

(19) 日本国特許庁(JP)

(12) 特 許 公 報(B2)

(11) 特許番号

特許第6079207号  
(P6079207)

(45) 発行日 平成29年2月15日(2017.2.15)

(24) 登録日 平成29年1月27日(2017.1.27)

(51) Int.Cl.

F I

G 0 6 F 17/30 (2006.01)

G 0 6 F 17/30 3 2 0 B

G 0 6 F 17/30 3 4 0 Z

請求項の数 4 (全 20 頁)

(21) 出願番号 特願2012-276261 (P2012-276261)  
(22) 出願日 平成24年12月18日(2012.12.18)  
(65) 公開番号 特開2014-120080 (P2014-120080A)  
(43) 公開日 平成26年6月30日(2014.6.30)  
審査請求日 平成27年8月4日(2015.8.4)

前置審査

(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 110002147  
特許業務法人酒井国際特許事務所  
(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 田中 秀樹

最終頁に続く

(54) 【発明の名称】 キーワード提示プログラム、キーワード提示方法及びキーワード提示装置

(57) 【特許請求の範囲】

【請求項1】

コンピュータに、  
キーワードの指定を受け付け、

複数のファイルもしくは複数のファイルの各々に含まれる文字列のインデックスを記憶する第1の記憶部、または、複数のファイルのメタデータもしくは複数のファイルの各々のメタデータに含まれる文字列のインデックスを記憶する第2の記憶部を参照して、前記指定のキーワードがヒットするファイルを検索し、

ファイルの検索結果にしたがって前記指定のキーワードの出現の有無をファイルもしくはファイルのメタデータごとに設定することによって出現マップを作成し、

前記指定のキーワードの出現マップと、前記指定のキーワードと組み合わせるキーワードの出現マップとを用いて、前記指定のキーワードとの組合せの検索結果が所定の件数を達成するキーワードを抽出し、

前記指定のキーワードに組み合わせるキーワード数を取得し、

前記指定のキーワードに組み合わせられるキーワードの各順序ごとに、当該順序に対応する回数にわたって前記指定のキーワードに組み合わせるキーワード数に対応する絞込み率を前記指定のキーワードによる検索結果の件数に乗算することによって当該順序における絞込み件数を算出し、

前記指定のキーワードに組み合わせられるキーワードの各順序ごとに当該順序までのキーワードがAND演算された場合の検索結果の件数が前記絞込み件数から所定の範囲内で

10

20

あるか否かを判定することによって、抽出されたキーワードが前記指定のキーワードと組み合わせられる妥当性を判定し、

妥当性があると判定されたキーワードを提示する  
各処理を実行させるキーワード提示プログラム。

【請求項 2】

前記キーワードを抽出する処理として、

前記指定のキーワードとの組合せの検索結果が第 1 の件数以下かつ前記第 1 の件数よりも小さい第 2 の件数以上であるキーワードを抽出することを特徴とする請求項 1 に記載のキーワード提示プログラム。

【請求項 3】

コンピュータが、  
キーワードの指定を受け付け、

複数のファイルもしくは複数のファイルの各々に含まれる文字列のインデックスを記憶する第 1 の記憶部、または、複数のファイルのメタデータもしくは複数のファイルの各々のメタデータに含まれる文字列のインデックスを記憶する第 2 の記憶部を参照して、前記指定のキーワードがヒットするファイルを検索し、

ファイルの検索結果にしたがって前記指定のキーワードの出現の有無をファイルもしくはファイルのメタデータごとに設定することによって出現マップを作成し、

前記指定のキーワードの出現マップと、前記指定のキーワードと組み合わせるキーワードの出現マップとを用いて、前記指定のキーワードとの組合せの検索結果が所定の件数を達成するキーワードを抽出し、

前記指定のキーワードに組み合わせるキーワード数を取得し、

前記指定のキーワードに組み合わせられるキーワードの各順序ごとに、当該順序に対応する回数にわたって前記指定のキーワードに組み合わせるキーワード数に対応する絞込み率を前記指定のキーワードによる検索結果の件数に乗算することによって当該順序における絞込み件数を算出し、

前記指定のキーワードに組み合わせられるキーワードの各順序ごとに当該順序までのキーワードが AND 演算された場合の検索結果の件数が前記絞込み件数から所定の範囲内であるか否かを判定することによって、抽出されたキーワードが前記指定のキーワードと組み合わせられる妥当性を判定し、

妥当性があると判定されたキーワードを提示する  
各処理を実行するキーワード提示方法。

【請求項 4】

複数のファイルもしくは複数のファイルの各々に含まれる文字列のインデックスを記憶する第 1 の記憶部、または、複数のファイルのメタデータもしくは複数のファイルの各々のメタデータに含まれる文字列のインデックスを記憶する第 2 の記憶部のうち少なくとも 1 つの記憶部と、

キーワードの指定を受け付ける受付部と、

前記第 1 の記憶部または前記第 2 の記憶部を参照して、前記指定のキーワードがヒットするファイルを検索する検索部と、

ファイルの検索結果にしたがって前記指定のキーワードの出現の有無をファイルもしくはファイルのメタデータごとに設定することによって出現マップを作成する作成部と、

前記指定のキーワードの出現マップと、前記指定のキーワードと組み合わせるキーワードの出現マップとを用いて、前記指定のキーワードとの組合せの検索結果が所定の件数を達成するキーワードを抽出する抽出部と、

前記指定のキーワードに組み合わせるキーワード数を取得する取得部と、

前記指定のキーワードに組み合わせられるキーワードの各順序ごとに、当該順序に対応する回数にわたって前記指定のキーワードに組み合わせるキーワード数に対応する絞込み率を前記指定のキーワードによる検索結果の件数に乗算することによって当該順序における絞込み件数を算出する算出部と、

10

20

30

40

50

前記指定のキーワードに組み合わせられるキーワードの各順序ごとに当該順序までのキーワードがAND演算された場合の検索結果の件数が前記絞込み件数から所定の範囲内であるか否かを判定することによって、抽出されたキーワードが前記指定のキーワードと組み合わせられる妥当性を判定する判定部と、

妥当性があると判定されたキーワードを提示する提示部と  
を有するキーワード提示装置。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、キーワード提示プログラム、キーワード提示方法及びキーワード提示装置に関する。 10

【背景技術】

【0002】

キーワードを手がかりに目的のデータを検索するキーワード検索が知られている。例えば、キーワード検索は、WebページまたはWebページ内の画像や音声などのファイルを検索するWeb検索やコンピュータ上に記憶されたファイルを検索するデスクトップ検索を行う場合などに利用される。

【0003】

かかるキーワード検索に用いられるキーワードが一般的である場合には、検索結果として膨大な件数がヒットする場合がある。このため、ユーザによって複数の指定がなされたキーワードを用いて目的のデータを検索するAND検索が用いられる場合がある。かかるAND検索では、入力されるキーワードと関連性の高い単語や共起関係の強い単語を候補として提示することによってキーワードの入力を補助することも行われている。 20

【先行技術文献】

【特許文献】

【0004】

【特許文献1】特開平10-260990号公報

【特許文献2】特開平5-054082号公報

【特許文献3】特開2003-108594号公報

【発明の概要】

30

【発明が解決しようとする課題】

【0005】

しかしながら、上記のキーワードの入力補助には、以下に説明するように、検索結果を精度よく絞り込むことができないという問題がある。

【0006】

例えば、上記の入力補助で関連性の高い単語が提示された場合には、入力されたキーワードの同義語や類義語を含むファイルが検索されることになる。このため、入力されたキーワード単独で検索する場合と大差のない検索結果しか得られない場合が多く、検索結果を絞り込むことができるとは限らない。また、上記の入力補助で共起関係の強い単語が提示された場合には、入力されたキーワードとともに出現する頻度が高い単語をさらに含むファイルが検索されることになる。ところが、共起関係の強い単語が提示された場合にも、入力されたキーワードと関連性の高い単語が提示されるケースが多く、入力されたキーワード単独で検索する場合と大差のない検索結果しか得られない場合が多い。 40

【0007】

1つの側面では、検索結果の絞込み精度を向上させることができるキーワード提示プログラム、キーワード提示方法及びキーワード提示装置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】

【0008】

一態様のキーワード提示プログラムは、コンピュータに、キーワードの指定を受け付ける処理を実行させる。さらに、前記コンピュータに、複数のファイルもしくは複数のファ 50

