

(19) 日本国特許庁(JP)

(12) 特 許 公 報(B2)

(11) 特許番号

特許第5334333号

(P5334333)

(45) 発行日 平成25年11月6日(2013.11.6)

(24) 登録日 平成25年8月9日(2013.8.9)

(51) Int.Cl.	F I
G 0 6 F 17/30 (2006.01)	G O 6 F 17/30 3 3 O B
G 0 6 F 12/00 (2006.01)	G O 6 F 17/30 3 5 O C
	G O 6 F 17/30 3 4 O D
	G O 6 F 12/00 5 1 2

請求項の数 19 (全 14 頁)

(21) 出願番号	特願2010-509482 (P2010-509482)	(73) 特許権者	500046438
(86) (22) 出願日	平成20年5月16日(2008.5.16)		マイクロソフト コーポレーション
(65) 公表番号	特表2010-528369 (P2010-528369A)		アメリカ合衆国 ワシントン州 9805
(43) 公表日	平成22年8月19日(2010.8.19)		2-6399 レッドモンド ワン マイ
(86) 国際出願番号	PCT/US2008/063991		クロソフト ウェイ
(87) 国際公開番号	W02008/147736	(74) 代理人	100140109
(87) 国際公開日	平成20年12月4日(2008.12.4)		弁理士 小野 新次郎
審査請求日	平成23年5月9日(2011.5.9)	(74) 代理人	100075270
(31) 優先権主張番号	11/805,380		弁理士 小林 泰
(32) 優先日	平成19年5月23日(2007.5.23)	(74) 代理人	100101373
(33) 優先権主張国	米国 (US)		弁理士 竹内 茂雄
		(74) 代理人	100118902
			弁理士 山本 修
		(74) 代理人	100153028
			弁理士 上田 忠

最終頁に続く

(54) 【発明の名称】 検索のためのユーザ定義の関連性ランク付け

(57) 【特許請求の範囲】

【請求項 1】

ユーザ定義の関連性関数を適用する方法であって、

複合的な検索クエリを、結合で接続され、同種の論理表現を含む単純な演算を含む、複数の単純な演算子に分解するステップと、

ユーザからの構造化クエリ言語 (SQL) で記述されたユーザ定義の関連性関数を受け取るステップと、

前記単純な演算子を前記ユーザ定義の関連性関数と関連付けて、以前は前記単純な演算子に関連していたデフォルトの関連性関数をオーバーライドするステップと、

前記複合的な検索クエリと合致する合致文書を取り出すステップであって、インデックスにアクセスして、単純な演算子に対応する部分的なクエリにマッチする1又は2以上の行であって、前記部分的なクエリにマッチする文書に対応する行を生成する、ストリーミングテーブルバリュードファンクション (STVF) を使用して、取り出すステップと、

前記ユーザ定義の関連性関数を使用して前記合致文書のランクを計算するステップとを含むことを特徴とする方法。

【請求項 2】

データベース中で前記合致文書を検索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 3】

前記合致文書は、前記ユーザ定義の関連性関数によって前記ランクの計算に使用される

10

20

情報を含むことを特徴とする請求項 1 に記載の方法。

【請求項 4】

前記計算するステップは、前記ユーザ定義の関連性関数が、前記合致文書に関連する情報を使用して前記ランクを計算するステップを含むことを特徴とする請求項 1 に記載の方法。

【請求項 5】

前記分解するステップは、同種の論理表現をマルチウェイスキップマージ結合演算子に変換するステップを含むことを特徴とする請求項 1 に記載の方法。

【請求項 6】

前記複合的な検索クエリはブール論理クエリを含むことを特徴とする請求項 1 に記載の方法。 10

【請求項 7】

前記分解するステップは前記複合的な検索クエリを演算子のツリーに分解するステップを含むことを特徴とする請求項 1 に記載の方法。

【請求項 8】

ユーザからの構造化クエリ言語 (SQL) で記述されたユーザ定義のランク付け関数を受け取るステップと、

検索クエリを受け取るステップと、

前記検索クエリを、結合で接続され、同種の論理表現を含む単純な演算を含む、複数の単純な演算に分解するステップと、 20

前記複数の単純な演算それぞれを前記ユーザ定義のランク付け関数と関連付けて、以前は前記複数の単純な演算に関連していた 1 又は 2 以上のデフォルトの関連性関数をオーバーライドするステップと、

前記複数の単純な演算に対応する複数の結果を取り出すステップであって、インデックスにアクセスして、単純な演算子に対応する部分的なクエリにマッチする 1 又は 2 以上の行であって、前記部分的なクエリにマッチする文書に対応する行を生成する、ストリーミングテーブルバリュードファンクション (STVF) を使用して、取り出すステップと、

