

(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2004-265014  
(P2004-265014A)

(43) 公開日 平成16年9月24日(2004.9.24)

(51) Int. Cl.<sup>7</sup>  
G06F 17/21

F I  
G06F 17/21 550A  
G06F 17/21 550J

テーマコード(参考)  
5B009

審査請求 有 請求項の数 22 O L (全 26 頁)

(21) 出願番号 特願2003-53258 (P2003-53258)  
(22) 出願日 平成15年2月28日(2003.2.28)

(71) 出願人 301022471  
独立行政法人情報通信研究機構  
東京都小金井市貫井北町4-2-1  
(74) 代理人 100103827  
弁理士 平岡 憲一  
(74) 代理人 100097836  
弁理士 福井 國敏  
(72) 発明者 村田 真樹  
東京都小金井市貫井北町4-2-1 独立  
行政法人通信総合研究所内  
Fターム(参考) 5B009 QA01 QA03 QA14 RB32

(54) 【発明の名称】 個人環境頻度記憶装置及び個人環境言語変換装置及び個人環境差分強調装置及びプログラム

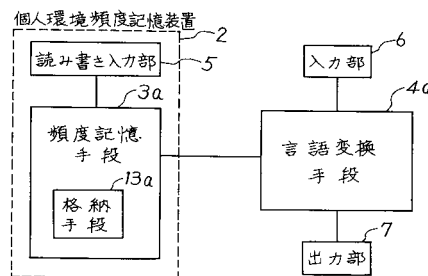
(57) 【要約】

【課題】 個人環境での文字列の頻度情報により、入力文の個人環境への言語変換や差分強調を容易に行えるようにすること。

【解決手段】 個人環境での読み書きの入力を行う読み書き入力部5と、文字列の頻度情報の検索を行う頻度記憶手段3aとを備え、前記頻度記憶手段3aは、前記読み書き入力部5から入力された個人環境での読み書きデータから任意の文字列を抽出し、該抽出した文字列毎の個数をカウントする。

【選択図】 図1

本発明の原理説明図



**【特許請求の範囲】****【請求項 1】**

個人環境での読み書きの入力を行う読み書き入力部と、  
文字列の頻度情報の検索を行う頻度記憶手段とを備え、  
前記頻度記憶手段は、前記読み書き入力部から入力された個人環境での読み書きデータから任意の文字列を抽出し、該抽出した文字列毎の個数をカウントすることを特徴とした個人環境頻度記憶装置。

**【請求項 2】**

前記頻度記憶手段に読み入力検出部と書き入力検出部とを備え、  
前記頻度記憶手段は、前記書き入力検出部で検出した前記任意の文字列の出現頻度の重みを前記読み入力検出部で検出した前記任意の文字列の出現頻度の重みより重くして前記任意の文字列の出現頻度を求めることを特徴とした請求項 1 記載の個人環境頻度記憶装置。 10

**【請求項 3】**

前記読み入力検出部は、表示時間の短いものは前記任意の文字列の出現頻度として検出しないようにして前記任意の文字列の出現頻度を求めることを特徴とした請求項 2 記載の個人環境頻度記憶装置。

**【請求項 4】**

前記頻度記憶手段は、前記読み書き入力部から入力された読み書きデータのうち古いものを削除して前記任意の文字列の出現頻度を求めることを特徴とした請求項 1 ~ 3 のいずれかに記載の個人環境頻度記憶装置。 20

**【請求項 5】**

前記頻度記憶手段は、前記読み書き入力部から入力された読み書きデータから古いものの重みを軽くして前記任意の文字列の出現頻度を求めることを特徴とした請求項 1 ~ 3 のいずれかに記載の個人環境頻度記憶装置。

**【請求項 6】**

前記請求項 1 ~ 5 のいずれかに記載の個人環境頻度記憶装置と、  
入力文を個人環境言語に変換して出力する言語変換手段とを備え、  
前記言語変換手段は、入力された文字列を前記個人環境頻度記憶装置に格納されている出現頻度の高い文字列に変換して出力することを特徴とした個人環境言語変換装置。 30

**【請求項 7】**

前記請求項 1 ~ 5 のいずれかに記載の個人環境頻度記憶装置と、  
入力文を個人環境言語に変換して出力する言語変換手段とを備え、  
前記言語変換手段は、入力された文字列のうち前記個人環境頻度記憶装置に格納されている出現頻度の高い文字列を括弧づけで補助表記して出力することを特徴とした個人環境言語変換装置。 30

**【請求項 8】**

前記請求項 1 ~ 5 のいずれかに記載の個人環境頻度記憶装置と、  
入力文の文字列の差分を強調して表示する差分強調装置とを備え、  
前記差分強調装置は、入力された文字列のうち前記個人環境頻度記憶装置に格納されている文字列の出現頻度がある閾値以下のものを強調表示することを特徴とした個人環境差分強調装置。 40

**【請求項 9】**

前記請求項 1 ~ 5 のいずれかに記載の個人環境頻度記憶装置と、  
入力文の文字列の差分を強調して表示する差分強調装置とを備え、  
前記差分強調装置は、入力された文字列のうち前記個人環境頻度記憶装置に格納されている文字列の出現頻度がある閾値以上のものを強調表示することを特徴とした個人環境差分強調装置。

**【請求項 10】**

個人環境での書きの入力を行う書き入力部と、  
文字列の頻度情報の検索を行う頻度記憶手段を備え、 50

前記頻度記憶手段は、前記書き入力部から入力された個人環境での書きデータから任意の文字列を抽出し、該抽出した文字列毎の個数をカウントする個人環境頻度記憶装置と、入力文を個人環境言語に変換して出力する言語変換手段とを備え、前記言語変換手段は、入力された文字列のうち前記個人環境頻度記憶装置に格納されている出現頻度の高い文字列を括弧づけで補助表記して出力することを特徴とした個人環境言語変換装置。

【請求項 1 1】

個人環境での書きの入力を行う書き入力部と、文字列の頻度情報の検索を行う頻度記憶手段を備え、前記頻度記憶手段は、前記書き入力部から入力された個人環境での書きデータから任意の文字列を抽出し、該抽出した文字列毎の個数をカウントする個人環境頻度記憶装置と、入力文の文字列の差分を強調して表示する差分強調装置とを備え、前記差分強調装置は、入力された文字列のうち前記個人環境頻度記憶装置に格納されている文字列の出現頻度がある閾値以下のものを強調表示することを特徴とした個人環境差分強調装置。

【請求項 1 2】

個人環境での書きの入力を行う書き入力部と、文字列の頻度情報の検索を行う頻度記憶手段を備え、前記頻度記憶手段は、前記書き入力部から入力された個人環境での書きデータから任意の文字列を抽出し、該抽出した文字列毎の個数をカウントする個人環境頻度記憶装置と、入力された文字列の差分として出力する対象の単位である抽出単位と入力された文字列の差分を検出するために比較する領域の単位である検出領域を設定し、抽出手段で入力された文字列の現在の前記検出領域以外の領域から全ての前記抽出単位に相当するものを抽出して前記格納手段に格納し、現在の前記検出領域において、前記格納手段に格納されていない前記抽出単位に相当するものを強調表示して現在の検出領域の文書出力することを、前記検出領域ごとに繰り返す差分強調装置とを備え、前記差分強調装置は、前記抽出手段で強調表示すべきと判断された前記入力された文字列のうち前記個人環境頻度記憶装置に格納されている文字列の出現頻度がある閾値以下のものを強調表示することを特徴とした個人環境差分強調装置。

【請求項 1 3】

個人環境での書きの入力を行う書き入力部と、文字列の頻度情報の検索を行う頻度記憶手段を備え、前記頻度記憶手段は、前記書き入力部から入力された個人環境での書きデータから任意の文字列を抽出し、該抽出した文字列毎の個数をカウントする個人環境頻度記憶装置と、入力された文字列の差分として出力する対象の単位である抽出単位と入力された文字列の差分を検出するために比較する領域の単位である検出領域を設定し、抽出手段で、入力された文書データの現在の前記検出領域において、格納手段に格納されていない前記抽出単位に相当するものを強調表示して現在の検出領域の文書出力し、前記強調表示したものを前記格納手段に格納することを、前記検出領域ごとに繰り返す差分強調装置とを備え、前記差分強調装置は、前記抽出手段で強調表示すべきと判断された前記入力された文字列のうち前記個人環境頻度記憶装置に格納されている文字列の出現頻度がある閾値以下のものを強調表示することを特徴とした個人環境差分強調装置。

【請求項 1 4】

個人環境での書きの入力を行う書き入力部と、文字列の頻度情報の検索を行う頻度記憶手段を備え、前記頻度記憶手段は、前記書き入力部から入力された個人環境での書きデータから任意の文字列を抽出し、該抽出した文字列毎の個数をカウントする個人環境頻度記憶装置と、入力文の文字列の差分を強調して表示する差分強調装置とを備え、前記差分強調装置は、入力された文字列のうち前記個人環境頻度記憶装置に格納されている文字列の出現頻度がある閾値以上のものを強調表示することを特徴とした個人環境差分

強調装置。

【請求項 15】

個人環境での書きの入力を行う書き入力部と、  
文字列の頻度情報の検索を行う頻度記憶手段を備え、  
前記頻度記憶手段は、前記書き入力部から入力された個人環境での書きデータから任意の文字列を抽出し、該抽出した文字列毎の個数をカウントする個人環境頻度記憶装置と、  
入力文の文字列の差分を強調して表示する差分強調装置とを備え、  
前記差分強調装置は、入力された文字列のうちユーザ個人での文字列出現確率が一般的なテキスト集合での出現確率より有意に大きいものだけを強調表示することを特徴とした個人環境差分強調装置。

10

【請求項 16】

個人環境での書きの入力を行う書き入力部と、  
文字列の頻度情報の検索を行う頻度記憶手段を備え、  
前記頻度記憶手段は、前記書き入力部から入力された個人環境での書きデータから任意の文字列を抽出し、該抽出した文字列毎の個数をカウントする個人環境頻度記憶装置と、  
入力文の文字列の差分を強調して表示する差分強調装置とを備え、  
前記差分強調装置は、入力された文字列のうちユーザ個人での文字列出現確率が一般的なテキスト集合での出現確率のある値倍より大きいものだけを強調表示することを特徴とした個人環境差分強調装置。

【請求項 17】

個人環境での読み書きの入力を行う読み書き入力手段と、  
該読み書き入力手段から入力された個人環境での読み書きデータから任意の文字列を抽出し、該抽出した文字列毎の個数をカウントする頻度記憶手段として、コンピュータを機能させるためのプログラム。

