

(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2010-92216

(P2010-92216A)

(43) 公開日 平成22年4月22日(2010.4.22)

(51) Int.Cl. F 1 テーマコード (参考)
G 0 6 F 1 7 / 3 0 (2006.01) G 0 6 F 1 7 / 3 0 3 8 0 E 5 B 0 7 5
 G 0 6 F 1 7 / 3 0 3 4 0 B

審査請求 未請求 請求項の数 7 O L (全 18 頁)

(21) 出願番号	特願2008-260695 (P2008-260695)	(71) 出願人	399035766 エヌ・ティ・ティ・コミュニケーションズ株式会社 東京都千代田区内幸町一丁目1番6号
(22) 出願日	平成20年10月7日(2008.10.7)	(74) 代理人	100121706 弁理士 中尾 直樹
		(74) 代理人	100128705 弁理士 中村 幸雄
		(74) 代理人	100066153 弁理士 草野 卓
		(72) 発明者	阿部 望 東京都千代田区内幸町一丁目1番6号 エヌ・ティ・ティ・コミュニケーションズ株式会社内

最終頁に続く

(54) 【発明の名称】 ドキュメントスコア算出装置、方法及びプログラム

(57) 【要約】

【課題】ユーザの属性を考慮してドキュメントにスコアを付与する。

【解決手段】リーダーランク計算部20が、アクセスログを用いて、各属性値を有するユーザの集合の各ドキュメントについてのリーダーランクを計算する。スコアを付与する際に考慮の対象となるユーザの1以上の属性のそれぞれから選択された属性値及びドキュメント情報が総合ランク計算部30に入力される。総合ランク計算部30は、それらの入力された各属性値を有するユーザの集合のそのドキュメント情報で特定されるドキュメントについての上記計算されたリーダーランクを統合して、それらの入力された属性値で特定されるセグメントに属するユーザの集合のそのドキュメントについての総合ランクを計算する。

【選択図】図1

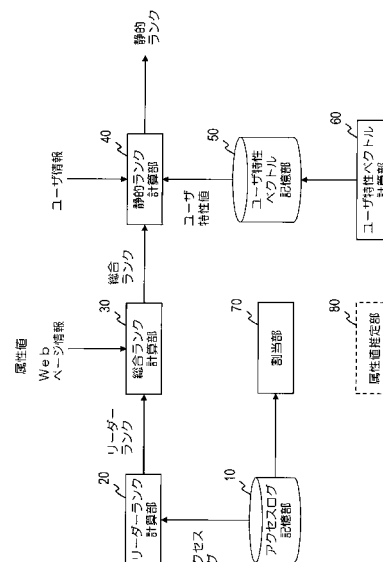


図1

【特許請求の範囲】**【請求項 1】**

複数の属性のそれぞれは固有の複数の属性値を持つとして、複数の属性のそれぞれにおいて少なくとも 1 つの属性値が与えられた各ユーザの、ドキュメントへのアクセスログが記憶されたアクセスログ記憶部と、

ある属性値を有するユーザの集合のあるドキュメントについての関連性の高さを表す指標をその属性値を有するユーザの集合のそのドキュメントについてのリーダーランクとして、上記アクセスログを用いて、各属性値を有するユーザの集合の各ドキュメントについてのリーダーランクを計算するリーダーランク計算部と、

1 以上の属性のそれぞれから選択される属性値で特定されるセグメントをそれらの属性値のすべてを有するユーザの集合とし、あるセグメントに属するユーザの集合のあるドキュメントについての関連性の高さを表す指標をそのセグメントに属するユーザの集合のそのドキュメントについての総合ランクとして、スコアを付与する際に考慮の対象となるユーザの 1 以上の属性のそれぞれから選択された属性値及びドキュメント情報が入力され、それらの入力された各属性値を有するユーザの集合のそのドキュメント情報で特定されるドキュメントについての上記計算されたリーダーランクを統合して、それらの入力された属性値で特定されるセグメントに属するユーザの集合のそのドキュメントについての総合ランクを計算する総合ランク計算部と、

を含むドキュメントスコア算出装置。

【請求項 2】

請求項 1 に記載のドキュメントスコア算出装置において、

あるユーザのあるドキュメントについての関心の高さを表す指標をそのユーザのそのドキュメントについてのユーザ特性値として、各ユーザの各ドキュメントについてのユーザ特性値が記憶されたユーザ特性ベクトル記憶部と、

上記スコアを付与する際に考慮の対象となるユーザを特定するユーザ情報が入力され、上記計算された総合ランクと、上記ユーザ特性ベクトル記憶部から読み込まれた、そのユーザの上記ドキュメント情報で特定されるドキュメントについてのユーザ特性値とを用いて静的ランクを計算する静的ランク計算部とを更に含むことを特徴とするドキュメントスコア算出装置。

【請求項 3】

請求項 1 又は 2 に記載のドキュメントスコア算出装置において、

上記リーダーランク計算部は、上記アクセスログを用いて、各属性値を有するユーザの集合に属するユーザが各ドキュメントを閲覧した回数、及び、各属性値を有するユーザ集合に属するユーザのうちの各ドキュメントを閲覧したユーザの人数を少なくとも求め、それらの求めた回数及び人数を用いて、上記リーダーランクを計算することを特徴とするドキュメントスコア算出装置。

