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

特許第5494066号
(P5494066)

(45)発行日　平成26年5月14日(2014.5.14)　(24)登録日　平成26年3月14日(2014.3.14)

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

請求項の数 8　(全 13 頁)

(21)出願番号　特願2010-61451(P2010-61451)
(22)出願日　平成22年3月17日(2010.3.17)
(65)公開番号　特開2011-197809(P2011-197809A)
(43)公開日　平成23年10月6日(2011.10.6)
審査請求日　平成25年1月8日(2013.1.8)

(73)特許権者　000005223
富士通株式会社
神奈川県川崎市中原区上小田中４丁目１番１号
(74)代理人　100089118
弁理士　酒井　宏明
(72)発明者　大堀　順也
福岡県福岡市早良区百道浜二丁目２番１号
株式会社富士通九州システムズ内

審査官　野崎　大進

最終頁に続く

(54)【発明の名称】検索装置、検索方法および検索プログラム

(57)【特許請求の範囲】
【請求項１】
意味を持つ単語毎に文字列を区切る単語区切方式に基づいて区切られ、文書データに関連付けられた単語インデックスと、文字毎に文字列を区切る文字区切方式に基づいて区切られ、文書データに関連付けられた文字インデックスと、記号を含む所定の文字列の特徴を定義したパターンファイルとを記憶する記憶部と、
検索文字列を受け付け、前記検索文字列と前記パターンファイルとを基にして、前記パターンファイルに定義された特徴が、前記検索文字列と一致する場合には、前記単語インデックスを用いて文書データの検索を行うと判定し、前記パターンファイルに定義された特徴が、前記検索文字列と一致しない場合には、前記文字インデックスを用いて文書データの検索を行うと判定する判定部と、
前記判定部の判定結果に基づいて、前記単語インデックスまたは前記文字インデックスを用いて文書データの検索を実行する検索部と
を備えたことを特徴とする検索装置。
【請求項２】
前記パターンファイルは前記所定の文字列に含まれる文字の形式を示す情報を含み、
前記判定部は、受け付けた前記検索文字列に含まれる文字の形式と前記パターンファイルに示される文字の形式とが一致するか否かに基づき、前記単語インデックスを用いて文書データの検索を行うのか、前記文字インデックスを用いて文書データの検索を行うのかを判定する

ことを特徴とする請求項１に記載の検索装置。
【請求項３】
前記検索文字列を、複数の部分文字列に分割する文字列分割部を更に有し、前記判定部は、部分文字列毎に、前記パターンファイルに定義された特徴が、前記部分文字列と一致する場合には、前記単語インデックスを用いて文書データの検索を行うと判定し、前記パターンファイルに定義された特徴が、前記部分文字列と一致しない場合には、前記文字インデックスを用いて文書データの検索を行うと判定することを特徴とする請求項１または２に記載の検索装置。
【請求項４】
意味を持つ単語毎に文字列を区切る単語区切方式に基づいて区切られ、文書データに関連付けられた単語インデックスと、文字毎に文字列を区切る文字区切方式に基づいて区切られ、文書データに関連付けられた文字インデックスと、記号を含む所定の文字列の特徴を定義したパターンファイルとを記憶する記憶装置を有する検索装置が、
検索文字列を受け付け、前記検索文字列と前記パターンファイルとを基にして、前記パターンファイルに定義された特徴が、前記検索文字列と一致する場合には、前記単語インデックスを用いて文書データの検索を行うと判定し、前記パターンファイルに定義された特徴が、前記検索文字列と一致しない場合には、前記文字インデックスを用いて文書データの検索を行うと判定する判定ステップと、
前記判定ステップの判定結果に基づいて、前記単語インデックスまたは前記文字インデックスを用いて文書データの検索を実行する検索ステップと
を含むことを特徴とする検索方法。
【請求項５】
意味を持つ単語毎に文字列を区切る単語区切方式に基づいて区切られ、文書データに関連付けられた単語インデックスと、文字毎に文字列を区切る文字区切方式に基づいて区切られ、文書データに関連付けられた文字インデックスと、記号を含む所定の文字列の特徴を定義したパターンファイルとを記憶する記憶装置を有するコンピュータに、
検索文字列を受け付け、前記検索文字列と前記パターンファイルとを基にして、前記パターンファイルに定義された特徴が、前記検索文字列と一致する場合には、前記単語インデックスを用いて文書データの検索を行うと判定し、前記パターンファイルに定義された特徴が、前記検索文字列と一致しない場合には、前記文字インデックスを用いて文書データの検索を行うと判定する判定手順と、
前記判定ステップの判定結果に基づいて、前記単語インデックスまたは前記文字インデックスを用いて文書データの検索を実行する検索手順と
を実行させることを特徴とする検索プログラム。
【請求項６】
第１の区切方式に基づいて区切られ、文書データに関連付けられた第１のインデックスと、第２の区切方式に基づいて区切られ、文書データに関連付けられた第２のインデックスと、文字列に含まれる文字の形式を示すパターンファイルとを記憶する記憶部と、
検索文字列を受け付け、前記検索文字列に含まれる文字の形式と前記パターンファイルに示される文字の形式とが一致するか否かに基づき、前記第１のインデックスを用いて文書データの検索を行うのか、前記第２のインデックスを用いて文書データの検索を行うのかを判定する判定部と、
前記判定部の判定結果に基づいて、前記第１のインデックスまたは前記第２のインデックスを用いて文書データの検索を実行する検索部と
を備えたことを特徴とする検索装置。
【請求項７】
第１の区切方式に基づいて区切られ、文書データに関連付けられた第１のインデックスと、第２の区切方式に基づいて区切られ、文書データに関連付けられた第２のインデックスと、文字列に含まれる文字の形式を示すパターンファイルとを記憶する記憶装置を有する検索装置が、

