

(19) 日本国特許庁(JP)

(12) 特 許 公 報(B2)

(11) 特許番号

特許第5910134号
(P5910134)

(45) 発行日 平成28年4月27日(2016.4.27)

(24) 登録日 平成28年4月8日(2016.4.8)

(51) Int. Cl. F 1
G06F 17/30 (2006.01) G06F 17/30 414B
 G06F 17/30 170A

請求項の数 6 (全 28 頁)

<p>(21) 出願番号 特願2012-24491 (P2012-24491) (22) 出願日 平成24年2月7日(2012.2.7) (65) 公開番号 特開2013-161371 (P2013-161371A) (43) 公開日 平成25年8月19日(2013.8.19) 審査請求日 平成27年2月2日(2015.2.2)</p>	<p>(73) 特許権者 000001443 カシオ計算機株式会社 東京都渋谷区本町1丁目6番2号 (74) 代理人 100095407 弁理士 木村 満 (72) 発明者 佐藤 勝彦 東京都羽村市栄町3丁目2番1号 カシオ 計算機株式会社 羽村技術センター内 審査官 吉田 誠</p>
---	---

最終頁に続く

(54) 【発明の名称】 テキスト検索装置及びプログラム

(57) 【特許請求の範囲】

【請求項1】

複数の文書カテゴリから構成される検索対象となる文書に含まれる文字列と、当該文字列の前記文書における出現位置と、を対応づけた索引情報を記憶している記憶手段と、

検索キーワードを一字ずつ入力するユーザの操作と、検索キーワードを確定するユーザの操作と、を受け付ける操作受付手段と、

前記操作受付手段が検索キーワードを一文字ずつ入力するユーザの操作を受け付ける度に、当該受け付けた操作が示す文字をそれまで受け付けた操作が示す文字又は文字列に加えて検索キーワードとして取得する取得手段と、

前記操作受付手段が検索キーワードを一文字ずつ入力するユーザの操作を受け付ける度に、前記文書のうち前方一致検索の対象となるカテゴリに属する文字列から、前記取得手段が取得した前記検索キーワードから始まる文字列を、前記索引情報を用いて検索する前方一致検索手段と、

前記操作受付手段が前記検索キーワードを確定するユーザの操作を受け付けたことに応じて、前記索引情報を用いて、前記文書から前記確定された検索キーワードを含む文字列を検索する部分一致検索手段と、

前記前方一致検索手段が検索した結果を示す情報と、前記部分一致検索手段が検索した結果を示す情報と、を出力する出力手段と、

を備えることを特徴とするテキスト検索装置。

【請求項2】

10

20

前記前方一致検索手段は、前記索引情報を用いて、前記検索キーワードが前記前方一致検索の対象となるカテゴリに属する文字列のうち、当該検索キーワードが先頭にある文字列を特定して、前記検索キーワードから始まる文字列を検索する、

ことを特徴とする請求項 1 に記載のテキスト検索装置。

【請求項 3】

前記記憶手段は、前記前方一致検索の対象となるカテゴリに属する文字列の先頭に、先頭であることを示すマークである先頭マークを挿入した上で、当該文字列を前記索引情報として記憶し、

前記前方一致検索手段は、前記取得手段が取得した検索キーワードの先頭に前記先頭マークを挿入して、当該先頭マークを含む検索キーワードを、前記先頭マークを挿入した文字列を含む索引情報を用いて前記文書を検索することにより、前記検索キーワードが先頭にある文字列を特定する、

ことを特徴とする請求項 2 に記載のテキスト検索装置。

【請求項 4】

前記索引情報は、前記前方一致検索の対象となるカテゴリに属する各文字列の先頭の位置を示す先頭情報をさらに含み、

前記前方一致検索手段は、前記索引情報を用いて前記検索キーワードの位置を特定し、当該特定した位置と、前記索引情報の先頭情報が示す位置と、を比較して前記検索キーワードが先頭にある文字列を特定する、

ことを特徴とする請求項 2 又は 3 に記載のテキスト検索装置。

【請求項 5】

複数の文書カテゴリから構成される検索対象となる文書に含まれる文字列と、当該文字列の前記文書における出現位置と、を対応づけた索引情報を記憶している記憶手段を備えたコンピュータを、

検索キーワードを一字ずつ入力するユーザの操作と、検索キーワードを確定するユーザの操作と、を受け付ける操作受付手段、

前記操作受付手段が検索キーワードを一字ずつ入力するユーザの操作を受け付ける度に、当該受け付けた操作が示す文字をそれまで受け付けた操作が示す文字又は文字列に加えて検索キーワードとして取得する取得手段、

前記操作受付手段が検索キーワードを一字ずつ入力するユーザの操作を受け付ける度に、前記文書のうち前方一致検索の対象となるカテゴリに属する文字列から、前記取得手段が取得した前記検索キーワードから始まる文字列を、前記索引情報を用いて検索する前方一致検索手段、

前記操作受付手段が前記検索キーワードを確定するユーザの操作を受け付けたことに応じて、前記索引情報を用いて、前記文書から前記確定された検索キーワードを含む文字列を検索する部分一致検索手段、

前記前方一致検索手段が検索した結果を示す情報と、前記部分一致検索手段が検索した結果を示す情報と、を出力する出力手段、

として機能させるためのプログラム。

【請求項 6】

検索キーワードを一字ずつ入力するユーザの操作と、検索キーワードを確定するユーザの操作と、を受け付ける操作受付手段と、

前記操作受付手段が検索キーワードを一字ずつ入力するユーザの操作を受け付ける度に、当該受け付けた操作が示す文字をそれまで受け付けた操作が示す文字又は文字列に加えて検索キーワードとして取得する取得手段と、

前記操作受付手段が検索キーワードを一字ずつ入力するユーザの操作を受け付ける度に、文書のうち前方一致検索の対象となるカテゴリに属する文字列から、前記取得手段が取得した前記検索キーワードから始まる文字列を検索する前方一致検索手段と、

前記操作受付手段が前記検索キーワードを確定するユーザの操作を受け付けたことに応じて、前記文書から前記確定された検索キーワードを含む文字列を検索する部分一致検索

10

20

30

40

50

手段と、

前記前方一致検索手段が検索した結果を示す情報と、前記部分一致検索手段が検索した結果を示す情報と、を出力する出力手段と、
を備えることを特徴とするテキスト検索装置。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、テキスト検索装置及びプログラムに関する。

【背景技術】

10

【0002】

複数の項目（フィールド）を含む文書（辞書等）を検索する装置が知られている。検索装置には、ユーザの利便性を高めるため、検索対象文書に文字列が存在するか否かだけでなく、その出現頻度等に注目して検索結果文書の順位付けを行なうものがある（例えば特許文献1）。特許文献1に記載の技術では、索引（インデックス）を用いて検索する。

【先行技術文献】

【特許文献】

【0003】

【特許文献1】特開平10-049549号公報

【発明の概要】

20

【発明が解決しようとする課題】

【0004】

電子辞書には、検索キーワードを一文字入れるたびに見出語を検索し、見出語一覧表示するインクリメント検索機能と、検索キーワードを入れ終わった後に全文を検索する全文検索（本検索）機能と、二つの検索機能を持つことでユーザの利便性が高めたものがある。このような電子辞書では、一般に前方一致検索でインクリメント検索し、部分一致検索により全文検索する。

【0005】

しかし、特許文献1の技術を用いて前方一致検索を用いる機能と部分一致検索を用いる機能を一つの装置に持たせるためには、それぞれに対応する索引情報を別々に記憶しておく必要がある。そのため、必要な記憶容量が大きくなってしまいうという問題がある。

30

【0006】

本発明はこのような事情に鑑みて成されたものであり、ユーザにとって利便性が高い検索機能を実現でき、かつ必要な記憶容量が小さいテキスト検索装置及びプログラムを提供することを目的とする。

【課題を解決するための手段】

【0007】

上記目的を達成するため、本発明に係るテキスト検索装置は、

複数の文書カテゴリから構成される検索対象となる文書に含まれる文字列と、当該文字列の前記文書における出現位置と、を対応づけた索引情報を記憶している記憶手段と、
検索キーワードを一文字ずつ入力するユーザの操作と、検索キーワードを確定するユーザの操作と、を受け付ける操作受付手段と、

40

前記操作受付手段が検索キーワードを一文字ずつ入力するユーザの操作を受け付ける度に、当該受け付けた操作が示す文字をそれまで受け付けた操作が示す文字又は文字列に加えて検索キーワードとして取得する取得手段と、

前記操作受付手段が検索キーワードを一文字ずつ入力するユーザの操作を受け付ける度に、前記文書のうち前方一致検索の対象となるカテゴリに属する文字列から、前記取得手段が取得した前記検索キーワードから始まる文字列を、前記索引情報を用いて検索する前方一致検索手段と、

前記操作受付手段が前記検索キーワードを確定するユーザの操作を受け付けたことに

50

じて、前記索引情報を用いて、前記文書から前記確定された検索キーワードを含む文字列を検索する部分一致検索手段と、

前記前方一致検索手段が検索した結果を示す情報と、前記部分一致検索手段が検索した結果を示す情報と、を出力する出力手段と、

を備えることを特徴とする。

【発明の効果】

【0008】

本発明によれば、ユーザにとって利便性が高い検索機能を実現でき、かつ必要な記憶容量が小さいテキスト検索装置及びプログラムを提供できる。

10

【図面の簡単な説明】

【0009】

【図1】本発明の実施形態1に係るテキスト検索装置の一例を示す斜視図である。

【図2】実施形態1に係るテキスト検索装置の一構成例を表すハードウェア構成図である。

。

【図3】実施形態1に係るテキスト検索装置が実行するデータ等生成処理の一例を表すフローチャートである。

【図4】実施形態1に係るテキスト検索装置が有する機能の一例を表す機能ブロック図である。

【図5】(a)は、テキスト検索装置が記憶する辞書データの一例を表す図である。(b)は、テキスト検索装置が記憶する再配置コンテンツテキストデータの一例を表す図である。

20

【図6】実施形態1に係るテキスト検索装置が表示するインクリメント検索の結果表示画面の一例を表す図である。

【図7】実施形態1に係るテキスト検索装置が表示する本検索の結果表示画面の一例を表す図である。

【図8】実施形態1に係るテキスト検索装置が記憶する辞書テーブルの一例を表す図である。

【図9】実施形態1に係るテキスト検索装置が記憶する転置インデックスを含む電子ファイルの一例を表す図である。

30

【図10】実施形態1に係るテキスト検索装置が実行するテキスト検索処理の一例を表すフローチャートである。

【図11】実施形態1に係るテキスト検索装置が実行する前方一致検索処理の一例を表すフローチャートである。

【図12】実施形態1に係るテキスト検索装置が実行する本検索処理の一例を表すフローチャートである。

【図13】実施形態1に係るテキスト検索装置が記憶する決定方法テーブルの一例を表す図である。

【図14】実施形態1に係るテキスト検索装置が実行する単一文字列検索処理の一例を表すフローチャートである。

40

【図15】実施形態1に係るテキスト検索装置が実行する出現位置取得処理の一例を表すフローチャートである。

【図16】実施形態1に係るテキスト検索装置が実行する複数文字列検索処理の一例を表すフローチャートである。

【図17】(a)は、検証文字列の特定出現位置が基準文字列の特定出現位置以降である場合における検証文字列と基準文字列との距離の一例を表す図である。(b)は、検証文字列の特定出現位置が基準文字列の特定出現位置よりも先である場合における検証文字列と基準文字列との距離の一例を表す図である。

【図18】検索キーワードの最小包含範囲の最小値と最大値との一例を表す図である。

【発明を実施するための形態】

50

【0010】

以下、本発明の実施形態に係るテキスト検索装置100について、添付図面を参照しつつ説明する。

【0011】

本発明の実施形態に係るテキスト検索装置100は、図1に示すような電子辞書で構成され、ユーザの操作に応じて検索キーワードを入力するキーボード100iと、検索キーワードに基づいて辞書を検索した検索結果を表示するLCD(Liquid Crystal Display)100hと、を備える。

