

(19) 日本国特許庁(JP)

## (12) 特許公報(B2)

(11) 特許番号

特許第5247336号  
(P5247336)

(45) 発行日 平成25年7月24日(2013.7.24)

(24) 登録日 平成25年4月19日(2013.4.19)

(51) Int.Cl.

F 1

G06F 17/30 (2006.01)  
G06F 3/0482 (2013.01)G06F 17/30 330Z  
G06F 3/048 654B  
G06F 17/30 340Z

請求項の数 9 (全 13 頁)

(21) 出願番号	特願2008-255501 (P2008-255501)
(22) 出願日	平成20年9月30日 (2008.9.30)
(65) 公開番号	特開2010-86332 (P2010-86332A)
(43) 公開日	平成22年4月15日 (2010.4.15)
審査請求日	平成23年8月4日 (2011.8.4)

(73) 特許権者	000001007 キヤノン株式会社 東京都大田区下丸子3丁目30番2号
(74) 代理人	100090273 弁理士 國分 孝悦
(72) 発明者	戸島 英一朗 東京都大田区下丸子3丁目30番2号 キ ヤノン株式会社内

審査官 野崎 大進

最終頁に続く

(54) 【発明の名称】検索方法及び検索装置

## (57) 【特許請求の範囲】

## 【請求項 1】

検索対象のデータ集合を検索する検索装置における検索方法であって、複数の検索式を組み合わせて検索式セットを生成する検索式セット生成工程と、前記検索式セット生成工程で生成された前記検索式セットに含まれる各検索式の前記データ集合内のヒット率又はヒット数に基づいて、前記ヒット率又はヒット数が均等に近い検索式セットが優先的に提示されるよう提示順位スコアを算出する検索式セットスコア算出工程と、

前記検索式セット生成工程で生成された検索式セットを前記検索式セットスコア算出工程で算出された提示順位スコア順に提示する検索式セット提示工程と、

前記検索式セット提示工程で提示された検索式セットからユーザに選択された検索式に基づいて前記データ集合に対して検索を実行する検索式実行工程と、  
を含むことを特徴とする検索方法。

## 【請求項 2】

前記検索式セット提示工程では、前記検索式セット生成工程で生成された検索式セットとともに、検索式セットに含まれる各検索式に対応する前記ヒット率又はヒット数を前記検索式セットスコア算出工程で算出された提示順位スコア順に提示することを特徴とする請求項 1 に記載の検索方法。

## 【請求項 3】

検索対象のデータ集合を検索する検索装置における検索方法であって、

10

20

複数の検索式を組み合わせて検索式セットを生成する検索式セット生成工程と、前記検索式セット生成工程で生成された前記検索式セットに含まれる各検索式の前記データ集合内のヒット率又はヒット数に基づいて、前記データ集合内の重複度が少ない検索式セットが優先的に提示されるよう提示順位スコアを算出する検索式セットスコア算出工程と、

前記検索式セット生成工程で生成された検索式セットを前記検索式セットスコア算出工程で算出された提示順位スコア順に提示する検索式セット提示工程と、

前記検索式セット提示工程で提示された検索式セットからユーザに選択された検索式に基づいて前記データ集合に対して検索を実行する検索式実行工程と、  
を含むことを特徴とする検索方法。

10

#### 【請求項 4】

検索対象のデータ集合を検索する検索装置であって、

複数の検索式を組み合わせて検索式セットを生成する検索式セット生成手段と、

前記検索式セット生成手段で生成された前記検索式セットに含まれる各検索式の前記データ集合内のヒット率又はヒット数に基づいて、前記ヒット率が均等に近い検索式セットが優先的に提示されるよう提示順位スコアを算出する検索式セットスコア算出手段と、

前記検索式セット生成手段で生成された検索式セットを前記検索式セットスコア算出手段で算出された提示順位スコア順に提示する検索式セット提示手段と、

前記検索式セット提示手段で提示された検索式セットからユーザに選択された検索式に基づいて前記データ集合に対して検索を実行する検索式実行手段と、  
を含むことを特徴とする検索装置。

20

#### 【請求項 5】

前記検索式セット提示手段は、前記検索式セット生成手段で生成された検索式セットとともに、検索式セットに含まれる各検索式に対応する前記ヒット率又はヒット数を前記検索式セットスコア算出手段で算出された提示順位スコア順に提示することを特徴とする請求項 4 に記載の検索装置。

#### 【請求項 6】

検索対象のデータ集合を検索する検索装置であって、

複数の検索式を組み合わせて検索式セットを生成する検索式セット生成手段と、

前記検索式セット生成手段で生成された前記検索式セットに含まれる各検索式の前記データ集合内のヒット率又はヒット数に基づいて、前記データ集合内の重複度が少ない検索式セットが優先的に提示されるよう提示順位スコアを算出する検索式セットスコア算出手段と、