検索文字列を受け付け、前記検索文字列に含まれる文字の形式と前記パターンファイルに示される文字の形式とが一致するか否かに基づき、前記第１のインデックスを用いて文書データの検索を行うのか、前記第２のインデックスを用いて文書データの検索を行うのかを判定する判定ステップと、
前記判定ステップの判定結果に基づいて、前記第１のインデックスまたは前記第２のインデックスを用いて文書データの検索を実行する検索ステップと
を含むことを特徴とする検索方法。
【請求項８】
第１の区切方式に基づいて区切られ、文書データに関連付けられた第１のインデックスと、第２の区切方式に基づいて区切られ、文書データに関連付けられた第２のインデックスと、文字列に含まれる文字の形式を示すパターンファイルとを記憶する記憶装置を有するコンピュータに、
検索文字列を受け付け、前記検索文字列に含まれる文字の形式と前記パターンファイルに示される文字の形式とが一致するか否かに基づき、前記第１のインデックスを用いて文書データの検索を行うのか、前記第２のインデックスを用いて文書データの検索を行うのかを判定する判定手順と、
前記判定手順の判定結果に基づいて、前記第１のインデックスまたは前記第２のインデックスを用いて文書データの検索を実行する検索手順と
を実行させることを特徴とする検索プログラム。
【発明の詳細な説明】
【技術分野】
【０００１】
本発明は、検索装置等に関する。
【背景技術】
【０００２】
複数の文書データから特定の文字列を検索する全文検索が知られている。この全文検索では、転置インデックスが用いられる。転置インデックスは、文字データに含まれる単語の位置情報等を格納する索引に対応する。転置インデックスを作成する方式には、大きく分けて文字区切方式と、単語区切方式とがある。
【０００３】
文字区切方式では、単語の意味を考えずに、文字単位で転置インデックスを作成するものである。文字区切方式で作成した転置インデックスを文字インデックスと表記する。文字インデックスを用いれば、完全な部分一致検索が可能である。しかし、検索キーワードと文字インデックスとを一文字ずつ比較する必要があり、検索時間を多く要してしまうという欠点がある。
【０００４】
単語区切方式では、意味のある単語単位で転置インデックスを作成するものである。単語区切方式で作成した転置インデックスを単語インデックスと表記する。単語インデックスを用いれば、検索キーワードを単語毎に比較するので、文字インデックスを利用する場合と比較して、検索時間を短縮することが可能となる。しかし、単語の区切り方によっては、検索漏れが発生する場合がある。
【０００５】
このように、文字区切方式および単語区切方式には、それぞれ長所、短所があるため、いかにして文字区切方式と単語区切方式とを使い分けるのかが重要になる。例えば、文字区切方式および単語区切方式を用いた従来技術として、検索キーワードの長さに応じて、文字インデックスと単語インデックスとを自動選択するという技術が開示されている。
【先行技術文献】
【特許文献】
【０００６】
【特許文献１】特開平１０－３０７８３５号公報

