

(19) 日本国特許庁(JP)

(12) 特 許 公 報(B2)

(11) 特許番号
特許第4698737号
(P4698737)

(45) 発行日 平成23年6月8日(2011.6.8)

(24) 登録日 平成23年3月11日(2011.3.11)

(51) Int.Cl.

G O 6 F 17/30 (2006.01)

F I

G O 6 F 17/30 3 5 O C

G O 6 F 17/30 3 8 O E

G O 6 F 17/30 3 4 O B

請求項の数 6 (全 20 頁)

(21) 出願番号	特願2008-532469 (P2008-532469)	(73) 特許権者	500046438
(86) (22) 出願日	平成18年9月20日 (2006. 9. 20)		マイクロソフト コーポレーション
(65) 公表番号	特表2009-509275 (P2009-509275A)		アメリカ合衆国 ワシントン州 9805
(43) 公表日	平成21年3月5日 (2009. 3. 5)		2-6399 レッドモンド ワン マイ
(86) 国際出願番号	PCT/US2006/037206		クロソフト ウェイ
(87) 国際公開番号	W02007/035919	(74) 代理人	100077481
(87) 国際公開日	平成19年3月29日 (2007. 3. 29)		弁理士 谷 義一
審査請求日	平成21年8月18日 (2009. 8. 18)	(74) 代理人	100088915
(31) 優先権主張番号	11/231, 955		弁理士 阿部 和夫
(32) 優先日	平成17年9月21日 (2005. 9. 21)	(74) 復代理人	100115624
(33) 優先権主張国	米国 (US)		弁理士 濱中 淳宏
早期審査対象出願		(74) 復代理人	100084191
前置審査			弁理士 合田 潔
		最終頁に続く	

(54) 【発明の名称】 ドキュメント利用統計を用いたランキング関数

(57) 【特許請求の範囲】

【請求項 1】

コンピュータプログラムを記憶したコンピュータ可読記憶媒体において、
前記コンピュータプログラムは、ユーザが入力した検索クエリによりコンピュータネットワーク上で複数のドキュメントを検索した結果のランキング方法を実施させるものであり、前記方法は、
前記ネットワーク上の前記複数のドキュメントの各々に対して、ネットワーク記憶システムが生成して前記ネットワーク記憶システムが記憶する前記ドキュメントの利用値Uを考慮に入れたクエリ非依存の利用パラメータQIDを生成するステップと、
前記ドキュメントの前記利用値Uは、もし実際の利用データがあるならば所与の時間期間内における前記ドキュメントの合計閲覧数であり、実際の利用データがないならば前記ドキュメントに割り当てられるデフォルトの利用値であり、前記クエリ非依存の利用パラメータQIDは、以下の2つの式(1)、式(2)のいずれか1つで表されるものであり、
(式1)

【数1】

$$QID(\text{ドキュメント}) = w_u \frac{k_u U}{k_u + U}$$

ただし、

Uは実際の利用値またはデフォルトの利用値を表し、 w_u および k_u は、前記利

用値に対する調整パラメータを表し、又は

(式2)

$$QID(\text{ドキュメント}) = w_u [1 + \exp(-k_u U - B)] + C$$

ただし、

Uは、実際の利用値またはデフォルトの利用値を表し、 w_u 、 k_u 、B、およびCは、利用値に対する調整パラメータを表し、ユーザにより入力された検索ストリングを含む検索クエリによりネットワーク上で前記複数のドキュメントを検索して、前記複数のドキュメントの各々についての前記検索ストリングへの関連性を表す適合性スコアQDを生成するステップと、前記クエリ非依存の利用パラメータQIDと前記適合性スコアQDとの和QID + QDからなるランキング関数の値に基づいて、検索された前記複数のドキュメントをランキングして、ユーザに前記ランキング付けされた検索結果を提供するステップと、を備えることを特徴とするコンピュータ可読記憶媒体。

10

【請求項2】

前記クエリ非依存の利用パラメータQIDは、以下の式の追加のQIDコンポーネントをさらに含む

【数2】

$$QID(\text{ドキュメント}) = w_{cd} \frac{k_{cd}}{b_{cd} \frac{CD}{k_{ew}} + b_{ud} UD} \frac{1}{k_{cd} + \frac{b_{cd} + b_{ud}}{k_{ew}}}$$

20

ただし、

w_{cd} はクエリ非依存コンポーネントの重みを表し、

b_{cd} はクリック距離の重みを表し、

b_{ud} はURL深さの重みを表し、

CDはドキュメントに対して計算されたクリック距離または割り当てられたバイアスされたクリック距離を表し、

k_{ew} はエッジ重みに関係する調整定数を表し、

UDはURL深さを表す、

30

ことを特徴とする請求項1に記載のコンピュータ可読記憶媒体。

【請求項3】

ネットワーク上の複数のドキュメントのランキングをするためにコンピュータが実施する方法であって、前記方法は、

前記ネットワーク上の前記複数のドキュメントの各々に対して、ネットワーク記憶システムが生成して前記ネットワーク記憶システムが記憶する前記ドキュメントの利用値Uを考慮に入れたクエリ非依存の利用パラメータQIDを生成するステップと、

前記ドキュメントの前記利用値Uは、もし実際の利用データがあるならば所与の時間期間内における前記ドキュメントの合計閲覧数であり、実際の利用データがないならば前記ドキュメントに割り当てられるデフォルトの利用値であり、前記クエリ非依存の利用パラメータQIDは、以下の2つの式(1)、式(2)のいずれか1つで表されるものであり、

40

(式1)

【数3】

$$QID(\text{ドキュメント}) = w_u \frac{k_u U}{k_u + U}$$

ただし、

Uは実際の利用値またはデフォルトの利用値を表し、 w_u および k_u は、前記利用値に対する調整パラメータを表し、又は

(式2)

50

$QID(\text{ドキュメント}) = w_u [1 + \exp(-k_u U - B)] + C$
 ただし、

U は、実際の利用値またはデフォルトの利用値を表し、 w_u 、 k_u 、B、および C は、利用値に対する調整パラメータを表し、ユーザにより入力された検索ストリングを含む検索クエリによりネットワーク上で前記複数のドキュメントを検索して、前記複数のドキュメントの各々についての前記検索ストリングへの関連性を表す適合性スコア QD を生成するステップと、前記クエリ非依存の利用パラメータ QID と前記適合性スコア QD との和 $QID + QD$ からなるランキング関数の値に基づいて、検索された前記複数のドキュメントをランキングして、ユーザに前記ランキング付けされた検索結果を提供するステップと、を備えることを特徴とする方法。

10

【請求項 4】

前記クエリ非依存の利用パラメータ QID は、以下の式の追加の QID コンポーネントをさらに含む

【数 4】

$$QID(\text{ドキュメント}) = w_{cd} \frac{k_{cd}}{b_{cd} \frac{CD}{k_{ew}} + b_{ud} UD} \frac{1}{k_{cd} + \frac{b_{cd} + b_{ud}}{k_{ew}}}$$

20

ただし、

w_{cd} はクエリ非依存コンポーネントの重みを表し、

b_{cd} はクリック距離の重みを表し、

b_{ud} は URL 深さの重みを表し、

CD はドキュメントに対して計算されたクリック距離または割り当てられたバイアスされたクリック距離を表し、

k_{ew} はエッジ重みに関係する調整定数を表し、

UD は URL 深さを表す、

ことを特徴とする請求項 3 に記載の方法。

【請求項 5】

30

ネットワーク上の複数のドキュメントのランキングをするコンピュータ・システムであって、前記システムは、

前記ネットワーク上の前記複数のドキュメントの各々に対して、ネットワーク記憶システムが生成して前記ネットワーク記憶システムが記憶する前記ドキュメントの利用値 U を考慮に入れたクエリ非依存の利用パラメータ QID を生成する手段と、