【0012】

テキスト検索装置100の内部には、図2に示すようなCPU(Central Processing Unit)100a、ROM(Read Only Memory)100b、RAM(Random Access Memory)100c、ハードディスク100d、メディアコントローラ100e、ビデオカード100g、及びスピーカ100jが内蔵され、図1に示したLCD100h及びキーボード100iとバスを介して接続されている。

10

【0013】

CPU100aは、ROM100b若しくはハードディスク100dに保存されたプログラムにより、下記する検索処理のためにテキスト検索装置100の各部を制御し、また内部処理を実行する。RAM100cは、CPU100aによるプログラムの実行時において、処理対象とするデータを一時的に記憶する等、作業領域として用いられる。

【0014】

20

ハードディスク100dは、各種のデータを保存したテーブルと、英和辞書等の辞書データと、を記憶する。尚、テキスト検索装置100は、ハードディスク100dの代わりに、フラッシュメモリを備えても良い。

【0015】

メディアコントローラ100eは、フラッシュメモリ、CD(Compact Disc)、DVD(Digital Versatile Disc)、及びブルーレイディスク(Blu-ray Disc)(登録商標)を含む記録媒体から各種のデータ及びプログラムを読み出す。

【0016】

ビデオカード100gは、CPU100aから出力されたデジタル信号に基づいて画像を描画(つまり、レンダリング)すると共に、描画された画像を表す画像信号を出力する。LCD100hは、ビデオカード100gから出力された画像信号に従って画像を表示する。尚、テキスト検索装置100は、LCD100hの代わりに、PDP(Plasma Display Panel)若しくはEL(Electroluminescence)ディスプレイを備えても良い。スピーカ100jは、CPU100aから出力された信号に基づいて音声を出力する。

30

【0017】

ユーザが、辞書データを記録した記録媒体を図2に示したメディアコントローラ100eへ当該記録媒体を挿入すると、CPU100aは、メディアコントローラ100eから辞書データの信号を受信する。CPU100aは、メディアコントローラ100eから辞書データを取得すると、辞書データをハードディスク100dに保存する。その後、CPU100aは、辞書データで表される辞書を検索キーワードに基づいて検索するために用いられるデータ及び電子ファイルを生成する、図3に示すようなデータ等生成処理を実行する。これにより、CPU100aはテキスト検索装置100のその他の構成要素と協働して、図4に示すような生成部120として機能する。図4は、CPU100aが実行する機能を示す機能ブロック図である。そして、CPU100aは、テキスト検索装置100のその他の構成要素、特にハードディスク100dと協働することで、情報記憶部110として機能する。

40

【0018】

情報記憶部110は、本実施形態で検索対象となる辞書データを記憶する。この辞書データは、図5(a)に示すように、見出語を表すテキスト(以下、見出語テキストという)と、見出語の解説を表すテキスト(以下、解説テキストという)と、例えば、成句や複

50

合語など（以下、イディオムという）といった見出語の用例を表すテキスト（以下、用例テキストという）と、で構成される。解説テキストと用例テキストとを合わせて本文テキストと言う。言い換えれば、辞書データに含まれるテキストは、見出語を示すテキスト（見出語テキスト）と、見出語を説明するための本文（本文テキスト）と、の2つのカテゴリ（見出語カテゴリ、本文カテゴリ）に分類できる。本文カテゴリに属するテキストは、さらに見出語の解説と、その用例と、の二つ（解説カテゴリ、用例カテゴリ）に分類できる。

また、辞書データのうち、見出語テキストが占める部分を見出部 C E、本文テキストがしめる部分を本文部 C B という。

解説テキストは、当該テキストで表される内容が解説であることを表す解説タグに囲まれており、用例テキストは、当該テキストで表される内容が用例であることを表す用例タグに囲まれている。

【 0 0 1 9 】

辞書データは見出部 C E と本文部 C B との組を一つの構成単位として、この構成単位が連なって構成されている。各構成単位は、見出部 C E に配置された見出語テキストのアルファベット順に並べられている。各構成単位において、見出語テキストを含む見出部 C E の直後に、その見出語を説明するための本文テキストを含む本文部 C B が配置されている。

また、各見出部 C E には、見出部 C E を識別する見出語番号が予め割り当てられている。辞書データには、この見出語番号を表す情報と、当該見出語番号で識別される見出部 C E が格納された情報記憶部 1 1 0 における記憶領域の先頭アドレスを表す情報と、当該見出部 C E の直後に格納された本文部 C B の先頭アドレスを表す情報と、が対応付けられた情報が、見出語の数だけ含まれる。

【 0 0 2 0 】

本文部 C B には、複数の解説テキストが、電子辞書の編集者が定めた並び順に従って配置されている。具体例としては、見出語のより一般的な意味内容を解説する解説テキストの方が、見出語のより特殊な意味内容を解説する解説テキストよりも先の位置に格納されている。あるいは、より使用頻度の高い意味内容を解説する解説テキストの方が、より使用頻度の低い意味内容を解説する解説テキストよりも先の位置に格納されているとしても良い。

【 0 0 2 1 】

本実施形態のテキスト検索装置 1 0 0 は、情報記憶部 1 1 0 に記憶された辞書データから検索キーワードを次の二つの方法で検索して、検索結果をユーザに提示する。

第 1 に、テキスト検索装置 1 0 0 は、ユーザがキーボード 1 0 0 i を用いて検索キーワードを一文字入力する毎に、その時点のキーワードを前方一致検索して、キーワードを先頭に含む見出語を表示する。この検索をインクリメント検索と呼ぶ。また、テキスト検索装置 1 0 0 がインクリメント検索を実行している状態を、インクリメント検索モードと呼ぶ。インクリメント検索の実行例を図 6 に示す。

【 0 0 2 2 】

ユーザが、検索語の先頭の一文字（図 6 の例では w）を入力すると検索対象となる複数の辞書データ（ここでは、英英辞典、英和辞典、百科事典）において、先頭が w から始まる見出語を検索する。そして、図 6 の左側に示したように、見出語一覧をユーザが選択可能な態様で提示する。

次の一文字（ここでは h）を入力すると、同様に w h を先頭に持つ見出語を検索して表示する（図 6 の中央）。以下、3 文字以上入力した場合についても、同様に検索して表示する（例えば図 6 の右）。ユーザは候補として所望の見出語（ここでは while）が表示されると、いつでも所望の見出語を選択してその本文テキストを表示できるようにすることができる。

なお、図 6 ではインクリメント検索の結果を各辞書データ（英英辞典、英和辞典、百科事典）毎に分けて表示する例を紹介した。しかし、この例に限られず、すべての検索結果

10

20

30

40

50

を文字順にソートしてまとめて表示する構成も可能である。このとき、重複する単語については一つだけ表示するとしても良い。

【 0 0 2 3 】

第 2 に、本実施形態のテキスト検索装置 1 0 0 は、ユーザがキーボード 1 0 0 i を用いて検索キーワードを入力して確定操作を実行すると、検索キーワードを含む見出テキスト・解説テキスト・用例テキストを部分一致検索して表示する（図 7）。この検索を本検索と呼ぶ。また、テキスト検索装置 1 0 0 が本検索を実行している状態を本検索モードと呼ぶ。

本実施形態では、インクリメント検索と本検索を一つの索引情報（転置インデックス）を用いて実行する。

【 0 0 2 4 】

ここで、インクリメント検索と本検索を一つの索引（転置インデックス）を用いて検索するための準備処理であるデータ等生成処理について、図 3 を参照して説明する。生成部 1 2 0 は、辞書データを取得すると図 3 に示すデータ等生成処理を開始する。

【 0 0 2 5 】

データ等生成処理では、まず生成部 1 2 0 が情報記憶部 1 1 0 に記憶された辞書コンテンツを表す辞書データを読み出す（ステップ S 1 0 1）。

【 0 0 2 6 】

そして、読み出した辞書データから、見出語テキストと本文テキスト（解説テキストと用例テキスト）とを抽出する（ステップ S 1 0 2）。具体的に説明する。本文部 C B には、解説テキストと用例テキストとが混在しているので、まず生成部 1 2 0 は、これらのテキストを内容に応じて分類分けをする。そのために、生成部 1 2 0 は、辞書データに含まれる見出語番号を表す情報と、見出部 C E の先頭アドレスを表す情報と、本文部 C B の先頭アドレスを表す情報と、を用いて、辞書データから見出語テキストと本文テキストとを複数抽出する。

さらに、生成部 1 2 0 はステップ S 1 0 2 で、抽出された見出語テキスト毎に、見出語テキストで表される見出語を解説する解説テキストを本文テキストから解説タグに基づいて複数抽出し、かつ見出語の用例を表す用例テキストを用例タグに基づいて本文テキストから複数抽出する。

【 0 0 2 7 】

そして、インクリメント検索の対象となるカテゴリのテキスト（ここでは見出語）の先頭に、先頭であることを示すマーク（先頭マーク、ここでは半角スペース）を挿入する（ステップ S 1 0 3）。

【 0 0 2 8 】

その後、生成部 1 2 0 は、図 5（b）に示すような再配置コンテンツテキストデータ（C T D）を生成して記憶する（ステップ S 1 0 4）。具体的には、見出語テキストに先頭マーク（ここでは「・」で代用して表示）を抽出した上で、構成単位ごとに抽出された解説テキストの並び順を変更せずにまとめて配置することで、抽出された複数の解説テキストが配置される部位（以下、解説部 C C という）を生成する。

同様に、生成部 1 2 0 は、構成単位毎に、抽出された複数の用例テキストの並び順を変更せずにまとめて配置することで、抽出された複数の用例テキストが配置された部位（以下、用例部 C X という）を生成する。

【 0 0 2 9 】

生成部 1 2 0 は、このようにして見出部 C E と、解説部 C C と、用例部 C X と、を対応付けた複数のデータ（再配置コンテンツテキストデータ、若しくは再配置 C T D）を生成する。そして、生成部 1 2 0 は、生成した再配置コンテンツテキストデータを情報記憶部 1 1 0 へ保存する（ステップ S 1 0 4）。

【 0 0 3 0 】

その後、生成部 1 2 0 は、再配置コンテンツテキストデータで表される辞書を識別する辞書番号を生成する。次に、生成部 1 2 0 は、図 8 に示す辞書テーブルへ、生成された辞

10

20

30

40

50

書番号を表す情報と、辞書の名称を表す情報と、再配置コンテンツテキストデータが格納された情報記憶部 110 の記憶領域の先頭アドレスを表す情報と、を対応付けた情報を情報記憶部 110 に保存する (ステップ S105)。

【0031】

その後、生成部 120 は、再配置コンテンツテキストデータで表されるテキスト (つまり、見出部 CE、解説部 CC、及び用例部 CX に配置されたテキスト) のそれぞれについて、先頭から 1 文字ずつずらしながら 1 文字を切り出すことで、モノグラム文字列パターンを抽出する。また、同様に、生成部 120 は、再配置コンテンツテキストデータで表されるテキストの先頭から 1 文字ずつずらしながら 2 文字を切り出すことで、バイグラム文字列パターンを抽出する。以下、モノグラム文字列パターン、バイグラム文字列パターンを、N グラム文字列パターンと総称する。尚、本実施形態では、文字は文字列に含まれるとし、特別に区別する記載がある場合を除き、文字と文字列とを区別しない。

10

【0032】

次に、生成部 120 は、各 N グラム文字列パターンについて、N グラム文字列パターンが再配置コンテンツテキストデータで表されるテキスト (以下、再配置テキストという) に出現する位置 (以下、出現位置という) を 1 又は複数特定する。その後、生成部 120 は、N グラム文字列パターン毎に、再配置テキストにおける N グラム文字列パターンの出現頻度を算出する。ここで、出現頻度は、例えば、再配置テキストに N グラム文字列パターンが現われる総回数であるとして説明するが、これに限定される訳ではない。

【0033】

20

次に、生成部 120 は、N グラム文字列パターン毎に、N グラム文字列パターンが出現する出現位置を示すアドレス (以下、出現位置アドレスという) を表す 1 又は複数の情報と、当該 N グラム文字列パターンの出現頻度を表す情報と、を対応付けた出現位置情報を生成する。