【特許文献２】特開２００１－３４６２３号公報
【特許文献３】特開２００８－７７６７３号公報
【発明の概要】
【発明が解決しようとする課題】
【０００７】
しかしながら、全文検索を行う文書データによっては、検索キーワードの長さが同じ場合でも、文字インデックスと単語インデックスとを使い分けた方が効率的な全文検索を行える場合がある。
【０００８】
例えば、バイオデータベースに記憶される文書データには、文書に加えて他のデータベースへリンクするためのＩＤ（Identification）が含まれている。一般的に、ＩＤ等の記号を有さない文書データに対しては、単語インデックスが有効であり、記号を有する文書データに対しては、文字インデックスが有効である。
【０００９】
ここで、「1.1.1.1ANDsuppressor」という検索式が与えられた場合を例にして説明する。かかる検索式に対して、文字インデックスを用いて全文検索を試みる場合には、「1.1.1.1」というＩＤを含み、かつ、「suppressor」という単語を含む文書データのみを検索することが好ましい。しかし、上記検索式に対して、文字インデックスを用いて全文検索を試みると、実際には、「1.1.1.11」、「1.1.1.12」等のＩＤを含む文書データもヒットしてしまう。
【００１０】
これに対して、上記検索式に対して、単語インデックスを用いて全文検索を試みると、「1.1.1.1」のＩＤを含む文書データのみを検索することが可能である。しかし、「suppressors」と「suppressor」とは完全に一致していないので、「suppressors」を含む文書データを検索することが出来なくなってしまう。
【００１１】
開示の技術は、上記に鑑みてなされたものであって、文書データの特性によらず、効率よく全文検索を実行することができる検索装置、検索方法および検索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】
【００１２】
本願の開示する検索装置は、一つの態様において、第１の区切方式に基づいて区切られ、文書データに関連付けられた第１のインデックスと、第２の区切方式に基づいて区切られ、文書データに関連付けられた第２のインデックスと、所定の文字の特徴を定義したパターンファイルを記憶する記憶部と、検索キーワードを受け付け、前記検索キーワードと前記パターンファイルとを基にして、前記第１のインデックスを用いて文書データの検索を行うのか、前記第２のインデックスを用いて文書データの検索を行うのかを判定する判定部と、前記判定部の判定結果に基づいて、前記第１のインデックスまたは前記第２のインデックスを用いて文書データの検索を実行する検索部とを備えたことを要件とする。
【発明の効果】
【００１３】
本願の開示する検索装置の一つの態様によれば、文書データの特性によらず、効率よく全文検索を実行することができるという効果を奏する。
【図面の簡単な説明】
【００１４】
【図１】図１は、本実施例１にかかる検索装置の構成を示す図である。
【図２】図２は、本実施例２にかかるシステムを示す図である。
【図３】図３は、本実施例２にかかる検索装置の構成を示す図である。
【図４】図４は、パターンファイルのデータ構造を示す図である。
【図５】図５は、本実施例２にかかる検索装置の処理手順を示すフローチャートである。

【図６】図６は、実施例にかかる検索装置を構成するコンピュータのハードウェア構成を示す図である。

【発明を実施するための形態】

【００１５】

以下に、本願の開示する検索装置、検索方法および検索プログラムの実施例を図面に基づいて詳細に説明する。なお、この実施例によりこの発明が限定されるものではない。

【実施例１】

【００１６】

図１は、本実施例１にかかる検索装置１００の構成を示す図である。図１に示すように、この検索装置１００は、記憶部１１０、判定部１２０、検索部１３０を有する。

【００１７】

記憶部１１０は、パターンファイル１１０ａ、第１のインデックス１１０ｂ、第２のインデックス１１０ｃを記憶する。パターンファイル１１０ａは、所定の文字の特徴を定義したデータである。第１のインデックス１１０ｂは、第１の区切方式に基づいて区切られ、文書データに関連付けられたデータである。第２のインデックス１１０ｃは、第２の区切方式に基づいて区切られ、文書データに関連付けられたデータである。

【００１８】

判定部１２０は、検索キーワードを受け付け、検索キーワードとパターンファイル１１０ａとを基にして、第１のインデックス１１０ｂを用いて検索を行うのか、第２のインデックス１１０ｃを用いて検索を行うのかを判定する。

【００１９】

検索部１３０は、判定部１２０の判定結果に基づいて、第１のインデックス１１０ｂまたは第２のインデックス１１０ｃを用いて文書データの検索を実行する。

【００２０】

上記の検索装置１００は、パターンファイル１１０ａを用いて、第１のインデックス１１０ｂを用いた検索を行うのか、第２のインデックス１１０ｃを用いた検索を行うのかを判定している。このため、検索キーワードの特徴に合わせて最適なインデックスを選択することができるので、文書データの特性によらず、効率よく全文検索を実行することができる。

