

(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2005-92881
(P2005-92881A)

(43) 公開日 平成17年4月7日(2005.4.7)

(51) Int.C1.⁷

F 1

G06F 17/30
G06F 12/00
G06F 13/00G06F 17/30 220Z
G06F 17/30 419B
G06F 12/00 546B
G06F 13/00 540F

テーマコード(参考)

5B075
5B082

審査請求 未請求 請求項の数 40 O L (全 28 頁)

(21) 出願番号 特願2004-265877 (P2004-265877)
 (22) 出願日 平成16年9月13日 (2004.9.13)
 (31) 優先権主張番号 10/663,933
 (32) 優先日 平成15年9月16日 (2003.9.16)
 (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 9805
 2-6399 レッドmond ワン マイ
 クロソフト ウェイ
 (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 マーク エー. ナジョーク
 アメリカ合衆国 94303 カリフォル
 ニア州 パロ アルト チューリップ レ
 ーン 67
 F ターム(参考) 5B075 ND35 ND36
 5B082 HA00 HA08

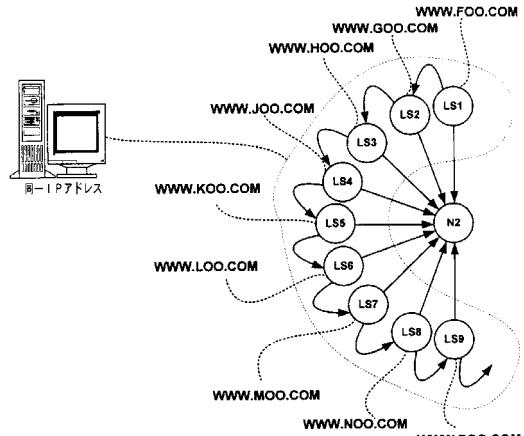
(54) 【発明の名称】構造的に相互関係のある情報に基づいて文書をランク付けするための改善されたシステムおよび方法

(57) 【要約】

【課題】 ネポティスティックリンクの影響を受け難いように、ハイパーリンク情報に基づいてウェブページをランク付けするシステムおよび方法を提供する。

【解決手段】 良質の問合せ結果を返すウェブ検索サービスが提供される。ターゲットページのスコアをつり上げるだけの目的で人為的に生成されるウェブページに対する、PageRank等の従来のランク付けアルゴリズムの脆弱性が解決される。直観的には、ランダムジャンプをすると、少数のページを有するウェブサーバ上で特定のページに到達するよりも多数のページを有するウェブサーバ上で特定のページに到達するほうが可能性が低い。即ち後者のようなページが他のページにリンク(他のページを推奨)することによって、当該他のページに及ぼす影響のほうが低い。そこで各ウェブページではなく各ウェブサーバに最低保証スコアを割り当て、そのウェブサーバ上のすべてのページ間で分配する。

【選択図】 図3G



【特許請求の範囲】**【請求項 1】**

複数の構造的にリンクされた文書のうちの1つの文書にスコアを割り当てる方法において、該文書は、(A)同じシンボリックホスト名を有する複数のウェブページを含むサーバ、(B)同じドメインに関連する複数のウェブページを含むサーバ、および(C)同じIPアドレスに関連する複数のウェブページを有するサーバ、のうちの少なくとも1つによって定義されるウェブサーバ上に位置し、該文書は、いずれに位置する場合であれ、前記複数の構造的にリンクされた文書のうちの少なくとも1つの他の文書からの少なくとも1つのバックリンクを有し、

前記ウェブサーバ上に位置する文書の数に反比例するスコアを前記文書に割り当てる
ことを特徴とする文書にスコアを割り当てる方法。

10

【請求項 2】

前記少なくとも1つの他の文書の数に比例するスコアを前記文書に割り当てる
こと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】

前記少なくとも1つの他の文書のうちの少なくとも1つに割り当てられた少なくとも1つのスコアに比例するスコアを割り当てる
ことをさらに含むことを特徴とする請求項1に記載の方法。

20

【請求項 4】

(A)前記少なくとも1つの他の文書の数、および(B)前記少なくとも1つの他の文書のうちの少なくとも1つに割り当てられた少なくとも1つのスコア、に比例するスコアを割り当てる
こと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 5】

前記少なくとも1つの他の文書のうちの少なくとも1つの文書のアウトリンク(outlinks)の数に反比例するスコアを前記文書に割り当てる
ことをさらに含むことを特徴とする請求項2に記載の方法。

30

【請求項 6】

前記割り当てることが、前記文書と同じドメインに位置する文書の数に反比例するスコアを前記文書に割り当てるること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】

前記割り当てることが、前記文書と同じシンボリックホスト名を有する文書の数に反比例するスコアを前記文書に割り当てるること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】

前記割り当てることが、前記文書と同じインターネットプロトコル(IP)アドレスに関連する文書の数に反比例するスコアを前記文書に割り当てるることを含むことを特徴とする請求項1に記載の方法。

【請求項 9】

前記第1の文書にリンクしている前記少なくとも1つの他の文書のスコアの合計を取ることに基づくスコアを前記文書に割り当てる

40

こと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 10】

前記複数の構造的にリンクされた文書が、ハイパーリンクを有するウェブページであり、前記文書がウェブページであることを特徴とする請求項1に記載の方法。

【請求項 11】

ウェブ検索サービスのコンポーネントに前記文書のスコアを出力すること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 12】

好みい文書の集合に、平均最低スコアよりも高いスコアを割り当てるなどをさらに含むことを特徴とする請求項1に記載の方法。

50

むことを特徴とする請求項 1 に記載の方法。

【請求項 1 3】

前記好み文書の集合が、ニールセン視聴率、人間により割り当てられた格付け、ISP プロキシログから抽出されたウェブページ利用パターン、検索エンジンから抽出されたウェブページ利用パターン、およびユーザ選好に従って指定された文書の少なくとも 1 つに基づくことを特徴とする請求項 1 2 に記載の方法。

【請求項 1 4】

第 2 のスコアリング技法に基づいて前記文書のスコアを変更すること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 1 5】

異常な結果を発見するために前記スコアを第 2 のスコアリング技法と比較すること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 1 6】

請求項 1 に記載の方法を実施するためのコンピュータ実行可能命令を有するコンピュータ実行可能モジュールを備えたことを特徴とするアプリケーションプログラミングインターフェース。

【請求項 1 7】

請求項 1 に記載の方法を実施する手段を備えたことを特徴とするコンピューティングデバイス。

【請求項 1 8】

請求項 1 に記載の方法を実行するためのコンピュータ実行可能命令を伝えることを特徴とする変調データ信号。

【請求項 1 9】

複数の構造的にリンクされた文書のうちの 1 つの文書にスコアを割り当てる方法において、該文書は、(A) 同じシンボリックホスト名を有する複数のウェブページを含むサーバ、(B) 同じドメインに関連する複数のウェブページを含むサーバ、および(C) 同じ IP アドレスに関連する複数のウェブページを有するサーバ、のうちの少なくとも 1 つによって定義されるウェブサーバ上に位置し、該文書は、前記複数の構造的にリンクされた文書のうちの少なくとも 1 つのソース文書からの少なくとも 1 つのバックリンクを有し、前記文書のスコアが、前記少なくとも 1 つのソース文書のうちの少なくとも 1 つに関連する少なくとも 1 つのスコアに比例して計算され、前記スコアが、前記ウェブサーバ上に位置する前記少なくとも 1 つのソース文書の数に反比例して計算される

ことを特徴とする文書にスコアを割り当てる方法。

【請求項 2 0】

前記スコアが、同じウェブサーバ上に位置する前記少なくとも 1 つのソース文書の数に反比例して計算されることを特徴とする請求項 1 9 に記載の方法。

【請求項 2 1】

前記スコアが、同じシンボリックホスト名を有する前記少なくとも 1 つのソース文書の数に反比例して計算されることを特徴とする請求項 2 0 に記載の方法。

【請求項 2 2】

前記スコアが、同じドメインに関連する前記少なくとも 1 つのソース文書の数に反比例して計算されることを特徴とする請求項 2 0 に記載の方法。

【請求項 2 3】

前記スコアが、同じインターネットプロトコル(IP)アドレスに関連する前記少なくとも 1 つのソース文書の数に反比例して計算されることを特徴とする請求項 2 0 に記載の方法。

【請求項 2 4】

前記複数の構造的にリンクされた文書が、ハイパーリンクを有するウェブページであり、前記文書がウェブページであることを特徴とする請求項 1 9 に記載の方法。

【請求項 2 5】

10

20

30

40

50

請求項 1 9 に記載の方法を実施するためのコンピュータ実行可能命令を有するコンピュータ実行可能モジュールを備えたことを特徴とするアプリケーションプログラミングインターフェース。

【請求項 2 6】

請求項 1 9 に記載の方法を実施する手段を備えたことを特徴とするコンピューティングデバイス。

【請求項 2 7】

請求項 1 9 に記載の方法を実行するためのコンピュータ実行可能命令を伝えることを特徴とする変調データ信号。

【請求項 2 8】

問合せ要求に基づいてウェブページ情報および関連するスコアのデータベースに問い合わせる問合せメカニズムを備えた検索エンジンとともに使用するためのアプリケーションプログラミングインターフェースであって、それによって、該データベースから取得された結果が各結果に関連するスコアによって順序付けられ、該スコアが前記問合せを満たすウェブページの品質を反映する、アプリケーションプログラミングインターフェースと、

前記ウェブページ情報に関連するスコアを生成するスコアリングオブジェクトであって、少なくとも 1 つの対応するソースウェブページへの少なくとも 1 つのバックリンクを有するウェブページに対して、該スコアリングオブジェクトが、該少なくとも 1 つの対応するソースウェブページのうちの少なくとも 1 つに関連する少なくとも 1 つのスコアに比例するスコアを該ウェブページに割り当て、該スコアが、同じウェブサーバ上に位置する前記少なくとも 1 つの対応するソースウェブページの数に反比例して計算される、スコアリングオブジェクトと

を備えたことを特徴とするサーバオブジェクト。

【請求項 2 9】

前記スコアリングオブジェクトが、前記問合せメカニズムとは独立に動作することを特徴とする請求項 2 8 に記載のサーバオブジェクト。

【請求項 3 0】

ウェブサーバが、共通のシンボリックホスト名に基づいて定義されることを特徴とする請求項 2 8 に記載のサーバオブジェクト。

【請求項 3 1】

ウェブサーバが、共通ドメインに基づいて定義されることを特徴とする請求項 2 8 に記載のサーバオブジェクト。

【請求項 3 2】

ウェブサーバが、共通のインターネットプロトコル (IP) アドレスに基づいて定義されることを特徴とする請求項 2 8 に記載のサーバオブジェクト。

【請求項 3 3】

複数の構造的にリンクされた文書のうちの 1 つの文書にスコアを割り当てるためのコンピュータ実行可能命令を含むコンピュータ実行可能モジュールを備えたコンピュータ可読媒体において、該文書は、ウェブサーバ上に位置し、前記複数の構造的にリンクされた文書のうちの少なくとも 1 つの他の文書からの少なくとも 1 つのバックリンクを有し、該モジュールが、

前記ウェブサーバ上に位置する文書の数に反比例するスコアを前記文書に割り当てる手段

を備えたことを特徴とするコンピュータ可読媒体。

【請求項 3 4】

前記少なくとも 1 つの他の文書の数に比例するスコアを前記文書に割り当てる手段をさらに含むことを特徴とする請求項 3 3 に記載のコンピュータ可読媒体。

【請求項 3 5】

前記少なくとも 1 つの他の文書のうちの少なくとも 1 つに割り当てられた少なくとも 1 つのスコアに比例するスコアを割り当てる手段

10

20

30

40

50

をさらに含むことを特徴とする請求項 3 3 に記載のコンピュータ可読媒体。

【請求項 3 6】

(A) 前記少なくとも 1 つの他の文書の数、および (B) 前記少なくとも 1 つの他の文書のうちの少なくとも 1 つに割り当てられた少なくとも 1 つのスコア、に比例するスコアを割り当てる手段

