

(19) 日本国特許庁(JP)

(12) 特 許 公 報(B2)

(11) 特許番号

特許第5312498号  
(P5312498)

(45) 発行日 平成25年10月9日(2013.10.9)

(24) 登録日 平成25年7月12日(2013.7.12)

(51) Int.Cl. F I  
**G06F 17/30 (2006.01)**  
 G06F 17/30 340B  
 G06F 17/30 350C  
 G06F 17/30 380E

請求項の数 22 (全 27 頁)

(21) 出願番号	特願2011-27886 (P2011-27886)	(73) 特許権者	305029922
(22) 出願日	平成23年2月10日(2011.2.10)		グーグル・インク
(62) 分割の表示	特願2006-533916 (P2006-533916) の分割		アメリカ合衆国、カリフォルニア州 94043、マウンテン・ビュー、アンフィシ アトル・パークウェイ 1600
原出願日	平成16年9月15日(2004.9.15)	(74) 代理人	110000408
(65) 公開番号	特開2011-159296 (P2011-159296A)		特許業務法人高橋・林アンドパートナーズ
(43) 公開日	平成23年8月18日(2011.8.18)	(72) 発明者	アチャーヤ・アマラグ
審査請求日	平成23年3月8日(2011.3.8)		アメリカ合衆国 カリフォルニア 95008 キャンプベル ポロードロード 1401
(31) 優先権主張番号	60/507, 617	(72) 発明者	カツ・マツ
(32) 優先日	平成15年9月30日(2003.9.30)		アメリカ合衆国 カリフォルニア 94040 マウンテンビュー ヒラリーアベニュー 116
(33) 優先権主張国	米国 (US)		
(31) 優先権主張番号	10/748, 664		
(32) 優先日	平成15年12月31日(2003.12.31)		
(33) 優先権主張国	米国 (US)		

最終頁に続く

(54) 【発明の名称】 ドキュメントをスコア付けする方法

(57) 【特許請求の範囲】

【請求項1】

一つまたは複数のサーバ装置によって実行されるドキュメントをランク付けする方法であって、

前記方法は、

検索結果のランキングにおいて時間に伴いドキュメントの位置が移動する量または比率を、前記検索結果のランキングにおける前記ドキュメントの一つ又は複数の前回の位置との比較によって決定し、

前記検索結果のランキングにおいて時間に伴い前記ドキュメントの位置が移動する前記量または前記比率を用いて、前記ドキュメントが時事的な現象またはスパムに関連するかを決定する所定の閾値に基づき、前記ドキュメントのスコアを生成し、

前記スコアに基づいて、前記ドキュメントが前記時事的な現象または前記スパムに関連するかの決定に基づき、前記ドキュメントを少なくとも一つの他のドキュメントとともにランク付けし、

前記ドキュメントのランク付けを、前記ドキュメントが前記スパムに関連する場合のスパムしようとする試みを回避するために用いることを含むことを特徴とするドキュメントをランク付けする方法。

【請求項2】

前記検索結果のランキングにおいて時間に伴い前記ドキュメントの位置が移動する前記量または前記比率を決定することは、

10

20

前記検索結果のランキングにおいて前記ドキュメントが移動する位置の動作量を決定することを含むことを特徴とする請求項 1 に記載のドキュメントをランク付けする方法。

【請求項 3】

前記検索結果のランキングにおいて時間に伴い前記ドキュメントの位置が移動する前記量または前記比率を決定することは、

前記検索結果のランキングにおいて前記ドキュメントが移動する位置の比率を決定することを含むことを特徴とする請求項 1 に記載のドキュメントをランク付けする方法。

【請求項 4】

前記ドキュメントが一セットの検索結果に含まれる際に前記ドキュメントが選択された頻度を決定することをさらに含み、

前記ドキュメントのスコアを生成することが、前記ドキュメントが一セットの検索結果に含まれる際に前記ドキュメントが選択された前記頻度に基づき前記スコアを生成することを含むことを特徴とする請求項 1 に記載のドキュメントをランク付けする方法。

【請求項 5】

前記ドキュメントがニュース記事に関連している場合に、前記ドキュメントを前記時事的な現象に関連するドキュメントであると決定することと、

前記ニュース記事に関連していない他のドキュメントを特定することと、

前記時事的な現象に関連するドキュメントに対して、前記他のドキュメントのスコアよりも高いスコアを決定することと、

をさらに含むことを特徴とする請求項 1 に記載のドキュメントをランク付けする方法。

【請求項 6】

一つまたは複数のサーバ装置によって実行されるドキュメントをランク付けする方法であって、

前記方法は、

時間に伴い前記ドキュメントのランクが変化する量を、前記ドキュメントのランクにおける前記ドキュメントの一つ又は複数の前回の位置との比較によって決定することと、

前記時間に伴い前記ドキュメントのランクが変化する前記量を用いて、前記ドキュメントが時事的な現象またはスパムに関連するかを決定する所定の閾値に基づき、前記ドキュメントのスコアを決定することと、

前記決定されたスコアに基づいて、前記ドキュメントが前記時事的な現象または前記スパムに関連するかの決定に基づき、前記ドキュメントを少なくとも一つの他のドキュメントとともにランク付けし、

前記ドキュメントのランク付けを、前記ドキュメントが前記スパムに関連する場合のスパムしようとする試みを回避するために用いることを含むことを特徴とするドキュメントをランク付けする方法。

【請求項 7】

前記ドキュメントのランクが所定の比率以上に変更されることを監視する手段を用いることをさらに含むことを特徴とする請求項 6 に記載のドキュメントをランク付けする方法。

【請求項 8】

前記ドキュメントのランクが所定期間の間に前記所定の閾値以上に増加することを監視することをさらに含むことを特徴とする請求項 6 に記載のドキュメントをランク付けする方法。

【請求項 9】

前記スコアを決定することは、前記時間に伴い前記ドキュメントのランクが変化する前記量に基づいて、前記ドキュメントのサーチクエリへの関連性に基づく関連スコアを組み合わせることで前記スコアを決定することを特徴とする請求項 6 に記載のドキュメントをランク付けする方法。

【請求項 10】

前記スコアを決定することは、前記ドキュメントのサーチクエリへの関連性に基づく関連

10

20

30

40

50

スコアを生成することなく前記スコアを決定することを特徴とする請求項 6 に記載のドキュメントをランク付けする方法。

【請求項 1 1】

ドキュメントをスコア付けするシステムであって、前記システムは、ネットワークを經由してクライアントと接続され、前記クライアントによって使用可能な一つまたは複数のサーバ装置を含み、

前記一つまたは複数のサーバ装置は、コンピュータに読込及び格納可能なデータを含む複数のドキュメントからドキュメントを特定し、

前記一つまたは複数のサーバ装置は、検索結果のランキングにおいて時間に伴い前記ドキュメントの位置が移動する量または比率を、前記検索結果のランキングにおける前記ドキュメントの一つ又は複数の前回の位置との比較によって決定し、

前記一つまたは複数のサーバ装置は、前記検索結果のランキングにおいて時間に伴い前記ドキュメントの位置が移動する前記量または前記比率を用いて、前記ドキュメントが時事的な現象またはスパムに関連するかを決定する所定の閾値に基づき、前記ドキュメントのスコアを生成し、

前記一つまたは複数のサーバ装置は、前記スコアに基づいて、前記ドキュメントが前記時事的な現象または前記スパムに関連するかの決定に基づき、前記ドキュメントを少なくとも一つの他のドキュメントとともにランク付けし、

前記ドキュメントのランク付けを、前記ドキュメントが前記スパムに関連する場合のスパムしようとする試みを回避するために用いることを特徴とするドキュメントをスコア付けするシステム。

【請求項 1 2】

前記検索結果のランキングにおいて時間に伴い前記ドキュメントの位置が移動する量または比率を決定するとき、前記一つまたは複数のサーバ装置は、前記ドキュメントが前記検索結果のランキングにおいて移動した位置の数を決定することを特徴とする請求項 1 1 に記載のドキュメントをスコア付けするシステム。

【請求項 1 3】

前記検索結果のランキングにおいて時間に伴い前記ドキュメントの位置が移動する量または比率を決定するとき、前記一つまたは複数のサーバ装置は、前記検索結果のランキングにおいて前記ドキュメントが移動する位置の比率を決定することを特徴とする請求項 1 1 に記載のドキュメントをスコア付けするシステム。

【請求項 1 4】

ドキュメントをランク付けするシステムであって、前記システムは、ネットワークを經由してクライアントと接続され、前記クライアントによって使用可能な一つまたは複数のサーバ装置を含み、

前記一つまたは複数のサーバ装置は、コンピュータに読込及び格納可能なデータを含む複数のドキュメントからドキュメントを特定し、

前記一つまたは複数のサーバ装置は、時間に伴い前記ドキュメントのランクが変化する量を、前記ドキュメントのランクにおける前記ドキュメントの一つ又は複数の前回の位置との比較によって決定し、

前記一つまたは複数のサーバ装置は、前記時間に伴い前記ドキュメントのランクが変化する前記量を用いて、前記ドキュメントが時事的な現象またはスパムに関連するかを決定する所定の閾値に基づき、前記ドキュメントのスコアを決定し、

前記一つまたは複数のサーバ装置は、前記決定されたスコアに基づいて、前記ドキュメントが前記時事的な現象または前記スパムに関連するかの決定に基づき、前記ドキュメントを少なくとも一つの他のドキュメントとともにランク付けし、

前記ドキュメントのランク付けを、前記ドキュメントが前記スパムに関連する場合のスパムしようとする試みを回避するために用いることを特徴とするドキュメントをランク付けするシステム。

【請求項 1 5】

前記一つまたは複数のサーバ装置は、前記ドキュメントのランクが所定の比率以上に変更されることを監視する手段を用いることを特徴とする請求項 14 に記載のドキュメントをランク付けするシステム。

【請求項 16】

前記一つまたは複数のサーバ装置は、前記ドキュメントのランクが所定期間の間に前記所定の閾値以上に増加することを監視することを特徴とする請求項 14 に記載のドキュメントをランク付けするシステム。

【請求項 17】

ドキュメントをスコア付けするコンピュータ実行可能命令を格納するコンピュータ読込可能媒体であって、

前記コンピュータ実行可能命令は、

検索結果のランキングにおいて時間に伴いドキュメントの位置が移動する量または比率を、前記検索結果のランキングにおける前記ドキュメントの一つ又は複数の前回の位置との比較によって決定する一つまたは複数の命令と、

前記検索結果のランキングにおいて時間に伴い前記ドキュメントの位置が移動する前記量または前記比率を用いて、前記ドキュメントが時事的な現象またはスパムに関連するかを決定する所定の閾値に基づき、前記ドキュメントのスコアを生成する一つまたは複数の命令と、

前記スコアに基づいて、前記ドキュメントが前記時事的な現象または前記スパムに関連するかの決定に基づき、前記ドキュメントを少なくとも一つの他のドキュメントとともにランク付けし、前記ドキュメントのランク付けを、前記ドキュメントが前記スパムに関連する場合のスパムしようとする試みを回避するために用いる一つまたは複数の命令と、  
を含むことを特徴とするコンピュータ読込可能媒体。