【0034】

次に、生成部 120 は、出現位置情報を 1 又は複数含む電子ファイル (以下、出現位置情報ファイル若しくは AP ファイルという) を生成し、情報記憶部 110 へ保存する (図 3 のステップ S106)。

図 9 では、出現位置情報ファイル (AP ファイル) は「position.idx」という名称 (ファイル名) で示されている。また、出現位置情報ファイルに保存された出現位置情報は、先頭アドレスから出現頻度用の所定バイト数までの領域に出現頻度を表す情報が格納され、当該領域の直後から出現位置用の所定バイト数毎に出現位置アドレスを表す情報が格納される。

30

【0035】

次に、生成部 120 は、N グラム文字列パターンを表す情報 (以下、N グラム文字列パターン情報という) と、当該 N グラム文字列パターンの出現位置情報が格納された情報記憶部 110 の記憶領域の先頭アドレス (以下、出現位置情報格納アドレスという) を表す情報と、を対応付けた情報を複数含む電子ファイル (以下、N グラム文字列パターンファイル若しくは S ファイルという) を生成する。その後、生成部 120 は、S ファイルを情報記憶部 110 へ保存する (ステップ S107)。

40

図 9 では、S ファイルは「pattern.idx」という名称 (ファイル名) で示されている。

【0036】

その後、生成部 120 は、見出語毎に、見出語の用例を表す用例テキストが分類された用例部 CX の先頭アドレスと、当該用例テキストの先頭アドレスと、の差分 (以下、用例部開始位置からの差分という) を、用例部 CX に分類された用例テキストの個数 (以下、用例個数という) だけ算出する。次に、生成部 120 は、見出語毎に、1 又は複数の用例部開始位置からの差分を表す情報で構成される用例開始位置情報を生成し、生成された用例開始位置情報を複数含む電子ファイル (以下、用例開始位置情報ファイル若しくは EP ファイルという) を生成する。

図 9 では、EP ファイルは「example.idx」という名称 (ファイル名) で示されている

50

その後、生成部 120 は、用例開始位置情報ファイルを情報記憶部 110 へ保存する（図 3 に示すステップ S108）。尚、EP ファイルに保存された用例開始位置情報は、用例部開始位置からの差分を表す情報が、先頭アドレスから当該差分用の所定バイト数までの領域に格納され、当該領域の直後から所定バイト数毎に他の差分を表す情報が格納される。

【0037】

次に、生成部 120 は、見出語テキストを含む見出部 CE を識別する見出番号と、当該見出部 CE が格納された情報記憶部 110 の記憶領域の先頭アドレス（以下、見出部 CE の開始位置を表すアドレスという）を表す情報と、当該見出語テキストで表される見出語を解説する解説テキストが配置された解説部 CC の先頭アドレス（以下、解説部 CC の開始位置を表すアドレスという）を表す情報と、当該見出語の用例を表す用例テキストが配置された用例部 CX の先頭アドレス（以下、用例部 CX の開始位置を表すアドレスという）を表す情報と、当該用例部 CX の用例個数を表す情報と、当該用例部 CX の用例部開始位置情報（EP ファイル、図 9 の example.idx）が情報記憶部 110 に格納された領域の先頭位置を表すアドレス（以下、用例部開始位置情報格納アドレスという）を表す情報と、当該見出語を掲載した辞書の辞書番号を表す情報と、を対応付けた情報を複数含む電子ファイル（以下、見出等ファイル若しくは T ファイルという）を生成する。

図 9 では、T ファイルは「number.idx」という名称（ファイル名）で示されている。

生成部 120 は、T ファイルを情報記憶部 110 へ保存した後に（図 3 に示すステップ S109）、データ等生成処理の実行を終了する。

【0038】

図 9（特に「pattern.idx」と「position.idx」）に示すような、文字列（ここでは N グラム文字列パタン）と、文字列の検索対象となる文書における出現位置とを、対応付けて記録した索引構造を転置インデックスと呼ぶ。本実施形態では、データ等生成処理により、インクリメント検索の対象となるカテゴリのテキストについて、先頭に先頭マークを付した再配置 CTD の転置インデックスを生成する。

【0039】

転置インデックスを情報記憶部 110 に記憶すると、この転置インデックスを用いて検索処理を実行する準備が整う。この状態で、ユーザが辞書の検索に用いる検索キーワードを入力する操作をキーボード 100i を用いて実行すると、CPU 100a がその情報を受け取り、図 10 に示すテキスト検索処理を実行する。

このテキスト検索処理を実行することにより、CPU 100a は図 4 に示すような取得部 130、インクリメント検索部 1410 と本検索部 1420 を含む検索部 140、算出部 150、決定部 160、及び出力部 170 として機能する。また、CPU 100a は、図 2 に示したビデオカード 100g 及び LCD 100h と協働して表示部 180 として機能する。

【0040】

ユーザがキーボード 100i を用いて、検索キーワードの入力を始めると、テキスト検索装置は待機モードからインクリメント検索モードに移行し、CPU 100a が図 10 に示すテキスト検索処理を開始する。テキスト検索処理では、まず取得部 130 が現在までに入力されたキーワード（入力キーワード）を取得する（ステップ S201）。

【0041】

そして、インクリメント検索部 1410 が、上述した複数の転置インデックスを用いて、入力キーワードが含まれる見出語を、前方一致検索により抽出する前方一致検索処理を実行する（ステップ S202）。

【0042】

ステップ S202 で実行される前方一致検索処理について、図 11 を参照して説明する。前方一致検索処理では、まず取得した入力キーワードの先頭に、先頭マーク（ここでは半角スペース）を挿入する（ステップ S301）。例えば、ステップ S201 で取得した

10

20

30

40

50

入力キーワードが「whi」の場合には、先頭に半角スペース（記号・で示す）を挿入した「・whi」を、以下の処理に用いる入力キーワードとする。以下、この例に沿って説明する。

【0043】

そして、検索部140（インクリメント検索部1410）は、取得された検索キーワードからNグラム文字列パターンを生成し、生成したNグラム文字列パターンを検索パターンとする（ステップS302）。

Nの値は、設定値（転置インデックスに対応するNの値）及び入力キーワードの文字数で変化する。例えば、入力キーワードが「・whi」であり、設定値がN=2である場合には、バイグラム文字列パターン「・w」、「wh」、「hi」を検索パターンとして生成する。N=3の場合は、「・wh」、「whi」を生成する。なお、（マーク挿入後の）入力キーワードの文字数が設定値以下である場合（例えば設定値がN=3であり、入力キーワードが「・w」である場合）、入力キーワードの文字数（ここでは2）をNの値とする。そして、一つのNグラム（バイグラム「・w」）のみからなるNグラム文字列パターンを生成する。ここでは、N=2とし、「・w」、「wh」、「hi」を生成するとする。

【0044】

そして、ステップS302で生成した検索パターンが一つであるか複数であるか判別する（ステップS303）。

【0045】

検索パターンが複数生成された場合（ステップS303；NO）、インクリメント検索部1410は、検索パターンの各出現位置を特定して、その中で注目出現位置を特定する（ステップS304）。

ここでは、生成した複数の検索パターンのうち最も出現頻度が少ない検索パターンの出現位置を注目出現位置とする。具体的には、まずインクリメント検索部1410がAPファイル（つまり、出現位置情報ファイル）及びSファイル（つまり、Nグラム文字列パターンファイル）から、検索パターンの出現頻度を検索する。即ち、インクリメント検索部1410は、Nグラム文字列パターンファイル（図9の「pattern.idx」）から、生成した検索パターンを表す情報に対応付けられた出現位置情報格納アドレスを表す情報を検索する。そして、インクリメント検索部1410は、出現位置情報ファイル（図9の「position.idx」）から、検索された情報で表される出現位置情報格納アドレスが指し示す位置を開始位置として、情報記憶部110に格納された出現位置情報を取得する。次に、インクリメント検索部1410は、取得した出現位置情報の先頭アドレスから上記出現頻度用の所定バイト数までの領域に格納された出現頻度を表す情報を抽出する。

【0046】

その後、インクリメント検索部1410は、生成した複数の検索パターンの内で、最少の出現頻度に対応付けられた検索パターンを注目検索パターンとして特定する。ここでは、「wh」を注目検索パターンとする。出現頻度が最も少ない検索パターンに基づいて辞書から検索キーワードを検索することにより、下記するループ（図11のステップS304からステップS316までのループ）が少ない回数で終了するので、より高速にインクリメント検索処理を実行出来るからである。

【0047】

S304ではさらに、インクリメント検索部1410が注目検索パターンの出現位置情報における出現頻度を表す情報が格納された領域の直後から上記出現位置用の所定バイト数を読み込み、注目検索パターンの最初の出現位置を注目出現位置として選択する（ステップS304）。

【0048】

そして、インクリメント検索部1410は注目検索パターン以外の検索パターン（他の検索パターン）の出現位置（特定出現位置）を特定する（ステップS305）。

具体的には、まず情報記憶部110が予め記憶する所定範囲を表す情報を取得する。さらに、インクリメント検索部1410は、注目検索パターン注目検索パターン「wh」以外の

10

20

30

40

50

検索パターン「・w」、「hi」の出現位置の中で、注目出現位置から所定範囲までの出現位置を、それぞれ特定出現位置とする（ステップS305）。

【0049】

次に、インクリメント検索部1410は、注目出現位置と、他の検索パタンの特定出現位置と、の連続性を評価して（ステップS306）、連続性の有無を判別し（ステップS307）、注目出現位置を含む位置に入力キーワードがあるか判別する。

具体的には、インクリメント検索部1410はステップS306で注目検索パターンと他の検索パターンについて、入力キーワードにおける相対位置（距離及び前後関係）を取得する。例えば、注目検索パターン「wh」は、入力キーワード（マーク挿入後）「・whi」において、他の検索パターン「・w」の一文字後にあり、「hi」の一文字前に位置する。相対位置とは、このような「n文字後（前）に位置する」という位置情報である。

10

同様に、ステップS305でSファイルとAPファイルを参照して取得した注目出現位置と特定出現位置を比較して、再配置コンテンツテキストデータ上の注目検索パターンと他の検索パターンとの相対位置を取得する。そして、全ての他の検索パターンと注目検索パタンの相対位置が、入力キーワードと再配置コンテンツテキストデータとで一致した場合に、連続性があると評価される。一方、何れかの他の検索パターンと注目検索パタンの相対位置が、入力キーワードと再配置コンテンツテキストデータとで異なる場合、連続性は無いと評価される。このようにして、注目出現位置における連続性を評価する（ステップS306）。

【0050】

20

インクリメント検索部1410は、ステップS306における連続性の評価で、連続性ありと評価されたか否か判別する（ステップS307）。連続性なしと評価された場合は（ステップS307；NO）、その注目出現位置については検索ヒット部位とせず、処理を終了する。そして、注目検索パタンの出現位置の全てについて上記処理を実行したか否かを判別する（ステップS308）。

【0051】

注目検索パタンの出現位置の全てについて上記処理済みではないと判別すると（ステップS308；No）、ステップS304に戻り、次の出現位置を選択して上記の処理を繰り返す。

【0052】

30

一方、全出現位置について上記処理済みであると判別すると（ステップS308；Yes）、ステップS318にスキップする。

【0053】

一方、ステップS307で連続性ありと評価された場合は（ステップS307；Yes）、現在の注目出現位置について、ステップS302で生成した全検索パタンのうち先頭にある検索パターン（本例では「・w」）の特定出現位置を、入力キーワード（本例では「・whi」）の再配置テキストにおける出現位置（検索ヒット部位）として特定する（ステップS309）。

【0054】

一方、ステップS303で検索パターンが一つと判別された場合（ステップS303；Yes）は、その一つの検索パターンを注目検索パターンとしてステップS304と同様に出現位置を求め、求めた出現位置のうち未処理のものを注目出現位置として選択する（ステップS310）。そして、注目出現位置を入力キーワードの再配置テキストにおける出現位置として特定する（ステップS311）。

40

【0055】

ステップS309又はステップS311で出現位置を特定すると、インクリメント検索部1410は、Tファイル（図9の「number.idx」）及び用例開始位置情報ファイル（EPファイル、図9の「example.idx」）と、を参照して、出現位置が、どの構成単位のどの部位（見出部CE、解説部CC、及び用例部CXのいずれか）に含まれるか特定する（ステップS312）。