【請求項 4】

請求項 1 から 3 の何れかに記載のドキュメントスコア算出装置において、

上記アクセスログによるとアクセスされたドキュメント（以下、リンク元ドキュメントとする。）のリンク先のドキュメントであって、上記アクセスログによるとアクセスされていないドキュメントの総合ランク又は静的ランクとして、上記リンク元ドキュメントの総合ランク又は静的ランクの A （ A は 0 より大 1 以下の実数）倍の値を割り当てる割当部を含むことを特徴とするドキュメントスコア算出装置。

【請求項 5】

請求項 1 から 4 の何れかに記載のドキュメントのスコア算出装置において、

上記ドキュメントは、Web ページであることを特徴とするドキュメントスコア算出装置。

【請求項 6】

アクセスログ記憶部には、複数の属性のそれぞれは固有の複数の属性値を持つとして、複数の属性のそれぞれにおいて少なくとも 1 つの属性値が与えられた各ユーザの、ドキュ

10

20

30

40

50

メントへのアクセスログが記憶されており、

リーダーランク計算部が、ある属性値を有するユーザの集合のあるドキュメントについての関連性の高さを表す指標をその属性値を有するユーザの集合のそのドキュメントについてのリーダーランクとして、上記アクセスログを用いて、各属性値を有するユーザの集合の各ドキュメントについてのリーダーランクを計算するリーダーランク計算ステップと、

総合ランク計算ステップが、1以上の属性のそれぞれから選択される属性値で特定されるセグメントをそれらの属性値のすべてを有するユーザの集合とし、あるセグメントに属するユーザの集合のあるドキュメントについての関連性の高さを表す指標をそのセグメントに属するユーザの集合のそのドキュメントについての総合ランクとして、スコアを付与する際に考慮の対象となるユーザの1以上の属性のそれぞれから選択された属性値及びドキュメント情報が入力され、それらの入力された各属性値を有するユーザの集合のそのドキュメント情報で特定されるドキュメントについての上記計算されたリーダーランクを統合して、それらの入力された属性値で特定されるセグメントに属するユーザの集合のそのドキュメントについての総合ランクを計算する総合ランク計算ステップと、を含むドキュメントスコア算出方法。

10

【請求項7】

請求項1から4の何れかに記載のドキュメントスコア算出装置の各部としてコンピュータを機能させるためのプログラム。

【発明の詳細な説明】

20

【技術分野】

【0001】

この発明は、Webページ等のテキストを含むドキュメントに対して、関連性の高さ、人気度合い等に応じたスコアをメタ情報として付与する技術に関する。

【背景技術】

【0002】

Webページに対して、その重要度に応じたスコアを付与する技術として、非特許文献1に記載された技術が知られている。

【0003】

非特許文献1に記載された技術（いわゆるPageRank（登録商標））では、ページAからページBへのリンクをページAによるページBへの支持投票とみなし、その投票数によりページBの重要度を判断し、その重要度に応じたスコアをページBに付与する。

30

【非特許文献1】グーグル株式会社、[online]、[平成20年10月7日検索]、インターネット<URL: http://www.google.co.jp/intl/ja/why_use.html>

【発明の開示】

【発明が解決しようとする課題】

【0004】

上記非特許文献1の技術は、スコアを付与する際に、ユーザの性別、年代、居住地及び趣味等の属性を考慮していない。したがって、ユーザの属性が違って、Webページに付与されるスコアは同じであった。

40

【0005】

この発明は、上記に鑑みて、ユーザの属性を考慮して、Webページ等のドキュメントにスコアを付与するドキュメントスコア算出装置、方法及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】

【0006】

この発明によるドキュメントスコア算出装置は、複数の属性のそれぞれは固有の複数の属性値を持つとして、複数の属性のそれぞれにおいて少なくとも1つの属性値が与えられた各ユーザの、ドキュメントへのアクセスログが記憶されたアクセスログ記憶部と、ある属性値を有するユーザの集合のあるドキュメントについての関連性の高さを表す指標をそ

50

の属性値を有するユーザの集合のそのドキュメントについてのリーダーランクとして、アクセスログを用いて、各属性値を有するユーザの集合の各ドキュメントについてのリーダーランクを計算するリーダーランク計算部と、1以上の属性のそれぞれから選択される属性値で特定されるセグメントをそれらの属性値のすべてを有するユーザの集合とし、あるセグメントに属するユーザの集合のあるドキュメントについての関連性の高さを表す指標をそのセグメントに属するユーザの集合のそのドキュメントについての総合ランクとして、スコアを付与する際に考慮の対象となるユーザの1以上の属性のそれぞれから選択された属性値及びドキュメント情報が入力され、それらの入力された各属性値を有するユーザの集合のそのドキュメント情報で特定されるドキュメントについての計算されたリーダーランクを統合して、それらの入力された属性値で特定されるセグメントに属するユーザの集合のそのドキュメントについての総合ランクを計算する総合ランク計算部と、を含む。

10

【発明の効果】

【0007】

ユーザが有する各属性値についてのリーダーランクを統合して総合ランクを計算することにより、ユーザの属性を考慮してドキュメントにスコア（総合ランク乃至静的ランク）を付与することが可能となる。