をさらに含むことを特徴とする請求項 3 3 に記載のコンピュータ可読媒体。

【請求項 3 7】

前記少なくとも 1 つの他の文書のうちの少なくとも 1 つの文書のアウトリンクの数に反比例するスコアを前記文書に割り当てる手段

をさらに含むことを特徴とする請求項 3 4 に記載のコンピュータ可読媒体。

10

【請求項 3 8】

前記割り当てる手段が、前記文書と同じシンボリックホスト名を有するウェブサーバ上に位置する文書の数に反比例するスコアを前記文書に割り当てる手段を含むことを特徴とする請求項 3 3 に記載のコンピュータ可読媒体。

【請求項 3 9】

前記割り当てる手段が、前記文書と同じドメインに位置する文書の数に反比例するスコアを前記文書に割り当てる手段を含むことを特徴とする請求項 3 3 に記載のコンピュータ可読媒体。

【請求項 4 0】

前記割り当てる手段が、前記文書と同じインターネットプロトコル (IP) アドレスに関連する文書の数に反比例するスコアを前記文書に割り当てる手段を含むことを特徴とする請求項 3 3 に記載のコンピュータ可読媒体。

20

【発明の詳細な説明】

【技術分野】

【0 0 0 1】

本発明は、構造的に相互関係のある情報に基づく文書のランク付けに関する。特に、本発明は、縁故主義的な (nepotistic)、つまり利己的な (self-serving) リンクに抵抗し、その影響を受け難くしたやり方での、ハイパーリンク情報に基づくウェブページのランク付けに関する。

【背景技術】

【0 0 0 2】

ウェブ検索サービスは、例えばユーザまたはアプリケーションから問合せを受け付け、その問合せを満たす結果 (例えば文書または文書へのリンク) のリストを返す。なお、本明細書で使用される場合の「文書」という用語は、検索することができる任意のコンテンツを指し、ワードプロセッサ文書やウェブページのようなファイルに限定されるものと解釈されてはならない。満足できる見聞を提供するため、この結果リストは、ユーザに最も関連性のある文書が最初に現れるべきであることを考慮して順序付けられるべきである。現在、文書をランク付けするための多数のアルゴリズムが存在し、ほとんどのウェブ検索エンジンは、このようなアルゴリズムのいくつかを使用して、さまざまなランク付けアルゴリズムによって割り当てられたランクの組合せに基づいて問合せの結果をランク付けしている。

30

【0 0 0 3】

従来の多数のランク付けアルゴリズムは、問合せ依存型 (動的ともいう) または問合せ非依存型 (静的ともいう) のいずれであるかに基づいて分類することができる。問合せ依存型ランク付けアルゴリズムは問合せ中の用語を使用するが、問合せ非依存型ランク付けアルゴリズムは問合せ中の用語を使用しない。すなわち、問合せ非依存型ランク付けアルゴリズムはウェブ上の各文書に品質スコア (得点) を割り当てる。したがって、問合せ非依存型ランク付けアルゴリズムは、前もって都合よく実行することができ、問合せが提出されるたびに再実行する必要がない。

40

【0 0 0 4】

50

ランク付けアルゴリズムは、コンテンツベース、使用量ベース、およびリンクベースの各ランク付けアルゴリズムに大別することもできる。コンテンツベースのランク付けアルゴリズムは、文書中の単語を用いて文書をランク付けする（例えば、問合せ依存型コンテンツベースランク付けアルゴリズムでは、文書の前の方に、または大きいフォントや太字のフォントで、問合せ用語を含む文書に、より高いスコアを与えるであろう）。使用量ベースのランク付けアルゴリズムは、ウェブページが何回閲覧されたかの見積値に基づいてウェブページをランク付けする。このような見積値は、ウェブプロキシログ（Web proxy logs）を調べることによって、または検索エンジンの結果ページでのクリックスルー（click-throughs）を監視することによって、生成することができる。最後に、リンクベースのランク付けアルゴリズムは、ウェブページ間のハイパーリンクを用いてウェブページをランク付けする。10

【0005】

例えば、非常に単純な静的リンクベースランク付けアルゴリズムでは、あるページを指している他のページからのリンクはそのページを「推奨（endorse）」していると考えて、各ウェブページに、そのページを指しているリンク（「バックリンク（backlink）」）の数に比例するスコアを割り当てるであろう。例えば、図1Aに示すように、ウェブページA、B、CおよびDがそれぞれ、ウェブページ内の黒い矩形で表されているように、他のウェブページへの3個のリンク（「アウトリンク（outlink）」）を含むとする。この例では、静的リンクベースランク付けアルゴリズムを用いると、ページDはページCよりも低いスコアを受ける。というのは、ページDにはバックリンクがないのに対して、ページCはページBからの1個のバックリンクL2およびページAからの1個のバックリンクL1を有するからである。なお、ページA、B、CおよびDをダウンロードしてしまえば、それぞれが何個のアウトリンクを有するか、およびそれらがどこへリンクしているかは、そのページを読むことができるので確定的であるが、アルゴリズムにおいて考慮に入れることができない未知のロケーション（記憶場所）からの、例えばバックリンクL1のような未知のバックリンクがまだ存在する可能性がある。この単純な手法の主たる欠点は、それぞれの「推奨（endorsement）」が同等に扱われることであるが、このことはこの手法を利用しやすいシステムにしている。20

【0006】

PageRank（ページランク）は、群を抜いて、最もよく知られている問合せ非依存リンクベースランク付けアルゴリズムであるので、その原理を本明細書に記載する。PageRankは、図1Aの素朴な静的リンクベースシステムの原理を基に、システムに再帰的レイヤを追加することによって作られている。図1Bに示すように、4個のウェブページが例示され、PageRankが直観的に示されている。PageRankでは、推奨される側のページ（被推奨側ページ）にスコアを割り当てる際に、推奨する側のページ（推奨側ページ）のスコアが考慮に入れられる。したがって、ウェブページGに与えられるスコアに対して、ウェブページE（そのスコアは100）からの推奨の重みのほうが、ウェブページF（そのスコアは9）からの推奨よりもはるかに大きい影響を及ぼす。直観的には、推奨側ページのスコアが、その被推奨側の間に分割されると考えができる。30

【0007】

数学的には、PageRankアルゴリズムの直観的概念は次のように説明することができる。既知のウェブページの集合およびそれらの間のリンクが、頂点集合V（各頂点がウェブページに対応する）および辺集合E（各辺（u, v）がページuからページvへのハイパーリンクに対応する）を有するグラフを引き起こすと仮定する。|V|で集合Vのサイズを表し、O(u)で頂点uの出次数（すなわち、ウェブページuに埋め込まれているハイパーリンクの数）を表し、pで0と1の間の数（例えば0.15）を表す。ウェブページvのPageRank R(v)は次のように定義される。40

【0008】

【数1】

$$R(v) = \frac{p}{|V|} + (1-p) \sum_{(u,v) \in E} \frac{R(u)}{O(u)}$$

【0009】

PageRank公式は、次のように説明されることが多い。ウェブ上でランダムウォーク(迷い歩き)しているウェブサーファを考える。ウォークにおける各ステップごとに、サーファは、次のアルゴリズムを用いて、あるウェブページから別のウェブページへ移動する：ある確率 p で、サーファは一様ランダムにウェブページを選択してそこへジャンプする；それ以外の場合、サーファは現在のページ内の出ハイパリンクのうちの1つを一様ランダムに選択してそれをたどる。この比喩のため、数 p は「ジャンプ確率」(すなわち、サーファが完全にランダムなページにジャンプする確率)と呼ばれることがある。ウェブサーファが確率 p でジャンプし、 $|V|$ 個のウェブページがある場合、ある特定のページにジャンプする確率は $p / |V|$ である。ジャンプによっていずれかのページに到達することができるので、各ページには少なくとも $p / |V|$ のスコアが保証される。

10

【0010】

PageRankスコアは、問合せ結果をランク付けするために用いることができる。他のすべてのファクタが同一である場合、PageRankを用いた検索エンジンは、PageRankスコアの低いページよりもPageRankスコアの高いページの方を上位にランク付けする。検索エンジンのほとんどのユーザは最初のいくつかの結果しか調べないので、商用ウェブサイトの運営者は、自分のサイトへのリンクが結果リストの前のほうに現れること、すなわち自分のウェブページが高いPageRankスコアを受けることに利害関係がある。すなわち、商用ウェブサイト運営者には、自分のウェブサイト上のページのPageRankスコアを人為的に上げようとする動機付けがある。

20

【0011】

PageRank公式を解析することにより、ウェブページ v のPageRankスコアを上げる1つの方法として、多くの他のページをそのページにリンクさせればよいことが明らかとなる。この理由は、ウェブページが自分のアウトリンクを通じて他のウェブページを推奨することができるという考え方がPageRankの核心にあるからである。 v にリンクしているすべてのページが低いPageRankスコアを有する場合、それぞれの個々のページはわずかしか寄与しないであろう。しかし、あらゆるページが最低限のPageRankスコアとして $p / |V|$ を有することが保証されているので、多くのこのような低品質のページからのリンクであっても、合計はかなりの寄与になることがある。これは、PageRankアルゴリズムの脆弱性を露呈する。

30

【0012】

実際、PageRankのこの脆弱性は、自分のメインホームページを「推奨」することのみを目的とする非常に多数のページを含むウェブサイトによって活用されている。通常、これらの推奨側ページは、推奨されるページへのリンクと、別の推奨側ページへの別のリンクとを含む。すべての推奨側ページはオンザフライで自動的に作成される。したがって、ウェブクローラ(We bページ巡回プログラム)が、たまたまいずれかの推奨側ページに遭遇すると、(推奨側ページが他の推奨側ページにリンクしているので)さらに多くの推奨側ページをダウンロードし続けることにより、それらのページを多数蓄積することになる。この多数のページは、すべて単一のページを推奨しているので、推奨されているページのPageRankスコアを人為的に上げる。PageRankスコアを人為的に上げるために用いられる技法は、俗に「リンクスパミング(link spamming)」あるいは「リンクスパム(link spam)」と呼ばれている。

40

【0013】

また、個別化(パーソナライズ)されたPageRankスコアが、特定の観点からのウェブの一覧(view)を作り出すことができることも知られている。例えば、ユーザのブックマークを取り、ユーザのブックマーク内にあるページのPageRankスコアをつ

50

り上げることにより、個別化された PageRank スコアリングシステムが実現される。本質的に、あるウェブページをブックマークとして指定しているユーザは、そのウェブページを、スコアリングシステムが基準にしてほしいウェブページとして暗黙のうちに推薦している。ユーザが「リンクスパム」ページをブックマークとして選択することはまれであり、まして多数の「リンクスパム」ページを選択することはないが、個別化された PageRank の考え方は、リンクスパミングの問題に明示的に対処するものではない。というのは、各リンクスパムウェブページに関連する最低スコアが依然として存在するからである。

【0014】

したがって、その基本的考え方は正しいが、PageRank の結果は、ネポティスティックリンク（縁故主義のリンク）により導入される妨害を受けやすい。すなわち、推薦側や被推薦側の真の価値を考慮せずに、自己推薦および昇格の目的の一群のページを作成されてしまうことがあり得る。PageRank スコアに関してリンクスパムの問題が存在することは既知であるが、当技術分野においてその解決法はいまだ見出されていない。

【発明の開示】

【発明が解決しようとする課題】

【0015】

したがって、問合せ非依存でリンクベースのランク付けアルゴリズムの改善が必要とされている。より具体的には、ネポティスティックリンクの影響を大幅に低減する改善されたランク付けのシステムおよび方法が必要とされている。さらに、ターゲットとなる被推薦側ウェブページに関連する PageRank スコアを人為的に上り上げる目的で一群の自己推薦ウェブページを作成しようとするリンクスパマ(link spammer)のやる気を低減する改善されたランク付けのシステムおよび方法が必要とされている。