前記ドキュメントの前記利用値 U は、もし実際の利用データがあるならば所与の時間期間内における前記ドキュメントの合計閲覧数であり、実際の利用データがないならば前記ドキュメントに割り当てられるデフォルトの利用値であり、前記クエリ非依存の利用パラメータ QID は、以下の 2 つの式 (1)、式 (2) のいずれか 1 つで表されるものであり、
 (式 1)

40

【数 5】

$$QID(\text{ドキュメント}) = w_u \frac{k_u U}{k_u + U}$$

ただし、

U は実際の利用値またはデフォルトの利用値を表し、 w_u および k_u は、前記利用値に対する調整パラメータを表し、又は

(式 2)

$$QID(\text{ドキュメント}) = w_u [1 + \exp(-k_u U - B)] + C$$

ただし、

50

U は、実際の利用値またはデフォルトの利用値を表し、 w_{ud} 、 k_{ud} 、B、および C は、利用値に対する調整パラメータを表し、ユーザにより入力された検索ストリングを含む検索クエリによりネットワーク上で前記複数のドキュメントを検索して、前記複数のドキュメントの各々についての前記検索ストリングへの関連性を表す適合性スコア QD を生成する手段と、前記クエリ非依存の利用パラメータ QID と前記適合性スコア QD との和 $QID + QD$ からなるランキング関数の値に基づいて、検索された前記複数のドキュメントをランキングして、ユーザに前記ランキング付けされた検索結果を提供する手段と、を備えることを特徴とするコンピュータ・システム。

【請求項 6】

10

前記クエリ非依存の利用パラメータ QID は、以下の式の追加の QID コンポーネントをさらに含む

【数 6】

$$QID(\text{ドキュメント}) = w_{cd} \frac{k_{cd}}{b_{cd} \frac{CD}{k_{ew}} + b_{ud} UD} \\ k_{cd} + \frac{b_{cd} + b_{ud}}{b_{cd} + b_{ud}}$$

ただし、

w_{cd} はクエリ非依存コンポーネントの重みを表し、

20

b_{cd} はクリック距離の重みを表し、

b_{ud} は URL 深さの重みを表し、

CD はドキュメントに対して計算されたクリック距離または割り当てられたバイアスされたクリック距離を表し、

k_{ew} はエッジ重みに関係する調整定数を表し、

UD は URL 深さを表す、

ことを特徴とする請求項 5 に記載のコンピュータ・システム。

【発明の詳細な説明】

【背景技術】

【0001】

30

所与の検索クエリに対するその適合性に従ってドキュメントをランク付けするランキング関数が知られている。知られたランキング関数を使用する検索エンジンによって生成される検索結果と比較して、所与の検索クエリに対する検索結果が良好であるランキング関数を開発する努力が当技術分野で続けられている。

【0002】

【特許文献 1】米国特許出願第 10 / 955983 号明細書

【特許文献 2】米国特許出願第 11 / 206286 号明細書

【特許文献 3】米国特許第 6463455 号明細書

【特許文献 4】米国特許第 6631369 号明細書

【特許文献 5】米国特許出願第 10 / 804326 号明細書

40

【発明の開示】

【課題を解決するための手段】

【0003】

本明細書では特に、ネットワーク上の所与のドキュメントに対してドキュメント適合性スコアを決定するための様々な技術が述べられている。ドキュメント適合性スコアは、1 つまたは複数のクエリ非依存コンポーネントを含むランキング関数により生成され、少なくとも 1 つのクエリ非依存コンポーネントは、ネットワーク上の 1 つまたは複数のドキュメントに対してウェブサーバ上で維持され、かつ記憶された実際のドキュメント利用データを考慮に入れる利用パラメータを含む。ランキング関数は、複数のドキュメントのドキュメント適合性スコアに基づき、複数のドキュメントを順番に（通常、降順に）ランク付

50

けするために検索エンジンにより使用され得る。

【0004】

この要約は、概して、以下の「詳細な説明」セクションで簡略化された形で述べた1つまたは複数の選択した概念に読者を導くために提供される。この要約では、特許請求される主題の重要な、かつ/または必要な特徴を特定することを意図していない。

【発明を実施するための最良の形態】

【0005】

本明細書に開示の方法およびプロセスの原理について理解を促進させるために、特有の諸実施形態の説明が行われ、またその特有の諸実施形態を述べるために特有の言語が使用される。そうではあるが、特有の言語を使用することにより開示の方法およびプロセスの範囲を限定することを意図していないことが理解されよう。論じられる開示の方法およびプロセスの原理の改変、さらなる変更、およびこのようなさらなる応用は、開示の方法およびプロセスが属する技術分野の当業者であれば普通に行うはずのものとして企図される。

【0006】

ネットワーク上のドキュメントに対するドキュメント適合性スコアを決定する方法が開示される。各ドキュメント適合性スコアは、望ましくは、1つまたは複数のクエリ非依存コンポーネント（例えば、所与の検索クエリまたは検索クエリタームに依存しない関数コンポーネント）、1つまたは複数のクエリ依存コンポーネント（例えば、所与の検索クエリまたは検索クエリタームの仕様に依存する関数コンポーネント）、またはその組合せを含むランキング関数を用いて計算される。ランキング関数により決定されたドキュメント適合性スコアは、各ドキュメント適合性スコアに従って、ネットワーク空間（例えば、企業のイントラネット空間）内のドキュメントをランク付けするために使用することができる。開示の方法を使用できる例示的な検索プロセスが、図1に例示的なプロセス10として示される。

【0007】

図1は、プロセスステップ80で開始する例示的な検索プロセス10を示しており、ユーザが検索クエリを入力する。ステップ80から、例示的な検索プロセス10はステップ200に進み、検索エンジンが、検索クエリの1つまたは複数のタームを求めてネットワーク空間内のドキュメントをすべて検索する。ステップ200から、例示的な検索プロセス10はステップ300に進み、検索エンジンのランキング関数が、1つまたは複数のクエリ非依存コンポーネント、1つまたは複数のクエリ依存コンポーネント、またはその組合せに基づいた各ドキュメントの適合性スコアに基づき、ネットワーク空間内のドキュメントをランク付けする。ステップ300から、例示的な検索プロセス10はステップ400に進み、ランク付けされた検索結果が、通常、降順でユーザに示され、検索クエリに最も適合するネットワーク空間内のドキュメントが特定される。

【0008】

以下でさらに詳細に論ずるように、ドキュメント適合性スコアを決定するいくつかの例示的な方法では、ドキュメント適合性スコアを決定するために使用されるランキング関数の少なくとも1つのクエリ非依存コンポーネントは、1人または複数のユーザによるネットワーク内の1つまたは複数のドキュメントの実際の利用に関連する「ドキュメント利用データ」、または「ドキュメント利用統計」を考慮に入れる。ドキュメント利用データおよび/または統計は、所与の検索エンジンとは別のウェブサーバ上のアプリケーションコードにより生成され、かつ記憶される。例えば、ドキュメント利用データは、ユーザがURLを要求する度にサーバが利用カウンタを更新するように、ウェブサイトにより維持することができる。利用カウンタは、先週、先月、前年、または所与のドキュメントもしくは1組のドキュメントの生存期間など、所与の時間間隔に対して取得されたドキュメント関連データを維持することができる。アプリケーションコードは、(i) 特殊なAPI (application programming interface)、もしくは(ii) ウェブサービス要求を介して、または(iii) ウェブサイト上のすべてのURL

に対して利用データを返す管理ウェブページを要求することにより、ウェブサイトから利用データを取得するために使用することができる。

【0009】

特定のウェブサイトは、ネットワーク空間内で利用データを生成し、かつ維持するように、ならびにローカルまたは遠隔の記憶システムに利用データを記憶するように使用することができる。ネットワーク空間内のドキュメントの利用データを生成し、維持し、かつ記憶するために適切なウェブサイトは、これだけに限らないが、W I N D O W S（登録商標）S H A R E P O I N T（登録商標）S e r v i c e s サイトを含む。