【実施例２】

【００２１】

次に、本実施例２にかかるシステムの一例について説明する。図２は、本実施例２にかかるシステムを示す図である。図２に示すように、このシステムは、利用者端末６０、検索装置２００を有する。利用者端末６０と検索装置２００は、ネットワーク５０を介して接続される。

【００２２】

利用者端末６０は、検索装置２００に検索キーワードを送信し、検索キーワードに対する検索結果を検索装置２００から受信する装置である。

【００２３】

検索装置２００は、文書データの全文検索を行う装置である。図３は、本実施例２にかかる検索装置２００の構成を示す図である。図３に示すように、この検索装置２００は、記憶部２１０、インデクシング処理部２２０、入力受付部２３０、検索式解析処理部２４０、スコアリング処理部２５０、検索結果出力部２６０を有する。

【００２４】

記憶部２１０は、パターンファール２１０ａ、文書データ群２１０ｂ、単語インデックス２１０ｃ、文字インデックス２１０ｄを記憶する。

【００２５】

パターンファイル２１０ａは、所定の文字の特徴を定義したデータである。図４は、パターンファイル２１０ａのデータ構造を示す図である。図４に示すように、このパターンファイルは、Ｎｏとパターンとを有する。Ｎｏは、各パターンを識別するものである。パ

ターンは、所定の文字の特徴を正規表現で示したものである。ここで、文字には、一般的な文字のほかに、数字や記号等も含まれるものとする。

【００２６】

ここで、パターンの記載方法の一例について説明する。パターン中の［］は、［と］の中に書かれたいずれかの一文字に一致する文字、数字、記号を意味する。例えば、［０－９］は、１桁の数字を意味する。パターン中の｛ｎ，ｍ｝は、直前の文字がｎ回からｍ回まで繰り返されることを意味する。例えば、［０－９］｛１，３｝は、１桁、２桁、３桁の数字を意味する。

【００２７】

また、パターン中の＋は、直前の文字が１回以上繰り返されることを意味する。例えば、［０－９］＋は、数字からなる文字列を意味する。パターン中の＊は、直前の文字が０回以上繰り返されることを意味する。例えば、［０－９］＊は、空文字または数字からなる文字列を意味する。

【００２８】

図３の説明に戻る。文書データ群２１０ｂは、複数の文書データを含む。また、各文書データは、固有のＩＤが割り当てられ、各種の文字列を含む。

【００２９】

単語インデックス２１０ｃは、文書データ群２１０ｂに含まれる各文書データの単語と、この単語の存在する文書データのＩＤとを対応付けた転置インデックスである。文字インデックス２１０ｄは、文書データ群２１０ｂに含まれる各文書データの文字と、この文字の存在する文書データのＩＤとを対応付けた転置インデックスである。

【００３０】

インデクシング処理部２２０は、文書データ群２１０ｂから単語インデックス２１０ｃと文字インデックス２１０ｄを生成する処理部である。インデクシング処理部２２０は、単語区切方式により、文書データ群２１０ｂから単語インデックス２１０ｃを生成する。また、インデクシング処理部２２０は、文字区切方式により、文書データ群２１０ｂから文字インデックス２１０ｄを生成する。なお、単語区切方式による単語インデックス２１０ｃの生成は、周知の単語区切方式と同様である。文字区切方式による文字インデックス２１０ｄの生成は、周知の文字区切方式と同様である。

【００３１】

入力受付部２３０は、利用者端末６０から検索キーワードを受け付け、この検索キーワードを検索式解析処理部２４０に出力する。なお、入力受付部２３０は、検索装置２００に接続された入力装置から、検索キーワードを取得してもよい。入力装置は、例えば、マウスやキーボードに対応する。

【００３２】

検索式解析処理部２４０は、検索キーワードとパターンファイル２１０ａとを比較して、単語インデックス２１０ｃを用いて文書データの検索を行うのか、文字インデックス２１０ｄを用いて文書データの検索を行うのかを判定する処理部である。以下において、単語インデックス２１０ｃを用いて文書データの検索を行うことを、単語区切方式の検索と表記する。文字インデックス２１０ｄを用いて文書データの検索を行うことを、文字区切方式の検索と表記する。

【００３３】

まず、検索式解析処理部２４０は、検索キーワードに対して構文解析を実行する。例えば、検索キーワードを「1.1.1.1ANDsuppressor」とする。検索式解析処理部２４０が、検索キーワード「1.1.1.1ANDsuppressor」に対して構文解析を実行することで、この検索キーワードに含まれる条件文「AND」と、条件文を挟む文字列「1.1.1.1」、「suppressor」が抽出される。