イルの各々に含まれる文字列のインデックスを記憶する第1の記憶部、または、複数のファイルのメタデータもしくは複数のファイルの各々のメタデータに含まれる文字列のインデックスを記憶する第2の記憶部を参照して、前記指定のキーワードがヒットするファイルを検索する処理を実行させる。さらに、前記コンピュータに、ファイルの検索結果にしたがって前記指定のキーワードの出現の有無をファイルもしくはファイルのメタデータごとに設定することによって出現マップを作成する処理を実行させる。さらに、前記コンピュータに、前記指定のキーワードの出現マップと、前記指定のキーワードと組み合わせるキーワードの出現マップとを用いて、前記指定のキーワードとの組合せの検索結果が所定の件数を達成するキーワードを抽出する処理を実行させる。さらに、前記コンピュータに、抽出されたキーワードが前記指定のキーワードと組み合わせられる妥当性を判定する処理を実行させる。さらに、前記コンピュータに、妥当性があると判定されたキーワードを提示する処理を実行させる。

10

【発明の効果】

【0009】

一実施形態によれば、検索結果の絞込み精度を向上させることができる。

【図面の簡単な説明】

【0010】

【図1】図1は、実施例1に係る検索システムの構成を示す図である。

【図2】図2は、実施例1に係るサーバ装置の機能的構成を示すブロック図である。

【図3】図3は、指定のキーワードの出現マップの一例を示す図である。

20

【図4】図4は、AND演算の一例を示す図である。

【図5】図5は、AND演算の一例を示す図である。

【図6】図6は、キーワードルートの一例を示す図である。

【図7】図7は、指定のキーワードによる検索結果の件数と指定のキーワードに組み合わせるキーワード数との関係の一例を示す図である。

【図8】図8は、クライアント端末に表示される画面遷移の一例を示す図である。

【図9】図9は、クライアント端末に表示される画面遷移の一例を示す図である。

【図10】図10は、実施例1に係る検索処理の手順を示すフローチャートである。

【図11】図11は、実施例1に係る提示キーワードルートの生成処理の手順を示すフローチャート(1)である。

30

【図12】図12は、実施例1に係る提示キーワードルートの生成処理の手順を示すフローチャート(2)である。

【図13】図13は、実施例1及び実施例2に係るキーワード提示プログラムを実行するコンピュータの一例について説明するための図である。

【発明を実施するための形態】

【0011】

以下に添付図面を参照して本願に係るキーワード提示プログラム、キーワード提示方法及びキーワード提示装置について説明する。なお、この実施例は開示の技術を限定するものではない。そして、各実施例は、処理内容を矛盾させない範囲で適宜組み合わせることが可能である。

40

【実施例1】

【0012】

[システム構成]

図1は、実施例1に係る検索システムの構成を示す図である。図1に示す検索システム1は、ネットワーク5を介して、WebページまたはWebページ内の画像や音声などのファイルを検索するWeb検索サービスをクライアント端末30A~30Cに提供するものである。

【0013】

図1に示すように、検索システム1には、サーバ装置10と、クライアント端末30A~30Cとが収容される。なお、図1には、3つのクライアント端末をそれぞれ図示した

50

が、検索システム 1 は任意の数のクライアント端末を収容できる。なお、以下では、クライアント端末 30 A ~ 30 C の各装置を区別なく総称する場合には、「クライアント端末 30」と記載する場合がある。

#### 【0014】

これらサーバ装置 10 及びクライアント端末 30 の間は、ネットワーク 5 を介して相互に通信可能に接続される。かかるネットワーク 5 には、有線または無線を問わず、インターネット (Internet) を始め、LAN (Local Area Network) や VPN (Virtual Private Network) などの任意の種類の通信網を採用できる。

#### 【0015】

サーバ装置 10 は、Web 検索サービスをクライアント端末 30 に提供するコンピュータである。例えば、サーバ装置 10 は、クライアント端末 30 を介して指定を受け付けたキーワードに組み合わせることによって検索結果の件数を絞り込む絞り込み用のキーワードを提示するキーワード提示サービスを上記の Web 検索サービスに併せて提供する。一態様としては、上記のキーワード提示サービスに関する処理を実行する Web サーバとして実装することとしてもよいし、また、上記のキーワード提示サービスをアウトソーシングにより提供するクラウドとして実装することもできる。他の一態様としては、パッケージソフトウェアやオンラインソフトウェアとして提供されるキーワード提示プログラムを所望のコンピュータにプリインストール又はインストールさせることによっても実装できる。

#### 【0016】

クライアント端末 30 は、上記のキーワード提示サービスの提供を受けるコンピュータである。かかるクライアント端末 30 の一態様としては、パーソナルコンピュータを始めとする固定端末の他、スマートフォン、携帯電話機、PHS (Personal Handyphone System) や PDA (Personal Digital Assistants) などの移動体端末も採用できる。

#### 【0017】

なお、以下では、サーバ装置 10 がネットワーク 5 上で公開された Web ページをクライアント端末 30 A ~ 30 C に検索させるポータルサイトとして機能する例を想定して説明を行うが、上記のキーワード提示サービスの適用範囲はサーバ装置 10 がポータルサイトとして機能する場合に限定されない。例えば、特定のサービス事業者が管理するサイト内のファイル、例えばマークアップ言語で記述された XML (Extensible Markup Language) や HTML (HyperText Markup Language) ファイル等をキーワードを用いて検索させる場合にも上記のキーワード提示サービスを同様に適用できる。また、企業や団体などの組織によってデータベース化されたファイルをアクセス権限を持つ組織の関係者を対象にキーワードを用いて検索させる場合にも上記のキーワード提示サービスを同様に適用できる。

#### 【0018】

##### [サーバ装置 10 の構成]

続いて、本実施例に係るサーバ装置 10 の機能的構成について説明する。図 2 は、実施例 1 に係るサーバ装置 10 の機能的構成を示すブロック図である。図 2 に示すように、サーバ装置 10 は、通信 I/F (InterFace) 部 11 と、記憶部 13 と、制御部 15 とを有する。なお、サーバ装置 10 は、図 2 に示した機能部以外にも既知のサーバ装置が有する各種の機能部、例えば各種の入力デバイスや音声出力デバイスなどの機能部を有することとしてもかまわない。

#### 【0019】

通信 I/F 部 11 は、他の装置、例えばクライアント端末 30 との間で通信制御を行うインタフェースである。かかる通信 I/F 部 11 の一態様としては、LAN (Local Area Network) カードなどのネットワークインタフェースカードを採用できる。例えば、通信 I/F 部 11 は、クライアント端末 30 から検索に用いるキーワードを受信したり、指定のキーワードを用いて検索された検索結果、さらには指定のキーワードに組み合わせて検索結果をさらに絞り込むための絞り込み用のキーワードをクライアント端末 30 へ送信し

10

20

30

40

50

たりする。

【 0 0 2 0 】

記憶部 1 3 は、制御部 1 5 で実行される O S (Operating System) やキーワード提示プログラムなどの各種プログラムを記憶する記憶デバイスである。記憶部 1 3 の一態様としては、フラッシュメモリなどの半導体メモリ素子、ハードディスク、光ディスクなどの記憶装置が挙げられる。なお、記憶部 1 3 は、上記の種類の記憶装置に限定されるものではなく、R A M (Random Access Memory)、R O M (Read Only Memory)であってもよい。

【 0 0 2 1 】

記憶部 1 3 は、制御部 1 5 で実行されるプログラムに用いられるデータの一例として、転置インデックス 1 3 a と、提示キーワードルート 1 3 b とを記憶する。これら転置インデックス 1 3 a 及び提示キーワードルート 1 3 b 以外にも、他の電子データ、例えば W e b 検索でアクセスされる W e b ページなどのファイルのキャッシュなども併せて記憶することもできる。

【 0 0 2 2 】

このうち、転置インデックス 1 3 a は、検索用の索引データである。かかる転置インデックス 1 3 a の一態様としては、索引とする見出し語ごとにファイル内のテキストに見出し語を含むファイル I D (Identifier) および見出し語がテキスト内で出現する位置情報に対応付けられたデータを採用できる。ここで言う「ファイル」の一例としては、H T M L ファイル、X M L ファイルや P D F (Portable Document Format) ファイルなどが挙げられる。これらのファイルには、その一部にテキストが含まれていればよく、画像や音声が含まれていてもかまわず、画像や音声を含んだファイルのテキストを対象にキーワード検索を実行することによって画像や動画像を検索することとしてもよい。また、「見出し語」の抽出には、任意の索引文字列の抽出方法、例えば形態素解析や N - G r a m 法などを用いることができる。