【0010】

ドキュメント適合性スコアを決定する開示の方法はさらに、1つまたは複数の追加のクエリ非依存コンポーネントを含むランキング関数を使用することができる。適切な追加のクエリ非依存コンポーネントは、これだけに限らないが、2004年8月30日に出願された、「SYSTEM AND METHOD FOR RANKING SEARCH RESULTS USING CLICK DISTANCE」と題する特許文献1に記載のネットワーク空間内の各ドキュメントに対するクリック距離を考慮に入れるクエリ非依存コンポーネント、2005年8月15日に出願された、「RANKING FUNCTIONS USING A BIASED CLICK DISTANCE OF A DOCUMENT ON A NETWORK」と題する特許文献2に記載のネットワーク空間内の各ドキュメントに対してバイアスされたクリック距離を考慮に入れるクエリ非依存コンポーネント、およびネットワーク空間内の各ドキュメントに対するURLを考慮に入れるクエリ非依存コンポーネントを含む。本特許出願の譲受人に譲渡された前述の米国特許出願のそれぞれの主題を、参照により本明細書にその全体を組み込む。

【0011】

さらに他の例示的な実施形態では、ドキュメント適合性スコアを決定する開示の方法は、上記で述べたドキュメント利用パラメータと、1つまたは複数の上記で述べた追加のクエリ非依存コンポーネントとを共に含む少なくとも1つのクエリ非依存コンポーネントを含むランキング関数を使用する。

【0012】

ドキュメント適合性スコアは、ネットワーク空間内のドキュメントをランク付けするために使用され得る。例えば、ネットワーク上のドキュメントをランク付けする方法は、上記で述べた方法を用いて、ネットワーク上の各ドキュメントに対するドキュメント適合性スコアを決定するステップと、各ドキュメントのドキュメント適合性スコアに基づき、所望の順番で（通常、降順で）ドキュメントをランク付けするステップとを含むことができる。

【0013】

ドキュメント適合性スコアはまた、検索クエリの検索結果をランク付けするために使用することができる。例えば、検索クエリの検索結果をランク付けする方法は、上記で述べた方法を用いて、検索クエリの検索結果における各ドキュメントに対するドキュメント適合性スコアを決定するステップと、各ドキュメントのドキュメント適合性スコアに基づいて、所望の順番で（通常、降順で）ドキュメントをランク付けするステップとを含むことができる。

【0014】

本明細書で開示された方法を使用するアプリケーションプログラムは、様々なハードウェアコンポーネントを備える様々なコンピュータシステム上にロードされ、実行され得る。本明細書で開示された方法を実施するための例示的なコンピュータシステムおよび例示的な動作環境を以下で述べる。

【0015】

例示的な動作環境

図2は、本明細書に開示の方法が実装され得る適切なコンピューティングシステム環境100の例を示す。コンピューティングシステム環境100は、適切なコンピューティング環境の一例にすぎず、本明細書に開示された方法の利用または機能の範囲に関して何ら

10

20

30

40

50

かの限定を示唆するものではない。コンピューティング環境 100 は、例示の動作環境 100 で示されたコンポーネントの任意の 1 つまたは組合せに関して何らかの依存性または要件を有するものと解釈すべきではない。

【0016】

本明細書に開示の方法は、数多くの他の汎用または専用のコンピューティングシステム環境もしくはコンフィギュレーションで動作可能である。本明細書に開示の方法と共に使用するのが適切であり得る周知のコンピューティングシステム、環境、および/またはコンフィギュレーションの諸例は、これだけに限らないが、パーソナルコンピュータ、サーバコンピュータ、ハンドヘルドもしくはラップトップ装置、マルチプロセッサシステム、マイクロプロセッサベースのシステム、セットトップボックス、プログラム可能な家庭用電子機器、ネットワーク PC、ミニコンピュータ、メインフレームコンピュータ、上記システムもしくは装置のうちの任意のものを含む分散コンピューティング環境などを含む。

【0017】

本明細書に開示の方法およびプロセスは、コンピュータ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記述することができる。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、または特定の抽象データ型を実施するルーチン、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。本明細書で開示の方法およびプロセスはまた、タスクが、通信ネットワークを介してリンクされた遠隔の処理装置により実施される分散コンピューティング環境中で実施することもできる。分散コンピューティング環境では、プログラムモジュールを、メモリ記憶装置を含むローカルおよび遠隔のコンピュータ記憶媒体中に配置することができる。

【0018】

図 2 を参照すると、本明細書で開示の方法およびプロセスを実施するための例示的なシステムは、コンピュータ 110 の形の汎用コンピューティング装置を含む。コンピュータ 110 のコンポーネントは、これだけに限らないが、処理装置 120、システムメモリ 130、およびこれだけに限らないが、システムメモリ 130 から処理装置 120 を含む様々なシステムコンポーネントを結合するシステムバス 121 を含むことができる。システムバス 121 は、メモリバスもしくはメモリコントローラ、周辺バス、および任意の様々なバスアーキテクチャを使用するローカルバスを含むバス構造のいくつかのタイプのうちの任意のものとしてすることができる。例のためであり、これだけに限らないが、このようなアーキテクチャは、ISA (Industry Standard Architecture) バス、MCA (Micro Channel Architecture) バス、EISA (Enhanced ISA) バス、VESA (Video Electronics Standards Association) ローカルバス、およびメザニン (Mezzanine) バスとしても知られる PCI (Peripheral Component Interconnect: 周辺コンポーネント相互接続) バスを含む。

【0019】

コンピュータ 110 は、通常、様々なコンピュータ可読媒体を含む。コンピュータ可読媒体は、コンピュータ 110 によりアクセスされ得る利用可能な任意の媒体とすることができる。揮発性および不揮発性媒体、取外し可能および取外し不能な媒体を共に含む。例のためであり、限定するものではないが、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施される揮発性および不揮発性、取外し可能および取外し不能媒体を含む。コンピュータ記憶媒体は、これだけに限らないが、RAM、ROM、EEPROM、フラッシュメモリもしくは他のメモリ技術、CD-ROM、DVD (digital versatile disks: 多用途デジタルディスク)、もしくは他の光ディスクストレージ、磁気カセット、磁気テープ、磁気ディスクストレージもしくは他の磁気記憶装置、あるいは所望の情報を記憶するのに使用することができ、コンピュータ 110 に

よりアクセスできる任意の他の媒体を含む。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、搬送波もしくは他の移送機構などの変調されたデータ信号で具現化し、また任意の情報送達媒体を含む。用語「変調されたデータ信号」とは、1つまたは複数のその特性を、信号中に情報を符合化するように設定もしくは変更させた信号を意味する。例示のためであり、限定するものではないが、通信媒体は、有線ネットワークや直接有線接続などの有線媒体、および音響、RF、赤外線、他の無線媒体などの無線媒体を含む。上記の任意のものの組合せもまた、本明細書で使用されるコンピュータ可読媒体の範囲内に含まれるべきである。

【0020】

システムメモリ130は、ROM（読取り専用メモリ）131、およびRAM（ランダムアクセスメモリ）132などの揮発性および/または不揮発性メモリの形のコンピュータ記憶媒体を含む。起動中などに、コンピュータ110内のエレメント間で情報を転送するのを助ける基本ルーチンを含むBIOS（基本入出力システム）133は、通常、ROM131に記憶される。RAM132は、通常、処理装置120が直ちにアクセス可能であり、かつ/または現在処理されているデータおよび/またはプログラムモジュールを含む。例のためであり、限定するものではないが、図2は、オペレーティングシステム134、アプリケーションプログラム135、他のプログラムモジュール136、およびプログラムデータ137を示す。