【００３４】

検索式解析処理部２４０は、検索キーワードから抽出した各文字列と、パターンファイル２１０ａのパターンとをそれぞれ比較し、各文字列に対して、単語区切方式の検索を行

うのか、文字区切方式の検索を行うのかを判定する。

【００３５】

具体的には、検索式解析処理部２４０は、パターンファイル２１０ａのパターンのいずれかに文字列がマッチする場合には、文字区切方式の検索を行うと判定する。例えば、文字列「1.1.1.1」は、図４に示したパターンファイル２１０ａのＮｏ「２」のパターンとマッチする。このため、検索式解析処理部２４０は、文字列「1.1.1.1」に対して単語区切方式の検索を行うと判定する。

【００３６】

また、文字列「suppressor」は、図４に示したパターンファイル２１０ａのパターンとマッチしない。このため、検索式解析処理部２４０は、文字列「suppressor」に対して文字区切方式の検索を行うと判定する。

【００３７】

検索式解析処理部２４０は、文字列と判定結果とを対応付けたデータをスコアリング処理部２５０に出力する。また、検索式解析処理部２４０は、検索キーワードに含まれる条件文も合わせてスコアリング処理部２５０に出力する。

【００３８】

スコアリング処理部２５０は、検索式解析処理部２４０の文字列、文字列の判定結果、条件文を取得し、取得したデータに基づいて、検索キーワードに対応する文書データを検索する処理部である。ここでは一例として、文字列「1.1.1.1」に対応する判定結果が「単語区切方式の検索を行う」であり、文字列「suppressor」に対応する判定結果が「文字区切方式の検索を行う」であり、条件文が「AND」とする。

【００３９】

この場合には、スコアリング処理部２５０は、文字列「1.1.1.1」と、文字インデックス２１０ｄとを比較して、文字列「1.1.1.1」に対応する文書データを特定し、特定した文書データを文書データ群２１０ｂから取得する。また、スコアリング処理部２５０は、文字列「suppressor」と、単語インデックス２１０ｃとを比較して、文字列「suppressor」に対応する文書データを特定し、特定した文書データを文書データ群２１０ｂから取得する。

【００４０】

そして、スコアリング処理部２５０は、条件文が「AND」であるため、文字列「1.1.1.1」に対応する文書データと文字列「suppressor」に対応する文書データとを比較し、重複する文書データを検索結果出力部２６０に出力する。なお、条件文が「OR」の場合には、スコアリング処理部２５０は、条件文が「OR」であるため、文字列「1.1.1.1」に対応する文書データと文字列「suppressor」に対応する文書データとを検索結果出力部２６０に出力する。

【００４１】

スコアリング処理部２５０は、文書データを検索した場合に、文書データに含まれる文字列の頻度に応じて、文書データにスコアを付与してもよい。

【００４２】

検索結果出力部２６０は、スコアリング処理部２５０から受け付けた文書データを、利用者端末６０に通知する。検索結果出力部２６０は、文書データのスコアに応じて、利用者端末６０に表示させる文書データの順番を調整してもよい。また、検索結果出力部２６０は、検索装置２００に接続された表示装置に文書データを出力してもよい。表示装置は、例えば、モニタや液晶ディスプレイに対応する。

【００４３】

次に、本実施例２にかかる検索装置２００の処理手順について説明する。図５は、本実施例２にかかる検索装置２００の処理手順を示すフローチャートである。図５に示すように、検索装置２００は、検索キーワードを取得し（ステップＳ１０１）、構文解析を実行する（ステップＳ１０２）。

【００４４】

検索装置２００は、パターンファイル２１０ａから未選択のパターンを取得し（ステップＳ１０３）、検索キーワードはパターンにマッチするか否かを判定する（ステップＳ１０４）。検索装置２００は、検索キーワードがパターンにマッチする場合には（ステップＳ１０４，Ｙｅｓ）、単語区切方式の検索を行うと判定し（ステップＳ１０５）、ステップＳ１０８に移行する。

【００４５】

一方、検索装置２００は、検索キーワードがパターンにマッチしない場合には（ステップＳ１０４，Ｎｏ）、未選択のパターンが存在するか否かを判定する（ステップＳ１０６）。検索装置２００は、未選択のパターンが存在する場合には（ステップＳ１０６，Ｙｅｓ）、ステップＳ１０３に移行する。