【 0 0 2 3 】

かかる転置インデックス 1 3 a を検索に用いることによって、クライアント端末 3 0 から指定されたキーワードを含むファイルをファイルそのものを記憶せずとも検索できる。なお、図 2 の例では、ファイルの検索に転置インデックス 1 3 a を用いる場合を例示したが、各ファイルそのものを記憶部 1 3 に記憶しておき、各ファイルに含まれるテキストから指定のキーワードを検索することとしてもかまわない。

【 0 0 2 4 】

制御部 1 5 は、各種の処理手順を規定したプログラムや制御データを格納するための内部メモリを有し、これらによって種々の処理を実行する。制御部 1 5 は、図 2 に示すように、受付部 1 5 a と、検索部 1 5 b と、作成部 1 5 c と、抽出部 1 5 d と、判定部 1 5 e と、提示部 1 5 f とを有する。

【 0 0 2 5 】

このうち、受付部 1 5 a は、キーワードの指定を受け付ける処理部である。一態様としては、受付部 1 5 a は、クライアント端末 3 0 からサーバ装置 1 0 が W e b 検索サービスを提供する U R L (Uniform Resource Locator) へのアクセスを受け付けた場合に、次のような検索画面をクライアント端末 3 0 に表示させる。かかる検索画面の一例としては、キーワードを入力可能なテキストボックスやキーワードを選択可能なプルダウンメニューなどが含まれる。このように、受付部 1 5 a は、クライアント端末 3 0 に検索画面を表示させることによってユーザが検索に用いるキーワードの指定を検索画面を介して受け付ける。

【 0 0 2 6 】

検索部 1 5 b は、転置インデックス 1 3 a を参照して、指定のキーワードを含むファイルを検索する処理部である。一態様としては、検索部 1 5 b は、転置インデックス 1 3 a に含まれる見出し語のうち、受付部 1 5 a によって指定が受け付けられたキーワードに対応する見出し語に対応付けられたファイル I D を検索する。その後、検索部 1 5 b は、検

索結果として得られたファイルIDの数、すなわち検索結果の件数が所定の目的件数の上限値、例えば1000件以下であるか否かを判定する。このとき、目的件数の上限値を超過している場合には、指定がなされたキーワードが一般的であり、目的の情報を閲覧できる程度まで検索結果が絞り込まれていない可能性が高い。このため、検索部15bは、指定のキーワードに組み合わせる絞り込み用のキーワードを提示するために、検索結果を作成部15cへ出力する。一方、目的件数の上限値以下である場合には、指定がなされたキーワードによってユーザが目的とする情報を閲覧できる程度まで検索結果が絞り込まれている可能性が高いと推定できる。この場合には、検索部15bは、絞り込み用のキーワードを生成することなく、検索結果を提示部15fへ出力する。

【0027】

10

作成部15cは、ファイルの検索結果にしたがって指定のキーワードの出現の有無をファイルごとに設定することによって出現マップを作成する処理部である。かかる出現マップのデータ構造の一例としては、ファイルIDごとに「0」または「1」のフラグが設定されたビット配列を採用できる。例えば、作成部15cは、指定のキーワードが含まれるファイルのファイルIDには「1」を付与する一方で、指定のキーワードが含まれないファイルのファイルIDには「0」を付与する。また、作成部15cは、指定のキーワードの出現マップの他、指定のキーワードに組み合わせる絞り込み用のキーワードの候補の出現マップも絞り込み用のキーワードの候補の検索結果をビット配列へ変換することによって作成する。

【0028】

20

抽出部15dは、指定のキーワードの出現マップと、指定のキーワードと組み合わせるキーワードの出現マップとを用いて、指定のキーワードとの組合せの検索結果が所定の件数を達成するキーワードを抽出する処理部である。一態様としては、抽出部15dは、指定のキーワードとの間で共起関係を持つキーワードを指定のキーワードと組み合わせる絞り込み用のキーワードの候補として用いる。このとき、抽出部15dは、指定のキーワードに絞り込み用のキーワードの候補を1つ付加する。そして、抽出部15dは、指定のキーワードの出現マップと絞り込み用のキーワードの候補の出現マップとを用いて、各々のキーワードのAND検索の検索結果が目標件数の上限値以下かつ目標件数の下限値以上となるまで、指定のキーワードに絞り込み用のキーワードの候補を付加する処理を繰り返し実行する。その後、抽出部15dは、AND検索の検索結果が目標件数の上限値以下かつ目標件数の下限値以上となった場合に、指定のキーワード、さらには、1番目からN番目までの絞り込み用のキーワードの候補とをキーワードルートとして抽出する。

30

【0029】

これを具体的に説明すると、抽出部15dは、作成部15cによって指定のキーワードの出現マップが作成された場合に、制御部15が有する内部メモリに指定のキーワードをキーワードルートの最初のキーワード0として登録する。このとき、抽出部15dは、キーワード0にキーワード0による検索結果の件数を対応付けて内部メモリに登録する。そして、抽出部15dは、指定のキーワードとの間で共起関係を持つキーワードを指定のキーワードと組み合わせる絞り込み用のキーワードの候補としてリスト化する。その上で、抽出部15dは、絞り込み用のキーワードの候補の中から新たなキーワードを1つ選択する。続いて、抽出部15dは、新たに選択されたキーワードの出現マップを作成部15cに作成させる。ここでは、新たにキーワードが選択される度に出現マップを作成部15cに作成させる場合を例示したが、処理時間を短縮するために、キーワードごとの出現マップを予め作成して記憶しておくこととしてもかまわない。

40

【0030】

その後、抽出部15dは、内部メモリに記憶されたキーワードルートの出現マップ、ここでは指定のキーワードの出現マップと、新たに選択されたキーワードの出現マップとの間でビット配列のAND演算を実行する。すなわち、抽出部15dは、互いの出現マップに含まれる同一のファイルIDに設定された値をAND演算する。例えば、同一のファイルIDに設定された値のうち少なくとも一方が「0」である場合には、当該ファイルID

50

の値に「0」が設定される。一方、同一のファイルIDに設定された値の両方が「1」である場合には、当該ファイルIDの値に「1」が設定される。その上で、抽出部15dは、AND演算によって得られた出現マップに含まれるビット配列のうち「1」に設定されたビットの数を集計する。これによって、キーワードルートに含まれる各キーワード、すなわちキーワード0及びキーワード1のAND検索の検索結果の件数が算出される。続いて、抽出部15dは、新たに選択されたキーワードをキーワードルートのキーワード1として内部メモリに追加する。このとき、抽出部15dは、キーワード1にキーワード0及びキーワード1のAND検索の検索結果の件数に対応付けて内部メモリに登録する。そして、抽出部15dは、キーワードルートに含まれる各キーワードのAND検索の検索結果の件数が目標件数の上限値以下であるか否かを判定する。

10

#### 【0031】

このとき、目標件数の上限値を超過する場合には、目的の情報を閲覧できる程度まで検索結果が絞り込まれていない可能性が高い。この場合には、抽出部15dは、絞り込み用のキーワードの候補の中から新たなキーワードをさらに選択した上でその出現マップを作成部15cに作成させ、内部メモリに記憶されたキーワードルートの出現マップとの間でビット配列のAND演算を実行する。そして、抽出部15dは、キーワードルートに含まれる各キーワード、すなわちキーワード0、キーワード1及びキーワード2のAND検索の検索結果の件数を算出する。さらに、抽出部15dは、新たに選択されたキーワードをキーワードルートのキーワード2として内部メモリに追加する。この場合には、抽出部15dは、キーワード2にキーワード0、キーワード1及びキーワード2のAND検索の検索結果の件数に対応付けて内部メモリに登録する。

20

#### 【0032】

このように、抽出部15dは、キーワードルートに含まれる各キーワードのAND検索の検索結果の件数が目標件数の上限値以下になるまで、絞り込み用のキーワードの候補の中から新たなキーワードを選択してキーワードルートに追加する処理を繰り返し実行する。

#### 【0033】

一方、目標件数の上限値以下である場合には、指定がなされたキーワードによってユーザが目的とする情報を閲覧できる程度まで検索結果が絞り込まれている可能性が高いと推定できる。この場合には、抽出部15dは、キーワードルートに含まれる各キーワードのAND検索の検索結果の件数が目標件数の下限値、例えば10件以上であるか否かを判定する。このように、目標件数の下限値以上であるか否かを判定するのは、目標件数の上限値以下であったとしても、検索結果の件数が過度に少ない場合には、ユーザにとって重要な情報を持つファイルが検索結果から漏れてしまっている可能性が高まるからである。