【0021】

コンピュータ110はまた、他の取外し可能/取外し不能、揮発性/不揮発性コンピュータ記憶媒体を含むことができる。例示のためだけであるが、図2は、取外し不能な、不揮発性の磁気媒体から読み取りもしくはそれに書き込むハードディスクドライブ140と、取外し可能な、不揮発性の磁気ディスク152から読み取りもしくはそれに書き込む磁気ディスクドライブ151と、CD-ROMもしくは他の光媒体など、取外し可能な、不揮発性の光ディスク156から読み取りもしくはそれに書き込む光ディスクドライブ155とを示す。例示的な動作環境で利用できる他の取外し可能/取外し不能な、揮発性/不揮発性のコンピュータ記憶媒体は、これだけに限らないが、磁気テープカセット、フラッシュメモリカード、多用途デジタルディスク、デジタルビデオテープ、固体RAM、固体ROMなどを含む。ハードディスクドライブ141は、通常、インターフェース140などの取外し不能なメモリインターフェースを介してシステムバス121に接続され、また磁気ディスクドライブ151および光ディスクドライブ155は、通常、インターフェース150などの取外し可能なメモリインターフェースにより、システムバス121に接続される。

【0022】

上記で論じた、また図2で示したドライブおよびその関連す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、およびコンピュータ110に対する他のデータの記憶を提供する。図2では、例えば、ハードディスクドライブ141は、オペレーティングシステム144、アプリケーションプログラム145、他のプログラムモジュール146、およびプログラムデータ147を記憶するものとして示されている。これらのコンポーネントは、オペレーティングシステム134、アプリケーションプログラム135、他のプログラムモジュール136、およびプログラムデータ137と同一にすることも、異なるものとする 것도可能であることに留意されたい。オペレーティングシステム144、アプリケーションプログラム145、他のプログラムモジュール146、およびプログラムデータ147は、少なくとも、それらが異なる複製であることを示すために、ここでは、異なる番号を与えている。

【0023】

ユーザは、キーボード162、および一般にマウス、トラックボール、もしくはタッチパッドと呼ばれる指示装置161などの入力装置を介して、コンピュータ110にコマンドおよび情報を入力することができる。他の入力装置（図示せず）は、マイクロフォン、ジョイスティック、ゲームパッド、サテライトディッシュ、スキャナなどを含むことがで

10

20

30

40

50

きる。これらのまた他の入力装置は、システムバス 121 に結合されたユーザ入力インターフェース 160 を介して、処理装置 120 に接続されることが多いが、パラレルポート、ゲームポート、または USB (universal serial bus) など、他のインターフェースおよびバス構造により接続することもできる。モニタ 191 または他のタイプの表示装置がまた、ビデオインターフェース 190 などのインターフェースを介して、システムバス 121 に接続される。モニタ 191 に加えて、コンピュータ 110 はまた、出力周辺インターフェース 195 を介して接続され得るスピーカ 197 およびプリンタ 196 など、他の周辺出力装置を含むこともできる。

【0024】

コンピュータ 110 は、遠隔コンピュータ 180 など、1 つまたは複数の遠隔コンピュータへの論理的な接続を用いて、ネットワーク化された環境中で動作することができる。遠隔コンピュータ 180 は、パーソナルコンピュータ、サーバ、ルータ、ネットワーク PC、同位 (peer) 装置、または他の共通のネットワークノードとすることができ、また、メモリ記憶装置 181 だけが図 2 に示されているが、通常、コンピュータ 110 に対して上記で述べたエレメントの多くのもの、またはすべてを含む。図 2 に示す論理的な接続は、LAN (ローカルエリアネットワーク) 171、および WAN (広域ネットワーク) 173 を含むが、他のネットワークを含むこともできる。このようなネットワーク化環境は、オフィス、企業規模のコンピュータネットワーク、イントラネット、およびインターネットにおいて普通のものである。

【0025】

LAN ネットワーク化環境で使用される場合、コンピュータ 110 は、ネットワークインターフェースまたはアダプタ 170 を介して、LAN 171 に接続される。WAN ネットワーク化環境で使用される場合、コンピュータ 110 は、通常、インターネットなどの WAN 173 を介して通信を確立するためのモデム 172 または他の手段を含む。モデム 172 は、内部または外部のものとして行うことができるが、ユーザ入力インターフェース 160、または他の適切な機構を介してシステムバス 121 に接続することができる。ネットワーク化された環境では、コンピュータ 110、またはその部分に関して示されたプログラムモジュールは、遠隔のメモリ記憶装置に記憶することができる。例のためであり、限定するものではないが、図 2 は、記憶装置 181 上に常駐する遠隔のアプリケーションプログラム 185 を示す。図示されたネットワーク接続は例示的なものであり、コンピュータ間で通信リンクを確立する他の手段を使用することも理解されよう。

【0026】

本明細書に開示の方法およびプロセスは、これだけに限らないが、サーバシステムソフトウェアアプリケーション (例えば、WINDOWS (登録商標) SERVER SYSTEM (商標) ソフトウェアアプリケーション)、検索ランキングアプリケーション、およびネットワーク空間内のドキュメントの利用データを生成し、維持し、かつ記憶するためのアプリケーション (例えば、WINDOWS (登録商標) SHAREPOINT (登録商標) Services アプリケーション) を含む 1 つまたは複数のアプリケーションプログラムを用いて実施することができ、そのアプリケーションプログラムのうちの任意の 1 つが、例示のシステム 100 におけるアプリケーションプログラム 135、アプリケーションプログラム 145、および遠隔のアプリケーションプログラム 185 として指定された数多くのアプリケーションプログラムのうちの 1 つとすることができる。

【0027】

上記で述べたように、所与のドキュメントに対するドキュメント適合性スコアを生成する開示の方法は、ハンドヘルド装置、マルチプロセッサシステム、マイクロプロセッサベースの、もしくはプログラム可能な家庭用電子機器、ネットワーク化されたパーソナルコンピュータ、ミニコンピュータ、メインフレームコンピュータなどを含む他のコンピュータシステムコンフィギュレーションで実装できることが当業者であれば理解されよう。所与のドキュメントに対するドキュメント適合性スコアを生成する開示の方法はまた、タスクが、通信ネットワークを介してリンクされた遠隔の処理装置により実施される分散コン

ピューティング環境で実行することもできる。分散コンピューティング環境では、プログラムモジュールを、ローカルおよび遠隔のメモリ記憶装置に配置することができる。

【0028】

例示的な諸実施形態の実装

上記で論じたように、ネットワーク上のドキュメントに対するドキュメント適合性スコアを決定する方法が提供される。開示された方法は、ネットワーク上の各ドキュメントのドキュメント利用値を考慮に入れるランキング関数を使用して、ネットワーク上のドキュメントをランク付けすることができる。

【0029】

ネットワーク上のドキュメントに対するドキュメント適合性スコアを決定する開示の方法は、いくつかのステップを含むことができる。例示的な一実施形態では、ネットワーク上のドキュメントに対するドキュメント適合性スコアを決定する方法は、サーバで維持され、かつ記憶された実際の利用データに基づいた実際の利用値 (U_A) を、 N 個のドキュメントを含むネットワーク上の 1 つまたは複数のドキュメントに割り当てるステップと、 N 個未満のドキュメントに実際の利用値 (U_A) が割り当てられる場合、関連する実際の利用データを有していないドキュメントに対して、デフォルトの利用値 (U_D) を割り当てるステップと、各ドキュメントに対して利用値 (すなわち、 U_A または U_D) を使用し、ネットワーク上の所与のドキュメントのドキュメント適合性スコアを決定するステップとを含む。

【0030】

本明細書で使用される場合、用語「実際の利用データ」は、1 人または複数のユーザによるドキュメントの「利用」に関連付けられたデータの 1 つまたは複数のタイプを表す。所与のドキュメントまたは 1 組のドキュメントに対する実際の利用データのタイプは、これだけに限らないが、所与の時間期間内のユーザすべてによるドキュメント閲覧数、所与の時間期間内のユーザごとの平均ドキュメント閲覧数、所与の時間期間内で特定のドキュメントに費やされた合計時間、所与の時間期間内で特定のドキュメントに費やされた平均時間などを含むことができる。所与の時間期間は、例えば、先週、先月、前年、そのドキュメントの生存期間、または他の任意の所望の時間期間とすることができる。