【００４６】

一方、検索装置２００は、未選択のパターンが存在しない場合には（ステップＳ１０６，Ｎｏ）、文字区切方式の検索を行うと判定し（ステップＳ１０７）、検索を実行する（ステップＳ１０８）。

【００４７】

上述してきたように、本実施例２にかかる検索装置２００は、パターンファイル２１０ａを用いて、単語区切方式の検索を行うのか、文字区切方式の検索を行うのかを判定している。このため、検索キーワードの特徴に合わせて最適なインデックスを選択することができるので、文書データの特性によらず、効率よく全文検索を実行することができる。

【００４８】

また、本実施例２では、検索キーワードが検索式の場合に、この検索式を複数の部分キーワードに分割し、部分キーワード毎に単語区切方式の検索を行うのか、文字区切方式の検索を行うのかを判定している。このため、既存の技術を踏襲した検索式をそそのまま利用して、全文検索を実行することができる。

【００４９】

また、本実施例２のパターンファイル２１０ａは、利用者単位の好みに合わせて容易にカスタマイズすることができる。

【００５０】

ところで、図３に示した検索装置２００の各構成要素は機能概念的なものであり、必ずしも物理的に図示の如く構成されていることを要しない。すなわち、検索装置２００の分散・統合の具体的形態は図示のものに限られず、その全部または一部を、各種の負荷や使用状況などに応じて、任意の単位で機能的または物理的に分散・統合して構成することができる。例えば、記憶部２１０を、着脱可能な外部装置または携帯端末等に搭載し、かかる外部装置または携帯端末等を検索装置２００に有線または無線で接続するようにしてもよい。

【００５１】

なお、検索装置２００は、既知のパーソナルコンピュータ、ワークステーション、携帯電話、ＰＨＳ端末、移動体通信端末またはＰＤＡなどの情報処理装置に、検索装置２００の各機能を搭載することによって実現することもできる。

【００５２】

図６は、実施例にかかる検索装置を構成するコンピュータのハードウェア構成を示す図である。図６に示すように、このコンピュータ３００は、各種演算処理を実行するＣＰＵ（Central Processing Unit）３０１と、ユーザからのデータの入力を受け付ける入力装置３０２と、モニタ３０３を有する。また、コンピュータ３００は、記憶媒体からプログラム等を読取る媒体読み取り装置３０４と、ネットワークを介して他のコンピュータとの間でデータの授受を行うネットワークインターフェース装置３０５を有する。また、コンピュータ３００は、各種情報を一時記憶するＲＡＭ（Random Access Memory）３０６と、ハードディスク装置３０７を有する。各装置３０１～３０７は、バス３０８に接続される。

【００５３】

そして、ハードディスク装置３０７には、図３に示した検索式解析処理部２４０、スコアリング処理部２５０、インデクシング処理部２２０と同様の機能を有する検索プログラム３０７ａを記憶する。また、ハードディスク装置３０７は、図３に示した各種データ２１０ａ～２１０ｄにそれぞれ対応する各種データ３０７ｂを記憶する。

【００５４】

ＣＰＵ３０１が検索プログラム３０７ａをハードディスク装置３０７から読み出してＲＡＭ３０６に展開することにより、検索プログラム３０７ａは、検索プロセス３０６ａとして機能するようになる。また、ＣＰＵ３０１は、各種データ３０７ｂをＲＡＭ３０６に読み出す。検索プロセス３０６ａは、各種データ３０６ｂを利用して、全文検索を実行する。

【００５５】

なお、上記の検索プログラム３０７ａは、必ずしもハードディスク装置３０７に格納されている必要はなく、ＣＤ－ＲＯＭ等の記憶媒体に記憶されたプログラムを、コンピュータ３００が読み出して実行するようにしてもよい。また、公衆回線、インターネット、ＬＡＮ（Local Area Network）、ＷＡＮ（Wide Area Network）等にこのプログラムを記憶させておき、コンピュータ３００がこれらからプログラムを読み出して実行するようにしてもよい。

【００５６】

以上の各実施例を含む実施形態に関し、さらに以下の付記を開示する。

【００５７】

（付記１）第１の区切方式に基づいて区切られ、文書データに関連付けられた第１のインデックスと、第２の区切方式に基づいて区切られ、文書データに関連付けられた第２のインデックスと、所定の文字の特徴を定義したパターンファイルとを記憶する記憶部と、

検索キーワードを受け付け、前記検索キーワードと前記パターンファイルとを基にして、前記第１のインデックスを用いて文書データの検索を行うのか、前記第２のインデックスを用いて文書データの検索を行うのかを判定する判定部と、

