
処理対象文字列の先頭文字を注目文字に設定
ST301
候補リストクリア
ST302
注目文字から処理対象文字列の末尾までの文字列を検索対象として設定
ST303
候補検索処理
ST304
候補抽出されたか？
ST305
YES
NO
注目文字を候補として抽出
ST306
抽出された候補を、注目文字の検索結果リストに格納
ST307
注目文字をつぎの文字に移動
ST309
注目文字は処理対象文字列の末尾か？
ST308
NO
YES
候補の最適な組み合わせを選択
ST310
解析データ出力
ST311
エンド

【図７】


候補検索処理
検索結果リストクリア
ST401
L←検索対象文字列の長さ
ST402
検索対象文字列を検索キーワードに設定
ST403
候補制御部に検索キーワードを出力
ST404
自動学習辞書に対する検索結果を取得
ST405
ユーザー辞書処理部に検索キーワードを出力
ST406
ユーザー辞書に対する検索結果を取得
ST407
辞書検索部に検索キーワードを出力
ST408
形態素解析用辞書に対する検索結果を取得
ST409
L←L－１
ST410
検索キーワードの末尾文字を削除
ST412
L＝０？
ST411
NO
YES
リターン

フロントページの続き

審査官　今村　剛

(56)参考文献　特開２００３－０５８５３７（ＪＰ，Ａ）
特開平０２－２９７１５１（ＪＰ，Ａ）
特開平０８－１５３０９８（ＪＰ，Ａ）

(58)調査した分野(Int.Cl.，ＤＢ名)
Ｇ０６Ｆ１７／２０－１７／２８