【請求項 18】

前記検索結果のランキングにおいて時間に伴い前記ドキュメントの位置が移動する前記量または前記比率を決定する一つまたは複数の命令は、前記ドキュメントが前記検索結果のランキングにおいて移動した位置の数を決定する一つまたは複数の命令を含むことを特徴とする請求項 17 に記載のコンピュータ読込可能媒体。

【請求項 19】

前記検索結果のランキングにおいて時間に伴い前記ドキュメントの位置が移動する前記量または前記比率を決定する一つまたは複数の命令は、前記検索結果のランキングにおいて前記ドキュメントが移動する位置の比率を決定する一つまたは複数の命令を含むことを特徴とする請求項 17 に記載のコンピュータ読込可能媒体。

【請求項 20】

ドキュメントをランク付けするコンピュータ実行可能命令を格納するコンピュータ読込可能媒体であって、

前記コンピュータ実行可能命令は、

時間に伴い前記ドキュメントのランクが変化する量を、前記ドキュメントのランクにおける前記ドキュメントの一つ又は複数の前回の位置との比較によって決定する一つまたは複数の命令と、

前記時間に伴い前記ドキュメントのランクが変化する前記量を用いて、前記ドキュメントが時事的な現象またはスパムに関連するかを決定する所定の閾値に基づき、前記ドキュメントのスコアを決定する一つまたは複数の命令と、

前記決定されたスコアに基づいて、前記ドキュメントが前記時事的な現象または前記スパムに関連するかの決定に基づき、前記ドキュメントを少なくとも一つの他のドキュメントとともにランク付けし、前記ドキュメントのランク付けを、前記ドキュメントが前記スパムに関連する場合のスパムしようとする試みを回避するために用いる一つまたは複数の命令と、

を含むことを特徴とするコンピュータ読込可能媒体。

【請求項 21】

10

20

30

40

50

前記コンピュータ実行可能命令が前記ドキュメントのランクが所定の比率以上に変更されることを監視する手段を用いる一つまたは複数の命令をさらに含むことを特徴とする請求項 20 に記載のコンピュータ読込可能媒体。

【請求項 22】

前記コンピュータ実行可能命令が前記ドキュメントのランクが所定期間の間に前記所定の閾値以上に増加することを監視する一つまたは複数の命令をさらに含むことを特徴とする請求項 20 に記載のコンピュータ読込可能媒体。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、一般的には、情報検索システム、より詳細には、関連するドキュメントに関連付けられたヒストリデータの少なくとも一部に基づいて検索結果を生成するためのシステムおよび方法に関する。

【背景技術】

【0002】

World Wide Web（「ウェブ」）は、膨大な量の情報を包含する。検索エンジンは、ウェブドキュメントをカタログ化することにより、この情報についての目標位置へとユーザーを到達させるように支援する。通常、ユーザーのリクエストに回答して、検索エンジンは、そのリクエストに関連するドキュメントへのリンクを返す。

【0003】

検索エンジンは、ユーザーによって提供された（サーチクエリと呼ばれる）検索用語で、ユーザーの関心事の決定の基礎を形成する。検索エンジンの目的は、サーチクエリに基づく高品質の関連する結果へのリンクを特定することである。通常、この検索エンジンは、サーチクエリ中の用語を、予め格納されたウェブドキュメントのコーパスへとマッチングさせることによって、これを達成する。ユーザーの検索用語を包含するウェブドキュメントは、「検索結果」とみなされ、ユーザーに返される。

【先行技術文献】

【特許文献】

【0004】

【特許文献 1】米国特許第 6, 285, 999 号

【発明の概要】

【発明が解決しようとする課題】

【0005】

理想的には、規定のユーザーのサーチクエリに回答して、検索エンジンは、もっとも関連性の高い結果をユーザーに提供することが望まれる。検索エンジンの一つのカテゴリーは、サーチクエリ用語とドキュメント中に含有される単語との比較に基づき、関連するドキュメントを特定する。検索エンジンの別のカテゴリーは、ドキュメント中のサーチクエリ用語の存在以外に、あるいはこの存在に加えて、ファクターを使用して関連するドキュメントを特定する。ある検索エンジンは、ドキュメントの相対的な重要度を決定するために、ドキュメントからあるいはドキュメントへのリンクに関連付けられた情報を使用する。

【0006】

検索エンジンの両カテゴリーとも、サーチクエリに対して高品質な結果の提供を図る。検索エンジンによって生成された結果の品質に影響を与えることができるいくつかのファクターが存在する。例えば、いくつかのウェブサイトのプロデューサーは、人為的にランクを上げるスパミング技法を使用する。さらに、「古くなった」ドキュメント（すなわち、一定期間に更新されていない、すなわち古いデータを包含するドキュメント）は、「より新鮮な」ドキュメント（すなわち、より最近に更新されている、つまりより最近のデータを包含するドキュメント）よりも高位置にランク付けされる可能性もある。個々の状況においては、古くなったドキュメントのより高位置のランク付けは、検索結果の品質

10

20

30

40

50

を低下させる。

【 0 0 0 7 】

従って、検索エンジンによって生成された結果の品質を向上させる必要性が、引き続き存在する。

【課題を解決するための手段】

【 0 0 0 8 】

本発明の原理にかかるシステムおよび方法は、ドキュメントに関連付けられたヒストリデータの少なくとも一部に基づいて、ドキュメントをスコア付けする。このスコア付けは、サーチクエリに関連して生成される検索結果を向上させるために使用することができる。

10

【 0 0 0 9 】

本発明の原理にかかる一実施態様においては、ドキュメントをスコア付けする方法が提供される。この方法は、ドキュメントの特定、および、ドキュメントに関連付けられた一つ以上のタイプのヒストリデータの獲得を含む。この方法は、さらに、一つ以上のヒストリデータのタイプの少なくとも一部に基づくドキュメントのスコアの生成を含むことができる。

【 0 0 1 0 】

別の実施態様においては、ドキュメントをスコア付けする方法が、提供される。この方法は、リンクされたドキュメントに関連付けられたリンケージ・データの年数の決定と、リンケージ・データの年数のランク下げ機能によって、リンクされたドキュメントのラン

20

【発明の効果】

【 0 0 1 1 】

ユーザーのサーチクエリに応答して、検索エンジンによりもっとも関連性の高い検索結果すなわち品質の高い検索結果をユーザーへ提供する。

【図面の簡単な説明】

【 0 0 1 2 】

本発明の一部に組み込まれ且つ本発明の一部を構成する、添付図面は、本発明の実施例を示し且つ記述と併せて本発明を説明する。

【図 1】本発明の原理にかかるシステムおよび方法が実施可能な、ネットワークの一例についてのダイアグラムである。

30

【図 2】本発明の原理にかかる一実施態様における、図 1 のクライアントおよび / またはサーバのダイアグラムの一例である。

【図 3】本発明の原理にかかる一実施態様における、図 1 の検索エンジンの機能ブロックダイアグラムの一例である。

【図 4】本発明の原理にかかる一実施態様における、ドキュメントをスコア付けするためのプロセッシングの一例についてのフローチャートである。

【発明を実施するための形態】

【 0 0 1 3 】

以下の本発明にかかる詳述は、添付の図面を参照する。異なる図面中の同一の参照番号は、同一または類似の要素とみなす。さらに、以下の詳述は、本発明を限定するものではない。

40

【 0 0 1 4 】

本発明の原理にかかるシステムおよび方法は、例えば、ドキュメントに関連付けられたヒストリデータを使用してドキュメントをスコア付けすることができる。このシステムおよび方法は、高品質の検索結果を提供するためにこれらのスコアを使用することができる。

【 0 0 1 5 】

ここで用いられる用語としての「ドキュメント」とは、コンピュータに読込および格納可能なあらゆる情報を含むと広義に解釈されるべきである。ドキュメントは、電子メール

50

、ウェブサイト、ファイル、ファイルの組み合わせ、他のファイルへ組み込まれたリンクを伴う一つ以上のファイル、ニュースグループへの投稿、ブログ、ウェブ広告等を含むことができる。インターネットにおいて、共通するドキュメントは、ウェブページである。ウェブページは、しばしばテキスト情報を含むことができ、さらに、（例えば、メタ情報、画像、ハイパーリンク等といった）組み込み情報、および/または、（例えば、Java Script等といった）組み込み命令を含むことができる。ページは、ドキュメントまたはドキュメントの一部に対応することができる。故に、「ページ」および「ドキュメント」という語は、場合によって同義に用いることができる。別のケースでは、ページは、例えばサブドキュメントといった、ドキュメントの一部を参照することができる。ページは、複数のドキュメントに対応する可能性もある。

10

**【0016】**

以下の記述において、ドキュメントは、他のドキュメントへのリンク、および/または、他のドキュメントからのリンクを有するものとして記述されることができる。例えば、ドキュメントが別のドキュメントへのリンクを含む場合、このリンクは、「前方のリンク」として参照されることができる。ドキュメントが別のドキュメントからのリンクを含む場合、このリンクは、「バック・リンク」として参照されることができる。「リンク」という語が使用される場合、前方のリンクまたは後方のいずれかのリンクを参照することができる。

**【0017】**

（ネットワーク構造の一例）

20

図1は、本発明の原理にかかるシステムおよび方法が実行可能な、ネットワーク100のダイアグラムの一例である。ネットワーク100は、ネットワーク150を通じて複数のサーバ120～140に接続される複数のクライアント110を含むことができる。ネットワーク150は、ローカルエリア・ネットワーク（Local Area Network: LAN）、広域ネットワーク（Wide Area Network: WAN）、公衆交換回線電話網（Public Switched Telephone Network: PSTN）、イントラネット、インターネット、メモリ装置、別タイプのネットワーク、またはネットワークの組み合わせを含むことができる。2つのクライアント110および三つのサーバ120～140は、簡易化のため、ネットワーク150に接続されるものとして示してある。実際には、クライアントおよびサーバは、これより少ない場合もあれば、多い場合もある。また、場合によっては、クライアントは、サーバの機能を実行し、サーバは、クライアントの機能を実行することもある。

30

**【0018】**

クライアント110は、クライアントエンティティを含むことができる。エンティティは、例えば無線電話、パーソナル・コンピュータ、携帯情報端末（PDA）、ノート型パソコン、もしくは別タイプの計算装置または通信装置、これらの装置のうち1つで実行されるスレッドまたはプロセス、および/またはこれらの装置の1つによって実行可能なオブジェクトといった、装置として定義することができる。サーバ120～140は、本発明の原理にかかる方法で、ドキュメントを集積し、処理し、検索し、および/または保存するサーバエンティティを含むことができる。クライアント110およびサーバ120～140は、有線、無線および/または光接続を経由してネットワーク150へと接続することができる。

40

**【0019】**

本発明の原理にかかる一つの実施態様において、サーバ120は、クライアント110によって使用可能な検索エンジン125を含むことができる。サーバ120は、ドキュメントコーパス（例えば、ウェブページ）をクロールし、ドキュメントをインデックス化し、クロールされたドキュメントのリポジトリ中のドキュメントに関連付けられた情報を格納する。サーバ130～140は、サーバ120によってクロールされることができるドキュメントを格納または保存することができる。サーバ120～140が個々のエンティティとして示される一方で、一つ以上のサーバ120 - 140が、別の一つ以上のサーバ

