

(19) 日本国特許庁(JP)

(12) 特 許 公 報(B2)

(11) 特許番号

特許第4574186号
(P4574186)

(45) 発行日 平成22年11月4日(2010.11.4)

(24) 登録日 平成22年8月27日(2010.8.27)

(51) Int.Cl. F 1
G 0 6 F 17/30 (2006.01)
 G 0 6 F 17/30 3 3 0 C
 G 0 6 F 17/30 1 7 0 A
 G 0 6 F 17/30 2 1 0 A
 G 0 6 F 17/30 3 2 0 D

請求項の数 3 (全 11 頁)

<p>(21) 出願番号 特願2004-40181 (P2004-40181) (22) 出願日 平成16年2月17日(2004.2.17) (65) 公開番号 特開2005-234688 (P2005-234688A) (43) 公開日 平成17年9月2日(2005.9.2) 審査請求日 平成19年1月5日(2007.1.5)</p>	<p>(73) 特許権者 000006747 株式会社リコー 東京都大田区中馬込1丁目3番6号 (72) 発明者 真野 博子 東京都大田区中馬込1丁目3番6号 株式 会社リコー内 審査官 長 由紀子</p>
---	--

最終頁に続く

(54) 【発明の名称】 重要言語識別方法、重要言語識別プログラム、重要言語識別装置、文書検索装置およびキーワード抽出装置

(57) 【特許請求の範囲】

【請求項1】

検索語候補の単語から重要言語を識別する重要言語識別装置であって、
 ユーザが入力した単語を含む検索条件を受け付ける検索条件受付手段と、
 前記検索条件受付手段で受け付けられた検索条件から検索語候補の単語を抽出する検索語候補抽出手段と、
 前記検索語候補抽出手段で抽出された前記検索語候補の単語が出現する出現頻度の累積値を記憶する記憶手段と、
 前記検索語候補抽出手段で抽出された前記検索語候補の単語の重要度を算定する単語重要度算定手段と、
 を有し、
 前記抽出された前記検索語候補の単語の数が所定数以下の場合、
 前記単語重要度算定手段が、前記検索語候補の単語を重要語とし、
 前記記憶手段が、前記検索語候補の単語に対応して前記検索語候補の単語の前記検索条件で出現した出現頻度を前記記憶手段に記憶されている前記累積値に加算した値を新たな累積値として記憶し、
 前記抽出された前記検索語候補の単語の数が所定数を超える場合は、
 前記単語重要度算定手段が、
 前記抽出された検索語候補の単語の前記出現頻度と、前記記憶手段に記憶されている前記検索語候補の単語に対応する前記累積値とを用いて所定の演算を行い、前記検索語候

補の単語の重要度を算定することを特徴とする重要言語識別装置。

【請求項 2】

検索語候補の単語から重要言語を識別する重要言語識別方法であって、
 重要言語識別装置が、前記検索語候補の単語が出現する出現頻度の累積値を記憶する記憶手段を備え、
 前記重要言語識別方法は、
 前記重要言語識別装置が、
 ユーザが入力した単語を含む検索条件を受け付ける検索条件受付工程と、
 前記検索条件受付工程で受け付けられた検索条件から検索語候補の単語を抽出する検索語候補抽出工程と、
 前記検索語候補抽出工程で抽出された前記検索語候補の単語が出現する出現頻度の累積値を記憶する記憶工程と、
 前記検索語候補抽出工程で抽出された前記検索語候補の単語の重要度を算定する単語重要度算定工程と、
 を含み、
 前記抽出された前記検索語候補の単語の数が所定数以下の場合には、
 前記単語重要度算定工程で、前記検索語候補の単語を重要語とし、
 前記記憶工程で、前記検索語候補の単語に対応して前記検索語候補の単語の前記検索条件で出現した出現頻度を前記記憶手段に記憶されている前記累積値に加算した値を新たな累積値として記憶し、
 前記抽出された前記検索語候補の単語の数が所定数を超える場合は、
 前記単語重要度算定工程で、
 前記抽出された検索語候補の単語の前記出現頻度と、前記記憶手段に記憶されている前記検索語候補の単語に対応する前記累積値とを用いて所定の演算を行い、前記検索語候補の単語の重要度を算定することを特徴とする重要言語識別方法。