前記ユーザ定義のランク付け関数を使用して前記複数の結果それぞれのランクを計算するステップと

を実施するためのコンピュータ実行可能命令を有することを特徴とするコンピュータ可読記憶媒体。 30

【請求項 9】

複数のユーザ定義のランク付け関数を受け取るステップと、

前記複数のユーザ定義のランク付け関数を前記複数の単純な演算と関連付けるステップとをさらに含むことを特徴とする請求項 8 に記載のコンピュータ可読記憶媒体。

【請求項 10】

前記複数の結果の前記ランクを組み合わせ、前記複数の結果それぞれの全体的なランクを生み出すステップをさらに含むことを特徴とする請求項 8 に記載のコンピュータ可読記憶媒体。

【請求項 11】

前記組み合わせるステップは、前記複数の結果の前記ランクから最大ランク値を計算するステップを含むことを特徴とする請求項 10 に記載のコンピュータ可読記憶媒体。

【請求項 12】

前記分解するステップは、

前記検索クエリを検索ツリーに変換するステップと、

前記検索ツリー中の同種の演算子を識別するステップと、

前記同種の演算子をマルチウェイスキップマージ結合演算子に変換するステップとを含むことを特徴とする請求項 8 に記載のコンピュータ可読記憶媒体。

【請求項 13】

前記マルチウェイスキップマージ結合演算子はストリーミングテーブル値関数 (STV 50

F)を含むことを特徴とする請求項12に記載のコンピュータ可読記憶媒体。

【請求項14】

前記検索クエリはベクトルクエリを含むことを特徴とする請求項8に記載のコンピュータ可読記憶媒体。

【請求項15】

前記ベクトルクエリは複数のキーワードを含み、各キーワードは検索重みに関連することを特徴とする請求項14に記載のコンピュータ可読記憶媒体。

【請求項16】

バスと、
前記バスに結合された、情報を処理するためのプロセッサと、
前記バスに結合された、情報を記憶するためのメモリと、
前記バスに結合された、情報を記憶するためのデータ記憶デバイスと
を備えるシステムであって、
ユーザからの構造化クエリ言語（SQL）で記述されたユーザ定義の関連性関数を受け取り、

複合的な検索クエリを受け取り、
前記複合的な検索クエリを、結合で接続され、同種の論理表現を含む単純な演算を含む、複数のより単純な演算に解析し、
前記複数のより単純な演算それぞれを前記ユーザ定義の関連性関数と関連付けて、以前は前記複数のより単純な演算のうちの1つの単純な演算に関連していたデフォルトの関連性関数をオーバーライドし、

インデックスにアクセスして、単純な演算子に対応する部分的なクエリにマッチする1又は2以上の行であって、前記部分的なクエリにマッチする文書に対応する行を生成する、ストリーミングテーブルバリュードファンクション（STVF）を使用して、前記複数のより単純な演算に対応する複数の結果を取り出し、

前記ユーザ定義の関連性関数を使用して前記複数の結果それぞれのランクを計算するように構成されたことを特徴とするシステム。

【請求項17】

前記複合的な検索クエリを検索ツリーに還元し、前記ツリー中の同種の演算子をマルチウェイスキップマージ結合演算子に変換し、前記検索ツリーの非リーフノードを関係演算子に変換することによって、前記複合的な検索クエリを解析するように構成されたことを特徴とする請求項16に記載のシステム。

【請求項18】

前記ユーザ定義の関連性関数を前記検索ツリーの一部として最適化するようにさらに構成されたことを特徴とする請求項17に記載のシステム。

【請求項19】

前記ユーザ定義の関連性関数は、宣言言語で書かれた関数を含むことを特徴とする請求項16に記載のシステム。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、検索のためのユーザ定義の関連性ランク付けに関する。

【背景技術】

【0002】

デジタル情報の根本的な問題の1つは、所与の時点で、最も関連性のある情報を見つけるためにどのように膨大なデータを効果的にソートするかである。検索エンジンは、この課題に対処するために、関連性またはランク付け関数を備える。これらの関連性関数を使用して、検索基準を必要とする文書に異なる重みを与える。次いで、割り当てられた重みまたはランクを使用して、例えばある順序で結果を表示することによって、情報のプールをさらに操作することができる。

【0003】

多くの検索エンジン、特にデータベースとの対話に使用される検索エンジンは、 $t f - i d f$ 重み方式の何らかの変形を使用するが、この方式では、特定の文書における用語の発生頻度が、逆文書頻度（その用語が文書のプール中でどれくらいの頻度で出現するかの尺度）に対して重み付けされる。異なる検索エンジンが、この方式の異なる変形を実施することになり、個々の検索エンジンは、その特定の関連性関数を使用するように最適化される。

【発明の概要】

【発明が解決しようとする課題】

【0004】

10

本明細書では、とりわけ、検索エンジンがユーザ定義の関連性関数を利用できるようにする技術について詳述する。

【課題を解決するための手段】

【0005】

この技術の一手法では、ユーザ定義の関連性関数を適用する方法について述べる。この手法では、複合的な検索クエリが単純な演算子に分解される。単純な演算子は、ユーザ定義の関連性関数と関連付けられる。検索クエリと合致する文書が取り出され、ユーザ定義の関連性関数を使用して文書のランクが計算される。