20

【請求項 18】

個人環境での読み書きの入力を行う読み書き入力手段と、  
該読み書き入力手段から入力された個人環境での読み書きデータから任意の文字列を抽出し、該抽出した文字列毎の個数をカウントして出現頻度を求める頻度記憶手段と、  
入力された文字列を前記頻度記憶手段に格納されている出現頻度の高い文字列に変換して出力する言語変換手段として、コンピュータを機能させるためのプログラム。

30

【請求項 19】

個人環境での読み書きの入力を行う読み書き入力手段と、  
該読み書き入力手段から入力された個人環境での読み書きデータから任意の文字列を抽出し、該抽出した文字列毎の個数をカウントして出現頻度を求める頻度記憶手段と、  
入力された文字列のうち前記頻度記憶手段に格納されている出現頻度の高い文字列を括弧づけで補助表記して出力する言語変換手段として、コンピュータを機能させるためのプログラム。

【請求項 20】

個人環境での読み書きの入力を行う読み書き入力手段と、  
該読み書き入力手段から入力された個人環境での読み書きデータから任意の文字列を抽出し、該抽出した文字列毎の個数をカウントして出現頻度を求める頻度記憶手段と、  
入力された文字列のうち前記頻度記憶手段に格納されている文字列の出現頻度がある閾値以下のものを強調表示する差分強調装置として、コンピュータを機能させるためのプログラム。

40

【請求項 21】

個人環境での読み書きの入力を行う読み書き入力手段と、  
該読み書き入力手段から入力された個人環境での読み書きデータから任意の文字列を抽出し、該抽出した文字列毎の個数をカウントして出現頻度を求める頻度記憶手段と、  
入力された文字列のうち前記頻度記憶手段に格納されている文字列の出現頻度がある閾値以上のものを強調表示する差分強調装置として、コンピュータを機能させるためのプログラム。

50

ラム。

【請求項 2 2】

個人環境での書きの入力を行う書き入力手段と、  
前記書き入力手段から入力された個人環境での書きデータから任意の文字列を抽出し、該抽出した文字列毎の個数をカウントする頻度記憶手段と、  
入力された文字列の差分として出力する対象の単位である抽出単位と入力された文字列の差分を検出するために比較する領域の単位である検出領域を設定し、抽出手段で、入力された文書データの現在の前記検出領域において、格納手段に格納されていない前記抽出単位に相当するものを強調表示して現在の検出領域の文書を出力し、前記強調表示したものを前記格納手段に格納することを、前記検出領域ごとに繰り返す差分強調装置と、  
前記抽出手段で強調表示すべきと判断された前記入力された文字列のうち、前記頻度記憶手段に格納されている文字列の出現頻度がある閾値以下のものを強調表示する前記差分強調装置として、コンピュータを機能させるためのプログラム。

10

【発明の詳細な説明】

【0001】

【発明の属する技術分野】

本発明は、個人環境での読み書きの入力データから文字列の頻度情報を求める個人環境頻度記憶装置と、該個人環境頻度記憶装置を使用する個人環境言語変換装置及び個人環境差分強調装置及びプログラムに関する。

【0002】

特に、個人環境の読み書きシステムにおいてその個人の良く知っている文字列（単語）を認定する個人環境頻度記憶装置及びプログラムと、なるべく個人の良く知っている文字列（単語）を使って表示するようにする個人環境言語変換装置及びプログラムと、個人環境での文字列の出現頻度により文字列を強調表示する個人環境差分強調装置及びプログラムに関する。

20

【0003】

【従来の技術】

従来、自然言語で記述された文または文章に関する表現の変換処理として典型的なものは、機械翻訳である。機械翻訳では、ある国の自然言語で記述された文または文章を他の国の自然言語で記述された文または文章に変換する。

30

【0004】

機械翻訳が他の国の言語に変換するのに対し、同一の自然言語間での文または文章の変換処理を行うシステムも用いられるようになってきている。例えば、要約文を自動生成したり、文章を推敲したりするシステムである。

【0005】

一般に同一自然言語間での文の変換処理では、変換前の語・句・文などのパターンと変換後の語・句・文などのパターンとの対からなる変換規則を大量に用意し、いわゆるパターン・マッチングによって入力文中に現れる変換前のパターンを探し出し、該当するパターンがあれば、それを変換後の語・句・文などのパターンに置き換える処理を行うものであった。

40

【0006】

また、各種の言い換えを統一的に扱うことができるようにするため、ある自然言語で記述された文字列を、同一の自然言語で記述された他の表現による文字列に変換するシステムとして次のものが提案されている。

【0007】

主要なモジュールとして変形処理手段と評価処理手段と、自然言語の文字列に関する変形の規則を記憶する変形規則記憶手段と、文字列を変形した結果が目的とするふさわしい変換であるかどうかを評価するための尺度を与える評価関数または評価規則を記憶する評価情報記憶手段とを持ち、これらを変換の目的に応じて交換できるようにする。さらに、変形規則記憶手段および評価情報記憶手段に、変形規則および評価関数等を複数種類用意し

50

、変換の目的に応じて選択できるようにする。

【0008】

これにより、複数の言語変換処理システムを開発する場合の開発コストを低減し、また、複数の言語変換処理システムを統一されたインタフェースで利用可能にすることができるものがあつた（例えば、非特許文献1及び特願2001-205889参照）。

【0009】

【非特許文献1】

村田真樹，井佐原均， 言い換えの統一的モデル - 尺度に基づく変形の利用 - ，2001年 3月30日言語処理学会第7回年次大会ワークショップ  
論文集，p. 21～26

10

【0010】

【発明が解決しようとする課題】

上記従来の要約文の自動生成、文章の推敲等の同一自然言語内での文または文章の変換処理では、一般に変換規則による一律な変換を行うものであり、平易文生成、要約文生成、文章の推敲といった変換の目的に応じて、それぞれ個別に独自のシステムを構築する必要があり、個人環境に対応した言語変換を自動で行えるものではなかつた。

【0011】

また、各種の言い換えを統一的に扱うことができるようにしたものは、個人の良く知っている単語を認定し、各個人の良く知っている単語を使って表示できるものではなかつた。

【0012】

本発明は上記問題点の解決を図り、個人環境の読み書きシステムにおいてその個人の良く知っている単語を認定し、なるべくその単語を使って表示するようにすることで、各個人にとってわかりやすい表現とすることを目的とする。

20

【0013】

また、個人環境での文字列の出現頻度により文字列を強調表示することで、自分の苦手な文字列や単語、自分の興味の高い文字列や単語を容易に特定できるようにすることを目的とする。

【0014】

【課題を解決するための手段】

図1は本発明の原理説明図である。図1中、2は個人環境頻度記憶装置、3aは頻度記憶手段、4aは言語変換手段、5は読み書き入力部、6は入力部、7は出力部、13aは格納手段である。

30

【0015】

本発明は、前記従来課題を解決するため次のような手段を有する。

【0016】

(1)：個人環境での読み書きの入力を行う読み書き入力部5と、文字列の頻度情報の検索を行う頻度記憶手段3aとを備え、前記頻度記憶手段3aは、前記読み書き入力部5から入力された個人環境での読み書きデータから任意の文字列を抽出し、該抽出した文字列毎の個数をカウントする。このため、個人環境での文字列の頻度情報により、入力文の個人環境への言語変換や差分強調を容易に行うことができる。

40

【0017】

(2)：前記(1)の個人環境頻度記憶装置において、前記頻度記憶手段3aに読み入力検出部と書き入力検出部とを備え、前記頻度記憶手段3aは、前記書き入力検出部で検出した前記任意の文字列の出現頻度の重みを前記読み入力検出部で検出した前記任意の文字列の出現頻度の重みより重くして前記任意の文字列の出現頻度を求める。このため、印象の多い「書く」ということを重要視することができる。

【0018】

(3)：前記(2)の個人環境頻度記憶装置において、前記読み入力検出部は、表示時間の短いものは前記任意の文字列の出現頻度として検出しないようにして前記任意の文字列の出現頻度を求める。このため、個人が読まないで単に表示するものを除くことができる

50

。

【0019】

(4) : 前記(1)~(3)の個人環境頻度記憶装置において、前記頻度記憶手段3aは、前記読み書き入力部5から入力された読み書きデータのうち古いものを削除して前記任意の文字列の出現頻度を求める。このため、最近の個人環境頻度を記憶でき、個人環境の変化に対応することができる。

【0020】

(5) : 前記(1)~(3)の個人環境頻度記憶装置において、前記頻度記憶手段3aは、前記読み書き入力部5から入力された読み書きデータから古いものの重みを軽くして前記任意の文字列の出現頻度を求める。このため、最近の個人環境頻度を重要視でき、個人環境の変化に対応することができる。

10

【0021】

(6) : 前記(1)~(5)のいずれかに記載の個人環境頻度記憶装置2と、入力文を個人環境言語に変換して出力する言語変換手段4aとを備え、前記言語変換手段4aは、入力された文字列を前記個人環境頻度記憶装置2に格納されている出現頻度の高い文字列に変換して出力する。このため、各個人にとって分かりやすい表現にすることができる。

【0022】

(7) : 前記(1)~(5)のいずれかに記載の個人環境頻度記憶装置2と、入力文を個人環境言語に変換して出力する言語変換手段4aとを備え、前記言語変換手段4aは、入力された文字列のうち前記個人環境頻度記憶装置2に格納されている出現頻度の高い文字列を括弧づけで補助表記して出力する。このため、完全に書き換えて、勘違いや文の意味が変わるのを防止することができる。

20

【0023】

(8) : 前記(1)~(5)のいずれかに記載の個人環境頻度記憶装置2と、入力文の文字列の差分を強調して表示する差分強調装置とを備え、前記差分強調装置は、入力された文字列のうち前記個人環境頻度記憶装置2に格納されている文字列の出現頻度がある閾値以下のものを強調表示する。このため、自分の苦手な文字列や単語を手際よく探すことができる。

【0024】

(9) : 前記(1)~(5)のいずれかに記載の個人環境頻度記憶装置2と、入力文の文字列の差分を強調して表示する差分強調装置とを備え、前記差分強調装置は、入力された文字列のうち前記個人環境頻度記憶装置2に格納されている文字列の出現頻度がある閾値以上のものを強調表示する。このため、自分の興味のある大きい文字列や単語を強調表示して、自分の興味のある段落や文を手際よく探すことができる。