【請求項 3】

請求項 2 に記載の重要言語識別方法をコンピュータに実行させることを特徴とする重要言語識別プログラム。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、入力された自然言語中の重要語を識別する重要言語識別方法、重要言語識別プログラム、重要言語識別装置、文書検索装置およびキーワード抽出装置に関する。

【背景技術】

【0002】

従来、自然言語による入力を検索条件とする文書検索においては、検索条件となっている文章中あるいは文書中から抽出された全単語から、真に検索に有用な重要語のみを識別する技術が採用されている。検索条件から抽出された各単語の重要度を算定するのに、従来から用いられてきたのは、単語の検索条件中での出現頻度である。すなわち、検索条件の中で繰り返し用いられている単語がそうでない単語よりも重要とみなされる。この場合、検索が行われるたびに、そこで与えられている検索条件のみに基づいて単語の重要度が判断される（例えば、特許文献 1、2 を参照。）。

【0003】

一方、それらとは別に、過去の検索時の情報を後の検索に利用する方法もいくつか提案されている（例えば、特許文献 3～5 を参照。）。特に、特許文献 4 に記載の技術は、過去の検索語の履歴を保存し、それを後の検索時に検索語候補として提示する。これにより、検索語が適切でなかった場合に検索条件を補正することができる。

【0004】

【特許文献 1】特開平 5 - 3 2 4 7 3 3 号公報

10

20

30

40

50

【特許文献2】特開2001-290840号公報

【特許文献3】特開2000-227922号公報

【特許文献4】特開2002-7450号公報

【特許文献5】特開2002-236699号公報

【発明の開示】

【発明が解決しようとする課題】

【0005】

しかしながら、上記従来技術のように、単語の検索条件中での出現頻度を基に重要語を識別するのでは、検索条件の中で繰り返し用いられている単語がそうでない単語よりも重要とは限らないので、重要語として識別するための精度に問題がある。

10

【0006】

例えば、特許文献4に開示された技術は、過去の検索語を検索語候補として提示する際に、検索対象や検索目的が限定されている状況での検索を想定している。したがって、このような制約のない場合には、実際に過去の検索語が適用できる場合は極めて限られるので、重要語として識別するための有効性は薄い。特に、検索条件として、単語だけでなく、ある程度長い文章も入力できるシステムでは、検索語候補である単語を予測提示しても、それが実際にユーザが入力しようとしている検索条件と一致する可能性は極めて低い。

【0007】

本発明は、上述した従来技術による問題点を解消するため、ユーザが検索に不可欠と考える単語を重要語と識別し、ユーザの検索意図が十分に反映された重要語に基づく検索を支援する重要言語識別方法および重要言語識別プログラム、重要言語識別装置を提供することを目的とする。さらに、その重要言語識別装置を備えた文書検索装置、キーワード抽出装置を提供することも本発明の目的としている。

20

【課題を解決するための手段】

【0008】

上述した課題を解決し、目的を達成するため、請求項1にかかる重要言語識別装置は、検索語候補の単語から重要言語を識別する重要言語識別装置であって、ユーザが入力した単語を含む検索条件を受け付ける検索条件受付手段と、前記検索条件受付手段で受け付けられた検索条件から検索語候補の単語を抽出する検索語候補抽出手段と、前記検索語候補抽出手段で抽出された前記検索語候補の単語が出現する出現頻度の累積値を記憶する記憶手段と、前記検索語候補抽出手段で抽出された前記検索語候補の単語の重要度を算定する単語重要度算定手段と、を有し、前記抽出された検索語候補の単語の数が所定数以下の場合には、前記単語重要度算定手段が、前記検索語候補の単語を重要語とし、前記記憶手段が、前記検索語候補の単語に対応して前記検索語候補の単語の前記検索条件で出現した出現頻度を前記記憶手段に記憶されている前記累積値に加算した値を新たな累積値として記憶し、前記抽出された前記検索語候補の単語の数が所定数を超える場合は、前記単語重要度算定手段が、前記抽出された検索語候補の単語の前記出現頻度と、前記記憶手段に記憶されている該検索語候補の単語に対応する前記累積値とを用いて所定の演算を行い、前記検索語候補の単語の重要度を算定することを特徴とする。

