

(19) 日本国特許庁(JP)

(12) 特許公報(B2)

(11) 特許番号

特許第4965086号  
(P4965086)

(45) 発行日 平成24年7月4日(2012.7.4)

(24) 登録日 平成24年4月6日(2012.4.6)

(51) Int. Cl. F I  
**G06F 17/30 (2006.01)**  
 G06F 17/30 220Z  
 G06F 17/30 180Z  
 G06F 17/30 350C

請求項の数 11 外国語出願 (全 14 頁)

(21) 出願番号	特願2005-141125 (P2005-141125)	(73) 特許権者	500046438
(22) 出願日	平成17年5月13日 (2005.5.13)		マイクロソフト コーポレーション
(65) 公開番号	特開2005-327293 (P2005-327293A)		アメリカ合衆国 ワシントン州 9805
(43) 公開日	平成17年11月24日 (2005.11.24)		2-6399 レッドモンド ワン マイ
審査請求日	平成20年5月13日 (2008.5.13)		クロソフト ウェイ
(31) 優先権主張番号	10/846,835	(74) 代理人	100077481
(32) 優先日	平成16年5月14日 (2004.5.14)		弁理士 谷 義一
(33) 優先権主張国	米国 (US)	(74) 代理人	100088915
			弁理士 阿部 和夫
		(72) 発明者	ベンユー チャン
			アメリカ合衆国 98052 ワシントン
			州 レッドモンド ワン マイクロソフト
			ウェイ マイクロソフト コーポレーシ
			ョン内

最終頁に続く

(54) 【発明の名称】 タイプ内およびタイプ間の関係に基づいてオブジェクトを格付けする方法およびシステム

(57) 【特許請求の範囲】

【請求項1】

プロセッサを備えたコンピュータが検索対象としての複数のデータオブジェクトの属性の属性値を決定する方法であって、

各データオブジェクトは一つのタイプに関連付けられ、該データオブジェクトのタイプは該タイプに固有の属性を有し、該属性は、該データオブジェクトの内向きリンクおよび外向きリンク、並びにWebログに基づいて決定され、

前記方法は、

前記プロセッサが複数のデータオブジェクトを識別するステップと、

前記識別されたデータオブジェクトに関連付けられた前記タイプの各々に対して、前記プロセッサが、該タイプに関連付けられる複数のデータオブジェクト間のタイプ内関係を識別し、および該タイプに関連付けられる一つのデータオブジェクトと他のタイプに関連付けられる他のデータオブジェクトとの間のタイプ間関係を識別するステップと、

前記識別されたデータオブジェクトに関連付けられた前記タイプの各々に対して、前記プロセッサが、該タイプの属性の属性値を計算する各タイプについての関数を使用して前記識別されたタイプ内関係およびタイプ間関係に基づいて前記データオブジェクトの属性の属性値を計算するステップであって、前記関数は、 $F_i$  がデータオブジェクト  $i$  に関連付けられた属性値を表し、 $R_i$  はデータオブジェクト  $i$  のタイプを持つ複数のデータオブジェクト間のタイプ内の関係を表し、および  $R_{j_i}$  はデータオブジェクト  $i$  のタイプを持つデータオブジェクトと他のデータオブジェクト  $j$  のタイプを持つデータオブジェクトと

10

20

の間のタイプ間の関係を表す、

$$F_i = F_i R_i + \sum_j F_j R_{j i}$$

として定義される、ステップと

前記プロセッサが、前記計算された属性値を格納するステップと  
を含み、

前記タイプ内関係は、同じタイプのデータオブジェクト間の関係であり、

前記タイプ間関係は、異なるタイプのデータオブジェクト間の関係であることを特徴とする方法。

【請求項 2】

プロセッサを備えたコンピュータが検索対象としての複数のデータオブジェクトの属性の属性値を決定する方法であって、

各データオブジェクトは一つのタイプに関連付けられ、該データオブジェクトのタイプは該タイプに固有の属性を有し、該属性は、該データオブジェクトの内向きリンクおよび外向きリンク、並びに Web ログに基づいて決定され、

前記方法は、

前記プロセッサが複数のデータオブジェクトを識別するステップと、

前記識別されたデータオブジェクトに関連付けられた前記タイプの各々に対して、前記プロセッサが、該タイプに関連付けられる複数のデータオブジェクト間のタイプ内関係を識別し、および該タイプに関連付けられる一つのデータオブジェクトと他のタイプに関連付けられる他のデータオブジェクトとの間のタイプ間関係を識別するステップと、

前記識別されたデータオブジェクトに関連付けられた前記タイプの各々に対して、前記プロセッサが、該タイプの属性の属性値を計算する各タイプについての関数を使用して前記識別されたタイプ内関係およびタイプ間関係に基づいて前記データオブジェクトの属性の属性値を計算するステップであって、前記関数は、

【数 1】

$$w_M = \alpha_M L_M^T w_M + \beta_{NM} \sum_{N \neq M} L_{NM}^T w_N$$

である、ステップと、

前記プロセッサが、前記計算された属性値を格納するステップと  
を含み、

前記タイプ内関係は、同じタイプのデータオブジェクト間の関係であり、

前記タイプ間関係は、異なるタイプのデータオブジェクト間の関係であることを特徴とする方法。

【請求項 3】

前記データオブジェクトのタイプは、オーソリティータイプ、ハブタイプ及び専門知識タイプを含むことを特徴とする請求項 1 または請求項 2 に記載の方法。

【請求項 4】

前記オーソリティータイプの複数のデータオブジェクト間の関係は、Web ページが別の Web ページからのリンクを有するかどうかに基づいて定まることを特徴とする請求項 3 に記載の方法。