【発明を実施するための最良の形態】

【0008】

以下、この発明の一実施形態を説明する。図1に、この発明によるドキュメントスコア算出装置の例の機能ブロックを示す。図2に、この発明によるドキュメントスコア算出方法の例を示す。以下、スコアを付与する対象となるドキュメントがWebページである場合を例に挙げて説明する。

20

【0009】

まず、「属性」の概念について説明をする。「属性」とは、ユーザの性別、年代、居住地、趣味、年収、学歴、職種等のユーザの特徴を示す枠組みである。各属性は、固有の属性値を有する。例えば、「性別」という属性は、「男」、「女」という2つの属性値を有し、「年代」という属性は、「10代」、「20代」、…、「80代」という8つの属性値を有するものとする。どのような属性を用いるか、それらの属性のそれぞれの属性値としてどのような属性値を用いるかは、この発明を用いるシステム要件等に応じて適宜定める。

30

【0010】

各ユーザは、複数の属性のそれぞれにおいて、少なくとも1つの属性値を有する。例えば、「性別」の属性においては、ユーザは「男性」又は「女性」の何れか1つの属性値を持つ。一方、例えば「趣味」の属性においては、「ゴルフ」「コンピュータ」等の複数の属性値を持ってもよい。

【0011】

アクセスログ記憶部10には、各ユーザのWebページへのアクセスログが記憶されている。例えば、インターネットサービスプロバイダー（ISP）におけるISP通信ログや、プロキシサーバーにおけるプロキシログが、アクセスログとして、アクセスログ記憶部10に記憶される。アクセスログ記憶部10に記憶されるアクセスログは、参考文献1に記載された手法により正規化されていてもよい。アクセスログは、後述するリーダーランクを計算することができるものであれば、その形式は問わない。

40

〔参考文献1〕[online]、[平成20年10月7日検索]、インターネット<URL: http://en.wikipedia.org/wiki/URL_normalization>

【0012】

<ステップS1>

リーダーランク計算部20は、アクセスログ記憶部10に記憶されたアクセスログを用いて、各属性値を有するユーザの集合の各Webページについてのリーダーランクを計算する（ステップS1）。ここで、ある属性値を有するユーザの集合のあるWebページについての関連性の高さを表す指標を、その属性値を有するユーザの集合のそのWebページ

50

ジについてのリーダーランクと定義する。計算されたリーダーランクは総合ランク計算部 30 に送られる。

【0013】

例えば属性として2つの属性「性別」「年代」が定義されており、「性別」の属性値として2つの属性値「男性」「女性」が、「年代」の属性値として4つの属性値「10代」「20代」「30代」「40代以上」が定義されているとする。この場合、リーダーランク計算部20は、「男性」「女性」「10代」「20代」「30代」「40代以上」の6つの属性値のそれぞれについて、その属性値を有するユーザの集合の各Webページについてのリーダーランクを計算する。すなわち、まず、「男性」という属性値を有するユーザについてのアクセスログを用いて、「男性」という属性値を有するユーザの集合の各Webページについてのリーダーランクを計算する。次に、「女性」という属性値を有するユーザについてのアクセスログを用いて、「女性」という属性値を有するユーザの集合の各Webページについてのリーダーランクを計算する。同様に、「10代」という属性値を有するユーザについてのアクセスログを用いて、「10代」という属性値を有するユーザの集合の各Webページについてのリーダーランクを計算する。「20代」「30代」「40代以上」の属性値についてもそれぞれ同様にリーダーランクを計算する。このように、属性値の数だけ、各Webページについてのリーダーランクを計算する。

【0014】

以下、Webページを、Webページ w_i ($i = 1, \dots, I$)と表記する。 I は、スコアを付与することが可能なWebページの総数である。アクセスログ記憶部10に記憶されたアクセスログによるとアクセスされたWebページが、リーダーランク、総合ランク、静的ランク等のスコアを付与することが可能なWebページとなる。

【0015】

以下、属性値 z を有するユーザの集合のWebページ w_i についてのリーダーランクを計算する方法の例を説明する。上記の例だと、「男性」「女性」「10代」「20代」「30代」「40代以上」のそれぞれが z に該当する。属性値 z を有するユーザのアクセスログから、閲覧頻度 $f(z, w_i)$ 、ユニークユーザ数 $u(z, w_i)$ 、アクセス日数 $d(z, w_i)$ 、滞在時間 $t(z, w_i)$ 、最後に閲覧してからの時間 $r(z, w_i)$ 、ドキュメントクリック数 $c(z, w_i)$ 、ページ到達ステップ数 $s(z, w_i)$ をそれぞれ計算する。

【0016】

図3に例示する、リーダーランク計算部20の閲覧頻度計算部21が閲覧頻度 $f(z, w_i)$ を計算し、ユニークユーザ数計算部22がユニークユーザ数 $u(z, w_i)$ を計算し、アクセス日数計算部23がアクセス日数 $d(z, w_i)$ を計算し、滞在時間計算部24が滞在時間 $t(z, w_i)$ を計算し、閲覧時間計算部25が最後に閲覧してからの時間 $r(z, w_i)$ を計算し、ドキュメントクリック数計算部26がドキュメントクリック数 $c(z, w_i)$ を計算し、ページ到達ステップ数計算部27がページ到達ステップ数 $s(z, w_i)$ を計算する。これらの詳細については後述する。