50

120 - 140のうち一つ以上の機能を実行する可能性もある。例えば、二つ以上のサーバ120 - 140が単独サーバとして実装される可能性もある。また、サーバ120 - 140のうち単独のサーバが二つ以上の個々の（あるいは分散型の）装置として実装される可能性もある。

【0020】

（クライアント/サーバ アーキテクチャの一例）

図2は、本発明の原理にかかる実施態様に従った、一つ以上のクライアント110およびサーバ120 - 140に対応することができる、クライアントエンティティまたはサーバエンティティ（これより「クライアント/サーバ エンティティ」と称する）についてのダイアグラムの一例である。このクライアント/サーバ エンティティは、バス210、プロセッサ220、メインメモリ230、リードオンリーメモリ（Read Only Memory：ROM）、ストレージデバイス250、一つ以上の入力装置260、一つ以上の出力装置270、通信インタフェース280を含むことができる。

バス210は、クライアント/サーバ エンティティのコンポーネント間の通信を許可する一つ以上の配線を含むことができる。

【0021】

プロセッサ220は、命令を翻訳し実行する一つ以上の従来型のプロセッサまたはマイクロプロセッサを含むことができる。メインメモリ230は、プロセッサ220による実行のための情報および命令を格納する、ランダムアクセスメモリ（Random Access Memory：RAM）または別タイプのダイナミック・ストレージデバイスを含むことができる。

ROM240は、プロセッサ220に使用するための情報および命令を格納する、従来型のROMまたは別タイプのスタティック・ストレージデバイスを含むことができる。ストレージデバイス250は、磁気記録装置、および/または、光学記録装置、並びにこれらに対応するドライブを含むことができる。

【0022】

入力装置260は、例えば、キーボード、マウス、ペン、音声認証機構および/または生体認証機構等といった、オペレータにクライアント/サーバ エンティティへと情報を入力するよう許可する一つ以上の従来型の機構を含むことができる。出力装置270は、一つ以上の従来型の機構を含み、オペレータへの出力情報はディスプレイ、プリンタ、スピーカなどを含むことができる。通信インタフェース280は、クライアント/サーバ エンティティがその他の装置および/またはシステムと通信ができるようにする、あらゆるトランシーバ様の機構を含む。例えば、通信インタフェース280は、例えばネットワーク150といったネットワークを経由し、別の装置またはシステムとの通信を行うための機構を含むことができる。

【0023】

以下に詳述するように、本発明の原理にかかるクライアント/サーバ エンティティは、一定の検索に関連する操作を実行する。このクライアント/サーバ エンティティは、例えばメモリ230といったコンピュータへ読み込み可能な媒体中に含有される、ソフトウェア命令を実行するプロセッサ220に回答してこれらの操作を実行する。コンピュータ読み込み可能な媒体は、一つ以上の物理メモリまたは論理メモリ装置および/または搬送波として決定される。

【0024】

ソフトウェア命令は、データ格納装置250といった別のコンピュータへ読み込み可能な媒体または通信インタフェース280を経由する別の装置から、メモリ230に読み込むことができる。メモリ230に含有されるソフトウェア命令は、プロセッサ220が後述するプロセスを実行する要因となる。もう一つの方法としては、ハードワイヤード回路は、本発明の原理にかかるプロセスを実行するためにソフトウェア命令の代わりまたはソフトウェア実行との組み合わせで、使用することができる。このようにして、本発明の原理にかかる実施態様は、ハードウェア回路およびソフトウェアのいかなる特定の組み合わ

10

20

30

40

50



せにも限定されない。

【0025】

(検索エンジンの一例)

図3は、本発明の原理にかかる実施態様における検索エンジン125の機能ブロックダイヤグラムの一例である。検索エンジン125は、ドキュメントロケータ310、ヒストリコンポーネント320、およびランキングコンポーネント330を含むことができる。図3に示すように、一つ以上のドキュメントロケータ310およびヒストリコンポーネント320は、ドキュメントコーパス340へ接続することができる。ドキュメントコーパス340は、先にクロールされ、インデックス化され、例えば検索エンジン125にアクセス可能なデータベースに格納された、ドキュメントに関連付けられた情報を含むことができる。ヒストリデータは、以下にさらに詳述されるように、ドキュメントコーパス340中のドキュメントのそれぞれに関連付けることができる。ヒストリデータは、ドキュメントコーパス340またはその他のあらゆる場所に格納することができる。

10

【0026】

ドキュメントロケータ310は、コンテンツがユーザー・サーチクエリに適合する一連のドキュメントを特定することができる。ドキュメントロケータ310は、最初に、コーパス中のドキュメントとユーザーのサーチクエリ中の用語を比較することによってドキュメントコーパス340からドキュメントを位置付けする。一般的に、検索された用語を包含する一連のドキュメントを返すために、ドキュメントをインデックス化し且つインデックス化された収集情報を検索するプロセスは、技術的に周知である。従って、このドキュメントロケータ310の機能性は、ここではさらに述べないでおく。

20

【0027】

ヒストリコンポーネント320は、ドキュメントコーパス340中のドキュメントに関連付けられたヒストリデータを収集することができる。本発明の原理にかかる実施態様において、ヒストリデータは、ドキュメント・インセプション日付、ドキュメント・コンテンツ更新/変更、クエリー分析、リンクに基づく基準、アンカー・テキスト(例えば、ハイパーリンクが組み込まれたテキストで、典型的にはドキュメント中にアンダーラインが引かれているかあるいはハイライトされているもの)、トラフィック、ユーザーの振る舞い、ドメイン関連の情報、ランキングヒストリ、ユーザーによって保存/生成されたデータ(例えば、ブックマーク)、アンカー・テキスト中の特有のワードやバイグラムおよびフレーズ、独立したピアのリンケージ、および/または、ドキュメント・トピックに関連するデータを含むことができる。これら異なるタイプのヒストリデータは、以下にさらに詳述されている。その他の実施態様において、ヒストリデータは、追加タイプまたは異なるタイプを含むことができる。

30

【0028】

ランキングコンポーネント330は、ランキングスコア(ここでは単に「スコア」と称する)を、ドキュメントコーパス340中の一つ以上のドキュメントに割り当てることができる。ランキングコンポーネント330は、サーチクエリより前、あるいはサーチクエリから独立し、もしくはサーチクエリと関連して、ランキングスコアを割り当てることができる。ドキュメントがサーチクエリに関連付けられている場合(例えば、サーチクエリに関連すると特定された場合)、検索エンジン125は、ランキングスコアに基づいてドキュメントをソートすることができる、さらにサーチクエリに提示されたクライアントへソートされた一連のドキュメントを返すことができる。本発明の態様にかかるにあたって、ランキングスコアは、ドキュメントの品質を定量化しようと試みる値である。本発明の原理にかかる実施態様において、スコアは、ヒストリコンポーネント320からのヒストリデータの少なくとも一部に基づいている。

40

【0029】

(ヒストリデータの一例)

[ドキュメント・インセプション日付]

本発明の原理にかかる一実施態様においては、ドキュメント・インセプション日付は、

50

そのドキュメントに関連付けられたスコアを生成（または変更）するために使用することができる。「日付」という語は、ここでは広義に使用され、故に、時間および日付の数値を含む。以下に記述されるように、ドキュメント・インセプション日付を決定するために使用できる技法がいくつか存在する。これらの技法のうちのいくつかは、ドキュメントに関連付けられたスコアを向上させようと望む第三者によって影響され得るという意味では、「バイアスがかかっている」。他の技法は、バイアスがかかっていない。これらの技法のいずれかまたはこれらの技法の組み合わせ、もしくはさらに別の技法は、ドキュメント・インセプション日付を決定するために使用することができる。

**【 0 0 3 0 】**

ある実施態様においては、ドキュメントのインセプション日付は、検索エンジン 1 2 5 がドキュメントを最初に認識したまたはインデックス化した日付から決定することができる。検索エンジン 1 2 5 は、クローリング、外部の情報源からの検索エンジン 1 2 5 へのドキュメントの提出（またはこれらの表示 / 要旨）、クローリングまたは提出に基づいたインデックス化技法の組み合わせ、もしくはその他の方法で、ドキュメントを発見することができる。別の方法としては、ドキュメントのインセプション日付は、検索エンジン 1 2 5 がドキュメントへのリンクを最初に発見した日付から決定することができる。

**【 0 0 3 1 】**

もう一つの実施態様においては、ドキュメントが登録されているドメインの日付は、このドキュメントのインセプション日付の表示として使用することができる。さらにもう一つの実施態様においては、ドキュメントが例えばニュース記事、ニュースグループ、メーリングリスト、あるいはこれらが一つ以上組み合わせられたドキュメントといった、別のドキュメント内で初めて参照される時間は、このドキュメントのインセプション日付を暗示するために使用することができる。さらなる実施態様においては、ドキュメントが少なくとも閾値のページ数を含む日付は、ドキュメントのインセプション日付の表示として使用することができる。もう一つの実施態様においては、ドキュメントのインセプション日付は、このドキュメントをホスティングするサーバによってドキュメントが関連付けられたタイムスタンプと等しくなることもできる。その他の本明細書で特に記載されていない技法、またはこれら技法の組み合わせは、ドキュメントのインセプション日付を決定または暗示するために使用できるだろう。

**【 0 0 3 2 】**

検索エンジン 1 2 5 は、ドキュメントをスコア付けするためにドキュメントのインセプション日付を使用することができる。例えば、かなり最近のインセプション日付のドキュメントは、その他のドキュメント（すなわち、バック・リンク）から相当に多くのリンクを有することはないと仮定することができる。スコアがドキュメントから / ドキュメントへのリンク数に基づいている既存のリンクに基づくスコア付け技法にとって、この最近のドキュメントは、より多数のリンク（すなわち、バック・リンク）を有するより古くなったドキュメントよりも低くスコア付けされることができる。しかしながら、ドキュメントのインセプション日付を考慮する場合、ドキュメントのスコアは、ドキュメントのインセプション日付に基づいて、（肯定的か否定的に）修正される場合もある。

**【 0 0 3 3 】**

1 0 個のバック・リンクによって参照される昨日のインセプション日付のドキュメントについての例を以下に考慮する。このドキュメントは、検索エンジン 1 2 5 によって 1 0 0 バック・リンクによって参照される 1 0 年前のインセプション日付のドキュメントよりも高くスコア付けすることができる。なぜならば、前者についてのリンク成長の比率が、後者より比較的高いからである。バック・リンクの数における成長が急上昇した比率は、ドキュメントをスコア付けする検索エンジン 1 2 5 によって使用されるファクターであることができるが、この比率はまた、検索エンジン 1 2 5 をスパミングしようという試みを示すこともできる。それに応じて、この場合においては、検索エンジン 1 2 5 は、スパミングの影響を減少させるためにドキュメントのスコアを実際に低下させることができる。

**【 0 0 3 4 】**

このようにして、本発明の原理にかかる一実施態様においては、検索エンジン 125 は、ドキュメントへのリンクが作成される（例えば、その期間におけるインセプション日付またはいくつかのウィンドウから作成されたリンクの数に基づく単位時間当たりの平均としての）比率を決定するためにドキュメントのインセプション日付を使用することができる。この比率はさらに、例えばリンクがより多く生成されたドキュメントにより加重を置いて、ドキュメントをスコア付けするために使用することができる。

