

(19) 日本国特許庁(JP)

(12) 特 許 公 報(B2)

(11) 特許番号

特許第5928248号  
(P5928248)

(45) 発行日 平成28年6月1日(2016.6.1)

(24) 登録日 平成28年5月13日(2016.5.13)

(51) Int.Cl. F 1  
**G 0 6 F 17/30 (2006.01)**  
 G 0 6 F 17/30 4 1 9 B  
 G 0 6 F 17/30 2 2 0 Z

請求項の数 7 (全 28 頁)

(21) 出願番号 特願2012-186316 (P2012-186316)  
 (22) 出願日 平成24年8月27日 (2012.8.27)  
 (65) 公開番号 特開2014-44552 (P2014-44552A)  
 (43) 公開日 平成26年3月13日 (2014.3.13)  
 審査請求日 平成27年5月12日 (2015.5.12)

(73) 特許権者 000005223  
 富士通株式会社  
 神奈川県川崎市中原区上小田中4丁目1番  
 1号  
 (74) 代理人 100092152  
 弁理士 服部 毅巖  
 (72) 発明者 清水 俊宏  
 神奈川県川崎市中原区上小田中4丁目1番  
 1号 富士通株式会社内  
 審査官 吉田 誠

最終頁に続く

(54) 【発明の名称】 評価方法、情報処理装置およびプログラム

(57) 【特許請求の範囲】

【請求項1】

複数のデータそれぞれが他のデータを指し示す関係性であるリンクをもっており、前記複数のデータそれぞれの評価値を、他のデータの評価値と他のデータがもつリンクとに基づいて評価する評価方法であって、情報処理装置が、

第1のデータの評価値が更新されると、前記第1のデータの評価値に基づいて、少なくとも前記第1のデータからリンクされる1または複数の第2のデータそれぞれの評価値を評価して更新し、

何れかのデータの評価値の参照要求を受け付けると、少なくとも当該データへのリンクをもつ1または複数の第3のデータそれぞれの評価値に基づいて、参照要求を受けた当該データの評価値を評価し出力する、評価方法。

【請求項2】

前記第1のデータの評価値が更新されると、前記第1のデータから前記第2のデータを介してリンクされるデータのうち、前記第1のデータの評価値の変化分に対して各データの評価値に反映される変化分の割合が第1の閾値以上であるデータの評価値を、前記第2のデータの評価値に基づいて更新し、

第4のデータの前記参照要求を受け付けると、前記第4のデータへ前記第3のデータを介してリンクするデータのうち、前記第4のデータの評価値に反映される各データの評価値の割合が第2の閾値以上であるデータの評価値に基づいて、前記第4のデータの評価値を評価する、請求項1記載の評価方法。

10

20

**【請求項 3】**

何れかのデータの評価値の更新がクライアント装置により要求される頻度と、クライアント装置から前記参照要求を受け付ける頻度と、に基づいて、前記第 1 の閾値および前記第 2 の閾値を変更する、請求項 2 記載の評価方法。

**【請求項 4】**

前記第 1 の閾値を増加させ前記第 2 の閾値を減少させるか、または、前記第 1 の閾値を減少させ前記第 2 の閾値を増加させるように、前記第 1 の閾値および前記第 2 の閾値を変更する、請求項 3 記載の評価方法。

**【請求項 5】**

前記第 1 の閾値と前記第 2 の閾値との積が一定になるように、前記第 1 の閾値および前記第 2 の閾値を変更する、請求項 3 または 4 記載の評価方法。

10

**【請求項 6】**

複数のデータそれぞれが他のデータを指し示す関係性であるリンクをもっており、前記複数のデータそれぞれの評価値を、他のデータの評価値と他のデータがもつリンクとに基づいて評価するために用いられる情報処理装置であって、

前記複数のデータおよび前記複数のデータそれぞれの評価値を記憶する記憶部と、

前記記憶部を参照して、第 1 のデータの評価値が更新されると、前記第 1 のデータの評価値に基づいて、少なくとも前記第 1 のデータからリンクされる 1 または複数の第 2 のデータそれぞれの評価値を評価して更新し、また、何れかのデータの評価値の参照要求を受け付けると、少なくとも当該データへのリンクをもつ 1 または複数の第 3 のデータそれぞれの評価値に基づいて、参照要求を受けた当該データの評価値を評価し出力する演算部と、を有する情報処理装置。

20

**【請求項 7】**

複数のデータそれぞれが他のデータを指し示す関係性であるリンクをもっており、前記複数のデータそれぞれの評価値を、他のデータの評価値と他のデータがもつリンクとに基づいて評価するためのプログラムであって、コンピュータに、

第 1 のデータの評価値が更新されると、前記第 1 のデータの評価値に基づいて、少なくとも前記第 1 のデータからリンクされる 1 または複数の第 2 のデータそれぞれの評価値を評価して更新し、

何れかのデータの評価値の参照要求を受け付けると、少なくとも当該データへのリンクをもつ 1 または複数の第 3 のデータそれぞれの評価値に基づいて、参照要求を受けた当該データの評価値を評価し出力する、処理を実行させるプログラム。

30

**【発明の詳細な説明】****【技術分野】****【0001】**

本発明は評価方法、情報処理装置およびプログラムに関する。

**【背景技術】****【0002】**

情報処理装置からネットワークを介して他の装置上に蓄積されたデータ（例えば、文書データ）にアクセスし、当該データの内容を閲覧することがある。データの蓄積量が大量になると、閲覧したい内容を含むデータをユーザ自身が探索するのに手間がかかる。そこで、検索システムが利用されている。検索システムは、例えばキーワードの入力を受け付け、キーワードを含む、または、キーワードに関連するデータを検索し出力する。

40

**【0003】**

例えば、文書データの一例に Web ページが挙げられる。Web ページの検索システムには Web ページをランク付けして提示するものがある。ランクとは人気度や重要度などを示す指標である。検索システムは、例えば検索された Web ページのうち、ランクの高いものを相対的に上位に提示し、ランクの低いものを相対的に下位に提示する。

**【0004】**

Web ページのランク付けの方法として、ページランク（PAGE RANK、登録商標

50

と呼ばれる方法がある。Webページはリンクをもつ。リンクは他のWebページを指し示す情報である。ページランクでは、各Webページのバックリンク (backlink) とフォワードリンク (forwardlink) とに基づいて各Webページをランク付けする。バックリンクはWebページが他のWebページから指し示されるリンクである。フォワードリンクはWebページから他のWebページを指し示すリンクである。例えば、Webページ間のリンクの関係性は隣接行列で表され得る。すると、ページランクによるランク付けは隣接行列 (または隣接行列に調整を加えた行列) の固有ベクトルを求める問題に帰着される。行列の固有ベクトルは、例えばベキ乗法 (Power Method) を用いて計算され得る。

【0005】

ところで、Webページの数膨大である。このため、ページランクの方法を用いて全てのWebページのランクを精度良く求めるとすると計算に時間がかかる。そこで、近似的なランクで代替することで計算コストを削減する提案がある。この提案では、収集されたページの近似PageRank値をリンク先に等分配することでリンク先の近似PageRank値を更新する。リンク先のページでも分配された分を更にそのリンク先へと分配し、それを繰り返す。分配を適当なところで打ち切ることによって計算コストを削減する。

10

【先行技術文献】

【特許文献】

【0006】

【特許文献1】米国特許第6285999号明細書

【非特許文献】

20

【0007】

【非特許文献1】Page、外3名、“The PageRank Citation Ranking: Bringing Order to the Web.”、[online]、2001年10月30日、The Stanford University InfoLab、[2012年2月23日検索]、インターネット<URL: <http://ilpubs.stanford.edu:8090/422/1/1999-66.pdf>>

【非特許文献2】山田、外3名、“インクリメンタルPageRankによる重要Webページの効率的な収集戦略”、情報処理学会論文誌: コンピューティングシステム、社団法人情報処理学会、2004年10月、Vol.45、No.SIG 12、p.465-473

【発明の概要】

【発明が解決しようとする課題】

30

【0008】

上記のように、複数のデータ (例えば、文書データ) それぞれの評価値 (例えば、ランク) を、他のデータの評価値と他のデータがもつリンクとに基づいて評価する方法が考えられる。また、このような方法において、リンク元のデータの評価値をリンク先のデータに分配して評価値の更新処理を行っていき、分配を途中で打ち切ることが考えられる。

【0009】

ところが、分配を途中で打ち切ると、評価値の更新の影響を受けるリンク先の範囲が制限されることになる。すると、その範囲外のデータの評価値を参照するとき、当該データの評価値を単に読み取るだけでは、他のデータの更新の影響が当該データの評価値に反映されていないために、読み取った評価値の精度が低い可能性があるという問題がある。

40

【0010】

一側面によれば、本発明は、評価の信頼性を高めることができる評価方法、情報処理装置およびプログラムを提供することを目的とする。

【課題を解決するための手段】

【0011】

一実施態様によれば、複数のデータそれぞれが他のデータを指し示す関係性であるリンクをもっており、複数のデータそれぞれの評価値を、他のデータの評価値と他のデータがもつリンクとに基づいて評価する評価方法が提供される。この評価方法では、情報処理装置が、第1のデータの評価値が更新されると、第1のデータの評価値に基づいて、少なくとも第1のデータからリンクされる1または複数の第2のデータそれぞれの評価値を評価

50

して更新する。何れかのデータの評価値の参照要求を受け付けると、少なくとも当該データへのリンクをもつ1または複数の第3のデータそれぞれの評価値に基づいて、参照要求を受けた当該データの評価値を評価し出力する。

【0012】

また、一実施態様によれば、複数のデータそれぞれが他のデータを指し示す関係性であるリンクをもっており、複数のデータそれぞれの評価値を、他のデータの評価値と他のデータがもつリンクとに基づいて評価するために用いられる情報処理装置が提供される。この情報処理装置は、記憶部と演算部とを有する。記憶部は、複数のデータおよび複数のデータそれぞれの評価値を記憶する。演算部は、記憶部を参照して、第1のデータの評価値が更新されると、第1のデータの評価値に基づいて、少なくとも第1のデータからリンクされる1または複数の第2のデータそれぞれの評価値を評価して更新する。演算部は、何れかのデータの評価値の参照要求を受け付けると、少なくとも当該データへのリンクをもつ1または複数の第3のデータそれぞれの評価値に基づいて、参照要求を受けた当該データの評価値を評価し出力する。

10

【0013】

また、一実施態様によれば、コンピュータによって実行されるプログラムであって、複数のデータそれぞれが他のデータを指し示す関係性であるリンクをもっており、複数のデータそれぞれの評価値を、他のデータの評価値と他のデータがもつリンクとに基づいて評価するためのプログラムが提供される。このプログラムは、コンピュータに、第1のデータの評価値が更新されると、第1のデータの評価値に基づいて、少なくとも第1のデータからリンクされる1または複数の第2のデータそれぞれの評価値を評価して更新し、何れかのデータの評価値の参照要求を受け付けると、少なくとも当該データへのリンクをもつ1または複数の第3のデータそれぞれの評価値に基づいて、参照要求を受けた当該データの評価値を評価し出力する、処理を実行させる。

20

【発明の効果】

【0014】

一実施態様によれば、評価の信頼性を高めることができる。

【図面の簡単な説明】

【0015】

【図1】第1の実施の形態の情報処理装置を示す図である。

30

【図2】第2の実施の形態の情報処理システムを示す図である。

【図3】第2の実施の形態のグラフの例を示す図である。

【図4】第2の実施の形態のサーバのハードウェア例を示す図である。

【図5】第2の実施の形態のサーバのソフトウェア例を示す図である。

【図6】第2の実施の形態のWebページ管理テーブルを示す図である。

【図7】第2の実施の形態のキューを示す図である。

【図8】第2の実施の形態の更新範囲の例を示す図である。

【図9】第2の実施の形態の参照範囲の例を示す図である。

【図10】第2の実施の形態のランク更新を示すフローチャートである。

【図11】第2の実施の形態の更新時影響度の例を示す図である。

40

【図12】第2の実施の形態のランク参照を示すフローチャートである。

【図13】第2の実施の形態の参照時影響度の例を示す図である。

【図14】第3の実施の形態のログテーブルの例を示す図である。

【図15】第3の実施の形態の更新・参照範囲の変更を示すフローチャートである。

【図16】第3の実施の形態の更新・参照範囲の変更例(その1)を示す図である。

【図17】第3の実施の形態の更新・参照範囲の変更例(その2)を示す図である。

【発明を実施するための形態】

【0016】

以下、本実施の形態を図面を参照して説明する。

[第1の実施の形態]