30

40

【0009】

この請求項1に記載の発明によれば、抽出された検索語候補のうちユーザの検索意図を十分に反映しているものをその重要度により把握することができる。

【0022】

また、請求項2にかかる重要言語識別方法は、検索語候補の単語から重要言語を識別する重要言語識別方法であって、重要言語識別装置が、前記検索語候補の単語が出現する出現頻度の累積値を記憶する記憶手段を備え、前記重要言語識別方法は、前記重要言語識別装置が、ユーザが入力した単語を含む検索条件を受け付ける検索条件受付工程と、前記検索条件受付工程で受け付けられた検索条件から検索語候補の単語を抽出する検索語候

50

補抽出工程と、前記検索語候補抽出工程で抽出された検索語候補の単語の数が所定数以下の場合、前記検索語候補の単語に対応して前記検索語候補の単語の前記検索条件で出現した出現頻度を前記記憶手段に記憶されている累積値に加算した値を新たな累積値として記憶手段に記憶する記憶工程と、を含み、前記抽出された検索語候補の単語の数が所定数以下の場合、前記単語重要度算定工程で、前記検索語候補の単語を重要語とし、前記記憶工程で、前記検索語候補の単語に対応して前記検索語候補の単語の前記検索条件で出現した出現頻度を前記記憶手段に記憶されている前記累積値に加算した値を新たな累積値として記憶し、前記抽出された前記検索語候補の単語の数が所定数を超える場合は、前記単語重要度算定工程で、前記抽出された検索語候補の単語の前記出現頻度と、前記記憶手段に記憶されている該検索語候補の単語に対応する前記累積値とを用いて所定の演算を行い、前記検索語候補の単語の重要度を算定することを特徴とする。

10

【0023】

この請求項8に記載の発明によれば、抽出された検索語候補のうちユーザの検索意図を十分に反映しているものをその重要度により把握することができる。

【0024】

また、請求項3に記載の重要言語識別プログラムは、請求項2に記載の重要言語識別方法をコンピュータに実行させることを特徴とする。

【0025】

この請求項3に記載の発明によれば、請求項2に記載の重要言語識別方法をコンピュータに実行させることができる。

20

【発明の効果】

【0040】

本発明にかかる重要言語識別方法および重要言語識別プログラム、重要言語識別装置によれば、短い検索条件で用いられた、ユーザが検索に不可欠と考える検索語を、長い検索条件においても重要とみなすことで、ユーザの検索意図が十分に反映された重要語を正確に識別できるという効果を奏する。さらに、この重要言語識別装置を文書検索装置に適用することで、ユーザの検索意図が十分に反映された検索語に基づく文書検索が可能になる。また、この重要言語識別装置をキーワード抽出装置に適用することで、ユーザの検索意図に合致したキーワード抽出ができる。

30

【発明を実施するための最良の形態】

【0041】

以下に添付図面を参照して、本発明にかかる重要言語識別方法および重要言語識別プログラム、重要言語識別装置、文書検索装置、キーワード抽出装置の好適な実施の形態を詳細に説明する。

【0042】

(重要言語識別装置のハードウェア構成)

まず、本発明の実施の形態にかかる重要言語識別装置のハードウェア構成について説明する。図1は、この重要言語識別装置のハードウェア構成を示す図である。この重要言語識別装置100は、CPU101、ROM102、RAM103、ハードディスク104、入力手段105、表示手段106、記憶媒体読取手段107、および通信インターフェース(I/F)108が、バス109によって接続され構成されている。