【0031】

ネットワーク空間内のドキュメントに対するドキュメント利用データまたは統計を生成し、維持し、記憶する諸ステップは、コンピューティングシステム上で一般に見出されるアプリケーションコードにより実施することができる。ドキュメント利用データは、所与の検索クエリまたは検索エンジンとは独立して生成され、維持され、かつ記憶され、通常、そのドキュメント (またはページ) を維持し、そのドキュメント (またはページ) をユーザに利用可能にするサーバ上のアプリケーションコードにより生成され、維持され、かつ記憶される。ドキュメント利用データまたは統計を生成し、維持し、かつ記憶するために適切なアプリケーションプログラムは、これだけに限らないが、WINDOWS (登録商標) SHAREPOINT (登録商標) Services、および他の同様のアプリケーションプログラムを含む。

【0032】

これらのサービスサイトならびに同様の機能を実施する他のウェブサイトで、記憶され維持されているドキュメント利用データは、上記で論じたアプリケーションコードを用いてアクセスすることができる。例えば、ドキュメント利用データは、所与のウェブサイト (例えば、WINDOWS (登録商標) SHAREPOINT (登録商標) Services サイト) から、(i) 特殊な API (application programming interface)、もしくは (ii) ウェブサービス要求を介して、または (iii) ウェブサイト上のすべての URL に対して利用データを返す管理ウェブページを要求することにより、アクセスすることができる。

【0033】

ネットワーク上のドキュメントに対するドキュメント適合性スコアを決定する開示され

10

20

30

40

50

た方法は、これだけに限らないが、実際のドキュメント利用を求めて、ネットワーク空間内の1つまたは複数のドキュメントをモニタするステップと、ローカルまたは遠隔のデータ記憶ファイルに、1つまたは複数のドキュメントに対する実際のドキュメント利用データを記憶するステップと、ドキュメントまたはそのドキュメントを含むフォルダに対する実際の利用データに基づいて、ドキュメントに対する実際の利用値 (U_A) を計算するステップと、1つまたは複数のドキュメントに対する実際の利用値 (U_A) を、ローカルまたは遠隔のデータ記憶ファイルに記憶するステップと、ローカルまたは遠隔のデータ記憶ファイルから記憶されたドキュメント利用データまたは実際の利用値 (U_A) を要求する (例えば、ユーザによる特定の検索クエリの後、検索エンジンからこのようなデータを要求する) ステップと、ローカルまたは遠隔のデータ記憶ファイルから、1つまたは複数のドキュメントに対する実際のドキュメント利用データまたは実際の利用値 (U_A) を取り出すステップと、任意選択で、ドキュメント利用値 (すなわち、実際のまたはデフォルトの) を、1つまたは複数のさらなるドキュメントプロパティとマージして、ドキュメントに対するドキュメント適合性スコアを決定するステップとを含む、いくつかの追加のステップを含むことができる。

10

【0034】

図3は、ネットワーク上のドキュメントに対して、実際のまたはデフォルトの利用値を提供し、その後、システム管理者による任意選択の格下げ/格上げ手順が行われる例示の方法の例示的な諸ステップを示す論理的な流れ図を表す。図3に示すように、例示の方法401は、ブロック402で開始し、ステップ403に進む。ステップ403では、ネットワーク上の第1のドキュメントが、実際の利用データを求めてクロール (crawl) される。

20

【0035】

実際の利用データを求めて第1のドキュメントをクロールするステップ (ステップ403) は、第1のドキュメントが、それに関連する何らかの実際の利用データを有するかどうかを判定できるクローラアプリケーションを用いて実施することができ、第1のドキュメントが、それに関連する実際の利用データを有する場合、その実際の利用データを取り出す。ネットワーク上のドキュメントに対する実際のまたはデフォルトの利用値を提供する開示の方法で使用するのに適したクローラアプリケーションは、これだけに限らないが、特許文献3および4に記載のクローラアプリケーションを含み、その両方の特許の主題を、参照により本明細書にその全体を組み込む。

30

【0036】

上記で論じたように、実際の利用データは、ネットワーク上の1つまたは複数のドキュメントに対する実際の利用データを記憶する1つまたは複数のファイルから取得され得る。実際の利用データは、ドキュメントのコンポーネントとして、そのドキュメントと共に記憶することができ、または実際のドキュメントから分離されたデータ記憶ファイルに記憶することもできる。適切な遠隔の記憶システムは、これだけに限らないが、Microsoft Corporation (マイクロソフト社) (ワシントン州レッドモンド) から市販のWINDOWS (登録商標) SHAREPOINT (登録商標) Services (WSS) 製品、ならびに他の同様な任意の遠隔記憶システムを含む。例えば、WSS遠隔記憶システムは、例えば、すべてのユーザにわたる所与のネットワーク上のすべてのドキュメントに対する要求数を含む実際の利用データを記録し、先週、先月、前年、またはそのドキュメントの全体の生存期間、あるいは他の任意の時間期間における、ドキュメントごとのクリック数の統計を生成する。さらに、上記で述べたように、本明細書で開示の方法は、WSS遠隔記憶システムに限定されるのではなく、WSS遠隔記憶システム、または開示の方法における他の同様な任意のドキュメントデータシステムを使用できることを理解されたい。

40

【0037】

ドキュメントがクロールされた後、例示の方法401は判断ブロック404に進む。判断ブロック404で、ドキュメントが、それに関連する実際の利用データを有するかどう

50

かに関して、判定がアプリケーションコードにより行われる。ドキュメントがそれに関連する実際の利用データを有すると判断された場合、例示の方法 401 はステップ 405 に進み、実際の利用に基づいた利用値 (U_A) がドキュメントに割り当てられる。実際の利用値 (U_A) は、ドキュメントに関連付けられた実際の利用データの 1 つまたは複数のコンポーネントを用いて決定することができる。例えば、いくつかの実施形態では、ドキュメントに割り当てられる実際の利用値 (U_A) は、ドキュメントを閲覧するユーザの数に対してだけ関連付けることができる。他の実施形態では、ドキュメントに割り当てられる実際の利用値 (U_A) は、所与の時間期間内のユーザすべてによるドキュメント閲覧数、所与の時間期間内のユーザごとの平均ドキュメント閲覧数、所与の時間期間内に特定のドキュメントに対して費やされた合計時間、所与の時間期間内に特定のドキュメントに対して費やされた平均時間、または上記基準の任意のものの組合せに関連付けることができるが、その所与の時間期間には、先週、先月、前年、ドキュメントの生存期間、または他の所望の任意の時間期間が含まれる。

【0038】

所与のドキュメントに関連付けられた実際の利用データが、所与の時間期間中にドキュメントが使用されていない、または閲覧されていなかったことを示唆する場合がある。このような場合、ドキュメントは、ゼロに等しい利用値 (U_A) が割り当てられて、その時間期間中に利用がなかったことを示すこともできるが、通常、実際の使用に基づく、または実際には使用されていないことに基づく利用値 (U_A) は、ゼロ以外の数が割り当てられる。

【0039】

さらに、いくつかの場合では、実際の利用データは、個々のドキュメントとは反対に、1 組のドキュメントと関連付けられ得る。例えば、フォルダは、1 組のドキュメントを含むことができ、また関連するサーバは、フォルダ内の個々のドキュメントではなく、フォルダ (の使用) にアクセスすることに関係する利用データを追跡するだけであり得る。この実施形態では、フォルダに関連付けられた実際の利用データがある場合、フォルダの実際の利用データに基づいて、利用値 (U_A) がフォルダ内の各ドキュメントに対して提供され得る。通常、各利用値 (U_A) は、フォルダ内の各ドキュメントで同じとなるが、しかし、もしそれが望ましい場合、フォルダ内の異なるドキュメントに対して、異なる利用値 (U_A) を割り当てることができる。