50

図1は、第1の実施の形態の情報処理装置を示す図である。情報処理装置1は、複数のデータそれぞれの評価値を評価するために用いられる。ここで、複数のデータそれぞれは他のデータを指し示す関係性であるリンクをもっている。例えば、データは文書データである。例えば、文書データがWebページであれば、リンクはWebページのHTML (HyperText Markup Language) ファイル内にハイパーテキストにより記述されたURL (Uniform Resource Locator) である。当該リンクはハイパーリンクと呼ばれることもある。また、例えば文書データが論文データであれば、リンクは引用文献として記述された他の文献の名称である。データは、文書データの識別情報と当該文書データがもつリンクとを管理できれば、文書データそのものでなくてもよい。データは、リンクの情報をもつものであれば、文書データ以外のものでもよい。

10

## 【0017】

情報処理装置1は、複数のデータそれぞれの評価値を、他のデータの評価値と他のデータがもつリンクとに基づいて評価する。評価値は、例えば、人気度、重要度およびお薦め度などを示す指標(ランク)である。具体的には、リンク元のデータの評価値を、リンク先のデータに分配する。あるデータが複数のリンク元のデータから評価値の分配を受ける場合、分配を受けた評価値の合計が当該データの評価値である。そして、当該評価値を当該データのリンク先のデータに更に分配する。この分配を繰り返した結果を得ることで、複数のデータそれぞれの評価値を得る。

## 【0018】

情報処理装置1は、記憶部1aおよび演算部1bを有する。記憶部1aはRAM (Random Access Memory) などのメモリである。演算部1bはCPU (Central Processing Unit) などのプロセッサである。例えば、記憶部1aに記憶されたプログラムを演算部1bが実行することで、第1の実施の形態の情報処理を実現できる。

20

## 【0019】

記憶部1aは、複数のデータおよび複数のデータそれぞれの評価値を記憶する。複数のデータそれぞれは、前述のように他のデータへのリンクの情報を有している。

例えば、記憶部1aはデータX, X1, X2, X11, X12, X13, X14, X15, X16, X17, ...を記憶する。データXはデータX1, X2へのリンクを有する。データX1はデータX11, X12へのリンクを有する。データX2はデータX12, X13へのリンクを有する。データX11はデータX14へのリンクを有する。データX12はデータX15, X16へのリンクを有する。データX13はデータX17へのリンクを有する。グラフ構造2はこれらデータ間のリンクの関係性を示している。

30

## 【0020】

更に、例えば、記憶部1aはデータY, Y1, Y2, Y11, Y12, Y13, Y14, Y15, Y16, Y17, ...を記憶する。データY1, Y2はデータYへのリンクを有する。データY11はデータY1へのリンクを有する。データY12はデータY1, Y2へのリンクを有する。データY13はデータY2へのリンクを有する。データY14はデータY11へのリンクを有する。データY15, Y16はデータY12へのリンクを有する。データY17はデータY13へのリンクを有する。グラフ構造3はこれらデータ間のリンクの関係性を示している。

40

## 【0021】

演算部1bは、記憶部1aを参照して、第1のデータの評価値が更新されると、第1のデータの評価値に基づいて、少なくとも第1のデータからリンクされる1または複数の第2のデータそれぞれの評価値を評価して更新する。

## 【0022】

例えば、データXの評価値が更新される。具体的には、データXに対する別のデータからのリンクが追加されると、データXの評価値が増大し得る。当該別のデータから評価値の分配を受けるからである。あるいは、データXに対する別のデータからのリンクが削除されると、データXの評価値が減少し得る。当該別のデータから評価値の分配を受けられなくなるからである。

50

## 【 0 0 2 3 】

演算部 1 b は、データ X の評価値が更新されると、データ X の評価値に基づいて、少なくともデータ X からリンクされるデータ X 1 , X 2 それぞれの評価値を評価して更新する。例えば、データ X 1 , X 2 から更にリンクされるデータ X 1 1 , X 1 2 , X 1 3 それぞれの評価値を評価して更新してもよい。例えば、データ X の評価値が増大する場合、評価値の増分をデータ X 1 , X 2 に分配する。データ X 1 1 , X 1 2 , X 1 3 にも分配する場合、データ X 1 , X 2 に分配された分を更にデータ X 1 1 , X 1 2 , X 1 3 に分配する。分配の繰り返しにより、データ X の評価値の増大による影響が他のデータに伝搬する。

## 【 0 0 2 4 】

例えば、他のデータが受ける当該影響の度合い（更新時影響度）は、データ X の評価値の変化分に対して他の各データの評価値に反映される変化分の割合により表せる。例えば、データ X の評価値の変化分の 25% がデータ X 1 1 の評価値の変化分となるならば、データ X からデータ X 1 1 が受ける更新時影響度は 25% ( = 0 . 25 ) である。

10

## 【 0 0 2 5 】

例えば、データ X はデータ X 1 , X 2 の 2 つに増分を分配する。このとき、データ X の増分によるデータ X 自身の影響度 1 に対して、データ X 1 , X 2 それぞれが受ける影響度は、 $1 / 2 = 0 . 5$  である。例えば、データ X 1 はデータ X 1 1 , X 1 2 の 2 つに更に分配する。このとき、データ X 1 1 , X 1 2 それぞれが受ける影響度は  $0 . 5 \times 0 . 5 = 0 . 25$  である。データ X の評価値が減少する場合は、評価値の減少分を同様に分配して各データの評価値を減少させればよい。このときも増加時と同様の考え方で更新時影響度を表せる。

20

## 【 0 0 2 6 】

ここで、更新範囲 2 a は更新対象となるデータの集合である。更新範囲 2 a は、更新時の影響度閾値（ $0 < \text{閾値} < 1$  の実数）によって定めることができる。更新時の影響度閾値は、各データに対する分配の影響をどこまで伝搬させるかを示す情報である。例えば、各データが受ける更新時影響度が 以上であるデータの集合が更新範囲 2 a である。

## 【 0 0 2 7 】

演算部 1 b は、何れかのデータの評価値の参照要求を受け付けると、少なくとも当該データへのリンクをもつ 1 または複数の第 3 のデータそれぞれの評価値に基づいて、参照要求を受けた当該データの評価値を評価し出力する。

30

## 【 0 0 2 8 】

例えば、データ Y の評価値の参照要求を受け付ける。すると、演算部 1 b は、少なくともデータ Y へのリンクをもつデータ Y 1 , Y 2 それぞれの評価値に基づいて、参照要求を受けたデータ Y の評価値を評価し出力する。ここで、例えばデータ Y はデータ Y 1 , Y 2 のみからリンクされているとする。例えば、データ Y 1 , Y 2 はデータ Y へのリンクのみを有するとする。このとき、例えばデータ Y 1 , Y 2 の評価値のみに基づいてデータ Y の評価値を評価するならば、データ Y 1 , Y 2 の評価値の和がデータ Y の評価値である。

## 【 0 0 2 9 】

データ Y 1 , Y 2 の評価値に加え、データ Y 1 , Y 2 へのリンクをもつデータ Y 1 1 , Y 1 2 , Y 1 3 それぞれの評価値に基づいて、データ Y の評価値を評価してもよい。例えば、データ Y 1 1 , Y 1 2 , Y 1 3 の評価値はデータ Y 1 , Y 2 を介して、データ Y の評価値に影響を及ぼす。例えば、この影響度（参照時影響度）は、データ Y の評価値に反映される他のデータの評価値の割合により表せる。例えば、データ Y 1 1 の評価値の 40% がデータ Y の評価値に反映されるならデータ Y に対するデータ Y 1 1 の参照時影響度は 40% ( = 0 . 4 ) である。参照時影響度とデータ Y 1 , Y 2 , Y 1 1 , Y 1 2 , Y 1 3 それぞれの評価値とを考慮した値に基づいて、データ Y の評価値を評価してもよい。

40

## 【 0 0 3 0 】

ここで、参照範囲 3 a は参照対象となるデータの集合である。参照範囲 3 a は、参照時の影響度閾値（ $0 < \text{閾値} < 1$  の実数）によって定めることができる。例えば、参照時影響度が 以上であるようなデータの集合が参照範囲 3 a である。

50

## 【 0 0 3 1 】

情報処理装置 1 によれば、演算部 1 b により、記憶部 1 a が参照されて、データ X の評価値が更新されると、データ X の評価値に基づいて、少なくともデータ X からリンクされるデータ X 1 , X 2 それぞれの評価値が評価されて更新される。また、演算部 1 b により、データ Y の評価値の参照要求が受け付けられると、少なくともデータ Y へのリンクをもつデータ Y 1 , Y 2 それぞれの評価値に基づいて、データ Y の評価値が評価されて出力される。

## 【 0 0 3 2 】

これにより、評価値を更新するための処理コストを軽減しながら、評価の信頼性を高めることができる。具体的には、評価値の更新時において更新範囲 2 a に含まれるデータのみを更新すると、処理コストを軽減できる反面、更新の影響の伝搬が制限されることになる。したがって、更新範囲 2 a の外部のデータについて評価値の精度が悪化し得る。このため、当該外部のデータについて、評価値の参照要求を受けた時に、記憶部 1 a から当該外部のデータの評価値を単に読み出して応答するのみでは、応答された評価値の信頼性は低いことになる。

10

## 【 0 0 3 3 】

そこで、情報処理装置 1 は、あるデータの評価値の参照時に参照範囲 3 a に含まれるデータの評価値を考慮して、参照対象のデータの評価値を求める。このように、参照対象のデータの評価値を、周囲のデータの評価値に基づいて再評価して応答する。したがって、あるデータの更新の影響が、参照対象のデータに伝搬していなかったとしても、周囲のデータに対して当該影響が伝搬していれば、当該影響を参照対象のデータに反映させて評価値を得ることができる。

20

## 【 0 0 3 4 】

より具体的には、データの全体集合 4 に、データ A , B , C , D , E が含まれているとする。データ A の評価値を更新するとき更新範囲 4 a に含まれるデータ（データ A の周囲のデータ）が更新対象となる。データ B の評価値を参照するとき参照範囲 4 b に含まれるデータ（データ B の周囲のデータ）が参照対象となる。データ A の更新時にデータ B まで更新の影響が伝搬しなかったとしても、更新範囲 4 a と参照範囲 4 b との両方に含まれるデータにより、データ B について当該更新の影響を反映させた評価値を得られる。

## 【 0 0 3 5 】

一方、データ C の評価値を参照するとき参照範囲 4 c に含まれるデータが参照対象となる。周囲のデータに何れのデータからの更新の影響も及んでいなければ、データ C の評価値を参照する際に、何れかのデータの更新の影響が及ぶことはない。この場合は、参照範囲 4 c に含まれるデータに有意な更新が発生していないと考えてよい。よって、周囲のデータの評価値に基づいてデータ C の評価値を求めても期待する信頼性は確保される。

30

## 【 0 0 3 6 】

同様に、データ D の評価値を更新するとき更新範囲 4 d に含まれるデータが更新対象となる。データ E の評価値を参照するとき参照範囲 4 e に含まれるデータが参照対象となる。データ A , D の更新時にデータ E まで更新の影響が伝搬しなかったとしても、更新範囲 4 a と参照範囲 4 e との両方に含まれるデータ、および、更新範囲 4 d と参照範囲 4 e との両方に含まれるデータにより、当該更新影響を考慮してデータ E の評価値を得られる。

40

## 【 0 0 3 7 】

このようにして、更新時の処理コストを軽減できる。そして、あるデータの評価値の更新の影響をリンク先の他のデータのうちの一部までにしかなんて伝搬させないとしても、参照する評価値の信頼性を高めることができる。

## 【 0 0 3 8 】

## [ 第 2 の実施の形態 ]

図 2 は、第 2 の実施の形態の情報処理システムを示す図である。第 2 の実施の形態の情報処理システムは、Web ページをランク付けする。第 2 の実施の形態の情報処理システムは、サーバ 1 0 0 , 1 0 0 a およびクライアント 2 0 0 , 2 0 0 a を有する。