40

【0043】

CPU101は、装置全体を制御する。ROM102には、基本入出力プログラムが記憶されている。RAM103は、CPU101のワークエリアとして使用される。ハードディスク104は、CPU101の制御にしたがいデータの書き込み/読み出しを行う。入力手段105は、キーボードやマウス等からなり、文字、数字、データの入力や、表示手段106に表示されるカーソルの移動や範囲選択、あるいはウィンドウの移動やサイズの変更等を行う。表示手段106は、カーソル、アイコンあるいはツールボックスをはじめ、文書、画像、機能情報等のデータを表示する。記憶媒体読取手段107は、CPU1

50

01の制御にしたがって、着脱可能な記憶媒体であるCD/DVDディスク110からのデータの読み取りを制御する。通信インターフェース108は、ネットワーク200に接続され、このネットワーク200と重要言語識別装置100とのインターフェースとして機能する。

【0044】

(重要言語識別装置を含むネットワーク構成)

次に、この重要言語識別装置を含むネットワーク構成を説明する。図2は、この重要言語識別装置を含むネットワーク構成を示す概略図である。図1に示したように、重要言語識別装置100は、ネットワーク200に接続可能な通信インターフェース108を備えている。そこで、図2に示すように、複数の重要言語識別装置100をネットワーク200に接続することで、ネットワーク200に接続されているサーバ201を仲介した各重要言語識別装置100間でのデータの送受信が可能になる。

10

【0045】

(重要言語識別装置の機能的構成)

続いて、この重要言語識別装置の機能的構成について説明する。図3は、この重要言語識別装置の機能的構成を示すブロック図である。重要言語識別装置100は、検索条件受付部301、検索語選出部302、および単語重要度算定部303を含み構成される。検索条件受付部301は、ユーザが入力手段105から入力した検索条件を受け付ける。検索語選出部302は、検索条件受付部301が受け付けた検索条件を形態素解析によって単語に分割する。そして、分割された単語群から助詞等検索に不適な単語を不要語として取り除き、残った単語を検索語候補として抽出する。単語重要度算定部303は、検索語選出部302で抽出された検索語候補の重要度(どの程度ユーザの検索意図が反映されているか)を算定する。そして、検索語選出部302は、単語重要度算定部303が算定した各検索語候補の重要度に基づいて、適切な検索語を選出する。

20

【0046】

ところで、この重要言語識別装置100において検索語候補の重要度を算定する際、次のことがいえる。すなわち、検索条件受付部301で受け付ける検索条件が短い場合(例えば、検索条件が数語程度からなる文節の形式である場合等)、検索条件中の各単語のほとんどが、ユーザが吟味の上、検索に不要な単語を除去し不可欠であるとみなした重要語であると考えられる。すなわち、検索条件中の各単語はユーザの検索意図が直接反映されている蓋然性が非常に高い。一方、検索条件が長い場合(例えば、検索条件が多く単語を含む文章全体である場合等)、そこに含まれる単語をすべて検索語とすることは、検索のためにはさほど重要でないかまたは不要である単語を含んでいる可能性が高いため、好ましくない。したがって、検索条件として入力された単語群のなかから重要である単語のみを検索語として選出することが好ましい。

30

【0047】

そこで、重要言語識別装置100では、検索条件受付部301が受け付けた検索条件が短く、検索語選出部302で抽出された検索語候補(単語)数が所定数(例えば3~5語程度)以下の場合、単語重要度算定部303が、それら検索語候補をすべて同一レベルの重要語とみなす。加えて、単語重要度算定部303は、重要語と認識された検索語候補の出現頻度を把握し、把握結果(累積値)をハードディスク104に記憶しておく。