前記検索式セット生成手段で生成された検索式セットを前記検索式セットスコア算出手段で算出された提示順位スコア順に提示する検索式セット提示手段と、

前記検索式セット提示手段で提示された検索式セットからユーザに選択された検索式に基づいて前記データ集合に対して検索を実行する検索式実行手段と、  
を含むことを特徴とする検索装置。

30

#### 【請求項 7】

検索対象のデータ集合を検索するコンピュータを、

40

複数の検索式を組み合わせて検索式セットを生成する検索式セット生成手段と、

前記検索式セット生成手段で生成された前記検索式セットに含まれる各検索式の前記データ集合内のヒット率又はヒット数に基づいて、前記ヒット率が均等に近い検索式セットが優先的に提示されるよう提示順位スコアを算出する検索式セットスコア算出手段と、

前記検索式セット生成手段で生成された検索式セットを前記検索式セットスコア算出手段で算出された提示順位スコア順に提示する検索式セット提示手段と、

前記検索式セット提示手段で提示された検索式セットからユーザに選択された検索式に基づいて前記データ集合に対して検索を実行する検索式実行手段と、  
して機能させることを特徴とするプログラム。

#### 【請求項 8】

50

検索対象のデータ集合を検索するコンピュータを、  
複数の検索式を組み合わせて検索式セットを生成する検索式セット生成手段と、  
前記検索式セット生成手段で生成された前記検索式セットに含まれる各検索式の前記データ集合内でのヒット率又はヒット数に基づいて、前記データ集合内での重複度が少ない検索式セットが優先的に提示されるよう提示順位スコアを算出する検索式セットスコア算出手段と、

前記検索式セット生成手段で生成された検索式セットを前記検索式セットスコア算出手段で算出された提示順位スコア順に提示する検索式セット提示手段と、

前記検索式セット提示手段で提示された検索式セットからユーザに選択された検索式に基づいて前記データ集合に対して検索を実行する検索式実行手段と、  
10  
して機能させることを特徴とするプログラム。

**【請求項 9】**

請求項 7 又は 8 に記載のプログラムを記憶したコンピュータにより読み取り可能な記憶媒体。

**【発明の詳細な説明】**

**【技術分野】**

**【0001】**

本発明は、検索方法及び検索装置に関する。

**【背景技術】**

**【0002】**

近年、記憶メディアの大容量化に伴って複合機等の機器が管理する文書の文書数が増大し、全貌が把握しづらいという問題が起きており、それを解決するために検索の需要が高まっている。

ところが、P C 等のテキスト情報機器においては検索クエリを入力するための U I 環境が充実しているので検索の遂行が比較的容易であるのに対し、複合機等の一般的機器においては、テキスト入力の環境が整備されているとはいえない。よって、検索クエリの指示を与えることが困難である。

**【0003】**

検索クエリ指示のユーザ負荷を軽減するためには、従来から検索式を自動生成するという取り組みがなされてきた。

例えば特許文献 1においては、データ集合を指定すると、それと同程度の内容の文書を検索できるように検索式を自動生成する手法が開示されている。

**【0004】**

**【特許文献 1】特開 2002 - 183194 号公報**

**【発明の開示】**

**【発明が解決しようとする課題】**

**【0005】**

ところが、機器から検索を指示するために必要な検索クエリは、上記のように、元文書と似た文書を検索するための検索であるとは限らない。一般的には、与えられたデータ集合を検索によって分割していく、それを繰り返すことによって目的の文書を見つけ出したい。このとき必要とされるのは、指定されたデータ集合をある程度の単位に分割するための検索式の集合体（検索式セット）であり、それを自動生成してくれればテキスト入力を回避することができるが、そのような提案は今までになされなかった。

そのため、自分で検索式を考案・入力していく必要があったため、複合機のような一般的機器において検索指示を出しづらく、著しく使い勝手が良くなかった。

**【0006】**

本発明はこのような問題点に鑑みなされたもので、データ集合をある程度の大きさに分割できる検索式のセットを自動で生成させ、ユーザはそのうちの一つを選択するだけで、検索が実行でき、データ集合を細分していくことができるようすることを目的とする。

**【課題を解決するための手段】**

10

20

30

40

50

**【0007】**

そこで、本発明は、検索対象のデータ集合を検索する検索装置における検索方法であつて、複数の検索式を組み合わせて検索式セットを生成する検索式セット生成工程と、前記検索式セット生成工程で生成された前記検索式セットに含まれる各検索式の前記データ集合内でのヒット率又はヒット数に基づいて、前記ヒット率又はヒット数が均等に近い検索式セットが優先的に提示されるよう提示順位スコアを算出する検索式セットスコア算出工程と、前記検索式セット生成工程で生成された検索式セットを前記検索式セットスコア算出工程で算出された提示順位スコア順に提示する検索式セット提示工程と、前記検索式セット提示工程で提示された検索式セットからユーザに選択された検索式に基づいて前記データ集合に対して検索を実行する検索式実行工程と、を含むことを特徴とする。