【0035】

ある実施態様において、検索エンジン 125 は、以下のようにドキュメントのリンクに基づいたスコアを修正することができる。

$$H = L / \text{Log}(F + 2)$$

ここで、H は、ヒストリ調整されたリンクスコアを参照し、L は、ドキュメントに与えられたリンクスコアを参照することができるが、これらは、ドキュメントからのノドキュメントへのリンクに基づいて、スコアをドキュメントに割り当てる、あらゆる既知のリンクスコア付け技法（例えば、特許文献 1 に記述されているスコア付け技法）を使用して派生させることも可能である。また、F は、ドキュメント（または、期間内のウィンドウ）に関連付けられたインセプション日付から測定された経過時間について言及することができる。

【0036】

いくつかのクエリーについては、より古くなったドキュメントは、より新しいドキュメントよりも好まれるだろう。その結果、結果セットの平均年数からの（年数の）違いに基づいてドキュメントのスコアを調整することは、有益だろう。言い換えれば、検索エンジン 125 は、（例えば、これらのインセプション日付を使用し）結果セット中の各ドキュメントの年数を決定し、そのドキュメントの平均年数を決定し、ドキュメントの年数と平均年数との違いに基づいてドキュメントのスコアを修正することができる。

【0037】

要約すれば、検索エンジン 125 は、ドキュメントのインセプション日付に関連する情報の少なくとも一部に基づいてドキュメントに関連付けられたスコアを生成（または変更）することができる。

【0038】

[コンテンツの更新 / 変更]

本発明の原理にかかる一実施態様においては、ドキュメントのコンテンツが時間とともに変化するような様式に関連する情報が、ドキュメントに関連付けられたスコアを生成（または変更）するために使用することができる。例えば、コンテンツが編集済みのドキュメントは、コンテンツが時間の経過とともにスタティックな状態のままであるドキュメントとは異なるスコア付けができることがしばしばある。また、時間とともに更新されたコンテンツを比較的大量に有するドキュメントは、時間とともに更新されたコンテンツを比較的小量に有するドキュメントとは異なるスコア付けができる。

【0039】

ある実施態様において、検索エンジン 125 は、以下のようなコンテンツ更新スコア（U）を生成することができる。

$$U = f(UF, UA)$$

ここで、f は、例えば和または加重和といった関数を参照することができるが、UF は、ドキュメント（またはページ）がどれくらいの頻度で更新されるかを表示する更新頻度スコアを参照することができるが、ならびに、UA は、ドキュメントが（またはページ）が時間とともにどれだけ変更されたかを表示する更新量スコアを参照することができる。UF は、更新されるまでの平均期間、既知の期間における更新の数等を含めた多くの方法で決定することができる。

【0040】

UA はまた、例えば期間とともにドキュメントに関連付けられた「新しい」または特有なページ数といった、一つ以上のファクターの関数として決定することもできる。別のフ

10

20

30

40

50

ファクターは、時間とともにドキュメントに関連付けられた新しいまたは特有なページ数と、このドキュメントに関連付けられたページの合計数との比を含むだろう。さらに別のファクターは、ドキュメントが $n$ 以上の期間（例えば、ドキュメントの可視コンテンツのうち $n\%$ は、期間 $t$ （例えば、最近 $m$ ヶ月間）で変化する）で更新される量を含むことができ、これが平均値となるであろう。さらなるファクターは、ドキュメント（またはページ）が $n$ 以上の期間（例えば、最近 $x$ 日間以内）で変化した量を含むだろう。

【0041】

ある実施態様の一例においては、 $UA$ は、ドキュメント・コンテンツの加重部分の異なる関数として決定することができる。例えば、JavaScript、コメント、広告、ナビゲーション要素、ポイラー・プレート・マテリアル、あるいは日付/時間タグといった、更新/変更されても重要でないとみなされるコンテンツは、 $UA$ を決定する際に比較的ほとんど加重が置かれず、むしろ無視されるであろう。その一方で、例えば、前方のリンクに関連付けられたタイトルまたはアンカー・テキストといった、（例えば、より頻繁に、より最近、より広範囲にわたって等）更新/変更されると重要であるみなされるコンテンツは、 $UA$ を決定する際、他のコンテンツへの変更よりも多くの加重が置かれるだろう。

10

【0042】

$UF$ および $UA$ は、ドキュメントに割り当てられたスコアに影響を与えるための他の方法として使用することができる。例えば、現在の期間における変更の比率は、加速傾向が存在するのかわかも減速傾向が存在するのかわかを決定する別の（例えば、前回の）期間における変更の比率と比較することができる。たとえ安定した変更の比率が比較的高い比率だとしても、その安定した変更の比率のドキュメントより、変更の比率に増加のあるドキュメントが、高くスコア付けされる。変更の量は、またこのスコア付けにおけるファクターとなることができる。例えば、その変更の量がいくつかの閾値より大きい場合に変更の比率に増加が見られるドキュメントは、安定した変更の比率が見られるドキュメントもしくは変更の量が閾値よりも低いドキュメントよりも高くスコア付けされることができる。

20

【0043】

場合によっては、データ保存リソースは、コンテンツの変更のためにドキュメントを監視する際、ドキュメントを格納するのに不十分である。この場合、検索エンジン125は、ドキュメントの表示を格納しさらに変更に対するこれらの表示を監視することができる。例えば、検索エンジン125は、ドキュメント・コンテンツに対する変更を検出するため、すべてのドキュメントそのものの代わりにドキュメントの「符号定数」を格納することができる。この場合、検索エンジン125は、ドキュメント（またはページ）に対する用語ベクトルを格納し、さらに比較的大きな変更のために監視することができる。別の実施態様においては、検索エンジン125は、重要であるもしくは（「ストップワード」を除いて）もっとも頻繁に出現すると決定されるドキュメントのうち（例えば、いくつかの用語といった）比較的小さな部分を格納且つ監視することができる。

30

【0044】

さらに別の実施態様においては、検索エンジン125は、ドキュメントの要約またはその他の表示を格納し且つ変更に対する情報を監視することができる。さらなる実施態様においては、検索エンジン125は、ドキュメントに対し（ほぼ重複するドキュメントを検出するために使用することができる）類似ハッシュを生成し且つ変更に対する類似ハッシュを監視することができる。類似ハッシュにおける変更は、それが関連付けられたドキュメントにおいて比較的大きな変更を示すと考えることができる。その他の実施態様において、さらに別の技法が、変更のためのドキュメントの監視に使用されることができる。適切なデータ保存リソースが存在する場合において、全ドキュメントは、ドキュメントについてのいくつかの表示というよりはむしろ変更を決定するために格納且つ使用されることができる。

40

【0045】

いくつかのクエリーに対して、最近変更がないコンテンツを有するドキュメントは、最

50

近変更されたコンテンツを有するドキュメントよりも好ましいだろう。その結果、結果セットの平均的な変更日からの相違に基づいて、ドキュメントのスコアを調整することは、有益である。言い換えれば、検索エンジン 125 は、結果セット中の各ドキュメントのコンテンツが最後に変更された日付を決定し、このドキュメントに対する変更の平均的な日付を決定し、さらに、ドキュメントの変更日および平均的な変更日との差異に基づいてドキュメントのスコアを（肯定的か否定的かに）修正することができる。

#### 【0046】

要約すれば、検索エンジン 125 は、ドキュメントのコンテンツが時間とともに変化する様式に関連する情報の少なくとも一部に基づいて、ドキュメントに関連付けられたスコアを生成（または変更）することができる。多数の個人または組織に属するコンテンツを含む非常に大量のドキュメントについて、そのスコアは、各サブドキュメント（すなわち、単独の個人または組織に属するかまたは更新されるコンテンツ）に対応させることができる。

10

#### 【0047】

##### [クエリー分析]

本発明の原理にかかると実施態様においては、一つ以上のクエリーに基づくファクターは、ドキュメントに関連付けられたスコアを生成（または変更）するために使用されることができる。例えば、あるクエリーに基づくファクターは、ドキュメントが一連の検索結果に含まれる場合にドキュメントが時間とともに選択される範囲内で関わるることができる。この場合、検索エンジン 125 は、ユーザーによって比較的頻繁に選択されている / ユーザーによる選択が増加しつつあるドキュメントを他のドキュメントよりも高くスコア付けすることができる。

20

#### 【0048】

別のクエリーに基づくファクターは、時間とともにクエリー中に現れる一定の検索用語の発生に関連させることができる。特定の一連の検索用語は、期間とともに次々とクエリーに現れることができる。例えば、人気を獲得している / 獲得した「最新の」トピックに関連する用語あるいはニュース速報は、一期間をとおして頻繁に現れるだろう。この場合、検索エンジン 125 は、これらの用語に関連付けられていないドキュメントよりも、これらの検索用語（またはクエリー）に関連付けられたドキュメントに高いスコア付けをすることができる。

30

#### 【0049】

さらなるクエリーに基づくファクターは、類似のクエリーによって生成された検索結果の数において、時間とともになう変化に関連させることができる。類似のクエリーによって生成された検索結果の数の著しい増加は、例えば、最新のトピックまたはニュース速報を示し、さらに検索エンジン 125 がそのようなクエリーに関連するドキュメントのスコアを増加させる原因となるであろう。

#### 【0050】

別のクエリーに基づくファクターは、時間とともに比較的一定さを保つクエリーに関連することができるが、しかし時間とともに変化する結果をもたらす。例えば、「ワールドシリーズ・チャンピオン」に関連するクエリーは、時間とともに変化する検索結果（例えば、特定のチームに関連するドキュメントは、既存の年またはシーズンにおける検索結果を左右する）をもたらす。この変化は、ドキュメントを適切にスコア付けするために監視され且つ使用することができる。

40

#### 【0051】

さらに別のクエリーに基づくファクターは、検索結果として返されるドキュメントの「古さ」に関連するであろう。ドキュメントの陳腐化は、例えば、ドキュメント作成日、アンカーの成長、トラフィック、コンテンツ変更、前方のリンク / バック・リンクの成長等といった、ファクターに基づくことができる。いくつかのクエリーについて、最近のドキュメントは、非常に重要である（例えば、FAQ (Frequently Asked Questions) 50

n) ファイルを検索する場合、最近のバージョンが非常に望ましい)。検索エンジン 125 は、検索結果中のどのドキュメントがユーザーのよって選択されるかを分析することにより、どのクエリーの最近の変更が最も重要であるかを学習することができる。より厳密にいうと、検索エンジン 125 は、ユーザーが検索結果中のより古くなったドキュメントより低くランク付けされているより最近のドキュメントをどのくらいの頻度で好むのかを考慮することができる。加えて、時間に伴う特定のドキュメントが主に時事問題のクエリー（例えば、「ワールドシリーズ・チャンピオン」）か、より限定的なクエリー（例えば、「ニューヨーク・ヤンキーズ」）のどちらかに含まれている場合、このクエリーに基づくファクター - それ自身または本明細書で述べたその他のファクター - は、陳腐であるように思われるドキュメントに対しスコアを低くするために使用されることができる。