【課題を解決するための手段】

【0016】

従来技術の上述の確認された短所に鑑みて、本発明は、文書の構造的相互関係に関する情報に基づいて文書をランク付けするシステムおよび方法を提供する。本発明のシステムおよび方法は、ネポティスティックリンクの影響を受けにくいように、ハイパーリンク情報に基づいてウェブページをランク付けするために用いることができる。種々の実施形態では、本発明は、良質の問合せ結果を返すためのウェブ検索サービスにおいて実施される。本発明は、ターゲットページのスコア（得点）を上り上げるだけの目的で人為的に生成されるウェブページに対する、PageRank 等の従来のランク付けアルゴリズムの脆弱性を解決する。直観的には、本発明は、ランダムジャンプをすると、少数のページを有するウェブサーバ上で特定のページに到達するよりも、多数のページを有するウェブサーバ上で特定のページに到達するほうが、その可能性が低いということを認識している。このことは、そのようなページが他のページにリンクする（すなわち他のページを推薦する）ことによって当該他のページに及ぼす影響が低いことを意味する。したがって、種々の非限定的実施形態では、本発明は、各ウェブページではなく各ウェブサーバに最低保証スコア (guaranteed minimum score) を割り当てる。そして、サーバに割り当てるこの最低スコアを、そのウェブサーバ上のすべてのページ間で分割することができる。

【0017】

本発明の他の利点および特徴は以下に記載される。

本発明による改善されたランク付けアルゴリズムを提供するシステムおよび方法は、添付図面を参照してさらに説明される。

【発明を実施するための最良の形態】

【0018】

【概要】

前述のように、PageRank アルゴリズムは、ウェブ上のあらゆるページを一つの数字、すなわちそのページの PageRank に凝縮するという大胆な仕事を引き受けている。PageRank は、すべてのウェブページについて、それらの内容とは無関係に

10

20

30

40

50

、ウェブのグラフ構造におけるその位置のみに基づいて決まるグローバル（広範囲）なランディングである。

【0019】

PageRankを用いると、検索結果は、より重要で中心的なウェブページほど優先されるように順序付けられる。PageRankの背後にある直観的概念は、ウェブページ自体の外部にある情報、すなわちそれらのバックリンクを使用することである。バックリンクは、一種のピアレビューを提供する。さらに、再帰的定義により、「重要」なページからのバックリンクは、平均的リンクからのバックリンクよりも有意義であるとみなされる。

【0020】

また、個別化されたPageRankスコアが、特定の観点からウェブの一覧を作り出すことができることも知られている。これは、例えば、ユーザのブックマークを取り、ユーザのブックマーク内にあるページのPageRankスコアをつり上げることにより行われる。しかし、個別化されたPageRankは、リンクスパミングの問題に明示的に対処するものではない。というのは、各リンクスパムウェブページに関連する最低スコアが依然として存在するからである。したがって、リンクスパムは依然として、単一のウェブサーバ上で、それぞれが自己の最低PageRankスコアを有する多数のウェブページを（必要であれば自動的に）作成することができ、それらのページが、互いを、およびターゲット被推奨側ウェブページを推奨することによって、ターゲット被推奨側ウェブページのスコアを人為的につり上げる。リンクスパムが作成する多数のウェブページは通常、（A）同じシンボリックホスト名を有するか、（B）同じドメインに関連付けられるか、または（C）同じIPアドレスに関連付けられるかのいずれか1つまたは複数である。

【0021】

本発明は、次のことを認識することによって、この脆弱性を解決する。すなわち、ランダムジャンプをすると、少数のページを有するウェブサーバ上で特定のページに到達するよりも、多数のページを有するウェブサーバ上で特定のページに到達するほうが、可能性が低い。このことは、そのようなページが他のページにリンクする（すなわち他のページを推奨する）ことによって当該他のページに及ぼす影響が低いことを意味する。したがって、種々の非限定的実施形態では、本発明は、各ウェブページではなく各ウェブサーバに最低保証スコアを割り当てる。そして、この最低スコアを、そのウェブサーバ上のすべてのページ間で分割することができる。

【0022】

[例示的なネットワーク環境および分散環境]

当業者には理解されるように、本発明は、任意のコンピュータまたは他のクライアントもしくはサーバデバイスとの関連において実施することができる。これらは、コンピュータネットワークの一部として配備されても、分散コンピューティング環境に配備されてもよい。この点で、本発明は、任意個数のメモリまたはストレージユニットを有する任意のコンピュータシステムまたは環境、ならびに任意個数のストレージユニットまたはボリュームにわたって生じる任意個数のアプリケーションおよびプロセスに関連し、これらを本発明による文書をランク付けするプロセスとの関連で用いることができる。本発明は、リモートまたはローカルストレージを有するネットワーク環境または分散コンピューティング環境に配備されたサーバコンピュータおよびクライアントコンピュータを有する環境に適用することができる。また、本発明は、リモートまたはローカルサービスとの関連で情報を生成、受信および送信するためのプログラミング言語機能、解釈および実行の能力を有する独立型コンピューティングデバイスにも適用することができる。ウェブページをダウンロードし解析することは、ネットワーク環境または分散コンピューティング環境で動作するコンピューティングデバイスにとって特に関連性があるため、本発明によるランク付けのアルゴリズムおよび技法は、そのような環境において特に有効に適用することができる。

【0023】

10

20

30

40

50

分散コンピューティングは、コンピューティングデバイスおよびシステムの間の交換によって、コンピュータリソースおよびサービスの共有を行う。このようなリソースおよびサービスとしては、情報の交換、キャッシュストレージ、およびファイルのためのディスク記憶装置がある。分散コンピューティングは、ネットワーク接続を利用することにより、クライアントの集団的能力を活用して企業全体の利益となることを可能にする。この点で、さまざまなデバイスが、本発明のランク付けのアルゴリズムおよびプロセスに関わり得るアプリケーション、オブジェクトまたはリソースを有することができる。

【0024】

図2Aは、例示的なネットワーク環境または分散コンピューティング環境の概略図を示している。分散コンピューティング環境は、コンピューティングオブジェクト10a、10b等およびコンピューティングオブジェクトまたはデバイス110a、110b、110c等を備える。これらのオブジェクトは、プログラム、メソッド、データストア、プログラム可能ロジック等を備えることができる。オブジェクトは、PDA (Personal Digital Assistant:携帯情報端末)、テレビ、MP3プレーヤ、パーソナルコンピュータ等の同じまたは異なるデバイスの各部分を含み得る。各オブジェクトは、通信ネットワーク14経由で他のオブジェクトと通信することができる。このネットワーク自体が、図2Aのシステムにサービスを提供する他のコンピューティングオブジェクトおよびコンピューティングデバイスを備えてもよく、ネットワーク自体は、複数の相互接続されたネットワークを表していてもよい。本発明の一態様によれば、各オブジェクト10a、10b等または110a、110b、110c等は、本発明によるランク付けプロセスの使用を要求するために、APIを利用し得るアプリケーション、または他のオブジェクト、ソフトウェア、ファームウェアおよび/またはハードウェアを含んでもよい。

【0025】

また、理解されるように、110c等のオブジェクトは、別のコンピューティングデバイス10a、10b等または110a、110b等をホストとしてもよい。したがって、図示されている物理的環境は、接続されているデバイスをコンピュータとして示しているかもしれないが、このような図示は単なる例示であり、物理的環境は、別法として、PDA、テレビ、MP3プレーヤ等の種々のデジタルデバイスや、インターフェース、COMオブジェクト等のソフトウェアオブジェクトを備えて図示または記述されてもよい。

【0026】

分散コンピューティング環境をサポートするさまざまなシステム、コンポーネント、およびネットワーク構成がある。例えば、コンピューティングシステムは、有線または無線システムによって、ローカルネットワークまたは広域分散ネットワークによってともに接続されてもよい。現在、ネットワークの多くはインターネットに接続されている。インターネットは、広域分散コンピューティングのためのインフラストラクチャを提供し、多くの異なるネットワークを包含する。いかなるインフラストラクチャも、本発明による相互関係のあるリンクを有する文書をランク付けすることに付随してなされる例示的通信のために用いることができる。

【0027】

ホームネットワーキング環境では、電力線、データ（無線および有線の両方）、音声（例えば電話）および娯楽メディアのように、それぞれが独自のプロトコルをサポートし得る少なくとも4つの異種のネットワークトランスポート媒体がある。電灯のスイッチや家庭電化製品のようなほとんどのホーム制御デバイスは、接続のために電力線を用いることができる。データサービスは、ブロードバンド（例えば、DSL（デジタル加入者回線）またはケーブルモ뎀のいずれか）として家庭に入っている場合があり、無線（例えばHomeRFまたは802.11B）または有線（例えばHomePNA、Cat5、イーサネット（登録商標）、そして電力線さえも）のいずれかの接続を用いて家庭内でアクセス可能である。音声トラフィックは、有線（例えばCat3）または無線（例えば携帯電話）のいずれによって家庭に入るとしてもよく、Cat3配線を用いて家庭内で配信されてもよい。娯楽メディア等のグラフィックデータは、衛星またはケーブルのいずれを通じて

10

20

30

30

40

50

家庭に入ることが可能であり、通常は同軸ケーブルを用いて家庭内に配信される。I E E E 1 3 9 4 およびD V I (デジタルビジュアルインターフェース)もまた、メディアデバイスのクラスタのためのデジタル相互接続である。プロトコル標準として現れ得るこれらのネットワーク環境等のすべてが相互接続されてインターネットのようなネットワークを形成し、これがインターネット経由で外部世界に接続されてもよい。要するに、データの記憶および伝送のためのさまざまな異種のソースが存在するため、先に進むと、コンピューティングデバイスは、本発明によるランク付け技法を利用するプログラムオブジェクトに付随してアクセスまたは利用されるデータ等のデータを共有する方法を必要とする。

【0028】

インターネットは、一般的に、コンピュータネットワーキングの技術分野で周知のT C P / I P プロトコルスイートを利用するネットワークおよびゲートウェイの集まりを指す。T C P / I P とは、「Transmission Control Protocol/Internet Protocol」(通信制御プロトコル/インターネットプロトコル)の頭字語である。インターネットは、ユーザがネットワークを通じて情報をやりとりし共有することを可能にするネットワーキングプロトコルを実行するコンピュータによって相互接続された、地理的に分散したリモートコンピュータネットワークのシステムとして記述することができる。このような広範囲の情報共有のために、インターネットのようなリモートネットワークはこれまで、オープンシステムへと全般的に進化してきた。オープンシステムのために、開発者は、本質的に制限なしに、特殊なオペレーションやサービスを実行するためのソフトウェアアプリケーションを設計することができる。

【0029】

こうして、ネットワークインフラストラクチャは、クライアント/サーバ、ピアツーピア(peer-to-peer)、またはハイブリッドアーキテクチャのような多数のネットワークトポロジーを可能にする。「クライアント」は、関係していない別のクラスまたはグループのサービスを使用するクラスまたはグループのメンバである。したがって、コンピューティングにおいては、クライアントは、別のプログラムによって提供されるサービスを要求するプロセス(すなわち、おおよそ命令またはタスクの集合)である。クライアントプロセスは、他のプログラムまたはサービス自体に関する作業の詳細を「知る」ことを必要とせずに、要求したサービスを利用する。クライアント/サーバ型アーキテクチャ、特にネットワークシステムでは、クライアントは通常、別のコンピュータ(例えばサーバ)によって提供される共有ネットワーキングリソースにアクセスするコンピュータである。図2Aの例では、コンピュータ110a、110b等はクライアントとみなすことができ、コンピュータ10a、10b等はサーバとみなすことができる。ここでサーバ10a、10b等において複製される。ただし、いかなるコンピュータも、状況に応じてクライアント、サーバ、またはその両方とみなすことができる。これらのコンピューティングデバイスのいずれも、本発明のランク付け技法に関わり得るデータを処理し、またはサービスもしくはタスクを要求することができる。