50

## 【 0 0 3 9 】

サーバ100, 100aおよびクライアント200, 200aは、ネットワーク10を介して接続されている。ネットワーク10は、例えばインターネットやWAN (Wide Area Network) などの広域ネットワークである。ネットワーク10は、LAN (Local Area Network) でもよい。

## 【 0 0 4 0 】

サーバ100, 100aは、Webページのランク付けを行うサーバコンピュータである。サーバ100, 100aは、クライアント200, 200aから各Webページにおけるランクの更新内容(例えば、増分や減少分)を受け付ける。すると、サーバ100, 100aは、ランクが更新されたWebページからリンクされている他のWebページに当該更新の影響を伝搬させ、他のWebページの評価値を更新する。

10

## 【 0 0 4 1 】

また、サーバ100, 100aは、クライアント200, 200aから何れかのWebページのランクの参照要求を受け付ける。すると、サーバ100, 100aは、参照対象のWebページのランクを評価してクライアント200, 200aに応答する。

## 【 0 0 4 2 】

サーバ100, 100aは、Webページの全体を分担して管理している。例えば、サーバ100は、全体のうちの一部のWebページのランクなどの情報を保持する。サーバ100aは、全体のうち他の一部のWebページのランクなどの情報を保持する。3以上のサーバコンピュータでWebページの管理を分担してもよい。

20

## 【 0 0 4 3 】

例えば、サーバ100の処理により、サーバ100aが保持する情報が更新される場合は、サーバ100はサーバ100aに該当の情報を更新させる。サーバ100の処理に、サーバ100aが保持する情報を用いる場合は、サーバ100はサーバ100aから該当の情報を取得する。逆の場合も同様である。

## 【 0 0 4 4 】

クライアント200, 200aは、ユーザが操作するクライアントコンピュータである。クライアント200, 200aは、ユーザの指示に応じて、Webページのランクの更新要求または参照要求をサーバ100, 100aに送信する。

## 【 0 0 4 5 】

図3は、第2の実施の形態のグラフの例を示す図である。図3(A)は有向グラフを例示している。有向グラフは、例えばノード21, 22および矢印23で表される。ノード21, 22は、それぞれが1つのデータ(あるいはレコード)に相当するものであり、第2の実施の形態ではWebページPx, Pyである。WebページPx, Pyは、例えばHTMLで記述された文書データである。例えば、WebページPxには、ハイパーテキストでWebページPyのURLが記述されることで、WebページPyへのリンクが張られている。

30

## 【 0 0 4 6 】

矢印23は、ノード21からノード22へ向けたリンクを示している。矢印23はノード21を起点としている。このため、矢印23はノード21のフォワードリンクである。矢印23はノード22を終点としている。このため、矢印23はノード22のバックリンクである。このように、ノード21, 22および矢印23を組み合わせたグラフ構造により、各Webページのリンクの関係性を表すことができる。

40

## 【 0 0 4 7 】

図3(B)はWebページのグラフ構造を例示している。このような関係性をもつ各Webページをランク付けする方法として、ページランクが知られている。以下の説明においてi, j, nは1以上の整数である。kは0以上の整数である。ページランクの単純な総和公式ではページPiのランクr(Pi)は式(1)により表される。

## 【 0 0 4 8 】



【数 1】

$$r(P_i) = \sum_{P_j \in B_{P_i}} \frac{r(P_j)}{|P_j|} \quad (1)$$

【0049】

ここで、 $B_{P_i}$ は、Webページ  $P_i$  へのリンクをもつ Webページ ( $P_i$  のバックリンク) の集合である。 $|P_j|$  は、Webページ  $P_j$  からのフォワードリンクの個数である。全ての Web ページのランクを求める場合、 $r(P_j)$  が未知である。そこで、全ての Web ページにランクの初期値 (Web ページの総数が  $n$  個であれば  $1/n$ ) を与え、反復法により各 Web ページのランクを収束させる。具体的には、 $k+1$  回目の手続における Web ページ  $P_i$  のランク  $r_{k+1}(P_i)$  は式 (2) により表される。

10

【0050】

【数 2】

$$r_{k+1}(P_i) = \sum_{P_j \in B_{P_i}} \frac{r_k(P_j)}{|P_j|} \quad (2)$$

【0051】

ここで、 $r_0(P_i) = 1/n$  である。式 (2) は Web ページのランクを 1 つずつ順番に計算する。行列を用いて同じ計算を行うことで、 $1 \times n$  の行ベクトル  $\pi^T$  により全ての Web ページの  $k+1$  回目の手続後のランクを保持できる。添え字  $T$  は転置を示し、ベクトル  $\pi^T$  が  $n$  列の要素をもつ行ベクトルであることを示す。Web ページのグラフ構造を行単位に正規化した  $n \times n$  行列  $H$  で表す。行列  $H$  はグラフ構造を表す隣接行列である。具体的には、Web ページ  $P_i$  から Web ページ  $P_j$  へのリンクが存在するとき、 $H_{ij} = 1/|P_i|$  である。それ以外の箇所は 0 である。行列  $H$  の一例を示せば式 (3) のようになる。

20

【0052】

【数 3】

$$H = \begin{matrix} & P_1 & P_2 & P_3 & P_4 & \cdots & P_n \\ \begin{matrix} P_1 \\ P_2 \\ P_3 \\ P_4 \\ \vdots \\ P_n \end{matrix} & \begin{pmatrix} 0 & 1/2 & 1/2 & 0 & \cdots & 0 \\ 1/3 & 0 & 1/3 & 1/3 & \cdots & 0 \\ 0 & 1/2 & 0 & 1/2 & \cdots & 0 \\ 0 & 0 & 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 0 \end{pmatrix} \end{matrix} \quad (3)$$

30

【0053】

行列  $H$  の行  $i$  における非ゼロの要素は Web ページ  $P_i$  からのフォワードリンクの Web ページに対応する。列  $i$  における非ゼロの要素は Web ページ  $P_i$  に対するバックリンクの Web ページに対応する。このとき、 $k+1$  回目の手続後のベクトル  $\pi^{(k+1)T}$  は式 (4) のように表せる。

40

【0054】

【数 4】

$$\pi^{(k+1)T} = \pi^{(k)T} H \quad (4)$$

【0055】

ベクトル  $\pi^{(k+1)T}$  の  $i$  列目の要素が  $r_{k+1}(P_i)$  に対応する。ただし、式 (4) では

50

ベクトル  $\pi^T$  の収束性において不十分な点がある。例えば、フォワードリンクをもたない Web ページのランクが過大となるランクシンク (Rank Sink) などの問題がある。そこで、行列  $H$  に調整を加え、このような問題を解消した行列  $G$  を用いる (式 (4) において、行列  $H$  を行列  $G$  に置き換える)。その場合、ページランクを求める計算は、行列  $G$  の固有ベクトル問題に帰着される。行列  $G$  の固有ベクトルがベクトル  $\pi^T$  に相当するからである。この固有ベクトル問題は、式 (5) (6) のように表せる。

【0056】

【数5】

$$\pi^T = \pi^T G \quad (5)$$

10

【0057】

【数6】

$$\pi^T e = 1 \quad (6)$$

【0058】

ここで、行列  $e$  は単位行列である。式 (6) は全 Web ページのランクの合計が 1 になるように正規化するものである。式 (5) について、例えばバッチ処理でベキ乗法を用いた計算を行い、式 (6) を用いてベクトル  $\pi^T$  の要素を正規化することで、全 Web ページのランクを所望の精度で算出することができる。しかし、Web ページの数は膨大であり、当該バッチ処理には時間を要する。よって、この方法は、Web ページの更新に対してランクを逐次更新したい場合などには不向きである。

20

【0059】

そこで、第 2 の実施の形態では、ページランクの厳密な方法を用いて求まるランクの代わりに、Web ページの近似的なランクを求める。以下の説明では、当該近似的なランクも含めてランクと称している。

【0060】

図 4 は、第 2 の実施の形態のサーバのハードウェア例を示す図である。サーバ 100 は、プロセッサ 101、RAM 102、HDD (Hard Disk Drive) 103、通信部 104、画像信号処理部 105、入力信号処理部 106、ディスクドライブ 107 および機器接続部 108 を有する。各ユニットがサーバ 100 のバスに接続されている。サーバ 100 a およびクライアント 200, 200 a のハードウェア例もサーバ 100 と同様である。

30

【0061】

プロセッサ 101 は、サーバ 100 の情報処理を制御する。プロセッサ 101 は、マルチプロセッサであってもよい。プロセッサ 101 は、例えば CPU、MPU (Micro Processing Unit)、DSP (Digital Signal Processor)、ASIC (Application Specific Integrated Circuit)、FPGA (Field Programmable Gate Array) または PLD (Programmable Logic Device) などである。プロセッサ 101 は、CPU、MPU、DSP、ASIC、FPGA、PLD のうちの 2 以上の要素の組み合わせであってもよい。

【0062】

RAM 102 は、サーバ 100 の主記憶装置である。RAM 102 は、プロセッサ 101 に実行させる OS (Operating System) のプログラムやアプリケーションプログラムの少なくとも一部を一時的に記憶する。また、RAM 102 は、プロセッサ 101 による処理に用いる各種データを記憶する。

40

【0063】

HDD 103 は、サーバ 100 の補助記憶装置である。HDD 103 は、内蔵した磁気ディスクに対して、磁氣的にデータの書き込みおよび読み出しを行う。HDD 103 には、OS のプログラム、アプリケーションプログラム、および各種データが格納される。サーバ 100 は、フラッシュメモリや SSD (Solid State Drive) などの他の種類の補助記憶装置を備えてもよく、複数の補助記憶装置を備えてもよい。

【0064】

50

通信部 104 は、ネットワーク 10 を介して他のコンピュータと通信を行えるインタフェースである。通信部 104 は、有線インタフェースでもよいし、無線インタフェースでもよい。

【0065】

画像信号処理部 105 は、プロセッサ 101 からの命令に従って、サーバ 100 に接続されたディスプレイ 11 に画像を出力する。ディスプレイ 11 としては、CRT (Cathode Ray Tube) ディスプレイや液晶ディスプレイなどを用いることができる。

【0066】

入力信号処理部 106 は、サーバ 100 に接続された入力デバイス 12 から入力信号を取得し、プロセッサ 101 に出力する。入力デバイス 12 としては、例えば、マウスやタッチパネルなどのポインティングデバイス、キーボードなどを用いることができる。

10

【0067】

ディスクドライブ 107 は、レーザ光などを利用して、光ディスク 13 に記録されたプログラムやデータを読み取る駆動装置である。光ディスク 13 として、例えば、DVD (Digital Versatile Disc)、DVD-RAM、CD-ROM (Compact Disc Read Only Memory)、CD-R (Recordable) / RW (ReWritable) などを使用できる。ディスクドライブ 107 は、例えば、プロセッサ 101 からの命令に従って、光ディスク 13 から読み取ったプログラムやデータを RAM 102 または HDD 103 に格納する。

【0068】

機器接続部 108 は、サーバ 100 に周辺機器を接続するための通信インタフェースである。例えば、機器接続部 108 にはメモリ装置 14 やリーダライタ装置 15 を接続できる。メモリ装置 14 は、機器接続部 108 との通信機能を搭載した記録媒体である。リーダライタ装置 15 は、メモリカード 16 へのデータの書き込み、またはメモリカード 16 からのデータの読み出しを行う装置である。メモリカード 16 は、カード型の記録媒体である。機器接続部 108 は、例えば、プロセッサ 101 からの命令に従って、メモリ装置 14 またはメモリカード 16 から読み取ったプログラムやデータを RAM 102 または HDD 103 に格納する。

20

【0069】

図 5 は、第 2 の実施の形態のサーバのソフトウェア例を示す図である。図 5 に示す各ユニットは、例えば、プロセッサ 101 が RAM 102 に記憶されたプログラムを実行することで実現できる。図 5 の各ユニットは ASIC や FPGA などによって実現されてもよい。また、サーバ 100 a のソフトウェア例もサーバ 100 と同様である。

30

【0070】

サーバ 100 は、ページ情報記憶部 110、キュー記憶部 120、受信部 130、キュー制御部 140、ランク処理部 150 および送信部 160 を有する。

ページ情報記憶部 110 は、Web ページに関する情報を記憶する。例えば、各 Web ページのリンク・被リンク関係や、各 Web ページのランクである。

【0071】

キュー記憶部 120 は、Web ページのランクの更新要求 (以下、単に更新要求と称する)、または、Web ページのランクの参照要求 (以下、単に参照要求と称する) を、キューを用いて記憶する。