10

**【0008】**

かかる構成とすることにより、データ集合をある程度の大きさに分割できる検索式のセットを自動で生成させ、ユーザはそのうちの一つを選択するだけで、検索が実行でき、データ集合を細分していくことができるようになることができる。

**【0009】**

また、本発明は、検索装置、プログラム及び記憶媒体としてもよい。

**【発明の効果】****【0010】**

本発明によれば、データ集合をある程度の大きさに分割できる検索式のセットを自動で生成させ、ユーザはそのうちの一つを選択するだけで、検索が実行でき、データ集合を細分していくことができるようになることができるるので使い勝手が向上する。

20

**【発明を実施するための最良の形態】****【0011】**

以下、本発明の実施形態について図面に基づいて説明する。

**【0012】****<実施形態1>**

図1は、検索装置の構成を示すブロック図である。

図示の構成において、CPUはマイクロプロセッサであり、画像処理、検索式セット提示処理、次候補処理、カーソル移動処理、検索処理、文書登録処理、表示処理等のための演算、論理判断等を行い、バスを介してバスに接続された各構成要素を制御する。

30

BUSはバスであり、CPUの制御対象である各構成要素を指示するアドレス信号、コントロール信号を転送する。また、BUSは各構成要素間のデータ転送を行う。

ROMは読み出し専用の不揮発性メモリである。ROMはCPUによるブートプログラム及び各種処理で使用する初期データ等を記憶する。ブートプログラムはシステム起動時に初期データをRAMにロードし、CPUに制御プログラムを実行させる。制御プログラムについては、後のフローチャートを参照して詳述する。

**【0013】**

入力装置はボタン、タッチパネル等である。

表示装置は液晶ディスプレイ等である。

RAMは読み書き可能なランダムアクセスメモリであって、各構成要素からの各種データの一次記憶に用いる。RAMには処理に必要な文書データを記憶するバッファ、表示に必要な表示データを一時記憶するバッファ、検索式セット候補テーブル等の処理中に値変更が行われる各種データが格納される。

40

通信装置はネットワークコントローラである。通信装置は通信回線を介して外部とのデータ交換を行う装置である。

HDはハードディスクであり、CPUにより実行される制御プログラム等が格納される。制御プログラムは必要に応じてRAMに示す主記憶にロードされ参照される。また、HDには本検索装置で管理する文書データベース、重要語を記憶する重要語テーブル等が格納される。

**【0014】**

50

かかる各構成要素からなる検索装置は、入力装置等からの各種イベントに応じて作動するものである。入力装置等からのインタラプトが供給されると信号が C P U に送られ、それに伴ってイベントが発生し、イベントに応じて C P U が R O M 又は R A M 内に記憶される各種命令（又はプログラム）を読み出し、その実行によって各種の制御が行われる。

#### 【0015】

図 2 は、検索画面の例を示す図である。ユーザが、所定の（又はユーザが指定した）、検索対象であるデータ集合の一例である文書集合に対して検索を指示すると、2 - 1 で示す検索の初期画面が表示される。初期画面にはその文書集合に対して最も推奨される検索式セットが表示される。2 - 1 - 1 は検索の属性、2 - 1 - 2 は属性に対する値を意味している。この画面の検索式セットは、「組織体 = 経理」「組織体 = 人事」「組織体 = 涉外」 「組織体 = 開発」及び上記以外の場合をまとめた「その他」の検索式からなる。2 - 1 - 3 はそれぞれの検索式に対応した文書集合内のヒット率が表示される。このヒット率によりユーザは自分が選択しようとしている検索式による絞込みが徒労に終わらないかどうかを判断することができる。2 - 1 - 4 は次候補のボタンであり、ユーザはこの検索式セットでは選択しづらい、或いは、選択しても無駄であると判断したときに、このボタンを押下し、次の選択式セットを表示させる。10

ユーザが、次候補を指示すると 2 - 2 の画面になり、次の選択式セットが表示される。ここで、更に次候補を指示すると 2 - 3 の画面が表示される。このように、検索装置は、後述する提示順位スコア順に（つまり、2 - 1、2 - 2、2 - 3 等の順に）検索式セットを表示装置等に表示することでユーザに提示する。20

もし、ユーザが 2 - 2 で「カメラ」を選択すると「製品 = カメラ」の検索が行われ、画面 2 - 4 となり、条件を充たす文書群が検索結果として一覧表示される。