【0040】

ステップ 405 から、例示の方法 401 は、以下で述べる判断ブロック 406 に進む。

【0041】

判断ブロック 404 に戻ると、ドキュメントが、それに関連する実際の利用データを有していないという判断が行われた場合、例示の方法 401 はステップ 407 に進み、デフォルトの利用値 (U_D) がドキュメントに割り当てられる。例えば、ドキュメント利用データを維持していないウェブサイトの一部であるドキュメントに対して、デフォルトの利用値 (U_D) を割り当てることができる。ドキュメントに割り当てられたデフォルトの利用値 (U_D) は、実際の利用データを有するドキュメントに対して、ドキュメントに初期の重要度を提供するために使用することができる。例えば、所与のドキュメントに対するより高い利用値が、ネットワーク内におけるドキュメントの相対的な重要度を示す場合、ドキュメントにより低いデフォルトの利用値 (U_D) を割り当てるとは、ネットワーク上の他のドキュメントに対して、そのドキュメントの重要度を格下げすることである。

【0042】

所与のドキュメントに対するより高い利用値が、ネットワーク内のドキュメントの相対的な重要度を示す例示的な一実施形態では、ドキュメントに割り当てられるデフォルトの利用値 (U_D) は、ネットワーク上の他のドキュメントに割り当てられた実際の利用値 (U_A) に対して相対的なものであり得る。例えば、ドキュメントの相対的な重要度を低くするために、上記で述べたように、ネットワーク上の他のドキュメントに割り当てられたものの実際の利用値 (U_A) にも満たないデフォルトの利用値 (U_D) をそのドキュメント

に割り当てることができる。ドキュメントの相対的な重要度を高めることが望ましい場合、デフォルトの利用値 (U_D) をドキュメントに割り当てることができるが、その場合、そのデフォルトの利用値 (U_D) は、ネットワーク上の他のドキュメントに割り当てられたものの実際の利用値 (U_A) よりも大きく、あるいはネットワーク上の他のドキュメントのいくつかに割り当てられた実際の利用値 (U_A) のいくつかよりも大きくする。

【0043】

他の諸実施形態では、割り当てられた実際の利用値 (U_A) を有するドキュメントと比較して、平均的な相対重要度がドキュメントに与えられるように、実際の利用データの無いドキュメントに対して、デフォルトの利用値 (U_D) を割り当てることができる。例えば、この実施形態では、実際の利用データを有しないドキュメントに対するデフォルトの利用値 (U_D) は、最小に割り当てられた実際の利用値 ($U_{A_{min}}$) から、最大に割り当てられた実際の利用値 ($U_{A_{max}}$) の範囲とすることができ、あるいは最小に割り当てられた実際の利用値 ($U_{A_{min}}$) と、最大に割り当てられた実際の利用値 ($U_{A_{max}}$) との間の特定の範囲内とすることもできる。この実施形態では、実際の利用データの無いドキュメントは、それに関連する実際のデータを有するドキュメントと比較して、中程度の利用を示唆する、平均的な相対重要度を備える。

【0044】

ステップ407から、例示の方法401は判断ブロック406に進む。判断ブロック406で、ネットワーク上のドキュメントのすべてが、実際の (U_A)、またはデフォルトの (U_D) 利用値を有するかどうかにに関して、アプリケーションコードにより判定が行われる。ネットワーク上のドキュメントのすべてが、実際の (U_A)、またはデフォルトの (U_D) 利用値を有していないという判断が行われた場合、例示の方法401はステップ408に進み、実際の利用データを求めて次のドキュメントがクロールされる。ステップ408から、例示の方法401は判断ブロック404に戻り、上記で論じたように進む。

【0045】

判断ブロック406に戻ると、ネットワーク上のドキュメントのすべてが、実際の (U_A)、またはデフォルトの (U_D) 利用値を有しているとアプリケーションコードにより判定が行われた場合、例示の方法401は判断ブロック409に進む。判断ブロック409で、ネットワーク空間内の所与のドキュメントの重要度をさらに正確に表すために、任意の実際の (U_A)、またはデフォルトの (U_D) 利用値を格下げすべきかどうか、システム管理者によって判定が行われる。ネットワーク空間内の1つまたは複数のドキュメントの重要度をさらに正確に表すために、1つまたは複数の実際の (U_A)、またはデフォルトの (U_D) 利用値を格下げするという判断が行われた場合、例示の方法401はステップ410に進み、1つまたは複数のドキュメント (またはURL) の実際の (U_A)、またはデフォルトの (U_D) 利用値が、負または正に調整される。ステップ410から、例示の方法401は、以下で述べるステップ411に進む。

【0046】

判断ブロック409に戻ると、1つまたは複数の実際の (U_A)、またはデフォルトの (U_D) 利用値を格下げ (または格上げ) しないという判断が行われた場合、例示の方法401は直接ステップ411に進む。ステップ411では、ネットワーク空間内の各ドキュメントに対する全体のドキュメント適合性スコアを決定するために、実際の (U_A)、およびデフォルトの (U_D) 利用値がランキング関数中で使用される。ステップ411から、例示の方法401は終了ブロック412に進む。

【0047】

すべての実際の (U_A)、またはデフォルトの (U_D) 利用値が決定され、かつ任意選択で、そうすることが望ましい場合、格下げ (または任意選択で格上げ) された後、各ドキュメントに対する実際の (U_A)、またはデフォルトの (U_D) 利用値がランキング関数中でパラメータとして使用され、各ドキュメントに対するドキュメント適合性スコアを提供することができる。このようなドキュメント適合性スコアは、検索クエリの検索結果をランク付けするために使用することができる。ドキュメント利用値パラメータを含むラ

10

20

30

40

50

ンキング関数を用いて、生成された検索結果をランク付けする例示の方法が図 4 に示される。

【 0 0 4 8 】

図 4 は、例示の方法 2 0 における例示的な諸ステップを示す論理的な流れ図を提供し、例示の方法 2 0 は、利用値パラメータを含むランキング関数を用いて、生成された検索結果をランク付けする方法を含む。図 4 に示すように、例示の方法 2 0 は、ブロック 2 0 1 で開始し、ステップ 2 0 2 に進む。ステップ 2 0 2 では、ユーザは、検索クエリを入力することにより検索を要求する。ステップ 2 0 2 の前に、ネットワーク上のドキュメントのそれぞれに対する実際のまたはデフォルトの利用値が、先行して計算されている。ステップ 2 0 2 から、例示の方法 2 0 は、ステップ 2 0 3 に進む。

10

【 0 0 4 9 】

ステップ 2 0 3 では、ネットワーク上の各ドキュメントに対する実際のまたはデフォルトの利用値が、インデックスに記憶された各ドキュメントに対する他の任意のドキュメント統計（例えば、他のクエリ非依存統計）とマージされる。他のドキュメント統計と、実際のまたはデフォルトの利用値とをマージすることにより、ランキングに関係するすべての情報が一緒に集められるので、クエリ応答時間をより速くすることができる。したがって、インデックス中にリストされた各ドキュメントは、マージされた後、関連する実際の、またはデフォルトの利用値を有する。マージが完了した後、例示の方法 2 0 は、ステップ 2 0 4 に進む。

【 0 0 5 0 】

20

ステップ 2 0 4 で、利用パラメータを含む所与のドキュメントに対するクエリ非依存ドキュメント統計が、ランキング関数のコンポーネントとして提供される。クエリ依存データもまた、通常、ランキング関数の別のコンポーネントとして所与のドキュメントに対して提供される。クエリ依存データまたはランキング関数のコンテンツ関連部分は、実際の検索タームおよび所与のドキュメントのコンテンツに依存する。

【 0 0 5 1 】

一実施形態では、ランキング関数は、利用パラメータを含む少なくとも 1 つのクエリ非依存（Q I D）コンポーネントを含む。一実施形態では、クエリ非依存（Q I D）コンポーネントは、以下の式により表すことができる。