40

【0072】

ページ情報記憶部 110 およびキュー記憶部 120 は、RAM 102 や HDD 103 の記憶領域を用いて実装できる。

受信部 130 は、クライアント 200, 200 a から更新要求や参照要求を受信して、キュー制御部 140 に出力する。

【0073】

キュー制御部 140 は、受信部 130 から取得した更新要求や参照要求のジョブをキュー記憶部 120 に格納する。キュー制御部 140 は、FIFO (First In / First Out) の手順に従って、キュー記憶部 120 に記憶された更新要求または参照要求のジョブを取

50

り出して、ランク処理部 150 に出力する。キュー制御部 140 は、ランク処理部 150 によって生成された更新や参照のジョブ内容を示す情報を、キュー記憶部 120 に格納することもある。

【0074】

ランク処理部 150 は、ページ情報記憶部 110 に記憶された Web ページの情報に基づいて、ランクの評価を行う。具体的には、ランク処理部 150 は、キュー制御部 140 から取得した更新要求のジョブに基づいて、更新対象となる Web ページのランクを更新する。このとき、上述のバッチ処理とは異なり、ランク処理部 150 は更新対象となる Web ページをページ情報記憶部 110 から検索しながら、ランクの更新処理を行うことになる。ランク処理部 150 は、ランクの更新処理が完了すると、更新完了の通知を送信部 160 に出力する。

10

【0075】

また、ランク処理部 150 は、キュー制御部 140 から取得した参照要求のジョブに基づいて、参照対象の Web ページのランクを評価し、送信部 160 に出力する。ランク処理部 150 は、参照対象の Web ページに対するリンク元の Web ページのランクに基づいて、参照対象の Web ページのランクを評価する。

【0076】

ランク処理部 150 は、更新や参照の処理に応じたジョブを生成し、当該ジョブ内容を示す情報をキュー制御部 140 に出力することもある。

送信部 160 は、ランク処理部 150 から取得した更新完了の通知を、更新要求の送信元であるクライアント（例えば、クライアント 200）に回答する。送信部 160 はランク処理部 150 から取得した Web ページのランクを、参照要求の送信元であるクライアントに回答する。

20

【0077】

図 6 は、第 2 の実施の形態の Web ページ管理テーブルを示す図である。Web ページ管理テーブル 111 は、ページ情報記憶部 110 に記憶される。Web ページ管理テーブル 111 は、URL、ランク、フォワードリンクおよびバックリンクの項目を含む。

【0078】

URL の項目には、Web ページの URL が登録される。ランクの項目には、ランクの値が登録される。フォワードリンクの項目には、当該 Web ページが指し示すリンク先 Web ページの URL のリストが登録される。バックリンクの項目には、当該 Web ページに対するリンク元 Web ページの URL のリストが登録される。

30

【0079】

例えば、Web ページ管理テーブル 111 には、URL が“P1”、ランクが“0.01”、フォワードリンクが“P2, P3”、バックリンクが“P21, P22”という情報が登録されている。これは、URL “P1” で示される Web ページのランクが“0.01”であることを示している。また、URL “P1” の Web ページから URL “P2, P3” の Web ページへのフォワードリンクが存在すること、URL “P1” の Web ページに対して URL “P21, P22” の Web ページからのバックリンクが存在することを示している。

40

【0080】

ランク処理部 150 は、定期的に（例えば、月次などで）ページランクの方法を用いて全 Web ページのランクを求め、Web ページ管理テーブル 111 を更新してもよい。全 Web ページのランクを定期的に一括更新することで、第 2 の実施の形態による評価の信頼性を向上し得るからである。

【0081】

なお、以下の説明では、URL “P1” に対応する Web ページを、Web ページ P1 のように表記することがある。

図 7 は、第 2 の実施の形態のキューを示す図である。キュー 121 は、キュー記憶部 120 に格納されている。キュー 121 には次のようなフォーマットのジョブデータが格納

50

される。当該フォーマットは（処理対象のWebページのURL、処理種別、処理内容、影響度）である。

【0082】

処理対象のWebページのURLには、更新または参照対象とするWebページのURLが設定される。処理種別には、更新または参照の何れかを示す情報が設定される。処理内容には、処理種別が更新の場合に、具体的なランクの増減値が設定される。処理内容は、処理種別が参照の場合、設定なしとなる。影響度は、ランク更新の場合、更新要求を受けたWebページのランク変化によって処理対象のWebページが受ける影響の度合いを示す。また、ランク参照の場合、参照対象のWebページに対して処理対象のWebページが与える影響の度合いを示す。クライアント200、200aから受信した更新要求および参照要求のジョブをキュー121に格納する場合、影響度は常に“1”である。

10

【0083】

例えば、キュー121には、“P3，更新，増分0.01，影響度1”という情報が登録されている。これは、WebページP3について、ランクの更新要求を受信しており、更新内容が0.01だけランクを増加させるものであること、この更新がWebページP3に対して100%の影響度をもつことを示す。

【0084】

また、例えば、キュー121には、“P2，参照，影響度1”という情報が登録されている。これは、WebページP2について、ランクの参照要求を受信していること、WebページP2のランク更新はWebページP2自身に対して100%の影響度をもつことを示す。

20

【0085】

更に、例えば、キュー121には、“P2，更新，減少分0.005，影響度1”という情報が登録されている。これは、WebページP2について、ランクの更新要求を受信しており、更新内容が0.005だけランクを減少させるものであること、この更新がWebページP2に対して100%の影響度をもつことを示す。

【0086】

キュー121には、クライアント200、200aから要求が到着するたびに、キュー制御部140により当該要求に関するジョブデータが挿入される。そして、キュー制御部140によりFIFOの手順で各要求に対応するジョブデータが取り出され、ランク処理部150に出力される。ランク処理部150により、各ジョブが処理される。

30

【0087】

なお、あるWebページでランクが増える場合としては、例えば、次のような場合が考えられる。第1には、当該Webページの既存のリンク元Webページがもつフォワードリンク数が減少することで、当該リンク元Webページから受け取るランクが増加する場合である。第2には、ランクが既知である他のWebページに当該Webページへのフォワードリンクが追加されることで、当該他のWebページからランクの分配を新たに受け取るようになる場合である。

【0088】

また、あるWebページでランクが減少する場合としては、例えば、当該Webページの既存のリンク元Webページがもつフォワードリンク数が増加することで、当該リンク元Webページから受け取るランクが減少する場合が考えられる。

40

【0089】

1つのWebページでフォワードリンクの追加・削除が行われると、複数の他のWebページにおいてランクの増減に係る更新が発生し得る。

また、キュー121に格納されるデータは、要求元のクライアントを示す情報（例えば、IP（Internet Protocol）アドレスなど）を含んでもよい。要求に対する応答先を判別可能にするためである。

【0090】

図8は、第2の実施の形態の更新範囲の例を示す図である。図8ではWebページの全

50

体集合300に対して更新範囲R1が例示されている。更新範囲R1は、WebページPxのランクの更新要求を受け付けたときに、その更新の影響を伝搬させる範囲である。更新範囲R1に含まれるWebページがランクの更新対象となる。このとき、各Webページが受ける影響度（更新時影響度）は、データPxのランクの変化分 $r$ に対して各Webページのランクに反映される変化分 $r'$ の割合 $= r' / r$ により表せる（これを百分率で表してもよい）。ここで、 $r$ は $-1 < r < 0$ 、 $0 < r < 1$ の実数である。 $r'$ は $-1 < r' < 0$ 、 $0 < r' < 1$ の実数である。 $\alpha$ は $0 < \alpha < 1$ の実数である。

【0091】

更新範囲R1は、影響度閾値 $\alpha$ （ $\alpha$ は $0 < \alpha < 1$ の実数）によって定めることができる。例えば、WebページPxのランク変化に対し、 $\alpha$ である他のWebページの集合が更新範囲R1である。例えば、 $\alpha = 0.05 = 5\%$ とする。

10

【0092】

図9は、第2の実施の形態の参照範囲の例を示す図である。図9ではWebページの全体集合300に対して参照範囲R2が例示されている。参照範囲R2は、WebページPxのランクの参照要求を受け付けたときに、ランクの参照対象とするWebページの範囲である。WebページPxのランクに対して、他のWebページが及ぼす影響度（参照時影響度）は、WebページPxのランクに反映される他のWebページのランクの割合 $\alpha'$ （ $\alpha'$ は $0 < \alpha' < 1$ の実数）により表せる。参照時影響度 $\alpha'$ は、あるWebページのフォワードリンク数と、当該WebページとWebページPxとの間に介入する他のWebページのフォワードリンク数と、に基づいて算出される。

20

【0093】

参照範囲R2は、影響度閾値 $\alpha'$ （ $\alpha'$ は $0 < \alpha' < 1$ の実数）によって定めることができる。例えば、WebページPxのランクの参照要求を受けたとき、WebページPxのランクに対して $\alpha'$ である他のWebページの集合が参照範囲R2である。例えば、 $\alpha' = 0.05 = 5\%$ とする。

【0094】

図10は、第2の実施の形態のランク更新を示すフローチャートである。以下、図10に示す処理をステップ番号に沿って説明する。なお、ステップS11の直前においてキュー121に格納されているジョブデータは存在しないものとする。

30

【0095】

（ステップS11）受信部130は、クライアント200から更新要求を受信する。受信部130は、当該更新要求をキュー制御部140に出力する。更新要求は、WebページP1について、 $r$ だけランクを増加させることを示しているとする。なお、WebページP1の変化分がWebページP1に及ぼす更新時影響度は前述の通り $r / r = 1$ である。すなわち、当該更新要求に相当するジョブは（P1，更新， $r$ ，1）である。

【0096】

（ステップS12）キュー制御部140は、受信部130から取得した更新要求のジョブ（P1，更新， $r$ ，1）をキュー121に挿入する。

（ステップS13）キュー制御部140は、キュー121が空であるか否かを判定する。キュー121が空である場合、キュー制御部140はランク処理部150にその旨を通知して、処理をステップS18に進める。キュー121が空でない場合、処理をステップS14に進める。

40

【0097】

（ステップS14）キュー制御部140は、キュー121から更新のジョブ（P，更新， $r$ ，1）を取り出して、ランク処理部150に出力する。取り出されたジョブは、キュー121から削除される。

【0098】

（ステップS15）ランク処理部150は、Webページ管理テーブル111を参照して、WebページPのランクに $x \cdot r$ を加算する。

50

(ステップS16) ランク処理部150は、 $\alpha < \beta$  であるか否かを判定する。 $\alpha < \beta$  である場合、処理をステップS13に進める。 $\alpha \geq \beta$  である場合、処理をステップS17に進める。

【0099】

(ステップS17) ランク処理部150は、更新のジョブ( $P'$ , 更新,  $r$ ,  $\beta$ )を生成して、キュー制御部140に出力する。キュー制御部140は、取得したジョブをキュー121に挿入する。ここで、Webページ $P'$ は、Webページ $P$ のリンク先(フォワードリンク)のWebページである。ランク処理部150は、Webページ管理テーブル111に基づいてWebページ $P'$ を取得する。 $\beta$ は、Webページ $P$ からWebページ $P'$ に対する影響度である。具体的には、 $\beta = \frac{1}{L}$  (Webページ $P$ のフォワードリンク数)である。なお、ここでいう $L$ はステップS14で取り出されたWebページ $P$ が受ける影響度である。Webページ $P$ のリンク先のWebページが複数ある場合は、複数のジョブが生成されて、キュー121に挿入されることになる。そして、処理がステップS13に進められる。なお、Webページ $P$ からのフォワードリンクが存在しない場合は、ステップS17をスキップして、ステップS13に進む。

10

【0100】

(ステップS18) ランク処理部150は、更新完了の通知を生成して送信部160に出力する。更新完了の通知には、クライアント200を示す情報(例えば、IPアドレスなど)が含まれる。送信部160は、更新完了の通知をクライアント200に应答する。