【0006】

この技術の別の手法では、コンピュータ実行可能命令を有するコンピュータ可読媒体について述べる。この手法では、ユーザ定義のランク付け関数が受け取られ、検索クエリが受け取られる。検索クエリは、いくつかの単純な演算に分解される。これらの単純な演算はそれぞれ、関連性関数と関連付けられる。より単純な演算に対応するいくつかの結果が取り出され、関連する関連性関数を使用してこれらの各結果のランクが計算される。

20

【0007】

この技術の別の手法では、バスとプロセッサとメモリとデータ記憶デバイスとを有するシステムが、複合的な検索クエリを受け取るように構成される。システムはさらに、この複合的な検索クエリをいくつかのより単純な演算に解析するように構成される。システムはまた、これらのより単純な演算に対応する結果を取り出し、ユーザ定義の関連性関数を適用することによって各結果のランクを計算するように構成される。

30

【0008】

「課題を解決するための手段」は、以下に「発明を実施するための形態」でさらに述べる概念の精選を、単純化した形で紹介するために提供するものである。この「課題を解決するための手段」は、特許請求する主題の鍵となる特徴または必須の特徴を識別するものとはせず、また、特許請求する主題の範囲を限定するために使用されるものともしない。

【0009】

本明細書に組み込まれその一部をなす添付の図面は実施形態を示し、記述と共に、特許請求する主題の原理を説明する働きをする。

【図面の簡単な説明】

【0010】

40

【図1】実施形態を実施できる例示的なコンピューティングシステムのブロック図である。

【図2】実施形態を実施できる例示的なデータベーステーブルの図である。

【図3】一実施形態による、ユーザ定義の関連性関数を適用する方法のフローチャートである。

【図4A】一実施形態による例示的なデータベースクエリの図である。

【図4B】一実施形態による検索ツリーを描いた図である。

【図4C】一実施形態による中間検索ツリーを描いた図である。

【図4D】一実施形態によるクエリツリーを描いた図である。

【発明を実施するための形態】

50

【0011】

次に、いくつかの実施形態を詳細に参照する。代替実施形態に関して主題を述べるが、これらは、特許請求する主題をこれらの実施形態に限定するものとはしないことは理解されるであろう。反対に、特許請求する主題は、代替、修正、および均等物をカバーするものとし、これらの代替、修正、および均等物は、添付の特許請求の範囲によって定義される特許請求する主題の趣旨および範囲に含めることができる。

【0012】

さらに、以下の詳細な記述では、特許請求する主題の完全な理解を提供するために多くの具体的詳細を示す。しかし、これらの具体的詳細がなくても、あるいはその均等物を用いても実施形態を実現できることは、当業者には理解されるであろう。他の場合では、主題の態様および特徴を不必要に曖昧にしないために、周知の方法、手順、コンポーネント、および回路については詳細に述べていない。

10

【0013】

以下の詳細な記述の一部については、方法の点から提示および考察する。この方法のステップおよび順序付けを、この方法の動作を記述する本明細書中の図（例えば図3）に開示するが、このようなステップおよび順序付けは例である。実施形態は、他の様々なステップを、または本明細書中の図のフローチャートに列挙するステップの変形を、本明細書に図示および記述する以外の順序で実施するのにもよく適する。

【0014】

詳細な記述のいくつかの部分は、コンピュータメモリ上で実施できるデータビットに対する操作の、手順、ステップ、論理ブロック、処理、および他の象徴的表現の点から提示する。これらの記述および表現は、データ処理技術分野の当業者によって、その作業の実体を他の当業者に最も効果的に伝えるのに使用される手段である。手順、コンピュータ実行ステップ、論理ブロック、プロセスなどは、本明細書では、また一般には、所望の結果に至る首尾一貫した一連のステップまたは命令と考えられる。ステップは、物理的量の物理的操作を必要とするステップである。必須ではないが通常、これらの量は、コンピュータシステム中で記憶、転送、結合、比較、および他の方法によって操作されることが可能な、電気信号または磁気信号の形をとる。これらの信号をビット、値、要素、記号、文字、用語、数字などとして言及することが、主に一般使用の理由で、しばしば好都合であることがわかっている。

20

30

【0015】

しかし、これらおよび類似の用語はすべて、適切な物理的量に関連付けられるべきであって、これらの量に適用される好都合なラベルに過ぎないことに留意されたい。以下の考察から明らかなように別段の指定がない限り、全体を通して、「アクセスする」、「書く」、「含む」、「記憶する」、「伝送する」、「横断する」、「関連付ける」、「識別する」などの用語を利用した考察は、コンピュータシステムのレジスタおよびメモリ内の物理的（電子的）量として表されるデータを操作して、コンピュータシステムメモリまたはレジスタ内、あるいは他のそのような情報記憶、伝送、または表示デバイス内の物理的量として同様に表される他のデータに変換するコンピュータシステムまたは同様の電子計算デバイスの、アクションおよびプロセスに言及することを理解されたい。