#### 【0016】

図 3 は、文書を管理するための情報である文書データベースの構成の一例を示した図である。

3 - 1 は文書 I D フィールドであり、その文書を特定する I D が格納される。3 - 2 は作成者フィールドであり、その文書の作成者が格納される。3 - 3 は作成日フィールドであり、その文書の作成日が格納される。3 - 4 は組織体フィールドであり、その文書に関係する組織体が格納される。3 - 5 は製品フィールドであり、その文書がどの製品に関係するかが格納される。3 - 2 から 3 - 5 まではメタデータと言われるのであり、その文書の付加すべき情報が格納される。3 - 6 は全文キーワードフィールドである。文書の附加情報でありメタデータではあるが、3 - 5 までの情報とは異なり、文書内容から自動抽出された情報が格納される。3 - 7 は文書内容であり、各レコードにぶら下がるように文書の本体が格納される。これらの情報は所謂リレーションナルデータベースに格納され、索引化されているので、検索装置は、属性に基づいてヒット件数を調べるだけであれば速やかに行うことができる。30

#### 【0017】

図 4 は、重要語テーブルの一例を示した図である。文書本体から自動抽出すべき単語のリストである。先の文書データベースの全文キーワード情報の作成は文書本体のテキストからここで示す単語が存在するかどうかで作成される。各運用組織によって自由にテーブルを構成すればよいが、ユーザは、例えば、その会社で扱う製品の名前等、検索が発生すると思われる単語を予め登録する。40

#### 【0018】

図 5 は、検索式セット候補テーブルの構成の一例を示した図である。検索装置は、検索が指示されたときに提示すべき検索式セットを作成していくが、その過程において検索式セット候補テーブルは作成・維持される。

検索式セット候補テーブルは、いくつかの検索式セット 5 - 1 の連なりからなる。各検索式は属性 5 - 2、一連の検索式 5 - 3 及びスコア 5 - 6 からなる。属性 5 - 2 はその検索式セットの観点を示し、文書データベースのメタデータとして管理されるフィールド名が使用され、指定文書集合内で実際に出現する属性が使用される。各検索式は値 5 - 4 と50

文書集合内ヒット率 5 - 5 とからなる。値 5 - 4 は指定文書集合内で属性 5 - 2 の値として実際に出現するものが使用される。文書集合内ヒット率 5 - 5 は、その文書集合内で「属性 = 値」を充たす文書の割合を格納する。最後のスコア 5 - 6 は、提示順位スコアのことであり、検索式セットを提示すべき順位の根拠として使用される。提示順位スコアは、各文書集合内ヒット率から後述する換算式で求められる。

#### 【0019】

図 6 は、検索式セットを提示する順位のもとになる提示順位スコアの算出式の一例を示した図である。

ある検索式セット内の各検索式による文書集合内ヒット率に対して、その二乗平均を提示順位スコアとしている。これは統計学でいう分散を示すものである。分散は、一般に二乗平均から平均の二乗を減じて算出されるが、この検索式セットの比較においては平均が概ね等しいと見なして、二乗平均を指標として用いている。この値が大きいほど統計データ、即ち、ヒット率が分散していることを示し、ヒット率のばらつきが多くことを意味している。逆にこの値が小さいほどヒット率が均等化していることを意味している。

10

なお、ここで算出の根拠とするヒット率は必ずしも百分率のような比のみに限定するわけではない。何件ヒットするかというヒット数の一例である絶対件数を使用しても、全体集合の枠内で計算しているので、本質的に同様の処理となる。本実施形態では絶対件数の両方を包含する概念として単にヒット率と呼んで発明を説明している。

#### 【0020】

以下、上述の動作をフローチャートに従って説明する。

20

図 7 は、検索装置の動作、より具体的には CPU の処理手順を示すフローチャートである。ステップ S7 - 1 は、システムの初期化処理であり、各種パラメータの初期化や初期画面の表示等を行う処理である。ステップ S7 - 2 は、ボタン押下等の入力装置から何らかのイベントが発生するのを CPU において待つ処理である。イベントが発生すると、ステップ S7 - 3 において、CPU がこのイベントを判別し、イベントの種類に応じて各種の処理に分岐する。各種イベントに対応した分岐先の複数の処理をステップ S7 - 4 という形でまとめて表現している。後述する図 8 で詳述する検索式セット提示処理、後述する図 10 で詳述する次候補処理、後述する図 11 で詳述するカーソル移動処理、後述する図 12 で詳述する検索実行処理、後述する図 13 で詳述する文書登録処理がこの分岐先の一部となる。他の処理としては、詳細は記述していないが、文書データベース管理処理、文書編集処理等の処理がある。ステップ S7 - 5 は、上記の各処理の処理結果や処理終了を通知し、画面を各処理の指示に従って表示する処理である。エラーがあった場合のエラー表示、正常な処理が行われた場合に設定されている表示データの画面表示への反映等、通常広く行われる処理である。