10

**【0052】**

場合によって、古くなったドキュメントは、より最近のドキュメントより好まれると考えられてもよい。その結果、検索エンジン 125 は、ドキュメントに対しスコアを生成する際に、どのドキュメントが時間とともに選択されるのかという範囲を考慮することができる。例えば、既知のクエリーに対し、ユーザーが時間とともにより低くランク付けされた比較的古くなったドキュメントを、より高くランク付けされた比較的最近のドキュメントよりも選択する傾向がある場合、これは検索エンジン 125 によって古くなったドキュメントのスコアを調整するための指標として使用されることができる。

**【0053】**

さらに別のクエリーに基づくファクターは、ドキュメントが、異なるクエリーに対しての結果で現れる範囲を関連させることができる。言い換えれば、一つ以上のドキュメントに対するクエリーのエントロピーは、スコア付けのための基本として監視されかつ使用されることができる。例えば、特定のドキュメントが一連の不均一なクエリーについてのヒットとして現れる場合、これは、（必ずしもではないが）このドキュメントがスパムであるシグナルと考えることができる。この場合検索エンジン 125 は、ドキュメントを比較的低くスコア付けすることができる。

20

**【0054】**

要約すると、検索エンジン 125 は、一つ以上のクエリーに基づくファクターの少なくとも一部に基づいてドキュメントに関連付けられるスコアを生成（または変更）することができる。

30

**【0055】****[リンクに基づく基準]**

本発明の原理にかかると実施態様においては、一つ以上のリンクに基づくファクターが、ドキュメントに関連付けられるスコアを生成する（または変更する）ために使用されることができる。ある実施態様において、リンクに基づくファクターは、新しいリンクがドキュメントに現れ且つ既存のリンクが消失する日付に関連することができる。リンクの出現日は、検索エンジン 125 がリンクを包含するリンクまたはドキュメントの日付（例えば、ドキュメントがリンクとともに検出された日付、もしくは、最後に更新された日付）を検出する最初の日付であることができる。リンクの消失日は、リンクを包含するドキュメントがリンクを落とすかまたはドキュメント自体が消失する最初の日付であることができる。

40

**【0056】**

これらの日付は、クローलまたはインデックス更新の操作中に検索エンジン 125 によって決定されることができる。この日付を参照として使用して検索エンジン 125 は、例えばリンクが現れるかまたは消える時間、リンクが時間とともに現れるかまたは消える比率、リンクが既存の期間の間に現れる数または消える数、ドキュメントに対する新しいリンクの出現と既存リンクの消失への傾向の存在の是非等といったドキュメントに対するリンクの時間依存性振る舞いを監視することができる。

**【0057】**

ドキュメントへの（および/または、ドキュメントからの）リンクの時間依存性振る舞い

50

を使用し、検索エンジン125は、ドキュメントを適切にスコア付けすることができる。例えば、時間とともに新しいリンクの数または比率における（例えば、最近の期間と古い期間における新しいリンクの数または比率の比較に基づく）減少傾向は、ドキュメントが古いというシグナルを検索エンジン125へと出すだろう。この場合、検索エンジン125はドキュメントのスコアを減少させることができる。逆に、増加傾向は、特別のシチュエーションおよび実施態様によって、より関連性があるとみなされる「新鮮な」ドキュメント（例えば、コンテンツが新鮮である - すなわち最近に作成されたか更新されたドキュメント）というシグナルを出すことができる。

#### 【0058】

時間とともにドキュメント（またはページ）に対するバック・リンクの増加/減少の数または比率の変化を分析することによって、検索エンジン125は、ドキュメントがどれだけ新鮮であるかについての価値あるシグナルを導くことができる。例えば、このような分析が急降下するカーブによって表されると、これは、このドキュメントが古い（例えば、もはや更新されず、重要性が衰え、別のドキュメントにとって代えられる等）というシグナルを出すことができる。

#### 【0059】

ある実施態様においては、この分析は、ドキュメントに対する新しいリンクの数に依存することができる。例えば、検索エンジン125は、ドキュメントが最初に検索されて以来からの新しいリンクの数と比較して、最近のn日間におけるドキュメントへの新しいリンクの数を監視することができる。あるいは、検索エンジン125は、第1のリンクが検索された年数と比較して、リンクでもっとも最近のy%の最も古い年数を決定することができる。

#### 【0060】

説明のために、 $y = 10$ 、および、両者とも100日前に最初に検索された2つのドキュメント（この例ではWebサイト）を考える。第1のサイトについて、リンクのうち10%は、10日以内に検索されていて、一方第2のサイトについては、リンクの0%がここ10日以内に検索された（言い換えると、このリンクはすべてより早い時期に検索された）。この場合、メトリックは、サイトAについては0.1、サイトBについては0という結果になる。このメトリックは、適切に尺度調整することができる。別の実施態様の一例において、メトリックは、リンク日付の分散についての比較的より詳細な分析を実行することにより修正することができる。例えば、特定の分散が特定のタイプのサイト（例えば、もはや更新されないサイト、人気が増加または減少しているサイト、別のサイトにとって代えられたサイト等）を表すか否かを予測するモデルを構築されることができる。

#### 【0061】

別の実施態様においては、この分析は、リンクに割り当てられた加重に依存することができる。この場合、各リンクは、リンクの新鮮度に伴い増加する関数によって加重されることができる。リンクの新鮮度は、リンクの出現/変更の日付、リンクに関連付けられたアンカー・テキストの出現/変更の日付、リンクを包含するドキュメントの出現/変更の日付によって決定することができる。リンクを包含するドキュメントの出現/変更の日付は、優良なリンクは変更されないままという理論に基づき、ドキュメントが更新された際には依然として関連するもの且つ優良であるかどうかの、リンクの新鮮度についてのより優良な指標となることができる。ドキュメントに関連のない些細な部分の微小な編集によるすべてのリンクの新鮮度を更新しない手段として、更新された各ドキュメントは、大きな変更（例えば、ドキュメントの大部分への変更またはドキュメントの多くの異なる部分への変更）について検査されることができ、さらに、リンクの新鮮度が適切に更新される（または更新されないようにする）ことができる。

#### 【0062】

リンクは、別の方法で加重されることができる。例えば、リンクを包含するドキュメントがどれだけ信頼されているかに基づいて加重されることができる（例えば、政府ドキュメントは、高い信頼性を与えられるであろう）。リンクはまた、あるいは、リンクを包含

10

20

30

40

50

するドキュメントがどれだけ権威あるものかどうかに基づいて、加重されてもよい（例えば、権威あるドキュメントは、特許文献1に記載の様式に類似の方法で決定されることができる。）。リンクはまた、あるいは、新鮮度を確立するために、いくつかの他の特徴を使用したリンクを含むドキュメントの新鮮度に基づいて加重される（例えば、頻繁に更新されるドキュメント（Yahooホームページなど）は、突然、リンクをドキュメントへ貼り付ける）。

#### 【0063】

検索エンジン125は、ドキュメントに張られたリンクの加重合計の関数として、リンクが存在するドキュメントのスコアを、高くあるいは低くすることができる。この技法は、再帰的に使用される。例えば、ドキュメントSが2年存在すると仮定する。Sへのリンクのうちn%が新鮮か、もしくは、Sへの前方のリンクを包含するドキュメントが新鮮とみなされる場合に、ドキュメントSは、新鮮とみなすことができる。後者は、ドキュメントの作成日の使用およびこの技法を再帰的な応用によって、チェックをかけることができる。

10

#### 【0064】

さらに別の実施態様においては、この分析は、ドキュメントへのリンクポイントに関連付けられた年数分布に依存することができる。言い換えると、ドキュメントへのリンクが作成された日付は、年数分布を決定する関数へ決定および入力することができる。古くなったドキュメントの年数分布は、新鮮なドキュメントの年数分布とはかなり異なるであろうと仮定されることができる。検索エンジン125は、ドキュメントに関連付けられた年数分布の少なくとも一部に基づいて、ドキュメントをスコア付けする。

20

#### 【0065】

リンクが出現する日付は、ドキュメントのオーナーまたはその同僚が、検索エンジンによって割り当てられたスコアを吊り上げる目的で独自のドキュメントを作成する、「スパム」を検出するためにも使用されることができる。典型的「合法的な」ドキュメントは、バック・リンクをゆっくりと引き付ける。バック・リンクの量における大きなスパイクは、時事的な現象（例えば、CDCウェブサイトは、SARSなどが勃発した直後に多数のリンクを発展することができる）を示すことができる。もしくは、シグナルは、リンクの交換、リンクの購入、またはマーキングされたリンクの編集決定権なしにドキュメントからリンクを得ることによって、（より高いランク付けを獲得し、さらにそれによって検索結果においてより優良な位置を獲得するために）検索エンジンにスパミングしようと試みる。編集決定権のないリンクを与えるドキュメントの例は、ゲストブック、参照者ログ、および全ての者にドキュメントにリンクを追加する「無料の」ページを含む。

30

#### 【0066】

さらなる実施態様においては、この分析は、リンクが消失した日付に依存することができる。多数のリンクの消失は、これらのリンクがポイントしたドキュメントが古くなっている（例えば、もはや更新されないか、または別のドキュメントにとって代えられている）ことを意味する。例えば、検索エンジン125は、ドキュメントに対する一つ以上のリンクが消失する日付、既知の時間のウィンドウに消失するリンクの数、もしくは、古くなったとみなされることができるドキュメントを特定するためのドキュメントに対するリンクの数におけるいくつかの他の時間依存性減少（または、そのようなリンクを包含するドキュメントへのリンク/更新）を、監視することができる。いったんドキュメントが古くなったと決定されると、そのドキュメントに含有されるリンクは、リンクによってポイントされたドキュメントに対するスコアを決定する際に検索エンジン125によってカウントされないかもしくは無視されるだろう。

40

#### 【0067】

別の実施態様においては、この分析は、ドキュメントに対するリンクの年数だけでなくリンクのダイナミック性に依存することができる。このようなものとして、検索エンジン125は、非常に新鮮なリンクを有するにもかかわらず、日々異なった特徴づけをされたリンクを有するドキュメントを、一貫して更新され且つ一貫して既存の目標ドキュメント

50



にリンクするドキュメントとは異なった（例えば、より低く）加重をすることができる。ある実施態様の一例において、検索エンジン125は、ドキュメントに対するリンクを伴ったドキュメントのスコアに基づき、時間のウィンドウ内のドキュメントの全バージョンについて、ドキュメントのスコアを生成することができる。これと別のバージョンは、ドキュメントの主な更新時間に基づいて、統合に向けた軽視/消失を考慮することができる。

【0068】

要約すると、検索エンジン125は、一つ以上のリンクに基づいたファクターの少なくとも一部に基づいて、ドキュメントに関連付けられたスコアを生成（または変更）することができる。

10

【0069】

[アンカー・テキスト]

本発明の原理にかかる一実施態様においては、アンカー・テキストが時間とともに変化する様式に関連する情報は、ドキュメントに関連付けられたスコアを生成（または変更）するために使用されることができる。例えば、ドキュメントに対するリンクに関連付けられたアンカー・テキストにおける時間に伴う変化は、更新またはさらにドキュメントのフォーカスの変化が存在するという指標として使用されることができる。