40

【0016】

コンピューティングシステム環境10などのコンピューティングデバイスは、通常、少なくとも何らかの形のコンピュータ可読媒体を備える。コンピュータ可読媒体は、コンピューティングデバイスによってアクセス可能な任意の利用可能な媒体とすることができる。限定ではなく例として、コンピュータ可読媒体は、コンピュータ記憶媒体および通信媒体を含むことができる。コンピュータ記憶媒体は、コンピュータ可読命令、データ構造、プログラムモジュール、または他のデータなどの情報を記憶するための任意の方法または技術で実現された、揮発性と不揮発性、取外し可能と取外し不可能の媒体を含む。コンピュータ記憶媒体は、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）、または他の光学記憶装置、

50

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デバイス、あるいは、所望の情報を記憶するのに使用できコンピューティングデバイスによってアクセスできる他の任意の媒体を含むが、これらに限定されない。通信媒体は通常、コンピュータ可読命令、データ構造、プログラムモジュール、または他のデータを、搬送波や他のトランスポート機構などの被変調データ信号に組み入れるものであり、任意の情報送達媒体を含む。用語「被変調データ信号」は、信号中の情報を符号化するようにしてその特性の1つまたは複数が設定または変更される信号を意味する。限定ではなく例として、通信媒体は、有線ネットワークまたは直接有線接続などの有線媒体と、音響、無線周波数、赤外線などのワイヤレス媒体および他のワイヤレス媒体とを含む。以上のいずれかの組合せも、コンピュータ可読媒体の範囲に含めるべきである。

10

【0017】

いくつかの実施形態については、1つまたは複数のコンピュータまたは他のデバイスによって実行されるプログラムモジュールなどのコンピュータ実行可能命令の一般的なコンテキストで述べることができる。一般に、プログラムモジュールは、特定のタスクを実施するかまたは特定の抽象データ型を実装するルーチン、プログラム、オブジェクト、コンポーネント、データ構造などを含む。通常、プログラムモジュールの機能は、様々な実施形態で望むように結合または分散させることができる。

【0018】

基本的なコンピューティングデバイス

図1を参照すると、実施形態を実施するための例示的なシステムが、コンピューティングシステム環境10などの汎用コンピューティングシステム環境を含む。コンピューティングシステム環境10は通常、その最も基本的な構成では、少なくとも1つの処理ユニット12と、メモリ14とを備える。コンピューティングシステム環境の厳密な構成およびタイプに応じて、メモリ14は、揮発性(RAMなど)、不揮発性(ROM、フラッシュメモリなど)、またはこの2つの何らかの組合せとすることができる。この最も基本的な構成は、図1では破線16で示されている。加えて、コンピューティングシステム環境10は、追加の機構/機能を有することもできる。例えば、コンピューティングシステム環境10は、追加の記憶装置(取外し可能および/または取外し不可能)を備えることができ、これらは磁気または光学のディスクまたはテープを含むが、これらに限定されない。このような追加の記憶装置は、図1では、取外し可能記憶装置18および取外し不可能記憶装置20で示されている。コンピュータ記憶媒体は、コンピュータ可読命令、データ構造、プログラムモジュール、または他のデータなどの情報を記憶するための任意の方法または技術で実現された、揮発性と不揮発性、取外し可能と取外し不可能の媒体を含む。メモリ14、取外し可能記憶装置18、および取外し不可能記憶装置20はすべて、コンピュータ記憶媒体の例である。

20

30

【0019】

コンピューティングシステム環境10はまた、コンピューティングシステム環境10が他のデバイスと通信できるようにする通信接続22を備えることもできる。通信接続22は、通信媒体の例である。図示の例では、コンピューティングシステム環境10が通信接続22を介してデータベース30と通信するのが示されている。データベース30の性質および内容は、種々の実施形態にわたって変わる場合がある。さらに、いくつかの実施形態では、データベース30は、例えば取外し可能記憶装置18および/または取外し不可能記憶装置20内に実装された、内部データベースとすることができる。

40

【0020】

コンピューティングシステム環境10はまた、キーボード、マウス、ペン、音声入力デバイス、タッチ入力デバイスなど、入力デバイス(複数可)24を有することもできる。また、表示装置、スピーカ、プリンタなど、出力デバイス(複数可)26を備えることもできる。本明細書で論じる特定の実施形態では、タッチ入力デバイスを、表示装置、例えばタッチスクリーンと組み合わせる。これらのデバイスはすべて当技術分野で周知であり、ここで詳細に論じる必要はない。