【0017】

スコア統合部28は、これら計算された閲覧頻度 $f(z, w_i)$ 、ユニークユーザ数 $u(z, w_i)$ 、アクセス日数 $d(z, w_i)$ 、滞在時間 $t(z, w_i)$ 、最後に閲覧してからの時間 $r(z, w_i)$ 、ドキュメントクリック数 $c(z, w_i)$ 、及び、ページ到達ステップ数 $s(z, w_i)$ を統合して、属性値 z を有するユーザの集合のWebページ w_i についてのリーダーランク $Pr(z, w_i)$ を計算する。

【0018】

閲覧頻度 $f(z, w_i)$ 、ユニークユーザ数 $u(z, w_i)$ 、アクセス日数 $d(z, w_i)$ 、滞在時間 $t(z, w_i)$ 、ドキュメントクリック数 $c(z, w_i)$ 、及び、ページ到達ステップ数 $s(z, w_i)$ は、これらの値が高いほど、属性値 z を有するユーザの集合のWebページ w_i についての関連性は高いと考えられる。一方、最後に閲覧してからの時間 $r(z, w_i)$ は、この値が高いほど、属性値 z を有するユーザの集合のWebページ

10

20

30

40

50

ージ w_i についての関連性は低いと考えられる。

【0019】

したがって、スコア統合部 28 は、閲覧頻度 $f(z, w_i)$ 、ユニークユーザ数 $u(z, w_i)$ 、アクセス日数 $d(z, w_i)$ 、滞在時間 $t(z, w_i)$ 、ドキュメントクリック数 $c(z, w_i)$ 、及び、ページ到達ステップ数 $s(z, w_i)$ についての単調増加関数であり、かつ、最後に閲覧してからの時間 $r(z, w_i)$ についての単調減少関数である関数を用いて、リーダーランク $Pr(z, w_i)$ を計算する。

【0020】

スコア統合部 28 による統合は、例えば以下の式に基づいて行う。 α 、 β 、 γ 、 ε 、 ζ 、 η 、 θ は、任意の正の実数であり、所望の結果を得ることができるよう適宜設定される。

【0021】

【数 1】

$$Pr(z, w_i) = \frac{\alpha \cdot f(z, w_i) \times \beta \cdot d(z, w_i) \times \gamma \cdot t(z, w_i) \times \varepsilon \cdot u(z, w_i) \times \zeta \cdot c(z, w_i) \times \eta \cdot s(z, w_i)}{\theta \cdot r(z, w_i)}$$

… (1)

この式は例示であり、適宜変形しても良い。例えば、 $f(z, w_i)$ 、 $d(z, w_i)$ 、 $\dots$ 、 $r(z, w_i)$ に、 $\log(\cdot)$ 等の関数を作作用させてもよい。例えば、上記式において $f(z, w_i)$ に代えて $\log(f(z, w_i))$ を用いることで、高い頻度のアクセスでもスコアへの反映を小さくすることができる。また、ユニットステップ関数を $U(\cdot)$ として、上記式において $u(z, w_i)$ に代えて $U(u(z, w_i) - u_0) \times u(z, w_i)$ を用いることで、ユニークユーザ数 $u(z, w_i)$ が u_0 以上である場合のみ、 $Pr(z, w_i)$ が非ゼロとなるようにしてもよい。 u_0 は任意の自然数である。

【0022】

上記 (1) 式において、値が高いほど属性値 z を有するユーザの集合の Web ページ w_i についての関連性が高いことを示す、閲覧頻度 $f(z, w_i)$ 、ユニークユーザ数 $u(z, w_i)$ 、アクセス日数 $d(z, w_i)$ 、滞在時間 $t(z, w_i)$ 、ドキュメントクリック数 $c(z, w_i)$ 、及び、ページ到達ステップ数 $s(z, w_i)$ が分子にあれば、上記 (1) 式の分子を積で構成しなくてもよい。上記 (1) 式の分子を、例えば $\alpha \cdot f(z, w_i) + \beta \cdot d(z, w_i) + \dots + \eta \cdot r(z, w_i)$ のように、 $f(z, w_i)$ 、 $d(z, w_i)$ 、 $\dots$ 、 $r(z, w_i)$ の和で構成してもよい。

【0023】

閲覧頻度 $f(z, w_i)$ は、属性値 z を有するユーザの集合に属するユーザが各 Web ページを閲覧した回数である。属性値 z を有するユーザの集合に属する各ユーザを j ($j = 1, \dots, J$) とする。 J は、属性値 z を有するユーザの集合に属するユーザの総数である。閲覧頻度 $f(z, w_i)$ は、各ユーザ j が Web ページ w_i を参照した回数 $f_j(z, w_i)$ の和である。すなわち、閲覧頻度 $f(z, w_i) = \sum_{j=1}^J f_j(z, w_i)$ である。

