

(19) 日本国特許庁(JP)

(12) 公開特許公報(A)

(11) 特許出願公開番号

特開2012-69152  
(P2012-69152A)

(43) 公開日 平成24年4月5日(2012.4.5)

(51) Int.Cl.	F I	テーマコード (参考)
<b>G06F 17/30 (2006.01)</b>	G06F 17/30 340Z	
	G06F 17/30 110C	
	G06F 17/30 180Z	

審査請求 有 請求項の数 13 O L (全 15 頁)

(21) 出願番号 特願2011-266249 (P2011-266249)  
 (22) 出願日 平成23年12月5日 (2011.12.5)  
 (62) 分割の表示 特願2005-175174 (P2005-175174) の分割  
 原出願日 平成17年6月15日 (2005.6.15)  
 (31) 優先権主張番号 10/951, 123  
 (32) 優先日 平成16年9月27日 (2004.9.27)  
 (33) 優先権主張国 米国 (US)

(71) 出願人 500046438  
 マイクロソフト コーポレーション  
 アメリカ合衆国 ワシントン州 98052-6399 レッドモンド ワン マイクロソフト ウェイ  
 (74) 代理人 110001243  
 特許業務法人 谷・阿部特許事務所  
 (72) 発明者 チャド クレイトン メリガン  
 アメリカ合衆国 98052 ワシントン州 レッドモンド ワン マイクロソフト ウェイ マイクロソフト コーポレーション内

最終頁に続く

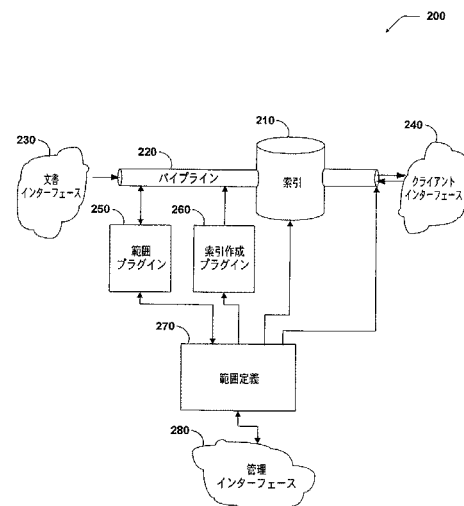
(54) 【発明の名称】 索引キーを使用して検索を絞込む方法および記録媒体

(57) 【要約】

【課題】レコードの検査は、時間がかかり、費用がかかる操作となる可能性がある。

【解決手段】1組の索引キーは、索引検索システムに含まれており、検索の対象である文書の内容ではなく、検索の範囲に関連付けられている。これらの範囲関連の索引キー、または範囲キーによって、検索の範囲を選択することができ、検索が結果を得るためにふるいにかかる必要がある文書の数が高減する。さらに、こうした複合検索範囲に基づいて検索の再ハッシュを無くすために複合検索範囲の索引が提供されるように、複合範囲が認識され、格納される。

【選択図】 図2



**【特許請求の範囲】****【請求項 1】**

文書のクエリのための検索範囲を設定する方法であって、  
前記文書に関連するプロパティを識別するステップであって、前記文書はクロールの間に識別される、ステップと、

前記識別されたプロパティに従って、範囲関連索引キーを生成するステップと、  
前記文書の内容に従って、内容関連索引キーを生成するステップであって、前記内容関連索引キーは、前記文書のキーワードを識別する、ステップと、

コンピューティングデバイスで索引を生成するステップであって、前記索引は、  
前記内容関連索引キーを、それぞれの前記内容関連索引キーにより識別される内容を含む、前記文書のサブセットに関連付けるデータを格納する第 1 のパーティションと、

前記範囲関連索引キーを、それぞれの前記範囲関連索引キーにより識別されるプロパティを有する、前記文書のサブセットに関連付けるデータを格納する第 2 のパーティションと

を含む、ステップと、

キーワードおよび範囲の限定を含む検索クエリに対する結果を提供するステップであって、

前記索引の前記第 2 のパーティションを使用して、前記範囲の限定に一致する、前記文書の前記サブセットを識別するステップ、および

前記索引の前記第 1 のパーティションを使用して、前記キーワードを含む、前記文書の前記サブセットの文書を識別するステップ

を含む、ステップと

を含むことを特徴とする方法。

**【請求項 2】**

前記文書はネットワーク上であることを特徴とする請求項 1 に記載の方法。

**【請求項 3】**

前記索引は、別の検索範囲が前記範囲関連索引キーの組合せによって定義されるように、前記組合せに従って構成された追加の文書レコードを含む第 3 のパーティションをさらに含むことを特徴とする請求項 1 に記載の方法。

**【請求項 4】**

前記範囲関連索引キーの組合せは、ブールの組合せに対応することを特徴とする請求項 3 に記載の方法。

**【請求項 5】**

追加の文書が前記プロパティに関連付けられると追加の文書レコードが前記索引の前記範囲関連索引キーに関連付けられるように、クエリの対象となる前記文書間の変更によって前記索引への更新が行われ、前記変更は、前記文書間に挿入された新しい文書、削除される前記文書間の文書、および前記文書間の文書への変更のうち少なくとも 1 つに対応することを特徴とする請求項 1 に記載の方法。

**【請求項 6】**

文書レコードは、前記クエリのための前記文書を識別する文書識別子を含むことを特徴とする請求項 1 に記載の方法。

**【請求項 7】**

クライアント生成クエリに範囲を提供するため、範囲選択がクライアントによって選択可能であるように、前記範囲関連の索引キーに従って前記範囲選択を生成するステップをさらに含むことを特徴とする請求項 1 に記載の方法。

**【請求項 8】**

前記文書の追加の検索範囲に関連付けられている追加のプロパティから追加の範囲関連索引キーを手動で生成し、処理するインターフェースを提供するステップをさらに含むことを特徴とする請求項 1 に記載の方法。