【請求項 5】

前記オーソリティータイプのデータオブジェクトと前記専門知識タイプのデータオブジェクトとの間の関係は、ユーザによる Web ページへのアクセスに基づいて定まることを特徴とする請求項 4 に記載の方法。

【請求項 6】

前記ハブタイプの複数のデータオブジェクト間の関係は、Web ページが別の Web ページへのリンクを有するかどうかに基づいて定まることを特徴とする請求項 3 に記載の方法。

【請求項 7】

10

20

30

40

50

前記プロセッサが、あるタイプのデータオブジェクトを該オブジェクトの属性値に基づいて格付けするステップをさらに含むことを特徴とする請求項 1 または請求項 2 に記載の方法。

【請求項 8】

前記データオブジェクトのタイプの属性値を定義する各データオブジェクトのタイプについての式が規定されることを特徴とする請求項 1 または請求項 2 に記載の方法。

【請求項 9】

前記計算するステップは、前記式を繰り返し解くステップを含むことを特徴とする請求項 8 に記載の方法。

【請求項 10】

プロセッサを備えたコンピュータが検索対象としてのデータオブジェクトの属性の属性値を決定する方法であって、

各データオブジェクトは一つのタイプに関連付けられ、該データオブジェクトのタイプは該タイプに固有の属性を有し、該属性は、該データオブジェクトの内向きリンクおよび外向きリンク、並びに Web ログに基づいて決定され、

所定の関数は、あるタイプのデータオブジェクトについてのタイプに固有の属性の属性値を、該タイプのデータオブジェクトと、別のタイプに固有の属性を有する別のタイプのデータオブジェクトとの間の関係に基づいて計算するための関数であり、

前記所定の関数は、該タイプの複数のデータオブジェクト間の関係にも基づいて計算するための関数であり、

前記関数は、 $F_i$  がデータオブジェクト  $i$  に関連付けられた属性値を表し、 $R_i$  はデータオブジェクト  $i$  のタイプを持つ複数のデータオブジェクト間のタイプ内関係を表し、および  $R_{j_i}$  はデータオブジェクト  $i$  のタイプを持つデータオブジェクトと他のデータオブジェクト  $j$  のタイプを持つデータオブジェクトとの間のタイプ間関係を表す、

$F_i = F_i R_i + \sum_j R_{j_i} F_j R_{j_i}$   
として定義され、

前記方法は、

前記プロセッサが、前記あるタイプの複数のデータオブジェクト間のタイプ内関係を特定する第 1 のデータを受け取るステップと、

前記プロセッサが、前記あるタイプのデータオブジェクトと前記別のタイプのデータオブジェクトとの間のタイプ間関係を特定する第 2 のデータを受け取るステップと、

前記第 1 のデータおよび前記第 2 のデータを受け取ったことに応答して、前記プロセッサが、前記所定の関数を計算して、前記あるタイプのデータオブジェクトの属性値を決定するステップと、

前記プロセッサが、前記決定された属性値を格納するステップと  
を含み、

前記タイプ内関係は、同じタイプのデータオブジェクト間の関係であり、

前記タイプ間関係は、異なるタイプのデータオブジェクト間の関係であることを特徴とする方法。

【請求項 11】

プロセッサを備えたコンピュータが検索対象としての複数のデータオブジェクトの属性の属性値を決定す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読み取り可能な記録媒体であって、

各データオブジェクトは一つのタイプに関連付けられ、該データオブジェクトのタイプは該タイプに固有の属性を有し、該属性は、該データオブジェクトの内向きリンクおよび外向きリンク、並びに Web ログに基づいて決定され、

前記コンピュータ実行可能命令は、前記プロセッサによって実行される時、前記プロセッサに、

複数のデータオブジェクトを識別するステップと、

前記識別されたデータオブジェクトに関連付けられた前記タイプの各々に対して、該タ

10

20

30

40

50

タイプに関連付けられる複数のデータオブジェクト間のタイプ内関係を識別し、および該タイプに関連付けられる一つのデータオブジェクトと他のタイプに関連付けられる他のデータオブジェクトとの間のタイプ間関係を識別するステップと、

前記識別されたデータオブジェクトに関連付けられた前記タイプの各々に対して、該タイプの属性の属性値を計算する各タイプについての関数を使用して前記識別されたタイプ内関係およびタイプ間関係に基づいて前記データオブジェクトの属性の属性値を計算するステップであって、前記関数は、 $F_i$  がデータオブジェクト  $i$  に関連付けられた属性値を表し、 $R_i$  はデータオブジェクト  $i$  のタイプを持つ複数のデータオブジェクト間のタイプ内の関係を表し、および  $R_{j_i}$  はデータオブジェクト  $i$  のタイプを持つデータオブジェクトと他のデータオブジェクト  $j$  のタイプを持つデータオブジェクトとの間のタイプ間の関係を表す、

