

(19) 日本国特許庁 (JP)

(12) 特 許 公 報 (B2)

(11) 特許番号
特許第5404563号
(P5404563)

(45) 発行日 平成26年2月5日 (2014.2.5)

(24) 登録日 平成25年11月8日 (2013.11.8)

(51) Int.Cl.

G O 6 F 17/30 (2006.01)

F I

G O 6 F 17/30 3 5 O C

G O 6 F 17/30 1 7 O A

G O 6 F 17/30 3 3 O C

G O 6 F 17/30 3 1 O Z

請求項の数 5 (全 10 頁)

(21) 出願番号	特願2010-203280 (P2010-203280)	(73) 特許権者	000006013
(22) 出願日	平成22年9月10日 (2010.9.10)		三菱電機株式会社
(65) 公開番号	特開2012-59126 (P2012-59126A)		東京都千代田区丸の内二丁目7番3号
(43) 公開日	平成24年3月22日 (2012.3.22)	(74) 代理人	100110423
審査請求日	平成24年11月20日 (2012.11.20)		弁理士 曾我 道治
		(74) 代理人	100094695
			弁理士 鈴木 憲七
		(74) 代理人	100111648
			弁理士 梶並 順
		(74) 代理人	100122437
			弁理士 大宅 一宏
		(74) 代理人	100147566
			弁理士 上田 俊一
		(74) 代理人	100161171
			弁理士 吉田 潤一郎

最終頁に続く

(54) 【発明の名称】 検索装置

(57) 【特許請求の範囲】

【請求項 1】

入力された文字列に基づいて、検索対象とする複数個の文書から所望の文書を検索する検索装置であって、

前記文字列を入力として、前記文字列と検索対象とする複数個の文書を照合し、前記文字列と部分一致または完全一致する複数個の文書と、前記文字列が複数個の文書中に出現する回数に応じた検索スコアとを検索結果として出力する検索手段と、

前記検索対象とする複数個の文書のそれぞれに対する形態素と、検索時に使用される重要度に応じて形態素毎に付与したペナルティ値とを保持する形態素辞書と、

前記文字列と前記検索手段の検索結果を入力とし、前記検索結果のそれぞれの文書に対し、前記形態素辞書を参照して前記文字列から形態素を抽出し、前記文書中には存在するが、前記文字列中からは抽出されなかった形態素に対し、前記ペナルティ値に基づいて前記検索スコアが低くなるように当該検索スコアを修正し、修正した検索スコアに基づいて検索結果の出力順位を再構成して出力する検索順位修正手段と、
を備えたことを特徴とする検索装置。

【請求項 2】

入力された文字列は、入力音声を音声認識手段により音声認識し、その認識結果が文字列として出力されたものであることを特徴とする請求項 1 記載の検索装置。

【請求項 3】

前記形態素に付与するペナルティ値は、当該文書を検索するときに入力文字列中から省

略される可能性が小さい形態素ほど大きなペナルティ値を付与しておくことを特徴とする、請求項 1 または 2 記載の検索装置。

【請求項 4】

前記検索順位修正手段は、前記形態素辞書を参照して前記文字列から形態素を抽出する方法として、文字列上の DP マッチングを用い、前記文字列と前記形態素辞書中の形態素が完全一致しない場合でも前記文字列中から形態素を抽出することを特徴とする請求項 1 または 2 記載の検索装置。

【請求項 5】

前記検索対象とする文書は複数個の施設名称であって、施設名称を構成する最後尾の形態素が「店」である施設名称の先頭の形態素に対し、他の形態素よりも大きなペナルティ値を付与することを特徴とする請求項 1 または 2 記載の検索装置。

10

【発明の詳細な説明】

【技術分野】

【0001】

この発明は、大量の文書や施設名中から、所望の文書や施設名の検索を効率よく行う大規模な検索装置に関するものである。

【背景技術】

【0002】

さまざまな施設名を検索対象とする検索システムを構築する場合、利用者は施設の正式名称を知らない場合があるので、施設名を形態素や音節に分解して、形態素や音節のユニグラムやバイグラムを照合単位として検索を行う技術が従来よりあり、このような技術として下記特許文献 1 に開示されたものがある。

20

【0003】

特許文献 1 では、単語や音節等を単位として、入力文字列と検索対象施設名を比較照合し、マッチした単語や音節のユニグラムやバイグラム数に基づいて検索スコアを算出し、スコアの高い順に候補を提示する技術が開示されている。