【0024】

図 4 を参照して、閲覧頻度 $f(z, w_i)$ の求め方の具体例を説明する。図 4 は、属性値 z を有するユーザのアクセスログであるとする。図 4 はアクセスログの例であり、アクセスログは図 4 のものに限定されない点に注意する。属性値 z を有するユーザは、12N3456 というユーザ ID で特定されるユーザ (以下、ユーザ 12N3456 とする。) 及び、54S1222 というユーザ ID で特定されるユーザ (以下、ユーザ 54S1222 とする。) の 2 人であるとし、これらの 2 人の所定の期間内におけるアクセスログの全てが、図 4 に掲げられているとする。「<http://sony.jp/products/ic-recorder>」を Web ページ w_i とすると、ユーザ 12N3456 は、Web ページ w_i に 1 回アクセスしている。また、ユーザ 54S1222 も、Web ページ w_i に 1 回アクセスしている。し

10

20

30

40

50

たがって、この例だと、閲覧頻度 $f(z, w_i) = 1 + 1 = 2$ となる。

【0025】

ユニークユーザ数 $u(z, w_i)$ は、属性値 z を有するユーザの集合の中で Web ページ w_i にアクセスしたユーザの人数のことである。例えば、属性値 z を有するユーザは A、B 及び C の 3 人であり、A、B 及び C のアクセスログから、A が Web ページ w_i を 1 回見ており、B が Web ページ w_i を 3 回見ており、C が Web ページ w_i を見ていないことがわかるとする。この場合、Web ページ w_i を見ているのは、A と B の 2 人なので、ユニークユーザ数 $u(z, w_i) = 2$ となる。同一のユーザが 2 回以上 Web ページ w_i を見ていたとしてもそのことは考慮されず、ユニークユーザ数 $u(z, w_i)$ としては 1 がカウントされるだけである。

10

【0026】

アクセス日数 $d(z, w_i)$ は、属性値 z を有するユーザが、一定期間内に、Web ページ w_i にアクセスした日数のことである。例えば、一定期間を一週間（7 日間）とする。月曜日から金曜日（5 日間）までは属性値 z を有するユーザの何れかが Web ページ w_i に毎日アクセスしているが、土曜日及び日曜日（2 日間）は属性値 z を有するユーザの何れもが Web ページ w_i にアクセスしていないとする。この場合、アクセス日数 $d(z, w_i) = 5$ となる。

【0027】

滞在時間 $t(z, w_i)$ は、属性値 z を有するユーザ j ($j = 1, \dots, J$) が Web ページ w_i を閲覧した閲覧時間 $t_j(z, w_i)$ の平均値である。すなわち、滞在時間 $t(z, w_i) = (\sum_{j=1}^J t_j(z, w_i)) / J$ である。

20

【0028】

図 5 を参照して、閲覧時間 $t_j(z, w_i)$ の求め方の例を説明する。図 5 はアクセスログの例であり、アクセスログは図 5 のものに限定されない点に注意する。図 5 において、ユーザ 54S1222 は、2006 年 6 月 20 日 15 時 40 分に「http://sony.jp/products/ic-recorder」にアクセスしており、2006 年 6 月 20 日 15 時 48 分に他のページに移動している。そのことは、移動先の Web ページの URL を表す「URL」の欄と、移動元の Web ページの URL を表す「Ref」とに「http://sony.jp/products/ic-recorder」があることから判断することができる。「http://sony.jp/products/ic-recorder」を Web ページ w_i とし、ユーザ 54S1222 を j とすると、閲覧時間 $t_j(z, w_i)$ はそれらの時間の差である 8 分である。

30

【0029】

なお、閲覧時間 $t_j(z, w_i)$ をその Web ページ w_i のデータ量で割り閲覧速度 $v_j(z, w_i)$ を求めて、その閲覧速度 $v_j(z, w_i)$ の値に応じて、閲覧時間 $t_j(z, w_i)$ を変化させてもよい。閲覧速度 $v_j(z, w_i)$ は、ユーザ j が Web ページ w_i を閲覧するに際し単位データ容量当たりにかかった閲覧時間を意味する。例えば、閲覧速度 $v_j(z, w_i)$ が所定の閾値を上回る場合には、その Web ページ w_i は実際にはユーザ j に閲覧されずに放置されていた可能性があるため、その閲覧時間 $t_j(z, w_i) = 0$ とする。また、例えば、閲覧速度 $v_j(z, w_i)$ が所定の閾値を下回る場合には、ユーザ j は、その Web ページ w_i は実質的に閲覧せずに、直ぐに他の Web ページに移動した可能性があるため、その閲覧時間 $t_j(z, w_i) = 0$ とする。

40

【0030】

最後に閲覧してからの時間 $r(z, w_i)$ は、属性値 z を有するユーザ j が同一の Web ページ w_i に対して 2 度アクセスしている場合の一度目のアクセスの閲覧時刻と二度目のアクセスの閲覧時刻との差 $d_j(z, w_i)$ の平均値である。すなわち、属性値 z を有する各ユーザ j について差 $d_j(z, w_i)$ を求めて、これらの差 $d_j(z, w_i)$ の平均値が最後に閲覧してからの時間 $r(z, w_i)$ となる。

【0031】

ドキュメントクリック数 $c(z, w_i)$ は、属性値 z を有するユーザ j が Web ページ w_i に含まれるリンクを所定の期間内にクリックした回数の平均値である。まず、Web