$F_i = F_i R_i + \sum_j R_{j_i} F_j R_{j_i}$   
として定義される、ステップと

前記計算された属性値を格納するステップと

を含む方法を実行させ、

前記タイプ内関係は、同じタイプのデータオブジェクト間の関係であり、

前記タイプ間関係は、異なるタイプのデータオブジェクト間の関係であることを特徴とするコンピュータ読み取り可能な記録媒体。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、一般には、オブジェクトの格付けに関し、より詳細には、オブジェクト関係に基づく格付けに関する。

【背景技術】

【0002】

GoogleおよびOvertureなど多くの検索エンジンサービスは、インターネットを介してアクセス可能な情報の検索を提供する。これらの検索エンジンサービスによりユーザは、Webページなど、ユーザにとって興味深いと思われる表示ページを検索することができる。検索語を含む検索要求（「クエリ」とも呼ばれる）をユーザが送信した後、検索エンジンサービスは、これらの検索語に関連していると思われるWebページを識別する。関連するWebページを迅速に識別するために、検索エンジンサービスは、Webページへのキーワードのマッピングを保持することができる。検索エンジンサービスは、Web（すなわちワールドワイドウェブ）を「クロール」することによりこのマッピングを生成して、各Webページのキーワードを抽出することができる。Webをクロールするために、検索エンジンサービスは、ルートのWebページのリストを使用し、これらのルートのWebページを通じてアクセス可能なすべてのWebページを識別することができる。特定のWebページのキーワードは、見出し語、Webページのメタデータで供給された語、ハイライト表示された語などを識別するような、様々な周知の情報検索技術を使用して抽出することができる。検索エンジンサービスにより、各々の合致の近さ、Webページの人気（例えばGoogleのPageRank）などに基づいて各Webページが検索要求にどの程度関連しているかを示す適合性の評点（score）を計算することができる。次に、検索エンジンサービスにより、これらのWebページへのリンクを適合性に基づいた順序でユーザに表示する。検索エンジンにより、さらに一般的に、あらゆる文書の集まりで情報の検索を提供することができる。例えば、文書の集まりには、すべての米国特許、すべての連邦の法廷意見、企業のすべての保管文書などを含めることもできる。

【0003】

Webページを格付けするための2つの周知の技術は、PageRankおよびHITS（「Hyperlinked-Induced Topic Search」）である。PageRankは、Webページが重要なWebページへのリンク（つまり「外向き

10

20

30

40

50

のリンク」)を持つという原則に基づく。したがって、Webページの重要度は、そのWebページにリンク(つまり「内向きのリンク」)する他のWebページの数および重要度に基づく。簡単な形式では、Webページ間のリンクは行列Aによって表すことができ、ここで $A_{ij}$ はWebページiからWebページjへの外向きのリンクの数を表す。Webページjの重要度の評点 $w_j$ は、以下の式で表すことができる。

$$w_j = \sum_i A_{ij} w_i$$

この式は、以下の式に基づく反復計算によって解くことができる。

$$A^T w = w$$

ただし、 $w$ はWebページの重要度の評点のベクトルであり、 $A^T$ の優固有ベクトル(principal eigenvector)である。

【0004】

HITS技術はさらに、他の重要なWebページへのリンクを数多く持つWebページはそれ自体が重要であるという原則に基づく。したがって、HITSは、Webページの「重要度」を、「ハブ」および「オーソリティー(authority)」という2つの関連する属性に分割する。「ハブ」は、WebページがリンクするWebページの「オーソリティー」の評点によって評価され、「オーソリティー」は、WebページにリンクするWebページの「ハブ」の評点によって評価される。Webページの重要度をクエリからは独立して計算するPage Rankとは対照的に、HITSでは、内向きおよび外向きのリンクをたどることによって結果のWebページおよび結果のWebページに関連するWebページに基づいて重要度を計算する。HITSは、検索エンジンサービスにクエリを送信し、結果のWebページをWebページの初期セットとして使用する。HITSは、内向きのリンクの宛先であるWebページおよび結果のWebページの外向きのリンクのソースであるWebページをセットに追加する。次にHITSは、反復アルゴリズムを使用して各Webページのオーソリティーおよびハブの評点を計算する。オーソリティーおよびハブの評点は、以下の式で表すことができる。

【0005】

【数2】

$$a(p) = \sum_{q \rightarrow p} h(q) \quad \text{および} \quad h(p) = \sum_{p \rightarrow q} a(q)$$

【0006】

ただし、 $a(p)$ はWebページpのオーソリティーの評点を表し、 $h(p)$ はWebページpのハブの評点を表す。HITSでは、隣接行列Aを使用してリンクを表す。隣接行列は、以下の式で表すことができる。

【0007】

【数3】

$$b_{ij} = \begin{cases} 1 & \text{ページiがページjへのリンクを有する場合、} \\ 0 & \text{ページiがページjへのリンクを有さない場合} \end{cases}$$

【0008】

ベクトルaおよびhはそれぞれ、セット内のすべてのWebページのオーソリティーおよびハブの評点に対応し、以下の式で表すことができる。

$$a = A^T h \quad \text{および} \quad h = A a$$

したがって、aおよびhは、行列 $A^T A$ および $A A^T$ の固有ベクトルである。HITSはまた、アクセス数によって評価されるWebページの人気を計算に入れるように変更することもできる。Webログの分析に基づいて、隣接行列の $b_{ij}$ は、ユーザがWebページiからWebページjに移動するときに増分することができる。