しかし特許文献 1 の技術では、例えば大船にある「ウミベ」という百貨店の正式名称が「ウミベ大船」である場合、「おーふなうみべ（大船ウミベ）」という入力文字列で音節バイグラム数に基づいて検索すると、正解の「うみべおーふな（ウミベ大船）」よりも、「えーしょぼーおーふなうみべてん（A 書房大船ウミベ店）」という不自然な検索結果が検索結果の上位に出力されるという課題があった。これは前記入力文字列中の音節バイグラム「なう」が前者の「うみべおーふな」ではマッチしないのに対し、後者の「えーしょぼーおーふなうみべてん」ではマッチし、検索スコアが後者のほうが高くなるためである。

30

【先行技術文献】

【特許文献】

【0004】

【特許文献 1】特開 2008-262279 号公報

【発明の概要】

【発明が解決しようとする課題】

40

【0005】

この発明は上記課題を解決するためになされたもので、前記のような不自然な検索結果を抑制し検索精度を向上させることを目的とする。

【課題を解決するための手段】

【0006】

この発明に係る検索装置は、

入力された文字列に基づいて、検索対象とする複数個の文書から所望の文書を検索する検索装置であって、

前記文字列を入力として、前記文字列と検索対象とする複数個の文書を照合し、前記文字列と部分一致または完全一致する複数個の文書と、前記文字列が複数個の文書中に出現

50

する回数に応じた検索スコアとを検索結果として出力する検索手段と、

前記検索対象とする複数個の文書のそれぞれに対する形態素と、検索時に使用される重要度に応じて形態素毎に付与したペナルティ値とを保持する形態素辞書と、

前記文字列と前記検索手段の検索結果を入力とし、前記検索結果のそれぞれの文書に対し、前記形態素辞書を参照して前記文字列から形態素を抽出し、前記文書中には存在するが、前記文字列中からは抽出されなかった形態素に対し、前記ペナルティ値に基づいて前記検索スコアが低くなるように当該検索スコアを修正し、修正した検索スコアに基づいて検索結果の出力順位を再構成して出力する検索順位修正手段とを備える。

【発明の効果】

【0007】

10

この発明による検索装置によれば、入力された文字列に基づいて検索手段で検索された検索対象の複数個の文書と、前記文字列が複数個の文書中に出現する回数に応じた検索スコアとの検索結果を、検索対象とする複数個の文書のそれぞれに対する形態素と、検索時に使用される重要度に応じて形態素毎に付与したペナルティ値とを保持する形態素辞書を参照し、前記検索対象文書中には存在するが、前記入力文字列中からは抽出されなかった形態素に対し、前記ペナルティ値を差し引いて前記検索スコアを修正し、修正した検索スコアに基づいて検索結果の出力順位を再構成して出力する検索順位修正手段により不自然な検索結果を抑制する効果がある。

【図面の簡単な説明】

【0008】

20

【図1】この発明による検索装置の実施の形態1の構成を示すブロック図である。

【図2】テキスト検索辞書作成用の検索対象施設名の説明図である。

【図3】検索対象施設名から作成したテキスト検索辞書の説明図である。

【図4】ペナルティ値が設定された形態素辞書例を示す説明図である。

【図5】検索手段の検索結果である施設名のID番号と検索スコアの対の中間検索結果の出力例を示す説明図である。

【図6】検索順位修正手段による処理手順のフロー図である。

【図7】検索順位修正手段による修正検索スコアの大きさ順並べ換え結果を示す説明図である。

【図8】この発明による検索装置の実施の形態2の構成を示すブロック図である。

30

【図9】実施の形態2の検索順位修正手段による修正検索スコアの大きさ順並べ換え結果を示す説明図である。

【発明を実施するための形態】

【0009】

実施の形態1.