【0070】

あるいは、ドキュメントのコンテンツが変更されると、バック・リンクに関連付けられたアンカー・テキストと著しく異なり、さらに、ドキュメントに関連付けられたドメインは、前回の内容とは著しく（完全に）変更されることができる。これは、ドメインが期限切れになる時や第三者がそのドメインを購入する時に起こるだろう。アンカー・テキストは、しばしば関連付けられたリンクポイントに関連付けられるドキュメントの一部とみなされるため、ドメインは、もはやトピック上に存在しないクエリーの検索結果に出現する可能性がある。これは、望ましくない結果である。

20

【0071】

この問題に対処する一つの方法は、ドメインが焦点を変更した日付を査定することである。これは、ドキュメントのテキストが著しく変更される際またはアンカー・テキストが著しく変更される際の日付を決定することによってなされる。この日付以前の全てのリンクおよび/またはアンカー・テキストは、無視されるかまたは軽視されるだろう。

30

【0072】

アンカー・テキストの新鮮度はまた、ドキュメントのスコア付けにおけるファクターとして使用することができる。アンカー・テキストの新鮮度は、例えば、アンカー・テキストの出現/変更の日付、アンカー・テキストに関連付けられたリンクの出現/変更の日付、および/または、リンクポイントに関連付けられたドキュメントの出現/変更の日付によって、決定されることができる。リンクにポイントされたドキュメントの出現/変更の日付は、優良なアンカー・テキストは変更されないままという理論に基づき、ドキュメントが更新された際には依然として関連するもの且つ優良であるかどうかの、アンカー・テキストの新鮮度についての優良な指標となることができる。ドキュメントに関連のない些細な部分の微小な編集によるすべてのアンカー・テキストの新鮮度を更新しないために、更新された各ドキュメントは、大きな変更（例えば、ドキュメントの大部分への変更またはドキュメントの多くの異なる部分への変更）について検査されることができ、さらに、アンカー・テキストの新鮮度が適切に更新される（または更新されないようにする）ことができる。

40

【0073】

要約すると、検索エンジン125は、アンカー・テキストが時間とともに変化する様式に関連する情報の少なくとも一部に基づいて、ドキュメントに関連付けられたスコアを生成（または変更）することができる。

【0074】

[トラフィック]

50

本発明の原理にかかると実施態様においては、時間とともにドキュメントに関連付けられたトラフィックに関連する情報は、ドキュメントに関連付けられたスコアを生成（または変更）するために使用することができる。例えば、検索エンジン125は、1人以上のユーザーによる、ドキュメントに対するトラフィックの時間依存性特徴、またはこのようなドキュメントの他の「使用」を、監視することができる。トラフィックの大きな削減は、このドキュメントは古くなった（例えば、もはや更新されないか、または別のドキュメントにとって代えられる場合もある）ことを表示することができる。

#### 【0075】

ある実施態様において、検索エンジン125は、最近j日間（例えば、 $j = 30$ ）にわたるドキュメントについての平均トラフィックと、ドキュメントがほとんどのトラフィックを受信し、季節変化を任意に調整する一ヶ月間の平均トラフィック、もしくは最近k日間の平均トラフィック（例えば、 $k = 365$ ）を比較することができる。任意で、検索エンジン125は、トラフィック・パターンの反復また、あるいは時間ともなうトラフィック・パターンにおける変化を特定することができる。例えば夏季期間中、週末、もしくはその他の季節の期間といった、ドキュメントが多かれ少なかれ人気の多いまたは少ない（すなわち、トラフィックが多いまたは少ない）期間が存在することが、見出されるだろう。トラフィック・パターンの反復またはトラフィック・パターンにおける変化を特定することによって、検索エンジン125は、これらの期間中およびこれらの期間以外に、ドキュメントのスコア付けを適切に調整することができる。

#### 【0076】

加えて、またあるいは、検索エンジン125は、特別のドキュメントに対する「広告トラフィック」に関連する時間依存性特徴を監視することができる。例えば、検索エンジン125は、以下のファクターのうち1つまたはこれらの組み合わせを監視することができる：（1）時間とともに既知のドキュメントによって広告が表示されたまたは更新された範囲、および、時間とともに既知のドキュメントによって広告が表示されたまたは更新された比率、（2）広告主の質（例えば、Amazon.comといった比較的高いトラフィックと信頼を有し、時間とともに検索エンジン125に対して既知のドキュメントにリンクする/を参照する広告を有するドキュメントは、例えばポルノサイトといった低いトラフィック/信頼できないドキュメントを参照する広告を有するドキュメントよりも、比較的大きい加重を与えられることができる）、ならびに（3）広告が、関連するドキュメントに対するユーザー・トラフィックを生成する範囲（例えば、クリックした比率）。検索エンジン125は、ドキュメントをスコア付けするための広告トラフィックに関連する、これらの時間依存性特徴を使用することができる。

#### 【0077】

要約すると、検索エンジン125は、時間とともにドキュメントに関連付けられたトラフィックに関連する情報の少なくとも一部に基づいて、ドキュメントに関連付けられたスコアを生成（または変更）することができる。

#### 【0078】

##### [ユーザー振る舞い]

本発明の原理にかかると実施態様においては、個人に対応する情報または時間とともにドキュメントに関連するユーザー振る舞いの集積は、ドキュメントに関連付けられたスコアを生成（または変更）するために使用することができる。例えば、検索エンジン125は、ドキュメントが一連の検索結果から選択される回数、および/または、1人以上のユーザーが、ドキュメントにアクセスするために費やす時間の量を監視することができる。検索エンジン125はさらに、この情報の少なくとも一部に基づいてドキュメントをスコア付けすることができる。

#### 【0079】

ドキュメントが一定のクエリーに対し、時間とともにあるいは既知の時間ウィンドウ内で返されると、ユーザーは、同一または同類のクエリーを与えられたドキュメントの平均時間より多い時間または少ない時間を費やすと、これは、ドキュメントが、それぞれ、新

10

20

30

40

50

鮮であるかもしくは古くなったという指標として使用することができる。例えば、「リバービュー・スイミング・スケジュール」というクエリーが「リバービュー・スイミング・スケジュール」というタイトルを伴うドキュメントを返すと仮定する。さらに、ユーザーは、そこにアクセスするのに30秒費やしたが、今はこのドキュメントを選択する全てのユーザーは、そこにアクセスするのに数秒だけしか費やさないと仮定する。検索エンジン125は、このドキュメントは古くなった（すなわち、日にちが過ぎてしまったスイミング・スケジュールを包含する）ということを決断するために情報を使用することができ、さらにドキュメントを適切にスコア付けすることができる。

【0080】

要約すると、検索エンジン125は、時間とともにドキュメントに関連する個人またはユーザー振る舞いの集積に対応する情報の少なくとも一部に基づいて、ドキュメントに関連付けられたスコアを生成（または変更）することができる。

【0081】

[ドメインに関連する情報]

本発明の原理にかかると一実施態様においては、ドキュメントに関連付けられたドメインに関連する情報は、ドキュメントに関連付けられたスコアを生成（または変更）するために使用されることができる。例えば、検索エンジン125は、ドキュメントがコンピュータ・ネットワーク内（例えば、インターネット、イントラネットまたは他のネットワーク、もしくはドキュメントのデータベース）でホスティングされる方法に関連する情報を監視することができ、さらに、ドキュメントをスコア付けするためにこの情報を使用することができる。

【0082】

検索エンジンを欺こうと（スパムしよう）試みる個人は、使い捨てまたは「ドアウェイ」ドメインをしばしば使用し、見つけ出される前にできるだけ大量のトラフィックを獲得しようとする。ドメインの合法性についての情報は、これらのドメインに関連付けられたドキュメントのスコア付けを行う際に検索エンジン125によって使用されることができる。

【0083】

一定のシグナルは、違法なドメインと合法的なドメインを区別するために使用されることができる。例えば、ドメインは、最高10年間で更新される。価値ある（合法的な）ドメインは、数年分前もって料金が支払われるが、その一方で、ドアウェイ（違法な）ドメインは、1年以上利用されることは稀である。従って、将来ドメインの有効期限が切れる日付は、ドメインの合法性を予測する上でのファクターとして使用することが可能であり、また同様に、そこに関連付けられたリンクを張られたドキュメントもドメインの合法性を予測する上でのファクターとして使用されることが可能である。

【0084】

また、あるいは、ドメインに対するドメインネーム・サーバ（DNS）記録は、このドメインが合法的であるかどうかを監視されることができる。DNS記録は、ドメイン登録した人物、アドミニストレイティブアドレスおよびテクニカルアドレス、およびネームサーバ（すなわち、IPアドレスへのドメインネームを解決するサーバ）のアドレスの詳細を包含する。ドメインに対して時間とともにこのデータを分析することによって、違法なドメインが特定される。例えば、検索エンジン125は、物理的に正しいアドレス情報が一期間の間に存在するかどうか、ドメインについてのコンタクト情報がしばしば変更されるかどうか、ならびに、異なるネームサーバとホスティング会社との間に比較的多数の変更があるかどうか等を、監視することができる。ある実施態様において、既知のコンタクトが取れない情報・ネームサーバ・および/またはIPアドレスのリストは、特定され、格納され、ドメインの合法性を予測するために使用され、またこうしてそこに関連付けられたドキュメントもドメインの合法性を予測するために使用される。

【0085】

また、あるいは、ドメインに関連付けられたネームサーバに関する、年数またはその他

10

20

30

40

50

の情報は、ドメインの合法性を予測するために使用されることができる。「優良な」ネームサーバは、異なるレジストラから異なるドメインの混合を有することができるし、さらにこれらのドメインのホスティングの歴史を有することができる。その一方で、「悪質な」ネームサーバは、ポルノドメインまたはドアウエイドメイン、コマーシャル用語を含むドメイン（スパムに共通する指標）、または主として単独のレジストラからのバルクドメイン、または真新しいものをホスティングする可能性がある。ネームサーバの真新しさは、関連付けられたドメインの合法性を決定するにあたって、自動的に否定的なファクターにはならないだろうが、例えば前述したようなその他のファクターと組み合わせると、否定的なファクターとなる可能性がある。

【0086】

要約すると、検索エンジン125は、ドキュメントに関連付けられたドメインの合法性に関連する情報の少なくとも一部に基づいて、ドキュメントに関連付けられたスコアを生成（または変更）することができる。

【0087】

[ランキングヒストリ]

本発明の原理にかかると実施態様においては、ドキュメントの前のランキングに関連する情報は、ドキュメントに関連するスコアを生成（または変更）するために使用することができる。例えば、検索エンジン125は、検索エンジン125によって提供されたサーチクエリに回答して、ドキュメントの時間依存性ランキングを監視することができる。検索エンジン125は、多数のクエリを通してランキングが飛ぶように変動するドキュメントは、時事的なドキュメント、または検索エンジン125にスパミングをかけようという試みであると決定できる。

【0088】

このように、ランキングにおける時間に伴うドキュメントの動作量または比率は、このドキュメントに割り当てられる将来のスコアに影響させるために使用される。ある実施態様において、一連の各検索結果について、ドキュメントは、検索結果のうち最上位N個における位置づけに従って加重されることができる。N = 30について、ある関数の一例は、 $[(N + 1) - \text{SLOT} / N]^4$ となるだろう。この場合、最上位結果は、1.0のスコアを受信することができ、N番目の結果に近づくにつれ0に近づく。