30

#### 【0021】

図 8 は、ステップ S7 - 4 の一部である検索式セット提示処理を詳細化したフローチャートである。ステップ S8 - 1 において、CPU は、処理対象となっている文書集合を特定する。文書データベースにある文書の全体を対象とする場合もあるが、例えば、検索結果を一度表示し、絞込み検索を指示している場合、CPU は、文書データベースの一部の絞り込まれた集合を指定文書集合ととらえて処理を行うことになる。ステップ S8 - 2 において、CPU は、文書集合内で使用されている属性の候補を抽出する。文書データベース全体では存在する属性であっても、その部分集合である指定文書集合では使用されていない場合があり、このような場合、当該属性は予め除外されることになる。ステップ S8 - 3 において、CPU は、属性候補のうち、ある属性を選択する。ステップ S8 - 4 において、CPU は、図 9 で詳述するように選択した属性についての検索式の候補を検索式セット候補テーブル中に作成し、設定する（検索式セット生成）。設定が終わるとステップ S8 - 5 において、CPU は、全ての属性を処理し尽くしたかを確認し、未処理の属性があるときはステップ S8 - 6 において次の属性を選択し、ステップ S8 - 3 にループする。ステップ S8 - 5 において、全ての属性を処理終了しているときは、CPU は、ステップ S8 - 7 に進み、検索式セット候補テーブル上に作成された検索式セット群をスコアが

40

50

小さい順にソートする。スコアが小さいほどヒット率群が集中、即ち、均等化しているので、ヒット率群が最も均等な検索式セットが先頭にくることになる。ステップS8-8において、CPUは、ソートされた先頭の検索式セットを初期検索式セットとして表示するために表示データに設定し、リターンする。

#### 【0022】

図9は、ステップS8-4における検索式セット候補設定の処理詳細を示したフローチャートである。ステップS9-1において、CPUは、必要な変数を初期設定する。ステップS9-2において、CPUは、指定文書集合において指定属性の値として出現するものを候補としてリストアップする。ステップS9-3において、CPUは、候補のうちから値を一つ取り出す。ステップS9-4において、CPUは、指定文書集合内で「属性=値」となる件数を求め、文書集合内ヒット率を算出する。ステップS9-5において、CPUは、求めた値と文書集合内ヒット率とを検索式セット候補テーブル内の該当検索式の然るべきフィールドに設定する。ステップS9-6において、CPUは、全ての値候補を処理し終えたかを確認し、未処理の値があるときはステップS9-7において次の値を選択し、ステップS9-3にループする。ステップS9-6において、全ての値の処理が終了しているときは、CPUは、ステップS9-8において、この検索式セットの提示順位スコアを図6に示した算出式に基づいて算出（検索式セットスコア算出）し、検索式セット候補テーブルに格納する。その後リターンする。ここで、図6通りの算出式で提示順位スコアを設定すればヒット率が均等に近い検索式セットが優先的に提示されることになる。或いは、データ集合内（文書集合内）での重複度が少ない検索式セットを優先的に提示したければ、提示順位スコアとして各文書集合内ヒット率の平均を採用することもできる。ヒット率の平均は、各検索式候補でヒットするデータの重複が全く無い場合に（全体集合数）/（検索式候補数）に一致するが、データの重複がある場合には、その分だけ値が大きくなるので、データ重複がない検索式セットの提示順位スコアが小さく算出されることになる。更に、ヒット率の均等化とデータの重複度との両方を加味して優先するには、提示順位スコアとして、図6の算出式とヒット率の平均を、それぞれの重複具合に応じて適切な重みをつけて加算した値を採用することにより行なうことができる。

#### 【0023】

図10は、ステップS7-4の一部である次候補処理の詳細を示したフローチャートである。ステップS10-1において、CPUは、提示順位スコアに基づいて、検索式セット候補テーブルから次の候補（次の検索式セット）を取得する。ステップS10-2において、CPUは、取得した検索式セットを表示データに展開して設定し、リターンする。

#### 【0024】

図11は、ステップS7-4の一部であるカーソル移動処理を詳細化したフローチャートである。ステップS11-1において、CPUは、ユーザの操作に基づいて、どの位置にカーソルが動いたかを求める。ステップS11-2において、CPUは、求まった方向に実際にカーソルを移動させる。ステップS11-3において、CPUは、カーソル移動後の表示データを設定し、リターンする。

#### 【0025】

図12は、ステップS7-4の一部である検索実行処理を詳細化したフローチャートである。ステップS12-1において、CPUは、検索式選択を行い、カーソル位置に相当する検索式を検索式セット候補テーブルから取得する。ステップS12-2において、CPUは、検索式実行処理を行い、指定された文書集合に対して取得した検索式を実行する。ステップS12-3において、CPUは、検索結果を表示データとして設定し、リターンする。