**【請求項 9】**

10

20

30

40

50

コンピュータに、文書のクエリのための検索範囲を設定する方法を実行させるためのプログラムを記録したコンピュータ読み取り可能な記録媒体であって、前記方法は、

前記文書に関連するプロパティを識別するステップであって、前記文書はクローラの間に識別される、ステップと、

前記識別されたプロパティに従って、範囲関連索引キーを生成するステップと、

前記文書の内容に従って、内容関連索引キーを生成するステップであって、前記内容関連索引キーは、前記文書のキーワードを識別する、ステップと、

コンピューティングデバイスで索引を生成するステップであって、前記索引は、

前記内容関連索引キーを、それぞれの前記内容関連索引キーにより識別される内容を含む、前記文書のサブセットに関連付けるデータを格納する第1のパーティションと、

前記範囲関連索引キーを、それぞれの前記範囲関連索引キーにより識別されるプロパティを有する、前記文書のサブセットに関連付けるデータを格納する第2のパーティションと

を含む、ステップと、

キーワードおよび範囲の限定を含む検索クエリに対する結果を提供するステップであって、

前記索引の前記第2のパーティションを使用して、前記範囲の限定に一致する、前記文書の前記サブセットを識別するステップ、および

前記索引の前記第1のパーティションを使用して、前記キーワードを含む、前記文書の前記サブセットの文書を識別するステップ

を含む、ステップと

を含むことを特徴とするコンピュータ読取可能な記録媒体。

#### 【請求項10】

前記索引を生成するステップは、別の検索範囲が前記範囲関連索引キーの組合せによって定義されるように、前記組合せに従って構成された追加の文書レコードを含む第3のパーティションを含むステップをさらに含むことを特徴とする請求項9に記載のコンピュータ読取可能な記録媒体。

#### 【請求項11】

前記第3のパーティションは、前記第1のパーティションおよび第2のパーティションの更新とは非同期に更新されることを特徴とする請求項10に記載のコンピュータ読取可能な記録媒体。

#### 【請求項12】

前記索引を生成するステップは、追加の文書が前記識別されたプロパティに関連付けられると追加の文書レコードが前記範囲関連索引キーに関連付けられるように、前記追加の文書がクエリの対象である前記文書間に挿入されると前記索引を更新するステップをさらに含むことを特徴とする請求項9に記載のコンピュータ読取可能な記録媒体。

#### 【請求項13】

文書レコードの各々は、前記クエリのための前記文書を識別する文書識別子を含むことを特徴とする請求項9に記載のコンピュータ読取可能な記録媒体。

#### 【発明の詳細な説明】

#### 【技術分野】

#### 【0001】

ネットワークおよびファイルシステム間で内容を検索する場合、多くの形式で検索が提供されてきたが、通常は、検索エンジンを変形して提供されるものである。検索エンジンは、指定されたキーワードについて文書を検索し、キーワードが見つかった文書のリストを返すプログラムである。

#### 【背景技術】

#### 【0002】

一般に、検索エンジンは、スパイダを送信してできるだけ多くの文書を取得してくることによって動作する。このようにして取得した文書は、インデクサと呼ばれる別のプログ

10

20

30

40

50

ラムによって読み取られ、各文書に含まれる単語に基づいて索引 ( i n d e x ) が作成される。索引とは、一意のレコードをそれぞれ識別するキーまたはキーワードのリストである。索引によって、特定のレコードを見つけ、レコードを索引フィールドでより速くソートすることができるようになる。検索エンジンは、理想的にはクライアントまたはユーザによるクエリごとに意味のある結果のみが返されるように、その索引を作成するアルゴリズムを使用する。

【発明の概要】

【発明が解決しようとする課題】

【0003】

こうしたクエリでは、一貫してキーワードまたは索引キーが使用される傾向にある。ユーザが検索クエリを長いテキスト文字列として入力しようと、ブール演算子の連結として入力しようと、検索エンジンは、キーワードに一致する文書に対応して入力されたキーワードについて、すべてのレコードを検査する。次いで、ブール演算子の制約を満たす、または長いテキスト文字列に対応するレコードのサブセットが戻される。これらのレコードの検査は、時間がかかり、費用がかかる操作となる可能性がある。さらに、クライアントは、特定のキーワードを含む文書の全レコード検索を望まない場合がある。

10

【課題を解決するための手段】

【0004】

本発明の実施態様は、単にキーワードを提供するのではなく、検索の範囲を定義する範囲キーと呼ばれる索引キーのクラスを提供することによって上記の制限を解消するシステムおよび方法に関連する。範囲キーが検索クエリに入力されると、範囲キーは、検索される索引レコードの範囲を限定する。例えば、範囲キーは、検索結果を、.mpgファイルなど何らかのファイルタイプに限定することによって検索の範囲を限定することができる。別の範囲キーは、URL(ユニフォームリソースロケータ)に従って検索の範囲を限定することができる。したがってそのURL下の文書のみが検索される。さらに別の範囲キーは、検索の範囲を、ユーザのコンピュータまたは他のネットワーク接続されたコンピュータ上の特定のデータベースに限定することができる。したがって本発明は、ユーザが特定のクラスの索引キーを使用して検索の範囲を限定することができるようにすることによって検索の時間および費用を大幅に低減することにより、上記の問題を解決する。

20

【0005】

本発明の別の態様によれば、複合範囲も認識され、格納される。この追加の索引パーティションは、基本的な範囲の組合せである範囲定義を含む。これらの複合範囲に対応する文書はすでに解決されており、これらの複合範囲を参照すると、より迅速な検索が可能となる。

30

【図面の簡単な説明】

【0006】

【図1】本発明の一実施形態例で使用することができるコンピューティング装置例を示す図である。

【図2】本発明による索引キーを使用して検索を絞込むためのシステム例を示すブロック図である。

40

【図3】本発明による索引の構造例を示すブロック図である。

【図4】本発明による複合範囲の管理を示すブロック図である。

【図5】本発明による索引の生成のプロセス例を示す論理フローチャートである。

【図6】本発明による索引のコンパイルのプロセス例を示す論理フローチャートである。

【図7】本発明によるクエリの処理のプロセス例を示す論理フローチャートである。

【発明を実施するための形態】

【0007】

次に、本明細書の一部を形成し、一例として本発明を実施する特定の実施形態例を示す添付の図面を参照して本発明について以下でより十分説明する。しかし、本発明は、異なる多くの形態で具体化することができ、本明細書に記載した実施形態に限定されるものと