本実施の形態では施設や観光スポットの名称（以後は簡単のため施設と観光スポットを総称して施設という）を検索する場合を例にとり説明する。

図1はこの発明による検索装置の実施の形態1の構成を示すブロック図である。

同図において、1は文字列の入力端、2は文字列、3は検索手段、4は検索辞書メモリ、5は中間検索結果、6は検索順位修正手段、7は形態素辞書メモリ、8は検索結果である。

40

【0010】

検索辞書メモリ4にはテキスト検索辞書を事前に作成して格納しておく。テキスト検索辞書の作成方法を説明する。例えば図2に示すとおり、検索対象施設名が「A書房大船ウミベ店（えーしょぼーおーふなうみべてん）」、「ウミベ大船（うみべおーふな）」等として説明する。（）内は施設名の読みを示している。ここで「ウミベ」は施設の固有名詞であり、本例では百貨店名とする。

【0011】

前記テキスト検索辞書は施設名を構成する言語単位を索引語として転置インデックスとして構成する。本例では索引語として施設名の読みの音節の2連鎖（音節バイグラム）を

50

用いる。「A書房大船ウミベ店（えーしょぼーおーふなうみべてん）」に含まれる音節バイグラムは、「えーしょ」、「しょぼー」、「ぼーおー」、「おーふ」「ふな」「なう」「うみ」「みべ」、「べて」「てん」の10種類である。また「ウミベ大船（うみべおーふな）」に含まれる音節バイグラムは「うみ」「みべ」「べおー」「おーふ」「ふな」の5種類である。検索辞書メモリ4は、これらの音節バイグラムを索引語として、索引語と施設名のID番号をテキスト検索辞書として保持する。前記の施設名から作成したテキスト検索辞書を図3に示す。

【0012】

形態素辞書メモリ7には形態素辞書を事前に作成して格納しておく。形態素辞書の作成方法を説明する。まず検索対象とする施設名を形態素解析器等を使用して形態素に分割する。必要に応じて形態素への分割結果を手で修正してもよい。また英語等のように元々単語に分割されている言語では分割処理は不要であり、この場合には単語を形態素とみなす。次に各形態素毎に検索時に使用される重要度に応じて所定のペナルティ値を付与し、形態素とともに形態素辞書として保持する。なお本実施の形態では前記ペナルティ値は当該施設を検索するときに省略される可能性の低い形態素ほど大きなペナルティ値を設定しておく。前記「A書房大船ウミベ店」、および「ウミベ大船」に対する形態素辞書の例を図4に示す。「A書房大船ウミベ店」の形態素辞書は、「えーしょぼー(3)」、「おーふな(1)」、「うみべ(1)」である。()内の値はペナルティ値である。「A書房大船ウミベ店」を検索する場合の文字列2としては、「えーしょぼー」という形態素を省略する可能性は低いと考えられるので、他の形態素よりも大きなペナルティ値を付与している。一方「ウミベ大船」に対する形態素辞書は、当該施設を検索する場合の発話としては、「うみべ」という形態素を省略する可能性は低いと考えられるので、他の形態素よりも大きなペナルティ値を付与している。

【0013】

次に検索の動作について説明する。

文字列の入力端1から文字列2を入力すると、検索手段3はまず文字列2を構成する音節バイグラムを全て抽出する。例えば入力文字列2を「おーふなうみべ」とすると、音節バイグラムとして、「おーふ」「ふな」「なう」「うみ」「みべ」という5個の音節バイグラムを抽出する。

【0014】

次に検索手段3は、検索辞書メモリ4に格納しているテキスト検索辞書を参照し、抽出した音節バイグラム毎に当該音節バイグラムを含む施設の検索スコアに1を加算する。抽出した全音節バイグラムに対しこのスコア加算処理を行う。本例では、施設ID=1の「A書房大船ウミベ店（えーしょぼーおーふなうみべてん）」は、「おーふ」「ふな」「なう」「うみ」「みべ」の5個の音節バイグラムが文字列2の音節バイグラムとマッチするので、検索スコアは5となる。一方施設ID=2の「ウミベ大船（うみべおーふな）」は「おーふ」「ふな」「うみ」「みべ」の4個の音節バイグラムが文字列2の音節バイグラムとマッチするので、検索スコアは4となる。上記加算処理終了後、検索手段3は中間検索結果5として、検索スコアが1以上のN個の施設名のID番号と検索スコアの対を出力する。ここでNは1以上の整数である。中間検索結果5の出力例を図5に示す。