50

ページ w_i に含まれる、PDF ファイル、画像、動画等のドキュメントへのリンクを抽出する。そして、属性値 z を有するユーザ j が、Web ページ w_i (その URL を $\text{http://www.xx.co.jp}$ とする。) へアクセスしてから予め定められた期間内に、上記抽出されたリンク (例えば、 $\text{http://www.xx.co.jp/aaa.pdf}$) をクリックした回数 $c_j(z, w_i)$ をカウントする。属性値 z を有する各ユーザ j について回数 $c_j(z, w_i)$ を求めて、これらの回数 $c_j(z, w_i)$ の平均値をドキュメントクリック数 $c(z, w_i)$ とする。

【0032】

ページ到達ステップ数 $s(z, w_i)$ は、簡単に到達することができない Web ページに対して高いスコアを与えようとするものである。属性値 z を有する各ユーザ j が、ページ到達ステップ数 $s(z, w_i)$ の計算の基礎となるハブとなる Web ページ (以下、ハブページ) から Web ページ w_i に到達するまでに辿ったステップ数 $step_j(z, w_i)$ 、及び、Web ページ w_i の閲覧時間 $t_j(z, w_i)$ を基にして、属性値 z を有する各ユーザ j のページ到達ステップ数 $s_j(z, w_i)$ を求め、これらのユーザ j のページ到達ステップ数 $s_j(z, w_i)$ の平均値を、ページ到達ステップ数 $s(z, w_i)$ とする。

【0033】

以下、図 6 を参照して、ユーザ j のページ到達ステップ数 $s_j(z, w_i)$ の求め方を説明する。ユーザ j のアクセスログのリファラーの欄 (図 4, 図 5 においては、「Ref」と表記している。) に一定の回数以上表れ、その Web ページを基点として移動したページの数がある一定の数以上の Web ページをハブページとする。図 6 においては、Web ページ w_A がハブページである。

【0034】

図 6 の例においては、ユーザ j は Web ページ w_A からまず Web ページ w_B に移動し Web ページ w_B を 0.5 秒閲覧している。その後、ユーザ j は Web ページ w_A に戻り、そこから Web ページ w_C に移動し Web ページ w_C を 60 秒閲覧している。その後、ユーザ j は Web ページ w_A に戻り、そこから Web ページ w_D に移動し Web ページ w_D を 4 秒閲覧している。その後、ユーザ j は Web ページ w_A に戻り、そこから Web ページ w_E に移動し Web ページ w_E を 4 秒閲覧している。その後、ユーザ j は、Web ページ w_E から Web ページ w_F に移動し Web ページ w_F を 300 秒閲覧している。その後、ユーザ j は Web ページ w_A に戻り、そこから Web ページ w_G に移動し Web ページ w_G を 0.5 秒閲覧している。したがって、Web ページ $w_B, w_C, w_D, w_E, w_F, w_G$ のそれぞれの $step_j(z, w_i), t_j(z, w_i)$ は次のようになる。

$$step_j(z, w_B) = 1, t_j(z, w_B) = 0.5$$

$$step_j(z, w_C) = 2, t_j(z, w_C) = 60$$

$$step_j(z, w_D) = 3, t_j(z, w_D) = 4$$

$$step_j(z, w_E) = 4, t_j(z, w_E) = 4$$

$$step_j(z, w_F) = 5, t_j(z, w_F) = 300$$

$$step_j(z, w_G) = 6, t_j(z, w_G) = 0.5$$

【0035】

$ss_j(z, w_i) = step_j(z, w_i) \times t_j(z, w_i)$ と定義し、Web ページ $w_B, w_C, w_D, w_E, w_F, w_G$ のそれぞれの $ss_j(z, w_i)$ を計算すると、次のようになる。

$$ss_j(z, w_B) = 0.5$$

$$ss_j(z, w_C) = 120$$

$$ss_j(z, w_D) = 12$$

$$ss_j(z, w_E) = 16$$

$$ss_j(z, w_F) = 1500$$

$$ss_j(z, w_G) = 3$$

【0036】

$srank_j(z, w_i)$ を、 $ss_j(z, w_i)$ が大きい順に Web ページ w_B, w

C, wD, wE, wF, wGを並び替えた場合のWebページwiの順位とすると、WebページwB, wC, wD, wE, wF, wGのそれぞれの $srank_j(z, wi)$ は次のようになる。

$srank_j(z, wB) = 6$
 $srank_j(z, wC) = 2$
 $srank_j(z, wD) = 4$
 $srank_j(z, wE) = 3$
 $srank_j(z, wF) = 1$
 $srank_j(z, wG) = 5$

【0037】

この $srank_j(z, wi)$ の逆数である $1/srank_j(z, wi) = sj(z, wi)$ とする。すなわち、WebページwB, wC, wD, wE, wF, wGのそれぞれの $sj(z, wi)$ は次のようになる。

$sj(z, wB) = 1/6$
 $sj(z, wC) = 1/2$
 $sj(z, wD) = 1/4$
 $sj(z, wE) = 1/3$
 $sj(z, wF) = 1$
 $sj(z, wG) = 1/5$