【0101】

20

このようにして、更新時影響度が影響度閾値よりも小さくなるまでWebページのフォワードリンクを辿り、ランクを更新する。上述の手順は、グラフの幅優先探索に基づいている。その探索のステップ数は、更新範囲 $R$ に含まれるWebページの数 $N$ ( $N$ は2以上の整数)およびリンクの数 $M$ ( $M$ は1以上の整数)とすれば、 $O(N+M)$ である。すなわち、更新範囲 $R$ が小さいほど更新の処理コストは低減し、更新範囲 $R$ が大きいほど更新の処理コストは増大する。

【0102】

なお、図10ではWebページ $P$ の更新要求を受け付けた場合を例示したが、他のWebページの更新要求を受け付けた場合も同様の手順となる。

図11は、第2の実施の形態の更新時影響度の例を示す図である。図11は更新時の影響度閾値 $\alpha = 5\%$ である場合を例示している。Webページ $P$ 1~ $P$ 17は、Webページの全体集合300の一部である。

30

【0103】

Webページ $P$ 1はWebページ $P$ 2,  $P$ 3へのフォワードリンクを有する。Webページ $P$ 2はWebページ $P$ 4,  $P$ 5へのフォワードリンクを有する。Webページ $P$ 3はWebページ $P$ 5,  $P$ 6へのフォワードリンクを有する。Webページ $P$ 5はWebページ $P$ 7,  $P$ 8へのフォワードリンクを有する。Webページ $P$ 8はWebページ $P$ 9,  $P$ 10へのフォワードリンクを有する。Webページ $P$ 9はWebページ $P$ 11,  $P$ 12へのフォワードリンクを有する。Webページ $P$ 11はWebページ $P$ 13,  $P$ 14へのフォワードリンクを有する。Webページ $P$ 12はWebページ $P$ 14,  $P$ 15へのフォワードリンクを有する。Webページ $P$ 13は、Webページ $P$ 16へのフォワードリンクを有する。Webページ $P$ 14はWebページ $P$ 16,  $P$ 17へのフォワードリンクを有する。

40

【0104】

図10で示した手順によれば、Webページ $P$ 1のランク更新に対して、各Webページに及ぶ更新時影響度は次のようになる。Webページ $P$ 1に対する影響度は100%である。

【0105】

Webページ $P$ 2,  $P$ 3に対する影響度は50%である。Webページ $P$ 1がWebページ $P$ 2,  $P$ 3への2つのフォワードリンクを有するからである。すなわち、 $100\% \div$

50

2 = 50%である。

【0106】

WebページP4, P5, P6に対する影響度は25%である。WebページP2, P3のそれぞれが2つのフォワードリンクを有するからである。すなわち、 $50\% \div 2 = 25\%$ である。

【0107】

WebページP7, P8に対する影響度は12.5%である。WebページP5がWebページP7, P8への2つのフォワードリンクを有するからである。すなわち、 $25\% \div 2 = 12.5\%$ である。

【0108】

WebページP9, P10に対する影響度は6.25%である。WebページP8がWebページP9, P10への2つのフォワードリンクを有するからである。すなわち、 $12\% \div 2 = 6.25\%$ である。

【0109】

WebページP11, P12に対する影響度は3.12%である。WebページP9がWebページP11, P12への2つのフォワードリンクを有するからである。すなわち、 $6.25\% \div 2 = 3.125\%$ である(図11では小数第2位までを表記)。

【0110】

WebページP13, P14, P15に対する影響度は1.56%である。WebページP11, P12のそれぞれが2つのフォワードリンクを有するからである。すなわち、 $3.125\% \div 2 = 1.5625\%$ である(図11では小数第2位までを表記)。

【0111】

WebページP16に対する影響度は1.56%である。WebページP13がWebページP16への1つのフォワードリンクを有するからである。すなわち、 $1.56\% \div 1 = 1.56\%$ である。ただし、WebページP16は、WebページP14からの影響度0.78%もある(WebページP14は2つのフォワードリンクをもつため)。この場合、WebページP16に対する影響度を $1.56\% + 0.78\% = 2.34\%$ と評価してもよい。

【0112】

WebページP17に対する影響度は0.78%である。WebページP14がWebページP16, P17への2つのフォワードリンクを有するからである。すなわち、 $1.5625\% \div 2 = 0.78125\%$ である(図11では小数第2位までを表記)。

【0113】

この場合、更新時影響度が よりも初めて小さくなるのは、WebページP1からフォワードリンクを辿っていくと、WebページP11, P12である。すなわち、WebページP1~P12が更新範囲R1に含まれるWebページとなる。よって、ランク処理部150は、WebページP1からフォワードリンクを辿り、WebページP2~P12のランクを更新する。一方、WebページP13~P17のランクは更新しない。

【0114】

図12は、第2の実施の形態のランク参照を示すフローチャートである。以下、図12に示す処理をステップ番号に沿って説明する。なお、ステップS21の直前においてキュー121に格納されているジョブデータは存在しないものとする。

【0115】

(ステップS21)受信部130は、クライアント200から参照要求を受信する。受信部130は、当該参照要求をキュー制御部140に出力する。当該参照要求では、WebページP16が指定されているものとする。なお、WebページP16のランクがWebページP16のランクに及ぼす参照時影響度を前述の通り“1”としている。すなわち、当該参照要求に相当するジョブは(P16, 参照, 1)である。

【0116】

(ステップS22)キュー制御部140は、WebページP16の参照要求を受け付け

10

20

30

40

50



たことをランク処理部 150 に通知する。ランク処理部 150 は、Web ページ P 16 のランク a を  $a = 0$  に初期化する。

【0117】

(ステップ S 23) キュー制御部 140 は、当該参照要求のジョブ (P 16, 参照, 1) をキュー 121 に挿入する。

(ステップ S 24) キュー制御部 140 は、キュー 121 が空であるか否かを判定する。キュー 121 が空である場合、キュー制御部 140 はランク処理部 150 にその旨を通知して、処理をステップ S 29 に進める。キュー 121 が空でない場合、処理をステップ S 25 に進める。

【0118】

(ステップ S 25) キュー制御部 140 は、キュー 121 から参照のジョブ (P, 参照, ) を取り出して、ランク処理部 150 に出力する。取り出されたジョブは、キュー 121 から削除される。

【0119】

(ステップ S 26) ランク処理部 150 は、Web ページ管理テーブル 111 を参照して、Web ページ P のランクを取得する。ただし、Web ページ P が Web ページ P 16 にする場合、当該ランク値は 0 である。ランク処理部 150 は、 $x$  (Web ページ P のランク) を Web ページ P 16 のランク a に加算する。

【0120】

(ステップ S 27) ランク処理部 150 は、 $<$  ' であるか否かを判定する。 $<$  ' である場合、処理をステップ S 24 に進める。 $>$  ' である場合、処理をステップ S 28 に進める。

【0121】

(ステップ S 28) ランク処理部 150 は、参照のジョブ (P', 参照, ' ) を生成して、キュー制御部 140 に出力する。キュー制御部 140 は、取得したジョブをキュー 121 に挿入する。ここで、Web ページ P' は、Web ページ P のリンク元 (バックリンク) の Web ページである。ランク処理部 150 は、Web ページ管理テーブル 111 に基づいて Web ページ P' を取得する。' は、Web ページ P' から Web ページ P 16 に対する影響度である。' は、' = / (Web ページ P' のフォワードリンク数) により得られる値である。は補正係数であり、 $0 < < 1$  の実数である。なお、ここでいう はステップ S 25 で取り出された Web ページ P からの影響度 である。Web ページ P へのリンク元の Web ページが複数ある場合は、複数のジョブが生成されて、キュー 121 に挿入されることになる。そして、処理がステップ S 24 に進められる。なお、Web ページ P へのバックリンクが存在しない場合は、ステップ S 28 をスキップして、ステップ S 24 に進む。

【0122】

(ステップ S 29) ランク処理部 150 は、Web ページ P 16 のランク a の通知を送信部 160 に出力する。当該通知には、クライアント 200 を示す情報 (例えば、IP アドレスなど) が含まれる。送信部 160 は、Web ページ P 16 のランク a をクライアント 200 に応答する。

【0123】

このようにして、Web ページ P 16 への参照時影響度が ' よりも小さくなるまで、バックリンクを辿り、各 Web ページのランクに参照時影響度を乗じた値を、Web ページ P 16 のランクに加算していく。なお、上述の手順はグラフの幅優先探索に基づいている。その探索のステップ数は、参照範囲 R 2 に含まれる Web ページの数を N (N は 2 以上の整数) およびリンクの数を M (M は 1 以上の整数) とすれば、 $O(N + M)$  である。すなわち、参照範囲 R 2 が小さいほど参照の処理コストは低減し、参照範囲 R 2 が大きいほど参照の処理コストは増大する。

【0124】

なお、図 12 では Web ページ P 16 の参照要求を受け付けた場合を例示したが、他の

10

20

30

40

50

Web ページの参照要求を受け付けた場合も同様の手順となる。

図 1 3 は、第 2 の実施の形態の参照時影響度の例を示す図である。図 1 3 は参照時の影響度閾値  $\theta = 5\%$ 、 $\alpha = 80\%$  である場合を例示している。Web ページ P 1 ~ P 1 7 のグラフ構造は、図 1 1 と同様である。

【 0 1 2 5 】

図 1 2 で示した手順によれば、Web ページ P 1 6 のランク参照に対して、各 Web ページから Web ページ P 1 6 に対する参照時影響度は次のようになる。Web ページ P 1 6 からの影響度は 100% である。

【 0 1 2 6 】

Web ページ P 1 3 からの影響度は 80% である。Web ページ P 1 3 は Web ページ P 1 6 への 1 つのフォワードリンクを有するからである。すなわち、 $100\% \div 1 \times 0.8 = 80\%$  である。

10

【 0 1 2 7 】

Web ページ P 1 4 からの影響度は 40% である。Web ページ P 1 4 は Web ページ P 1 6、P 1 7 への 2 つのフォワードリンクを有するからである。すなわち、 $100\% \div 2 \times 0.8 = 40\%$  である。

【 0 1 2 8 】

Web ページ P 1 1、P 1 2 からの影響度は 16% である。Web ページ P 1 1、P 1 2 のそれぞれは 2 つのフォワードリンクを有するからである。すなわち、 $40\% \div 2 \times 0.8 = 16\%$  である。

20

【 0 1 2 9 】

Web ページ P 9 からの影響度は 6.4% である。Web ページ P 9 は Web ページ P 1 1、P 1 2 への 2 つのフォワードリンクを有するからである。すなわち、 $16\% \div 2 \times 0.8 = 6.4\%$  である。

【 0 1 3 0 】

Web ページ P 8 からの影響度は 2.56% である。Web ページ P 8 は Web ページ P 9、P 1 0 への 2 つのフォワードリンクを有するからである。すなわち、 $6.4\% \div 2 \times 0.8 = 2.56\%$  である。

【 0 1 3 1 】

Web ページ P 5 からの影響度は 1.02% である。Web ページ P 5 は Web ページ P 7、P 8 への 2 つのフォワードリンクを有するからである。すなわち、 $2.56\% \div 2 \times 0.8 = 1.024\%$  である (図 1 3 では小数第 2 位までを表記)。

30

【 0 1 3 2 】

Web ページ P 2、P 3 からの影響度は 0.40% である。Web ページ P 2、P 3 のそれぞれは 2 つのフォワードリンクを有するからである。すなわち、 $1.024\% \div 2 \times 0.8 = 0.4096\%$  である (図 1 3 では小数第 2 位までを表記)。

【 0 1 3 3 】

Web ページ P 1 からの影響度は 0.16% である。Web ページ P 1 は Web ページ P 2、P 3 への 2 つのフォワードリンクを有するからである。すなわち、 $0.4096\% \div 2 \times 0.8 = 0.16384\%$  である (図 1 3 では小数第 2 位までを表記)。

40

【 0 1 3 4 】

上記以外の Web ページ P 4、P 6、P 7、P 1 0、P 1 5、P 1 7 は、Web ページ P 1 6 に繋がるフォワードリンクを有していない。このため、Web ページ P 4、P 6、P 7、P 1 0、P 1 5、P 1 7 から Web ページ P 1 6 に対する影響度は 0% である。