50

解釈されないものとする。むしろ、これらの実施形態は、この開示が完全で欠落がなく、本発明の範囲を当業者に十分伝えるように提供される。特に、本発明は、方法または装置として具体化することができる。したがって、本発明は、完全にハードウェアの実施形態、完全にソフトウェアの実施形態、またはソフトウェアおよびハードウェアの態様を組み合わせた実施形態の形態をとることができる。したがって、以下の詳細な説明を限定の意味で理解すべきではない。

#### 【0008】

##### 動作環境例

図1を参照すると、本発明を実施する1つのシステム例は、コンピューティング装置100などのコンピューティング装置を含んでいる。コンピューティング装置100は、クライアント、サーバ、モバイル装置、または他の任意のコンピューティング装置として構成することができる。極めて基本的な構成では、コンピューティング装置100は一般に、少なくとも1つの処理ユニット102およびシステムメモリ104を含む。コンピューティング装置の正確な構成および種類に応じて、システムメモリ104は、揮発性(RAMなど)、不揮発性(ROM、フラッシュメモリなど)、またはこれら2つの何らかの組合せとすることができる。システムメモリ104は、一般にオペレーティングシステム105、1つまたは複数のアプリケーション106を含んでおり、またプログラムデータ107を含んでいてもよい。一実施形態において、アプリケーション106は、本発明の機能を実施する検索絞込みアプリケーション120を含む。この基本的な構成は、図1の点線108内の構成要素によって示している。

10

20

#### 【0009】

コンピューティング装置100は、追加の特徴および機能を有していてもよい。例えば、コンピューティング装置100は、磁気ディスク、光ディスク、テープなどの追加のデータ記憶装置(取外式および/または固定式)を含むこともできる。こうした別の記憶装置を、図1に、取外式記憶装置109および固定式記憶装置110で示している。コンピュータ記憶媒体には、コンピュータ読取可能命令、データ構造、プログラムモジュール、他のデータなど、情報を記憶するための任意の方法または技術で実施される揮発性および不揮発性の取外式および固定式媒体などがある。システムメモリ104、取外式記憶装置109、および固定式記憶装置110は、すべてコンピュータ記憶装置媒体の例である。コンピュータ記憶媒体には、それだけには限定されないが、RAM、ROM、EEPROM、フラッシュメモリまたは他のメモリ技術、CD-ROM、デジタル多用途ディスク(DVD)または他の光ディスク記憶装置、磁気カセット、磁気テープ、磁気ディスク記憶装置または他の磁気記憶装置、または所望の情報の格納に使用でき、コンピューティング装置100からアクセスできる他の任意の媒体などがある。こうした任意のコンピュータ記憶媒体は、装置100の一部とすることができる。コンピューティング装置100は、キーボード、マウス、ペン、音声入力装置、タッチ入力装置などの入力装置112を有することもできる。ディスプレイ、スピーカ、プリンタなどの出力装置114を含むこともできる。

30

#### 【0010】

コンピューティング装置100は、装置が例えばネットワークを介して他のコンピューティング装置118と通信できるようにする通信接続116も含む。通信接続116は、通信媒体の一例である。通信媒体は一般に、コンピュータ読取可能命令、データ構造、プログラムモジュール、または他のデータによって搬送波または他の移送機構などの変調されたデータ信号に組み込まれる。これには任意の情報配送媒体がある。「変調されたデータ信号」という用語は、信号に情報を符号化するように1つまたは複数のその特性が設定または変更された信号を意味する。通信媒体には、それだけには限定されないが一例として、有線ネットワーク、直接配線された接続などの有線媒体、および音響、RF、赤外線、その他の無線媒体などの無線媒体がある。コンピュータ読取可能な媒体という用語は、本明細書で使用する場合、記憶媒体および通信媒体の両方を含む。

40

#### 【0011】

50

### 検索を絞込むための実施形態例

次の説明および特許請求の範囲を通じて、「文書」という用語は、ネットワーク文書、ファイル、フォルダ、Webページ、および他のリソースなど、ネットワークの検索クエリまたはクロールの結果として返される任意の考え得るリソースを指す。「索引キー」という用語は、索引の検索クエリまたは作成で1組の文書を標的にするために使用される検索に関連付けられている任意のキーワードまたはキーを指す。「範囲キー」という用語は、検索が開始される前に検索される文書の数が高減されるように、検索の範囲を絞るために使用することができる任意の索引キーを指す。検索の範囲は、ファイルタイプなどの属性、何らかのデータベースまたはURLなどの場所に従って、または検索される文書の高減する他の基準によって絞ることができる。

10

#### 【0012】

本発明の実施形態は、索引内の各項目の範囲を範囲キーで表し、適切な範囲キーを別の制限としてクエリに追加することによって絞込まれた(scoped)すべてのクエリのクエリ効率を上げることに関連する。この方法は、文書のプロパティ(URLまたは他のメタデータなど)に基づいてユーザキーワードに一致する文書ごとに範囲条件を再計算する方法とは対照的に提供される。ほとんどの範囲は、すべてクロールされた1組の項目の細かい1片を表すため、絞込まれたクエリの効率は、その1片の細さより少ないが、それに関連する倍数だけ向上する。一例において、ポータル(すなわち検索を含む多数のサービスを提供するWebサイト)の管理者は、ポータルのユーザが機能上の仕様に特に興味があると決定することができる。したがって、他の文書内のこうした無用の単語をふるいにかけることなく、キーワードについてこれらの仕様を検索するだけの検索機構を提供することができるという点で有利となる。管理インターフェースを介して、管理者は、rule:profile="spec"を有する範囲を定義することができる。またはこの範囲は、基本的な範囲としてすでに定義されていてもよい。管理者は、基本的な範囲を使用して、rule:profile="spec" AND filetype="Text" AND author!="John Doe"を有する複合範囲を定義することができる。この範囲にクライアントのユーザインターフェースのわかりやすい名前を提供し、この範囲がどのサイトから入手可能であるかを指定した後、範囲は、クライアントのドロップダウン範囲リストに表示される。この範囲選択は、そのプロパティ値を含む文書のみをクエリ結果として返す。このように管理者によって定義された範囲は、作成された範囲と呼ぶこともできる。