50

【0021】

ユーザ定義の関連性関数

関連性関数は検索エンジンの必要かつ貴重な部分だが、単一の関連性関数があらゆる状況に適するわけではないかもしれない。以下の実施形態で述べるように、検索エンジンによって適用されている関連性関数をユーザが修正できる検索エンジンを有することは、価値があるであろう。さらに、いくつかの実施形態では、ユーザは、ユーザの特定の必要性をよりよく反映するよう検索エンジンを調整するために、関連性関数のいくらかまたはすべてを完全に置き換えることができる場合もある。

【0022】

したがって、以下の実施形態では、検索エンジンと共に使用するためのランク付け関数をユーザが指定できるようにする手法について述べる。いくつかの実施形態では、これは、関連性関数の基本的な構成単位、例えばマルチウェイマージ結合論理演算子(multiway merge join logical operators)を、ランク付け関数の書き手に公開することによって達成される。

【0023】

データベーステーブル

次に図2を参照すると、一実施形態により、例示的なデータベーステーブル200が示されている。具体的な列挙された特徴および要素を組み込むものとしてデータベーステーブル200を示すが、実施形態は、追加の、より少ない、または異なる特徴または要素を含む適用例にもよく適することを理解されたい。特に、2つの列だけを含む単純なテーブルとしてデータベーステーブル200を示すが、実施形態は、幅広い検索タイプおよびデータタイプにわたる適用例に適することを理解されたい。

【0024】

データベーステーブル200は、いくつかのデータベース行またはエントリ、例えば行210から260を含むものとして示されている。各データベース行は、文書キー列、例えばdockey212を含み、これは、その行に記憶された文書に対する一意の識別子を含む。各データベース行はまた、文書列、例えばdocument214を含み、これは、その特定のdockeyに対応する文書の文書テキストを含む。図示の実施形態では、複数の文書列エントリが複数の用語を含むものとして示されており、これらを使用して以下にさらに他の例を示す。

【0025】

関連性関数

種々の実施形態で、種々のタイプのクエリまたは検索が受け入れられる。例えば、一実施形態では、ブール検索、例えばAND、OR、またはAND NOTなどのブール論理演算子を使用して用語を組み合わせる検索がサポートされる。別の実施形態では、ベクトル指定、例えば用語のいくつかもまたはすべてに異なる重みを割り当てて複数の用語を検索することがサポートされる。別の実施形態、例えばウェブ検索では、返される検索文書の最終的なランク付けは、文書に関連する外部考慮事項、例えばメタデータに、部分的に依存することがある。別の実施形態では、他の検索手法が利用される。

【0026】

多くの適用例では、特定の検索エンジンに組み込まれた関連性関数にユーザが依拠するのではなく、検索エンジンによって適用される関連性関数をユーザが操作または修正することが有利である。例えば、ユーザは、異なる関連性関数を異なるシナリオまたはデータセットに適用することを選択することができ、あるいは、検索結果に適用されるランク付けを改変して何らかの望ましい基準を満たすために、ランク付け関数を修正することができる。この操作する自由は、ユーザに利点をもたらす。

【0027】

いくつかの実施形態では、ユーザは、利用可能な異なる演算子に対応する、複数のランク付け関数を供給することを選択することができる。例えば、ブール論理演算子を扱うとき、ユーザは、AND演算子とOR演算子に対して別々の関連性関数を供給することがで

10

20

30

40

50

きる。さらに、いくつかのこのような実施形態では、ユーザ提供の関連性関数によってオーバーライドされない演算子と共に使用されるように、検索エンジンによってデフォルトの関連性関数を供給することができる。

【0028】

ユーザ定義の関連性関数の適用

次に図3を参照すると、一実施形態により、ユーザ定義の関数を適用する方法のフローチャート300が示されている。フローチャート300には特定のステップを開示するが、このようなステップは例である。すなわち、実施形態は、様々な他の（追加の）ステップを、またはフローチャート300に列挙したステップの変形を実施するのにもよく適する。フローチャート300中のステップは、提示したのと異なる順序で実施してもよいこと、および、フローチャート300中のすべてのステップを実施しなくてもよいことを理解されたい。

10

【0029】

ここでステップ310を参照すると、ユーザ指定のランク付け関数が受け取られる。種々の実施形態で、このような関連性関数を供給するための種々の方法が利用される。例えば、いくつかの実施形態では、ユーザは、プログラミング言語、例えばT-SQLで書かれた関連性関数を供給することができる。いくつかのこのような実施形態では、宣言言語、例えばSQLまたはダイアレクト(dialect)が、ユーザ定義の関連性関数に使用される。また、種々の実施形態で、ランク付け関数の動作は変動する場合がある。例えば、一実施形態では、ランク付け関数は、ブール演算子ANDに適用される場合があり、TF-IDF方式を検索結果に適用する場合がある。