40

【0048】

一方、検索条件受付部301が受け付けた検索条件が長く、検索語選出部302で抽出された検索語候補数が所定数を超える場合は、単語重要度算定部303は、まず、長い検索条件から検索語候補として新たに検索語選出部302において抽出された各単語の出現頻度を算出する(第1段階)。次に、単語重要度算定部303は、第1段階で算出された各単語の出現頻度と、前記短い検索条件下で重要語と認識された同一検索語候補の出現頻度(累積値)との間で所定の演算(例えば、加法や乗法等)を行い(第2段階)、この結果を各検索語候補の重要度とする。すなわち、本発明の重要言語識別装置は、検索語候補の数が所定数を超える場合には、前記第1段階および前記第2段階の処理を行って、各検

50

索語候補の重要度を求める。

【0049】

そして、検索語選出部302は、前記検索語候補が所定数以下の場合、単語重要度算定部303が同一レベルの重要語とみなしたすべての検索語候補を検索語として選出する。一方、前記検索語候補が所定数を超える場合は、単語重要度算定部303が算定した重要度に基づき前記検索語候補のなかから所定値以上の重要度を有するものを検索語として選出する。なお、所定値以上の重要度とは、例えば上位3～5位までというように、ユーザがあらかじめ設定するものとする。また、この場合は、検索語として選出された単語やその出現頻度をハードディスク104に記憶させることはしない。長い検索条件において選出された検索語は、ユーザの意図が直接反映されているわけではないからである。

10

【0050】

例えば、第1の検索条件として、「サーバー」、「クライアント」が入力されたとする。この第1の検索条件は2語であるため、短い検索条件となり、各単語の出現頻度として「サーバー」=1、「クライアント」=1が算出され、各値がそのまま各単語の出現頻度累積値として記憶される。

【0051】

次に、第2の検索条件として、「サーバー」、「クライアント」、「J A V A (R)」が入力されたとする。この第2の検索条件も3語であるため、短い検索条件となり、各単語の出現頻度として「サーバー」=1、「クライアント」=1、「J A V A (R)」=1が算出される。そして、各単語の出現頻度の累積値として、「サーバー」=2、「クライアント」=2、「J A V A (R)」=1が記憶される。

20

【0052】

さらに、第3の検索条件として、「サーバー」、「クライアント」、「J A V A (R)」、「インターネット」、「ウェブ」が入力されたとする。この第3の検索条件は5語であるため、長い検索条件となる。したがって、まず第1段階として第3の検索条件のみにおける各単語の出現頻度として、「サーバー」=1、「クライアント」=1、「J A V A (R)」=1、「インターネット」=1、「ウェブ」=1が算出される。続いて、第2段階として、この第3の検索条件のみから求められた各単語の出現頻度と、同一単語の出現頻度の累積値との間で所定の演算を行い、これを各検索語候補の重要度とする。ここでは、各検索語候補の重要度を求めるための演算に乗法を用いるとすると、「サーバー」=2、「クライアント」=2、「J A V A (R)」=1、「インターネット」=0、「ウェブ」=0が各検索語候補の重要度となる。

30

【0053】

最後に、検索語が選出される。ここに示した例では、各検索語候補の重要度は、「サーバー」=2、「クライアント」=2、「J A V A (R)」=1、「インターネット」=0、「ウェブ」=0であるので、重要度上位3位までの検索語候補を検索語として選出するとすれば、「サーバー」、「クライアント」、「J A V A (R)」が検索語として選出されることになる。

【0054】

このようにすることにより、本発明の重要言語識別装置は、ユーザの検索意図を十分反映された単語を検索語として選出することが可能になる。

40

【0055】

(重要語識別処理の手順)

次に、本発明の重要言語識別装置を用いた重要言語識別処理の手順について説明する。図4は、この重要言語識別処理の手順を示すフローチャートである。

【0056】

まず、検索条件受付部301が、入力手段105からユーザにより入力された検索条件を受け付ける(ステップS401)。ここで入力される検索条件は、人間が特別な訓練なしに自然に習得し使用できる自然言語により設定される。

【0057】

50

次に、検索語選出部 302 が、検索条件受付部 301 によって受け付けられた検索条件から検索語候補を抽出する（ステップ S402）。具体的には、ユーザにより入力された検索条件を形態素解析によって単語に分割する。そして、分割された単語群から助詞等検索に不適な単語を不要語として取り除き、残った単語を検索語候補とする。