前記判定部の判定結果に基づいて、前記第１のインデックスまたは前記第２のインデックスを用いて文書データの検索を実行する検索部と

を備えたことを特徴とする検索装置。

【００５８】

（付記２）前記検索キーワードを、複数の部分キーワードに分割するキーワード分割部を更に有し、前記判定部は、部分キーワード毎に前記第１のインデックスを用いて文書データの検索を行うのか、前記第２のインデックスを用いて文書データの検索を行うのかを判定することを特徴とする付記１に記載の検索装置。

【００５９】

（付記３）前記第１の区切方式は、意味を持つ単語毎に文字列を区切る単語区切方式であり、前記判定部は、前記パターンファイルに定義された特徴が、前記検索キーワードと一致する場合には、前記第１のインデックスを用いて文書データの検索を行うと判定することを特徴とする付記１または２に記載の検索装置。

【００６０】

（付記４）第１の区切方式に基づいて区切られ、文書データに関連付けられた第１のインデックスと、第２の区切方式に基づいて区切られ、文書データに関連付けられた第２のインデックスと、所定の文字の特徴を定義したパターンファイルとを記憶する記憶装置を有する検索装置が、

検索キーワードを受け付け、前記検索キーワードと前記パターンファイルとを基にして、前記第１のインデックスを用いて文書データの検索を行うのか、前記第２のインデックスを用いて文書データの検索を行うのかを判定する判定ステップと、

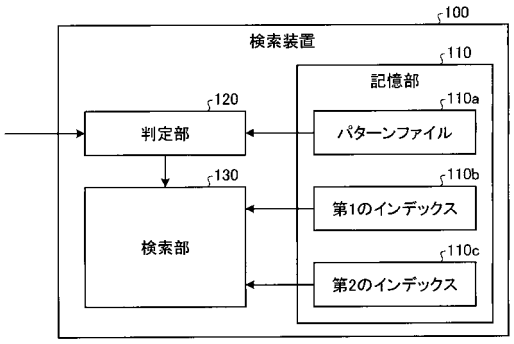
前記判定ステップの判定結果に基づいて、前記第１のインデックスまたは前記第２のインデックスを用いて文書データの検索を実行する検索ステップと

を含むことを特徴とする検索方法。

【００６１】
（付記５）前記検索キーワードを、複数の部分キーワードに分割するキーワード分割ステップを更に含み、前記判定ステップでは、部分キーワード毎に前記第１のインデックスを用いて文書データの検索を行うのか、前記第２のインデックスを用いて文書データの検索を行うのかを判定することを特徴とする付記４に記載の検索方法。
【００６２】
（付記６）前記第１の区切方式は、意味を持つ単語毎に文字列を区切る単語区切方式であり、前記判定ステップでは、前記パターンファイルに定義された特徴が、前記検索キーワードと一致する場合には、前記第１のインデックスを用いて文書データの検索を行うと判定することを特徴とする付記４または５に記載の検索方法。
【００６３】
（付記７）第１の区切方式に基づいて区切られ、文書データに関連付けられた第１のインデックスと、第２の区切方式に基づいて区切られ、文書データに関連付けられた第２のインデックスと、所定の文字の特徴を定義したパターンファイルとを記憶する記憶装置を有するコンピュータに、
検索キーワードを受け付け、前記検索キーワードと前記パターンファイルとを基にして、前記第１のインデックスを用いて文書データの検索を行うのか、前記第２のインデックスを用いて文書データの検索を行うのかを判定する判定手順と、
前記判定ステップの判定結果に基づいて、前記第１のインデックスまたは前記第２のインデックスを用いて文書データの検索を実行する検索手順と
を実行させることを特徴とする検索プログラム。
【００６４】
（付記８）前記検索キーワードを、複数の部分キーワードに分割するキーワード分割手順を更にコンピュータに実行させ、前記判定手順は、部分キーワード毎に前記第１のインデックスを用いて文書データの検索を行うのか、前記第２のインデックスを用いて文書データの検索を行うのかを判定することを特徴とする付記７に記載の検索プログラム。
【００６５】
（付記９）前記第１の区切方式は、意味を持つ単語毎に文字列を区切る単語区切方式であり、前記判定手順は、前記パターンファイルに定義された特徴が、前記検索キーワードと一致する場合には、前記第１のインデックスを用いて文書データの検索を行うと判定することを特徴とする付記７または８に記載の検索プログラム。
【符号の説明】
【００６６】
１００　検索装置
１１０ａ　パターンファイル
１１０ｂ　第１のインデックス
１１０ｃ　第２のインデックス
１２０　判定部
１３０　検索部