30

#### 【0034】

そして、抽出部15dは、目標件数の下限値未満である場合に、内部メモリに記憶されたキーワードルートを消去する。一方、抽出部15dは、目標件数の下限値以上である場合に、後述のキーワードルートの妥当性を判定する対象として、内部メモリに記憶されたキーワードルートをそのまま残す。その後、抽出部15dは、指定のキーワードに絞り込み用のキーワードの候補を組み合わせる探索が全ての組合せについて終了するまで、すなわち既に探索されたキーワードルートおよび破棄されたキーワードルートを構成する各キーワードとの間で重複しない組合せがなくなるまで、指定のキーワードに絞り込み用のキーワードの候補を組み合わせる探索を繰り返し実行する。このとき、指定のキーワードがキーワードルートの最初のキーワード0であることは明らかである。このため、抽出部15dは、絞り込み用のキーワードの候補の中からキーワードルートのキーワード1に対応するキーワードを選択する段階から繰り返し処理を開始する。

40

#### 【0035】

図3は、指定のキーワードの出現マップの一例を示す図である。図4及び図5は、AND演算の一例を示す図である。図3には、キーワードとして「protein」が指定された場合の出現マップが図示されている。また、図4には、キーワード0「protein」の出現マップとキーワード1「metabolic」の出現マップとの間で実行されるビット配列のAND

50



演算が図示されている。さらに、図 5 には、キーワード 0「protein」及びキーワード 1「metabolic」の出現マップとキーワード 2「ec」の出現マップとの間で実行されるビット配列の AND 演算が図示されている。

#### 【 0 0 3 6 】

図 3 に示すキーワード「protein」の出現マップは、ファイル ID「1」、「3」、「6」及び「7」の各ファイルにキーワード「protein」が含まれており、その他のファイル ID のファイルにはキーワード「protein」が含まれていないことを意味する。かかるキーワード「protein」は、クライアント端末 30 から指定を受け付けたキーワードであるので、キーワードルートのキーワード 0 として内部メモリに登録される。そして、絞り込み用のキーワードの候補「metabolic」がキーワード 1 として選択された場合には、キーワード 0「protein」の出現マップとキーワード 1「metabolic」の出現マップとの間でビット配列の AND 演算が実行される。図 4 に示すように、互いの出現マップの間で同一のファイル ID の値にビット「1」が設定されたファイル ID「1」、「3」及び「6」の値が AND 演算によって「1」と算出される一方で、その他のファイル ID の値は「0」と算出される。ここで、指定のキーワードである「protein」だけの場合には、ファイル ID「7」のビットは「1」であったが、キーワード「metabolic」との AND 演算によってファイル ID「7」のビットが「0」になっている。これは、キーワード「metabolic」の追加によってファイル ID「7」のファイルが除外されて絞り込みが進んだことを意味する。

#### 【 0 0 3 7 】

さらに、絞り込み用のキーワードの候補「ec」がキーワード 2 として選択された場合には、キーワード 0「protein」及びキーワード 1「metabolic」の出現マップとキーワード 2「ec」の出現マップとの間でビット配列の AND 演算が実行される。図 5 に示すように、互いの出現マップの間で同一のファイル ID の値にビット「1」が設定されたファイル ID「3」及び「6」の値が AND 演算によって「1」と算出される一方で、その他のファイル ID の値は「0」と算出される。ここで、キーワード 1「protein」及びキーワード 2「metabolic」の AND 演算の場合には、ファイル ID「1」のビットは「1」であったが、キーワード「ec」との AND 演算によってファイル ID「1」のビットが「0」になっている。これは、キーワード「ec」の追加によってファイル ID「1」のファイルが除外されてさらに絞り込みが進んだことを意味する。

#### 【 0 0 3 8 】

図 6 は、キーワードルートの一例を示す図である。図 6 に示す 1 番目のレコードでは、指定のキーワードであるキーワード 0「protein」によって検索結果の件数が 5 0 0 0 0 件まで絞り込まれたことを意味する。さらに、キーワード 1「metabolic」によって検索結果の件数が 1 5 0 0 0 件まで絞り込まれたことを意味する。さらに、キーワード 2「ec」によって検索結果の件数が目標件数の上限値「1 0 0 0 件」以下かつ目標件数の下限値「1 0 件」以上である 9 0 0 件にまで絞り込まれていることを意味する。また、図 6 に示す 2 番目のレコードの例では、指定のキーワードであるキーワード 0「protein」によって検索結果の件数が 5 0 0 0 0 件まで絞り込まれたことを示す。さらに、キーワード 1「substrate」によって検索結果の件数が 2 7 3 0 0 件まで絞り込まれたことを示す。さらに、キーワード 2「abc」によって検索結果の件数が 1 2 5 0 0 件まで絞り込まれたことを示す。さらに、キーワード 3「transport」によって検索結果の件数が目標件数の上限値「1 0 0 0 件」以下かつ目標件数の下限値「1 0 件」以上である 1 0 0 0 件にまで絞り込まれていることを示す。

#### 【 0 0 3 9 】

なお、ここでは、目標件数の下限値未満である場合に、内部メモリに記憶されたキーワードルートを破棄することとしたが、ユーザが目的とする情報をピンポイントで閲覧できるレクエリが入力されるケースも考えられるので、検索結果の件数が目標件数の上限値以上であるか否かしか判定を実行せず、検索結果の件数が目標件数の下限値未満であるキーワードルートを残すこととしてもかまわない。

## 【 0 0 4 0 】

図2の説明に戻り、判定部15eは、抽出部15dによって抽出されたキーワードが指定のキーワードと組み合わせられる妥当性を判定する処理部である。一態様としては、判定部15eは、指定のキーワードに組み合わせるキーワード数に合わせてキーワードが追加される度に検索結果が絞り込まれる絞り込み率を設定する。例えば、指定のキーワードによる検索結果の件数を「Z件」、目標件数を「Q件」、キーワードを1つ追加する度に検索結果が絞り込まれる絞り込み率を「 $\frac{Q}{Z}$ 」とする。なお、絞り込み率「 $\frac{Q}{Z}$ 」は、一定であるものとする。このとき、キーワード0に組み合わせるキーワードがキーワード1のみである場合には、キーワード1によって絞り込まれる絞り込み件数は「 $Z \times \frac{Q}{Z}$ 」となる。また、キーワード0に組み合わせるキーワードがキーワード1及びキーワード2である場合には、キーワード1及びキーワード2によって絞り込まれる絞り込み件数は「 $Z \times \frac{Q}{Z} \times \frac{Q}{Z}$ 」となる。そして、キーワード0に組み合わせるキーワードがキーワード1からキーワードNまでのN個である場合には、キーワード1～キーワードNによって絞り込まれる絞り込み件数は「 $Z \times \left(\frac{Q}{Z}\right)^N$ 」となる。このことから、N個のAND検索でQ件まで絞り込むには、下記の式(1)が成立する。

## 【 0 0 4 1 】

$$Q = Z \times \left(\frac{Q}{Z}\right)^N \cdots \text{式(1)}$$

## 【 0 0 4 2 】

ここで、指定のキーワードに組み合わせる絞り込み用のキーワードの数ごとに指定のキーワードによる検索結果の件数「Z件」及び目標件数「Q件」の理想のサンプル値を設定する。かかる理想のサンプル値は、実験値を設定することもできるし、クライアント端末30にて任意の値を設定することもできる。その上で、判定部15eは、上記の式(1)の導出式にNとともに理想のサンプル値のZ及びQを代入する。これによって、指定のキーワードに組み合わせるキーワード数に合わせてキーワードが追加される度に検索結果が絞り込まれる絞り込み率を導出しておくことができる。図7は、指定のキーワードによる検索結果の件数と指定のキーワードに組み合わせるキーワード数との関係の一例を示す図である。図7には、指定のキーワードによる検索結果の件数Zを「10000件」、目標件数Qを「1000件」としたときの検索結果の件数が指定のキーワードに組み合わせるキーワード数ごとに図示されている。図7に示すグラフの縦軸は、検索結果の件数を指し、横軸は、指定のキーワードに組み合わせるキーワード数を指す。図7に示すように、指定のキーワードに組み合わせるキーワード数が1である場合には、絞り込み率は0.01程度が好ましいことを示している。さらに、キーワード数が2である場合には、絞り込み率は0.1程度が好ましいことを示している。さらに、キーワード数が3である場合には、絞り込み率は0.215程度が好ましいことを示している。さらに、キーワード数が4である場合には、絞り込み率は0.316程度が好ましいことを示している。さらに、キーワード数が5である場合には、絞り込み率は0.398程度が好ましいことを示している。さらに、キーワード数が6である場合には、絞り込み率は0.464程度が好ましいことを示している。さらに、キーワード数が7である場合には、絞り込み率は0.517程度が好ましいことを示している。