50

【 0 0 5 6 】

具体的には、インクリメント検索部 1 4 1 0 は、入力キーワード（本例では「・whi」）について特定した出現位置を表すアドレス以前のアドレスが指し示す位置から開始する見出部 C E であって、開始位置を指し示すアドレスが最大の見出部 C E を特定する。特定された見出部 C E を特定見出部と呼ぶ。即ち、特定見出部とは、出現位置が属する構成単位の見出部 C E である。

【 0 0 5 7 】

次に、インクリメント検索部 1 4 1 0 は、T ファイル（見出等ファイル、図 9 の「number.idx」）を参照して、特定見出部の開始位置を指し示すアドレスを表す情報、当該特定見出部に分類された見出語テキストが表す見出語を解説する解説テキストが分類された解説部（以下、当該特定見出部に対応する解説部という）C C の開始位置を指し示すアドレスを表す情報、当該特定見出部に対応する用例部 C X の開始位置を指し示すアドレスを表す情報、当該特定見出部に対応する用例開始位置情報格納アドレスを表す情報、当該見出語を掲載した辞書を識別する辞書番号を表す情報、及び当該特定見出部を識別する見出番号を表す情報を取得する。これらの情報から、特定見出部を示す識別子（見出番号）と、各部位の位置を特定できる情報を生成する。特定見出部を含む構成単位を特定構成単位と呼ぶ。また、各部位の位置から、特定した出現位置が特定構成単位のどの部位に位置するか定めることができる。この部位を特定部位とする。

【 0 0 5 8 】

ステップ S 3 1 2 では、さらに、特定部位が見出部 C E である場合に、算出部 1 5 0 が下記式（1）を用いて出現位置の評価値（ $Est_{idxid, cgy}$ ）を算出する。

【 0 0 5 9 】

【数 1】

$$Est_{idxid, cgy} = Est_{idxid, 0} = Pos_{stdstr, 0} - PosSt_{idxid, headline, hdlid} \dots (1)$$

【 0 0 6 0 】

ただし、 $idxid$ は得られた見出番号、 cgy は特定部位の識別子である。 cgy には、見出部 C E について 0 が、解説部 C C について 1 が、用例部 C X について 2 が、それぞれ割り振られている。ここでは、特定した位置が見出部 C E 以外である場合には評価値を算出しないため、 cgy は 0 のみを取る。また、 $Pos_{stdstr, 0}$ は入力キーワードの特定した出現位置、 $PosSt_{idxid, headline, hdlid}$ は見出番号 $idxid$ の見出部 C E の開始位置を示す。即ち、評価値（ $Est_{idxid, cgy}$ ）は、特定した出現位置が、見出語の先頭から何文字離れているかを示す値である。評価値（ $Est_{idxid, cgy}$ ）は、見出語に先頭マーク（半角スペース）が含まれている場合には 0 にならない場合がある。

【 0 0 6 1 】

そして、決定部 1 6 0 が特定した出現位置が見出部 C E の先頭にあるか判別する（ステップ S 3 1 3）。即ち、特定部位が見出部 C E であり、且つ評価値が 0 である場合は（ステップ S 3 1 3；Yes）、特定した出現位置は見出部 C E の先頭であるため、その見出部 C E を前方一致検索のヒット部位とみなせる。そのため決定部 1 6 0 がこの見出部 C E を検索結果に反映する（ステップ S 3 1 4）。一方、特定部位が見出部 C E でないか、あるいは評価値が 0 でない場合は（ステップ S 3 1 3；No）、特定した出現位置は検索結果から除外され（ステップ S 3 1 5）、以下の処理の対象とならない。

【 0 0 6 2 】

次に、インクリメント検索部 1 4 1 0 は、ステップ S 3 0 4 で求めた注目検索パタンの出現位置の全てについて、注目出現位置として上記処理を実行したか否かを判別する（ステップ S 3 1 6）。未処理の出現位置が残っている場合（ステップ S 3 1 6；No）、次の出現位置について処理を繰り返す。このとき、まずステップ S 3 0 3 で検索パターンが一つであるか否かを判別する（ステップ S 3 1 7）。複数であった場合（ステップ S 3 1 7；

10

20

30

40

50

No)、ステップS304に戻る。一方、一つであった場合(ステップS317; Yes)、ステップS310に戻る。

【0063】

一方、全ての出現位置について上記処理済みであった場合(ステップS316; Yes)、ステップS314で検索結果に反映した見出部CEを、見出し番号順にソートする(ステップS318)。この処理を、図8に示した辞書テーブルに登録された各辞書データについて実行する。

【0064】

図10に戻って、ステップS202で前方一致検索して見出語を抽出すると、出力部170がこの結果を示す情報を表示部180に出力して、図6に示したように検索結果を表示する。(ステップS203)。

そして、検索キーワードが確定したか判別する(ステップS204)。具体的には、現在の入力キーワードを検索キーワードとして確定する操作(例えば実行キーの押下)を検出すると、現在の入力キーワードを検索キーワードとする。この操作を検出しない場合(ステップS204; NO)、新たに一文字追加する操作を待ち受ける。そして、新たな操作を受け付けたことにより、新たな入力キーワードが得られた場合には、その入力キーワードについてステップS201から処理を繰り返す。

【0065】

一方、上記の確定操作を検出して検索キーワードが確定すると(ステップS204; Yes)、テキスト検索装置100は本検索モードに移行し、検索キーワードを用いた本検索処理を実行する(ステップS205)。

【0066】

本検索処理では、再配置コンテンツデータに含まれる全テキストについて、部分一致検索を実行し(全文検索)、ヒットした部位を優先度順に表示する(図7)。図7の例では、検索結果を各カテゴリ毎に優先度をつけて表示している。この処理は、検索対象となる文書を全文して、部分一致した結果に優先度を付けて表示する既知の任意の方法を用いて実行して良い。ここでは、以下のような処理によって本検索を実行する。なお、必ずしも全文を検索する必要は無く、例えば検索対象となる文書の範囲(辞書の種類やカテゴリ、部位等)をユーザが選択できるとしても良い。

【0067】

ステップS205で実行される本検索処理を、図12を参照して説明する。

【0068】

図12に示すテキスト検索処理では、まず取得部130が確定した検索キーワード(1又は複数の検索キーワード)を取得する(ステップS11)。ここでは、確定した入力キーワードが区切り文字(スペース等)を含む場合、その区切り文字で入力された文字を区切って、複数の検索キーワードを取得したとする。以降、1つの検索キーワード「while」が取得された場合を例に挙げて説明を行う。

【0069】

ステップS11の後に、本検索部1420は、図11のステップS302と同様に、取得された検索キーワード(本例では「while」)からNグラム文字列パターン(「wh」、「hi」、「il」、「le」)を生成し、生成したNグラム文字列パターンを検索パターンとする(ステップS12)。なお、取得した検索キーワードが複数である場合は、各キーワードについてNグラム文字列パターンを生成する。

【0070】

次に、本検索部1420は、検索パターン毎に、図11のステップS304で説明した方法で、各検索パターンの出現頻度を取得する(ステップS13)。そして、出現頻度が最も少ない検索パターンを特定する(ステップS14)。このような検索パターンに基づいて辞書から検索キーワードを検索すれば、出現頻度がより多い検索パターンに基づいて辞書を検索するよりもテキスト検索処理が短時間で終了するためである。ここでは、「il」の出現頻度が最も少ないとして説明を行う。

10

20

30

40

50

【 0 0 7 1 】

その後、本検索部 1 4 2 0 は、ステップ S 1 1 で取得した検索キーワードが一つであるか、複数であるか判別する（ステップ S 1 5）。検索キーワードが一つである場合（ステップ S 1 5 ; Y e s）、その検索キーワードを基準文字列（優先度を定めるにあたって基準となる検索キーワード）とする（ステップ S 1 6）。そして、単一文字列検索処理を実行する（ステップ S 1 7）。

一方、検索キーワードが複数である場合（ステップ S 1 5 ; N o）、ステップ S 1 2 で生成した検索パタンのうち、最少の出現頻度の検索パターンを含む検索キーワードを基準文字列とし、基準文字列以外のキーワードを検証文字列とする（ステップ S 1 8）。検索に要する計算量を削減するためである。そして、複数の検索キーワードを検索する処理（複数文字列検索処理）を実行する（ステップ S 1 9）。ステップ S 1 7 乃至ステップ S 1 9 の検索処理を終えると、図 1 2 の本検索処理は終了する。

【 0 0 7 2 】

ステップ S 1 7 で実行される単一文字列検索処理を、図 1 4 を参照して説明する。単一文字列検索処理では、まず本検索部 1 4 2 0 が図 1 2 のステップ S 1 2 で生成した、基準文字列の N グラム文字列パターンを検索パターンとして取得する（ステップ S 2 1）。

【 0 0 7 3 】

そして、基準文字列の再配置 C T D における出現位置を全て取得する処理（出現位置取得処理）を実行する（ステップ S 2 2）。

【 0 0 7 4 】

出現位置取得処理について、図 1 5 を参照して説明する。まず出現位置を取得すべき対象となる文字列（対象文字列、ここでは基準文字列）について生成された検索パターンが一つであるか判別する（ステップ S 3 1）。検索パターンが複数生成された場合（ステップ S 3 1 ; N O）、本検索部 1 4 2 0 は、図 1 1 のステップ S 3 0 4 ~ ステップ S 3 0 7 と同様にステップ S 3 2 ~ ステップ S 3 5 を実行する。即ち、注目検索パタンの出現位置を求めて注目出現位置を選択し（ステップ S 3 2）、他の検索パタンの出現位置（特定出現位置）を求め（ステップ S 3 3）、その連続性を評価して（ステップ S 3 4）、連続しているか判別する（ステップ S 3 5）。

【 0 0 7 5 】

そして、ステップ S 3 5 で注目出現位置において連続性が無いと判別すると（ステップ S 3 5 ; N o）、処理はステップ S 3 6 に移行し、未処理の出現位置があるか判別する（ステップ S 3 6）。そして、未処理の出現位置があれば（ステップ S 3 6 ; N o）、次の出現位置についてステップ S 3 2 から処理を繰り返す。また、全ての出現位置について処理済なら（ステップ S 3 6 ; Y e s）、出現位置取得処理を終了する。

【 0 0 7 6 】

一方、連続性があると判別すると（ステップ S 3 5 ; Y e s）、図 1 1 の S 3 0 9 と同様に、対象文字列の全検索パタンのうち先頭にある検索パタンの特定出現位置を、対象文字列の再配置テキストにおける出現位置として特定する（ステップ S 3 7）。

【 0 0 7 7 】

一方、ステップ S 3 1 で検索パターンが一つと判別された場合（ステップ S 3 1 ; Y e s）は、その一つの検索パターンを注目検索パターンとしてステップ S 3 2 と同様に出現位置を求め、注目出現位置を選択する（ステップ S 3 8）。そして、注目出現位置を入力キーワードの再配置テキストにおける出現位置として特定する（ステップ S 3 9）。

【 0 0 7 8 】

次に、本検索部 1 4 2 0 は、注目検索パタンの出現位置の全てを注目出現位置として上記処理を実行したか否かを判別する（ステップ S 4 0）。未処理の出現位置が残っている場合（ステップ S 4 0 ; N o）、次の出現位置について処理を繰り返す。このとき、まずステップ S 3 1 で検索パターンが一つであると判別したか否かを判別する（ステップ S 4 1）。検索パターンが一つであった場合（ステップ S 4 1 ; Y e s）、ステップ S 3 8 に戻る。一方、検索パターンが複数であった場合（ステップ S 4 1 ; N o）、ステップ S 3 2

10

20

30

40

50

に戻る。

【0079】

一方、全ての出現位置について上記処理済みであった場合（ステップS40；Yes）、対象文字列の全ての出現位置を特定したので、出現位置取得処理は終了する。

【0080】

図14に戻って、ステップS22で基準文字列の再配置CTDにおける出現位置を全て取得すると、次に本検索部1420が、各出現位置が、どの構成単位のどの部位（見出部CE、解説部CC、及び用例部CXのいずれか）に含まれるか、特定する（ステップS23）。具体的な特定方法は図11のステップS312と同様である。