【0030】

サーバは通常、インターネットのようなリモートまたはローカルネットワークを通じてアクセス可能なりモートコンピュータシステムである。クライアントプロセスは第1のコンピュータシステムにおいてアクティブであり、サーバプロセスは第2のコンピュータシステムにおいてアクティブであるとしてよい。これらが互いに通信媒体を通じて通信することにより、分散された機能を提供し、複数のクライアントがサーバの情報収集能力を利用することが可能となる。本発明のランク付け技法に従って利用されるいかなるソフトウェアオブジェクトも、複数のコンピューティングデバイスまたはオブジェクトにわたって分散されてよい。

【0031】

クライアントおよびサーバは、プロトコルレイヤによって提供される機能を利用して互いに通信する。例えば、HyperText Transfer Protocol (H T T P:ハイパーテキスト転送

10

20

30

40

50

プロトコル)は、ワールドワイドウェブ(WWW)(あるいは単に「ウェブ」ともいう)と関連して使用される一般的なプロトコルである。通常、インターネットプロトコル(IP)アドレス等のコンピュータネットワークアドレスや、ユニバーサルリソースロケータ(URL)等の他の参照を用いて、サーバまたはクライアントコンピュータを互いに識別することができる。ネットワークアドレスは、URLアドレスとも呼ぶことができる。通信は、通信媒体を通じて提供することができる。例えば、クライアントおよびサーバは、大容量通信用のTCP/IPコネクション経由で互いに接続されることが可能である。

【0032】

このように、図2Aは、サーバがネットワーク/バス経由でクライアントコンピュータと通信する、例示的なネットワーク環境または分散環境を示しており、本発明はこの環境において用いることができる。さらに詳細には、本発明によれば、いくつかのサーバ10a、10b等が、通信ネットワーク/バス14(これは、LAN、WAN、インターネット、インターネット等でよい)経由で、いくつかのクライアントまたはリモートコンピューティングデバイス110a、110b、110c、110d、110e等(例えば、ポータブルコンピュータ、ハンドヘルドコンピュータ、シンクライアント、ネットワーク家庭電化製品、またはVCR、TV、オープン、電灯、ヒータ等の他のデバイス)と相互接続される。このように、本発明は、構造的に相互関係のあるリンクを有する文書のランク付けを実施するのが望ましいかなるコンピューティングデバイスにも適用することができると考えられる。

【0033】

例えば、通信ネットワーク/バス14がインターネットであるネットワーク環境において、サーバ10a、10b等はウェブサーバであることが可能であり、クライアント110a、110b、110c、110d、110e等は、HTTP等のいくつかの既知のプロトコルのいずれかによりこのウェブサーバと通信する。サーバ10a、10b等は、分散コンピューティング環境に特徴的なように、クライアント110a、110b、110c、110d、110e等としての役割を果たすこともある。

【0034】

通信は、適宜、有線でも無線でもよい。クライアントデバイス110a、110b、110c、110d、110e等が通信を行う場合、通信ネットワーク/バス14経由であってもそうでなくともよく、それらに関連する独立の通信を有してもよい。例えば、TVまたはVCRの場合、その制御にネットワークの側面がある場合もない場合もある。各クライアントコンピュータ110a、110b、110c、110d、110e等およびサーバコンピュータ10a、10b等は、種々のアプリケーションプログラムモジュールまたはオブジェクト135および種々のタイプのストレージ要素またはオブジェクトへのコネクションまたはアクセスを備えてもよい。これらにより、ファイルまたはデータストリームが記憶され、あるいはファイルまたはデータストリームの一部がダウンロード、伝送または移動されてもよい。コンピュータ10a、10b、110a、110b等のいずれか1つまたは複数は、本発明により処理されるデータを記憶するためのデータベース20等の記憶要素(例えば、データベースまたはメモリ20)の維持および更新を担当してもよい。このように、本発明は、コンピュータネットワーク/バス14にアクセスしてやりとりすることができるクライアントコンピュータ110a、110b等と、クライアントコンピュータ110a、110b等のデバイスおよびデータベース20とやりとりすることができるサーバコンピュータ10a、10b等とを有するコンピュータネットワーク環境において利用することができる。

【0035】

[例示的コンピューティングデバイス]

図2Bおよび以下の説明は、本発明を実施し得る好適なコンピューティング環境の簡単な概要を提供するためのものである。しかし、理解されるように、すべての種類のハンドヘルド型、ポータブル型等のコンピューティングデバイスおよびコンピューティングオブジェクトが、本発明との関連で使用するために、すなわち、コンピューティング環境にお

10

20

30

40

50

いてウェブページ等の構造的に相互関係のある文書と接するいかなる場合にも、考えられる。以下では汎用コンピュータについて説明するが、これは単なる一例にすぎず、本発明は、ネットワーク／バス相互運用性および相互作用を有するシンクライアントを用いて実施してもよい。したがって、本発明は、わずかな、あるいは最小限のクライアントリソースしか関与しないネットワークホストサービスの環境において実施することができる。そのような環境は、例えば、家庭電化製品内にあるオブジェクトのように、クライアントデバイスが単にネットワーク／バスへのインターフェースとしてのみ働くようなネットワーク環境である。本質的に、データを記憶しておくことができる場所や、そこからデータを取り出すことができる場所、または別のコンピュータへデータを送信することができる場所はいずれも、本発明によるランク付け技法の動作にとって望ましい、あるいは好適な環境である。

【0036】

必須要件ではないが、本発明は、デバイスまたはオブジェクトに対するサービスの開発者が使用するためにオペレーティングシステムを通じて実施されること、および／または本発明のランク付け技法との関連で動作するアプリケーションソフトウェア内に含まれることが可能である。ソフトウェアは、クライアントワークステーション、サーバ等のデバイスのような1つまたは複数の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等のコンピュータ実行可能命令との一般的な関連で記述することができる。一般的に、プログラムモジュールは、特定のタスクを実行し、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、データ構造等を含む。通常、プログラムモジュールの機能は、種々の実施形態において望み通りに組み合わせ、または分散させることができる。さらに、当業者には理解されるように、本発明は、他のコンピュータシステム構成およびプロトコルとともに実施してもよい。本発明とともに使用するのに好適であり得る他の周知のコンピューティングシステム、環境、および／または構成としては、以下のものに限定されないが、パーソナルコンピュータ（PC）、現金自動預け払い機（ATM）、サーバコンピュータ、ハンドヘルド型またはラップトップ型デバイス、マルチプロセッサシステム、マイクロプロセッサ方式のシステム、プログラム可能な家庭電化製品、ネットワークPC、電気器具、照明、環境制御要素、ミニコンピュータ、メインフレームコンピュータ（大型汎用コンピュータ）等がある。また、本発明は、通信ネットワーク／バス等のデータ伝送媒体を通じてリンクされたリモート処理デバイスによってタスクが実行されるような分散コンピューティング環境において実施してもよい。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はメモリ記憶デバイスを含むローカルおよびリモートの両方のコンピュータ記憶媒体に配置されてもよく、その場合、クライアントノードはサーバノードとして動作してもよい。

【0037】

このように、図2Bは、本発明を実施し得る好適なコンピューティングシステム環境100の一例を示しているが、上記で明らかに、コンピューティングシステム環境100は好適なコンピューティング環境の単なる一例であり、本発明の使用および機能の範囲に関するいかなる限定を示唆することも意図していない。また、コンピューティング環境100は、例示的オペレーティング環境100に示されるいかなるコンポーネントまたはその組合せに関するいかなる従属性または要件を有するとも解釈されてはならない。

【0038】

図2Bを参照すると、本発明を実施するための例示的システムは、コンピュータ110の形態の汎用コンピューティングデバイスを含む。コンピュータ110のコンポーネントとしては、処理ユニット120、システムメモリ130、およびシステムメモリを含む種々のシステムコンポーネントを処理ユニット120に接続するシステムバス121が挙げられるが、これらには限定されない。システムバス121は、さまざまなバスアーキテクチャのいずれかを使用するメモリバスまたはメモリコントローラ、周辺バス、およびローカルバスを含む、いくつかのタイプのバス構造のいずれでもよい。例として、限定ではないが、このようなアーキテクチャとしては、Industry Standard Architecture（ISA）

10

20

30

40

50

バス、Micro Channel Architecture(M C A)バス、Enhanced ISA(E I S A)バス、Video Electronics Standards Association(V E S A)ローカルバス、および(Mezzanine バスとも呼ばれる) Peripheral Component Interconnect(P C I)バスがある。

【 0 0 3 9 】

コンピュータ 1 1 0 は通常、さまざま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 1 1 0 がアクセスすることができるいかなる利用可能な媒体であってもよく、揮発性および不揮発性媒体、リムーバブルおよび非リムーバブル媒体の両方がある。例として、限定ではないが、コンピュータ可読媒体としては、コンピュータ記憶媒体および通信媒体が挙げられる。コンピュータ記憶媒体としては、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のような情報の記憶のための任意の方法または技術で実現された揮発性および不揮発性、リムーバブルおよび非リムーバブル媒体の両方がある。コンピュータ記憶媒体としては、以下のものに限定されないが、RAM、ROM、EEPROM、フラッシュメモリ等のメモリ技術、CD-ROM、デジタル多用途ディスク(DVD)等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等の磁気記憶デバイス、または所望の情報を記憶するために使用可能でありコンピュータ 1 1 0 によりアクセス可能な任意の他の媒体がある。通信媒体は通常、キャリア波等の変調データ信号または他のトランスポートメカニズムで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具現化し、いかなる情報配信媒体も含む。「変調データ信号」という用語は、信号中に情報を符号化するよう1つまたは複数の信号の特性が設定または変更された信号を意味する。例として、限定ではないが、通信媒体としては、有線ネットワークまたは直接有線コネクションのような有線媒体、および音響、RF(無線周波数)、赤外線等の無線媒体のような無線媒体がある。上記のいずれかの組合せもまた、コンピュータ可読媒体の範囲内に含まれるべきである。

【 0 0 4 0 】

システムメモリ 1 3 0 は、読み出し専用メモリ(ROM) 1 3 1 およびランダムアクセスメモリ(RAM) 1 3 2 のような揮発性および/または不揮発性メモリの形態のコンピュータ記憶媒体を含む。起動中等にコンピュータ 1 1 0 内の要素間で情報を転送するのに役立つ基本ルーチンを含む基本入出力システム 1 3 3 (BIOS)が通常 ROM 1 3 1 に記憶されている。RAM 1 3 2 は通常、処理ユニット 1 2 0 から直ちにアクセス可能な、および/または処理ユニット 1 2 0 が現在作用し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。例として、限定ではないが、図 2 B は、オペレーティングシステム 1 3 4 、アプリケーションプログラム 1 3 5 、他のプログラムモジュール 1 3 6 、およびプログラムデータ 1 3 7 を示している。

【 0 0 4 1 】

また、コンピュータ 1 1 0 は、他のリムーバブル(取り外し可能)/非リムーバブル(取り外し不可能)、揮発性/不揮発性のコンピュータ記憶媒体を含んでもよい。単なる例として、図 2 B は、非リムーバブル不揮発性磁気媒体の読み書きを行うハードディスクドライブ 1 4 1 、リムーバブル不揮発性磁気ディスク 1 5 2 の読み書きを行う磁気ディスクドライブ 1 5 1 、および CD - ROM 等の光媒体のようなリムーバブル不揮発性光ディスク 1 5 6 の読み書きを行う光ディスクドライブ 1 5 5 を示している。例示的オペレーティング環境で使用可能な他のリムーバブル/非リムーバブル、揮発性/不揮発性のコンピュータ記憶媒体としては、以下のものに限定されないが、磁気テープカセット、フラッシュメモリカード、デジタル多用途ディスク、デジタルルビデオテープ、固体 RAM 、固体 ROM 等がある。ハードディスクドライブ 1 4 1 は通常、インターフェース 1 4 0 のような非リムーバブルメモリインターフェースを通じてシステムバス 1 2 1 に接続され、磁気ディスクドライブ 1 5 1 および光ディスクドライブ 1 5 5 は通常、インターフェース 1 5 0 のようなリムーバブルメモリインターフェースによりシステムバス 1 2 1 に接続される。