【 0 1 3 5 】

この場合、参照時影響度が  $\theta$  よりも初めて小さくなるのは、Web ページ P 1 6 からバックリンクを辿っていくと、Web ページ P 8 である。すなわち、Web ページ P 8、P 9、P 1 1、P 1 2、P 1 4、P 1 6 が参照範囲 R 2 に含まれる Web ページとなる。よって、ランク処理部 150 は、Web ページ P 1 6 からバックリンクを辿り、Web ページ P 1 4、P 1 3、P 1 2、P 1 1、P 9、P 8 のランクに基づいて、Web ページ P

50

16のランクを計算する。

【0136】

これにより、ランクを更新するための処理コストを軽減しながら、ランク評価の信頼性を高めることができる。具体的には、ランクの更新時において更新範囲R1に含まれるWebページのランクのみを更新することで、処理コストを軽減できる。ただし、更新の影響の伝搬が制限されることになるので、更新範囲R1の外部のWebページ(図11の例でいえばWebページP13~P17)についてランクの精度が悪化し得る。このため、当該Webページについて、ランクの参照要求を受けた時に、当該Webページのランクを単に抽出して応答するのみでは、応答されたランクの信頼性は低いことになる。

【0137】

そこで、サーバ100, 100aは、あるWebページのランクの参照時に参照範囲R2に含まれるWebページのランクを考慮して、参照対象のWebページのランクを求める。例えば、上記のようにWebページP16のランクを参照する際に、WebページP16のランクに影響を及ぼす所定範囲のWebページのランクに基づいて、当該WebページP16のランクを評価して応答する。このため、例えばWebページP1の更新の影響がWebページP16に伝搬していなかったとしても、関連するWebページに対して当該影響が伝搬していれば、当該影響を参照対象のWebページP16に反映させてランクを得ることができる。すなわち、その時点においてWebページP16につき、より信頼性の高いランクを得られることになる。このとき、参照するランクの誤差(式(5), (6)に基づいて厳密に算出されるランクに対する誤差)は、影響度閾値 $\theta$ を用いて $\theta$ 程度となることが期待される。ランクの分配を途中で打ち切る場合の誤差が更新時の影響度閾値 $\theta$ 程度とすると、 $0 < \theta < 1$ であるから、 $\theta < 1$ である。すなわち、ランクの分配を途中で打ち切る場合に、あるWebページのランクを単に読み出す方法に比べて誤差を改善し得る。よって、ランク評価の信頼性を向上し得る。

【0138】

上記の方法は、例えばあるWebページの更新に対し、より短時間で更新の影響を他のWebページに反映させたい場合に有用である。Webページの更新の頻度が比較的高ければ、更新時の影響度閾値 $\theta$ を参照時の影響度閾値 $\theta'$ よりも大きくしておく(更新範囲R1を狭める)。すると、更新処理の処理コストを軽減でき、頻繁に発生するWebページの更新に追従して、各Webページのランクをリアルタイムに更新し得る。その場合、例えば、 $\theta'$ を $\theta$ よりも小さくしておく(参照範囲R2を広げる)。更新時の影響度閾値と参照時の影響度閾値との積 $\theta\theta'$ が一定になるように、 $\theta'$ を変化させれば、変化の前後で、参照するランクについて期待される誤差をおおよそ一定に保つことができる。なお、例えば参照の頻度が比較的多ければ、逆に $\theta'$ を大きくし、 $\theta$ を小さくすることも考えられる。

【0139】

[第3の実施の形態]

以下、第3の実施の形態を説明する。前述の第2の実施の形態との相違点を主に説明し、共通する事項の説明を省略する。

【0140】

第3の実施の形態では、更新要求の頻度および参照要求の頻度に応じて、影響度閾値 $\theta$ ,  $\theta'$ を動的に変化させる機能を提供する。ここで、第3の実施の形態の情報処理システムは、図2で説明した第2の実施の形態の情報処理システムと同様である。また、第3の実施の形態のサーバのハードウェア例およびソフトウェア例は、図4, 5で説明した第2の実施の形態のサーバ100のハードウェア例およびソフトウェア例と同様である。このため、第3の実施の形態の各装置を第2の実施の形態と同一の名称・符号を付して示す。

【0141】

ただし、ページ情報記憶部110は、ログテーブルを更に記憶する点が第2の実施の形態と異なる。ログテーブルは、更新範囲R1の外部との境界に位置する更新範囲R1内のWebページに対するランクの更新内容を記録したものである。

## 【 0 1 4 2 】

また、ランク処理部 1 5 0 は、過去の所定期間（例えば 1 日、1 週間など）における更新要求の頻度および参照要求の頻度に応じて、影響度閾値  $\theta$  を変化させる点が第 2 の実施の形態と異なる。

## 【 0 1 4 3 】

図 1 4 は、第 3 の実施の形態のログテーブルの例を示す図である。ログテーブル 1 1 2 は、ページ情報記憶部 1 1 0 に記憶される。ログテーブル 1 1 2 は、開始ノード、終了ノードおよびランク増分の項目を含む。

## 【 0 1 4 4 】

開始ノードの項目には、更新要求で指定された Web ページの URL が登録される。終了ノードの項目には、更新処理の終端となった Web ページの URL が登録される。更新処理の終端の Web ページとは、更新範囲 R 1 の外部との境界に位置する更新範囲 R 1 内の Web ページである。ランク増分の項目には、ランクの増分値が登録される。ランクの増分値は負の値でもよい（その場合はランクの減少分を示すことになる）。

## 【 0 1 4 5 】

例えば、ログテーブル 1 1 2 には、開始ノードが “ P 1 ”、終了ノードが “ P 1 1 ”、ランク増分が “ 0 . 0 0 2 ” という情報が登録されている。これは、Web ページ P 1 のランクの更新要求を受信し、その際に更新処理の終端となった Web ページが Web ページ P 1 1 であったことを示す。また、このときの Web ページ P 1 1 のランクの増分値が “ 0 . 0 0 2 ” であったことを示す。

## 【 0 1 4 6 】

ある開始ノードに対して複数の終了ノードが登録されることもある。ログテーブル 1 1 2 の例では開始ノードである Web ページ P 1 に対して、終了ノードである 2 つの Web ページ P 1 1 , P 1 2 が登録されている。

## 【 0 1 4 7 】

なお、ランク処理部 1 5 0 は、図 1 0 のステップ S 1 8 において、更新処理の結果に基づいてログテーブル 1 1 2 にレコードを登録する。図 1 1 の例でいえば、Web ページ P 1（開始ノード）に対して、更新範囲 R 1 に含まれる Web ページ P 1 1 , P 1 2（終了ノード）の情報を、ログテーブル 1 1 2 に登録することになる。

## 【 0 1 4 8 】

図 1 5 は、第 3 の実施の形態の更新・参照範囲の変更を示すフローチャートである。以下、図 1 5 に示す処理をステップ番号に沿って説明する。

（ステップ S 3 1）ランク処理部 1 5 0 は、クライアント 2 0 0 , 2 0 0 a から過去の所定期間内（例えば、1 日）に受け付けた更新要求の頻度および参照要求の頻度の比率を確認する。ランク処理部 1 5 0 は、参照要求の頻度の比率が、デフォルトの比率よりも増大しているか否かを判定する。デフォルトの比率よりも増大している場合、処理をステップ S 3 2 に進める。デフォルトの比率よりも増大していない場合、処理をステップ S 3 5 に進める。例えば、更新要求の頻度：参照要求の頻度のデフォルトの比率を、1 : 2 などのようにランク処理部 1 5 0 に予め与えておく。デフォルトの比率に応じて影響度閾値  $\theta$  についてもデフォルトの値が定められる。例えば、上記比率 1 : 2 に対して、 $\theta = 5 \%$  ,  $\theta' = 1 0 \%$  とする。

## 【 0 1 4 9 】

（ステップ S 3 2）ランク処理部 1 5 0 は、更新時の影響度閾値  $\theta$  を現在値よりも小さくする。例えば、ステップ S 3 1 で例示したデフォルトの比率に対して、更新要求の頻度：参照要求の頻度が 1 : 4（参照要求の頻度が 2 倍）になっているとする。その場合、例えば、 $\theta$  を 2 分の 1 にする。後述のように  $\theta'$  を 2 倍にして参照の処理コストを軽減し、かつ、誤差のオーダを維持するためである。すなわち、 $\theta = 5 \% \div 2 = 2 . 5 \%$  とする。これは、更新範囲 R 1 を広げることに相当する。

## 【 0 1 5 0 】

（ステップ S 3 3）ランク処理部 1 5 0 は、ログテーブル 1 1 2 に基づいて、拡大した

10

20

30

40

50

後の更新範囲と拡大前の更新範囲との差分の範囲に含まれる各Webページのランクを更新する。詳細は後述する。

【0151】

(ステップS34)ランク処理部150は、参照時の影響度閾値  $\theta$  を現在値よりも大きくする。ステップS32での例示を継続して説明すれば、 $\theta$  を2倍にする。すなわち、 $\theta = 10\% \times 2 = 20\%$ とする。これは、参照範囲R2を狭めることに相当する。そして、処理を終了する。

【0152】

(ステップS35)ランク処理部150は、更新要求の頻度の比率が、デフォルトの比率よりも増大しているか否かを判定する。デフォルトの比率よりも増大している場合、処理をステップS36に進める。デフォルトの比率よりも増大していない場合、処理を終了する。

10

【0153】

(ステップS36)ランク処理部150は、更新時の影響度閾値  $\theta$  を現在値よりも大きくする。例えば、ステップS31で例示したデフォルトの比率に対して、更新要求の頻度：参照要求の頻度が2：2(更新要求の頻度が2倍)になっているとする。その場合、例えば、 $\theta$  を2倍にする。すなわち、 $\theta = 5\% \times 2 = 10\%$ とする。これは、更新範囲R1を狭めることに相当する。更新の処理コストを軽減するためである。

【0154】

(ステップS37)ランク処理部150は、参照時の影響度閾値  $\theta$  を現在値よりも小さくする。ステップS36での例示を継続して説明すれば、 $\theta$  を2分の1にする。誤差のオーダを維持するためである。すなわち、 $\theta = 10\% \div 2 = 5\%$ とする。これは、参照範囲R2を広げることに相当する。そして、処理を終了する。

20

【0155】

このようにして、ランク処理部150は定期的に影響度閾値  $\theta$  ,  $\theta$  を調整する。このようにすれば、更新要求の頻度および参照要求の頻度に応じて、更新範囲R1および参照範囲R2を変更でき、要求の受信状況に応じて効率的に処理コストを削減できるようになる。特に、例示したように更新時の影響度閾値と参照時の影響度閾値との積  $\theta$  が一定になるように調整すれば、 $\theta$  ,  $\theta$  の変更前後においてランクの評価精度を一定に保つことができる。

30

【0156】

なお、図15の手順において、ステップS33の処理をスキップしてステップS34を先に実行しておき、ステップS33の処理をランクの更新・参照の通常の処理と並行して行ってもよい。

【0157】

図16は、第3の実施の形態の更新・参照範囲の変更例(その1)を示す図である。図16では更新範囲を広げて、参照範囲を狭める場合を例示する。図15のステップS32~S34に相当する処理である。

【0158】

図16(A)は、影響度閾値の変更前の状態の更新範囲R10および参照範囲R20を例示している。このとき、例えば、更新範囲R10に対応する更新時の影響度閾値  $\theta = 5\%$ である。また、例えば、参照範囲R20に対応する参照時の影響度閾値  $\theta = 10\%$ である。

40

【0159】

図16(B)は、図16(A)よりも  $\theta$  を小さくした状態を例示している。例えば、 $\theta = 2.5\%$ である。これにより、更新範囲R10は更新範囲R11に広がる。すると、更新範囲R10, R11について差分領域  $R$ が生ずることになる。そこで、ランク処理部150は、差分領域  $R$ に含まれるWebページにつき、ログテーブル112に基づいてランクの更新を行う。ランクの参照時の評価精度を高めるためである。

【0160】

50

具体的には、ログテーブル 112 に記録された開始ノードを起点とした変更の影響を、終了ノードのフォワードリンク先まで伝搬させる。伝搬の範囲（更新範囲 R11）は開始ノードを起点としたフォワードリンクのツリーと変更後の影響度閾値とにより定まる。このとき、ログテーブル 112 には、終了ノードのランク増分が登録されている。よって、終了ノードのフォワードリンク先を起点に終了ノードのランク増分の影響を伝搬させればよい（更新範囲 R10 内の Web ページのランクは改めて計算しなくてよい）。