【0015】

次に検索順位修正手段6は、文字列の入力端1からの文字列2と検索手段3からの中間検索結果5を入力とし、中間検索結果5のN個の施設名それぞれに対し形態素辞書メモリ7に格納されている当該施設名の形態素辞書を用いて、文字列2と照合することにより文字列2に含まれる形態素を抽出する。抽出した形態素と、形態素辞書メモリ7に格納されている当該施設の形態素辞書を比較し、形態素辞書中には存在するが、認識結果の音素列からは抽出されなかった形態素に対し、図4に示す形態素辞書に予め設定されたペナルティ値を付与して検索スコアをリスクアリングする。

【0016】

以下に図6を参照し、検索順位修正手段6の具体的な処理手順を述べる。

手順 1) $k=1$ とおく (図 6 の st101)

手順 2) 形態素辞書メモリ 7 に保持している形態素辞書を参照し、図 5 に示す検索手段 3 の中間検索結果 5 の第 k 位 (この場合は $k=1$ であるから 1 位) の施設名の形態素と文字列 2 の照合処理を行い、文字列 2 に含まれる形態素を抽出する (図 6 の st102)。ここで前記照合処理とは、形態素辞書中の 1 個以上の形態素の組み合わせが文字列 2 と一致するかどうかを調べることであり、一致する場合は前記 1 個以上の形態素が文字列 2 に含まれると判定し、前記 1 個以上の形態素を抽出する。

例えば $k=1$ の場合は、前述のように 1 位の検索結果は施設 ID = 1 の施設名であり、図 4 に示すとおり形態素辞書中の形態素は、「えーしょぼー」、「おーふな」、「うみべ」、「てん」となる。これらの形態素と文字列 2 である「おーふなうみべ」との間で照合を行うと、「おーふな」と「うみべ」の 2 個の形態素が抽出される。

【 0 0 1 7 】

手順 3) 手順 2 で抽出した文字列 2 に含まれる形態素と、 k 位の検索結果の形態素辞書中の形態素を比較し、前記形態素辞書中には存在するが文字列 2 中には存在しない形態素に対し、形態素辞書中のペナルティ値を累積したペナルティ累積値 $P(k)$ を算出する (図 6 の st103)。

例えば $k=1$ の場合は、上述のとおり文字列 2 に含まれる形態素は「おーふな」と「うみべ」の 2 個、形態素辞書中の形態素は、「えーしょぼー」、「おーふな」、「うみべ」、「てん」なので、形態素辞書中には存在するが文字列 2 中には存在しない形態素は「えーしょぼー」と「てん」の 2 個である。これらの形態素に対するペナルティ値は図 4 に示すとおり、それぞれ 3 と 0 なので、前記ペナルティ累積値 $P(k)$ の値は、 $P(k) = 3+0 = 3$ となる。

【 0 0 1 8 】

手順 4) 手順 3 で算出したペナルティ累積値 $P(k)$ と、検索スコア $S(k)$ から下記の (1) 式によって修正検索スコア $S'(k)$ を算出する (図 6 の st104)。(1) 式中で は実験的に予め決めた定数であり、本実施の形態例では $=0.5$ とする。

【 0 0 1 9 】

$$S'(k) = S(k) - P(k) \cdot \dots \quad (1)$$

この結果、上述の $k=1$ の例では、 $S'(1) = 5 - 0.5 \cdot 3 = 3.5$ となる。

【 0 0 2 0 】

手順 5) $k = N$ なら、手順 6 へ進む。 $k < N$ なら、 $k=k+1$ とし、手順 2 に戻る。(図 6 の st105, st106)。

手順 6) 手順 4 で修正した修正スコア $S'(k)$ ($k=1 \sim N$) を用い、修正スコア $S'(k)$ の大きい順に検索結果を並べ換え、検索結果 8 として出力する。(図 5 の st107)

【 0 0 2 1 】

処理手順は以上である。上記処理の結果、検索手段の出力結果で第 2 位の施設 ID=2 では、図 4 に示すとおり形態素辞書中の形態素は「うみべ」、「おーふな」なので、これらの形態素と文字列 2 である「おーふなうみべ」との間で照合を行うと「おーふな」と「うみべ」の 2 個の形態素が抽出される。この結果、形態素辞書中の形態素が認識結果中に全て存在するので、ペナルティ累積値 $P(k)$ の値は 0 となり、(1) 式で計算される修正後の検索スコア $S'(2) = S(2) = 4$ となる。