【 0 0 4 2 】

前述し図 2 B に示したドライブおよびそれらに関連するコンピュータ記憶媒体は、コンピュータ 1 1 0 のための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

10

20

30

40

50

び他のデータの記憶を行う。例えば図2Bにおいて、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を記憶するように示されている。なお、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と同じでも異なってもよい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別のコピーであることを示すためにここでは異なる番号が与えられている。ユーザは、キーボード162、およびマウス、トラックボールまたはタッチパッドと一般的に呼ばれるポインティングデバイス161のような入力デバイスを通じてコンピュータ110にコマンドおよび情報を入力することができる。他の入力デバイス(図示せず)としては、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、スキヤナ等が挙げられる。これらおよび他の入力デバイスは、システムバス121に結合したユーザ入力インターフェース160を通じて処理ユニット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ユニバーサルシリアルバス(USB)等の他のインターフェースおよびバス構造により接続されてもよい。ノースブリッジ(North bridge)等のグラフィックスインターフェース182もまたシステムバス121に接続されてもよい。ノースブリッジは、CPU、すなわちホスト処理ユニット120と通信するチップセットであり、アクセラレーテッドグラフィックスポート(AGP)通信を担当する。1つまたは複数のグラフィックス処理ユニット(GPU)184がグラフィックスインターフェース182と通信してもよい。この点で、GPU184は、一般的に、レジスタストレージ等のオンチップメモリストレージを含み、GPU184は、本発明のアプリケーション変数が影響を及ぼし得るビデオメモリ186と通信する。しかし、GPU184は、コプロセッサ(coprocessor)の単なる一例であり、さまざまなコプロセッシングデバイスがコンピュータ110に含まれてもよく、また、ピクセルシェーダ(shader:陰影処理)や頂点シェーダ等のさまざまな手続き型シェーダを含んでもよい。モニタ191または他のタイプのディスプレイデバイスもまた、ビデオインターフェース190のようなインターフェースを通じてシステムバス121に接続され、このインターフェースがさらにビデオメモリ186と通信してもよい。モニタ191に加えて、コンピュータはスピーカ197やプリンタ196のような他の周辺出力デバイスを含んでもよく、これらは出力周辺インターフェース195を通じて接続され得る。

【0043】

コンピュータ110は、リモートコンピュータ180のような1つまたは複数のリモートコンピュータへの論理結合を用いたネットワーク環境または分散環境で動作してもよい。リモートコンピュータ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、ピアデバイス(peer device)または他の普通のネットワークノードであってよく、通常、コンピュータ110に関して前述した要素の多くまたはすべてを含む。ただし、図2Bにはメモリ記憶デバイス181のみが示されている。図2Bに示す論理結合は、ローカルエリアネットワーク(LAN)171および広域ネットワーク(WAN)173を含むが、他のネットワーク/バスを含んでもよい。このようなネットワーキング環境は、家庭、オフィス、企業規模のコンピュータネットワーク、インターネットおよびインターネットで一般的である。

【0044】

LANネットワーキング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すなわちアダプタ170を通じてLAN171に接続される。WANネットワーキング環境で使用される場合、コンピュータ110は通常、インターネットのようなWAN173を通じて通信を確立するためのモデム172等の手段を含む。モデム172は、内蔵でも外付けでもよいが、ユーザ入力インターフェース160等の適当なメカニズムを通じてシステムバス121に接続され得る。ネットワーク環境では、コンピュータ110に関して図示したプログラムモジュールまたはその部分は、リモートメモリ記憶デバイ

10

20

30

40

50

スに記憶されてもよい。例として、限定ではないが、図2Bは、リモートアプリケーションプログラム185がメモリデバイス181上に存在するように示している。理解されるように、図示したネットワークコネクションは例示であり、コンピュータ間に通信リンクを確立する他の手段を使用してもよい。

【0045】

[例示的分散コンピューティングフレームワークまたはアーキテクチャ]

パーソナルコンピューティングとインターネットの集中を考慮に入れて、種々の分散コンピューティングフレームワークがすでに開発され、現在も開発され続けている。個人ユーザにも企業ユーザにも区別なく、アプリケーションおよびコンピューティングデバイスのため、シームレスに相互運用可能な、ウェブ対応のインターフェースが提供されている。これにより、コンピューティング活動はますますウェブブラウザ指向あるいはネットワーク指向になっている。

10

【0046】

例えば、MICROSOFT（登録商標）のマネージドコードプラットフォーム、すなわち、.NETは、サーバと、ウェブベースのデータストレージやダウンロード可能なデバイスソフトウェアのようなビルディングブロックサービスを含む。一般的に言えば、.NETプラットフォームにより以下のことが実現される。（1）すべてのコンピューティングデバイスを協調動作させ、それらすべての上でユーザ情報を自動的に更新し同期させることができる。（2）HTMLではなくXMLのほうをより多く活用することによって、ウェブページのインタラクティブ（対話式）の能力を向上させる。（3）例えば電子メールのような種々のアプリケーションやOffice-.NETのようなソフトウェアの管理のために、スターティングポイント（中央基点）からユーザへの製品およびサービスのカスタマイズされたアクセスおよび配布を特徴とするオンラインサービスを提供する。（4）集中化されたデータストレージにより、情報へのアクセスと、ユーザおよびデバイスの間の情報の同期を効率化し容易化する。（5）電子メール、ファックス、および電話のような種々の通信媒体を統合することができる。（6）開発者にとっては、再利用可能なモジュールを作成することができるため、生産性が向上しプログラミングエラーの数が減る。そして、（7）多くの他のクロスプラットフォーム機能および言語統合機能もある。

20

【0047】

本明細書では、いくつかの例示的実施形態が、コンピューティングデバイス上に存在するソフトウェアとの関連で説明されるが、本発明の1つまたは複数の部分は、オペレーティングシステム、アプリケーションプログラミングインターフェース（API）または「仲介役（middle man）」オブジェクト、コントロールオブジェクト、ハードウェア、ファームウェア、中間言語命令またはオブジェクト等によって実施されてもよい。それによって、本方法は、.NETコードのようなマネージドコードにより可能となるすべての言語およびサービス、ならびに他の分散コンピューティングフレームワークに、含まれ、サポートされ、またはそれらによってアクセスされることが可能となる。

30

【0048】

[ウェブページをランク付けするシステムおよび方法]

背景技術のところで既に述べたように、本発明は、ターゲットウェブページのスコアをつり上げるだけの目的で人為的に生成されるウェブページに対する、PageRank等の従来のランク付けアルゴリズムの脆弱性を解決する。本発明では、ウェブサーファが特定のサーバへランダムにジャンプしようとしていると仮定すると、少数のページを有するウェブサーバ上でサーファが特定のページに「着地」するよりも、多数のページを有するウェブサーバ上でサーファが特定のページに「着地」するほうが、可能性が低いであろう、と認識している。この原理を適用してランク付けアルゴリズムを適用すると、本発明は、複数の推奨が同一サーバから発している場合、それらの推奨の影響を低減する。この影響の低減を達成するために、種々の非限定的実施形態では、本発明は、各ウェブページではなく、各ウェブサーバに最低保証スコアを割り当てる。そして、そのウェブサーバに割り当てるその最低スコアを、そのウェブサーバ上のすべてのページ間で分割すること

40

50

ができる。従って、あるウェブサーバ上の複数のページにより任意多数個のネポティスティックリンクを作成しても、それにより達成される「推奨値」は、ウェブサーバ上の少数のページ上の比較的少数のネポティスティックリンクと同程度である。一般的に、任意多数個のネポティスティックリンクは、(A)同じシンボリックホスト名を有するか、(B)同じドメインに関連付けられるか、または(C)同じIPアドレスに関連付けられるかのいずれか1つまたは複数である。一実施形態では、本発明は、良質の問合せ結果をユーザに返すためにウェブ検索サービスにおいて実施される。

【0049】

なお、何がウェブサーバを構成するかについて、いくつかの可能な定義が存在する。ウェブサーバは、シンボリックホスト名(例えば、www.google.com)によって定義されてもよく、共通ドメインによって定義されてもよく、あるいは、1つ(または複数)のIPアドレス(例えば、207.46.134.222)によって定義されてもよい。ウェブサーバのいずれの定義を利用する実施態様もリンクスパム(link spammers)問題を解決するが、本発明の目的上、後の2つの定義のほうが好適であり、リンクスパムに対して、より良い抑止力として働く。というのは、ほとんど無限個のホスト名を解決するようにDNS(ドメイン名システム)サーバを設定することが可能であって(RFC1035は、ホスト名の長さを高々255キャラクタに限定しており、各キャラクタは文字、数字、またはハイフンである。したがって、 37^{255} 個の可能なホスト名が存在し、すべての実用上の目的にとっては「ほとんど無限」である)、無限個のウェブページをサービスするようにウェブサーバを設定することが可能であることとちょうど同じだからである。以下の説明では、2つの代替実施形態について順に各定義を使用する。

【0050】

本発明によるランク付けメトリック(metric)の第1実施形態では、ウェブサーバはそのシンボリックホスト名によって定義される。h(u)でURL uのホスト名コンポーネントを表し、Hを全ホストの集合、すなわち

$$H = \{ h(v) : v \in V \}$$

とする。V^H(h)を、ホストhによってサービスされるURLの集合、すなわち

$$V^H(h) = \{ v : v \in V \text{ and } h(v) = h \}$$

とする。そして、これらの2つの定義を、本発明の第1実施形態による第1の新しいランク付けメトリックR^Hの定義において用いる。これは次のように表される。

【0051】

【数2】

$$R^H(v) = \frac{p}{|H|} \frac{1}{|V^H(h(v))|} + (1-p) \sum_{(u,v) \in E} \frac{R^H(u)}{O(u)}$$

【0052】

本発明によるランク付けメトリックの第2実施形態では、ウェブサーバはそのドメイン名によって定義される。d(u)でURL uのドメイン名コンポーネントを表し、Dを全ドメインの集合、すなわち

$$D = \{ d(v) : v \in V \}$$

とする。V^D(h)を、ドメインd内のウェブサーバによってサービスされるURLの集合、すなわち

$$V^D(d) = \{ v : v \in V \text{ and } d(v) = d \}$$

とする。そして、これらの2つの定義を、本発明の第2実施形態による第2の新しいランク付けメトリックR^Dの定義において用いる。これは次のように表される。

【0053】

【数3】

$$R^D(v) = \frac{p}{|D|} \frac{1}{|V^D(d(v))|} + (1-p) \sum_{(u,v) \in E} \frac{R^D(u)}{O(u)}$$

10

20

30

40

50

【0054】

本発明の第3実施形態では、ウェブサーバはそのIPアドレスの集合によって定義される。A(u)で、そこからURL uがサービスされ得るIPアドレスの集合(すなわち、h(u)が解決される先のIPアドレスの集合)を表す。Aを全IPアドレスの集合、すなわち、

【0055】

【数4】

$$A = \bigcup_{v \in V} A(v)$$

【0056】

とする。V^A(a)をIPアドレスaによってサービスされるURLの集合、すなわち
 $V^A(a) = \{v : v \in V \text{ 且し } a \in A(v)\}$