20

【0030】

次にステップ320を参照すると、クエリが分解される。いくつかの実施形態では、複合的なクエリが、より単純な演算子のツリーに分解される。いくつかの他の実施形態では、クエリが、いくつかのより単純な演算に分解される。例えば、一実施形態では、フルテキストクエリが、結合で接続されたいくつかの単純な演算を含むツリーに分離される。このような一実施形態では、結合のタイプは、単純な演算を接続する論理演算子に基づく。例えば、ANDは内部結合として扱われ、ORは外部結合として扱われ、AND NOTは存在しないとして対処される。いくつかの実施形態では、より単純な演算はそれぞれ、同種の論理演算子、例えばANDまたはORのみを含む。このような一実施形態では、同種の論理演算子に対するランク付け関数は可換として扱われる。

30

【0031】

次にステップ330を参照すると、単純な演算がランク付け関数と関連付けられる。いくつかの実施形態では、単純な演算は、ユーザ供給のランク付け関数と関連付けることができる。いくつかの実施形態では、単純な演算を、ユーザ供給のランク付け関数が利用可能であればそれとリンクさせることができ、あるいはデフォルトのランク付け関数とリンクさせることができる。いくつかの実施形態では、ランク付け関数は、共通のシグネチャを有し、インデックスアクセスのための言語要素を基本ツールとして使用する。このような実施形態では、コンテキスト情報がランク付け関数に提供され、ランク付け関数は、インデックスアクセスを実施する要素に渡される。さらに、コンテキスト情報を他のテーブル値関数(TVF)に渡して、ランク関数にとって重要だがインデックス中に存在しない情報を得ることができる。

40

【0032】

次にステップ340を参照すると、検索が実施され、対応する結果が取り出される。種々の実施形態で、この検索および取出しの動作は、種々の方法で実施される。実施形態は、多くの異なる検索手法と結果取出し手法の両方をういた適用例によく適することを理解されたい。いくつかの実施形態では、取り出された結果は、関連性関数が結果のランク付けを計算するのに必要な情報を含む。このような情報は、文書カウント、用語頻度、用語重み、および他の情報などの要素を含むことができる。いくつかの実施形態では、前述の分解された単純な演算を使用して、複数の検索が実施される。このような一実施形態では

50

、コンテキスト情報を提供することにより、インデックスアクセス要素は、より単純な演算に対応する部分的なクエリに合致する行を取り出すことができる。

【0033】

次にステップ350を参照すると、適切な1つまたは複数のランク付け関数に従って、取り出された結果がランク付けされる。いくつかの実施形態では、ランク付け関数は、集合体およびスカラユーザ定義関数など、標準的なSQLプログラミングコンストラクトを使用して、各文書のスコアを計算する。

【0034】

いくつかの実施形態では、クエリ分解、およびランク付け関数を単純な演算子に関連付けることは、コンパイル段階の間に実施され、検索、取出し、およびランク付けは、ランタイム段階の間に実施される。他の実施形態では、これらの動作のタイミングは変動する可能性がある。

【0035】

ストリーミングテーブル値関数

前述の方法を実施するいくつかの実施形態では、ストリーミングテーブル値関数（STVF）を使用してインデックスアクセスが提供される。このような一実施形態では、STVFの使用は、いくつかの利点をもたらす。これらの利点の中には、反転インデックスにわたるマルチウェイANDおよびORのための効率的な実施、ならびに、効率的な反転があり、文書識別子が反転インデックス中で内部列であっても、出力は、文書識別子の小さい順にソートされる。他の実施形態では、他の検索結果順序付け、例えば関連性順を利用することができる。さらに、STVFの使用は、ユーザ要求に基づいて異なるレベルで圧縮解除することを可能にし、例えば、文書識別子圧縮解除の場合はレベル1での圧縮解除、あるいは発生圧縮解除の場合はレベル2での圧縮解除を可能にする。加えて、STVFの使用は、ランク付けに必要なとされる情報、例えば文書カウント、用語頻度、用語重み、または列重みを、効率的に取り出すことを可能にする。さらに、STVFの使用は正確な濃度推定を可能にし、このことは、データベースクエリオプティマイザによる効果的な最適化を可能にする。

【0036】

クエリの分解

次に図4Aを参照すると、一実施形態により、例示的なデータベースクエリ400が示されている。クエリ400は、特定の定義された要素を組み込むものとして示されているが、実施形態は、追加の、より少ない、または異なる要素を含む適用例、あるいは異なる要素配置を特徴とする適用例にもよく適することを理解されたい。

【0037】

クエリ400は、図示の実施形態では、SQLデータベースに向けたフルテキストクエリとして示されている。この実施形態では、クエリ400は、クエリが複数のより単純な演算に還元されるようにして構築されたいくつかの変換を経て、ユーザ供給のランク付け関数の包含を可能にする。