【0081】

ステップS23ではさらに、算出部150が図13に示す決定方法テーブルを参照して、現在の検索条件、各出現位置に対する評価値の決定方法（順位決定方法）を取得する。

例えば、検索キーワードが「単一」（つまり、1個）であり、見出部CEが出現部位である場合については、数式（2）を用いて評価値を算出するという情報を取得する。このとき、その見出部CEに含まれるテキストを特定テキストと呼ぶ。

さらに、算出部150は、評価値を算出する方法の情報を取得すると、その情報が示す方法で各出現部位の評価値を算出する（ステップS23）。この評価値の値が小さい程、出力優先度は高い。

【0082】

本実施形態では、検索キーワードが一つであり、特定部位が見出部CEであった場合の出力優先度（ $Est_{idxid,0}$ ）は、以下の式（2）を用いて算出する。

【0083】

【数2】

$$Est_{idxid,0} = \frac{NumStr_{index}}{NumStr_{stdstr}} \dots (2)$$

【0084】

なお、 $NumStr_{index}$ は特定部位（見出部）の文字列の長さであり、 $NumStr_{stdstr}$ は基準文字列の長さである。

数式（2）で算出される評価値は、基準文字列と特定部位（見出部）の文字列（つまり、見出テキスト）とが完全に一致する場合に最小の値となり、特定見出部の見出テキストに基準文字列「while」以外の文字が多く含まれる程大きい値となる。これは、通常、ユーザは、検索キーワードである基準文字列と完全一致する見出テキストの表示を所望することが多いからである。また、通常、ユーザは、検索キーワード以外の文字を多く含む見出テキストよりも検索キーワード以外の文字を少なく含む見出テキストの表示を所望することが多いからである。

【0085】

また、検索キーワードが一つであり、特定部位が解説部CCであった場合の出力優先度（ $Est_{idxid,1}$ ）は、以下の式（3）を用いて算出する。

【0086】

【数3】

$$Est_{idxid,1} = Pos_{stdstr} - PosSt_{idxid,comentary} \dots (3)$$

【0087】

なお、 $PosSt_{idxid,comentary}$ は、特定見出部に対応する解説部C

10

20

30

40

50

Cの開始位置である。数式(3)で算出される評価値は、基準文字列の特定出現位置が、解説部CCの開始位置に近い程小さい値となる。例えば、基準文字列をそれぞれ含む複数の解説テキストが同じ解説部CCに分類されている場合には、解説部CCにおいて先の位置に格納された解説テキストの評価値程小さい値となる。再配置コンテンツテキストデータの解説部CCに分類された解説テキストは、例えば、見出語の一般的な意味内容若しくは見出語の使用頻度の高い意味内容を解説する解説テキスト程先の位置に格納されており、通常、ユーザは、一般的な解説を行う若しくは使用頻度の高いテキストの表示を所望することが多いためである。つまり、解説部CCにおいて先の位置に格納された解説テキスト程ユーザが所望するテキストである可能性が高いと考えられるためである。このときには、その解説部CCのうち、特定部位を含む解説テキストの特定部位を含む所定の文字数の文字列を特定テキストと呼ぶ。

10

【0088】

また、検索キーワードが一つであり、特定部位が用例部CXであった場合の出力優先度($Est_{idxid,2}$)は、以下の式(4)を用いて算出する。

【0089】

【数4】

$$Est_{idxid,2} = Pos_{stdstr} - PosSt_{idxid,example,expid} \dots (4)$$

【0090】

20

なお、 $PosSt_{idxid,example,expid}$ は、特定見出部idxidに対応する用例部CXにおける、基準文字列を含む用例テキスト(用例番号expidとする)の開始位置である。

なお、検索部140は、 $PosSt_{idxid,example,expid}$ は、次の手順で取得する。

まず、ステップS23で取得した用例開始位置情報格納アドレスに基づいて、図9に示したファイル名が「example.idx」のEPファイル(つまり、用例開始位置情報ファイル)から用例開始位置情報を参照して求める。さらに、用例番号用例開始位置情報に含まれる用例部開始位置からの差分を表す情報の内で、算出された用例部CXの開始位置と、処理対象となる特定出現位置と、の差分以下であり、かつ最大の差分を求める。

30

次に、検索部140は、検索した差分を表す情報が用例開始位置情報において何番目の情報であるかを上記差分用の所定バイト数に基づいて特定し、特定した番号を用例番号とする。

その後、求めた差分を、特定見出部に対応した用例部CXの開始位置に加算することで、特定した用例番号の用例テキストの開始位置($PosSt_{idxid,example,expid}$)を算出する。

【0091】

数式(4)で算出される評価値は、基準文字列の特定出現位置が、用例テキストの開始位置に近い程小さい値となる。例えば、基準文字列をそれぞれ含む複数の解説テキストが同じ用例部CXに分類されている場合には、用例部CXにおける格納位置が先であるか後であるかではなく、用例テキストにおいて基準文字列「while」が使用される位置が先である程評価値が小さい値となる。通常、例えば、検索キーワードである基準文字列「while」の一般的な使用例若しくは使用頻度の高い使用例を表す用例テキスト程先の位置で基準文字列「while」を使用することが多いからである。また、通常、ユーザは、一般的な使用例を表すテキスト若しくは使用頻度の高いテキストの表示を所望することが多いためである。つまり、先の位置に基準文字列「while」が格納された用例テキスト程ユーザが所望するテキストである可能性が高いと考えられるためである。このときには、この用例テキストのうち特定部位を含む所定の長さの文字列を特定テキストと呼ぶ。すなわち、特定テキストとは、検索ヒット部位を含む、検索結果として表示されるべきテキストである。特定テキストは、上記説明に限らず、自由に設定された設定情報に基づき検索結果をふ

40

50

くむテキストから抽出されるとしてもよい。

【0092】

そして、ステップS23で全ての出現位置について評価値を算出すると、決定部160が、特定テキストのカテゴリ毎に、算出した特定テキストの評価値に基づき、1又は複数の特定テキストの表示順位を決定する(ステップS24)。本実施形態では、決定部160は、特定テキストの評価値が小さい特定テキストの表示順位程、先に表示される順位に決定する。

【0093】

その後、出力部170が、カテゴリ毎に、決定された表示順位で1又は複数の特定テキストを表示する、図7に示すような検索結果表示画面を表す信号を表示部180へ出力する(ステップS25)。そして、単一文字列検索処理の実行を終了する。尚、表示部180は、出力された信号に基づいて検索結果表示画面を表す。

【0094】

次に、ステップS19で実行される複数文字列検索処理について、図16を参照して説明する。

【0095】

複数文字列検索処理では、まず本検索部1420が図12のステップS12で生成した、基準文字列のNグラムを取得する(ステップS51)。

【0096】

そして、基準文字列の再配置CTDにおける出現位置を全て取得する処理(出現位置取得処理、図15)を実行する(ステップS52)。

【0097】

さらに、図12のS18で定められた検証文字列の一つについて、図12のステップS12で生成した、基準文字列のNグラムを取得する(ステップS53)。

【0098】

そして、検証文字列の再配置CTDにおける出現位置を全て取得する処理(出現位置取得処理、図15)を実行する(ステップS54)。

【0099】

ある検証文字列の出現位置を全て取得すると、未処理の検証文字列があるか判別する(ステップS55)。未処理の検証文字列が有る場合(ステップS55; No)、次の未処理の検証文字列について、ステップS53から処理を繰り返す。

【0100】

一方、全ての検証文字列について出現位置を取得している場合(ステップS55; Yes)、次にステップS52で取得した基準文字列の出現位置の一つを、注目出現位置として選択する(ステップS56)。

【0101】

そして、注目出現位置から所定の範囲に、すべての検証文字列が存在するか、ステップS54で取得した検証文字列の出現位置に基づいて判別する(ステップS57)。ここで、所定の範囲とは、注目出現部位が、ある見出番号の見出部CEに有る場合は、その見出部全体を言う。また、解説部CCに有る場合は、その解説部CC全体の中で、所定の文字数の範囲を言う。また、用例部CXに有る場合は、その注目出現部位が属する用例テキスト(図5(b)の用例テキスト11、用例テキスト12、等のそれぞれ)が占める範囲を言う。

なお、この「所定の範囲」はこれに限らず、自由に設定可能である。また、ユーザ操作等により設定変更することが出来るとしても良い。

【0102】

注目出現位置から所定の範囲に、すべての検証文字列が存在しない場合(ステップS7; No)、その注目出現部位は複数の検索キーワードに対する検索(AND検索)のヒット部位では無いとの判断のもと、注目出現部位を検索結果から排除する(ステップS8)。

10

20

30

40

50

【 0 1 0 3 】

一方、注目出現位置から所定の範囲に、すべての検証文字列が存在する場合（ステップ S 5 7 ; Y e s ）、注目出現部位を検索ヒット部位とする（ステップ S 5 9 ）。そして、その注目出現部位について評価値を算出する（ステップ S 6 0 ）。

【 0 1 0 4 】

ここで、本処理における評価値の算出方法を説明する。まず、算出部 1 5 0 が図 1 3 の決定方法テーブルを参照して、注目出現部位のカテゴリと、検索キーワードの単語数に合致する順位決定方法（算出方法）を取得する。

【 0 1 0 5 】

注目出現部位が見出部 C E に位置する場合には、算出部 1 5 0 は、図 1 3 に示す決定方法テーブルから、検索キーワードが「複数」であることを表す情報と、特定カテゴリ「見出部」を表す情報と、に対応付けられた順位決定方法「数式 5 」を表す情報を取得する。

10

【 0 1 0 6 】

次に、算出部 1 5 0 は、基準文字列と、各検証文字列の文字数を取得する。次に、そして、各数値を以下の数式（ 5 ）に代入して、注目出現部位の特定テキストの評価値を算出する。

【 0 1 0 7 】

【 数 5 】

$$Est_{idxid, cgy} = Est_{idxid, 0} = EstDist_{WithoutOrder} \dots (5)$$

20

$$EstDist_{WithoutOrder} = EstDist = \max Pos - \min Pos$$

maxPos: 全ての $Dist_{stdstr, vjstr_k}$ が最小となる場合の全ての検索文字列を含む範囲の上限値

minPos: 全ての $Dist_{stdstr, vjstr_k}$ が最小となる場合の全ての検索文字列を含む範囲の下限値

$$Dist_{stdstr, vjstr_k} = \begin{cases} (Pos_{vjstr_k} + NumStr_{vjstr_k}) - Pos_{stdstr} & [Pos_{stdstr} \leq Pos_{vjstr_k}] \\ (Pos_{stdstr} + NumStr_{stdstr}) - Pos_{vjstr_k} & [Pos_{stdstr} > Pos_{vjstr_k}] \end{cases}$$

$EstDist$: 検索キーワード間距離

30

$EstDist_{WithoutOrder}$: 検索キーワード間距離(入力順を考慮しない場合)

$Dist_{stdstr, vjstr_k}$: 基準文字列とk番目の検証文字列との距離

$NumStr_{vjstr_k}$: k番目の検証文字列の文字数

$NumStr_{stdstr}$: 基準文字列の文字数

Pos_{vjstr_k} : k番目の検証文字列の特定出現位置

Pos_{stdstr} : 基準文字列の特定出現位置

40

【 0 1 0 8 】

ここで、数式（ 5 ）で算出される基準文字列と k 番目の検証文字列との距離について、基準文字列を「while」とし、1番目の検証文字列を「for」とした例を用いて説明する。図 1 7 (a) に示すように、基準文字列「while」の特定出現位置が、1番目の検証文字列「for」の特定出現位置よりも先である場合には、基準文字列「while」と1番目の検証文字列「for」との距離は、基準文字列「while」の先頭から検証文字列の末尾までの距離を表す。これに対して、図 1 7 (b) に示すように、基準文字列「while」の特定出現位置が、1番目の検証文字列「for」の特定出現位置よりも後である場合には、基準文字列「while」と1番目の検証文字列「for」との距離は、検証文字列「for」の先頭から基準文字列「while」の末尾まで距離を表す。