## 【 0 0 4 3 】

このような絞り込み率が導出されている構成の下、判定部15eは、キーワード1からキーワードNまでの各順序ごとに当該順序までのキーワードがAND演算された場合の検索結果の件数と比較する基準となる絞り込み件数 $W_n$ を算出する。例えば、判定部15eは、指定のキーワードに組み合わせるキーワード数を取得する。そして、判定部15eは、キーワードルートに含まれるキーワード1からキーワードNまでの各順序ごとに、当該順序に対応する回数、すなわち順序と同一の回数にわたって指定のキーワードに組み合わせるキーワード数に対応する絞り込み率を指定のキーワードによる検索結果の件数に乗算する。これによって、当該順序における絞り込み件数 $W_n$ を算出する。その上で、判定部15eは、キーワードルートに含まれるキーワード1からキーワードNまでの各順序ごとに当該順序までのキーワードがAND演算された場合の検索結果の件数が絞り込み件数 $W_n$ から

所定の範囲、例えば  $\pm 20\%$  の範囲内であるか否かを判定する。このとき、検索結果の件数が絞込み件数  $W_n \pm$  内にない場合には、絞込み件数  $W_n \pm$  内にないキーワード  $N$  が一般的な単語で絞込みへ有効に機能していないのか、あるいはキーワード  $N$  によって取りこぼしが発生するおそれがあることがわかる。この場合には、キーワード  $N$  を含むキーワードルートに妥当性がないと推定する。一方、検索結果の件数が絞込み件数  $W_n \pm$  を逸脱するキーワードがない場合には、キーワード 1 ~ キーワード  $N$  を用いても取りこぼしが抑制でき、各キーワード 1 ~  $N$  が絞込みへ有効に機能する可能性が高いと推定できる。この場合には、キーワード  $N$  を含むキーワードルートに妥当性があると推定できる。判定部 15 e は、キーワード 0 及びキーワード 1 ~ キーワード  $N$  をクライアント端末 30 へ提示する対象とする提示キーワードルート 13 b として記憶部 13 へ登録する。

10

#### 【0044】

提示部 15 f は、クライアント端末 30 に対する情報の提示制御を行う処理部である。一態様としては、提示部 15 f は、指定のキーワードによる検索結果の件数が目的件数の上限値以下である場合には、指定のキーワードを含むファイルの一覧を検索結果としてクライアント端末 30 へ出力する。このとき、提示部 15 f は、検索結果として提示するファイルの一覧にタイトル、URL 等のリンクや広告を含めて出力することもできる。他の一態様としては、提示部 15 f は、指定のキーワードを含むファイルの一覧を検索結果として出力するとともに、判定部 15 e によって記憶部 13 に登録された提示キーワードルート 13 b に含まれるキーワードのうち指定を受け付けたキーワードの次のキーワードをクライアント端末 30 へ出力する。その後、提示部 15 f は、次のキーワードの選択を受け付けると、キーワードルートにさらに続きのキーワードが存在するか否かを判定する。そして、提示部 15 f は、続きのキーワードがなくなるまで、先に選択された次のキーワードが AND 検索された検索結果を出力するとともに、さらに次のキーワードをクライアント端末 30 へ出力する。その後、提示部 15 f は、続きのキーワードがなくなると、最後のキーワードが AND 検索された検索結果をクライアント端末 30 へ出力する。

20

#### 【0045】

図 8 及び図 9 は、クライアント端末 30 に表示される画面遷移の一例を示す図である。このうち、図 8 には、上記の従来技術によるキーワードの入力補助がなされた場合の画面遷移が図示されている。また、図 9 には、実施例 1 に係るサーバ装置 10 によって提示されるキーワードルートが図示されている。図 8 及び図 9 に示すように、クライアント端末 30 から指定を受け付けたキーワード「protein」を受け付けた段階では、従来技術および本実施例のいずれにおいても検索結果の件数が 50000 件という膨大な数となっている。

30

#### 【0046】

このとき、図 8 の例では、キーワード「protein」に続いて入力補助されたキーワード「kinase」が選択されている。このように、キーワード「kinase」でさらに AND 検索を行った場合でも、50000 件から 40000 件となったが、10000 件しか絞り込まれていない。さらに、図 8 の例では、キーワード「kinase」に続いて入力補助されたキーワード「cancer」が選択されている。この結果、キーワード「cancer」でさらに AND 検索を行うことによって検索結果の件数が 40000 件から 30000 件となったが、結局は 10000 件しか絞り込まれていない。このように、従来技術では、「protein」、「kinase」及び「cancer」の 3 つのキーワードで AND 検索を行っても、依然として、30000 件という膨大な検索結果が表示され、所定の絞込み精度が得られない場合がある。

40

#### 【0047】

一方、図 9 の例では、キーワード「protein」に続いて提示された次のキーワード「metabolic」が選択されている。このように、キーワード「metabolic」でさらに AND 検索を行った結果、従来技術では 40000 件までしか絞り込めていなかったが、本実施例では 50000 件から 15000 件まで絞り込むことができています。さらに、図 9 の例では、キーワード「metabolic」に続いて提示されたキーワード「ec」が選択されている。このように、キーワード「ec」でさらに AND 検索を行った結果、従来技術では 30000

50

件までしか絞り込めていなかったが、本実施例では900件まで絞り込むことができている。このように、本実施例では、「protein」、「metabolic」及び「ec」の3つのキーワードでAND検索を行うことによって、ユーザが目的とする情報を閲覧できる程度の件数まで検索結果が絞り込まれている。

#### 【0048】

なお、制御部15には、各種の集積回路や電子回路を採用できる。また、制御部15が有する機能部の一部を別の集積回路や電子回路とすることもできる。例えば、集積回路としては、ASIC (Application Specific Integrated Circuit) が挙げられる。また、電子回路としては、CPU (Central Processing Unit) やMPU (Micro Processing Unit) などが挙げられる。

10

#### 【0049】

##### [ 処理の流れ ]

続いて、本実施例に係るサーバ装置10の処理の流れについて説明する。なお、ここでは、サーバ装置10によって実行される(1)検索処理を説明した後に検索処理のサブルーチンとして実行される(2)提示キーワードルートの生成処理を説明することとする。

#### 【0050】

##### (1) 検索処理

図10は、実施例1に係る検索処理の手順を示すフローチャートである。この処理は、クライアント端末30からキーワードの指定を受け付けた場合に処理が起動される。図10に示すように、受付部15aによってクライアント端末30からキーワードの指定が受け付けられると(ステップS101)、検索部15bは、転置インデックス13aに含まれる見出し語のうち、指定のキーワードに対応する見出し語に対応付けられたファイルIDを検索する(ステップS102)。

20

#### 【0051】

その後、検索部15bは、検索結果として得られたファイルIDの数、すなわち検索結果の件数が目的件数の上限値以下であるか否かを判定する(ステップS103)。このとき、目的件数の上限値を超過している場合(ステップS103No)には、指定がなされたキーワードが一般的であり、目的の情報を閲覧できる程度まで検索結果が絞り込まれていない可能性が高い。この場合には、指定のキーワードに組み合わせるキーワードとしてクライアント端末30へ提示する提示キーワードルートの生成処理が実行される(ステップS104)。

30

#### 【0052】

そして、提示部15fは、指定のキーワードを含むファイルの一覧を検索結果として出力するとともに、提示キーワードルート13bに含まれるキーワードのうち指定を受け付けたキーワードの次のキーワードをクライアント端末30へ出力する(ステップS105)。

#### 【0053】

その後、提示部15fは、次のキーワードの選択を受け付けると(ステップS106)、キーワードルートにさらに続きのキーワードが存在するか否かを判定する(ステップS107)。

40

#### 【0054】

そして、続きのキーワードが存在する場合(ステップS107No)には、提示部15fは、ステップS106で選択された次のキーワードがさらにAND検索された検索結果を出力するとともに、さらに続く次のキーワードをクライアント端末30へ出力し(ステップS105)、ステップS105で続きのキーワードとして出力された次のキーワードの選択をさらに受け付ける(ステップS106)。