Webページwiについて各ユーザjの $sj(z, wi)$ を求め、これらの平均値を属性値zを有するユーザの集合のWebページwiについてのページ到達ステップ数 $s(z, wi)$ とする。

【0038】

<ステップS2>

次に、総合ランク計算部30の説明で用いる概念である「セグメント」「総合ランク」について説明する。「セグメント」とは、1以上の属性のそれぞれから選択された属性値のすべてを有するユーザの集合のことである。すなわち、1以上の属性のそれぞれから選択される属性値で特定されるセグメントをそれらの属性値のすべてを有するユーザの集合とする。「総合ランク」とは、あるセグメントに属するユーザの集合のあるWebページについての関連性の高さを表す指標である。すなわち、あるセグメントに属するユーザの集合のあるWebページについての関連性の高さを表す指標をそのセグメントに属するユーザの集合のそのWebページについての総合ランクとする。

【0039】

総合ランク計算部30には、スコアを求める対象となるユーザの1以上の属性のそれぞれから選択された属性値、及び、Webページ情報が入力される。Webページ情報とは、Webページを特定するための情報である。総合ランク計算部30は、それらの入力された属性値で特定されるセグメントに属するユーザの集合の、入力されたWebページ情報で特定されるWebページについての総合ランクを計算する(ステップS2)。その際、総合ランク計算部30は、それらの入力された各属性値を有するユーザの集合の上記Webページ情報で特定されるWebページについての上記計算されたリーダーランクを統合することにより、総合ランクを計算する。

【0040】

総合ランク計算部30に、入力された属性値が「男性」及び「20代」であり、入力されたWebページ情報がWebページwiを特定するものであるとする。「男性」は「性別」という属性から選択された属性値であり、「20代」は「年代」という属性から選択された属性値である。この場合、総合ランク計算部30は、「男性」かつ「20代」、すなわち「男性20代」のセグメントに属するユーザの集合のWebページwiについての総合ランクを計算する。その際、総合ランク計算部30は、リーダーランク計算部20が計算した、属性値「男性」を有するユーザの集合のWebページwiについてのリーダーランク $Pr(\text{男性}, wi)$ 、及び、属性値「20代」を有するユーザの集合のWebページ

10

20

30

40

50

ジ w_i についてのリーダーランク $Pr(20代, w_i)$ を統合して、その総合ランクを求める。

【0041】

リーダーランクの統合は、例えば次の式に基づいて行う。 α 、 β は、任意の正の実数であり、所望の結果を得ることができるように適宜設定される。 $\sum_{年代} Pr(年代, w_i)$ は、属性「年代」の各属性値についてのリーダーランクの和である。すなわち、「年代」の属性値が、「10代」「20代」「30代」「40代以上」である場合には、 $\sum_{年代} Pr(年代, w_i) = Pr(10代, w_i) + Pr(20代, w_i) + Pr(30代, w_i) + Pr(40代以上, w_i)$ である。

【0042】

【数2】

$$Sr(\text{男性}20代, w_i) = \alpha \cdot \frac{Pr(\text{男性}, w_i)}{Pr(\text{男性}, w_i) + Pr(\text{女性}, w_i)} + \beta \cdot \frac{Pr(20代, w_i)}{\sum_{年代} Pr(年代, w_i)} \quad \dots (2)$$

【0043】

この例では、リーダーランク $Pr(\text{男性}, w_i)$ 及びリーダーランク $Pr(20代, w_i)$ のそれぞれを正規化した後に、正規化したリーダーランク $Pr(\text{男性}, w_i)$ 及びリーダーランク $Pr(20代, w_i)$ を統合している。このように、リーダーランクを正規化した後に統合することにより、ある属性の各属性値を有するユーザにおいて万遍なく評価が高い Web ページよりも、ある属性の特定の属性値を有するユーザにおいて評価が高い Web ページに高いスコアを付与することが可能となる。属性が「性別」であり、その属性値が「男性」「女性」である場合を例に挙げると、「男性」のユーザ及び「女性」のユーザの両方から評価されているような性差の特徴が表れていない Web ページのスコアを下げて、「男性」のユーザ又は「女性」の一方から評価されているような性差の特徴が表れている Web ページのスコアを上げることができる。

【0044】

もちろん、正規化せずに、例えば下記の式により、総合ランク $Sr(\text{男性}20代, w_i)$ を求めても良い。

$$Sr(\text{男性}20代, w_i) = \alpha \cdot Pr(\text{男性}, w_i) + \beta \cdot Pr(20代, w_i) \quad \dots (3)$$

【0045】

総合ランクの計算の他の例を説明する。総合ランク計算部30に、入力された属性値が「男性」「20代」「コンピュータ」であり、入力された Web ページ情報が Web ページ w_i を特定するものであるとする。「男性」は「性別」という属性から選択された属性値であり、「20代」は「年代」という属性から選択された属性値であり、「コンピュータ」は「趣味」という属性から選択された属性値である。この場合、総合ランク計算部30は、「男性」かつ「20代」かつ「コンピュータ」、すなわち「男性20代コンピュータ」のセグメントに属するユーザの集合の Web ページ w_i についての総合ランクを計算する。その際、総合ランク計算部30は、リーダーランク計算部20が計算した、属性値「男性」を有するユーザの集合の Web ページ w_i についてのリーダーランク $Pr(\text{男性}, w_i)$ 、属性値「20代」を有するユーザの集合の Web ページ w_i についてのリーダーランク $Pr(20代, w_i)$ 、及び、属性値「コンピュータ」を有するユーザの集合の Web ページ w_i についてのリーダーランク $Pr(\text{コンピュータ}, w_i)$ を統合して、その総合ランクを求める。