【0058】

続いて、単語重要度算定部 303 が、検索語選出部 302 により抽出された検索語候補の数が所定数以下か否かを判断する（ステップ S403）。基準となる検索語候補の数はユーザが任意に設定できるが、一般的には 3～5 語程度が好ましい。抽出された検索語候補の数が所定数以下である場合（ステップ S403：Yes）は、単語重要度算定部 303 が当該検索語候補をすべて重要語とみなし、重要語とされた各検索語候補の出現頻度（累積値）をハードディスク 104 に記憶する（ステップ S404）。そして、検索語選出部 302 は、単語重要度算定部 303 が重要語とみなした検索語候補をすべて検索語として選出し（ステップ S405）、処理を終了する。一方、抽出された検索語候補の数が所定数を超える場合（ステップ S403：No）は、ステップ S406 へ進む。

10

【0059】

ステップ S403 において、抽出された検索語候補の数が所定数を超える場合は（ステップ S403：No）、単語重要度算定部 303 が前述した方法で検索語候補の重要度を算出する（ステップ S406）。そして、検索語選出部 302 が、単語重要度算定部 303 が算定した検索語候補の重要度（例えば、上位 3～5 位程度）に基づき、実際に用いる検索語を選出し（ステップ S405）、処理を終了する。

20

【0060】

本発明の重要言語識別方法では、以上のような手順を経ることにより、ユーザの検索意図が十分反映された単語を検索語として選出することが可能になる。

【0061】

（文書検索装置）

次に、本発明の重要言語識別装置を備えた文書検索装置について説明する。図 5 は、この文書検索装置の機能的構成の一例を示すブロック図である。この文書検索装置は、重要言語識別装置 100、文書検索部 501、および文書データベース 502 を含み構成される。そして、文書データベース 502 に格納されている文書データに対して、前述したような処理を経て重要言語識別装置 100 で選出された検索語をもとに公知の方法により文書検索を行う。このように、重要言語識別装置 100 で選出された検索語に基づいた文書検索を行うことで、よりユーザの検索意図に合致した文書検索が可能になる。

30

【0062】

また、図 5 に示した文書検索装置において、検索語と関連する単語を関連語として選出し、この関連語による再検索を実行させることもできる。図 6 は、関連語による検索を可能とする文書検索装置の機能的構成を示すブロック図である。この文書検索装置は、図 5 に示した構成に加え、検索語選出部 302 で選出された検索語の関連語候補を抽出し、この関連語候補の重要度を単語重要度算定部 303 に算定させ、重要度の高い関連語候補を関連語として選出する関連語選出部 601 を備えている。そして、文書検索部 501 がその関連語に基づいて再検索を行う。このようにすることで、さらにユーザの検索意図に合致した文書検索が可能になる。

40

【0063】

（キーワード抽出装置）

続いて、本発明の重要言語識別装置を備えたキーワード抽出装置について説明する。図 7 は、このキーワード抽出装置の機能的構成の一例を示すブロック図である。このキーワード抽出装置は、重要言語識別装置 100、入力文書受付部 701、およびキーワード抽出部 702 を含み構成される。このキーワード抽出装置は、文書を検索するのではなく、入力された文章からキーワードを抽出するところが、図 5 および図 6 に示した文書検索装置とは異なる。このキーワード抽出装置では、まず、入力文書受付部 701 において、入力文章から形態素解析によりキーワード候補となる単語を抽出する。そして、キーワード

50

抽出部 702 において、重要言語識別装置 100 によって算定された単語の重要度を参照して、前記キーワード候補として抽出された各単語の重要度を算定し、重要度の高い単語をキーワードとして抽出する。このようにすることで、よりユーザの意図に合致したキーワード抽出が可能になる。