#### 【0055】

一方、続きのキーワードが存在しない場合(ステップS107Yes)、提示部15fは、最後のキーワードがAND検索された検索結果をクライアント端末30へ出力し(ステップS108)、処理を終了する。

50

## 【 0 0 5 6 】

また、目的件数の上限値以下である場合（ステップ S 1 0 3 Y e s ）には、指定がなされたキーワードによってユーザが目的とする情報を閲覧できる程度まで検索結果が絞り込まれている可能性が高いと推定できる。この場合には、検索部 1 5 b は、絞り込み用のキーワードを生成することなく、検索結果を提示部 1 5 f へ出力し（ステップ S 1 0 8 ）、処理を終了する。

## 【 0 0 5 7 】

## （ 2 ）提示キーワードルートの生成処理

図 1 1 及び図 1 2 は、実施例 1 に係る提示キーワードルートの生成処理の手順を示すフローチャートである。この処理は、図 1 0 に示したステップ S 1 0 3 の判定で N o （否定）と判定された場合に処理が起動する。

## 【 0 0 5 8 】

図 1 1 に示すように、作成部 1 5 c は、内部メモリに記憶されたキーワードルートを空にする初期化を実行する（ステップ S 2 0 1 ）。続いて、作成部 1 5 c は、指定のキーワードによる検索結果にしたがって指定のキーワードの出現の有無をファイルごとに設定することによって出現マップを作成する（ステップ S 2 0 2 ）。

## 【 0 0 5 9 】

そして、抽出部 1 5 d は、指定のキーワードをキーワードルートの最初のキーワード 0 として内部メモリに登録する（ステップ S 2 0 3 ）。このとき、抽出部 1 5 d は、キーワード 0 にキーワード 0 による検索結果の件数を対応付けて内部メモリに登録する。

## 【 0 0 6 0 】

そして、抽出部 1 5 d は、指定のキーワードとの間で共起関係を持つキーワードを指定のキーワードと組み合わせる絞り込み用のキーワードの候補としてリスト化する（ステップ S 2 0 4 ）。その上で、抽出部 1 5 d は、絞り込み用のキーワードの候補の中から新たなキーワードを 1 つ選択する（ステップ S 2 0 5 ）。続いて、抽出部 1 5 d は、ステップ S 2 0 5 で選択されたキーワードの出現マップを作成部 1 5 c に作成させる（ステップ S 2 0 6 ）。

## 【 0 0 6 1 】

その後、抽出部 1 5 d は、内部メモリに記憶されたキーワードルートの出現マップと、ステップ S 2 0 5 で選択されたキーワードの出現マップとの間でビット配列の A N D 演算を実行する（ステップ S 2 0 7 ）。その上で、抽出部 1 5 d は、キーワードルートに含まれる各キーワードの A N D 検索の検索結果の件数を算出する（ステップ S 2 0 8 ）。

## 【 0 0 6 2 】

続いて、抽出部 1 5 d は、ステップ S 2 0 5 で選択されたキーワードをキーワードルートのキーワード 1 として内部メモリに追加する（ステップ S 2 0 9 ）。このとき、抽出部 1 5 d は、キーワード 1 にキーワード 0 及びキーワード 1 の A N D 検索の検索結果の件数を対応付けて内部メモリに登録する。

## 【 0 0 6 3 】

そして、抽出部 1 5 d は、キーワードルートに含まれる各キーワードの A N D 検索の検索結果の件数が目標件数の上限値以下であるか否かを判定する（ステップ S 2 1 0 ）。このとき、目標件数の上限値を超過する場合（ステップ S 2 1 0 N o ）には、目的の情報を閲覧できる程度まで検索結果が絞り込まれていない可能性が高い。この場合には、抽出部 1 5 d は、絞り込み用のキーワードの候補の中から新たなキーワードをさらに選択し（ステップ S 2 0 5 ）、ステップ S 2 0 6 ～ステップ S 2 0 9 までの処理を繰り返し実行する。

## 【 0 0 6 4 】

一方、目標件数の上限値以下である場合（ステップ S 2 1 0 Y e s ）には、指定がなされたキーワードによってユーザが目的とする情報を閲覧できる程度まで検索結果が絞り込まれている可能性が高いと推定できる。この場合には、抽出部 1 5 d は、キーワードルートに含まれる各キーワードの A N D 検索の検索結果の件数が目標件数の下限値以上であるか否かを判定する（ステップ S 2 1 1 ）。

## 【 0 0 6 5 】

そして、目標件数の下限値未満である場合（ステップ S 2 1 1 N o ）には、抽出部 1 5 d は、内部メモリに記憶されたキーワードルートを消去する（ステップ S 2 1 2 ）。一方、目標件数の下限値以上である場合（ステップ S 2 1 1 Y e s ）には、抽出部 1 5 d は、キーワードルートの妥当性を判定する対象として、内部メモリに記憶されたキーワードルートをそのまま残し、ステップ S 2 1 3 へ移行する。

## 【 0 0 6 6 】

その後、抽出部 1 5 d は、指定のキーワードに絞込み用のキーワードの候補を組み合わせる探索が全ての組合せについて終了するまで（ステップ S 2 1 3 N o ）、上記のステップ S 2 0 5 ~ ステップ S 2 1 2 までの処理を繰り返し実行する。

10

## 【 0 0 6 7 】

そして、指定のキーワードに絞込み用のキーワードの候補を組み合わせる探索が全ての組合せについて終了すると（ステップ S 2 1 3 Y e s ）、判定部 1 5 e は、内部メモリに登録されたキーワードルートを 1 つ選択する（ステップ S 2 1 4 ）。

## 【 0 0 6 8 】

続いて、判定部 1 5 e は、指定のキーワードに組み合わせるキーワード数を取得する（ステップ S 2 1 5 ）。そして、判定部 1 5 e は、キーワード 1 からキーワード N までの各順序ごとに、当該順序に対応する回数にわたって指定のキーワードに組み合わせるキーワード数に対応する絞込み率 を指定のキーワードによる検索結果の件数に乗算することによって各順序における絞込み件数  $W_n$  を算出する（ステップ S 2 1 6 ）。

20

## 【 0 0 6 9 】

その後、判定部 1 5 e は、内部メモリが持つキーワードカウンタ  $n$  を初期値「 1 」に設定する初期化を実行する（ステップ S 2 1 7 ）。その上で、判定部 1 5 e は、キーワードカウンタ  $n$  に対応する順序までのキーワードが A N D 演算された場合の検索結果の件数が絞込み件数  $W_n$  から所定の範囲、例えば  $\pm$  の範囲内であるか否かを判定する（ステップ S 2 1 8 ）。

## 【 0 0 7 0 】

このとき、検索結果の件数が絞込み件数  $W_n \pm$  内である場合（ステップ S 2 1 8 Y e s ）には、キーワードカウンタ  $n$  の値が N になったか否かを判定する（ステップ S 2 1 9 ）。そして、キーワードカウンタ  $n$  の値が N 未満である場合（ステップ S 2 1 9 N o ）には、キーワードカウンタ  $n$  をインクリメントし（ステップ S 2 2 0 ）、上記のステップ S 2 1 8 の処理を再び実行する。なお、検索結果の件数が絞込み件数  $W_n \pm$  内にない場合（ステップ S 2 1 8 N o ）には、ステップ S 2 1 9 ~ ステップ S 2 2 1 までの処理をとばし、ステップ S 2 2 2 の処理へ移行する。

30

## 【 0 0 7 1 】

一方、キーワードカウンタ  $n$  の値が N になった場合（ステップ S 2 1 9 Y e s ）には、判定部 1 5 e は、ステップ S 2 1 4 で選択されたキーワードルートをクライアント端末 3 0 へ提示する対象とする提示キーワードルート 1 3 b として記憶部 1 3 へ追加する（ステップ S 2 2 1 ）。

## 【 0 0 7 2 】

40

その後、内部メモリに登録された全てのキーワードルートについて妥当性の判定を終了するまで（ステップ S 2 2 2 N o ）、上記のステップ S 2 1 4 ~ ステップ S 2 2 1 までの処理を繰り返し実行する。そして、内部メモリに登録された全てのキーワードルートについて妥当性の判定を終了すると（ステップ S 2 2 2 Y e s ）、処理を終了する。

## 【 0 0 7 3 】

## 〔 実施例 1 の効果 〕

上述してきたように、本実施例に係るサーバ装置 1 0 は、入力されるキーワードとの間で組み合わせるキーワードを提示する場合に、検索結果の件数が目標の件数を満たし、かつ組合せに妥当性があるキーワードを提示する。したがって、本実施例に係るサーバ装置 1 0 によれば、検索結果の絞込み精度を向上させることができる。