修正後の検索スコアの大きい順に検索順位を並べ換えた結果を図 7 に示す。「ウミベ大船」が検索順位の第 1 位になっていることがわかる。

【 0 0 2 2 】

このように本実施の形態によれば、各施設名毎に形態素辞書を備え、各形態素には当該形態素が文字列 2 に含まれなかった場合に付与するペナルティ値を設定する。このペナルティ値として当該施設を検索するときに省略される可能性の低い形態素ほど大きなペナルティ値を設定しておき、上述したとおりペナルティ累積値 $P(k)$ を差し引いた修正スコア $S'(k)$ の大きい順に検索結果を出力するように構成したので、「大船ウミベ」という発話に対し、「ウミベ大船」よりも「A 書房大船ウミベ店」が上位に検索されるという不自然

10

20

30

40

50

な結果を抑制する効果がある。

【 0 0 2 3 】

なお、本例では検索手段3では、音節バイグラムを転置インデックスの索引語としたが、索引語は任意の単位でよい。例えば単語のバイグラムや、単語または音節のユニグラムでもよい。また本例では検索手段3における検索方式として転置インデックスを用いる方式を説明したが、文字列2と検索対象との部分マッチングを許す任意の検索方式を用いてもよい。

【 0 0 2 4 】

また、形態素辞書の各形態素に付与するペナルティ値としては、施設名を構成する最後尾の形態素が「店」である施設名の先頭の形態素に対し、他の形態素よりも大きなペナルティ値を付与してもよい。これは一般に、公園や百貨店内にある施設名は「施設のブランド名等の固有名詞＋（公園名または百貨店名）＋店」というパターンが多く、最後尾の形態素が「店」である施設名の先頭の形態素は、当該施設を検索する場合に省略することがほぼ無いと考えられるからである。このようにペナルティ値を付与することによりペナルティ付与作業の効率化を図る効果が得られる。

【 0 0 2 5 】

実施の形態2 .

本実施の形態では、実施の形態1と同様に施設名を検索する場合を例にとり説明する。

図8はこの発明による検索装置の実施の形態2の構成を示すブロック図である。

同図において、実施の形態1と同等部分には同一番号を付し、説明を省略する。9は音声の入力端、10は入力音声、11は音声認識手段、12は言語モデルメモリ、13は音響モデルメモリである。

【 0 0 2 6 】

言語モデルメモリ12には統計言語モデルを事前に作成して格納しておく。本例では検索対象とする全施設名の表記の音節列を学習データとして、音節を単位としたトライグラムを学習して格納しておく。なお音節を単位とすることの利点は、学習データとする施設数に関わらず、音節の種類数は数百個以下におさまるので、認識時の演算量増加を抑えた言語モデルを作成できることである。

音響モデルメモリ13には音声の特徴をモデル化した音響モデルを格納している。本実施の形態では音響モデルは例えばHMM(Hidden Markov Model)とする。

【 0 0 2 7 】

次に音声認識と検索の動作について説明する。

音声の入力端9から音声10を入力すると音声認識手段11は言語モデルメモリ12に保存されている言語モデルと音響モデルメモリ13に保存されている音響モデルを用いて、例えばビタピアルゴリズムによって音声認識を行い音声認識結果として、文字列2を出力する。文字列2は本例ではひらがな表記とする。

例えば音声10の発話内容が「大船ウミベ」である音声認識手段11の出力は、例えば「おーふなうみで」となる。本例では、「うみベ」の最後の1音節を「で」に誤認識したものとする。

【 0 0 2 8 】

次に検索手段3は文字列2である「おーふなうみで」を入力として以下のように検索処理を行う。まず文字列2である「おーふなうみで」を構成する音節バイグラムを全て抽出する。本例では「おーふ」「ふな」「なう」「うみ」「みで」という5個の音節バイグラムを抽出する。次に検索辞書メモリ4に格納しているテキスト検索辞書を参照し、抽出した音節バイグラム毎に当該音節バイグラムを含む施設の検索スコアに1を加算する。抽出した全音節バイグラムに対しこの検索スコア加算処理を行う。本例では、施設ID=1の「A書房大船ウミベ店（えーしょぼーおーふなうみべてん）」は、「おーふ」「ふな」「なう」「うみ」の4個の音節バイグラムが文字列2の音節バイグラムとマッチするので、検索スコアは4となる。一方施設ID=2の「ウミベ大船（うみベおーふな）」は「おーふ」「ふな」「うみ」の3個の音節バイグラムが文字列2の音節バイグラムとマッチするので、検