とする。そして、これらの2つの定義を、本発明の第3実施形態による第3の新しいランク付けメトリックR^Aの定義において用いる。これは次のように表される。

【0057】

【数5】

$$R^A(v) = \frac{p}{|A|} \sum_{a \in A(v)} \frac{1}{|V^A(a)|} + (1-p) \sum_{(u,v) \in E} \frac{R^A(u)}{O(u)}$$

【0058】

次に、本発明について、図3A～図3Gを参照して説明する。図3Aは、第1ノードN1(例えば、文書またはウェブページ)を示しており、この第1ノードN1は第2ノードN2へのリンクを有する。PageRankの直観的概念は、N2のほうが、全く入リンク(incoming link)のないノードよりも優れているページであるということである。というのは、N1は、N2にリンクすることによって、N2を推薦している、すなわちN2の存在を「認証(validate)」しているからである。PageRankの直観的概念は、図3Bを図3Aと比較することによって、さらに説明することができる。図3AではN1はただ1つのノードN2にリンクしているのに対して、図3BではノードN1は9個のノードN2～N10にリンクしているので、PageRankはN1の推薦の価値を低下させる。というのは、N1は、図3AのノードN1に比べて、他のウェブページを無差別に推薦しているように見えるからである。図3AのノードN1は他のノードをより選択的に推薦しているので、図3Aでは、対応するN2には、(すべての他のファクタは同一として)N2へのN1のリンクの結果として、より高いスコアが割り当てられる。PageRankの直観的概念は、図3Cを図3Aと対比することによって、さらに説明することができる。この対比では、図3Cの推薦側ノードN1に割り当てられるスコアは、図3Aの推薦側ノードN1に割り当てられるスコアの10倍である。したがって、推薦側ノードの品質(スコア)は図3Cのほうが高いため、図3AのノードN2に割り当てられるスコアに比べて、図3CのノードN2に割り当てられるスコアのほうが高い。これらの直観的概念を組み合わせることにより、PageRankアルゴリズムが得られる。

【0059】

しかし、背景技術のところで注意したように、PageRankは図3Dに例示する状況に対して脆弱である。というのは、あらゆるノードは、それが他の何個のノードを推薦しているか、あるいは自分のPageRankスコアがどれほど小さいかにかかわらず、最低限のスコアを受けるからである。この最低スコアは小さいが、それが多数になると、小さいスコアが増倍されて相当の大きさになり得る。したがって、ノードN2のオーナは、多数の推薦側ページを手動または自動で作成し、例えば図3Dに示すような何らかの方法または構成で各推薦側ページがそれぞれノードN2および互いを推薦することによって、ノードN2のスコアを人為的に上げることができる。一般的に言えば、費用対効果の高い方法でこれを行うために、リンクスパムは通常、リンクスパムノードLS1～LS9等のそれぞれを同じホスト(例えば、www.foo.com)上に置く。というのは、同じホス

10

20

30

40

50

ト上に、www.foo.com/page1、www.foo.com/page2、www.foo.com/page3等の追加のウェブページを生成するのには、ほとんどまたは全くコストがかからないからである。したがって、利口なリンクスパマは、この技法によって、ウェブ上で、人気がないウェブページの露出を増大させることができる。ウェブサーバには無限個のページを動的に作成してサービスするものがあるので、ウェブページの数は無限である。明らかに、もう1つのウェブページを作成するための増分コストは低い。

【0060】

これに応じて、本発明は、このようなリンクスパム技法の使用を不利にする。リンクスパムノード LS 1 ~ LS 9 のそれぞれがすべて同じホスト www.foo.com 上にある場合、LS 1 ~ LS 9 のそれぞれに最低スコアを割り当てるのではなく、一実施形態では、本発明は、各ホストに最低スコアを割り当て、そのホスト上の各ノード間でその最低スコアを分配する。したがって、ホスト www.foo.com 上に 1000 個のリンクスパムノード LS 1 ~ LS 1000 が存在するか、それとも 1 個のリンクスパムノード LS 1 が存在するかにかかわらず、ノード N 2 に対する推奨への寄与は（他のすべてのファクタを同一と仮定すれば）同一である。こうして、図 3 E に例示するように、本発明では、ホスト www.foo.com に最低スコアを割り当て、ページ LS 1 ~ LS 9 の間でその最低スコアを分割する。

【0061】

しかし、確信的なリンクスパマは、本発明のこの改良型（variant）の裏をかくことができる。リンクスパマは、類似の品質スコアを有する他のウェブページよりも高いスコアを得るために、自分のウェブページに関連する品質スコアを比較的少量だけ増大させることに決めるかもしれない。例えば、ウェブページのランクが典型的な検索エンジン結果において上昇し、ランキングにおいて 5 番目にリストされていたのが最初にリストされるようになれば、比較的小さい品質スコアの変化でも、リンクスパマにとっては大きな利益となるであろう。現在の形態のインターネットドメインネーミングシステムでは、ユーザは各ドメイン名について年間利用料（約 25 ドル）を支払い、そのドメイン内に任意個数のシンボリックホスト名を作成することができる。したがって、リンクスパマは、少数のドメインを取得し、それらのドメイン内の任意の可能なホスト名を決定する（resolve）ように DNS サーバを設定することができる。そして、スパマは、これらのドメイン内の多くの異なるホストから来ているように見えるリンクスパムページを設けることにより、非常に多数のページの最低スコアを累積することによって、ページを推奨する能力を取り戻すことができる。この状況を図 3 F に示す。図 3 F において、複数のウェブページが確信的なリンクスパマによって生成されている。各ウェブページは、固有のシンボリックホスト名 www1.foo.com、www2.foo.com、www3.foo.com 等からもたらされる。しかし、各ウェブページは、共通ドメイン foo.com からもたらされる。したがって、本発明の第 2 実施形態では、最低スコアを、各ウェブページやそれぞれの異なるホスト名ではなく、それぞれの異なるドメイン名に割り当てる。

【0062】

結局のところ、ドメイン名は金がかかるが、極端に高くつくわけではない。無限個のウェブページおよび潜在的に多数（37²⁵⁵）のドメイン名があるが、現実に存在するのは 1500 万個程度のドメイン名である。新しいホスト名を作成するには、（ドメイン内の任意の可能なホスト名を決定する（resolve）ように DNS サーバを設定することによって）、無料であるが、新しいドメイン名を作成するには 25 ドル程度かかる。したがって、約 25 ドルでドメイン名を取得することができるので、確信的リンクスパマにとっては、図 3 G に例示するシステムを実装することは依然として実現可能である。図 3 G において、確信的リンクスパマはリンクスパムノード LS 1 ~ LS 9 等を作成しているが、この場合、リンクスパマは、各ノードを、固有のシンボリックホスト名 www.foo.com、www.goo.com、www.hoo.com 等を有する固有のドメイン上に置いている。しかし、リンクスパマは、www.foo.com、www.goo.com、www.hoo.com 等のそれぞれを同じ IP アドレスに置いているであろう。IP アドレスには、約 40 億個の異なる IP アドレスがある。IP アドレスは比較的安価に取得できるが、リンクスパマはリンクスパムウェブページごとに異なる

10

20

30

40

50

IPアドレスを割り当てるシステムを実装しないであろう。というのは、異なるIPアドレスを無限に供給することはできないからである。したがって、本発明の第3実施形態では、最低スコアを、各ウェブページ、各ホスト名、または各ドメインではなく、それぞれの異なるIPアドレスに割り当て、確信的リンクスパマから防護する。この理由は、リンクスパマがサーバ上にリンクスパムノードを有する100個の異なるドメインを置くか、リンクスパマがサーバ上に1個のリンクスパムノードを有する1個のドメインを置くかにかかわらず、ノードN2に対する推奨に関する効果は同一だからである（今度も、他のすべてのファクタは一定のままであるとする）。

【0063】

本発明は、任意の構造的に相互関係のある文書をランク付けするために適用することができるが、図4A（システム図）、図4Bおよび図4C（流れ図）は、検索エンジンアプリケーションのためのウェブページのランク付けに対する本発明の例示的適用形態を示している。本発明のアルゴリズムは、問合せ非依存であるので、本発明による文書をランク付けするプロセスは、文書に割り当たられるスコアに基づいて、文書を要求する検索エンジン等のアプリケーションとは独立に行うことができる。例えば、図4Aに示すように、1つまたは複数のクローラ（crawler）410が、インターネット上のウェブページ等の構造的に相互関係のある文書のソース400をクロール（crawl：這い回って巡回）して、リポジトリ（repository：貯蔵所）420に格納するために文書または文書に関する関連情報を取り込むことができる。文書に関する関連情報は、他のソース415からもたらされることも可能である。

10

20

20

【0064】

クローラとは、検索エンジンインデックスのエントリを作成するために、ウェブサイトを訪問してそれらのページその他の情報を読み出すプログラムである。クローラは、サーバからサーバへとハイパーテキストリンクをたどり、検索基準に基づいて情報をインデックス付けすることによって、新しい文書および新しいサイトを探索するために使用される。

【0065】

ウェブ上の主要な検索エンジンはすべてこのようなプログラムを有し、それは「スパイダ（spider）」、「アント（ant）」、「ロボット（robot）」（「ボット（bot：検索ロボット）」）または「インテリジェントエージェント（intelligent agent）」とも呼ばれる。クローラは通常、サイトのオーナによって新規なサイトまたは更新されたサイトとして登録されたサイトを訪問するようにプログラムされる。サイト全体または特定のページを選択的に訪問しインデックス付けすることができる。クローラは、サイト内を一度に1ページずつ巡回（crawl）し、全ページを読み終えるまで、そのサイト上の他のページへのリンクをたどることから、その名前が付いたようである。通常、クローラは、同時に多くのウェブサイトをクロールする。一般的に、クローラは、Standard for Robot Exclusion（SRE）に規定されるウェブクローラの行儀作法（politeness）の規則に従う。

30

【0066】

図4Bは、文書を収集するプロセスを流れ図により示している。450で、ウェブページを見つけ、ウェブページおよび／またはそのウェブページに関する情報をクローラにより収集する。これは、矢印で示されているように、反復的な、または進行中のプロセスとすることができます。460で、文書の集まりの構造が分かった任意の与えられた時点において、本発明を適用することにより、収集された各文書にスコアを割り当て、それにより、ウェブページ（またはそれへのリンク）のリポジトリがそれぞれ存在し、そのそれぞれがR^Aメトリック、R^Dメトリック、またはR^Hメトリックを用いて関連付けられた品質スコアを有するようにしている。オブジェクト430が、API432経由で、リポジトリ420内の各文書にスコアを割り当てるための計算を実行することができる。また、以下でさらに詳細に説明するが、理解されるように、R^A、R^D、および／またはR^Hメトリックは、特定のアプリケーションまたはユーザに関して文書に割り当たられるスコアの品質を上げるために、他のメトリックと組み合わせることもできる。

40

50

【0067】

文書およびスコアの初期リポジトリ 420 (これは、反復的に、連続的に、または定期的に更新することができる)が生成された後、検索エンジンの例示的アプリケーションを適用することができる。例えば、470で、検索エンジン (または他のアプリケーション) オブジェクト 440 は、ユーザから入力問合せを受け取ることができる。480で、その問合せに基づいて、問合せ基準に従って問合せ用語を含むウェブページを検索し、それらに関連するスコアに従ってランク付けすることができ、それにより、最も品質の高い文書を最初に、またはより目立つようにユーザに対して表示する。一例示的実施態様では、図 4A に示すように、問合せは検索エンジンサーバオブジェクト 440 から受け取られる。検索エンジンサーバオブジェクト 440 は、API 434 経由でオブジェクト 430 と 10 インタフェースをとる。オブジェクト 430 は、問合せに基づいて、API 432 経由でリポジトリ 420 から関連するウェブページ (またはウェブページへのリンク) を検索して順序付ける。そして、オブジェクト 430 は、ユーザに対して表示するために、順序付けられた結果のリストをアプリケーションオブジェクト 440 に返す。別法として、オブジェクト 440 が、関連するスコアに基づいて結果の順序付けを実行することができる。