50

## 【実施例 2】

## 【0074】

さて、これまで開示の装置に関する実施例について説明したが、本発明は上述した実施例以外にも、種々の異なる形態にて実施されてよいものである。そこで、以下では、本発明に含まれる他の実施例を説明する。

## 【0075】

## [ デスクトップ検索 ]

上記の実施例 1 では、Web 検索サービスを提供する場合を例示したが、例えば、コンピュータ上に記憶されたファイルを検索するデスクトップ検索を行う場合にも上記のキーワード提示サービスを同様に適用することができる。このように、デスクトップ検索を行う場合には、例えば、クライアントサーバシステムでなく、制御部 15 内の各機能部を有するコンピュータをスタンドアローンで動作させることができる。この場合には、例えば、ファイルの代わりにファイルのメタデータを記憶部 13 に記憶しておき、各ファイルのメタデータに含まれるテキストから指定のキーワードを検索することとしてもかまわない。また、複数のファイルの各々のメタデータに含まれる単語のインデックスをメタデータの転置インデックスとして記憶部 13 に記憶しておき、メタデータの転置インデックスを用いて、指定のキーワードを含むファイルを検索することとしてもかまわない。これによって、例えば、キーワードとしてファイル名を構成する文字列の一部が指定された場合に、目的の文書ファイル、画像ファイル、あるいは音声ファイルなどをさらに絞り込む文字列を提示できる。

## 【0076】

## [ 分散および統合 ]

また、図示した各装置の各構成要素は、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、各装置の分散・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、受付部 15 a、検索部 15 b、作成部 15 c、抽出部 15 d、判定部 15 e または提示部 15 f をサーバ装置 10 の外部装置としてネットワーク経由で接続するようにしてもよい。また、受付部 15 a、検索部 15 b、作成部 15 c、抽出部 15 d、判定部 15 e または提示部 15 f を別の装置がそれぞれ有し、ネットワーク接続されて協働することで、上記のサーバ装置 10 の機能を実現するようにしてもよい。

## 【0077】

## [ キーワード提示プログラム ]

また、上記の実施例で説明した各種の処理は、予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などのコンピュータで実行することによって実現することができる。そこで、以下では、図 13 を用いて、上記の実施例と同様の機能を有するキーワード提示プログラムを実行するコンピュータの一例について説明する。

## 【0078】

図 13 は、実施例 1 及び実施例 2 に係るキーワード提示プログラムを実行するコンピュータの一例について説明するための図である。図 13 に示すように、コンピュータ 100 は、操作部 110 a と、スピーカ 110 b と、カメラ 110 c と、ディスプレイ 120 と、通信部 130 とを有する。さらに、このコンピュータ 100 は、CPU 150 と、ROM 160 と、HDD 170 と、RAM 180 とを有する。これら 110 ~ 180 の各部はバス 140 を介して接続される。

## 【0079】

HDD 170 には、図 13 に示すように、上記の実施例 1 で示した受付部 15 a、検索部 15 b、作成部 15 c、抽出部 15 d、判定部 15 e 及び提示部 15 f と同様の機能を発揮するキーワード提示プログラム 170 a が予め記憶される。このキーワード提示プログラム 170 a については、図 2 に示した各々の受付部 15 a、検索部 15 b、作成部 15 c、抽出部 15 d、判定部 15 e 及び提示部 15 f の各構成要素と同様、適宜統合又は

分離しても良い。すなわち、HDD 170に格納される各データは、常に全てのデータがHDD 170に格納される必要はなく、処理に必要なデータのみがHDD 170に格納されれば良い。

#### 【0080】

そして、CPU 150が、キーワード提示プログラム170aをHDD 170から読み出してRAM 180に展開する。これによって、図13に示すように、キーワード提示プログラム170aは、キーワード提示プロセス180aとして機能する。このキーワード提示プロセス180aは、HDD 170から読み出した各種データを適宜RAM 180上の自身に割り当てられた領域に展開し、この展開した各種データに基づいて各種処理を実行する。なお、キーワード提示プロセス180aは、図2に示した受付部15a、検索部15b、作成部15c、抽出部15d、判定部15e及び提示部15fにて実行される処理、例えば図10～図12に示す処理を含む。また、CPU 150上で仮想的に実現される各処理部は、常に全ての処理部がCPU 150上で動作する必要はなく、処理に必要な処理部のみが仮想的に実現されれば良い。

10

#### 【0081】

なお、上記のキーワード提示プログラム170aについては、必ずしも最初からHDD 170やROM 160に記憶させておく必要はない。例えば、コンピュータ100に挿入されるフレキシブルディスク、いわゆるFD、CD-ROM、DVDディスク、光磁気ディスク、ICカードなどの「可搬用の物理媒体」に各プログラムを記憶させる。そして、コンピュータ100がこれらの可搬用の物理媒体から各プログラムを取得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、WANなどを介してコンピュータ100に接続される他のコンピュータまたはサーバ装置などに各プログラムを記憶させておき、コンピュータ100がこれらから各プログラムを取得して実行するようにしてもよい。

20

#### 【符号の説明】

#### 【0082】

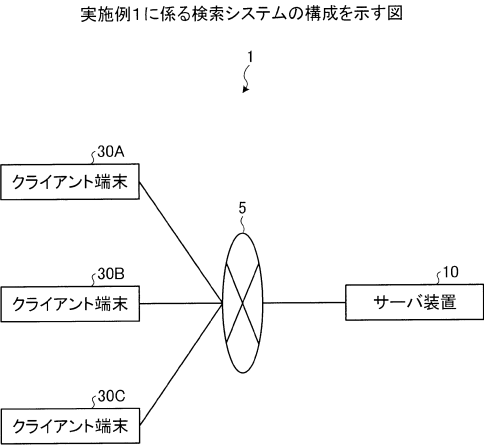
- 1 検索システム
- 5 ネットワーク
- 10 サーバ装置
- 11 通信I/F部
- 13 記憶部
- 13a 転置インデックス
- 13b 提示キーワードルート
- 15 制御部
- 15a 受付部
- 15b 検索部
- 15c 作成部
- 15d 抽出部
- 15e 判定部
- 15f 提示部
- 30A, 30B, 30C クライアント端末