【0089】

クエリセット（例えば、コマーシャルクエリ）は、反復され、ランク付けで上位M%超を獲得したドキュメントは、フラグ付けをされることができ、あるいは、ランク付けの百分率成長は、ドキュメントに対するスコアを決定する上でのシグナルとして使用することができる。例えば、検索エンジン125は、上位結果の平均（中間）スコアが比較的高い場合、クエリは商業的だろうと決定され、上位結果では、月ごとに著しい変化が存在する。検索エンジン125はまた、コマーシャル・クエリの指標として、大量生産を監視することができる。コマーシャル・クエリについては、スパムの可能性が高いため、検索エンジン125は、コマーシャル・クエリに関連付けられたドキュメントを、適切に取り扱うことができる。

【0090】

既知のクエリに対するドキュメントの位置（またはランキング）についてのヒストリに加えて、検索エンジン125は、ドキュメントが時間とともに検索結果として選択された、クエリの数および（増加する/減少する）比率や、時間に伴い検索結果として選択されたドキュメント、季節性、突発性、および検索結果として選択されたドキュメントの時間に伴うその他のパターン、ならびに/もしくは、URL-クエリ・ペアに対する時間に伴うスコアの変化といった、一つ以上の他のファクターを、（ページ、ホスト、ドキュメント、および/またはドメインベースで）監視することができる。

【0091】

加えて、またあるいは、検索エンジン125は、時間とともにクエリに基づく基準から独立した多数のドキュメント（例えば、URL）を監視することができる。例えば、検

10

20

30

40

50

検索エンジン 125 は、既知のクエリーまたは一連のクエリーに回答して生成された一連の最上位の結果間の平均スコアを監視し、さらにその一連の結果のスコア、および/または、既知のクエリーまたは一連のクエリーに回答して生成されたその他の結果についてのスコアを調整することができる。さらには、検索エンジン 125 は、時間に伴う特別なクエリーまたは一連のクエリーに対して生成された結果の数を監視することができる。検索エンジン 125 が結果の数が増加するもしくは増加の比率に変化が存在する（例えば、増加が「最新のトピック」またはその他の現象の指標であるようなこと）のを決定すると、検索エンジン 125 は、将来これらの結果をより高くスコア付けすることができる。

#### 【0092】

加えて、またあるいは、検索エンジン 125 は、ドキュメントのランクにおける突然の急激な上昇を検出するために時間とともにドキュメントのランクを監視することができる。この急激な上昇は、時事的な現象（例えば、最新のトピック）か、または、例えばリンクをトレードまたは購入するかによって検索エンジン 125 をスパムしようとする試みを示すことができる。検索エンジン 125 は、例えば、一定の比率にランクを成長させるためにヒステリシスを採用するという試みによって、スパムを回避するための尺度をとることができる。別の実施態様において、既知のドキュメントについてのランクは、時間について予め定義されたウィンドウに伴う一定の最大閾値の成長を与えることができる。スパムドキュメントと時事的な現象に関連するドキュメントとを区別するためのさらなる尺度として、検索エンジン 125 は、スパムドキュメントは、例えば、ニュースで寸評されることはないだろうという理論に基づいて、ニュース記事や討論グループのドキュメントについて寸評を考慮することができる。これらの技法のいずれかまたは組み合わせは、スパム試行を削減するために使用されることができる。

#### 【0093】

検索エンジン 125 が、例えば政府の文書や、ウェブ・ディレクトリ（例えば、Yahoo）および時間を通して比較的一定かつ高ランクにあるドキュメントといった、いくつかの点で権威あると決定されたドキュメントについての例外を作成する可能性はある。例えば、権威あるドキュメントに対するリンクの増加の数または比率において通常でない急激な上昇が発生すると、検索エンジン 125 は、このようなドキュメントはスパムでないとなす事ができる、またこのようにして、（時間に伴う）ランク（の成長）に対し比較的高いまたは閾値なしとすることができる。

#### 【0094】

加えて、またあるいは、検索エンジン 125 は、これらのドキュメントが「支持の範囲外」であるかまたは古くなっているかという指標としてドキュメントのランクにおける著しい急落を考慮することができる。例えば、時間とともにドキュメントのランクが著しく急落すると、検索エンジン 125 は、このドキュメントを時代遅れとみなし、適切にドキュメントをスコア付けする。

#### 【0095】

要約すると、検索エンジン 125 は、ドキュメントの先のランク付けに関連する情報の少なくとも一部に基づいて、ドキュメントに関連付けられたスコアを生成（または変更）することができる。

#### 【0096】

##### [ ユーザーが保持/生成するデータ ]

本発明の原理にかかると実施態様においては、ユーザーが保持または生成するデータは、ドキュメントに関連付けられたスコアを生成（または変更）するために使用することができる。例えば、各検索エンジン 125 は、「ブックマーク」「お気に入り」または、ユーザーに気に入られるかあるいは関心を持たれたドキュメントのうちいくつかの表示を提供する、その他のデータのタイプといった、ユーザーが保持または生成するデータを監視することができる。検索エンジン 125 は、直接的（例えば、ブラウザアシスタント）か、または間接的（例えば、ブラウザ）に、このデータを獲得することができる。検索エンジン 125 は、ドキュメントの重要性を決定するためにドキュメントが関連付けられたブッ

10

20

30

40

50

クマーク／お気に入りの数を時間とともに分析することができる。

【 0 0 9 7 】

検索エンジン 1 2 5 は、ブックマーク／お気に入りリストから、ドキュメント（あるいはより具体的には、ドキュメントへのパス）を追加または削除するための上昇傾向または下降傾向や、ドキュメントがブックマーク／お気に入りリストに加えられたまたは削除された比率、および／または、ドキュメントがブックマーク／お気に入りリストに追加・削除されるかもしくはブックマーク／お気に入りを通じてアクセスされるかどうかを分析することができる。多数のユーザーが、特別のドキュメントをブックマーク／お気に入りリストに加えるかもしくは時間とともにそのようなリストを通してこのドキュメントにしばしばアクセスする場合、これは、このドキュメントが比較的重要であるという指標となることができる。その一方で、多数のユーザーが、ブックマーク／お気に入りリストに表示されたドキュメントにアクセスするのが少なくなるかもしくはこれらのリスト由来のそのようなドキュメントへのパスが次第に削除／置換される場合、これは、ドキュメントが時代遅れで、人気がない等の指標であるとみなされることができる。検索エンジン 1 2 5 は、ドキュメントを適切にスコア付けすることができる。

10

【 0 0 9 8 】

他の実施態様において、時間とともに特別なドキュメントのユーザーの関心における増減を示すことができるユーザーデータの他のタイプは、ドキュメントをスコア付けするために検索エンジン 1 2 5 によって使用されることができる。例えば、ユーザーに関連付けられた「テンプ」またはキャッシュファイルは、時間とともに加えられたドキュメントにおける増加または減少が存在するかどうかを特定するために検索エンジン 1 2 5 によって監視されることができるだろう。

20

同様に、特別なドキュメントに関連付けられたクッキーは、ドキュメントの関心における上昇傾向または下降傾向が存在するかどうかを決定するために、検索エンジン 1 2 5 によって監視されるだろう。

【 0 0 9 9 】

要約すると、検索エンジン 1 2 5 は、ユーザーが保持または生成するデータの少なくとも一部に基づいて、ドキュメントに関連付けられたスコアを生成（または変更）することができる。

【 0 1 0 0 】

[ アンカー・テキストにおける特有なワード、バイグラム、フレーズ ]  
本発明の原理にかかると実施態様においては、アンカー・テキストにおける特有なワード、バイグラム、およびフレーズに関連する情報は、ドキュメントに関連付けられたスコアを生成（または変更）するために使用されることができる。例えば、検索エンジン 1 2 5 は、時間に伴うウェブ（またはリンク）・グラフおよびその振る舞いを監視することができる、さらに、この情報をスコア付け、スパム検出、またはその他の目的のために使用することができる。自ずと発展したウェブ・グラフ

30

は、典型的には個々の決定を伴う。統合的に生成されたウェブ・グラフは（通常スパム目的の暗示であるが）、調整された決定に基づいており、比較的急激な上昇しそうなアンカー・ワード／バイグラム／フェーズにおける成長のプロファイルの原因となる。

40

【 0 1 0 1 】

そのような急激な上昇についての一つの理由は、多くのドキュメントからの多数の同一のアンカー・テキストの加算だろう。もう一つの可能性は、多くのドキュメントからの異なるアンカーの故意の追加だろう。検索エンジン 1 2 5 は、アンカーを監視し、さらに、リンクポイントが関連付けられるドキュメントのスコア付けを考慮に入れることができる。例えば、検索エンジン 1 2 5 は、関連付けられたドキュメントのスコアについて推測するアンカーの影響を捕らえる。あるいは、検索エンジン 1 2 5 は、合成生成の可能性についての継続的尺度を使用することができるし、さらに、ドキュメントに対するスコアをスケールするために倍数因子を導出することができる。

【 0 1 0 2 】

50

要約すると、検索エンジン 125 は、ドキュメントにポイントされる一つ以上のリンクに関連付けられたアンカー・テキストにおける、特有なワード・バイグラム・フレーズに関する情報の少なくとも一部に基づいて、ドキュメントに関連付けられたスコアを生成（または変更）することができる。

【0103】

[独立したピアのリンケージ]

本発明の原理にかかると一実施態様においては、独立したピア（例えば、関係のないドキュメント）のリンケージに関する情報は、ドキュメントに関連付けられたスコアを生成（または変更）するために使用されることができる。

【0104】

個々のドキュメントの大量の出入りのいずれも/いずれかを伴った、明らかに独立したピアの数の急成長は、スパムへの意図の指標となる潜在合成ウェブ・グラフを示すことができる。この指標は、成長が著しく整合的または一致しないアンカー・テキストに対応する場合、強化することができる。二分決定項目（例えば、固定した量でスコアを降格する）または倍数因子といったリンクに基づいたスコア付け技法を使用する際に、この情報は、そのようなリンクの影響を降格するために使用することが可能である。

【0105】

要約すると、検索エンジン 125 は、独立したピアのリンケージに関する情報の少なくとも一部に基づいてドキュメントに関連付けられたスコアを生成（または修正）することができる。

【0106】

[ドキュメント・トピック]

本発明の原理にかかると一実施態様においては、ドキュメント・トピックに関する情報は、ドキュメントに関連付けられたスコアを生成（または変更）するために使用されることができる。例えば、検索エンジン 125 は、（例えば、カテゴリー化・URL 分析・コンテンツ分析・クラスタリング・要約化・一連の特有な低頻度用語、もしくはその他のタイプのトピック抽出を通して）トピック抽出を実行することができる。検索エンジン 125 は、時間とともにドキュメントのトピックを監視し且つスコア付け目的のためにこの情報を使用することができる。