20

30

#### 【0013】

さらに、いくつかのデフォルトの範囲選択を使用することができる。一般のクライアントに、単に絞込まれていないクエリを示す「すべての内容」を検索する選択肢を示すことができる。「このサイト」の範囲選択は、現在のサイトおよびそのサブサイトのすべての文書を検索する。特定のサブセットに含まれる文書の範囲を指定するだけというよりむしろ、いくつかの結果を除外する範囲を生成することもできる。例えば、範囲定義は、テキスト(.txt)文書を除くすべての文書タイプについての検索に対応することができる。特定のWebサイトURLに関連付けられているものを除いて、ネットワークにおけるすべての文書の検索に対応することができる別の範囲定義を考えることもできる。範囲選択の数は、デフォルトであろうと、作成されたものでであろうと、その他の場合であろうと、本明細書に記載したものに限定されない。

40

#### 【0014】

別の実施形態においては、クライアントに、検索要求に直接範囲キーを入力する選択肢が与えられる場合もある。クライアントによって使いやすい名前が範囲キーに与えられ、検索要求にその名前を入力すると、関連する範囲によって検索が限定される。

#### 【0015】

図2は、本発明による索引キーを使用して検索を絞込む(scoping)システム例の機能ブロック図を示している。システム200は、索引210、パイプライン220、文書インターフェース230、クライアントインターフェース240、範囲プラグイン250、索引作成プラグイン260、範囲の記述270、および管理インターフェース28

50

0を含む。

【0016】

索引210は、内容キー（すなわちキーワード）および範囲キーの別々の索引を含むように構成される。索引210の構成のより詳細な説明については、図3の説明で後述する。これらの索引のレコードは、クライアントのクエリに結果を提供するのに使用される。一実施形態において、索引210は、索引レコードの記憶域をまとめて提供する複数のデータベースに対応する。

【0017】

パイプライン220は、索引作成のために文書または文書のレコードを取得する収集機構の表現例である。パイプライン220は、データに対応するレコードが索引210に入力される前に様々なプラグイン（範囲プラグイン250など）によるデータのフィルタリングを可能にする。

【0018】

文書インターフェース230は、複数のデータベースおよびネットワーク場所を介して文書を取り出すためのプロトコル、ネットワークアクセスポイント、およびデータベースアクセスポイントを提供する。例えば、文書インターフェース230は、インターネットへのアクセスを提供するとともに、ローカルサーバのデータベースへのアクセスおよび現在のコンピューティング装置上のデータベースへのアクセスを提供することができる。他の実施形態は、本発明の意図または範囲から逸脱することなく、様々なプロトコルを使用して他の文書位置にアクセスすることができる。

【0019】

クライアントインターフェース240は、検索を定義し、開始するためのクライアントによるアクセスを提供する。検索は、キーワードおよび/または範囲キーに従って定義することができる。検索クエリを処理する方法例については、以下、図7を参照して詳しく説明する。

【0020】

範囲プラグイン250は、いくつかの収集パイプラインプラグインのうちの1つである。範囲プラグイン250は、範囲キー（すなわち範囲索引内で索引付けされる項目）として再度発行されるプロパティ値を識別する。範囲（例えばファイルタイプ、URLなど）に対して興味のあるプロパティとして識別されたプロパティは、文書インターフェース230を介して提供された文書がクロールされる間に範囲プラグイン250によって集められる。これらのプロパティは、範囲プラグイン250によってパイプライン220に再度発行されて、索引210に含まれる。これらのプロパティは、これらのプロパティに従ってクライアントに範囲選択を提供する管理者または他の操作者（entity）によっても使用可能である。

【0021】

索引作成プラグイン260は、パイプライン220に接続された別のプラグインである。索引作成プラグインは、索引210を生成し、パーティションし、更新する機構を提供する。索引210を生成する方法例については、以下の図5の説明でより詳しく説明する。索引210を更新する方法例については、以下の図6の説明でより詳しく説明する。一実施形態において、索引作成プラグイン260は、キーワードを一時的にキャッシュし、これらの結果が索引210にフラッシュされる前にクロールされた文書から生成されたキーを絞込む単語リストを提供する。索引210のレコードは、これらの単語リストに含まれるクロール結果からのデータで埋められる。

【0022】

範囲の記述270は、範囲についての情報を格納するテーブルを提供する。例えば、範囲の記述270は、範囲、範囲のルール、可視性、および範囲関連のプロパティ、および検索クエリで使用するために生成された範囲選択に対応する他の属性など、管理情報および内部情報を含み得る。範囲の記述270は、範囲プラグイン250を介して範囲選択を生成するプロパティを受信する。範囲の記述270は、索引210内の範囲索引の生成お

10

20

30

40

50

よび構成のために、索引作成プラグイン 260 によってもアクセスされる。索引 210 も、複合範囲別索引の生成および更新のために、範囲の記述 270 にアクセスする（以下の図 3 および図 4 を参照）。また、範囲の記述 270 は、クライアントが検索クエリに含めるための範囲、または検索に適用するための範囲選択を選択できるように、クライアントインターフェース 240 でもアクセスされる。

#### 【0023】

管理インターフェース 280 も、管理者または他の制御機構（例えば自動プログラム）が、範囲プラグイン 250 によって提供されたプロパティを取得し、検索クエリを使用するための範囲選択を作成することができるように、範囲の記述 270 にアクセスする。管理インターフェース 280 は、（例えばインターネットログインアクセスを介して）範囲選択の作成、および範囲の記述の操作を可能にする任意のフォーマットに従って提供され得る。