30

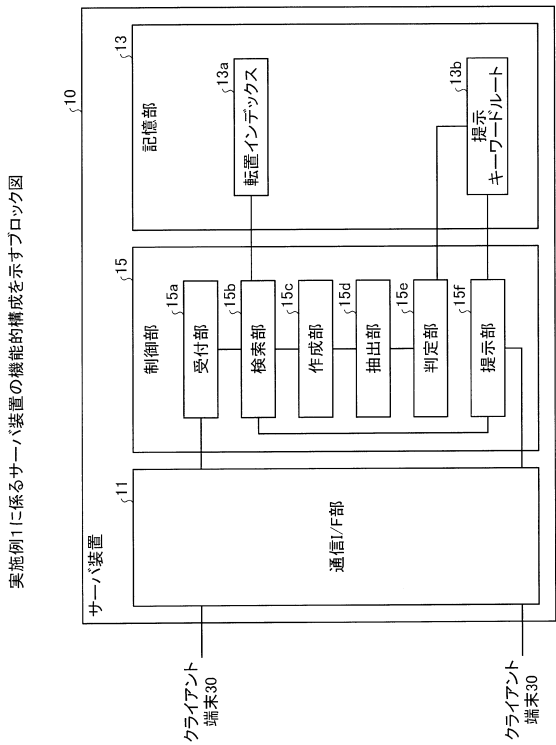
40



【 図 1 】



【 図 2 】



【 図 3 】

指定のキーワードの出現マップの一例を示す図

キーワード	1	2	3	4	5	6	7	...	N
protein	1	0	1	0	0	1	1	...	0

【 図 5 】

AND 演算の一例を示す図

キーワード	1	2	3	4	5	6	7	...	N
Protein metabolic	1	0	1	0	0	1	0	...	0

AND 演算

キーワード	1	2	3	4	5	6	7	...	N
ec	0	1	1	0	0	1	0	...	1

↓

キーワード	1	2	3	4	5	6	7	...	N
Protein Metabolic ec	0	0	1	0	0	1	0	...	0

【 図 4 】

AND 演算の一例を示す図

キーワード	1	2	3	4	5	6	7	...	N
protein	1	0	1	0	0	1	1	...	0

AND 演算

キーワード	1	2	3	4	5	6	7	...	N
metabolic	1	0	1	0	0	1	0	...	0

↓

キーワード	1	2	3	4	5	6	7	...	N
Protein metabolic	1	0	1	0	0	1	0	...	0

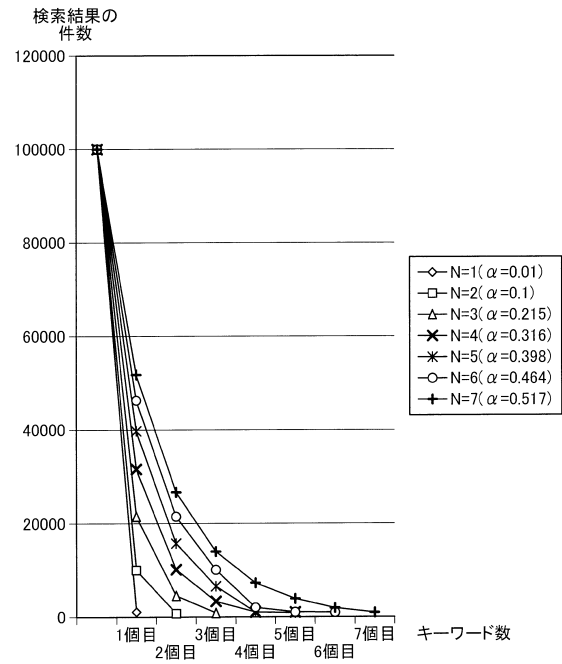
【図 6】

キーワードルートの一例を示す図

キーワード0	件数	キーワード1	件数	キーワード2	件数	キーワード3	件数	...
protein	50000	metabolic	15000	ec	900			...
protein	50000	substrate	27300	abc	12500	transport	1000	...
...	...	...	...	...	...	...	...	...

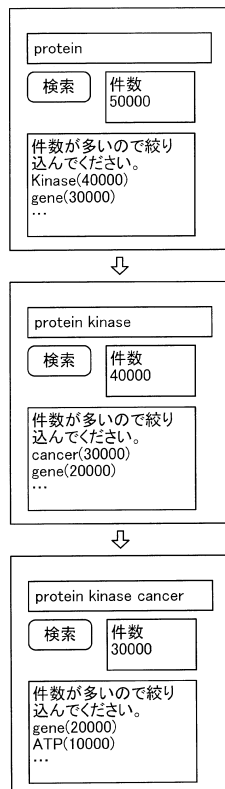
【図 7】

指定のキーワードによる検索結果の件数と指定のキーワードに組み合わせるキーワード数との関係の一例を示す図



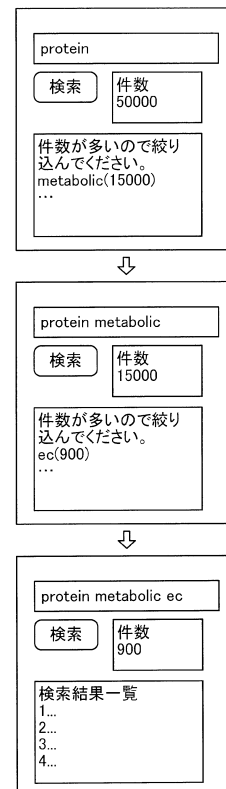
【図 8】

クライアント端末に表示される画面遷移の一例を示す図



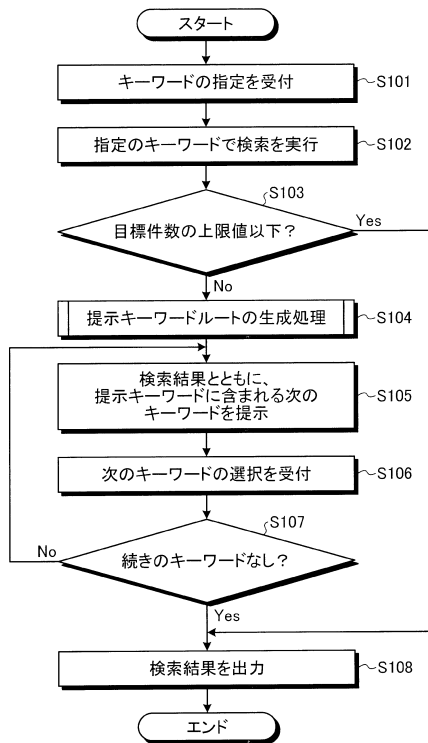
【図 9】

クライアント端末に表示される画面遷移の一例を示す図



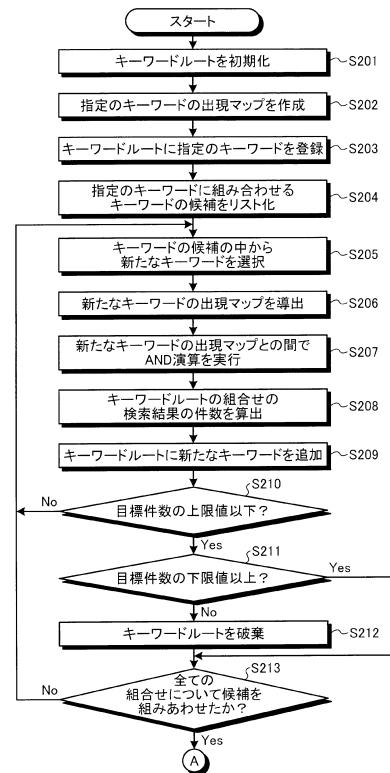
【図 10】

実施例1に係る検索処理の手順を示すフローチャート



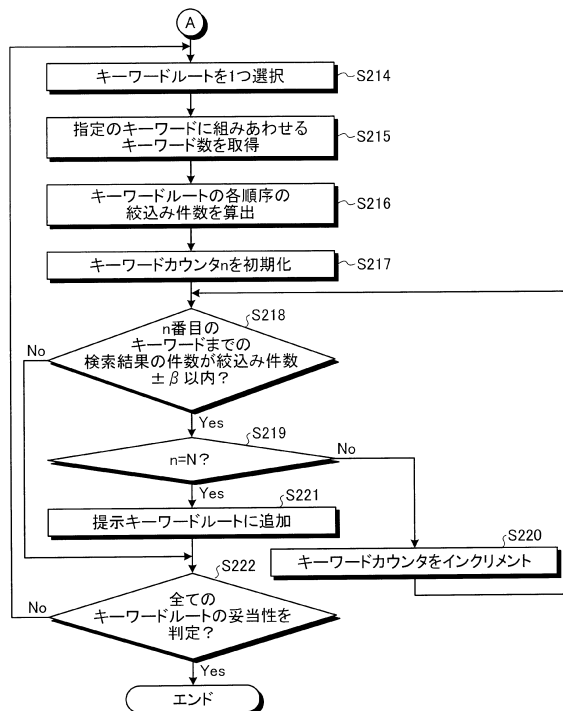
【図 11】

実施例1に係る提示キーワードルートの生成処理の手順を示すフローチャート(1)



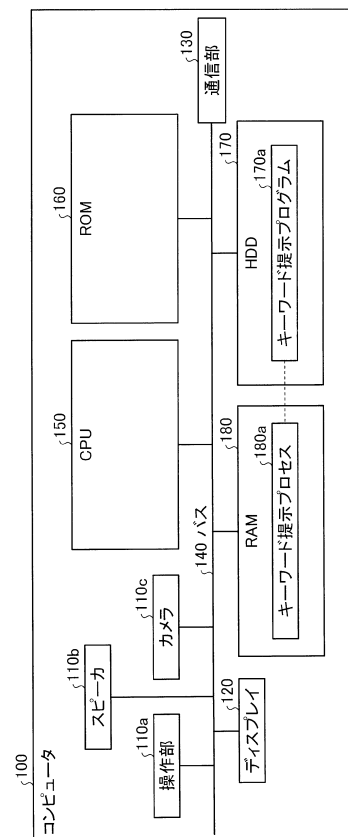
【図 12】

実施例1に係る提示キーワードルートの生成処理の手順を示すフローチャート(2)



【図 13】

実施例1及び実施例2に係るキーワード提示プログラムを実行するコンピュータの一例について説明するための図



---

フロントページの続き

(56)参考文献 特開平 0 9 - 2 1 8 8 8 1 ( J P , A )  
特開 2 0 0 4 - 2 4 6 4 2 2 ( J P , A )  
特開 2 0 0 0 - 2 2 7 9 2 1 ( J P , A )

(58)調査した分野(Int.Cl. , D B 名)  
G 0 6 F 1 7 / 3 0