50

【0109】

次に、数式(5)で算出される上限値 $maxPos$ 及び下限値 $minPos$ 並びにこれらを用いて算出される検索キーワード間距離(入力順を考慮しない場合)について、基準文字列「while」と1番目の検証文字列「for」と2番目の検証文字列「a」とを具体例として挙げて説明を行う。図18に示すように、上限値 $maxPos$ 及び下限値 $minPos$ は、基準文字列と、基準文字列「while」との距離を最小とする1番目の検証文字列「for」と、基準文字列「while」との距離を最小とする2番目の検証文字列「a」と、基準文字列「while」、1番目の検証文字列「for」、及び2番目の検証文字列「a」の全てを含む最小の範囲(以下、最小包含範囲という)の下限値及び上限値である。また、検索キーワード間距離(入力順を考慮しない場合)は、当該下限値 $minPos$ と上限値 $maxPos$ との差異である。

10

【0110】

このため、数式(5)で算出される評価値は、最小包含範囲に含まれる文字列が、入力された検索キーワードのみである場合に最小値となる。また、検索キーワードの全てがより狭い最小包含範囲に含まれている程(つまり、検索キーワードのいずれかの使用が開始される位置が、全ての検索キーワードの使用が終了される位置により近い程)算出される評価値が小さい値となる。例えば、入力された複数の検索キーワードが「for」、「a」、及び「while」である場合には、「for a while」を含む用例テキストの評価値の方が、「for a brief while」を含む用例テキストの評価値よりも小さい値となる。また、「for a brief while」を含む用例テキストの評価値の方が、「for a job while」を含む「This will allow you to look 『for a job while』 you are working at a part-time job」というテキストの評価値よりも小さい値となる。つまり、通常、ユーザは、検索キーワードのみで構成されるイディオムの使用例を表す用例テキストの表示を所望することが多いからであり、かつ検索キーワード以外の文字列を多く含む用例テキストよりも検索キーワード以外の文字を少なく含む用例テキストの表示を所望するからである。また、入力された複数の検索キーワードの間に文字が多く使用されている用例テキスト程、入力された複数の検索キーワードで構成されるイディオムとは異なるイディオムの使用例を表すことが多いためである。

20

【0111】

また、特定テキストの特定カテゴリが解説部CCである場合には、算出部150は、決定方法テーブルから、検索キーワードが「複数」であることを表す情報と、特定カテゴリ「解説部」を表す情報と、に対応付けられた順位決定方法「数式6」を表す情報を取得する。

30

【0112】

次に、算出部150は、数式(5)を用いて評価値を算出する場合と同様に、検索キーワード間距離及び下限値 $minPos$ を算出する。また、算出部150は、数式(3)を用いて評価値を算出する場合と同様に、特定見出部に対応する解説部CCの開始位置を算出する。その後、算出部150は、検索キーワード間距離及び下限値 $minPos$ と解説部CCの開始位置 $Pos_{idxid,comentary}$ とを、以下の数式(6)に用いて特定テキストの評価値を算出する。

40

【0113】

【数6】

$$\begin{aligned} Est_{idxid, cgy} &= Est_{idxid, 1} \\ &= EstDist_{WithoutOrder} + (\min Pos - Pos_{idxid, comentary}) \cdots (6) \end{aligned}$$

【0114】

数式(6)で算出される評価値は、検索キーワード間距離(入力順を考慮しない場合)が同じ解説テキストであれば、検索キーワードを包含する最小包含範囲の下限値 $minPos$ が解説部CCの開始位置に近い程小さい値となる。また、数式(6)で算出される評

50

価値は、下限値 $minPos$ と解説部 CC の開始位置との距離が同じ解説テキストであれば、検索キーワード間距離（入力順を考慮しない場合）が短い程小さい値となる。

【0115】

次に、特定テキストの特定カテゴリが用例部 CX である場合には、決定方法テーブルから、検索キーワードが「複数」であることを表す情報と、特定カテゴリ「用例部」を表す情報と、に対応付けられた順位決定方法「数式7」を表す情報を取得する。

【0116】

この場合、数式(5)を用いて評価値を算出する場合と同様に、検索キーワード間距離及び下限値 $minPos$ を算出し、数式(6)を用いて評価値を算出する場合と同様に、用例テキストの開始位置を算出する。その後、算出部150は、検索キーワード間距離及び下限値 $minPos$ と用例テキストの開始位置 $Pos_{idxid,example,expid}$ とを、以下の数式(7)に用いて特定テキストの評価値を算出する。

10

【0117】

【数7】

$$Est_{idxid,egy} = Est_{idxid,2} = EstDist_{WithoutOrder} + (\min Pos - Pos_{idxid,exaple,expid}) \dots(7)$$

【0118】

数式(7)で算出される評価値は、同じ並び順の複数の検索キーワードの検索キーワード間距離（入力順を考慮しない場合）が同じであれば、下限値 $minPos$ が用例テキストの開始位置に近い程小さい値となり、下限値 $minPos$ と用例テキストの開始位置との距離が同じ解説テキストであれば、検索キーワード間距離（入力順を考慮しない場合）が短い程小さい値となる。

20

【0119】

以上のようにして注目出現部位の特定テキストについて評価値を算出すると、次に、S52で求めた基準文字列の出現位置の全ての出現位置を注目出現位置として上記処理を実行したかしたか否かを判別する（ステップS61）。未処理の出現位置が残っている場合（ステップS61；No）、次の未処理の出現位置についてステップS56から処理を繰り返す。

30

一方、全ての出現位置について上記処理済みであった場合（ステップS61；Yes）、決定部160が、カテゴリ毎に、ステップS60で算出された評価値に基づき、1又は複数の特定テキストの表示順位を決定する（ステップS62）。本実施形態では、決定部160は、特定テキストの評価値が小さい特定テキストの表示順位程、先に表示される順位に決定する。

【0120】

その後、出力部170が、カテゴリ毎に、決定された表示順位で1又は複数の特定テキストを表示する、検索結果表示画面を表す信号を表示部180へ出力する（ステップS63）。そして、単一文字列検索処理の実行を終了する。尚、表示部180は、出力された信号に基づいて検索結果表示画面を表す。

40

【0121】

以上説明したとおり、本実施形態のテキスト検索装置によれば、前方一致検索とその他の検索（部分一致検索）とを、一つの索引（転置インデックス）を用いて実行することが出来る。そのため、ユーザに対して前方一致検索と部分一致検索の結果を提示できる、ユーザの利便性の高いテキスト検索機能を、少ない必要記憶容量で実現できる。

【0122】

また、式(1)の評価値や、先頭マークを含めた検索により、部分一致検索用の索引情報を用いて、検索キーワードが前方一致検索の対象となるテキストの先頭にある文字列を特定する手段（特定手段）を実現している。このため、利便性の高いテキスト検索機能を実現するために必要な計算量の増加が少ない。

50

【 0 1 2 3 】

例えば、先頭マークを含める検索により、先頭にある文字列を特定することにより、先頭マークを含めない場合よりも検索ヒット候補を抑えることが出来、計算量が少なくてすむ。

【 0 1 2 4 】

また、本実施形態では、ユーザがキーワードを一文字ずつ入力している場合に前方一致検索（インクリメント検索）を実施し、検索キーワードを確定するとより詳細かつ表示される検索結果が多い部分一致検索を実施する。このように、ユーザが必要とする検索結果を、キーワード入力中とキーワード確定後といった状況に合わせて提示することが出来る。

10

【 0 1 2 5 】

(変形例)

以上、本発明の実施形態 1 について説明したが、本発明の実施形態はこれに限られない。

例えば、上記実施形態 1 では、先頭マークは半角スペースであるとしたが、先頭マークはこれに限られず自由に選択可能である。例えば、先頭マークはシステムに予約されたユニークなマーク（その他で前方一致検索の対象となるテキストに現れないマーク）であってよい。このような場合、ユニークな先頭マークを挿入した転置インデックスを用いて、先頭マークを挿入した検索キーワードを検索すれば、検索ヒット部位は必ず対象となる文字列の先頭である。そのため、この場合には検索ヒット部位（特定部位）が前方一致検索の対象となるテキストの先頭であるかを、先頭を示すアドレスと特定部位とを比較して評価する必要はない。具体的には、図 11 のステップ S 3 1 2 で評価値を算出し、ステップ S 3 1 3 での判別処理に反映する必要は無い。このような構成によれば、少ない必要計算量で前方一致検索と部分一致検索を実現できる。

20

【 0 1 2 6 】

一方、前方一致検索において、先頭を示すアドレスと検索ヒット部位（特定部位）とを比較して評価する場合、転置インデックスに先頭マークが含まれておらず、検索キーワードの先頭に先頭マークを挿入しなくても、転置インデックスを用いて前方一致検索を実行することが出来る。この場合は、転置インデックスが外部サーバにあり、先頭マークを含んでいない一般的なものであった場合でも、その転置インデックスを用いて前方一致検索と部分一致検索とを実行することが出来る。

30

【 0 1 2 7 】

また、実施形態 1 では再配置 C T D 及び転置インデックスを生成する処理（データ等生成処理）をテキスト検索装置 1 0 0 が実行するとしたが、これに限らず、予め外部装置が生成した再配置 C T D 及び転置インデックスを記憶部に記憶していることで足りる。例えば、工場出荷時において、各辞書データに対応する再配置 C T D と転置インデックスとを記憶するとしても良い。

【 0 1 2 8 】

尚、実施形態 1 およびその変形例において、電子辞書は、電子辞典をも含む。つまり、電子辞書は、国語辞典、英和辞典、和英辞典、及び百科事典であっても良い。また、本発明では、テキスト検索装置 1 0 0 は、検索キーワードに基づいて辞書を検索するとして説明したが、検索の対象とされる文書は、辞書に限定される訳ではなく、複数のカテゴリに分類されたテキストで構成される文書であれば、どのような文書でも良い。また、インクリメント検索（前方一致検索）と、全文検索（部分一致検索）における具体的な方法、特に順位付けの方法については、任意に設定可能である。

40

【 0 1 2 9 】

検索の対象とされる文書は、例えば、「発明の名称」及び「特許請求の範囲」などのカテゴリにそれぞれ分類されたテキストで構成される特許明細書であっても良い。このとき、インクリメント検索の対処を、発明の名称とし、全文を本検索の対象とすることが考えられる。

50

また、検索の対象とされる文書は、例えば、ある製品が有する機能の名称を表すテキストが分類されるカテゴリ（以下、機能名カテゴリという）と、当該機能を利用するための操作方法を表すテキストが分類されるカテゴリ（以下、操作方法カテゴリという）とを有する説明書であっても良い。このとき、製品名と、カテゴリと、をインクリメント検索の対象とし、全文を本検索の対象とすることが考えられる。

【 0 1 3 0 】

尚、本実施形態において、検索キーワードの検索方法は、Nグラム検索を用いるとして説明したが、検索キーワードの検索方法は、これに限定される訳でない。

【 0 1 3 1 】

尚、本発明は、実施形態1、若しくはその変形例に係るテキスト検索装置100として提供できることはもとより、プログラムの適用により、既存のテキスト検索装置をテキスト検索装置100として機能させることもできる。すなわち、テキスト検索装置100による各機能構成を実現させるためのテキスト検索プログラムを、既存のテキスト検索装置を制御するコンピュータ（CPUなど）が実行できるように適用することで、実施形態1、若しくは変形例に係るテキスト検索装置100として機能させることができる。

【 0 1 3 2 】

このようなプログラムの配布方法は任意であり、例えば、メモ리카ード、CD-ROM、又はDVD-ROMなどの記録媒体に格納して配布できる他、インターネットなどの通信媒体を介して配布することもできる。また、再配置CTD及び転置インデックス等の上記処理に必要なデータの一部又は全部は、外部サーバに記憶されており、通信機能によりこれらのデータを取得して上記処理を実行する構成も可能である。

【 0 1 3 3 】

また、本発明は、本発明の広義の精神と範囲を逸脱することなく、様々な実施形態及び変形が可能とされるものである。つまり、本発明のいくつかの実施形態を説明したが、上述した実施形態は本発明を説明するためのものであり、本発明の範囲を限定するものではない。本発明の範囲は、実施形態ではなく、特許請求の範囲に記載された発明とその均等の範囲を含む。