#### 【0026】

図13は、ステップS7-4の一部である文書登録処理を詳細化したフローチャートである。ステップS13-1において、CPUは、文書の本体を文書データベースに登録する。ステップS13-2において、CPUは、文書に付加されているメタデータを文書データベースのメタデータ部に登録する。ステップS13-3において、CPUは、文書本

10

20

30

40

50

体からテキストデータを抽出する。ステップS13-4において、CPUは、前記抽出したテキストから重要語テーブルに基づいて重要語を抽出する。ステップS13-5において、CPUは、抽出した重要語をその文書のメタデータとして文書データベースの全文キーワード部に登録し、その後リターンする。

【0027】

<実施形態2>

実施形態1においては、全文キーワードは文書テキストから抽出された重要語を区別無く文書データベースに登録していたが、ある程度分類して登録することも可能である。例えば、製品名、人名、地名、電話番号、企業名、等の区分けを重要語テーブルに持っておき、別々のメタデータとして保持することも出来る。このようにすると細分化して検索指定ができるので、ユーザの操作性が一段と向上する。10

【0028】

<その他の実施形態>

また、本発明の目的は、以下のようにすることによって達成される。即ち、上述した実施形態の機能を実現するソフトウェアのプログラムコードを記録した記憶媒体（又は記録媒体）を、システム或いは装置に供給する。そして、そのシステム或いは装置の中央演算処理手段（CPUやMPU）が記憶媒体に格納されたプログラムコードを読み出し実行する。この場合、記憶媒体から読み出されたプログラムコード自体が上述した実施形態の機能を実現することになり、そのプログラムコードを記録した記憶媒体は本発明を構成することになる。20

【0029】

また、システム或いは装置の前記中央演算処理手段が読み出したプログラムコードを実行することにより、そのプログラムコードの指示に基づき、システム或いは装置上で稼働しているオペレーティングシステム（OS）等が実際の処理の一部又は全部を行う。その処理によって上述した実施形態の機能が実現される場合も含まれる。

【0030】

更に、記憶媒体から読み出されたプログラムコードが、前記システム或いは装置に挿入された機能拡張カードや、接続された機能拡張ユニットに備わるメモリに書込まれたとする。その後、そのプログラムコードの指示に基づき、その機能拡張カードや機能拡張ユニットに備わるCPU等が実際の処理の一部又は全部を行い、その処理によって上述した実施形態の機能が実現される場合も含まれる。30

【0031】

本発明を前記記憶媒体に適用する場合、その記憶媒体（コンピュータ読み取り可能な記憶媒体）には、先に説明したフローチャートに対応するプログラムコードが格納されることになる。

【0032】

以上、上述した各実施形態によれば、データ集合を一定程度の大きさに分割できる検索式のセットを自動で生成させ、ユーザはそのうちの一つを選択するだけで、検索が実行でき、データ集合を細分していくことができるようになりますので使い勝手が向上する。40

【0033】

以上、本発明の好ましい実施形態について詳述したが、本発明は係る特定の実施形態に限定されるものではなく、特許請求の範囲に記載された本発明の要旨の範囲内において、種々の変形・変更が可能である。