10

#### 【0024】

機能ブロック間の一方向および双方向の通信のシステム 200 を示しているが、こうした通信タイプの任意のものを、本発明の意図および範囲から逸脱することなく、別のタイプに変更することができる（例えばすべての通信は、一方向通信ではなく双方向を必要とする受信通知メッセージを有し得る）。

#### 【0025】

図 3 は、本発明による索引の構造例の機能ブロック図を示している。索引 300 は、内容別索引 (content index) (.ci) 310、基本的範囲別索引 (basic scopes index) (.bsi) 320、および複合範囲別索引 (compound scopes index) (.csi) 330 を含む。

20

#### 【0026】

内容別索引 310 は、検索クエリで使用されるキーワードおよび他の索引キーに対応する文書を列挙する逆引き索引 (inverted index) で構成されるレコードを含む。しかし、範囲キーは、基本的範囲別索引 320 に流用される。

#### 【0027】

基本的範囲別索引 320 は、基本的な範囲に対応する文書のレコードを含む。基本的な範囲は、一般に、文書の単一の範囲関連プロパティに対応する範囲選択を指す。例えば、サイト <http://www.example.com> でクロールされた文書の数字 ID は、プロパティ (サイト) および値 ("example.com") を包含するように範囲プラグイン 250 によって構成された範囲キーの文書リストに記録される。

30

#### 【0028】

複合範囲別索引 330 は、基本的範囲別索引 320 の基本的な範囲の組合せから生成された範囲を含む。例えば、1つの複合範囲は、特定の URL にも関連する特定のファイルタイプの文書のレコードを含むことができる。

#### 【0029】

図 4 は、本発明による複合範囲を管理するためのブロック図例を示している。ブロック図 400 は、基本的範囲別索引 410、元の複合範囲別索引 420、および新しい複合範囲別索引 430 を含む。

40

#### 【0030】

基本的範囲別索引が追加の基本的な範囲で更新されると（以下の図 6 参照）、複合範囲別索引も更新されなければならない。元の複合範囲別索引 420 の位置線 422 は、新しい複合範囲が含まれるべき位置を示している。元の複合範囲別索引 420 のコピーが作成されて、新しい複合範囲別索引 430 が生成される。元の複合範囲別索引 420 のコピーは、新しい複合範囲が含まれるべき位置に到達するまで作成される。次いで新しい複合範囲 432 は、新しい複合範囲別索引 430 に書き込まれる。新しい複合範囲 432 が含まれた後、引き続き元の複合範囲別索引 420 のコピーが続行する。新しい複合範囲 432 に続く複合範囲 434 は、新しい複合範囲 432 の包含を補うオフセットだけ相殺して元の複合範囲別索引 420 からコピーされる。

50



## 【 0 0 3 1 】

図5は、本発明による索引を生成するプロセス例の論理フローチャートを示している。プロセス500は、ブロック502で開始し、アクセスが文書のコーパスに提供される。処理はブロック504に進む。

## 【 0 0 3 2 】

ブロック504で、文書のコーパスがクロールされて、存在する文書、およびこうした文書に関連付けられているプロパティ（ファイルタイプなど）を決定する。次いで文書ごとの識別子またはID、およびその関連のプロパティがクロールの結果として転送される。処理はブロック506に進む。

## 【 0 0 3 3 】

ブロック506で、範囲に関連する文書に関連付けられているプロパティが範囲プラグインによって取得される。範囲プラグインは、プロパティから範囲定義を作成する。範囲定義は、クライアントがその範囲に従って検索を限定することができる範囲選択を作成するために管理者が使用することができる。処理はブロック508に進む。

## 【 0 0 3 4 】

ブロック508で、取得されたプロパティから作成された範囲定義は、クロールの結果の間に範囲キーとして発行される。これらの範囲キーは、文書の内容よりむしろ検索の範囲を対象としながら、クロールから生成されたキーワードおよび他の索引キーと同じように動作する。範囲キーの生成のために取得されたプロパティの一部は、文書タイプ、文書のURL、文書の作成者、および他のプロパティを含む。範囲キーは、範囲キーのタイプの識別子（ID）、および特定の範囲キーを識別するテキスト文字列を含むように生成される。例えば、範囲キー関連のURLのIDが237である場合、<http://www.example.com>内の文書に対応する範囲キーは「[237] <http://www.example.com>」となる。この範囲キーは、パイプライン内に発行され、実質的に索引内の文書に関連付けられる。範囲キーが発行されると、処理はブロック510に進む。

## 【 0 0 3 5 】

ブロック510で、範囲キー、キーワード、およびすべての文書内で見つけられた他の蓄積されたプロパティが索引に送込まれる（flushed）。この送込みによって、キーおよびプロパティがディスクに書き込まれる。送込みの間、範囲キーは、分けられ、基本的範囲別索引に送信され、残りのデータは内容別索引に送信される。処理はブロック512に進む。

## 【 0 0 3 6 】

ブロック512で、複合範囲別索引が索引内に生成される。一実施形態において、索引について開始されたコンパイルプロセスに回答して複合範囲別索引が生成される。複合範囲別索引の生成に使用される1つのプロセス例については、以下の図6の説明で説明する。一実施形態において、複合範囲は、クライアントからのクエリによって定義される。別の実施形態において、索引がインスタンス化される前に管理者によって複合範囲のリストが生成される。複合範囲別索引が生成されると、処理はブロック514に進み、プロセス500が終了する。

## 【 0 0 3 7 】

一実施形態においては、クロールが開始されると基本的範囲別索引はデータで埋められるが、複合範囲別索引は、クロールが完了し、基本的範囲別索引が完全に構築されるまで埋められない。複合範囲別索引の構築を待つことで、基本的範囲別索引へのクエリを低減することによってオーバーヘッドが低減する。