【0064】

以上説明したように、本発明の重要言語識別方法および重要言語識別プログラム、重要言語識別装置によれば、短い検索条件で用いられた、ユーザが検索に不可欠と考える検索語を、長い検索条件においても重要とみなすことで、ユーザの検索意図が十分に反映された重要語を正確に識別できる。さらに、この重要言語識別装置を文書検索装置に適用することで、ユーザの検索意図が十分に反映された検索語に基づく文書検索が可能になる。また、この重要言語識別装置をキーワード抽出装置に適用することで、ユーザの検索意図に合致したキーワード抽出ができる。

10

【0065】

なお、本実施の形態で説明した重要言語識別方法は、あらかじめ用意されたプログラムをパーソナル・コンピュータやワークステーション等のコンピュータで実行することにより実現することができる。このプログラムは、ハードディスク、フレキシブルディスク、CD-ROM、MO、DVD等のコンピュータで読み取り可能な記録媒体に記録され、コンピュータによって記録媒体から読み出されることによって実行される。またこのプログラムは、インターネット等のネットワークを介して配布することが可能な伝送媒体であってもよい。

20

【産業上の利用可能性】

【0066】

以上のように、本発明にかかる重要言語識別方法および重要言語識別プログラム、重要言語識別装置は、ユーザが検索に不可欠と考える重要語をより正確に識別することに有用であり、特に、文書検索装置、キーワード抽出装置に適している。

【図面の簡単な説明】

【0067】

【図1】本発明の実施の形態にかかる重要言語識別装置のハードウェア構成を示す図である。

【図2】重要言語識別装置を含むネットワーク構成を示す概略図である。

30

【図3】重要言語識別装置の機能的構成を示すブロック図である。

【図4】重要言語識別処理の手順を示すフローチャートである。

【図5】重要言語識別装置を備えた文書検索装置の機能的構成の一例を示すブロック図である。

【図6】重要言語識別装置を備えた文書検索装置の機能的構成の一例を示すブロック図である。

【図7】重要言語識別装置を備えたキーワード抽出装置の機能的構成の一例を示すブロック図である。

【符号の説明】

【0068】

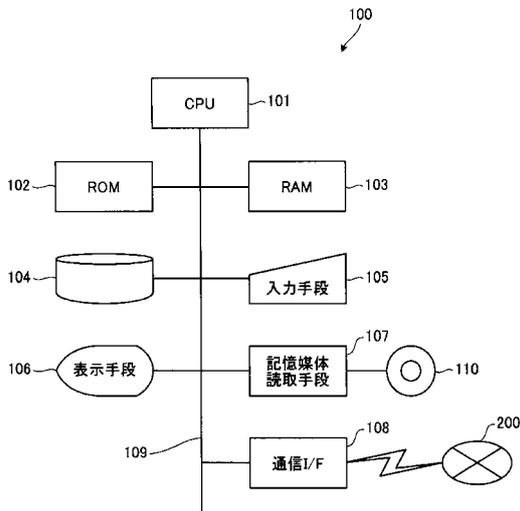
40

- 100 重要言語識別装置
- 101 CPU
- 102 ROM
- 103 RAM
- 104 ハードディスク
- 105 入力手段
- 106 表示手段
- 107 記憶媒体読取手段
- 108 通信インターフェース(I/F)
- 109 バス

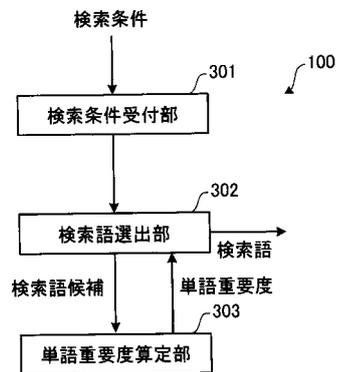
50

- 1 1 0 CD/DVDディスク
- 2 0 0 ネットワーク
- 2 0 1 サーバ
- 3 0 1 検索条件受付部
- 3 0 2 検索語選出部
- 3 0 3 単語重要度算定部
- 5 0 1 文書検索部
- 5 0 2 文書データベース
- 6 0 1 関連語選出部
- 7 0 1 入力文書受付部
- 7 0 2 キーワード抽出部