【発明の開示】

【発明が解決しようとする課題】

【0009】

上述のWebページ格付け技術では、Webページの格付けを主としてWebページ自

10

20

30

40

50

体の属性に基づいて行う。Web ページ自体の属性は、一方の Web ページから別の Web ページへのリンク、および一方の Web ページから別の Web ページへの移動を含む。格付け技術では、Web ページに直接関連がない属性を計算に入れることができない。例えば、Web ページの重要度は、Web ページにアクセスするユーザの専門知識が計算に入れられる場合、より正確に決定することができる。Web ページに直接関連がない属性に基づいて Web ページの重要度を計算するための技術を有することが望ましい。より一般的には、一方のタイプのオブジェクト（例えば、Web ページ）の評点を、別のタイプのオブジェクト（例えば、ユーザ）との関係に基づいて生成することが望ましい。

【課題を解決するための手段】

【0010】

異なるオブジェクトタイプのオブジェクトとの関係に基づいてオブジェクトを格付けするための方法およびシステムを提供する。格付けシステムにより、オブジェクトの各タイプの属性ごとに式を定義する。式は、属性値を定義し、属性と、同じタイプのオブジェクトおよび異なるタイプのオブジェクトに関連付けられる属性との間の関係に基づく。属性値は、一方の属性が別の属性に関して定義され、逆の場合も同じであるように相互依存することができるので、式は属性の再帰的定義を表す。格付けシステムにより、属性値が解に収束するまで式を使用して、オブジェクトの属性値を繰り返し計算する。次に、格付けシステムにより、属性値に基づいてオブジェクトの格付けを行う。

【発明を実施するための最良の形態】

【0011】

同じまたは別のデータオブジェクトタイプのデータオブジェクトとの関係に基づいてあるデータオブジェクトタイプのデータオブジェクトを格付けするための方法およびシステムを提供する。一実施形態において、格付けシステムにより、様々なデータオブジェクトタイプのデータオブジェクトを識別する。例えば、一方のデータオブジェクトタイプが Web ページであり、別のデータオブジェクトタイプがクエリであり、および別のデータオブジェクトタイプがユーザである場合がある。各データオブジェクトタイプは、様々なタイプ固有の属性を有することができる。例えば、Web ページがオーソリティー属性を有し、およびユーザが専門知識属性を有する場合がある。Web ページのオーソリティー属性は、Web ページの内向きのリンクの数に基づくことができる。ユーザの専門知識属性は、ユーザが高いオーソリティー属性値を持つ Web ページにアクセスする場合に増加される。格付けシステムにより、データオブジェクトの属性値を計算し、およびデータオブジェクトの属性値に基づいてデータオブジェクトを格付けすることができる。

【0012】

格付けシステムにより、各オブジェクトが単一の属性を含むように、オブジェクトの「タイプ」を定義する。例えば、格付けシステムにより、Web ページのオーソリティー属性に対応するタイプ、および Web ページのハブ属性に対応する別のタイプを定義することができる。したがって、2つのタイプは、同じ基礎をなすデータオブジェクト（例えば Web ページ）を表すことができる。格付けシステムにより、タイプ内の関係と呼ばれる同じタイプのオブジェクト間の様々な関係、およびタイプ間の関係と呼ばれる異なるタイプのオブジェクト間の様々な関係を決定する。例えば、クエリが送信される場合、格付けシステムにより、その結果をオーソリティータイプのオブジェクトとして使用し、および Web ログを使用して、これらの Web ページにアクセスしたユーザを専門知識タイプのオブジェクトとして識別することができる。オーソリティータイプのタイプ内の関係のオブジェクトは、Web ページの内向きのリンクおよび外向きのリンクの関係を含むことができる。例えば、Web ページが別の Web ページへのリンクを有する場合、Web ページはその他の Web ページへの外向きのリンクの関係を有し、およびその他の Web ページは Web ページへの内向きのリンクの関係を有する。オーソリティータイプおよび専門知識タイプのオブジェクト間のタイプ間の関係は、Web ページへのユーザアクセスに基づく。例えば、ユーザが Web ページにアクセスする場合、Web ページおよびユーザはアクセスの関係を有する。格付けシステムにより、タイプ内の関係および他のタイプのオ

10

20

30

40

50

プロジェクトの属性値と組み合わせられたタイプ間の関係を使用して、所定のタイプのオブジェクトに対する属性の値を導き出す。例えば、格付けシステムにより、内向きのリンクおよび外向きのリンクの関係およびユーザーアクセスの関係を使用して、Webページのオーソリティー属性およびハブ属性ならびにユーザの専門知識属性を導き出すことができる。

【0013】

一実施形態において、格付けシステムにより、一次方程式など一連の式を使用して関係および属性を表す。格付けシステムにより、別のタイプの属性値に基づいて再帰的に定義される一次方程式を使用して、各タイプの属性値を表す。例えば、オーソリティー属性の一次方程式が専門知識属性の属性値に基づいて定義されることも、またその逆の場合もある。一次方程式は再帰的に定義することができるので、格付けシステムにより、属性値が解に収束するまで、各一次方程式の属性値を繰り返し計算することによって一次方程式を解く。一次方程式を解いた後、格付けシステムにより、属性値に基づいてデータオブジェクトを格付けすることができる。例えば、格付けシステムにより、Webページのオーソリティー属性に基づいてWebページを格付けすることができる。