【0038】

図示の実施形態では、Tは、テーブル200などのデータベーステーブルである。documentおよびdockeyは、データベーステーブル内の列である。検索ストリングは、図示の実施形態では「apples AND oranges OR grapes AND pears」である。CONTAINSTABLEは、検索ストリングを適格とする行、ならびにこの検索ストリングに対する文書のランクを返す、T-SQL T VFである。図示の実施形態では、このクエリは、document列がキーワードapplesとorangesの両方、またはキーワードgrapesとpearsの両方を含む、テーブルTのすべての行およびそれらのランク値を返すように、検索エンジンに命令する。

【0039】

一実施形態では、クエリは解析されバインドされる。解析およびバインドのプロセスの

間、CONTAINSTABLE関数が識別され、フルテキストクエリコンパイラが呼び出されて検索ストリングが処理される。フルテキストクエリコンパイラは、検索ストリングを、クエリキーワードとそれらをバインドする論理演算子とを表すツリーに変換する。そのようなツリーの1つであるツリー401を、図4Bに示す。

【0040】

ツリー401に示すように、クエリ400からの検索ストリングが、ツリー構造に変換されている。ツリー401は、検索ストリングによって関連付けられるツリー要素を論理演算子がバインドするように構成されている。このようにして、OR410が、AND420をAND425とバインドする。AND420は、apples431をoranges433とバインドし、AND425は、grapes435をpears437とバインドする。

10

【0041】

いくつかの実施形態では、検索ストリングが単純化されたツリー構造に還元された後、ANDおよびOR演算子など同種の論理演算子が識別される。いくつかのこのような実施形態では、単純な用語、例えばキーワードをバインドする同種の論理表現は、STVFとして実施されるマルチウェイスリップマージ結合演算子に変換される。このような一実施形態を、図4Cに中間ツリー441として示す。

【0042】

中間ツリー441に示すように、ツリー401のAND印字は、STVFとして実施されるマルチウェイスリップマージ結合演算子で置き換えられている。OR450は今や、要素460と465をバインドし、これらはそれぞれマルチウェイスリップマージ結合演算子である。

20

【0043】

いくつかの実施形態では、クエリがこのようにして変換された後、システムは、適切な演算子に対してユーザから提供されたランク付け関数をロードすることになる。例えば、中間ツリー441を参照すると、システムは、AND演算子に対してユーザから提供されたランク付け関数をロードすることになる。ランク付け関数のコンパイル中に、システムは適宜、実際のマルチウェイマージ結合演算子を挿入することになる。いくつかの実施形態では、特定の演算子に対してユーザからランク付け関数が提供されない場合は、代わりにデフォルトのシステム提供の関数がロードされコンパイルされる。

30

【0044】

次いで、ランク付け関数によって返されたランク値が、より高いレベルで、例えば図4CのOR450のレベルで組み合わせられる。いくつかの実施形態では、この組合せは、デフォルトのシステム動作によって対処される。例えば、OR演算子の場合は、組み合わせられたランクを最大値として計算し、AND演算子の場合は最小値として計算する。他の実施形態では、ユーザは、ユーザ指定のランク関数と同様にして、ユーザ指定の組合せ関数を提供することができる。

【0045】

図4Dに、一実施形態によりクエリツリー471を示す。クエリツリー471は、前述のような組合せ関数と共に、クエリ400の最終状態を示す。図示の実施形態では、フルテキストクエリツリー中の非リーフノードが、関係演算子に、例えばANDの場合は内部結合、ORの場合は外部結合、またはAND NOTの場合は半結合に変換される。図示のように、OR450が外部結合485で置き換えられている。ランク付け関数490および495が、それぞれマルチウェイスリップマージ結合演算子492および497に対してロードされている。ブロック480が組合せ関数を示すが、ここでは、ランク付け関数490および495の返却ランク値の最大値として示されている。

40

【0046】

いくつかの実施形態では、このクエリツリーは、SQLクエリ全体に組み込まれて、全体として最適化される。いくつかのこのような実施形態では、この最適化により、ユーザ定義のランク付け関数は、どんなに複雑であっても、ユーザ提供のクエリの一部であるか

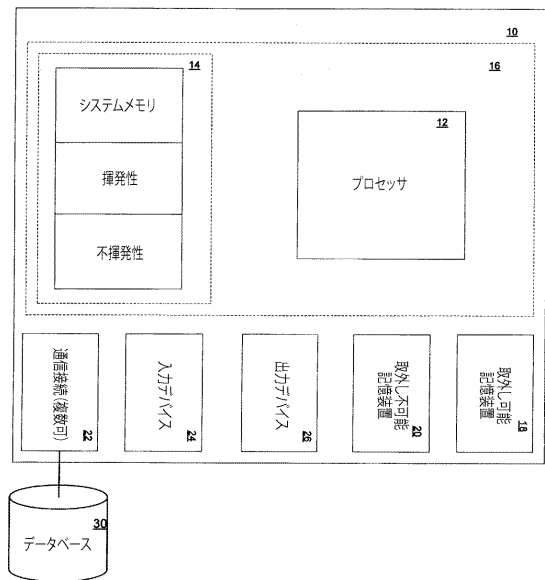
50

のような程度と同程度まで、システムによって最適化されることが確実になる。

【0047】

構造上の特徴および／または方法上の行為に特有の言葉で主題について述べたが、添付の特許請求の範囲に定義する主題は、必ずしも前述の特定の特徴または行為に限定されないことを理解されたい。そうではなく、前述の特定の特徴および行為は、特許請求の範囲を実施する例示的な形として開示する。