以下に本願出願の当初の特許請求の範囲に記載された発明を付記する。

【 0 1 3 4 】

（付記1）

複数の文書カテゴリから構成される検索対象となる文書に含まれる文字列と、当該文字列の前記文書における出現位置と、を対応づけた索引情報を記憶する記憶手段と、

検索キーワードを取得する取得手段と、

前記文書のうち前方一致検索の対象となるカテゴリに属する文字列から、前記検索キーワードから始まる文字列を、前記索引情報を用いて検索する前方一致検索手段と、

前記索引情報を用いて、前記文書から前記検索キーワードを含む文字列を検索する部分一致検索手段と、

前記前方一致検索手段が検索した結果を示す情報と、前記部分一致検索手段が検索した結果を示す情報と、を出力する出力手段と、

を備えることを特徴とするテキスト検索装置。

【 0 1 3 5 】

（付記2）

前記前方一致検索手段は、前記索引情報を用いて、前記検索キーワードが前記前方一致検索の対象となるカテゴリに属する文字列のうち、当該検索キーワードが先頭にある文字列を特定して、前記検索キーワードから始まる文字列を検索する、

ことを特徴とする付記1に記載のテキスト検索装置。

【 0 1 3 6 】

（付記3）

前記記憶手段は、前記前方一致検索の対象となるカテゴリに属する文字列の先頭に、先頭であることを示すマークである先頭マークを挿入した上で、当該文字列を前記索引情報

10

20

30

40

50

を記憶し、

前記前方一致検索手段は、前記取得手段が取得した検索キーワードの先頭に前記先頭マークを挿入して、当該先頭マークを含む検索キーワードを、前記先頭マークを挿入した文字列を含む索引情報を用いて前記文書を検索することにより、前記検索キーワードが先頭にある文字列を特定する、

ことを特徴とする付記 2 に記載のテキスト検索装置。

【 0 1 3 7 】

(付記 4)

前記索引情報は、前記前方一致検索の対象となるカテゴリに属する各文字列の先頭的位置を示す先頭情報をさらに含み、

前記前方一致検索手段は、前記索引情報を用いて前記検索キーワードの位置を特定し、当該特定した位置と、前記索引情報の先頭情報が示す位置と、を比較して前記検索キーワードが先頭にある文字列を特定する、

ことを特徴とする付記 2 又は 3 に記載のテキスト検索装置。

【 0 1 3 8 】

(付記 5)

検索キーワードを一字ずつ入力するユーザの操作と、検索キーワードを確定するユーザの操作と、を受け付ける操作受付手段を更に備え、

前記取得手段は、

前記操作受付手段が検索キーワードを入力する操作を受け付ける度に、当該操作が示す文字をそれまで受け付けた操作が示す文字又は文字列に加えて、前記前方一致検索手段が検索する検索キーワードとして取得し、

前記操作受付手段が前記検索キーワードを確定するユーザの操作を受け付けると、当該確定された検索キーワードを、前記部分一致検索手段が検索する検索キーワードとして取得する、

ことを特徴とする付記 1 乃至 4 の何れか一つに記載のテキスト検索装置。

【 0 1 3 9 】

(付記 6)

コンピュータに、

複数の文書カテゴリから構成される検索対象となる文書に含まれる文字列と、当該文字列の前記文書における出現位置と、を対応づけた索引情報を取得する処理、

検索キーワードを取得する処理、

前記文書のうち前方一致検索の対象となるカテゴリに属する文字列から、前記検索キーワードから始まる文字列を、前記索引情報を用いて検索する処理、

前記索引情報を用いて、前記文書のうち前記検索キーワードを含む文字列を検索する処理、

前記検索した結果を示す情報を出力する処理、

を実行させることを特徴とするプログラム。

【 符号の説明 】

【 0 1 4 0 】

1 0 0 . . . テキスト検索装置、 1 0 0 a . . . CPU、 1 0 0 b . . . ROM、 1 0 0 c . . . RAM、 1 0 0 d . . . ハードディスク、 1 0 0 e . . . メディアコントローラ、 1 0 0 g . . . ビデオカード、 1 0 0 h . . . LCD、 1 0 0 i . . . キーボード、 1 0 0 j . . . スピーカ、 1 1 0 . . . 情報記憶部、 1 2 0 . . . 生成部、 1 3 0 . . . 取得部、 1 4 0 . . . 検索部、 1 4 1 0 . . . インクリメント検索部、 1 4 2 0 . . . 本検索部、 1 5 0 . . . 算出部、 1 6 0 . . . 決定部、 1 7 0 . . . 出力部、 1 8 0 . . . 表示部

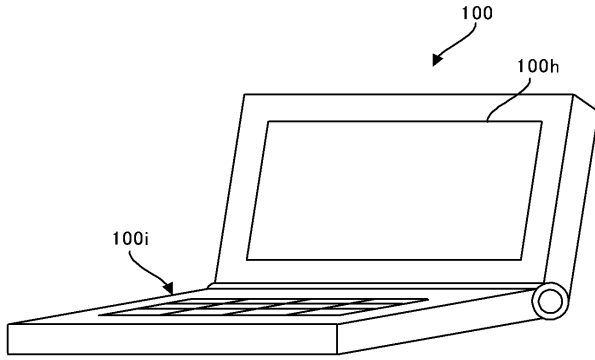
10

20

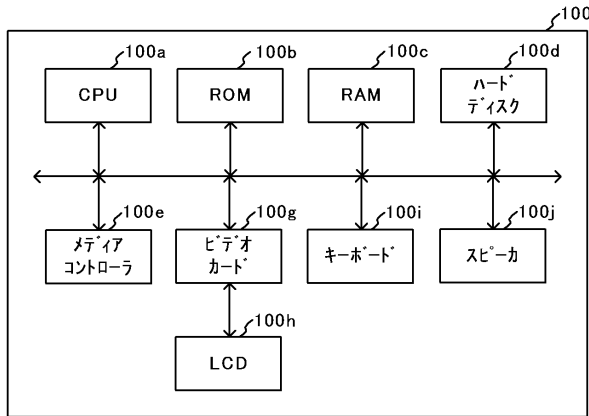
30

40

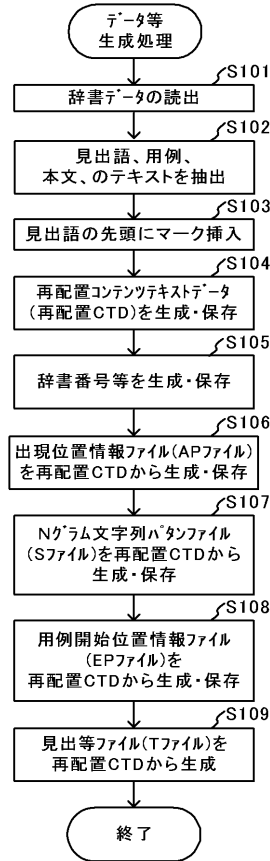
【図1】



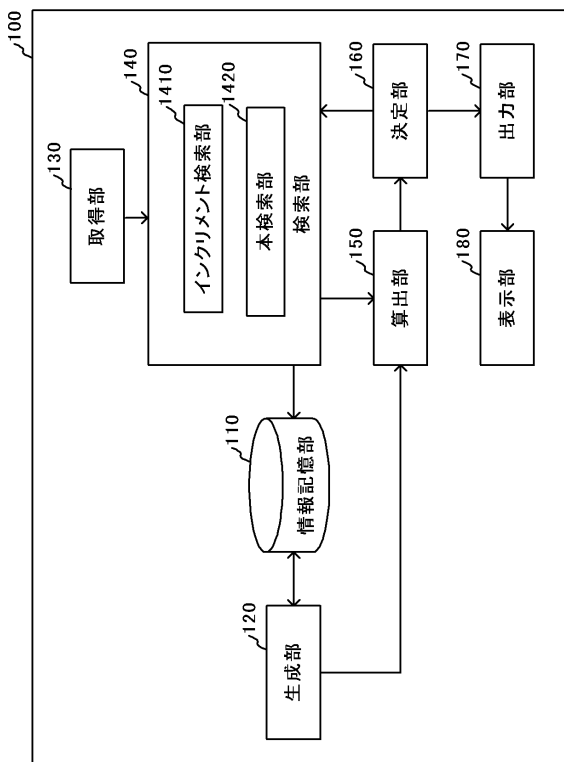
【図2】



【図3】



【図4】



【図5】

辞書データ

見出語1		
解説テキスト11	用例テキスト11	解説テキスト12
用例テキスト12	解説テキスト13	用例テキスト13
見出語2		
解説テキスト21	用例テキスト21	解説テキスト22
用例テキスト22	解説テキスト23	用例テキスト23
⋮	⋮	⋮

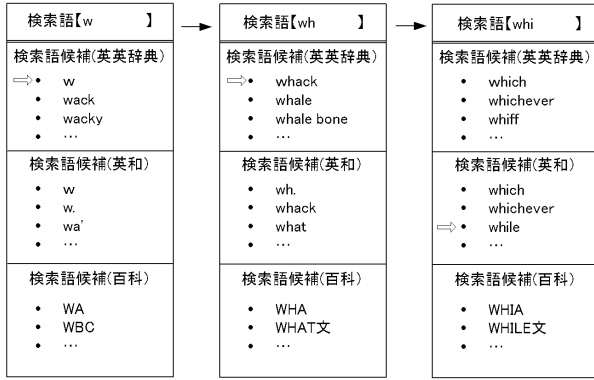
(a)

再配置コンテンツテキストデータ

・見出語1		
解説テキスト11	解説テキスト12	解説テキスト13
用例テキスト11	用例テキスト12	用例テキスト13
・見出語2		
解説テキスト21	解説テキスト22	解説テキスト23
用例テキスト21	用例テキスト22	用例テキスト23
⋮	⋮	⋮

(b)