10

【0014】

格付けシステムにより、オブジェクトのタイプ内の関係およびタイプ間の関係に基づいて、オブジェクトの属性値を表す。属性の値は、以下の数式によって表すことができる。

$$F_i = F_i R_{ij} + F_j R_{ji}$$

【0015】

ただし、 $F_i$  はオブジェクト  $i$  に関連付けられる属性値を表し、 $R_{ij}$  はオブジェクト  $i$  のタイプのオブジェクト間のタイプ内関係を表し、 $R_{ji}$  はオブジェクト  $i$  のタイプと他のオブジェクト  $j$  のタイプとのオブジェクト間のタイプ間の関係を表す。 $x = \{x_1, x_2, \dots, x_m\}$  および  $y = \{y_1, y_2, \dots, y_n\}$  の2つのオブジェクトのタイプがある場合、これらのタイプ内の関係は  $R_x$  および  $R_y$  として表すことができ、タイプ間の関係は  $R_{xy}$  および  $R_{yx}$  によって表すことができる。格付けシステムでは、隣接行列を使用して関係情報を表す。 $L_x$  および  $L_y$  は、それぞれ集合  $X$  および  $Y$  のタイプ内の関係の隣接行列を表す。 $L_{xy}$  および  $L_{yx}$  は、それぞれ、 $X$  のオブジェクトから  $Y$  のオブジェクトへのタイプ間の関係の隣接行列および  $Y$  のオブジェクトから  $X$  のオブジェクトへのタイプ間の関係の隣接行列を表す。格付けシステムでは、以下のように隣接行列を表す。

20

【0016】

【数4】

$$L_{xy}(i, j) = \begin{cases} 1 & \text{オブジェクト } x_i \text{ からオブジェクト } y_j \text{ へ関係がある場合} \\ 0 & \text{オブジェクト } x_i \text{ からオブジェクト } y_j \text{ へ関係がない場合} \end{cases}$$

【0017】

ただし、 $L_{xy}(i, j)$  は、集合  $X$  のオブジェクト  $i$  から集合  $Y$  のオブジェクト  $j$  への関係（「リンク」とも呼ばれる）が存在するかどうかを示す。属性値の一次方程式は、以下の式のように表すことができる。

【0018】

【数5】

$$w_y = L_y^T w_y + L_{xy}^T w_x \quad (1)$$

$$w_x = L_x^T w_x + L_{yx}^T w_y$$

【0019】

ただし、 $w_x$  は、 $X$  のオブジェクトの属性ベクトルであり、 $w_y$  は  $Y$  のオブジェクトの属性ベクトルである。式1は、以下の形式に一般化することができる。

【0020】

30

40

## 【数 6】

$$W_M = L_M^T W_M + \sum_{\forall N \neq M} L_{NM}^T W_N \quad (2)$$

## 【0021】

ただし、Mは属性ベクトルの行列を表す。

## 【0022】

オブジェクトの間の相互に補強し合う関係がオブジェクトに過度の属性値をもたらすことがあるので、格付けシステムにより、オブジェクトが1つの隣接行列内のn個の他のオブジェクトに関連する場合、各関連するオブジェクトはその1/n番目の属性値を受け取るような方法で、バイナリの隣接行列を正規化することができる。格付けシステムにより、Page Rankのランダムサーファーマデルを導入して、ランダムな関係をシミュレートし、したがって、以下に説明する計算時のシンクノードを回避することができる。さらに、異なるタイプからの属性が各々他の属性に対して異なる重要度を有する可能性があるため、格付けシステムにより、タイプの各々の組み合わせに重み付けを使用することができる。したがって、格付けシステムにより、正規化、ランダムサーファーマデル、および重み付けを計算に入れて、以下の式によって属性値を表すことができる。

10

## 【0023】

## 【数 7】

$$w_M = \alpha_M L'_M{}^T w_M + \beta_{NM} \sum_{\forall N \neq M} L'_{NM}{}^T w_N$$

20

ただし、

$$\alpha_M + \sum_{\forall N \neq M} \beta_{NM} = 1; \alpha_M > 0 \beta_{NM} > 0; \quad (3)$$

$$L'_M = \varepsilon U + (1 - \varepsilon) L_M; 0 < \varepsilon < 1;$$

$$L'_{NM} = \delta_N U + (1 - \delta_N) L_{NM}; 0 < \delta_N < 1;$$

## 【0024】

ただし、Uは一様な推移確率の推移行列であり(すべてのi、jについて $U_{ij} = 1/n$ 、ただし、nはデータ空間Nのオブジェクトの合計数)、 $L_M$ および $L_{NM}$ は正規化された隣接行列であり、 $\varepsilon$ および $\delta_N$ は行列 $L_M$ および $L_{NM}$ 内のランダムな関係をシミュレートするために使用される平滑化因数であり、 $\alpha_M$ および $\beta_{NM}$ は関係の重み付けを表す。格付けシステムにより、収束するまで式3を繰り返し計算する。式3は、以下の式で表される統合された正方行列Aによって表すことができる。

30

## 【0025】

## 【数 8】

$$A = \begin{pmatrix} \alpha_1 L'_1 & \beta_{12} L'_{12} & \cdots & \beta_{1n} L'_{1n} \\ \beta_{21} L'_{21} & \alpha_2 L'_2 & \cdots & \beta_{2n} L'_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \beta_{n1} L'_{n1} & \beta_{n2} L'_{n2} & \cdots & \alpha_n L'_n \end{pmatrix} \quad (4)$$