30

【0025】

(10) : 個人環境での書きの入力を行う書き入力部と、文字列の頻度情報の検索を行う頻度記憶手段3aを備え、前記頻度記憶手段3aは、前記書き入力部から入力された個人環境での書きデータから任意の文字列を抽出し、該抽出した文字列毎の個数をカウントする個人環境頻度記憶装置2と、入力文を個人環境言語に変換して出力する言語変換手段4aとを備え、前記言語変換手段4aは、入力された文字列のうち前記個人環境頻度記憶装置2に格納されている出現頻度の高い文字列を括弧づけで補助表記して出力する。このため、完全に書き換えて、勘違いや文の意味が変わるのを防止することができる。

40

【0026】

(11) : 個人環境での書きの入力を行う書き入力部と、文字列の頻度情報の検索を行う頻度記憶手段3aを備え、前記頻度記憶手段3aは、前記書き入力部から入力された個人環境での書きデータから任意の文字列を抽出し、該抽出した文字列毎の個数をカウントする個人環境頻度記憶装置2と、入力文の文字列の差分を強調して表示する差分強調装置とを備え、前記差分強調装置は、入力された文字列のうち前記個人環境頻度記憶装置2に格納されている文字列の出現頻度がある閾値以下のものを強調表示する。このため、自分の苦手な文字列や単語を手際よく探すことができる。

50

## 【0027】

(12) : 個人環境での書きの入力を行う書き入力部と、文字列の頻度情報の検索を行う頻度記憶手段3aを備え、前記頻度記憶手段3aは、前記書き入力部から入力された個人環境での書きデータから任意の文字列を抽出し、該抽出した文字列毎の個数をカウントする個人環境頻度記憶装置2と、入力された文字列の差分として出力する対象の単位である抽出単位と入力された文字列の差分を検出するために比較する領域の単位である検出領域を設定し、抽出手段で入力された文字列の現在の前記検出領域以外の領域から全ての前記抽出単位に相当するものを抽出して前記格納手段に格納し、現在の前記検出領域において、前記格納手段に格納されていない前記抽出単位に相当するものを強調表示して現在の検出領域の文書を出力することを、前記検出領域ごとに繰り返す差分強調装置とを備え、前記差分強調装置は、前記抽出手段で強調表示すべきと判断された前記入力された文字列のうち前記個人環境頻度記憶装置2に格納されている文字列の出現頻度がある閾値以下のものを強調表示する。このため、強調表示される箇所が少なく、自分の苦手な表現が初出に現れた箇所を容易に特定できる。

10

## 【0028】

(13) : 個人環境での書きの入力を行う書き入力部と、文字列の頻度情報の検索を行う頻度記憶手段3aを備え、前記頻度記憶手段3aは、前記書き入力部から入力された個人環境での書きデータから任意の文字列を抽出し、該抽出した文字列毎の個数をカウントする個人環境頻度記憶装置2と、入力された文字列の差分として出力する対象の単位である抽出単位と入力された文字列の差分を検出するために比較する領域の単位である検出領域を設定し、抽出手段で、入力された文書データの現在の前記検出領域において、格納手段に格納されていない前記抽出単位に相当するものを強調表示して現在の検出領域の文書を出し、前記強調表示したものを前記格納手段に格納することを、前記検出領域ごとに繰り返す差分強調装置とを備え、前記差分強調装置は、前記抽出手段で強調表示すべきと判断された前記入力された文字列のうち、前記個人環境頻度記憶装置2に格納されている文字列の出現頻度がある閾値以下のものを強調表示する。このため、強調表示される箇所が少なく、自分の苦手な表現が初出に現れた箇所を容易に特定できる。

20

## 【0029】

(14) : 個人環境での書きの入力を行う書き入力部と、文字列の頻度情報の検索を行う頻度記憶手段3aを備え、前記頻度記憶手段3aは、前記書き入力部から入力された個人環境での書きデータから任意の文字列を抽出し、該抽出した文字列毎の個数をカウントする個人環境頻度記憶装置2と、入力文の文字列の差分を強調して表示する差分強調装置とを備え、前記差分強調装置は、入力された文字列のうち前記個人環境頻度記憶装置2に格納されている文字列の出現頻度がある閾値以上のものを強調表示する。このため、自分の興味の高い文字列や単語を強調表示して、自分の興味のある段落や文を手際よく探すことができる。

30

## 【0030】

(15) : 個人環境での書きの入力を行う書き入力部と、文字列の頻度情報の検索を行う頻度記憶手段3aを備え、前記頻度記憶手段3aは、前記書き入力部から入力された個人環境での書きデータから任意の文字列を抽出し、該抽出した文字列毎の個数をカウントする個人環境頻度記憶装置2と、入力文の文字列の差分を強調して表示する差分強調装置とを備え、前記差分強調装置は、入力された文字列のうちユーザ個人での文字列出現確率が一般的なテキスト集合での出現確率より有意に大きいものだけを強調表示する。このため、出現確率が有意に大きくない一般的な名詞等は強調表示されなくなり、その人の興味の高い文字列や単語が強調表示され、自分の興味のある文字列や単語を含む段落や文を手際よく探すことができる。

40

## 【0031】

(16) : 個人環境での書きの入力を行う書き入力部と、文字列の頻度情報の検索を行う頻度記憶手段3aを備え、前記頻度記憶手段3aは、前記書き入力部から入力された個人環境での書きデータから任意の文字列を抽出し、該抽出した文字列毎の個数をカウントす

50



る個人環境頻度記憶装置 2 と、入力文の文字列の差分を強調して表示する差分強調装置とを備え、前記差分強調装置は、入力された文字列のうちユーザ個人での文字列出現確率が一般的なテキスト集合での出現確率のある値倍より大きいものだけを強調表示する。このため、出現確率がある値倍より大きくない一般的な名詞等は比率的に強調表示されなくなり、その人の興味より大きい文字列や単語が強調表示され、自分の興味のある文字列や単語を含む段落や文を手際よく探すことができる。

【0032】

【発明の実施の形態】

§ 1：個人環境言語変換の説明

この発明は、基本的には同一言語内の言語変換を行うものであるが、特に、個人の環境にマッチした言語変換を行うものである。 10

【0033】

つまり、個人環境の読み書きを行うシステムと同一言語内での言語変換装置をくっつけるものである。一般に、個人が文章を作成するためのワードプロセッサや文章を読むためにディスプレイに表示するための個人環境の読み書きシステムにおいてその個人の良く知っている単語を認定し、なるべくその単語を使って表示するようにするものである。

【0034】

(1)：個人環境言語変換装置の説明

図 2 は個人環境言語変換装置の説明図である。図 2 において、個人環境言語変換装置には、頻度記憶部 3、言語変換部 4、読み書き入力部 5、入力部 6、出力部 7 が設けてある。 20

【0035】

頻度記憶部 3 は、個人環境での読み書き入力部 5 から入力された文字列の出現頻度を求めるものである。読み書きシステムが使用されると読み書き入力部 5 から読み書きが入力され頻度記憶部 3 はいつも書き換えられるものである。

【0036】

言語変換部 4 は、頻度記憶部 3 の変形規則によって変換の候補を獲得し、出現頻度等の評価の尺度（評価関数など）によって評価し、最もふさわしい変換の候補を選択するものである。

【0037】

読み書き入力部 5 は、読みシステム、書きシステム又は読み書きが一体になったシステム等の読み書きシステムから読み書きが入力されるものである。入力部 6 は、変換対象文を入力するものである。出力部 7 は、言語変換結果を出力するものである。 30

【0038】

(2)：言語補助変換部を用いる場合の説明

図 3 は言語補助変換部を用いる場合の説明図である。図 3 の個人環境言語変換装置は、図 2 の言語変換部 4 の代わりに言語補助変換部 8 を用いるものである。図 3 において、個人環境言語変換装置には、頻度記憶部 3、読み書き入力部 5、入力部 6、出力部 7、言語補助変換部 8 が設けてある。

【0039】

頻度記憶部 3 は、個人環境での読み書きシステム等から入力された文字列の出現頻度を求めるものである。読み書き入力部 5 は読み書きシステムからの読み書きが入力されるものである。入力部 6 は、変換対象文を入力するものである。出力部 7 は、言語変換結果を出力するものである。 40

【0040】

言語補助変換部 8 は、頻度記憶部 3 の変形規則によって変換の候補を獲得し、出現頻度等の評価の尺度（評価関数など）によって評価し、最もふさわしい変換の候補を選択し括弧づけで補助表記するものである。

【0041】

(3)：頻度記憶部の説明

a) 頻度記憶部の例 (1)

図4は頻度記憶部の説明図(1)である。図4において、頻度記憶部には、読み入力検出部11、読みデータ格納部13、全文検索エンジン14、書き入力検出部21、書きデータ格納部23、全文検索エンジン24が設けてある。

【0042】

読み入力検出部11は、読み書き入力部5から入力された読み入力を検出するものである。読みデータ格納部13は、読み入力検出部11が検出した読み入力を格納するものである。全文検索エンジン14は、読みデータ格納部13に格納された任意の文字列の個数をカウントするものである。書き入力検出部21は、読み書き入力部5から入力された書き入力を検出するものである。書きデータ格納部23は、書き入力検出部21が検出した書き入力を格納するものである。全文検索エンジン24は、書きデータ格納部23に格納された任意の文字列の個数をカウントするものである。なお、全文検索エンジン14と24は、一つの検索エンジンを用いることもできる。

10

【0043】

図5は頻度記憶部の処理フローチャート(1)である。以下、頻度記憶部の処理を図5のステップS1～S5に従って説明する。

【0044】

S1：読み書き入力部5から読み入力されたデータを読み入力検出部11で検出する。具体的には、画面上に何分か以上連続して表示された部分を読み入力されたデータとして認識する。

【0045】

S2：読み入力検出部11で検出された読み入力された文字列をそのまま読みデータ格納部13に格納する。

20

【0046】

S3：読み書き入力部5から書き入力されたデータを書き入力検出部21で検出する。具体的には、キーボード入力などで入力された文字列を書き入力されたデータとして認識する。

【0047】

S4：書き入力検出部21で検出された書き入力された文字列をそのまま書きデータ格納部23に格納する。

【0048】

S5：文字列の存在また個数を高速に検索する、全文検索エンジン14、24を用いて任意の文字列の個数をカウントできるようにしておく。

30

【0049】

b) 頻度記憶部の例(2)

図6は頻度記憶部の説明図(2)である。図6において、頻度記憶部には、読み入力検出部11、形態素解析器12、読みデータ格納部13、単語検索エンジン14、書き入力検出部21、形態素解析器22、書きデータ格納部23、全文検索エンジン24が設けてある。