【 0 0 5 2 】

30

【数 1】

$$QID(\text{ドキュメント}) = w_u \frac{k_u U}{k_u + U} \quad (1)$$

【 0 0 5 3 】

ただし、

U は、実際の利用値またはデフォルトの利用値を表し、また、

w_u および k_u は、利用値に対する調整パラメータを表す。他の実施形態では、クエリ非依存（Q I D）コンポーネントは、以下の式により表すことができる。

【 0 0 5 4 】

40

$$QID(\text{ドキュメント}) = w_u U + k_u \quad (2)$$

ただし、

U は、実際の利用値またはデフォルトの利用値を表し、また、

w_u および k_u は、利用値に対する調整パラメータを表す。さらに他の実施形態では、クエリ非依存（Q I D）コンポーネントは、以下の式により表すことができる。

【 0 0 5 5 】

$$QID(\text{ドキュメント}) = w_u [1 + \exp(-k_u U - B)] + C \quad (3)$$

ただし、

U は、実際の利用値またはデフォルトの利用値を表し、また、

w_u 、 k_u 、B、および C は、利用値に対する調整パラメータ（すなわち、スカラー定

50

数)を表す。

【0056】

他の実施形態では、ランキング関数は、上記で述べたクエリ非依存(QID)コンポーネントと、少なくとも1つのクエリ依存(QD)コンポーネントの合計を含み、次のようになる。

【0057】

スコア = QD(ドキュメント、クエリ) + QID(ドキュメント)

QDコンポーネントは、任意のドキュメントスコアリング関数とすることができる。一実施形態では、QDコンポーネントは、2004年3月18日に出願された、「FIELD WEIGHTING IN TEXT DOCUMENT SEARCHING」と題する特許文献5に記載のフィールド重み付けスコアリング関数(field weighted scoring function)に対応し、その主題を参照により本明細書にその全体を組み込む。フィールド重み付けスコアリング関数の表現として使用できる1つの式は次のようになる。

【0058】

【数2】

$$QD(\text{ドキュメント、クエリ}) = \sum \frac{wtf'(k_1 + 1)}{k_1 + wtf'} \times \log\left(\frac{N}{n}\right)$$

【0059】

ただし、

wtf'は、すべてのフィールド(例えば、ドキュメントのタイトル、本体など)にわたる重みで乗算され、かつ各フィールド長および対応する平均長さにより正規化された、検索クエリの所与のタームの重み付けされたターム頻度、またはターム頻度の合計を表し、

Nはネットワーク上のドキュメント数を表し、

nはクエリタームを含むドキュメント数を表し、また

k₁は調整可能な定数である。

上記の項および式はさらに、特許文献5で詳細に述べられており、その主題を参照により本明細書にその全体を組み込む。

【0060】

いくつかの実施形態では、ランキング関数はさらに、その両方の主題が全体として参照により本明細書に組み込まれる(i)特許文献1に開示された方法により決定されるクリック距離値、(ii)特許文献2に開示された方法により決定されるバイアスされたクリック距離値、(iii)ドキュメントのURL深さ、または(iv)(i)もしくは(ii)および(iii)の組合せ、を考慮に入れるQIDコンポーネントを含むことができる。例えば、この任意選択の追加のQIDコンポーネントは、以下の関数を含むことができる。

【0061】

【数3】

$$QID(\text{ドキュメント}) = w_{cd} \frac{k_{cd}}{b_{cd} \frac{CD}{k_{ew}} + b_{ud} UD} \frac{1}{k_{cd} + \frac{b_{cd} + b_{ud}}{k_{ew}}}$$

【0062】

ただし、

w_{cd}は、クリック距離またはバイアスされたクリック距離パラメータを含むコンポーネントなどのクエリ非依存コンポーネントの重みを表し、

b_{cd}は、URL深さに対するクリック距離またはバイアスされたクリック距離の重みを表し、

10

20

30

40

50

b_{ud} は、URL 深さの重みを表し、

CD は、ドキュメントに対する、計算されたもしくは割り当てられたクリック距離、またはバイアスされたクリック距離を表し、

k_{ew} は、他の調整パラメータと同様の、ランキング関数の精度を最適化することにより決定される調整定数を表し（すなわち、 k_{ew} は、すべてのエッジが同じエッジ重み値を有する場合、エッジ重み値を表し、エッジ重み値が互いに異なる場合、平均のエッジ値を表すことができる）、

UD は、URL 深さを表し、また

k_{cd} は、クリック距離飽和 (saturation) 定数である。

【0063】

重み付けされた項 (w_{cd} 、 b_{cd} 、および b_{ud}) は、その関係する各項（すなわち、クリック距離もしくはバイアスされたクリック距離パラメータ、所与のドキュメントに対するクリック距離もしくはバイアスされたクリック距離値、および所与のドキュメントの URL 深さをそれぞれ含むコンポーネント）の重要度を、かつ最終的にスコアリング関数の結果の重要度を定義するのを支援する。

【0064】

URL 深さ (UD) は、上記で参照したクエリ非依存コンポーネントに対する任意選択の追加であり、クリック距離またはバイアスされたクリック距離値がスコアリング関数に与え得る影響を均等化することができる。例えば、いくつかの場合では、あまり重要ではない（すなわち、URL 深さが大である）ドキュメントが、短いクリック距離またはバイアスされたクリック距離値を有することができる。URL 深さは、ドキュメントの URL 中で、スラッシュの数により表される。例えば、 $www.example.com/d1/d2/d3/d4.htm$ は、4 つのスラッシュを含んでおり、したがって URL 深さ 4 を有するはずである。しかし、このドキュメントは、比較的低いクリック距離またはバイアスされたクリック距離値を与える、メインページ $www.example.com$ から直接のリンクを有することもできる。上記で参照された関数に URL 深さの項を含めること、およびクリック距離またはバイアスされたクリック距離値に対して URL 深さの項を重み付けすることは、ネットワーク内でのドキュメントの重要度をさらに正確に反映させるように、比較的高いクリック距離またはバイアスされたクリック距離値を補償することになる。ネットワークに応じて、3 以上の URL 深さは、深いリンクと見なすことができる。

【0065】

一実施形態では、所与のドキュメントに対して、ドキュメント適合性スコアを決定するために使用されるランキング関数は、以下の関数を含む。

【0066】

【数 4】

$$\text{スコア} = \sum \frac{w_{tf}^i (k_1 + 1)}{k_1 + w_{tf}^i} \times \log\left(\frac{N}{n}\right) + w_{cd} \frac{k_{cd}}{b_{cd} \frac{CD}{k_{ew}} + b_{ud} UD} + w_u \frac{k_u U}{k_u + U}$$

【0067】

ただし、諸項は上記で述べたものである。

【0068】

他の諸実施形態では、クエリ依存コンポーネント、クエリ非依存コンポーネント、またはその両方の正確さを改善するために、URL 深さがランキング関数から除かれ、または他のコンポーネントがランキング関数に追加されることもあり得る。さらに、利用パラメータを含む上記で述べたクエリ非依存コンポーネントは、検索結果のランク付けを改善するために、他のランキング関数（示していない）に組み込むこともできる。

【0069】

ステップ204で、所与のドキュメントに対するドキュメント統計がランキング関数に提供された後、例示の方法20はステップ205に進む。ステップ205では、ドキュメント適合性スコアが所与のドキュメントに対して決定され、メモリに記憶され、所与のドキュメントと関連付けられる。ステップ205から、例示の方法20は判断ブロック206に進む。

【0070】

判断ブロック206で、ネットワーク内の各ドキュメントに対して、ドキュメント適合性スコアが計算されているかどうか、アプリケーションコードにより判定が行われる。ネットワーク内の各ドキュメントに対して、ドキュメント適合性スコアが計算されていないと判定された場合、例示の方法20はステップ204に戻り上記で述べたように進む。ネットワーク内の各ドキュメントに対して、ドキュメント適合性スコアが計算されていると判定された場合、例示の方法20はステップ207に進む。