## 【 0 0 3 8 】

図6は、本発明による索引をコンパイルするプロセス例の論理フローチャートを示している。プロセス600は、ブロック602で開始し、コンパイルプロセスが開始する。一実施形態において、プロセス600は、任意の既存の複合範囲を更新するために、ある時間間隔（例えば15分ごと）で非同期的に開始される。別の実施形態において、図5のプロセス500がブロック512に入ると、プロセス600が開始されて、新しく生成され

10

20

30

40

50

た基本的範囲別索引から複合範囲別索引が生成される。さらに別の実施形態において、プロセス600は、この索引への別の送込みに応答して開始される。プロセス600が開始すると、処理は決定ブロック604に進む。

#### 【0039】

決定ブロック604で、複合範囲別索引内の各範囲に対応する変更レコードは、現在の複合範囲が変わったことを示しているかどうかの決定が行われる。一実施形態において、初めて複合範囲別索引を生成するには、同様のプロセスを使用し、そのデフォルトの設定は、すべての範囲が変わったということを前提とする。したがって、新しい複合範囲別索引の生成、および複合範囲別索引の更新は、同じコンパイルプロセスによって処理される。複合範囲が変わると、処理はブロック606に進む。

10

#### 【0040】

ブロック606で、コンパイルプロセスは、ユーザクエリに類似のクエリを実行し、クエリプロセスが複合範囲別索引内の範囲に対応する文書のリストを更新することができるようにする。次いで文書のリストは、複合範囲別索引が（前のバージョンが存在するとき）その前のバージョンからコピーされると、複合範囲別索引に追加される。処理は決定ブロック610に進む。

#### 【0041】

あるいは、変更レコードが特定の範囲は変わっていないことを示す場合、処理はブロック608に進む。ブロック608で、範囲は変わっていないため、複合範囲別索引の前のバージョンの範囲に対応する文書IDのリストがそっくりそのままコピーされる。処理は決定ブロック610に進む。

20

#### 【0042】

決定ブロック610で、コンパイルプロセス中により多くの複合範囲を新しい複合範囲別索引にコピーする必要があるかどうかの決定が行われる。より多くの複合範囲をコピーすべき場合、処理は決定ブロック610に戻って、複合範囲が変わったかどうかの決定が行われる。しかし、より多くの複合範囲を新しい複合範囲別索引に転送する必要がない場合、処理はブロック612に進み、処理600が終了する。

#### 【0043】

文書のコーパスに対する更新は、いつでも行うことができる。文書のコーパスの文書IDは、メモリ内単語リスト、または複数の単語リストで連続的に更新される。単語リストを埋めることは、クライアントにより開始された検索、コーパスが再度クロールされるようにするリフレッシュ動作（action）、または文書のコーパス間での変更の発見をもたらす他の様々な操作のいずれかから生じる。文書が変更される（追加、削除、変更など）と、変更された文書の文書IDは、次いで、変更の種類とともにメモリ内単語リストに転送される。更新された文書IDを含む単語リストは、次いで索引に送込まれる。文書への変更によって、内容別索引への更新が行われるとともに、基本的範囲別索引も更新される。変更を発見した漸進的なクロールが完了し、基本的範囲別索引が更新されると、ネットワーク上の文書への変更を反映するように複合範囲別索引も更新される。プロセス600は、更新が検索される文書のコーパス間での新しい文書であるか、コーパスからの文書の削除であるか、文書の範囲に影響を与える文書への変更であるかに関係なく、複合範囲別索引での更新を反映するために使用される。複雑なプロセスを非同期に稼働させることによって、ネットワーク上の文書への変更を反映するために複合範囲別索引が頻繁に更新される。

30

40

#### 【0044】

図7は、本発明によるクエリを処理するプロセス例の論理フローチャートを示している。プロセス700は、ブロック702で開始し、索引がインスタンス化され、クライアントからクエリを受け付ける用意ができています。処理は決定ブロック704に進む。

#### 【0045】

決定ブロック704で、検索クエリがクライアントによって開始されたかどうかの決定が行われる。クライアントは、クエリを開始するユーザ、または検索を要求しているプロ

50

グラムに対応している。検索が開始されていない場合、検索クエリが開始されるのを待っている間、処理はブロック704に戻る。しかし、検索クエリが開始すると、処理は決定ブロック706に進む。

【0046】

決定ブロック706で、範囲キーが検索要求に使用されたかどうかの決定が行われる。範囲キーが存在しない場合、処理はブロック716に進む。しかし、範囲キーが検索要求に存在する場合、処理は決定ブロック708に進む。

【0047】

決定ブロック708で、範囲キーのインスタンスが複合範囲の一部であるかどうかの決定が行われる。範囲キーが複合範囲の一部として使用されていない場合、処理は決定ブロック712に進む。しかし、範囲キーが複合範囲の一部である場合、処理はブロック710に進む。

10

【0048】

ブロック710で、検索クエリに含まれる複合範囲に対応するとして、その文書IDによって識別される文書について、複合範囲別索引(.csi)が調べられる。次いで、検索の完了まで、これらの文書に対応する文書IDが戻されて、検索結果に追加される。処理は決定ブロック712に進む。

【0049】

決定ブロック712で、範囲キーのインスタンスが複合範囲の一部であるかどうかの決定が行われる。範囲キーが基本的な範囲に対応していない場合、処理はブロック716に進む。しかし、範囲キーが基本的な範囲に対応している場合、処理はブロック714に進む。

20

【0050】

ブロック714で、検索クエリに含まれる範囲キーに対応するとして、その文書IDによって識別される文書について基本的範囲別索引(.bsi)が調べられる。次いで、検索の完了まで、これらの文書に対応する文書IDが戻されて、検索結果に追加される。処理は決定ブロック716に進む。

【0051】

決定ブロック716で、文書の内容に関連するキーワードまたは他の索引キーが検索要求に含まれているかどうかの決定が行われる。キーワードが検索要求に含まれていない場合、処理は決定ブロック720に進む。しかし、キーワードが検索要求に含まれている場合、処理はブロック718に移動する。