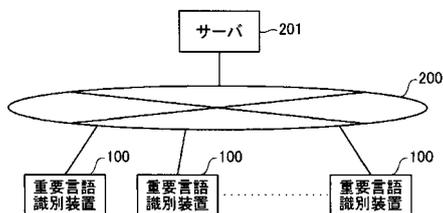
【図1】



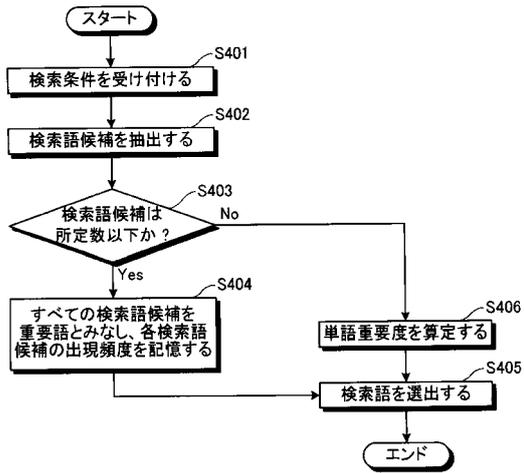
【図3】



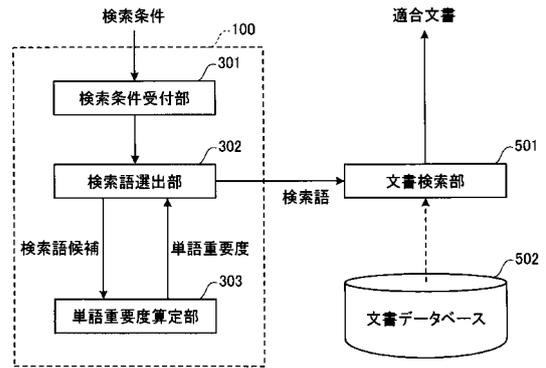
【図2】



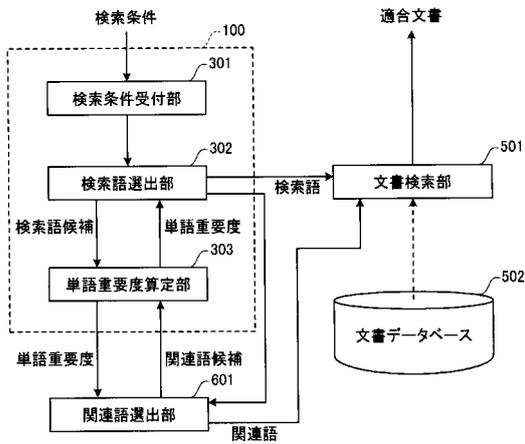
【図4】



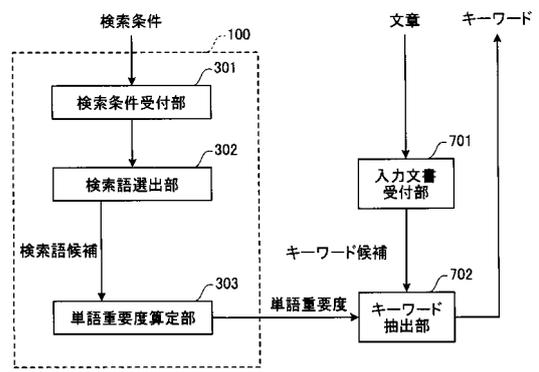
【図5】



【図6】



【図7】



フロントページの続き

- (56)参考文献 特開平06 - 208588 (JP, A)
特開2002 - 245062 (JP, A)
特開2002 - 073681 (JP, A)
特開2000 - 250585 (JP, A)
特開平11 - 073419 (JP, A)
特開2003 - 208433 (JP, A)
特開2002 - 099573 (JP, A)
特開平08 - 614964 (JP, A)

- (58)調査した分野(Int.Cl., DB名)
G06F 17/30