40

## 【0026】

行列Aは、対角線上に

## 【0027】

## 【数 9】

$$L'_M$$

## 【0028】

、統合された行列の他の部分に

## 【0029】

## 【数 1 0】

 $L'_{NM}$ 

## 【0 0 3 0】

を有する。格付けシステムにより、行列  $A$  を使用して、異なるデータ空間にあるすべてのデータオブジェクトの属性ベクトルである、ベクトル  $w$  を変換するために反復アプローチを使用する（例えば、 $w = A^T w$ ）。反復が収束する場合、ベクトル  $w$  は行列  $A$  の優固有ベクトルである。

## 【0 0 3 1】

$M$  および  $N$  が異質のデータ空間である場合、格付けシステムでは、ランダムな関係を使用して無関係を表す。 $M$  のオブジェクトが  $N$  のオブジェクトにリンクの関係を持たない場合、部分行列

10

## 【0 0 3 2】

## 【数 1 1】

 $L'_{NM}^T$ 

## 【0 0 3 3】

はゼロであり、「シンクノード」を表す。計算では、シンクノードにすべての属性値を割り当てることがある。これを防ぐために、格付けシステムにより、部分行列

## 【0 0 3 4】

## 【数 1 2】

20

 $L'_{NM}^T$ 

## 【0 0 3 5】

の対応する行のすべての要素を  $1/n$  に設定する。ただし、 $n$  はデータ空間  $N$  のオブジェクトの合計数である。代替として、格付けシステムにより、望ましくないタイプ内およびタイプ間の関係に対して、対応する重み付けを 0 に設定することができる。しかし、 $M_N$  が 0 よりも大きい場合、反復計算が収束するのであれば  $N_M$  は 0 よりも大きい必要がある。したがって、

## 【0 0 3 6】

## 【数 1 3】

30

 $L'_{NM}^T$ 

## 【0 0 3 7】

の関係が望ましくない場合、格付けシステムにより、 $N_M$  を非常に小さい正の重みに設定して

## 【0 0 3 8】

## 【数 1 4】

 $L'_{NM}^T$ 

40

## 【0 0 3 9】

の影響を低減する。

## 【0 0 4 0】

すべての隣接行列を使用して統合された行列を構築することにより、格付けシステムでは、異なるオブジェクトのタイプを含む、統合されたデータ空間を構築する。したがって、以前のタイプ間の関係は、統合された空間のタイプ間の関係と見なすことができ、格付けシステムにより、効率的に単一のデータ空間のリンク解析を行う。

## 【0 0 4 1】

図 1 は、一実施形態における格付けシステムのコンポーネントを例示するフロー図である。格付けシステム 110 は、通信リンク 102 を介して様々な Web サイト 101 に接

50

続されている。格付けシステムには、オブジェクト収集コンポーネント 1 1 2、関係確立コンポーネント 1 1 3、評点計算コンポーネント 1 1 4、およびオブジェクト順序付けコンポーネント 1 1 5 を呼び出してオブジェクトを格付けするオブジェクト格付けコンポーネント 1 1 1 を含む。オブジェクト格付けコンポーネント 1 1 1 は、Web ページのセットを受け取り、タイプ内およびタイプ間の関係に基づいて Web ページを格付けすることができる。オブジェクト収集コンポーネント 1 1 2 は、様々なタイプのオブジェクトに関連する関係情報を取り出す。例えば、オブジェクト収集コンポーネント 1 1 2 は、Web サイトの Web ログにアクセスして、どのユーザがどの Web ページにアクセスするのかを識別することができる。関係確立コンポーネント 1 1 3 は、タイプ内およびタイプ間の関係の行列を作成する。例えば、関係の行列は、ユーザがアクセスする Web ページにユーザをマッピングすることができる。評点計算コンポーネント 1 1 4 は、属性値が解に収束するまで、式 3 を使用して属性値を再帰的に計算する。オブジェクト順序付けコンポーネント 1 1 5 は、属性値に基づいてデータオブジェクトをソートする。例えば、オブジェクト順序付けコンポーネント 1 1 5 は、Web ページの オーソリティー 属性の値を使用して、Web ページをソートすることができる。

10

**【 0 0 4 2 】**

格付けシステムが実装されるコンピューティング装置には、中央処理装置、メモリ、入力装置（例えば、キーボードおよびポインティングデバイスなど）、出力装置（例えば、ディスプレイ装置など）、および記憶装置（例えば、ディスクドライブなど）を含めることができる。メモリおよび記憶装置は、格付けシステムを実装する命令を含めることができるコンピュータ読取可能な媒体である。さらに、データ構造およびメッセージ構造を、通信リンク上の信号などのデータ伝送媒体を介して格納または送信することができる。インターネット、ローカルエリアネットワーク、ワイドエリアネットワーク、またはポイントツーポイントのダイヤルアップ接続など、様々な通信リンクを使用することができる。

20

**【 0 0 4 3 】**

格付けシステムを、様々なオペレーティング環境において実装することができる。使用に最適な様々な周知のコンピューティングシステム、環境、および構成には、パーソナルコンピュータ、サーバコンピュータ、ハンドヘルドまた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に基づくシステム、プログラム可能な家庭用電化製品、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、上記のシステムまたは装置のいずれかを含む分散コンピューティング環境などが含まれる。