【0050】

読み入力検出部11は、読み書き入力部5から入力された読み入力を検出するものである。形態素解析器12は、読み入力を単語に分割するものである。読みデータ格納部13は、形態素解析器12で分割した単語を格納するものである。単語検索エンジン14は、任意の単語の出現回数をカウントするものである。書き入力検出部21は、読み書き入力部5から入力された書き入力を検出するものである。形態素解析器22は、書き入力を単語に分割するものである。書きデータ格納部23は、形態素解析器22で分割した単語を格納するものである。単語検索エンジン24は、任意の単語の出現回数をカウントするものである。なお、形態素解析器12と22、読みデータ格納部13と書きデータ格納部23及び全文検索エンジン14と24は、それぞれ一つの形態素解析器、一つの格納部及び一つの検索エンジンを用いることもできる。

40

【0051】

50

図7は頻度記憶部の処理フローチャート(2)である。以下、形態素解析器を用いる頻度記憶部の処理を図7のステップS11~S15に従って説明する。

【0052】

S11: 読み書き入力部5から読み入力されたデータを読み入力検出部11で検出する。具体的には、画面上に何分か以上連続して表示された部分を読み入力されたデータとして認識する。

【0053】

S12: 読み入力検出部11で検出された読み入力された文字列を形態素解析器12で単語に分割し、単語ごとに読みデータ格納部13に格納する。各単語ごとに何回出現したかの回数のデータも同時に格納する。すでに格納してある単語と同じ単語のものを格納する場合は出現回数のデータのみを更新する。

10

【0054】

S13: 読み書き入力部5から書き入力されたデータを書き入力検出部21で検出する。具体的には、キーボード入力などで入力された文字列を書き入力されたデータとして認識する。

【0055】

S14: 書き入力検出部21で検出された書き入力された文字列を形態素解析器22で単語に分割し、単語ごとに書きデータ格納部23に格納する。各単語ごとに何回出現したかの回数のデータも同時に格納する。すでに格納してある単語と同じ単語のものを格納する場合は出現回数のデータのみを更新する。

20

【0056】

S15: 単語検索エンジン14、24では、任意の単語の出現回数をカウントできるようにしておく。

【0057】

(4) 言語変換部の説明

図8は言語変換部の説明図である。図8において、言語変換部には、言語変換処理部31、変換規則部(言語変換の辞書)32、変換用尺度33が設けてある。言語変換処理部31は、変換規則を用いて変形の候補をあげ、変換用尺度により変換の妥当性のチェックをし、最も妥当であると判断されたものに変換するものである。即ち、変形規則を用いて変換用尺度が大きくなるような変換を行う。変形規則部(言語変換の辞書)32は、「罷免する」を「やめさせる」に変形する等の規則である。変換用尺度33は、類似度、長さ、出現頻度(出現確率)等があるが、ここでは頻度記憶部の出現頻度を用いる。

30

【0058】

(5) 個人環境言語変換処理の説明

a) 頻度記憶部と言語変換部を用いる個人環境言語変換の説明

1: ユーザの日々の読み書きの行動から、頻度記憶部のフローチャート(図5、図7参照)の処理で、ユーザの読みデータ、書きデータを頻度記憶部3に記憶する。

【0059】

2: 言語変換部4には言語変換規則が蓄えられているものとする。適用可能な言語変換規則があるとき、その変換をした後の文字列の頻度と、変換前の文字列の頻度を、頻度記憶部3から求めて、変換をした後の文字列の頻度の方が大きい場合、変換を行なう。

40

【0060】

また、変換前の文字列の頻度の方が大きい場合は、変換を行なわない。変換を行ないうる言語変換規則が複数ある場合は、その変換をした後の文字列の頻度がもっとも大きい規則を用いて変換を行なう。

【0061】

このときの文字列の頻度は、読みデータと書きデータの両方を組み合わせたもので、概ね以下のような式で求めなおしたのを用いる。

【0062】

具体的には、読みシステムにおいての各単語  $t$  の出現頻度を  $f_r(t)$ 、書きシス

50

テムにおいての各単語  $t$  の出現頻度を  $f_w(t)$  とするとき、その個人の単語出現頻度分布を

$$\alpha \times f_r(t) + (1 - \alpha) \times f_w(t) \quad (\text{ただし、} 0 \leq \alpha \leq 1)$$

として、この頻度が多くなるように単語を変換する。すなわち、言語変換をする際に用いる尺度として、個人環境の読み書きシステムにおけるその個人の単語出現頻度分布を用いるものである。ここで、 $\alpha$  を設けるのは、「読む」ということは「書く」ということより印象が少ないので、「書く」事への重みを高める（重み付ける）ためである。つまり、 $\alpha$  は 0.5 より小さいものとなる。なお、 $\alpha$  などの定数はユーザが設定変更できるようにしておくものである。

【0063】

例えば、その個人が政治家であるとするようになるべく政治家の専門用語を使った表現に変換するようになる。一般に自分が良く使う表現で書かれた文ほど、その人にとってわかりやすい。この発明の個人環境言語変換装置を用いると、他の分野の専門家の文章も、自分の分野でよく使われる表現に変更することができる。

【0064】

（具体的な例による説明）

同一言語内の言語変換については、すでに出願済であるが、ここでは、個人環境の読み書きシステムと同一言語内での言語変換装置をくっつけたものである。個人環境の読み書きシステムをくっつけることで、その個人がよく知っている単語の頻度分布を容易に取得でき、その情報を言語変換に用いることができる点が重要なところである。

【0065】

例えば、以下の矢印のように変換する。

【0066】

「世界知識を用いた照応解析の研究」 「常識を用いた指示詞の指示先の推定の研究」  
言語変換の辞書（変換規則）に次のものがあるとする。

【0067】

世界知識 = 常識

照応解析 = 指示詞の指示先の推定

ここで、言語処理の分野で記述される「世界知識を用いた照応解析の研究」という文が入力されたとする。

【0068】

（場合1）

ユーザが言語処理の研究者とする。

【0069】

「常識」よりも「世界知識」、「指示詞の指示先の推定」よりも「照応解析」の方がそのユーザの読み書きシステムでの利用頻度（出現頻度）が高いとする。

【0070】

これは、言語処理の研究者のためにそのような専門用語を使う頻度が多いためである。

【0071】

この場合、個人環境言語変換装置は、書き換えずにそのまま「世界知識を用いた照応解析の研究」を出力することになる。このユーザーにとってはこの表現の方が自然なのでこのままの出力を見る方が都合が良い。

【0072】

（場合2）

ユーザーが言語処理の分野をあまり知らない人とする。

【0073】

この場合、より一般的な用語の「常識」や「指示詞の指示先の推定」の方が利用頻度が高くなると思われる。

【0074】

この場合、個人環境言語変換装置は、書き換えを行ない「常識を用いた指示詞の指示先の

10

20

30

40

50

推定の研究」を出力する。

【0075】

このユーザーは、「照応解析」や「世界知識」という用語を知らないのでだいぶんわかりやすい表現を見ることがになり、文章の理解が容易になる。

【0076】

ただし、完全に書き換えたのでは、勘違いをしたり、文の意味が変わる可能性がある。

【0077】

b) 頻度記憶部と言語補助変換部を用いる個人環境言語変換の説明

1 : ユーザの日々の読み書きの行動から、頻度記憶部の処理フローチャート(図5、図7参照)の処理で、ユーザの読みデータ、書きデータを頻度記憶部3に記憶する。

10

【0078】

2 : 言語補助変換部8には言語変換規則が蓄えられているものとする。適用可能な言語変換規則があるとき、その変換をした後の文字列の頻度と、変換前の文字列の頻度を、頻度記憶部から求めて、変換をした後の文字列の頻度の方が大きい場合、補助変換を行なう。

【0079】

また、変換前の文字列の頻度の方が大きい場合は、補助変換を行なわない。変換を行ないうる言語変換規則が複数ある場合は、その変換をした後の文字列の頻度がもっとも大きい規則を用いて補助変換を行なう。

【0080】

このときの文字列の頻度は、読みデータと書きデータの両方を組み合わせたもので、概ね以下のような式で求めなおしたものを用いる。

20

【0081】

具体的には、読みシステムにおける各単語  $t$  の出現頻度を  $f_r(t)$ 、書きシステムにおける各単語  $t$  の出現頻度を  $f_w(t)$  とするとき、その個人の単語出現頻度分布を

$$\alpha \times f_r(t) + (1 - \alpha) \times f_w(t) \quad (\text{ただし、} 0 \leq \alpha \leq 1)$$

として、この頻度が多くなるように単語を変換する。すなわち、言語変換をする際に用いる尺度として、個人環境の読み書きシステムにおけるその個人の単語出現頻度分布を用いるものである。ここで、 $\alpha$  を設けるのは、「読む」ということは「書く」ということより印象が少ないので、「書く」事への重みを高める(重み付ける)ためである。つまり、 $\alpha$  は 0.5 より小さいものとなる。なお、 $\alpha$  などの定数はユーザが設定変更できるようにしておくものである。

30

【0082】

ところで、補助変換とは、文字列を変換してしまうのではなく、変換先の文字列を括弧づけで補助表記することを意味する。

【0083】

(具体的な例による説明)

「世界知識(常識)を用いた照応解析(指示詞の指示先の推定)の研究」

のように、完全に書き換えてしまうのではなく、括弧づけで補足的な表示をする。なお、本文中に括弧が使われている場合は、それと区別するため異なる括弧を使用することもできる。

40

【0084】

このときも、専門の研究者など、「世界知識」「照応解析」など、用語をよく知っている人にはこの補助変換(表示)をするとむしろ不便であるので、ユーザの用語の使用頻度によって出すか出さないかなどを判断した方がよい。

【0085】

この括弧づけで判断する(補助変換)方法は以下で説明する。

【0086】

・前記のように、その個人の単語出現頻度分布を  $\alpha \times f_r(t) + (1 - \alpha) \times f_w(t)$

50

$w$  (  $t$  ) ( ただし、  $0 < w < 1$  ) として、この頻度が多くなるような書き換え候補の語を括弧付けで付ける。即ち、書き換える語の使用頻度が大きくなる方を括弧付けで付ける。

【 0 0 8 7 】

・前記のように、その個人の単語出現頻度分布 [  $w \times f_r$  (  $t$  ) + (  $1 - w$  )  $\times f_w$  (  $t$  ) ( ただし、  $0 < w < 1$  ) ] として、この頻度が減らないような書き換え候補の語で、かつ、もとの語の頻度がある閾値よりも小さい語を括弧付けで付ける。即ち、個人の使用頻度が多ければ括弧付けは行わないが、頻度が 0 とか少ない場合には括弧付けで付ける。