30

【0052】

ブロック708で、検索クエリに含まれるキーワードに対応するとして、その文書IDによって識別される文書について内容別索引(.ci)が調べられる。一実施形態において、内容別索引をキーワードについて検索するとき、検索は、基本的範囲別索引および/または複合範囲別索引に従って予め定義されている範囲に限定される。次いで、検索の完成まで、これらの文書に対応する文書IDが戻されて、検索結果に追加される。処理は決定ブロック710に進む。

【0053】

40

ブロック720で、異なる索引パーティションで重複する文書の文書IDの集まりがクエリ結果として返される。例えば、文書IDは、基本的範囲別索引の範囲に対応し、特定のキーワードを含む可能性もある。検索結果がこの特定の範囲によって限定され、キーワードを含んでいる場合、文書IDは、索引パーティション間で重複することになる。これらの重複するIDは、検索の結果を表す。結果に含まれる各文書へのポイントは、次いで検索要求に回答してクライアントに提供することができる。一般に、文書が特定の範囲を有するかどうかを確認するために、文書プロパティを調べる代わりに、索引間の重複する文書を決定するのがかなり迅速である。一般に文書プロパティが無作為にデータベース内にある場合、索引は、キー(範囲キーまたはキーワード)に従ってディスク上にクラスタ化される。したがって本発明は、範囲を検索クエリに適用することができる速度が大幅に

50

向上し、およびより容易となる。結果が提供されると、処理はブロック 7 2 2 に移動して、処理 7 0 0 は終了する。

【 0 0 5 4 】

一実施形態において、操作ブロック 7 0 8 ~ 7 1 8 で提供されたプロセスステップは、順次的なものではない。代わりに、基本的範囲別索引、または複合範囲別索引、または内容別索引は、通常のキー、およびクエリの範囲キーに基づいて調べられ、それらが調べられる順序は、キーに対応する文書 ID のソート順によって決まる。さらに、検索クエリには複数の範囲、つまり複合範囲および基本的な範囲がある場合があるため、操作ブロック 7 0 8 ~ 7 1 8 で提供された処理ステップを複数回繰り返す必要がある場合がある。

【 0 0 5 5 】

別の実施形態において、新しい複合範囲が要求によって作成される場合、複合範囲別索引は、各検索要求の後に新しい複合範囲で更新される。複合範囲別索引は、上記の図 4 の説明で提供された方法に従って更新される。

【 0 0 5 6 】

さらに別の実施形態において、検索要求の絞込まれた部分は、クライアントによってなされた検索選択である。検索選択は、提供された範囲のリストに従って予め選択された結果の範囲に対応する。範囲のこのリストは、範囲定義に従って管理者によって生成することができる。

【 0 0 5 7 】

上記の仕様、例、およびデータにより、製造する際の完全な説明および本発明の構成の使用が提供されている。本発明の多くの実施形態は、本発明の意図および範囲から逸脱することなく作成することができるので、本発明は特許請求の範囲にある。

【 符号の説明 】

【 0 0 5 8 】

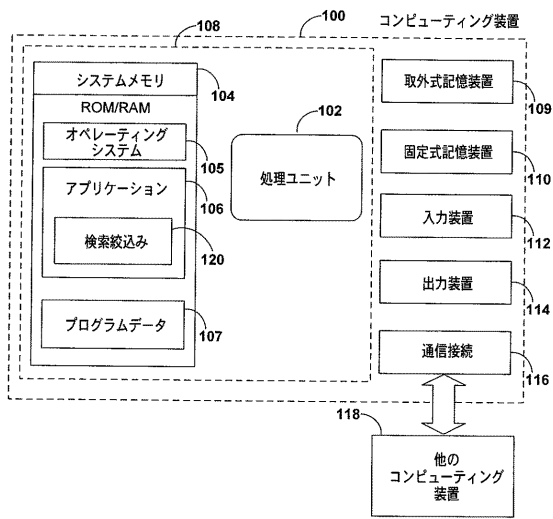
- 1 0 0 コンピューティング装置
- 1 0 2 処理ユニット
- 1 0 4 システムメモリ
- 1 0 5 オペレーティングシステム
- 1 0 6 アプリケーション
- 1 0 7 プログラムデータ
- 1 0 9 取外式記憶装置
- 1 1 0 固定式記憶装置
- 1 1 2 入力装置
- 1 1 4 出力装置
- 1 1 6 通信接続
- 1 1 8 他のコンピューティング装置
- 1 2 0 検索絞込み