【0161】

図 16 (C) は、図 16 (B) よりも  $\theta$  を大きくした状態を例示している。例えば、 $\theta = 20\%$  である。これにより、参照範囲 R20 は参照範囲 R21 に狭まる。その結果、参照時の処理コストを軽減できる。その代わりに、更新範囲 R10 を更新範囲 R11 のように拡張しているため、ランクの参照精度を、図 16 (A) の状態と同程度に維持することができる。

10

【0162】

図 17 は、第 3 の実施の形態の更新・参照範囲の変更例（その 2）を示す図である。図 17 では更新範囲を狭めて、参照範囲を広げる場合を例示する。図 15 のステップ S36、S37 に相当する処理である。

【0163】

図 17 (A) は、影響度閾値の変更前の状態の更新範囲 R10a および参照範囲 R20a を例示している。このとき、例えば、更新範囲 R10a に対応する更新時の影響度閾値  $\theta = 5\%$  である。また、例えば、参照範囲 R20a に対応する参照時の影響度閾値  $\theta = 10\%$  である。

20

【0164】

図 17 (B) は、図 17 (A) よりも  $\theta$  を大きくした状態を例示している。例えば、 $\theta = 10\%$  である。これにより、更新範囲 R10a は更新範囲 R11a に狭まる。

図 17 (C) は、図 17 (B) よりも  $\theta$  を小さくした状態を例示している。例えば、 $\theta = 5\%$  である。これにより、参照範囲 R20a は参照範囲 R21a に広がる。

【0165】

このように、更新範囲 R10a を更新範囲 R11a に狭めることで、更新時の処理コストを軽減できる。その代わりに、参照範囲 R20a を参照範囲 R21a のように拡張しているため、ランクの参照精度を、図 17 (A) の状態と同程度に維持することができる。

30

【0166】

以上のように、第 3 の実施の形態では更新要求の頻度および参照要求の頻度に応じて、Web ページのランクの更新範囲および参照範囲を変更するようにした。具体的には、更新要求の頻度がデフォルトよりも高まれば、更新範囲を狭める（縮小する）。これにより、更新要求の頻度が高い場合に、更新処理の処理コストを軽減することができる。一方、参照要求の頻度がデフォルトよりも高まれば、参照範囲を狭める（縮小する）。これにより、参照要求の処理コストを軽減して、より早くランクの評価結果を応答することができる。

【0167】

このとき、更新範囲を縮小したならば参照範囲を拡大し、また、参照範囲を縮小したならば更新範囲を拡大するように制御する。これによって、参照時に得られるランクの信頼性が低下するのを抑制できる。

40

【0168】

ここで、第 2、第 3 の実施の形態では、Web ページをランク付けする場合を例示したが、他の場合にも利用できる。例えば、論文を評価する際に利用することが考えられる。論文は、他の論文を引用することが多い。このため、第 1 の論文（例えば、論文の文書データ）を有向グラフのノード 21 に、第 2 の論文を有向グラフのノード 22 に、第 1 の論文における第 2 の論文の引用を有向グラフの矢印 23（リンク）に対応付けることができる。これによって、論文ごとの重要度を示す評価値を評価することが考えられる。同様に、特許文献の評価に利用することも考えられる。特許文献も先行技術文献を引用する

50

からである。

【0169】

更に、文書以外を評価する場合にも利用できる。例えば、オンラインの通信販売などでお薦めの商品を提示するような場合である。例えば、ユーザが購入した第1, 第2の商品を示す情報を有向グラフのノード21, 22に、第1の商品が購入された場合、同じユーザにより第2の商品も購入されるという関連を矢印23(リンク)に対応付けることができる。すなわち、第1の商品を示す情報に、第2の商品を示す情報を指し示すリンクを含めたデータを作成し得る。このデータを用いて、商品ごとのお薦め度を示す評価値を評価することが考えられる。

【0170】

なお、前述のように、第1の実施の形態の情報処理は、情報処理装置1にプログラムを実行させることで実現できる。また、第2の実施の形態の情報処理は、サーバ100, 100aにプログラムを実行させることで実現できる。プログラムは、コンピュータ読み取り可能な記録媒体(例えば、光ディスク13、メモリ装置14およびメモリカード16など)に記録できる。

【0171】

プログラムを流通させる場合、例えば、当該プログラムを記録した可搬記録媒体が提供される。また、プログラムを他のコンピュータの記憶装置に格納しておき、ネットワーク経由でプログラムを配布することもできる。コンピュータは、例えば、可搬記録媒体に記録されたプログラムまたは他のコンピュータから受信したプログラムを、記憶装置に格納し、当該記憶装置からプログラムを読み込んで実行する。ただし、可搬記録媒体から読み込んだプログラムを直接実行してもよく、他のコンピュータからネットワークを介して受信したプログラムを直接実行してもよい。

【0172】

また、上記の情報処理の少なくとも一部を、DSP、ASIC、PLDなどの電子回路で実現することもできる。

【符号の説明】

【0173】

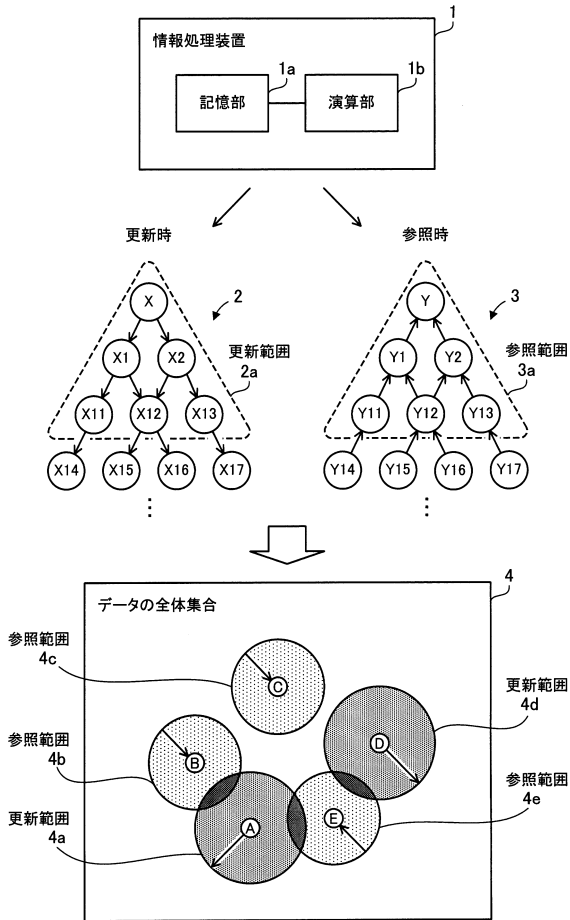
- 1 情報処理装置
- 1 a 記憶部
- 1 b 演算部
- 2, 3 グラフ構造
- 2 a, 4 a, 4 d 更新範囲
- 3 a, 4 b, 4 c, 4 e 参照範囲
- 4 データの全体集合

10

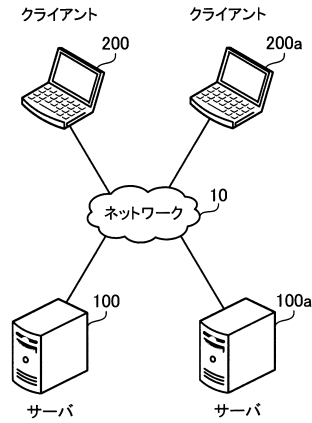
20

30

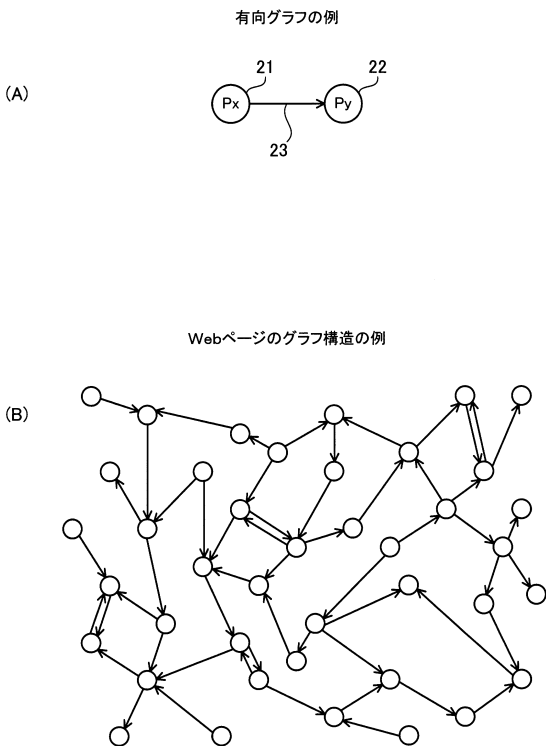
【図1】



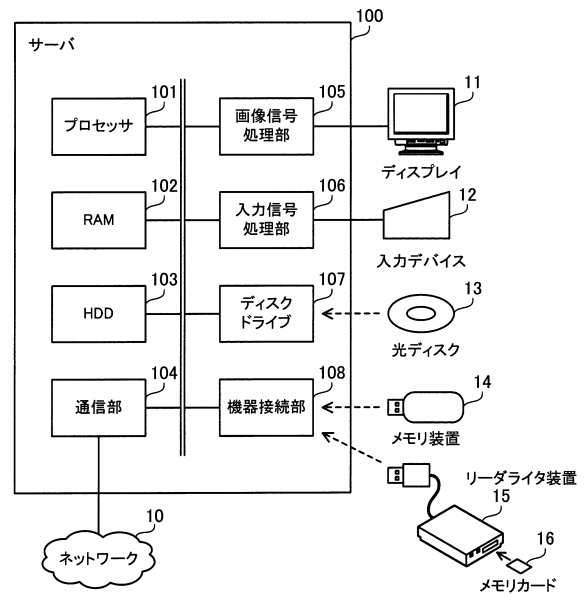
【図2】



【図3】

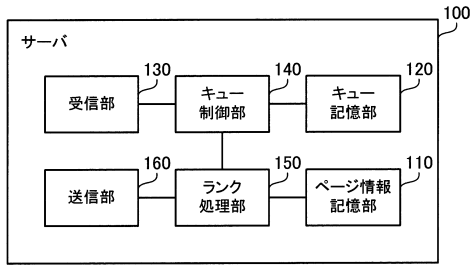


【図4】





【図5】

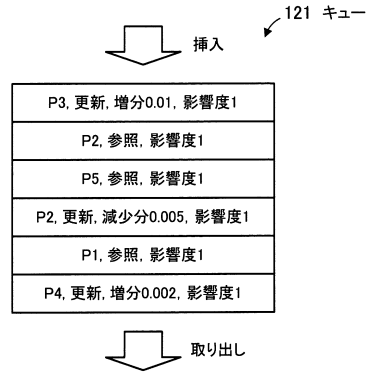


【図6】

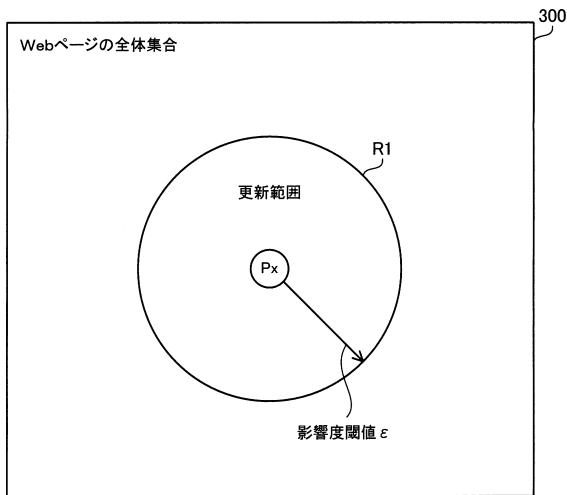
Webページ管理テーブル			
URL	ランク	フォワードリンク	バックリンク
P1	0.01	P2, P3	P21, P22
P2	0.02	P4, P5	P1, P31, P32
P3	0.01	P5, P6	P1, P41, P42
...	...	...	...