【図面の簡単な説明】

【0034】

【図1】検索装置の構成を示すブロック図である。

【図2】検索画面の例を示す図である。

【図3】文書を管理するための情報である文書データベースの構成の一例を示した図である。50

【図4】重要語テーブルの一例を示した図である。

【図5】検索式セット候補テーブルの構成の一例を示した図である。

【図6】検索式セットを提示する順位のもとになる提示順位スコアの算出式の一例を示した図である。

【図7】検索装置の動作、より具体的にはCPUの処理手順を示すフローチャートである。

【図8】ステップS7-4の一部である検索式セット提示処理を詳細化したフローチャートである。

【図9】ステップS8-4における検索式候補設定の処理詳細を示したフローチャートである。 10

【図10】ステップS7-4の一部である次候補処理の詳細を示したフローチャートである。

【図11】ステップS7-4の一部であるカーソル移動処理を詳細化したフローチャートである。

【図12】ステップS7-4の一部である検索実行処理を詳細化したフローチャートである。

【図13】ステップS7-4の一部である文書登録処理を詳細化したフローチャートである。

#### 【符号の説明】

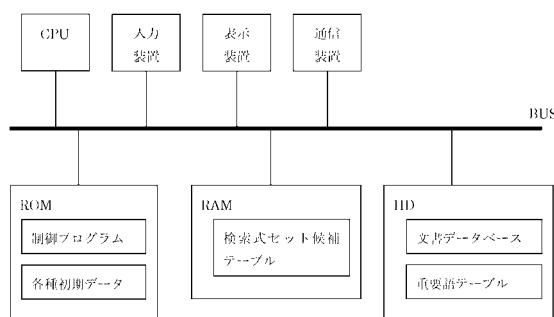
【0035】

2-1 初期画面

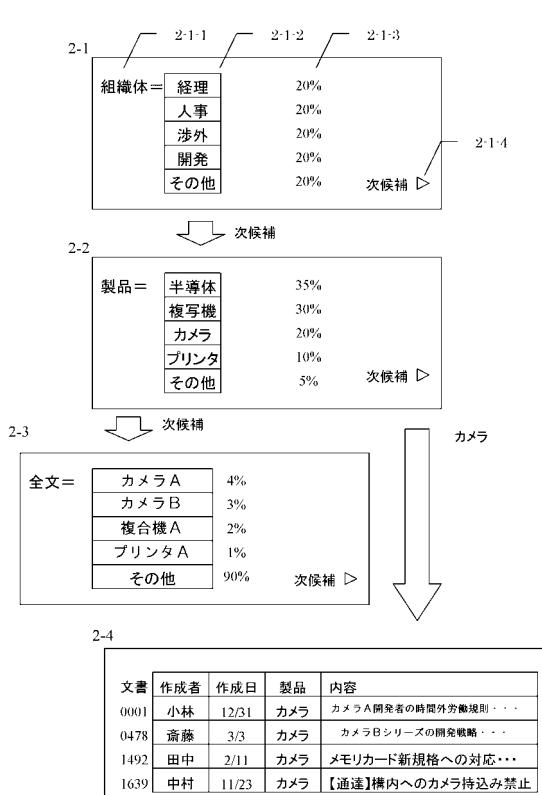
10

20

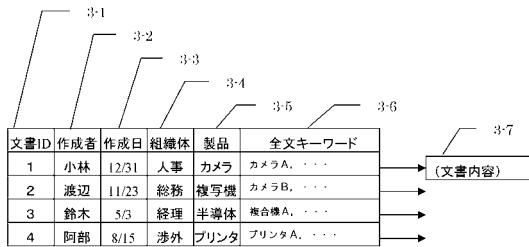
【図1】