10

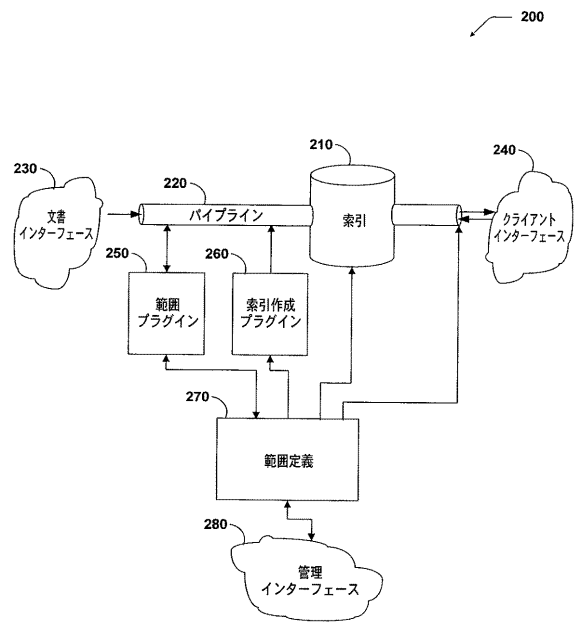
20

30

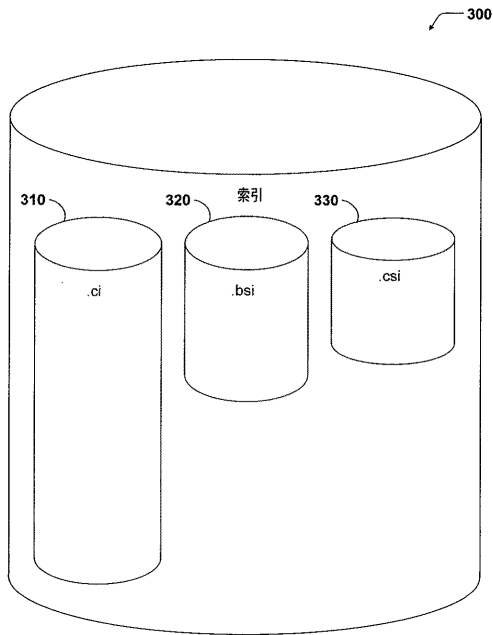
【 図 1 】



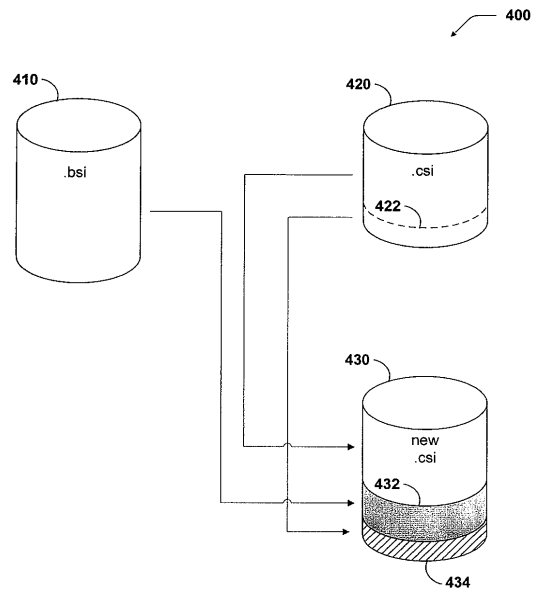
【 図 2 】



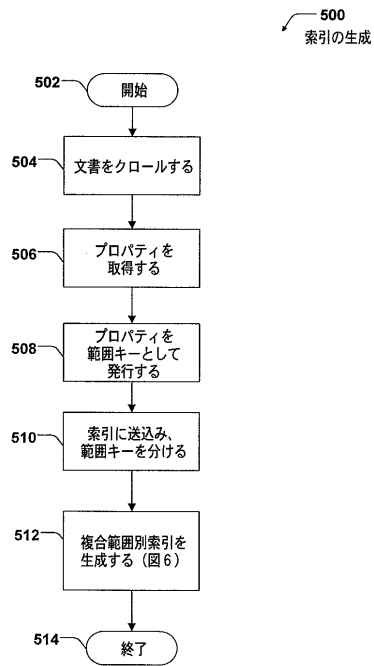
【 図 3 】



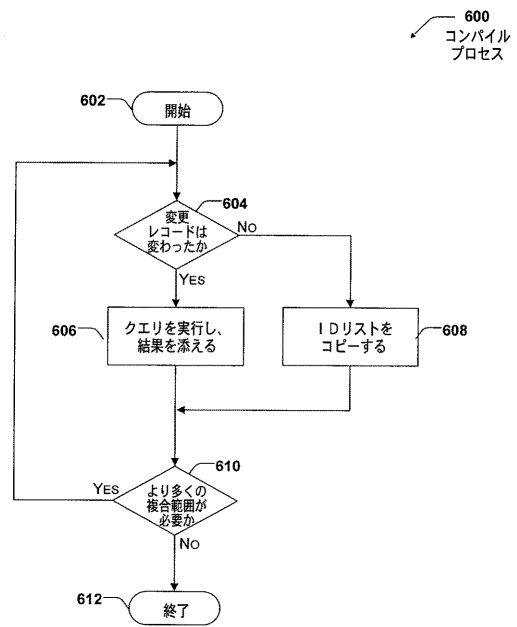
【 図 4 】



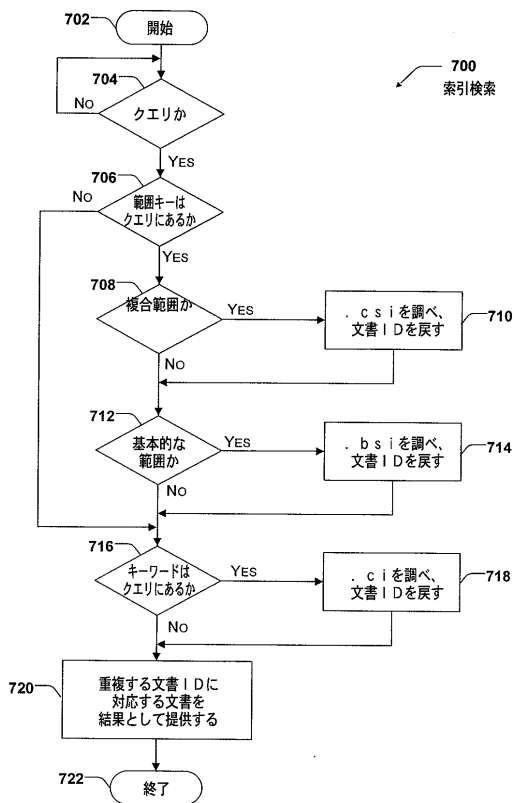
【 図 5 】



【 図 6 】



【 図 7 】



---

フロントページの続き

- (72)発明者 デビッド ジェイ・リー  
アメリカ合衆国 98052 ワシントン州 レッドモンド ワン マイクロソフト ウェイ マ  
イクロソフト コーポレーション内
- (72)発明者 ドミトリー マイヤーゾン  
アメリカ合衆国 98052 ワシントン州 レッドモンド ワン マイクロソフト ウェイ マ  
イクロソフト コーポレーション内
- (72)発明者 カイル ジー・ペルトネン  
アメリカ合衆国 98052 ワシントン州 レッドモンド ワン マイクロソフト ウェイ マ  
イクロソフト コーポレーション内