10

【0071】

ステップ207で、数多くのドキュメントを含むクエリの検索結果は、その関連付けられたドキュメント適合性スコアに従ってランク付けされる。その結果得られたドキュメント適合性スコアは、ネットワーク内の各ドキュメントの実際のまたはデフォルトの利用値を考慮に入れる。検索結果がランク付けされた後、例示の方法20はステップ208に進み、ランク付けされた結果がユーザに表示される。ステップ208から、例示の方法20はステップ209に進み、最高にランク付けされた結果がユーザにより選択され、閲覧される。ステップ209から、例示の方法20はステップ210に進み、例示の方法20は終了する。

20

【0072】

ネットワーク内のドキュメントに対してドキュメント適合性スコアを生成し、またドキュメント適合性スコアを用いて、検索クエリの検索結果をランク付けする上記で述べた方法に加えて、上記で述べた方法を実施するためのコンピュータ実行可能命令をその上に記憶させたコンピュータ可読媒体がまた、本明細書で開示される。

【0073】

コンピューティングシステムがさらに本明細書で開示される。例示的なコンピューティングシステムは、そのコンピューティングシステムで使用可能な少なくとも1つのアプリケーションモジュールを含み、その少なくとも1つのアプリケーションモジュールが、その上にロードされたアプリケーションコードを含み、そのアプリケーションコードが、ネットワーク内のドキュメントに対するドキュメント適合性スコアを生成する方法を実施する。ネットワーク内のドキュメントに対するドキュメント適合性スコアを生成し、かつドキュメント適合性スコアを用いて、上記で述べたように検索クエリの検索結果をランク付けするためのコンピュータ実行可能命令をその上に有する上記で述べたコンピュータ可読媒体のうちのいずれかを用いて、アプリケーションコードをコンピューティングシステム上にロードすることができる。

30

【0074】

本明細書がその特定の諸実施形態に関して詳細に述べられてきているが、当業者であれば、前述の内容の理解が得られると、これらの実施形態に対する改変、変形、および均等な形態を容易に考え付くことができることを理解されたい。したがって、開示された方法、コンピュータ可読媒体、およびコンピューティングシステムの範囲は、添付の特許請求の範囲として、また請求項に対する任意の均等な形態の範囲として評価されるべきである。

40

【図面の簡単な説明】

【0075】

【図1】ユーザにより入力された検索クエリに応じてランク付けされた検索結果を生成する方法における例示的な諸ステップを示す例示の論理的な流れ図である。

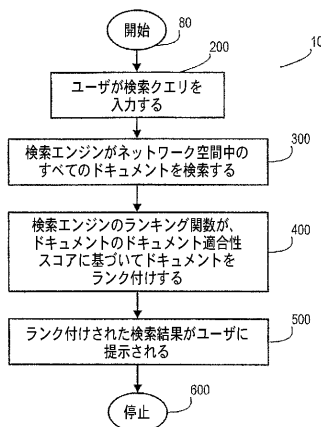
【図2】本明細書で開示の方法およびプロセスを実施するための例示的な動作環境の主要なコンポーネントのいくつかのブロック図である。

50

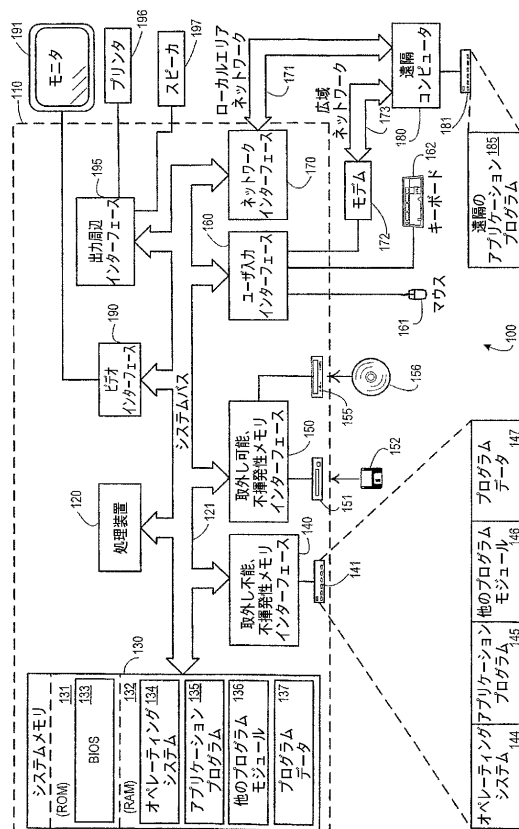
【図 3】ネットワーク上のドキュメントに対するドキュメント適合性スコアを決定するための例示の方法における例示的な諸ステップを示す論理的な流れ図である。

【図 4】ドキュメント利用パラメータを含むランキング関数を用いて生成された検索結果をランク付けする方法における例示的な諸ステップを示す論理的な流れ図である。

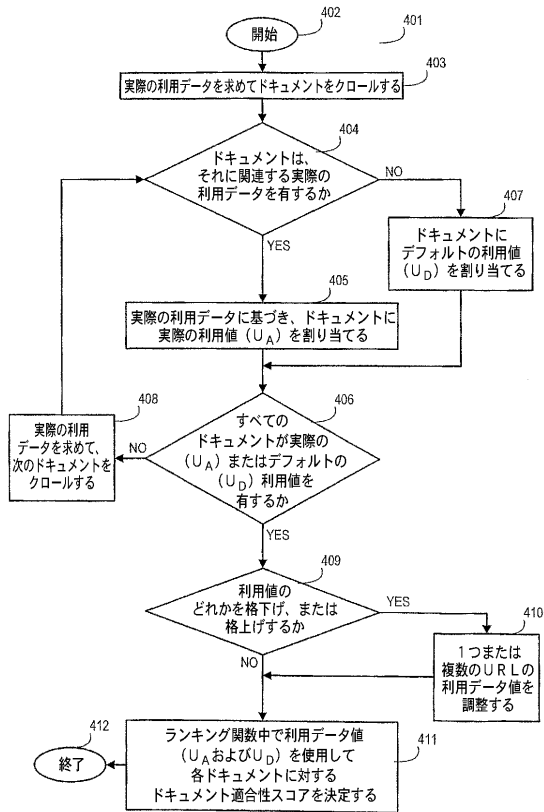
【図 1】



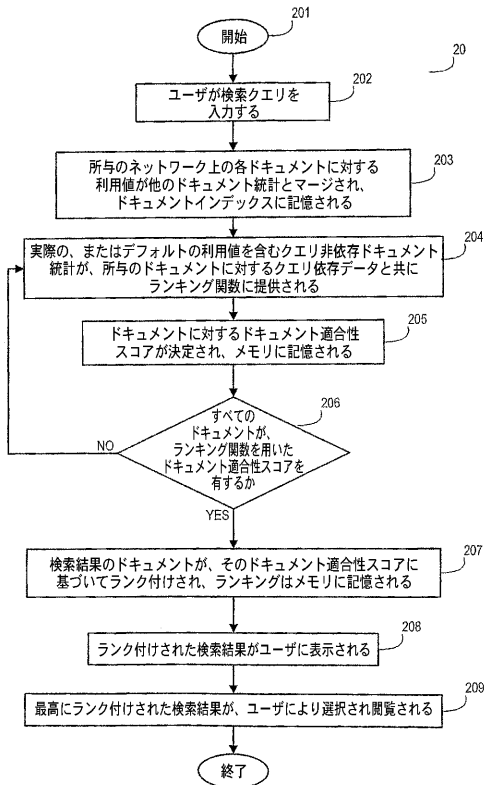
【図 2】



【図 3】



【図 4】



フロントページの続き

- (72)発明者 ドミトリー メイヤーゾン
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 インターナショナル パテンツ内
- (72)発明者 ヒューゴー サラゴサ
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 カイル 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 インターナショナル パテンツ内
- (72)発明者 アンドリュー デブルイネ
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2002-202992(JP, A)
特開2003-076715(JP, A)
特開2002-157271(JP, A)
米国特許出願公開第2005/0210079(US, A1)

- (58)調査した分野(Int.Cl., DB名)
G06F 17/30