111

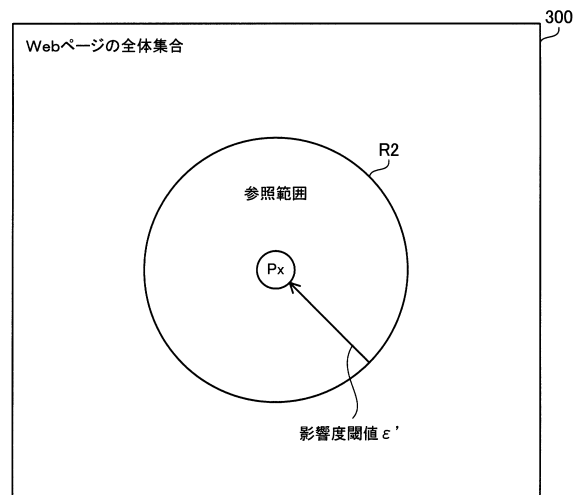
【図7】



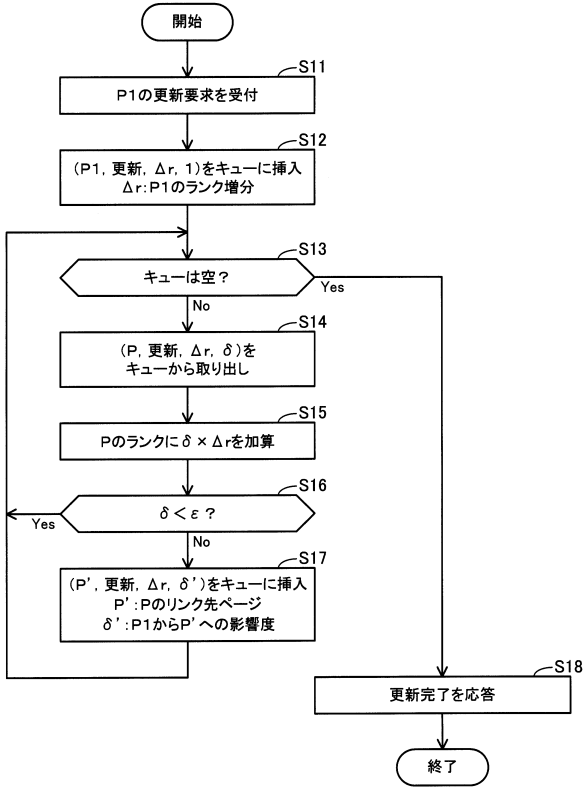
【図8】



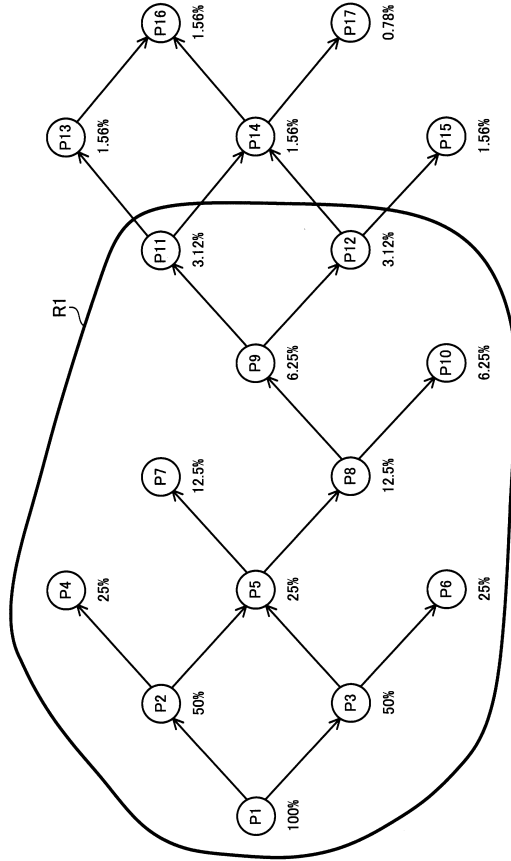
【図9】



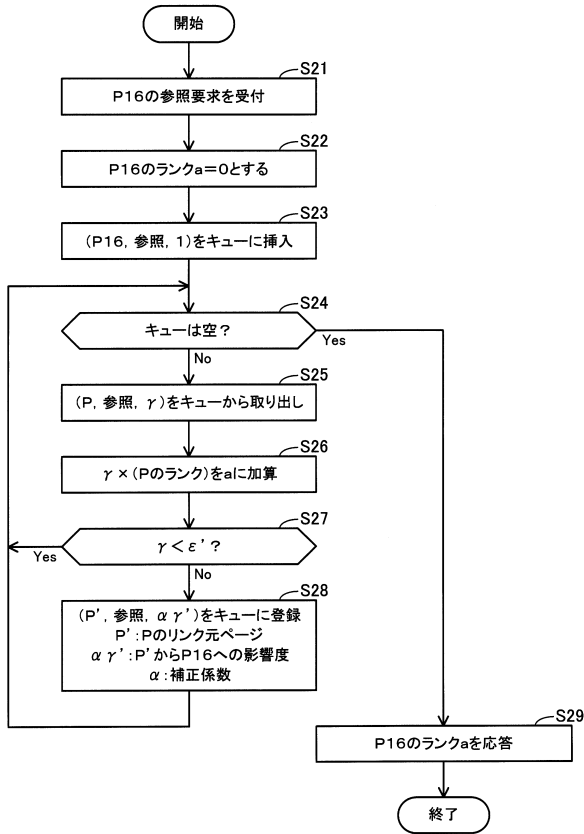
【図10】



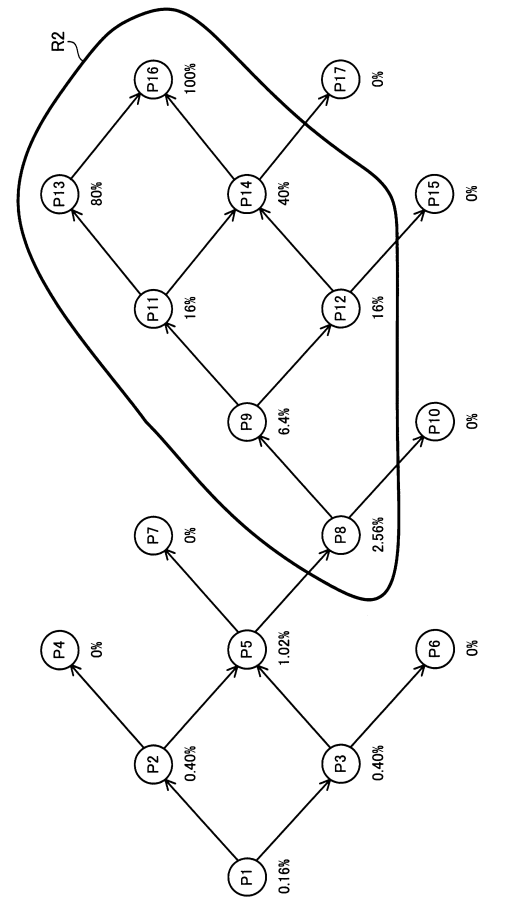
【図11】



【図12】



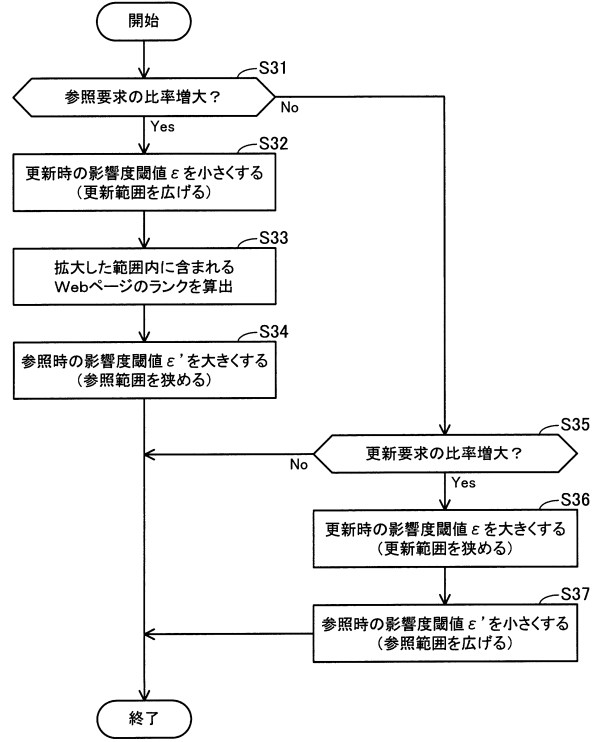
【図13】



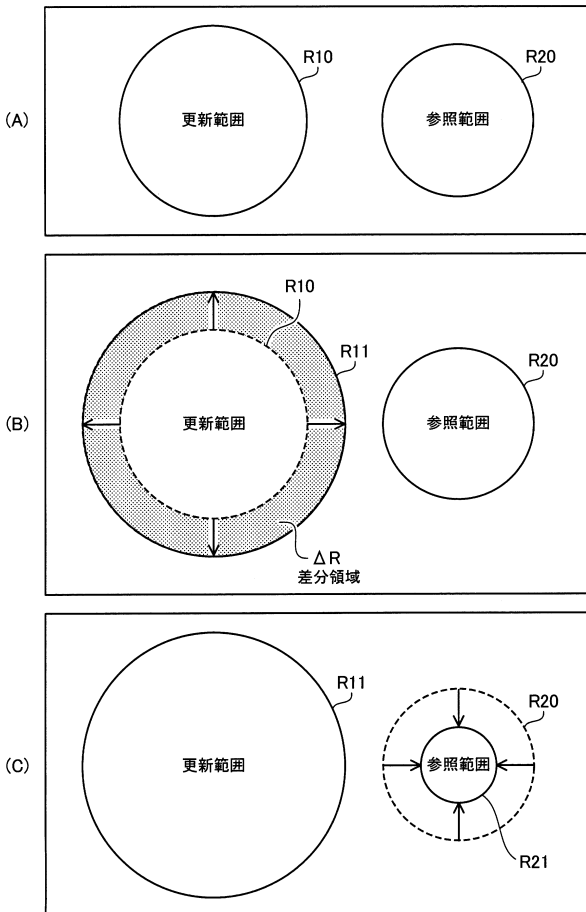
【図14】

ログテーブル		
開始ノード	終了ノード	ランク増分
P1	P11	0.002
P1	P12	0.001
...	...	...

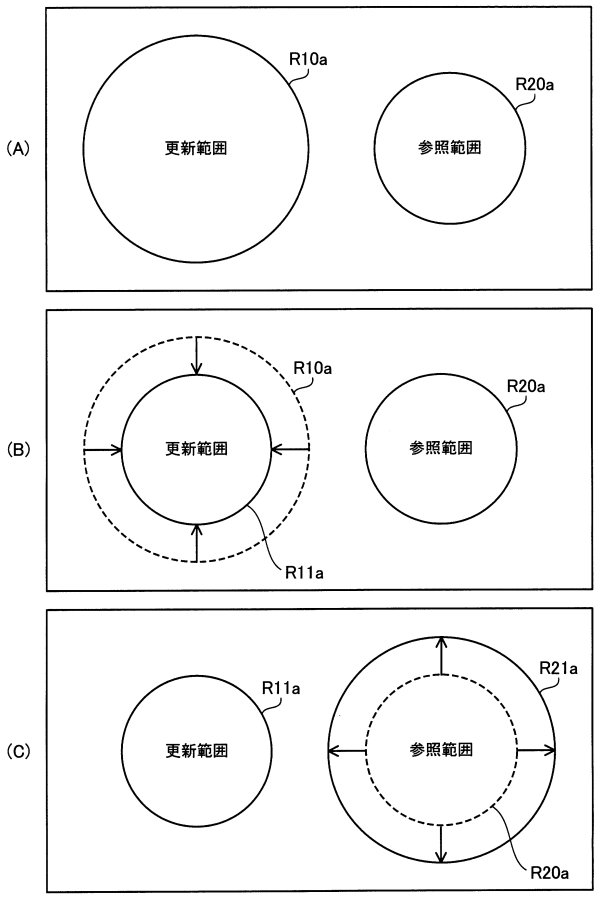
【図15】



【図16】



【図17】



---

フロントページの続き

- (56)参考文献 特開2008-243050(JP,A)  
特表2007-511815(JP,A)  
特開2007-272888(JP,A)  
山田 雅信, インクリメンタルPageRankによる重要Webページの効率的な収集戦略,  
情報処理学会論文誌, 日本, 社団法人情報処理学会, 2004年10月15日, 第45巻 No  
.SIG11(ACS7), 465-473ページ

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