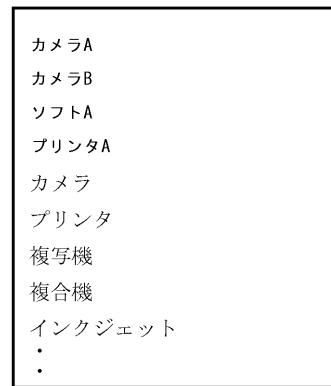
【図2】



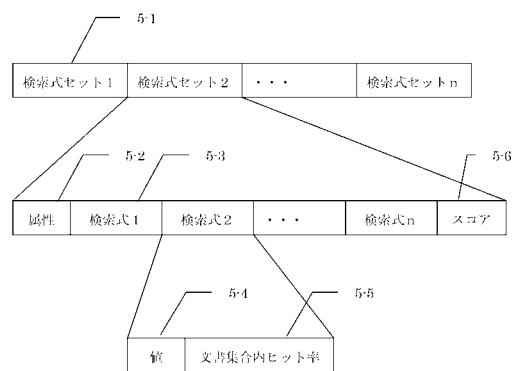
【図3】



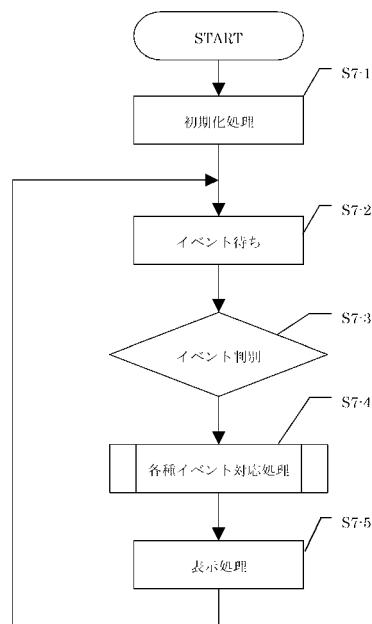
【図4】



【図5】



【図7】

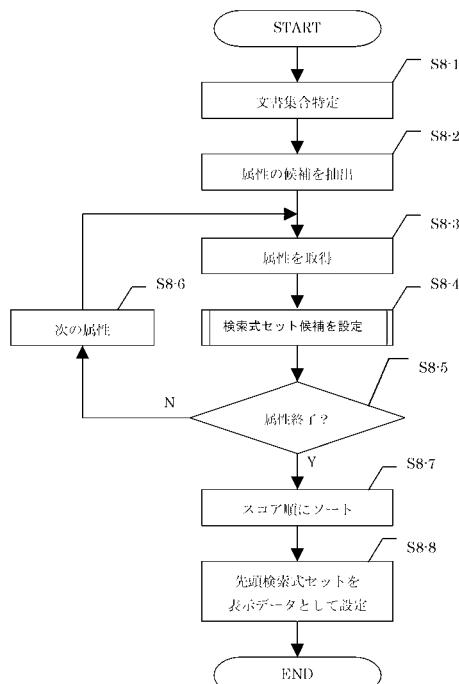


【図6】

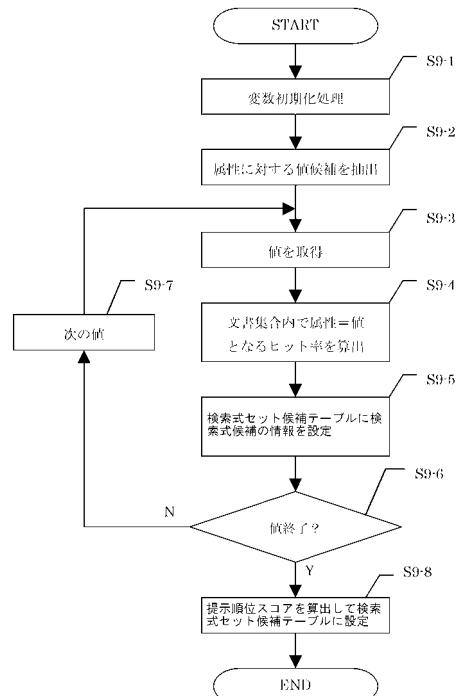
各検索式の文書集合内ヒット率を  $h_k$  とする、  
提示順スコア  $S$  は、以下に求められる。

$$S = \frac{\sum_{k=1}^n (h_k)^2}{n}$$

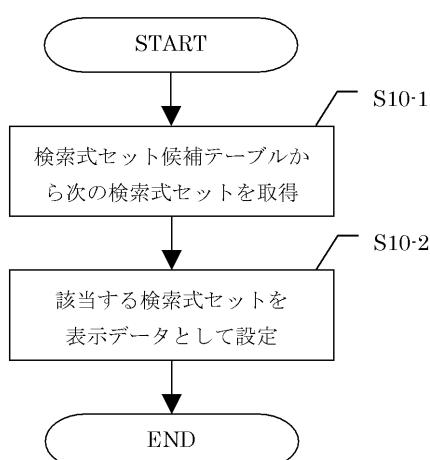
【図8】



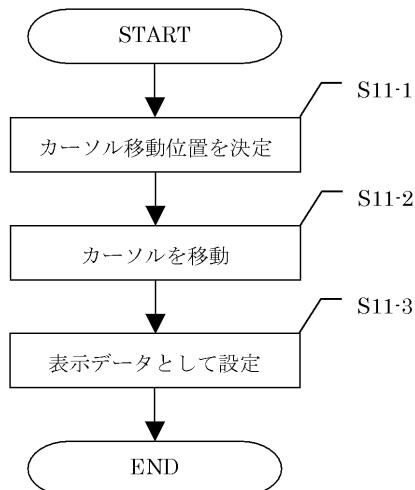
【図9】



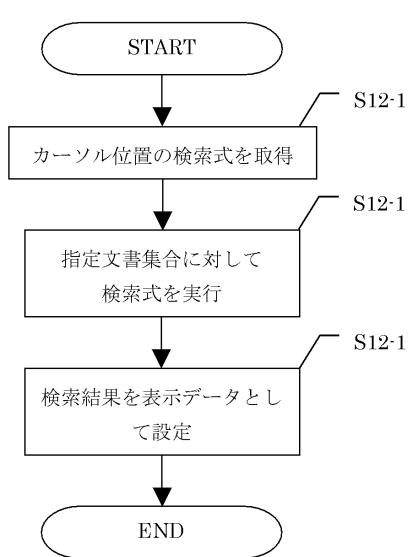
【図10】



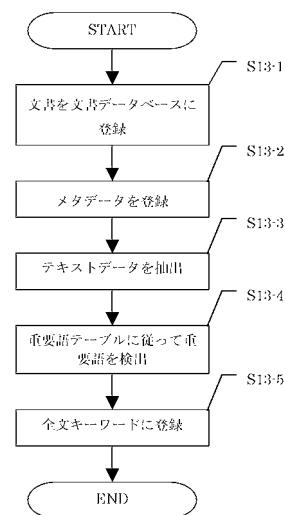
【図11】



【図12】



【図13】



---

フロントページの続き

(56)参考文献 特開平05-314182(JP,A)  
特開平11-161658(JP,A)  
特開2003-067226(JP,A)  
特開2006-350915(JP,A)

(58)調査した分野(Int.Cl., DB名)

G 06 F 17 / 30  
G 06 F 3 / 0482