【0068】

明らかなように、本発明のメトリックは、互いに組み合わせてもよく、全体的メトリックに対する他の改善と組み合わせてもよい。例えば、「信頼できる」文書の基本セットを用いて、任意のメトリックにより達成されるスコアの全体的品質を改善することができる。このような改善としては、ニールセン視聴率 (Nielsen rating) を考慮に入れることができる。例えば、上位のニールセンウェブページは、他のスコアを評価するためのウェブページの信頼できる基礎として用いることができる。さらに、ニールセン視聴率は、重み付けされたスコアを割り当てるために、本発明のメトリックと組み合わせて用いることができる。あるいは、異常な結果を発見するために、本発明のメトリックの結果をニールセン視聴率と比較することができる。

【0069】

基本的に、ウェブページ等の文書の信頼性に関する外部情報源は、本発明のメトリックの結果と組み合わせて、またはその結果を確認するために、用いることができる。他の例としては、ISP からのユーザ情報の集まりに基づく情報を用いるものがある。例えば、ISP は、ユーザがある特定のウェブページをどのくらいの頻度で訪問しているかに関して、ニールセン情報に類似の直接的情報を収集することができ、したがって、利用パターンに基づいて品質、人気、または信頼性のスコアを割り当てることができる。ISP プロキシログ (proxy logs) を、この目的のために調べることができる。もう 1 つの例は、検索エンジン上で人々を観察することを含む。ユーザがウェブページを選択する可能性が最も高い理由は、単にそのウェブページが最高スコアを取っていることだけではない。したがって、検索エンジン上でのユーザの挙動は、ウェブページの品質を検証するのに役立つ。また、例えば 10,000 件の優良なウェブページを認定する仕事を何人かの編集者に割り当てるともできる。職務の進行中に、それらのウェブページが信頼できるウェブページであるということが知られるため、これらの 10,000 件のスコアがつり上げられることにより、本発明のメトリックによるそれらの推薦能力がつり上げられる可能性がある。良質のウェブページに関するもう 1 つの情報源として、ユーザのブックマークのようなユーザの選好からもたらされるものがあり得る。要約すれば、本発明のメトリックは、最良のユーザ見聞が提供されることを保証するために、任意の他の既知の品質メトリックと組み合わせることができる。有利な点として、本発明のメトリックを含むいかなる組合せも、リンクスパムの努力を挫折させるであろう。

【0070】

本発明を実施する方法としては、アプリケーションおよびサービスが本発明のランク付けのシステムおよび方法を使用することを可能にする適当な API、ツールキット、ドライバコード、オペレーティングシステム、コントロール、独立型の、あるいはダウンロード可能なソフトウェアオブジェクト等の多くの方法がある。本発明は、API (または他

のソフトウェアオブジェクト)の観点からとともに、本発明によるランク付け技法の適用のためにウェブページまたはウェブページに関連する構造的情報を受け取るソフトウェアまたはハードウェアオブジェクトの観点からの本発明の使用を考慮している。したがって、本明細書に記載されている本発明の種々の実施態様は、完全にハードウェアによる態様、部分的にハードウェアで部分的にソフトウェアによる態様、およびソフトウェアによる態様を有し得る。

【0071】

上記のように、本発明の例示的実施形態は種々のコンピューティングデバイスおよびネットワークアーキテクチャとの関連で説明されているが、基礎となる概念は、構造的に相互関係のある文書をランク付けすることが望ましいとなるコンピューティングデバイスまたはシステムにも適用することができる。例えば、本発明のアルゴリズムおよびハードウェア実施態様は、コンピューティングデバイスのオペレーティングシステムに適用され、デバイス上の別個のオブジェクトとして、別のオブジェクトの一部として、再利用可能コントロールとして、サーバからダウンロード可能なオブジェクトとして、デバイスまたはオブジェクトとネットワークとの間の「仲介役」として、分散オブジェクトとして、ハードウェアとして、メモリ内に、あるいは以上のいずれかの組合せとして提供されてもよい。例示的なプログラミング言語、名前および例は、本明細書では、種々の選択肢を代表するものとして選択されているが、これらの言語、名前および例は限定的であることを意図していない。当業者には理解されるように、本発明の種々の実施形態によって達成されるのと同一、類似または等価な機能を達成するオブジェクトコードおよび命名法を提供する数多くの方法がある。

【0072】

上記のように、本明細書に記載されている種々の技法は、ハードウェアもしくはソフトウェア、または適宜その両方の組合せとの関連で実施することができる。したがって、本発明の方法および装置、またはそのある特定の態様または部分は、フロッピー(登録商標)ディスクケット、C D - R O M、ハードドライブ等のマシン可読記憶媒体のような有形媒体に具現化されたプログラムコード(すなわち命令)の形態をとってもよい。ここで、プログラムコードがコンピュータ等のマシンにロードされて実行される場合、そのマシンが本発明を実施する装置となる。プログラム可能なコンピュータ上でプログラムコードを実行する場合、コンピューティングデバイスは一般的に、プロセッサ、そのプロセッサが読むことができる記憶媒体(揮発性および不揮発性メモリおよび/または記憶要素を含む)、少なくとも1つの入力デバイス、および少なくとも1つの出力デバイスを含む。例えばデータ処理A P I、再利用可能コントロール等の使用を通じて本発明のランク付け技法を実施または利用することができる1つまたは複数のプログラムは、好ましくは、コンピュータシステムと通信する高レベルの手続き型またはオブジェクト指向プログラミング言語として実装される。しかし、プログラムは、必要であれば、アセンブリ言語またはマシン語で実装することができる。いずれの場合でも、言語は、コンパイルされた言語または解釈された言語であってもよく、ハードウェア実施態様と組み合わされてもよい。

【0073】

本発明の方法および装置は、電気配線またはケーブル、光ファイバ、または任意の他の形態の伝送等の何らかの伝送媒体を通じて伝送されるプログラムコードの形態で具現化される通信により実施されてもよい。ここで、プログラムコードがE P R O M、ゲートアレイ、プログラマブルロジックデバイス(PL D)、クライアントコンピュータ等のマシンに受信され、ロードされて実行される場合、そのマシンが本発明を実施する装置となる。汎用プロセッサ上で実施される場合、プログラムコードがプロセッサと組み合わされて、本発明の機能を呼び出すように動作する固有の装置を提供する。また、本発明との関連で使用される任意の記憶技法は必然的にハードウェアとソフトウェアの組合せとなるであろう。

【0074】

本発明は、種々の図の好ましい実施形態との関連で説明されているが、理解されるよう

10

20

30

40

50

に、本発明から逸脱することなく、本発明と同じ機能を実行するために、他の類似の実施形態を用いてもよく、記載されている実施形態に変更または追加をしてよい。例えば、本発明の例示的ネットワーク環境は、ピアツーピアネットワーク環境等のネットワーク環境との関連で説明されているが、当業者には認識されるように、本発明はこれに限定されない。本明細書に記載の方法は、有線か無線かを問わず、ゲーム用コンソール、ハンドヘルドコンピュータ、ポータブルコンピュータ等のいかなるコンピューティングデバイスまたは環境に適用することもできるし、また、通信ネットワーク経由で接続され、ネットワーク越しにやりとりをする任意個数のこのようなコンピューティングデバイスに適用することもできる。さらに、強調しておかなければならぬが、特に無線ネットワークデバイスの数が増大し続けるとともに、ハンドヘルドデバイス用オペレーティングシステム等のアプリケーション固有のオペレーティングシステムを含むさまざまなコンピュータプラットフォームが考えられる。

【0075】

例示的実施形態は、ウェブ検索サービスの関連で本発明を利用することに関するものであるが、本発明は、ウェブ検索サービスの場合に限定されず、何らかの形で互いを参照する文書またはコンテンツの任意の集合にある種の品質メトリックを提供するために実施することができる。例えば、ユーザのコンピュータに（または複数のコンピューティングデバイスにわたって）、何らかの形で（人、場所、時、イベント、アーティスト、アルバム、タイトル、俳優等によって）構造的に相互関係のある画像、映画、歌等の集合が記憶されている場合、本発明のスコアリングを、コンテンツの構造的相互関係に基づいて画像に適用することができる。例えば、第1のコンテンツと第2のコンテンツの間の類似性等の関係を、第1のコンテンツから第2のコンテンツへのリンクとみなしてもよい。また、A P I 4 3 2 は上記ではスコアリングプロセスとの関連で説明されているが、その目的のために別個のA P I を実装してもよいことが注意される。すなわち、問合せに基づいて検索および順序付けを行うためのA P I 4 3 2 の使用は、必ずしもスコアリングプロセスに関連するものである必要はない。さらに、本明細書で用いられる場合の「～に比例する（して）」という用語は、一方のエンティティが増大すると、すべての既知の数学的関係に従って他方が増大するような2つのエンティティ間の任意の数学的関係を指す。そのような数学的関係としては、以下のものに限定されないが、幾何、線形、指数、対数等の関係がある。同じことは「反比例する」または「～に反比例する（して）」という用語、すなわち、一方のエンティティが増大すると他方が減少する場合にも当てはまる。また、サーバという用語はさまざまな文脈でさまざまなものを意味し得るが、本明細書で種々に用いられる場合の「ウェブサーバ」という用語は、（A）同じシンボリックホスト名を有する複数のウェブページ、（B）同じドメインに関連する複数のウェブページ、および（C）同じIPアドレスに関連する複数のウェブページ、のいずれか1つまたは複数を含むサーバコンピュータおよび/またはサーバオブジェクトを少なくとも指すことを意図している。さらに、本発明は、複数の処理チップまたはデバイスにおいて、またはそれら全体にわたって実施することができ、記憶（格納）も同様に複数のデバイスにわたって実行することができる。したがって、本発明は、いかなる単一の実施形態にも限定されてはならず、特許請求の範囲に基づく広さおよび範囲において解釈されなければならない。

【図面の簡単な説明】

【0076】

【図1A】静的リンクベースアルゴリズムによりウェブページをランク付けする従来技術の技法を示す概念図である。

【図1B】PageRankアルゴリズムによりウェブページをランク付けする従来技術の技法を示す概念図である。

【図2A】本発明を実装し得るさまざまなコンピューティングデバイスを有する例示的ネットワーク環境を表すブロック図である。

【図2B】本発明を実装し得る例示的な非限定的コンピューティングデバイスを表すブロック図である。

10

20

30

40

50

【図3A】本発明によるリンクスパムの防止の背後にある直観的概念を示す模式図である。

【図3B】本発明によるリンクスパムの防止の背後にある直観的概念を示す模式図である。

【図3C】本発明によるリンクスパムの防止の背後にある直観的概念を示す模式図である。

【図3D】本発明によるリンクスパムの防止の背後にある直観的概念を示す模式図である。

【図3E】本発明によるリンクスパムの防止の背後にある直観的概念を示す模式図である。

10

【図3F】本発明によるリンクスパムの防止の背後にある直観的概念を示す模式図である。

【図3G】本発明によるリンクスパムの防止の背後にある直観的概念を示す模式図である。

【図4A】検索エンジンアプリケーションにおける本発明の種々の実施形態の例示的実施態様を示すシステム図である。

【図4B】検索エンジンアプリケーションにおける本発明の種々の実施形態の例示的実施態様を示す流れ図である。

【図4C】検索エンジンアプリケーションにおける本発明の種々の実施形態の例示的実施態様を示す流れ図である。

20

【符号の説明】

【0077】

10a, 10b コンピューティングオブジェクト(サーバ)

14 通信ネットワーク

20 データベース

100 コンピューティングシステム環境

110 コンピュータ

110a, 110b, 110c コンピューティングオブジェクト(クライアント)

110d, 110e コンピューティングオブジェクト(クライアント)

120 処理ユニット

30

121 システムバス

130 システムメモリ

131 読み出し専用メモリ(ROM)