【図 6】

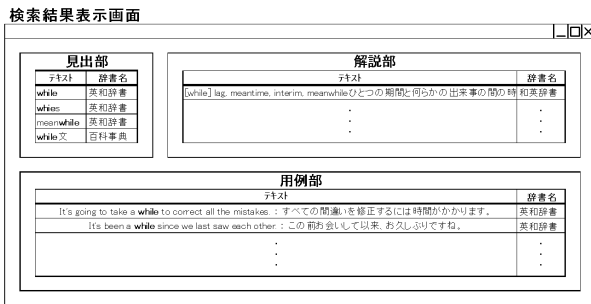


【図 8】

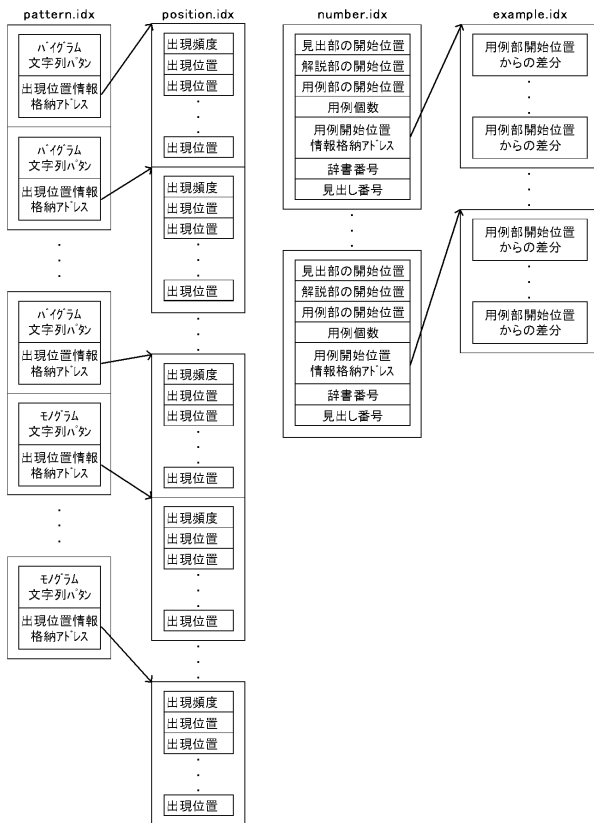
辞書テーブル

辞書番号	辞書名	先頭アドレス
1	国語辞典	0x8000
2	英和辞典	0x8800
3	和英辞典	0x9000
4	百科辞典	0x9800
⋮	⋮	⋮

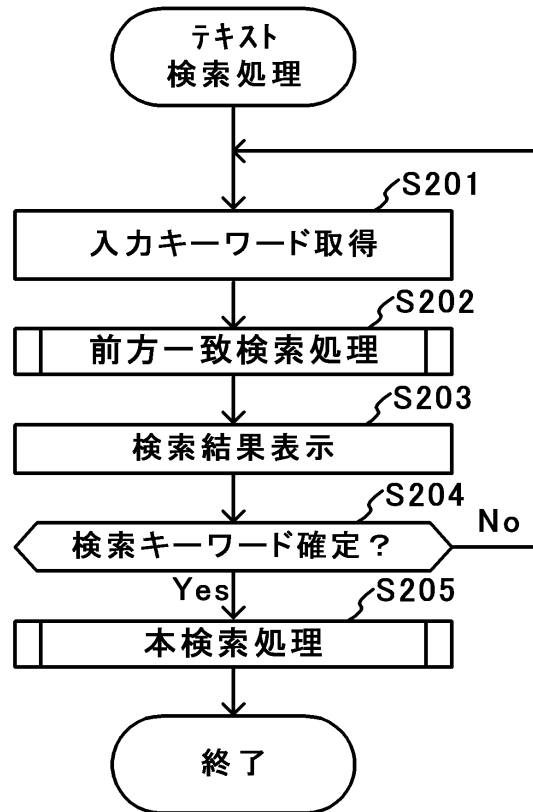
【図 7】



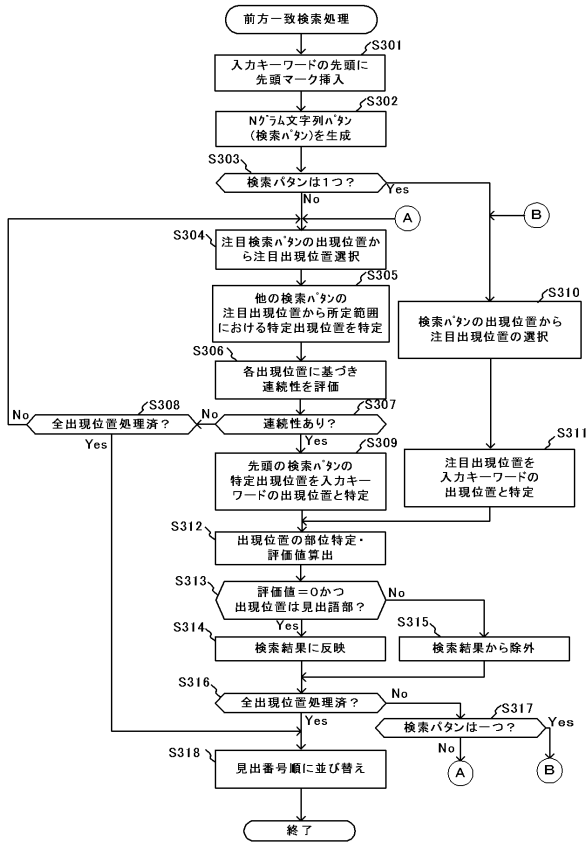
【図 9】



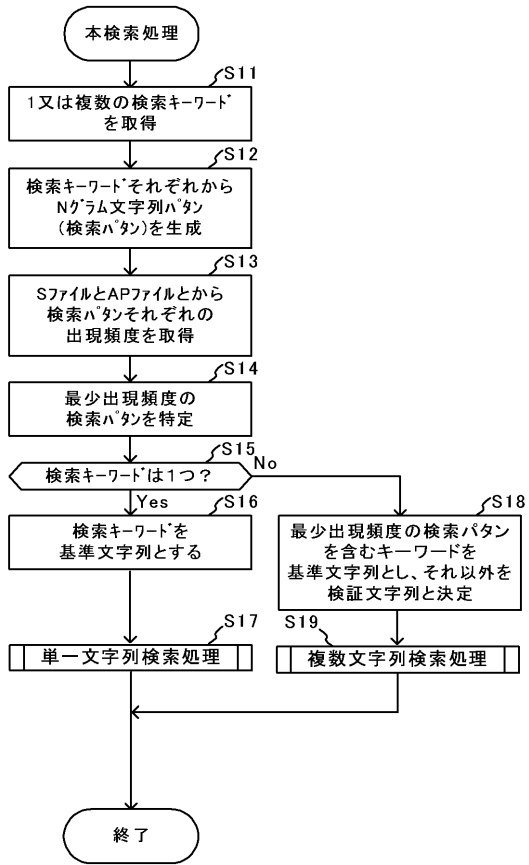
【図 10】



【図11】



【図12】

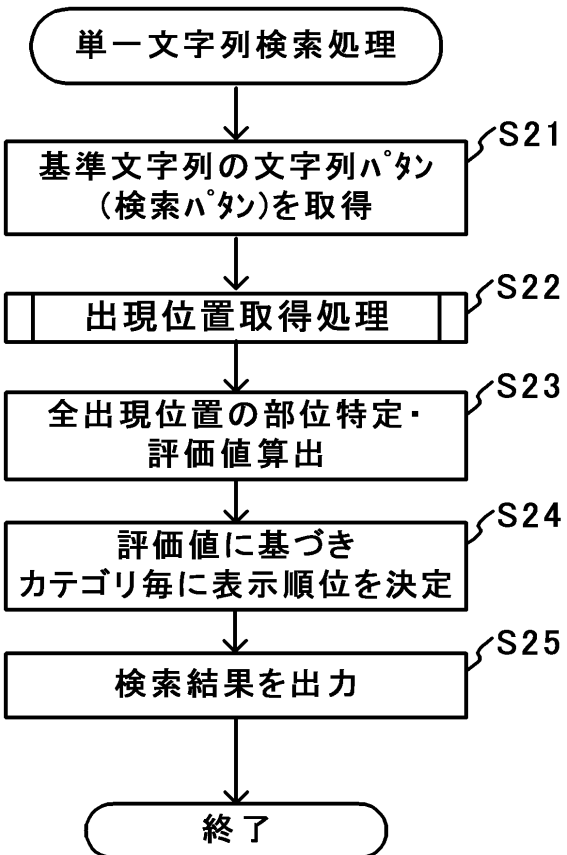


【図13】

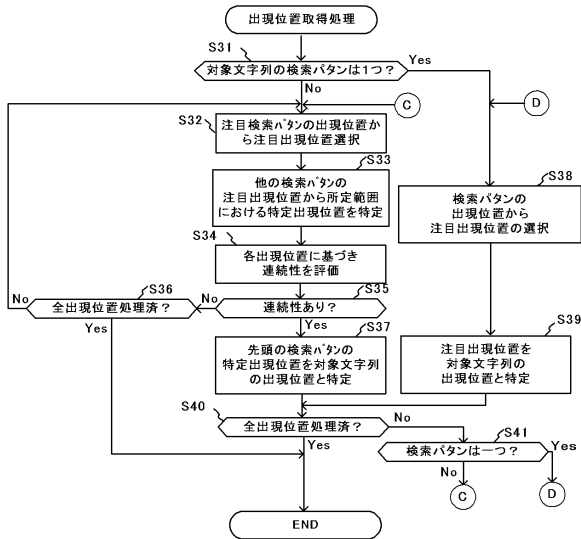
決定方法テーブル

検索キーワード数	特定カテゴリ	順位決定方法
単一	見出部	数式2
単一	解説部	数式3
単一	用例部	数式4
複数	見出部	数式5
複数	解説部	数式6
複数	用例部	数式7
.	.	.
.	.	.
.	.	.

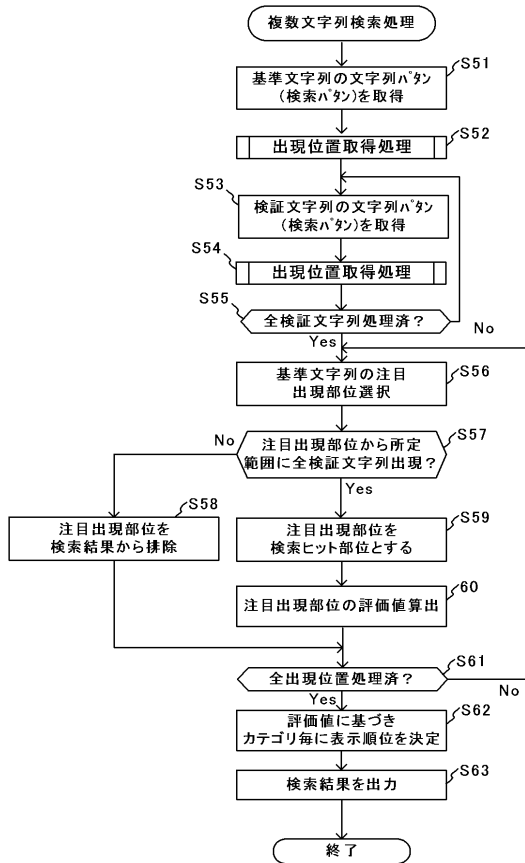
【図14】



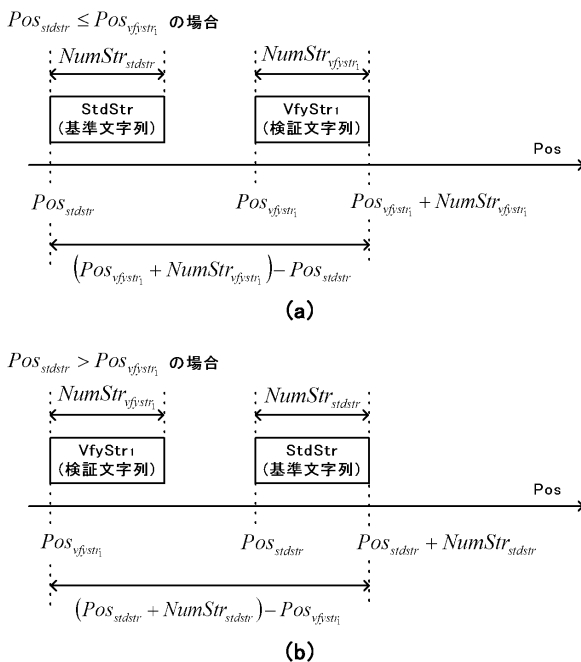
【図15】



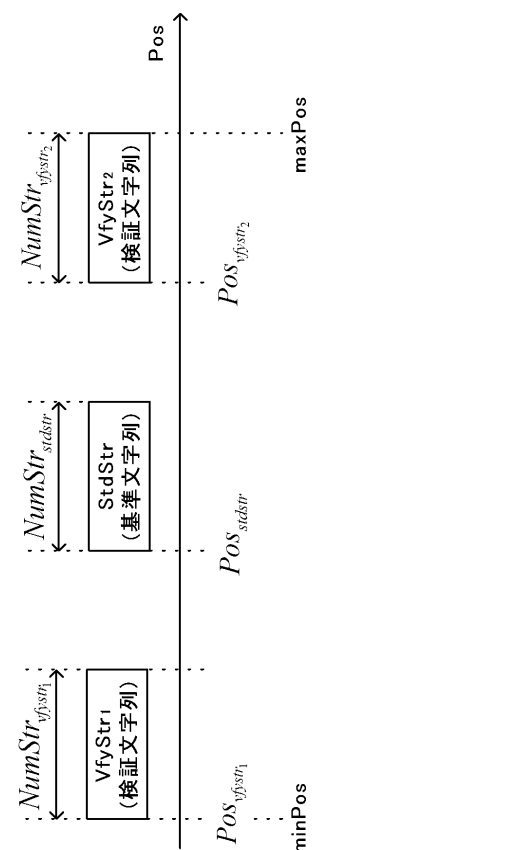
【図16】



【図17】



【図18】



フロントページの続き

- (56)参考文献 特開平08 - 314948 (JP, A)
特開2002 - 269139 (JP, A)
特開平08 - 241332 (JP, A)
特開2011 - 107986 (JP, A)

- (58)調査した分野(Int.Cl., DB名)
G06F 17/30