【図 1】

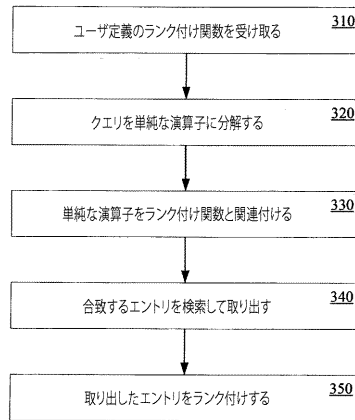


【図 2】

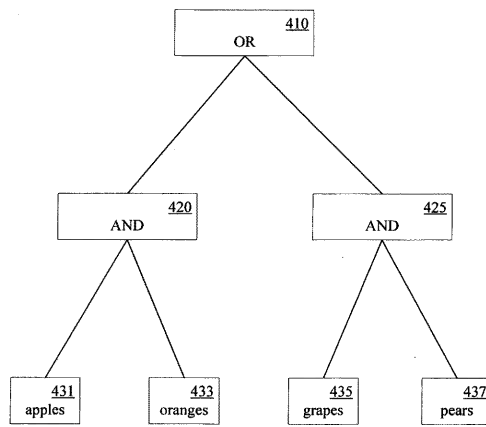
<dockey>	<document>	
212	...apples, oranges, grapes, pears....	214
222	...apples, grapes....	224
232	...oranges, grapes, pears....	234
242	...apples, oranges, pears....	244
252	...oranges, pears....	254
262	...apples, bananas, oranges, grapes, figs, pears....	264

200 データベーステーブル

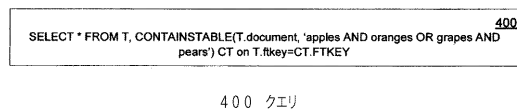
【図 3】



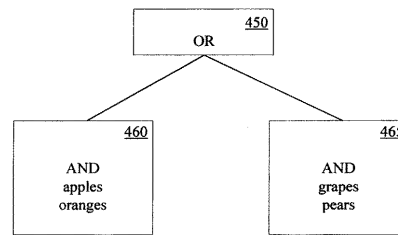
【図 4 B】



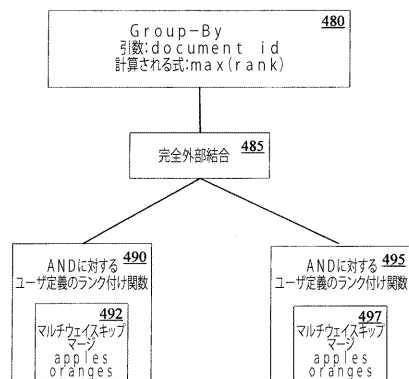
【図 4 A】



【図 4 C】



【図 4 D】



フロントページの続き

- (74)代理人 100120112
弁理士 中西 基晴
- (74)代理人 100147991
弁理士 鳥居 健一
- (74)代理人 100119781
弁理士 中村 彰吾
- (74)代理人 100162846
弁理士 大牧 綾子
- (74)代理人 100173565
弁理士 末松 亮太
- (74)代理人 100138759
弁理士 大房 直樹
- (74)代理人 100091063
弁理士 田中 英夫
- (74)代理人 100077481
弁理士 谷 義一
- (74)代理人 100088915
弁理士 阿部 和夫
- (72)発明者 デニス チュリン
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 インターナショナル パテンツ内
- (72)発明者 ジンウェイ リュー
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 インターナショナル パテンツ内
- (72)発明者 ニミシュ カノルカー
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 インターナショナル パテンツ内
- (72)発明者 スンダラム ナラヤナン
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 インターナショナル パテンツ内
- (72)発明者 タバス ナヤク
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 インターナショナル パテンツ内
- (72)発明者 フェン ティエン
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 インターナショナル パテンツ内

審査官 鈴木 和樹

- (56)参考文献 特開2005-316980 (JP, A)
特表2007-529820 (JP, A)
特開2001-273327 (JP, A)
米国特許出願公開第2005/0222975 (US, A1)
米国特許出願公開第2005/0038775 (US, A1)
米国特許第07181438 (US, B1)
米国特許第06326962 (US, B1)

(58)調査した分野(Int.Cl., DB名)

G 0 6 F 1 7 / 3 0
G 0 6 F 1 2 / 0 0