【 0 0 8 8 】

なお、読み書きシステムには、読みシステム、書きシステム又は読み書きが一体になったシステムがある。読みシステムにはメーラ、インターネット・エクスプローラ、読むために開いた(表示した)ワード文章(文章を作成システムの種類)等の文章を読むためのシステムである。書きシステムには文字を入力して文章を作成するワード文章等の文章を書くためのシステムである。また、読みシステムにおいては、ディスプレイ等に表示される文章の量が多くなるので、表示時間の短いものは除くようにすることもできる。

【 0 0 8 9 】

さらに、読みシステムにおいて、頻度記憶部に格納する単語の重み付けを変えることもできる。例えば、文章を作成システムであるワード文章等を読む場合は丁寧に読むものと考えられるので、インターネット等で画面を見る場合と比べ重みを高くすることができる。

【 0 0 9 0 】

また、頻度記憶部に格納されている単語は、古いものを除くようにすることができる。例えば、個人の趣味が変わるとか、ある分野の専門家になる等で個人環境も変化する場合があるので、古いものは削除するか重み付けを低くするものである。

【 0 0 9 1 】

§ 2 : 個人環境差分強調装置の説明

図 9 は個人環境差分強調装置の説明図である。図 9 において、個人環境差分強調装置には、頻度記憶部 3、読み書き入力部 5、入力部 6、出力部 7、差分強調装置 9 が設けてある。

【 0 0 9 2 】

頻度記憶部 3 は、個人環境での読み書きシステム等から入力された文字列の出現頻度を求めるものである。読み書きシステムが使用されると頻度記憶部 3 はいつも書き換えられるものである。

【 0 0 9 3 】

読み書き入力部 5 は、読み書きシステムから読み書きが入力されるものである。入力部 6 は、変換対象文を入力するものである。出力部 7 は、言語変換結果を出力するものである。

【 0 0 9 4 】

差分強調装置 9 は、入力部 6 から入力された文章の文字列と頻度記憶部 3 の一定頻度の文字列等との差分から、入力部 6 から入力された文章の文字列の強調表示を行うものである。

【 0 0 9 5 】

( 1 ) : 個人環境差分強調装置の動作説明 ( 1 )

1 : ユーザの日々の読み書きの行動から、頻度記憶部 3 のフローチャートの処理 ( 図 5、図 7 参照 ) で、ユーザの読みデータ、書きデータを記憶する。

【 0 0 9 6 】

2 : 差分強調装置 9 で、入力された文章のうち、ユーザ個人の単語出現頻度がある閾値以下のものだけを強調表示する。

【 0 0 9 7 】

( 例による説明 )

例えば、次のようなテキストが入力部 6 から入力されたとする。

10

20

30

40

50

## 【0098】

『自然言語では、動詞を省略するということがある。この省略された動詞を復元することは、対話システムや高品質の機械翻訳システムの実現には不可欠なことである。そこで本研究では、この省略された動詞を表層の表現（手がかり語）と用例から補完することを行なう。解析のための規則を作成する際、動詞の省略現象を補完する動詞がテキスト内にあるかいないかなどで分類した。小説を対象にして実験を行なったところ、テストサンプルで再現率84%、適合率82%の精度で解析できた。このことは本手法が有効であることを示している。テキスト内に補完すべき動詞がある場合は非常に精度が良かった。それに比べ、テキスト内に補完すべき動詞がない場合はあまり良くなかった。しかし、テキスト内に補完すべき動詞がない場合の問題の難しさから考えると、少しでも解析できるだけでも価値がある。また、コーパスが多くなり、計算機の性能もあがり大規模なコーパスが利用できるようになった際には、本稿で提案した用例を利用する手法は重要になるだろう。』

10

ユーザは、言語学の専門の人とする。その場合、工学的な表現「テストサンプル」「再現率」「適合率」の使用頻度が極端に小さいとする。

## 【0099】

その場合、次のように強調表示される。即ち、強調表示箇所は《 》（二重山括弧）で囲って表示している。

## 【0100】

『自然言語では、動詞を省略するということがある。この省略された動詞を復元することは、対話システムや高品質の機械翻訳システムの実現には不可欠なことである。そこで本研究では、この省略された動詞を表層の表現（手がかり語）と用例から補完することを行なう。解析のための規則を作成する際、動詞の省略現象を補完する動詞がテキスト内にあるかいないかなどで分類した。小説を対象にして実験を行なったところ、《テストサンプル》で《再現率》84%、《適合率》82%の精度で解析できた。このことは本手法が有効であることを示している。テキスト内に補完すべき動詞がある場合は非常に精度が良かった。それに比べ、テキスト内に補完すべき動詞がない場合はあまり良くなかった。しかし、テキスト内に補完すべき動詞がない場合の問題の難しさから考えると、少しでも解析できるだけでも価値がある。また、コーパスが多くなり、計算機の性能もあがり大規模なコーパスが利用できるようになった際には、本稿で提案した用例を利用する手法は重要になるだろう。』

20

30

このようにして、自分の苦手な語を手際よく探すことができ、それをなんらかの辞書などで調べるとよいとすぐわかり便利である。

## 【0101】

(2) : 個人環境差分強調装置の動作説明(2)

1 : ユーザの日々の読み書きの行動から、頻度記憶部3のフローチャートの処理(図5、図7参照)で、ユーザの読みデータ、書きデータを記憶する。

## 【0102】

2 : 差分強調装置9で、入力部5から入力された文章において、その文章のうちで初めての文字列を強調表示すべきと判断され、かつ、ユーザ個人の単語出現頻度がある閾値以下のものだけを強調表示する。

40

## 【0103】

なお、差分強調装置で、入力された文章のうちで初めての文字列を強調表示すべきと判断する手法は、次の手法1、2がある(特願2002-290946参照)。

## 【0104】

手法1

1 入力部等により、予め抽出の単位(抽出単位)、検出領域の単位を定める。抽出単位とは、差分として出力する対象の単位である。抽出単位には、「単語」「漢字」「名詞句」などが考えられる。検出領域の単位とは、差分を検出するために比較する領域の単位のことである。検出領域の単位には、「文字」「単語」「文」「箇条書の項目」「段落」

50

「特許の請求項」などが考えられる。

【0105】

2 差分強調装置は、すべての入力データを記憶手段（差分強調装置内の）に記憶させる。

【0106】

3 差分強調装置は、入力されたデータを左から調べて左の検出領域から 1 で定めた検出領域ごとに以下の処理 4 と処理 5 を繰り返す。

【0107】

4 差分強調装置は、現在の検出領域以外の領域すべてから、すべての抽出単位に相当するもの（例えば単語）を抽出し、それを抽出物記憶手段（差分強調装置内の）に格納する。

10

【0108】

5 差分強調装置は、現在の検出領域において、抽出物記憶手段（差分強調装置内の）に格納されていない抽出単位に相当するもの（例えば単語）を強調表示して現在の検出領域の文章を出力する。

【0109】

手法 2

1 入力部 1 等により、予め抽出の単位（抽出単位）、検出領域の単位を定める。抽出単位とは、差分として出力する対象の単位である。抽出単位には、「単語」「漢字」「名詞句」などが考えられる。検出領域の単位とは、差分を検出するために比較する領域の単位のことである。検出領域の単位には、「文字」「単語」「文」「箇条書の項目」「段落」「特許の請求項」などが考えられる。

20

【0110】

2 入力部から前記 1 で定めた検出領域ごとに入力データが入力され、差分強調装置は、以下の処理 3 と処理 4 を繰り返す。

【0111】

3 差分強調装置は、現在の検出領域において、抽出物記憶装置（差分強調装置内の）に格納されていない抽出単位に相当するもの（例えば単語）を強調表示して現在の検出領域の文章を出力する。ただし、抽出物記憶装置（差分強調装置内の）は最初は空である。

【0112】

4 前記処理 1 で強調表示した表現を抽出物記憶装置に格納する。

30

【0113】

（例による説明）

ここでは手法 2 に適用した例を説明する。差分強調装置は、上記手法 2 により、入力された文章のうち、以下の二重山括弧で囲った部分を強調表示すべきと判断する。なお、抽出の単位、検出領域の単位ともに単語である。

【0114】

『《本研究の目的は、日本語》の《受け身文》、《使役》文《を能動》文《に変換する際に《変更され》《るべき格助詞》を《機械学習》を《用いて自動》変換する《ことである。》日本語の受け身文、使役文の《例》を《図 1 と》図《 2 》に《あげる》。図 1 の文の日本語の《接尾辞「れ《た》」は《受動態》を《示す助動詞》で《あり》、《この》文は受け身文である。図 2 の文の日本語の接尾辞「《せ》た」は使役を示す助動詞であり、この文は使役文である。《これら》の文に《対》《応》する能動文を図《 3 》に示す。図 1 の文《が》能動文に変換さ《れるとき》は、《（ i ）》格助詞「に」は格助詞「が」に《（ i i ）》格助詞「が」は格助詞「を」に変換される。図 2 の文が能動文に変換されるときは、（ i ）格助詞「が」の《部分》「《彼》が」の《文節》が《消去》され、（ i i ）格助詞「に」が格助詞「が」に変換され、《（ i i i ）》格助詞「を」は変換され《ず》に《そのまま残る》。本研究では、これらの格助詞の変換《（ 》例《：》格《助》《詞》「に」の格助詞「が」《へ》の変換《 》と、《不要》部分の消去（例：「彼が」の消去）を、研究の《対象》とする。（《以降》、《本稿

40

50



》では《便宜上》「彼が」《など》の消去の部分《も》格助詞の変換と《呼ぶ》.)  
 受け身文, 使役文の能動文への変換は, 文《生成》, 《言い換え》, 文の《平易化/言語  
 》《運用支援》, 《自然》言語文《から》の《知識獲得や情報抽出》, 《質問応答システ  
 ム》と《多く》の研究《分野》で《役に立つもの》である. 《例えば》, 質問応答システ  
 ムでは, 質問文が《能》《動》文で《答え》が《受動》文で《書か》れて《いる場合》,  
 質問文と答えを《含む》文で, 文の《構造》が《異なるため》に, 質問の答えを《取り出  
 す》のが《困難な》場合がある. この《よう》な《問》《題》も受け身文, 使役文の能動  
 文への変換が《できる》ように《なる》と《解決》する《のであ》る. このように受け身  
 文, 使役文の能動文への変換は, 自然言語《処理》で《重要》なものである. 』