30

**【 0 0 4 4 】**

格付けシステムにより、1 つまたは複数のコンピュータまたは他の装置によって実行されるプログラムモジュールなど、コンピュータ実行可能命令の一般的なコンテキストにおいて説明することができる。一般に、プログラムモジュールには、特定のタスクを実行するかまたは特定の抽象データ型を実装するルーチン、プログラム、オブジェクト、コンポーネント、データ構造などが含まれる。通常、プログラムモジュールの機能を、様々な実施形態において望ましいように組み合わせまたは分散することができる。

**【 0 0 4 5 】**

図 2 は、一実施形態におけるオブジェクト格付けコンポーネントの処理を例示するフロー図である。コンポーネントは、オブジェクト情報を収集し、オブジェクト間の関係を確立し、オブジェクトの属性値を計算し、属性に基づいてオブジェクトを順序付ける。ブロック 2 0 1 において、コンポーネントは、様々なオブジェクトに関連する情報を収集する。ブロック 2 0 2 において、コンポーネントは、関係確立コンポーネントを呼び出して、隣接行列を生成する。関係確立コンポーネントはまた、および の重みを取り出して調整することもできる。ブロック 2 0 3 において、コンポーネントは、評点計算コンポーネントを呼び出して、属性値が解に収束するまで属性値を繰り返し計算する。ブロック 2 0 4 において、コンポーネントは、属性の値に基づいてデータオブジェクトを順序付ける。例えば、コンポーネントは、オーソリティー 属性に基づいて Web ページを順序付けることができる。

40

50

## 【0046】

図3は、一実施形態における関係確立コンポーネントの処理を例示するフロー図である。ブロック301から303において、コンポーネントは、各タイプに対する隣接行列を確立するループを行う。ブロック301において、コンポーネントは次のタイプを選択する。判断ブロック302において、すべてのタイプがすでに選択されている場合、コンポーネントは戻り、別の場合、コンポーネントはブロック303において続行する。ブロック303において、コンポーネントは、選択されたタイプのオブジェクトとすべてのタイプのオブジェクトとの間の関係を確立する。例えば、コンポーネントは、オーソリティータイプとハブタイプとの間の関係、およびオーソリティータイプと専門知識タイプとの間の関係を確立する。次に、コンポーネントはブロック301へのループを行い、次のタイプを選択する。

10

## 【0047】

図4は、一実施形態における評点計算コンポーネントの処理を例示するフロー図である。コンポーネントは、属性値が収束するまで式を繰り返し計算する。ブロック401において、コンポーネントは、隣接行列によって表されたオブジェクトの関係を取り出す。ブロック402において、コンポーネントは、タイプ内およびタイプ間の関係に対する重みおよび初期化を取り出す。ブロック403において、コンポーネントは、各タイプのベクトル $w$ を初期化して、そのタイプの各オブジェクトに対し等しい属性値を持つようにする。コンポーネントは、各値を $1/m$ に設定することができる。ただし、 $m$ はタイプのオブジェクトの数である。例えば、10のユーザがある場合、コンポーネントは専門知識タイプの初期の属性値を $1/10$ に設定する。コンポーネントはまた、判断ブロック405の検査を最初に通過するように、各タイプの差異変数を大きい値に初期化する。コンポーネントは、計算が解に収束しているかどうかを決定するために、各反復の最後において各差異変数に新しい値を計算する。ブロック404から409において、コンポーネントは、計算が解に収束するまで式3の計算を実行する。ブロック404において、コンポーネントは次の反復を開始する。判断ブロック405において、前回の反復中に計算された差異の合計が差異のしきい値よりも小さい場合、計算は解に収束しており、コンポーネントは戻り、別の場合、コンポーネントはブロック406において続行する。ブロック406において、コンポーネントは次のタイプを選択する。判断ブロック407において、すべてのタイプがすでに選択されている場合、コンポーネントはブロック404にループを行い、次の反復を開始し、別の場合、コンポーネントはブロック408において続行する。ブロック408において、コンポーネントは、以前の反復において計算された値に基づいて、選択されたタイプの値を計算する。ブロック409において、コンポーネントは、この反復の値と選択されたタイプの以前の反復の値との間の差異を計算する。次に、コンポーネントはブロック406にループを行い、次のタイプを選択する。

20

30

## 【0048】

本明細書において格付けシステムの特定の実施形態について例示のために説明しているが、本発明の精神および範囲を逸脱することなく、様々な変更を行うことができることを、当事業者であれば理解するであろう。例えば、当事業者であれば、属性値を表すために非線形方程式を使用することができることを理解するであろう。また、格付けシステムは、相互に何らかの関係を有するすべてのタイプのオブジェクトに使用することができる。例えば、格付けシステムにより、学生または志願者と教授との関係を使用する「重要度」に基づいて大学を格付けするために使用することもできる。ここで、大学、学生、および教授は異なるタイプのオブジェクトを表す。したがって、本発明は、添付の請求の範囲による場合を除き、限定されることはない。