132 ランダムアクセスメモリ(RAM)

133 基本入出力システム(BIOS)

134 オペレーティングシステム

135 アプリケーションプログラム

136 他のプログラムモジュール

137 プログラムデータ

140 非リムーバブルメモリインタフェース

40

141 ハードディスクドライブ

144 オペレーティングシステム

145 アプリケーションプログラム

146 他のプログラムモジュール

147 プログラムデータ

150 リムーバブルメモリインタフェース

151 磁気ディスクドライブ

152 磁気ディスク

155 光ディスクドライブ

156 光ディスク

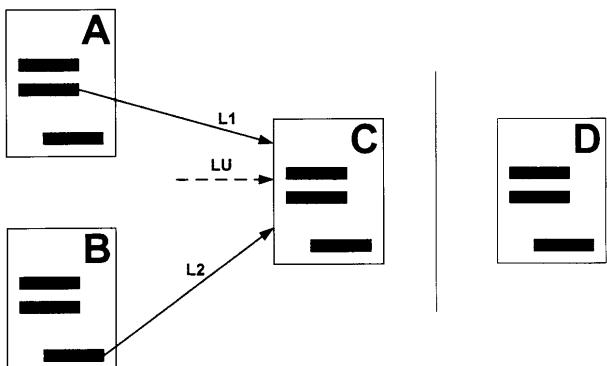
50

1 6 0 ユーザ入力インターフェース
 1 6 1 ポインティングデバイス
 1 6 2 キーボード
 1 7 0 ネットワークインターフェース
 1 7 1 ローカルエリアネットワーク (L A N)
 1 7 2 モデム
 1 7 3 広域ネットワーク (W A N)
 1 8 0 リモートコンピュータ
 1 8 1 メモリ記憶デバイス
 1 8 2 グラフィクスインターフェース
 1 8 4 グラフィクス処理ユニット (G P U)
 1 8 5 リモートアプリケーションプログラム
 1 8 6 ビデオメモリ
 1 9 0 ビデオインターフェース
 1 9 1 モニタ
 1 9 5 出力周辺インターフェース
 1 9 6 プリンタ
 1 9 7 スピーカ
 4 0 0 構造的に相互関係のある文書のソース
 4 1 0 クローラ
 4 2 0 リポジトリ
 4 3 0 オブジェクト
 4 3 2 , 4 3 4 A P I
 4 4 0 アプリケーションオブジェクト

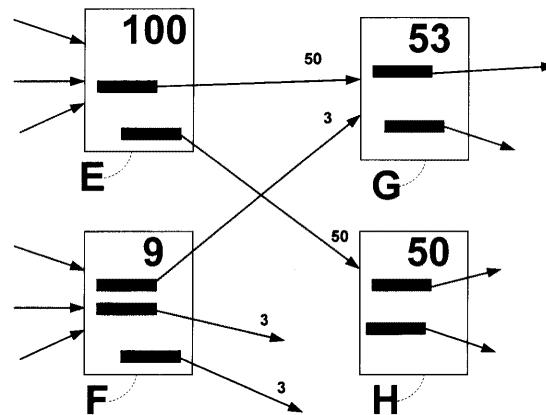
10

20

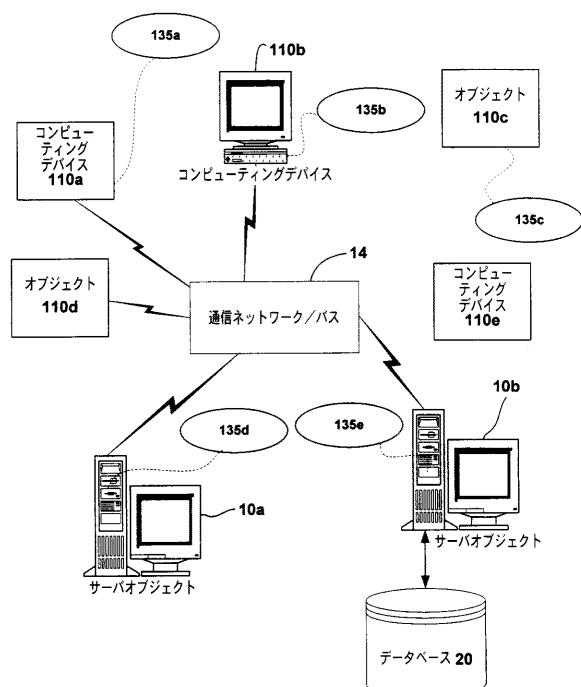
【図 1 A】



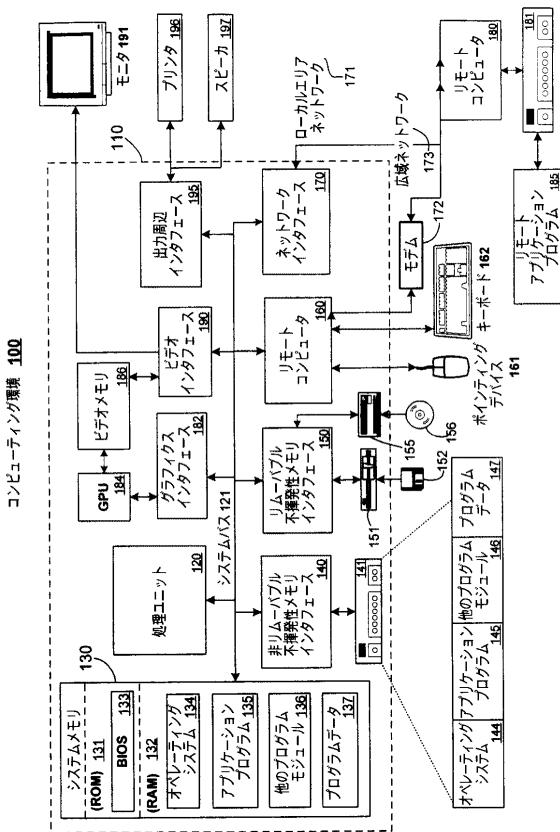
【図 1 B】



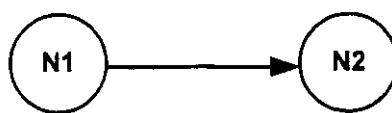
【図2A】



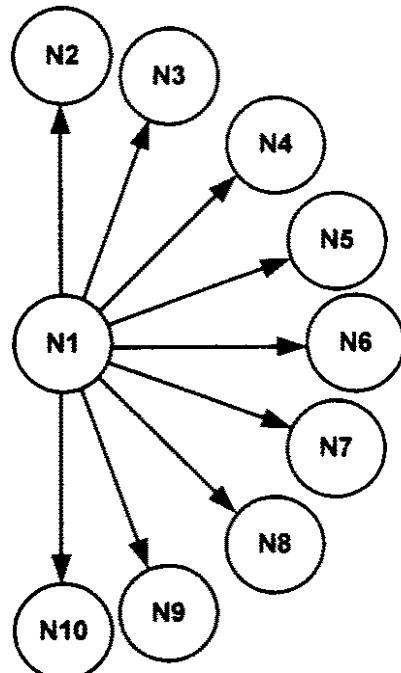
【図2B】



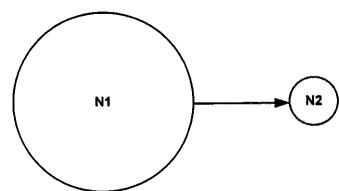
【図3A】



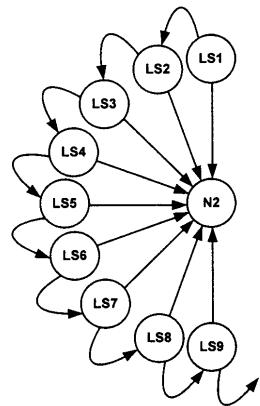
【図3B】



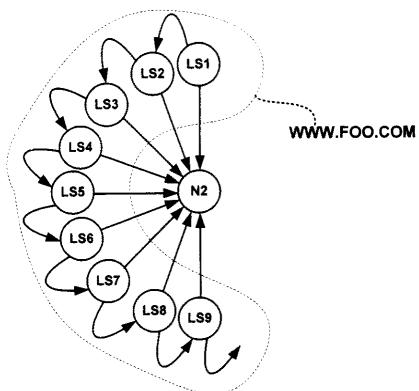
【図3C】



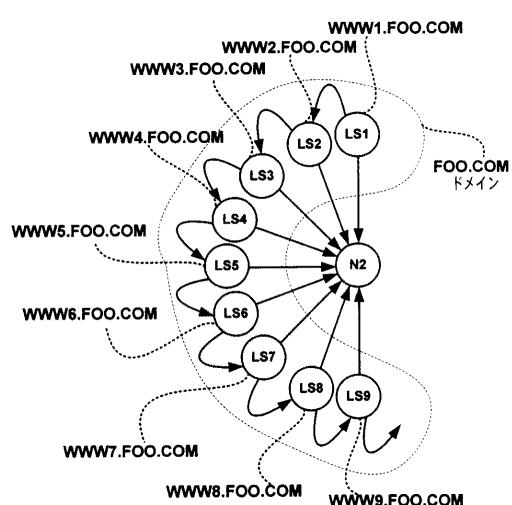
【図3D】



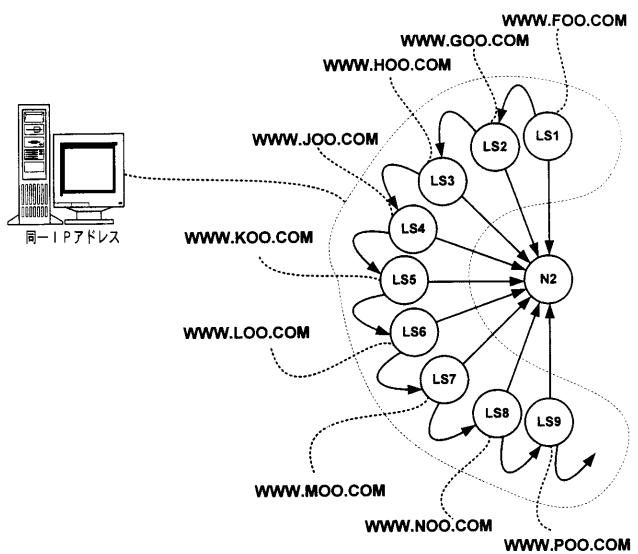
【図3E】



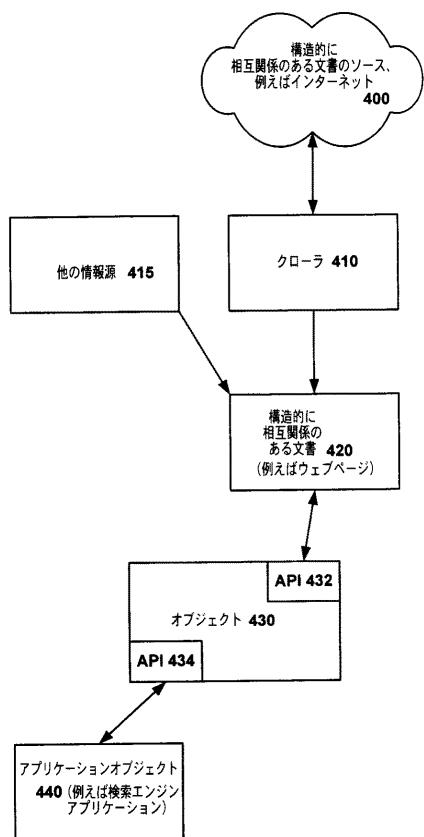
【図3F】



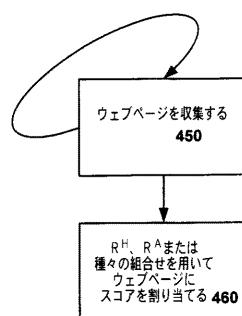
【図3G】



【図4A】



【図4B】



【図4C】