ここでユーザは言語学にも言語処理も詳しくない人とする. そうすると、言語学、言語処  
 理の専門用語は出現確率が低く、「受け身」「使役」「能動」「態」「言い換え」が頻度  
 が0であったとする。また、上述の閾値も0であったとする。そうすると、個人環境差分  
 強調装置の出力として、これらの語と上記で強調されている部分の重なったところのみが  
 次のように強調表示される。

【0115】

『本研究の目的は、日本語の《受け身》文, 《使役》文を《能動》文に変換する際に変更  
 されるべき格助詞を機械学習を用いて自動変換することである. 日本語の受け身文, 使役  
 文の例を図1 と図2 にあげる. 図1 の文の日本語の接尾辞「れた」は《受動態》を  
 示す助動詞であり, この文は受け身文である. 図2 の文の日本語の接尾辞「せた」は使  
 役を示す助動詞であり, この文は使役文である. これらの文に対応する能動文を図3 に  
 示す. 図1 の文が能動文に変換される時は, (i) 格助詞「に」は格助詞「が」に  
 (i i) 格助詞「が」は格助詞「を」に変換される. 図2 の文が能動文に変換される  
 ときは, (i) 格助詞「が」の部分「彼が」の文節が消去され, (i i) 格助詞「に」が  
 格助詞「が」に変換され, (i i i) 格助詞「を」は変換されずにそのまま残る. 本研  
 究では, これらの格助詞の変換(例: 格助詞「に」の格助詞「が」への変換) と, 不  
 要部分の消去(例: 「彼が」の消去) を, 研究の対象とする. (以降, 本稿では便  
 宜上「彼が」などの消去の部分も格助詞の変換と呼ぶ.)

受け身文, 使役文の能動文への変換は, 文生成, 《言い換え》, 文の平易化/言語運用支  
 援, 自然言語文からの知識獲得や情報抽出, 質問応答システムと多くの研究分野で役に立  
 つものである. 例えば, 質問応答システムでは, 質問文が能動文で答えが受動文で書かれ  
 ている場合, 質問文と答えを含む文で, 文の構造が異なるために, 質問の答えを取り出す  
 のが困難な場合がある. このような問題も受け身文, 使役文の能動文への変換ができるよ  
 うになると解決するのである. このように受け身文, 使役文の能動文への変換は, 自然言  
 語処理で重要なものである. 』

上記のようになると、これは先にあげた単純に前回出願(手法2だけ)を使ったときに比  
 べて強調表示される箇所が少なく、見やすい。また、単純に今回(手法2を用いない)だ  
 けだと、すべての「受け身」「使役」「能動」「態」「言い換え」が強調表示されるが、  
 前回出願(手法1又は2)と併用することで、初出の「受け身」「使役」「能動」「態」  
 「言い換え」だけが強調表示されることになる。

【0116】

これにより、自分の苦手な表現が初出にあらわれた箇所を容易に特定でき便利である。

【0117】

さらに、原理的に、この個人環境差分強調装置が行なっていることを考えてみると、閾値  
 0の場合は、このテキストの最初に、その個人が読み書きしてきた全テキストをくっつけ  
 て前回出願の手法2を行なったことを意味する。即ち、もし、その人が、読み書きシステ  
 ムでした文字を見たり書いたりすることがないとすると、ここで強調表示されるものは、  
 その人が全生涯通じて初めて見た単語を意味する。手法2を生涯にまで拡張したものと見  
 ることができる。

【0118】

(3) : 削除用単語記憶部を用いる個人環境差分強調装置の説明(個人環境差分強調装置

10

20

30

40

50

の動作説明(3))

図10は削除用単語記憶部を用いる個人環境差分強調装置の説明図である。図10において、個人環境差分強調装置には、頻度記憶部3、読み書き入力部5、入力部6、出力部7、差分強調装置9、削除用単語記憶部10が設けてある。

【0119】

図10の個人環境差分強調装置は、図9の個人環境差分強調装置に予め指定している単語(名詞以外の単語や一般的な名詞)は強調表示しないようにする削除用単語記憶部10を追加したものである。

【0120】

(動作説明)

1 : ユーザの日々の読み書きの行動から、頻度記憶部3のフローチャートの処理(図5、図7参照)で、ユーザの読みデータ、書きデータを記憶する。

【0121】

2 : 差分強調装置9で、入力された文章のうち、ユーザ個人の単語出現頻度がある閾値以上のものだけを強調表示する。ただし、あらかじめ指定している単語(名詞以外の単語や一般的な名詞)は強調表示しないようにする。

【0122】

(例による説明)

例えば、次のようなテキストが入力部6から入力されたとする。

【0123】

『自然言語では、動詞を省略するということがある。この省略された動詞を復元することは、対話システムや高品質の機械翻訳システムの実現には不可欠なことである。そこで本研究では、この省略された動詞を表層の表現(手がかり語)と用例から補完することを行なう。解析のための規則を作成する際、動詞の省略現象を補完する動詞がテキスト内にあるかいないかなどで分類した。小説を対象にして実験を行なったところ、テストサンプルで再現率84%、適合率82%の精度で解析できた。このことは本手法が有効であることを示している。テキスト内に補完すべき動詞がある場合は非常に精度が良かった。それに比べ、テキスト内に補完すべき動詞がない場合はあまり良くなかった。しかし、テキスト内に補完すべき動詞がない場合の問題の難しさから考えると、少しでも解析できるだけでも価値がある。また、コーパスが多くなり、計算機の性能もあがり大規模なコーパスが利用できるようになった際には、本稿で提案した用例を利用する手法は重要になるだろう。』

ここで、ユーザは言語学の専門の人とする。その場合、言語学的な表現、「言語」「動詞」「省略」の使用頻度が極端に高いとする。ただし、名詞以外の語や「こと」「内」などの一般的な名詞はあらかじめ強調表示しない単語として削除用単語記憶部10に登録しておく。その場合、次のように強調表示される。

【0124】

『自然《言語》では、《動詞》を《省略》するということがある。この《省略》された《動詞》を復元することは、対話システムや高品質の機械翻訳システムの実現には不可欠なことである。そこで本研究では、この《省略》された《動詞》を表層の表現(手がかり語)と用例から補完することを行なう。解析のための規則を作成する際、《動詞》の省略現象を補完する《動詞》がテキスト内にあるかいないかなどで分類した。小説を対象にして実験を行なったところ、テストサンプルで再現率84%、適合率82%の精度で解析できた。このことは本手法が有効であることを示している。テキスト内に補完すべき《動詞》がある場合は非常に精度が良かった。それに比べ、テキスト内に補完すべき《動詞》がない場合はあまり良くなかった。しかし、テキスト内に補完すべき《動詞》がない場合の問題の難しさから考えると、少しでも解析できるだけでも価値がある。また、コーパスが多くなり、計算機の性能もあがり大規模なコーパスが利用できるようになった際には、本稿で提案した用例を利用する手法は重要になるだろう。』

強調表示箇所は《 》で囲っている。このように、自分のよく使う単語、つまり、その

10

20

30

40

50

人の興味の大きい単語が強調表示されることになる。自分の興味のある単語を含む段落や文を手際よく探すことができ、それを中心に読むなどのことができるので便利である。

【0125】

(4) : 個人環境差分強調装置の動作説明(4)

1 : ユーザの日々の読み書きの行動から、頻度記憶部3のフローチャートの処理(図5、図7参照)で、ユーザの読みデータ、書きデータを記憶する。

【0126】

2 : 差分強調装置9で、入力された文章のうち、ユーザ個人での単語出現確率が、一般的なテキスト集合での出現確率よりも有意に大きいものだけを強調表示する。

【0127】

単語Aの出現確率は、単語Aの出現回数を全単語の総出現回数で割ったものである。一般的なテキスト集合をもっていればこれは実現できる。有意に大きいかどうかの判定には一般に統計的検定が用いられる。

【0128】

統計的検定とは、検定すべき仮説と客観的証拠としての標本データとを比較して、その間に矛盾がなければ仮説を受け入れ、矛盾が生じた場合には仮説を棄却するものである(参考文献の一例として、心理教育統計学 培風館 肥田他、参照)。

【0129】

(例による説明)

例えば、次のようなテキストが入力部6から入力されたとする。

【0130】

『自然言語では、動詞を省略するということがある。この省略された動詞を復元することは、対話システムや高品質の機械翻訳システムの実現には不可欠なことである。そこで本研究では、この省略された動詞を表層の表現(手がかり語)と用例から補完することを行なう。解析のための規則を作成する際、動詞の省略現象を補完する動詞がテキスト内にあるかいないかなどで分類した。小説を対象にして実験を行なったところ、テストサンプルで再現率84%、適合率82%の精度で解析できた。このことは本手法が有効であることを示している。テキスト内に補完すべき動詞がある場合は非常に精度が良かった。それに比べ、テキスト内に補完すべき動詞がない場合はあまり良くなかった。しかし、テキスト内に補完すべき動詞がない場合の問題の難しさから考えると、少しでも解析できるだけでも価値がある。また、コーパスが多くなり、計算機の性能もあがり大規模なコーパスが利用できるようになった際には、本稿で提案した用例を利用する手法は重要になるだろう。』

ユーザは、言語学の専門の人とする。その場合、言語学的な表現、その個人環境での「言語」「動詞」「省略」の使用頻度が、一般的テキストでの出現頻度に比べて有意に高いとする。

【0131】

この方法の場合、有意差検定の利用(統計的検定)により、自動的に「こと」「内」などの一般的な名詞は有意に高いとはでないので、強調表示されないことになる。その場合、次のような強調表示となる。

【0132】

『自然《言語》では、《動詞》を《省略》するということがある。この《省略》された《動詞》を復元することは、対話システムや高品質の機械翻訳システムの実現には不可欠なことである。そこで本研究では、この《省略》された《動詞》を表層の表現(手がかり語)と用例から補完することを行なう。解析のための規則を作成する際、《動詞》の省略現象を補完する《動詞》がテキスト内にあるかいないかなどで分類した。小説を対象にして実験を行なったところ、テストサンプルで再現率84%、適合率82%の精度で解析できた。このことは本手法が有効であることを示している。テキスト内に補完すべき《動詞》がある場合は非常に精度が良かった。それに比べ、テキスト内に補完すべき《動詞》がない場合はあまり良くなかった。しかし、テキスト内に補完すべき《動詞》がない場

10

20

30

40

50

合の問題の難しさから考えると、少しでも解析できるだけでも価値がある。また、コーパスが多くなり、計算機の性能もあがり大規模なコーパスが利用できるようになった際には、本稿で提案した用例を利用する手法は重要になるだろう。』