40

## 【図面の簡単な説明】

## 【0049】

【図1】一実施形態における格付けシステムのコンポーネントを例示するフロー図である。

【図2】一実施形態におけるオブジェクト格付けコンポーネントの処理を例示するフロー

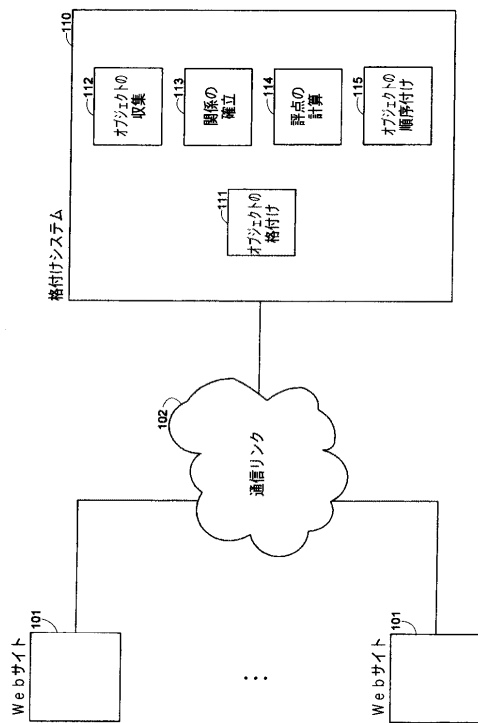
50

図である。

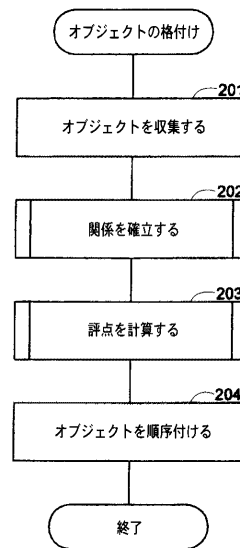
【図3】一実施形態における関係確立コンポーネントの処理を例示するフロー図である。

【図4】一実施形態における評点計算コンポーネントの処理を例示するフロー図である。

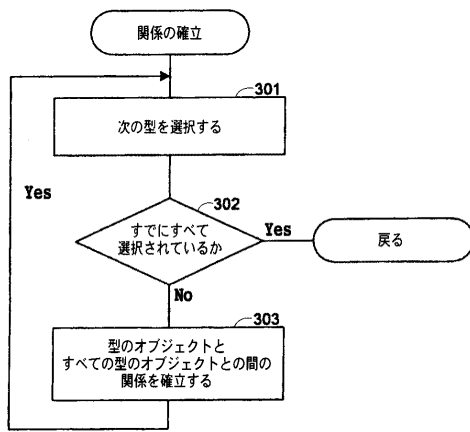
【図1】



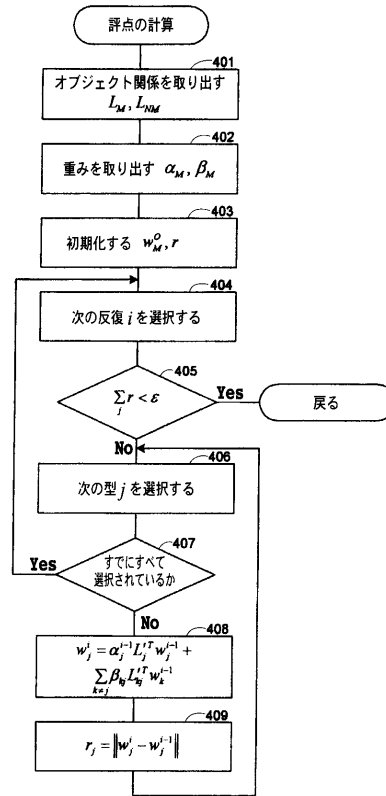
【図2】



【図3】



【図4】



## フロントページの続き

- (72)発明者 ホア - ジュン チェン  
アメリカ合衆国 98052 ワシントン州 レッドモンド ワン マイクロソフト ウェイ マ  
イクロソフト コーポレーション内
- (72)発明者 ウェイ - イン マ  
アメリカ合衆国 98052 ワシントン州 レッドモンド ワン マイクロソフト ウェイ マ  
イクロソフト コーポレーション内
- (72)発明者 ウェンシー シー  
アメリカ合衆国 98052 ワシントン州 レッドモンド ワン マイクロソフト ウェイ マ  
イクロソフト コーポレーション内
- (72)発明者 ツェン チェン  
アメリカ合衆国 98052 ワシントン州 レッドモンド ワン マイクロソフト ウェイ マ  
イクロソフト コーポレーション内

審査官 長 由紀子

- (56)参考文献 平林 真実 外5名, リンクに基づいた明示的Webページ評価法, 情報処理学会論文誌 第4  
3巻 No. SIG12(TOD16), 日本, 社団法人情報処理学会, 2002年12月1  
5日, 第43巻, p.92-102  
向 亨 外2名, 利用履歴に基づくPageRankアルゴリズムの改良, 第13回データ工学  
ワークショップ(DEWS2002)論文集 [online], 日本, 電子情報通信学会デー  
タ工学研究専門委員会, 2002年 5月15日, 論文番号A1-2  
絹川 達也 外4名, 作成者責任の概念を用いたWebページ評価法のサイトへの適用実験, 電  
子情報通信学会技術研究報告 Vol. 101 No. 192, 日本, 社団法人電子情報通信学  
会, 2001年 7月11日, 第101巻第192号, p.113-120

(58)調査した分野(Int.Cl., DB名)

G06F 17/30