【0107】

ドキュメントに関連付けられた一連のトピックにおける時間に伴う著しい変化は、このドキュメントは所有者が変更され、スコアやアンカー・テキスト等といった前回のドキュメント指標はもはや信頼できないことを示すことができる。同様に、トピック数における急激な上昇は、スパムを示す可能性がある。例えば、特定のドキュメントが「一定の時間」とみなされる期間において、一つ以上の一連のトピックに関連付けられる場合、ドキュメントに関連付けられるトピック数において（急激な）上昇が発生する。このことは、このドキュメントが「ドアウェイ」ドキュメントに支配されているという指標となることができる。別の指標は、ドキュメントに関連付けられたオリジナルトピックの消失を含むことができる。これらの状態が一つ以上検出されると、検索エンジン 125 は、そのようなドキュメントおよび/またはリンク、アンカー・テキスト、もしくはドキュメントに関連

【0108】

要約すると、検索エンジン 125 は、ドキュメントに関連付けられた一つ以上のトピックにおける変更の少なくとも一部に基づいて、ドキュメントに関連付けられたスコアを生成（または変更）することができる。

【0109】

(処理の一例)

図 4 は、本発明の原理にかかると一実施態様によるドキュメントをスコア付けするための処理の一例についてのフローチャートである。処理は、ドキュメントを特定するサーバ 120 から開始される（動作 410）。このドキュメントは、例えば、サーチクエリに関連し

10

20

30

40

50

ていると特定されるといったような、サーチクエリに関連付けられた一つ以上のドキュメントを含むことができる。もう一つの方法として、このドキュメントは、いかなるサーチクエリからも独立したドキュメント（例えば、ネットワークのクローリングによって特定され且つレポジトリに格納されるドキュメント）のコーパスまたはレポジトリにおける一つ以上のドキュメントを含むことができる。

#### 【0110】

検索エンジン125は、特定されたドキュメントに関連付けられた履歴データを獲得することができる（動作420）。上記記載のように、履歴データは異なるフォームをとることができる。例えば、履歴データは、ドキュメント・インセプション日付に関連するデータ、ドキュメント・コンテンツの更新/変更、クエリー分析、リンクに基づき基準、アンカー・テキスト、トラフィック、ユーザー振る舞い、ドメイン関連の情報、ランキング履歴、ユーザーによって保存/生成されたデータ（例えば、ブックマークおよび/またはお気に入り）、特有なワード、バイグラム、アンカー・テキスト中のフレーズ、独立したピアのリンケージ、および/またはドキュメント・トピックを含むことができる。検索エンジン125は、これらの種類の履歴データのうち一つまたは組み合わせを獲得することができる。

10

#### 【0111】

検索エンジン125は、さらに、履歴データの少なくとも一部に基づいて特定されたドキュメントをスコア付けすることができる（動作430）。特定されたドキュメントがサーチクエリに関連付けられる際、検索エンジン125はまた、例えばそれらがどの程度サーチクエリに関連するかに基づいて、ドキュメントについての関連スコアも生成することができる。検索エンジン125は、ドキュメントについての全体スコアを獲得するために、履歴スコアと関連スコアを組み合わせることができる。スコアを組み合わせる代わりに、検索エンジン125は、履歴データに基づいてドキュメントに対する関連スコアを変更することができ、この方法によってスコアを上げたりまたは下げたり、場合によっては、スコア等位にしておく。あるいは、検索エンジン125は、関連スコアを生成することなしに履歴データに基づいてドキュメントをスコア付けすることができる。いずれにしても、検索エンジン125は、履歴データのタイプのうち一つまたは組み合わせを使用してドキュメントをスコア付けすることができる。

20

#### 【0112】

特定されたドキュメントがサーチクエリに関連付けられる際、検索エンジン125はまた、スコア付けされたドキュメントからの検索結果を形成することもできる。例えば、検索エンジン125は、そのスコアに基づいてドキュメントをソートすることができる。検索エンジン125は、ドキュメントに対する参照を形成することができるが、ここで参照は、（実際のドキュメントに選択された際にユーザーに指令するハイパーテキスト・リンクを包含することができる）ドキュメントのタイトルと、ドキュメント由来の断片（すなわちテキストからの引用）を含むだろう。その他の実施態様において、参照は異なって形成される。検索エンジン125は、多数の最上位にスコア付けされたドキュメント（例えば、規定の数のドキュメント、閾値を越えたスコアを伴うドキュメント、すべてのドキュメント等）に対応する参照を、サーチクエリを送信したユーザーに示すことができる。

30

40

#### 【0113】

（結論）

本発明の原理にかかるシステムおよび方法は、スコア・ドキュメントに履歴データを使用することができ、また、高品質の検索結果を形成することができる。

#### 【0114】

本発明の好適な実施例についての先行する記述は、具体例および記述を提供するが、網羅的になるように意図されておらず、または、本発明を開示された明確な形式に限定しない。修正または変化は、上記技法を考慮して可能であり、もしくは、本発明の実践を取得することができる。例えば、一連の動作が図4に関して記述される一方で、動作の手順は、本発明の原理にかかるその他の実施態様において修正することができる。また、依存し

50



ない動作が、平行して実行されることができる。

【0115】

さらに、サーバ120は、一般的には、すべてではないにしても、図4の処理に関して記述する動作のほとんどを実行することを記述した。本発明の原理にかかるともう一つの実施態様において、すべての動作のうちの一つ以上は、例えば別のサーバ130および/または140、もしくはクライアント110といった、別のエンティティによって実行されることができる。

【0116】

当業者にとって、本発明の態様はまた、上述したように図に示した実施態様におけるソフトウェア、ファームウェア、およびハードウェアの中で、多くの異なる形式で実施できるのは、明白であろう。本発明の原理にかかると態様を実行するのに使用される実際のソフトウェア・コードまたは特殊化された制御ハードウェアは、本発明において限定されるものではない。このように、この態様における操作および振る舞いは、特定のソフトウェア・コードへの参照なしに記述されていた - 当業者は、本明細書の記述に基づく態様を実施するために、ソフトウェアをデザインし且つハードウェアを制御することが可能となるだろう。

10

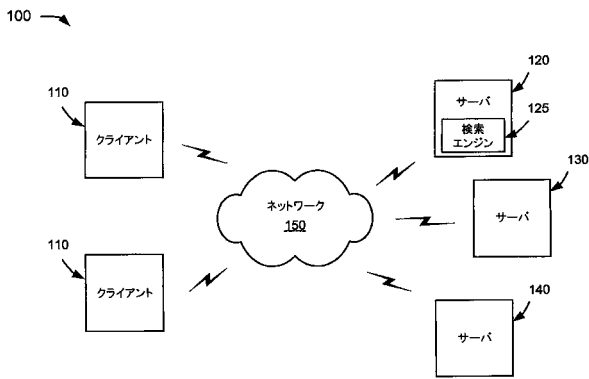
【符号の説明】

【0117】

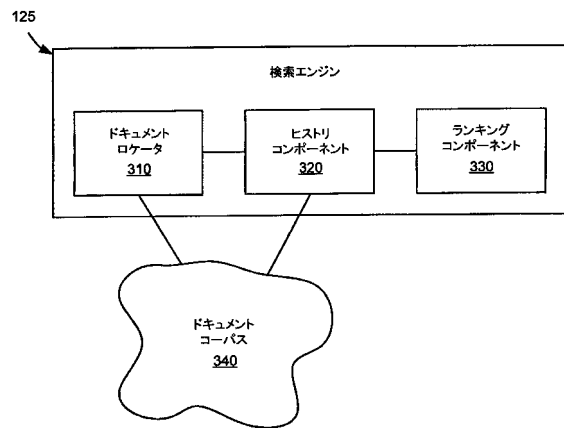
- 125 検索エンジン
- 310 ドキュメントロケータ
- 320 ヒストリコンポーネント
- 330 ランキングコンポーネント
- 340 ドキュメントコーパス

20

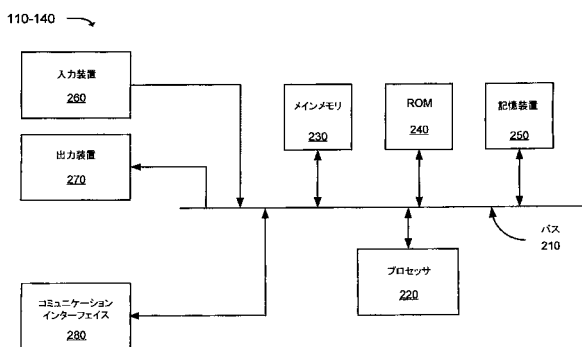
【図1】



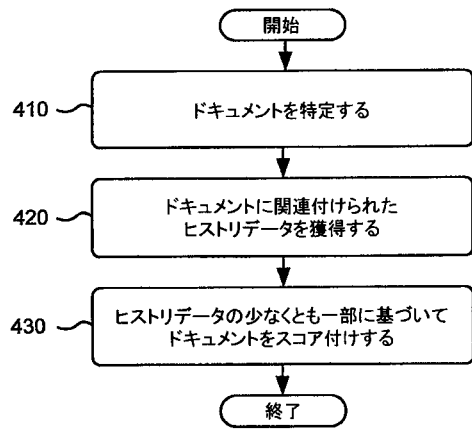
【図3】



【図2】



【 図 4 】



## フロントページの続き

- (72)発明者 ディーン・ジェフリー  
アメリカ合衆国 カリフォルニア 94303 パロアルト スtockトンプレイス 3179
- (72)発明者 ハーアー・ポール  
アメリカ合衆国 カリフォルニア 94114 サンフランシスコ 22番ストリート 4222
- (72)発明者 ヘンジンガー・モニカ  
スイス連邦 コルセオークス 1802 チェミンデュチャノ 18
- (72)発明者 ホエルツル・ウース  
アメリカ合衆国 カリフォルニア 94306 パロアルト コーネルストリート 2298
- (72)発明者 ローレンス・スティープ  
アメリカ合衆国 カリフォルニア 94040 マウンテンビュー 2400 ウェストエルカミノリアル アパートメント #204
- (72)発明者 ブフェレガー・カール  
アメリカ合衆国 カリフォルニア 94040 マウンテンビュー デルメディオアベニュー 450
- (72)発明者 セルチノグル・オルカン  
アメリカ合衆国 カリフォルニア 94040 マウンテンビュー 2400 ウェストエルカミノリアル アパートメント #716
- (72)発明者 トン・シモン  
アメリカ合衆国 カリフォルニア 94040 マウンテンビュー 541 デルメディオアベニュー アパートメント #319

審査官 鈴木 和樹

- (56)参考文献 特開2001-005705(JP,A)  
特開2003-122669(JP,A)  
特開平10-247201(JP,A)  
特開2001-209570(JP,A)  
特開2002-140361(JP,A)  
特開2002-183216(JP,A)  
特開2002-007801(JP,A)  
特表2002-507794(JP,A)  
米国特許出願公開第2002/0123988(US,A1)  
米国特許第06421675(US,B1)  
特開2000-339316(JP,A)  
特開2003-228581(JP,A)  
灰原清太郎、外1名、サイト評価情報を用いたWWW検索表示方法、情報処理学会論文誌、日本、社団法人情報処理学会、1999年 8月15日、第40巻、第SIG6(TOD3)号、p.22-30

- (58)調査した分野(Int.Cl., DB名)  
G06F 17/30