強調表示箇所は《 》で囲っている。このように、自分のよく使う単語、つまり、その人の興味の高い単語が強調表示されることになる。自分の興味のある単語を含む段落や文を手際よく探すことができ、それを中心に読むなどのことができ、便利である。

【0133】

(5) : 個人環境差分強調装置の動作説明(5)

1 : ユーザの日々の読み書きの行動から、頻度記憶部3のフローチャートの処理(図5、図7参照)で、ユーザの読みデータ、書きデータを記憶する。

10

【0134】

2 : 差分強調装置9で、入力された文章のうち、ユーザ個人での単語出現確率が、一般的なテキスト集合での出現確率のある値倍よりも大きいものだけを強調表示する。

【0135】

単語Aの出現確率は、単語Aの出現回数を全単語の総出現回数で割ったものである。一般的なテキスト集合をもっていればこれは実現できる。この事例では有意かどうかの判定をしないので統計的検定などの難しい方法を使わなくて済む。

【0136】

単に、ユーザ個人での単語出現確率と一般的なテキスト集合での出現確率を計算し、その割り算が、つまり、ユーザ個人での単語出現確率を一般的なテキスト集合での出現確率で割った値があらかじめ定めたある値よりも大きいものだけを強調表示する。

20

【0137】

(例による説明)

例えば、次のようなテキストが入力部6から入力されたとする。

【0138】

『自然言語では、動詞を省略するということがある。この省略された動詞を復元することは、対話システムや高品質の機械翻訳システムの実現には不可欠なことである。そこで本研究では、この省略された動詞を表層の表現(手がかり語)と用例から補完することを行なう。解析のための規則を作成する際、動詞の省略現象を補完する動詞がテキスト内にあるか否かなどで分類した。小説を対象にして実験を行なったところ、テストサンプルで再現率84%、適合率82%の精度で解析できた。このことは本手法が有効であることを示している。テキスト内に補完すべき動詞がある場合は非常に精度が良かった。それに比べ、テキスト内に補完すべき動詞がない場合はあまり良くなかった。しかし、テキスト内に補完すべき動詞がない場合の問題の難しさから考えると、少しでも解析できるだけでも価値がある。また、コーパスが多くなり、計算機の性能もあがり大規模なコーパスが利用できるようになった際には、本稿で提案した用例を利用する手法は重要になるだろう。』

30

ここでユーザは言語学の専門の人とする。その場合、言語学的な表現、その個人環境での「言語」「動詞」「省略」の使用頻度が、一般的テキストでの出現頻度のあらかじめ定めた値倍したものよりも高いとする。この方法の場合も、自動的に「こと」「内」などの一般的な名詞は比率的にはそれほど出現しないので強調表示されないことになる。その場合、次のような強調表示となる。

40

【0139】

『自然《言語》では、《動詞》を《省略》するということがある。この《省略》された《動詞》を復元することは、対話システムや高品質の機械翻訳システムの実現には不可欠なことである。そこで本研究では、この《省略》された《動詞》を表層の表現(手がかり語)と用例から補完することを行なう。解析のための規則を作成する際、《動詞》の省略現象を補完する《動詞》がテキスト内にあるか否かなどで分類した。小説を対象にして実験を行なったところ、テストサンプルで再現率84%、適合率82%の精度で解析できた。このことは本手法が有効であることを示している。テキスト内に補完すべき《

50

動詞》がある場合は非常に精度が良かった。それに比べ、テキスト内に補完すべき《動詞》がない場合はあまり良くなかった。しかし、テキスト内に補完すべき《動詞》がない場合の問題の難しさから考えると、少しでも解析できるだけでも価値がある。また、コーパスが多くなり、計算機の性能もあがり大規模なコーパスが利用できるようになった際には、本稿で提案した用例を利用する手法は重要になるだろう。』

強調表示箇所は《 》で囲っている。このように、自分のよく使う単語、つまり、その人の興味の高い単語が強調表示されることになる。自分の興味のある単語を含む段落や文を手際よく探すことができ、それを中心に読むなどのことができ、便利である。

【0140】

なお、全文検索技術の参考資料の一例として、日本語全文検索システムの構築と活用、馬場肇 S O F T B A N K がある。 10

【0141】

また、前記実施の形態では、強調表示として、2重山括弧で囲む説明をしたが、下線、色分け、背景の変更、字体の変更、点滅等他の強調表示を行うこともできる。

【0142】

さらに、前記実施の形態では、頻度記憶部3で、個人環境での読み書きシステムから入力された文字列の出現頻度を求める説明をしたが、読みシステムのみ、又は、書きシステムのみから入力された文字列の出現頻度を求めるようにしてもよい。

【0143】

### §3：プログラムインストールの説明 20

個人環境頻度記憶装置2、頻度記憶手段3a、頻度記憶部3、言語変換手段4a、言語変換部4、読み書き入力部5、入力部6、出力部7、言語補助変換部8、差分強調装置9、削除用単語記憶部10、格納手段13a等は、プログラムで構成でき、主制御部(CPU)が実行するものであり、主記憶に格納されているものである。このプログラムは、一般的な、コンピュータで処理されるものである。このコンピュータは、主制御部、主記憶、ファイル装置、表示装置、キーボード等の入力手段である入力装置などのハードウェアで構成されている。このコンピュータに、本発明のプログラムをインストールする。このインストールは、フロッピィ、光磁気ディスク等の可搬型の記録(記憶)媒体に、これらのプログラムを記憶させておき、コンピュータが備えている記録媒体に対して、アクセスするためのドライブ装置を介して、或いは、LAN等のネットワークを介して、コンピュータに設けられたファイル装置にインストールされる。そして、このファイル装置から処理に必要なプログラムステップを主記憶に読み出し、主制御部が実行するものである。 30

【0144】

### 【発明の効果】

以上説明したように、本発明によれば、次のような効果がある。

【0145】

(1)：頻度記憶手段で、読み書き入力部から入力された個人環境での読み書きデータから任意の文字列を抽出し、該抽出した文字列毎の個数をカウントするため、個人環境での文字列の頻度情報により、入力文の個人環境への言語変換や差分強調を容易に行うことができる。 40

【0146】

(2)：頻度記憶手段で、書き入力検出部で検出した任意の文字列の出現頻度の重みを読み入力検出部で検出した前記任意の文字列の出現頻度の重みより重くして前記任意の文字列の出現頻度を求めるため、印象のより多い「書く」ということを重要視することができる。

【0147】

(3)：読み入力検出部で、表示時間の短いものは任意の文字列の出現頻度として検出しないようにして前記任意の文字列の出現頻度を求めるため、個人が読まないで単に表示するものを除くことができる。

【0148】

(4) : 頻度記憶手段で、読み書き入力部から入力された読み書きデータのうち古いものを削除して前記任意の文字列の出現頻度を求めるため、最近の個人環境頻度を記憶でき、個人環境の変化に対応することができる。

【0149】

(5) : 頻度記憶手段で、読み書き入力部から入力された読み書きデータから古いものの重みを軽くして前記任意の文字列の出現頻度を求めるため、最近の個人環境頻度を重要視でき、個人環境の変化に対応することができる。

【0150】

(6) : 言語変換手段で、入力された文字列を個人環境頻度記憶装置に格納されている出現頻度の高い文字列に変換して出力するため、各個人にとって分かりやすい表現にすることができる。 10

【0151】

(7) : 言語変換手段で、入力された文字列のうち個人環境頻度記憶装置に格納されている出現頻度の高い文字列を括弧づけで補助表記して出力するため、完全に書き換えて、勘違いや文の意味が変わるのを防止することができる。

【0152】

(8) : 差分強調装置で、入力された文字列のうち個人環境頻度記憶装置に格納されている文字列の出現頻度がある閾値以下のものを強調表示するため、自分の苦手な文字列や単語を手際よく探すことができる。

【0153】

(9) : 差分強調装置で、入力された文字列のうち個人環境頻度記憶装置に格納されている文字列の出現頻度がある閾値以上のものを強調表示するため、自分の興味のある文字列や単語を強調表示して、自分の興味のある段落や文を手際よく探すことができる。 20

【0154】

(10) : 個人環境頻度記憶装置で、個人環境での書きの入力を行う書き入力部から入力された個人環境での書きデータから任意の文字列を抽出し、該抽出した文字列毎の個数をカウントし、言語変換手段で、入力された文字列のうち前記個人環境頻度記憶装置に格納されている出現頻度の高い文字列を括弧づけで補助表記して出力するため、完全に書き換えて、勘違いや文の意味が変わるのを防止することができる。

【0155】

(11) : 個人環境頻度記憶装置で、個人環境での書きの入力を行う書き入力部から入力された個人環境での書きデータから任意の文字列を抽出し、該抽出した文字列毎の個数をカウントし、差分強調装置で、入力された文字列のうち前記個人環境頻度記憶装置に格納されている文字列の出現頻度がある閾値以下のものを強調表示するため、自分の苦手な文字列や単語を手際よく探すことができる。 30

【0156】

(12) : 入力された文字列の差分として出力する対象の単位である抽出単位と入力された文字列の差分を検出するために比較する領域の単位である検出領域を設定し、抽出手段で入力された文字列の現在の前記検出領域以外の領域から全ての前記抽出単位に相当するものを抽出して前記格納手段に格納し、現在の前記検出領域において、前記格納手段に格納されていない前記抽出単位に相当するものを強調表示して現在の検出領域の文書出力することを、前記検出領域ごとに繰り返す差分強調装置で、前記抽出手段で強調表示すべきと判断された前記入力された文字列のうち個人環境頻度記憶装置に格納されている文字列の出現頻度がある閾値以下のものを強調表示するため、強調表示される箇所が少なく、自分の苦手な表現が初出に現れた箇所を容易に特定できる。 40

【0157】

(13) : 入力された文字列の差分として出力する対象の単位である抽出単位と入力された文字列の差分を検出するために比較する領域の単位である検出領域を設定し、抽出手段で、入力された文書データの現在の前記検出領域において、格納手段に格納されていない前記抽出単位に相当するものを強調表示して現在の検出領域の文書出力し、前記強調表 50

示したものを前記格納手段に格納することを、前記検出領域ごとに繰り返す差分強調装置で、前記抽出手段で強調表示すべきと判断された前記入力された文字列のうち、個人環境頻度記憶装置に格納されている文字列の出現頻度がある閾値以下のものを強調表示するため、強調表示される箇所が少なく、自分の苦手な表現が初出に現れた箇所を容易に特定できる。