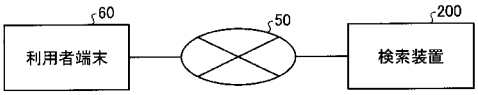
【図１】

本実施例1にかかる検索装置の構成を示す図



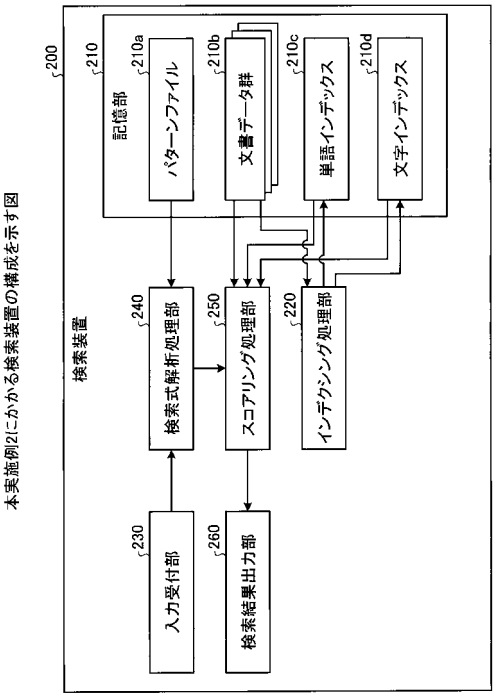
【図２】

本実施例2にかかるシステムを示す図



【図３】

本実施例2にかかる検索装置の構成を示す図



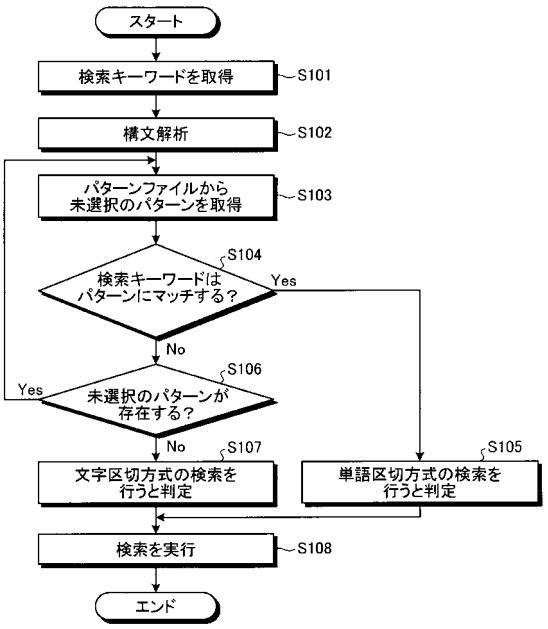
【図４】

パターンファイルのデータ構造を示す図

No	パターン
1	[0-9]{3}[0-9]*
2	[0-9]{1,3}.[0-9]{1,3}.[0-9]{1,3}.[0-9]{1,3}
3	ENSG[0-9]+
4	NP_[0-9]+
5	IPI[0-9]+
6	[a-z]{3,4}:[a-zA-z0-9]+
…	…

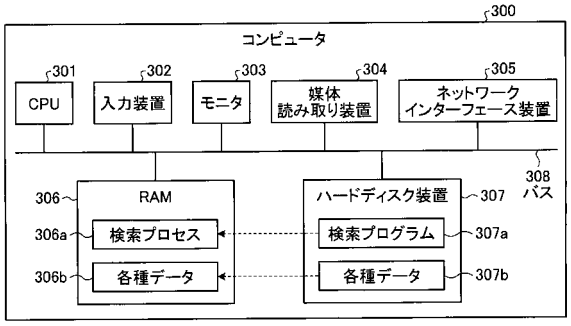
【図５】

本実施例2にかかる検索装置の処理手順を示すフローチャート



【図６】

実施例にかかる検索装置を構成するコンピュータのハードウェア構成を示す図



フロントページの続き

(56)参考文献 特開平０９－２５９１３２（ＪＰ，Ａ）
特開２００８－０７７６７３（ＪＰ，Ａ）
特開２００１－０３４６２３（ＪＰ，Ａ）
特開平１０－３０７８３５（ＪＰ，Ａ）

(58)調査した分野(Int.Cl.，ＤＢ名)
Ｇ０６Ｆ　１７／３０
ＪＳＴＰｌｕｓ（ＪＤｒｅａｍＩＩＩ）