10

20

30

40

50

索スコアは3となる。上記加算処理終了後、検索手段3は中間検索結果5として、検索スコアが1以上のN個の施設名のID番号と検索スコアの対を出力する。ここでNは1以上の整数である。

【0029】

次に検索順位修正手段6は、文字列2と中間検索結果5を入力とし、中間検索結果5のN個の施設名それぞれに対し当該施設名の形態素辞書を用いて、文字列2と照合することにより文字列2に含まれる形態素を抽出する。抽出した形態素と、当該施設の形態素辞書を比較し、形態素辞書中には存在するが、認識結果の音素列からは抽出されなかった形態素に対し、予め設定したペナルティ値を付与して検索スコアをリスコアリングする。

【0030】

検索順位修正手段6の具体的な処理手順は実施の形態1とほぼ同等である。違いは実施の形態1で述べた検索順位修正手段6の処理手順2における検索結果の施設名の形態素と文字列2との照合処理の方法である。実施の形態1では、形態素辞書中の1個以上の形態素の組み合わせが文字列2と一致するか否かを調べることによって照合処理を行ったが、本実施例では、形態素辞書中の1個以上の形態素の組み合わせと、文字列2との間で音節あるいは音素の置換または脱落または挿入を許したDP(Dynamic Programming)マッチングによる照合処理を行う。そして置換または脱落または挿入の個数が予め定めた所定の個数c以下なら、前記1個以上の形態素が文字列2に含まれると判定し、前記1個以上の形態素を抽出する。本実施の形態では前記所定の個数c=1とする。DPマッチングを用いる理由は、文字列2に音声認識誤りがあり、形態素辞書中の形態素と音節または音素が完全一致しない場合でも、形態素を抽出できるようにするためである。

【0031】

例えばk=1の場合は、k(=1)位の検索結果は施設ID=1の施設名であり、図4に示すとおり形態素辞書中の形態素は、「えーしょぼー」、「おーふな」、「うみべ」、「てん」となる。これらの形態素と音声認識結果である「おーふなうみで」との間でDPマッチングを用いた照合処理を行う。これによって文字列2である「おーふなうみで」から、「おーふな」と「うみべ」の2個の形態素が抽出される。このうち「うみべ」は音声認識結果の文字列2である「おーふなうみで」中には完全一致する音節列が存在しないが、音節「べ」と「で」の置換が1個なので、DPマッチングを行うことによって抽出が可能になる。

【0032】

またk=2の場合は、k(=2)位の検索結果は施設ID=2では、図4に示すとおり形態素辞書中の形態素は「うみべ」、「おーふな」なので、これらの形態素と音声認識結果の文字列2である「おーふなうみで」との間でDPマッチングを行うと「おーふな」と「うみべ」の2個の形態素が抽出される。

手順3以降の処理は実施の形態1と同一なので説明を省略する。

【0033】

以上の処理によって修正検索スコアの大きい順に検索順位を並べ換えた結果を図9に示す。図9によれば「ウミベ大船」が検索順位の第1位になっていることがわかる。なお図7に示した実施の形態1における検索スコアおよび修正検索スコアと比較して、本実施例の検索スコアおよび修正検索スコアの値がそれぞれ1小さいが、これは上述したとおり音声認識結果である文字列2の「おーふなうみで」の最後の1音節「で」は「べ」を誤認識したものであり、その結果検索手段3における検索スコア算出時にマッチする音節パイグラム数が1個少なくなったためである。

【0034】

なお、形態素辞書メモリ7に保持している形態素辞書の各形態素に付与するペナルティ値としては、施設名を構成する最後尾の形態素が「店」である施設名の先頭の形態素に対し、他の形態素よりも大きなペナルティ値を付与してもよい。これは一般に、公園や百貨店内にある施設名は「施設のブランド名等の固有名詞+(公園名または百貨店名)+店」というパターンが多く、最後尾の形態素が「店」である施設名の先頭の形態素は、当該施設を検索する場合に省略することがほぼ無いと考えられるからである。このようにペナル

10

20

30

40

50

ティ値を付与することによりペナルティ付与作業の効率化を図る効果が得られる。

【産業上の利用可能性】

【0035】

この発明は文字列により大量の文書や施設名中から、所望の文書や施設名の大規模な検索を効率よく行う検索装置に関し、携帯端末やカーナビゲーションシステム等各種のナビゲーションシステムに適用が可能である。