【0158】

(14)：個人環境頻度記憶装置で、個人環境での書きの入力を行う書き入力部から入力された個人環境での書きデータから任意の文字列を抽出し、該抽出した文字列毎の個数をカウントし、差分強調装置で、入力された文字列のうち個人環境頻度記憶装置に格納されている文字列の出現頻度がある閾値以上のものを強調表示するため、自分の興味の高い文字列や単語を強調表示して、自分の興味のある段落や文を手際よく探すことができる。

10

【0159】

(15)：差分強調装置で、入力された文字列のうちユーザ個人での文字列出現確率が一般的なテキスト集合での出現確率より有意に大きいものだけを強調表示するため、出現確率が有意に大きくない一般的な名詞等は強調表示されなくなり、その人の興味の高い文字列や単語が強調表示され、自分の興味のある文字列や単語を含む段落や文を手際よく探すことができる。

【0160】

(16)：差分強調装置で、入力された文字列のうちユーザ個人での文字列出現確率が一般的なテキスト集合での出現確率のある値倍より大きいものだけを強調表示するため、出現確率がある値倍より大きくない一般的な名詞等は比率的に強調表示されなくなり、その人の興味の高い文字列や単語が強調表示され、自分の興味のある文字列や単語を含む段落や文を手際よく探すことができる。

20

【0161】

(17)：個人環境での読み書きの入力を行う読み書き入力手段と、該読み書き入力手段から入力された個人環境での読み書きデータから任意の文字列を抽出し、該抽出した文字列毎の個数をカウントする頻度記憶手段として、コンピュータを機能させるためのプログラム又はプログラムを記録した記録媒体とするため、このプログラムをコンピュータにインストールすることで、個人環境での文字列の頻度情報を得ることができる個人環境頻度記憶装置を容易に提供することができる。

30

【0162】

(18)：個人環境での読み書きの入力を行う読み書き入力手段と、該読み書き入力手段から入力された個人環境での読み書きデータから任意の文字列を抽出し、該抽出した文字列毎の個数をカウントして出現頻度を求める頻度記憶手段と、入力された文字列を前記頻度記憶手段に格納されている出現頻度の高い文字列に変換して出力する言語変換手段として、コンピュータを機能させるためのプログラム又はプログラムを記録した記録媒体とするため、このプログラムをコンピュータにインストールすることで、各個人にとって分かりやすい表現にすることができる個人環境言語変換装置を容易に提供することができる。

【0163】

(19)：個人環境での読み書きの入力を行う読み書き入力手段と、該読み書き入力手段から入力された個人環境での読み書きデータから任意の文字列を抽出し、該抽出した文字列毎の個数をカウントして出現頻度を求める頻度記憶手段と、入力された文字列のうち前記頻度記憶手段に格納されている出現頻度の高い文字列を括弧づけで補助表記して出力する言語変換手段として、コンピュータを機能させるためのプログラム又はプログラムを記録した記録媒体とするため、このプログラムをコンピュータにインストールすることで、完全に書き換えて、勘違いや文の意味が変わるのを防止することができる個人環境言語変換装置を容易に提供することができる。

40

【0164】

(20)：個人環境での読み書きの入力を行う読み書き入力手段と、該読み書き入力手段から入力された個人環境での読み書きデータから任意の文字列を抽出し、該抽出した文字

50

列毎の個数をカウントして出現頻度を求める頻度記憶手段と、入力された文字列のうち前記頻度記憶手段に格納されている文字列の出現頻度がある閾値以下のものを強調表示する差分強調装置として、コンピュータを機能させるためのプログラム又はプログラムを記録した記録媒体とするため、このプログラムをコンピュータにインストールすることで、自分の苦手な文字列や単語を手際よく探すことができる個人環境差分強調装置を容易に提供することができる。

【0165】

(21)：個人環境での読み書きの入力を行う読み書き入力手段と、該読み書き入力手段から入力された個人環境での読み書きデータから任意の文字列を抽出し、該抽出した文字列毎の個数をカウントして出現頻度を求める頻度記憶手段と、入力された文字列のうち前記頻度記憶手段に格納されている文字列の出現頻度がある閾値以上のものを強調表示する差分強調装置として、コンピュータを機能させるためのプログラム又はプログラムを記録した記録媒体とするため、このプログラムをコンピュータにインストールすることで、自分の興味のある段落や文を手際よく探すことができる個人環境差分強調装置を容易に提供することができる。

10

【0166】

(22)：個人環境での書きの入力を行う書き入力手段と、前記書き入力手段から入力された個人環境での書きデータから任意の文字列を抽出し、該抽出した文字列毎の個数をカウントする頻度記憶手段と、入力された文字列の差分として出力する対象の単位である抽出単位と入力された文字列の差分を検出するために比較する領域の単位である検出領域を設定し、抽出手段で、入力された文書データの現在の前記検出領域において、格納手段に格納されていない前記抽出単位に相当するものを強調表示して現在の検出領域の文書を出し、前記強調表示したものを前記格納手段に格納することを、前記検出領域ごとに繰り返す差分強調装置と、前記抽出手段で強調表示すべきと判断された前記入力された文字列のうち、前記頻度記憶手段に格納されている文字列の出現頻度がある閾値以下のものを強調表示する前記差分強調装置として、コンピュータを機能させるためのプログラム又はプログラムを記録した記録媒体とするため、このプログラムをコンピュータにインストールすることで、強調表示される箇所が少なく、自分の苦手な表現が初出に現れた箇所を容易に特定できる個人環境差分強調装置を容易に提供することができる。

20

【図面の簡単な説明】

30

【図1】本発明の原理説明図である。

【図2】実施の形態における個人環境言語変換装置の説明図である。

【図3】実施の形態における言語補助変換部を用いる場合の説明図である。

【図4】実施の形態における頻度記憶部の説明図(1)である。

【図5】実施の形態における頻度記憶部の処理フローチャート(1)である。

【図6】実施の形態における頻度記憶部の説明図(2)である。

【図7】実施の形態における頻度記憶部の処理フローチャート(2)である。

【図8】実施の形態における言語変換部の説明図である。

【図9】実施の形態における個人環境差分強調装置の説明図である。

【図10】実施の形態における削除用単語記憶部を用いる個人環境差分強調装置の説明図である。

40

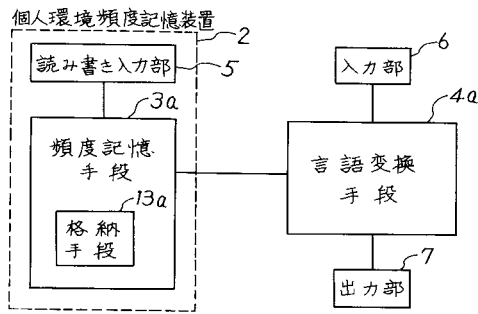
【符号の説明】

- 2 個人環境頻度記憶装置
- 3 a 頻度記憶手段
- 4 a 言語変換手段
- 5 読み書き入力部
- 6 入力部
- 7 出力部
- 13 a 格納手段



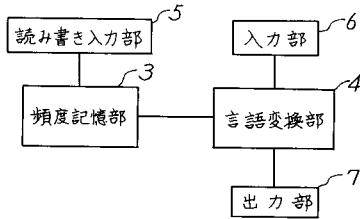
【図1】

本発明の原理説明図



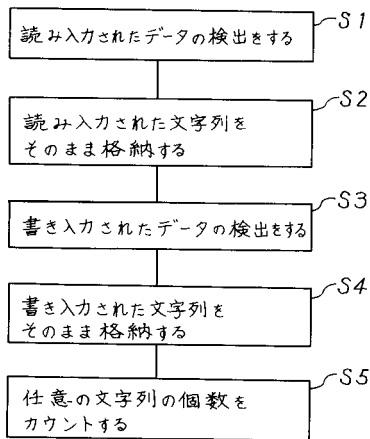
【図2】

個人環境言語変換装置の説明図



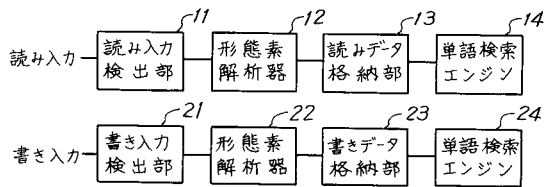
【図5】

頻度記憶部の処理フローチャート(1)



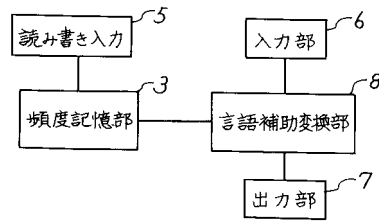
【図6】

頻度記憶部の説明図(2)



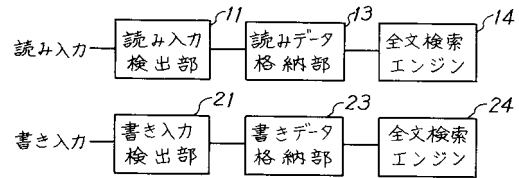
【図3】

言語補助変換部を用いる場合の説明図



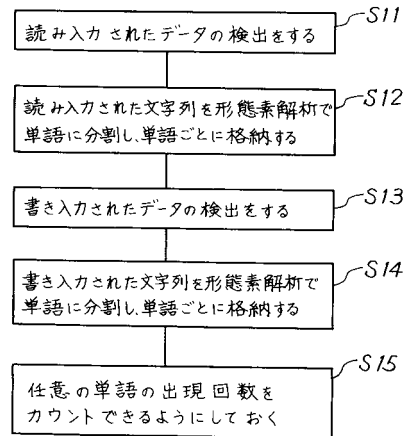
【図4】

頻度記憶部の説明図(1)



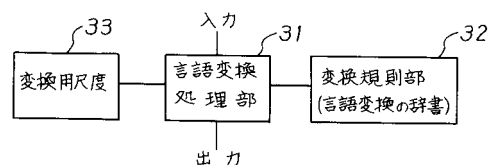
【図7】

頻度記憶部の処理フローチャート(2)



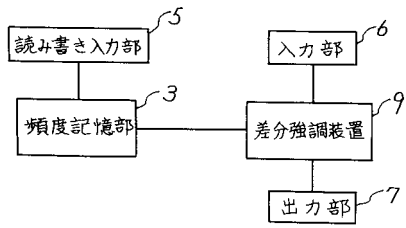
【図8】

言語変換部の説明図



【図 9】

個人環境差分強調装置の説明図



【図 10】

削除用単語記憶部を用いる個人環境差分強調装置の説明図