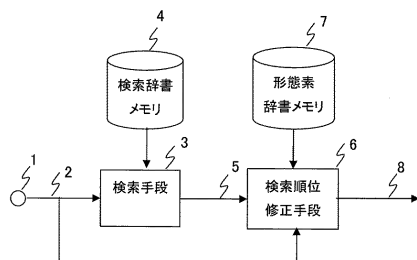
【符号の説明】

【0036】

1、9；文字列の入力端、2；文字列、3；検索手段、4；検索辞書メモリ、5；中間検索結果、6；検索順位修正手段、7；形態素辞書メモリ、8；検索結果、10；入力音声、11；音声認識手段、12；言語モデルメモリ、13；音響モデルメモリ。

10

【図1】



【図2】

施設ID番号	施設名	読み
1	A 書房大船ウミベ店	えーしょぼーおーふなるうみべてん
2	ウミベ大船	うみべおーふな
...

【図3】

索引語	ID 番号
えーしょ	1
しょぼー	1
ぼーおー	1
おーふ	1,2
ふな	1, 2
なう	1,
うみ	1, 2
みべ	1, 2
べて	1
てん	1
べおー	2
...	...

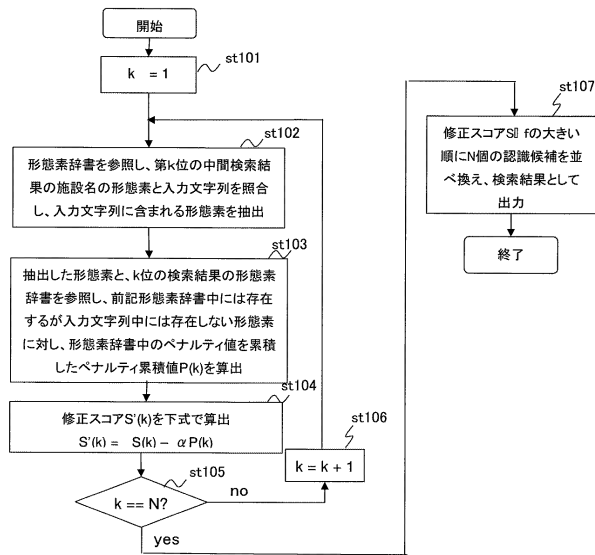
【図4】

施設ID番号	施設名	形態素辞書（ペナルティ値）
1	A 書房大船ウミベ店	えーしょぼー(3), おーふな(1), うみべ(1), てん (0)
2	ウミベ大船	うみべ(3), おーふな(1)
...

【図 5】

施設 I D 番号	検索スコア
1	5
2	4
...	...

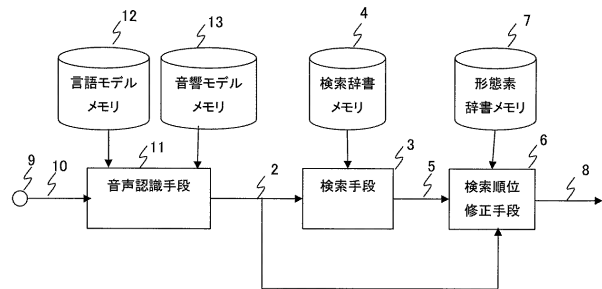
【図 6】



【図 7】

施設 I D 番号	検索スコア	ペナルティ累積値	修正検索スコア
2	4	0	4
1	5	3	3.5
...

【図 8】



【図 9】

施設 I D 番号	検索スコア	ペナルティ累積値	修正検索スコア
2	3	0	3
1	4	3	2.5
...

フロントページの続き

(74)代理人 100161115

弁理士 飯野 智史

(72)発明者 花沢 利行

東京都千代田区丸の内二丁目7番3号 三菱電機株式会社内

(72)発明者 岡登 洋平

東京都千代田区丸の内二丁目7番3号 三菱電機株式会社内

審査官 野崎 大進

(56)参考文献 特開2002-259426(JP,A)

特開平03-116377(JP,A)

特開平06-208588(JP,A)

特開2008-262279(JP,A)

特開2000-250589(JP,A)

特開2008-090401(JP,A)

特開平10-097545(JP,A)

(58)調査した分野(Int.Cl., DB名)

G06F 17/30

JSTPlus(JDreamIII)