【0046】

リーダーランクの統合は、例えば次の式に基づいて行う。 α 、 β 、 γ 、 δ は、任意の正の実数であり、所望の結果を得ることができるように適宜設定される。 $\sum_{年代} Pr(年代, w_i)$ は、属性「年代」の各属性値についてのリーダーランクの和である。すなわち、「年代」の属性値が、「10代」「20代」「30代」「40代以上」である場合には、

$\sum_{\text{年代}} \text{Pr}(\text{年代}, wi) = \text{Pr}(10\text{代}, wi) + \text{Pr}(20\text{代}, wi) + \text{Pr}(30\text{代}, wi) + \text{Pr}(40\text{代以上}, wi)$ である。同様に、 $\sum_{\text{趣味}} \text{Pr}(\text{趣味}, wi)$ は、属性「趣味」の各属性値についてのリーダーランクの和である。

【0047】

【数3】

$Sr(\text{男性}20\text{代コンピュータ}, wi)$

$$= \left[\alpha \cdot \frac{\text{Pr}(\text{男性}, wi)}{\text{Pr}(\text{男性}, wi) + \text{Pr}(\text{女性}, wi)} + \beta \cdot \frac{\text{Pr}(20\text{代}, wi)}{\sum_{\text{年代}} \text{Pr}(\text{年代}, wi)} + \gamma \cdot \frac{\text{Pr}(\text{コンピュータ}, wi)}{\sum_{\text{趣味}} \text{Pr}(\text{趣味}, wi)} \right]^\delta \quad 10$$

… (4)

このように、総合ランク計算部30にはユーザの1以上の属性のそれぞれから選択された属性値が入力され、総合ランク計算部30は、これらの入力された各属性値についてのリーダーランクから、総合ランクを求める。したがって、求まった総合ランクは、ユーザの属性を考慮したものとなる。

上記(2)から(4)式も、上記(1)式と同様に例示であり、適宜変形しても良い。

【0048】

20

<ステップS3>

静的ランク計算部40の説明で用いる「ユーザ特性値」の概念について説明する。「ユーザ特性値」は、ユーザのWebページについての関心の高さを表す指標である。すなわち、あるユーザのあるWebページについての関心の高さを表す指標をそのユーザのそのWebページについてのユーザ特性値とする。各ユーザの各Webページについてのユーザ特性値はユーザ特性ベクトル計算部60により計算されて、ユーザ特性ベクトルの形でユーザ特性ベクトル記憶部50に記憶される。

【0049】

ユーザjのWebページwi(i=1, ..., I[#](I[#]はユーザ特性値を求めようとするWebページの総数))についてのユーザ特性値jnw_iの求め方の例を説明する。fd_j(wi)を、ユーザjがWebページwiを参照した回数f_j(wi)と、ユーザjが所定の期間内にWebページwiにアクセスした日数d_j(wi)とをかけたものとする。各Webページwi(i=1, ..., I[#])について、fd_j(wi)を求める。

30

$$fd_j(wi) = f_j(wi) \times d_j(wi)$$

【0050】

このとき、ユーザ特性値jnw_iは、次式のように計算される。下記式において、IはIを意味する。αは任意の正の実数であり、所望の結果を得ることができるよう適宜設定される。

$$jnw_i = \alpha \cdot (fd_j(wi) / \sum_{i=1}^{I^{\#}} fd_j(wi))$$

【0051】

40

ユーザ特性ベクトルpersonal_jは、jnw_i(i=1, ..., I[#])を各成分に持つベクトルのことである。

$$personal_j = (jnw_1, jnw_2, \dots, jnw_{I^{\#}})$$

【0052】

静的ランク計算部40による静的ランクの計算例を以下に説明する。静的ランク計算部40には、ユーザjを特定するためのユーザ情報が入力される。静的ランク計算部40は、そのユーザ情報から特定されるユーザjの、スコアを付与しようとするWebページwiについてのユーザ特性値jnw_iを、ユーザ特性ベクトル記憶部50から読み出す。静的ランク計算部40は、総合ランク計算部30が計算した総合ランクSr(s, wi)に、ユーザ特性値jnw_iをかけて、静的ランクST(j, wi)を求める(ステップS3

50

）。ここで、 s は、上記ユーザ情報から特定されるユーザ j の 1 以上の属性から選択された属性値の直積である。すなわち、総合ランク計算部 30 に入力された属性値が「男性」及び「20代」である場合には、 $s = \text{「男性 20代」}$ となる。

$$ST(j, w_i) = Sr(s, w_i) \times jnw_i$$

【0053】

ユーザ j のスコアを付与しようとする Web ページ w_i についてのユーザ特性値が、ユーザ特性ベクトル記憶部 50 に記憶されていない場合には、すなわち、 $w_i = njw_i$ となる njw_i が存在しない場合には、総合ランクを静的ランクとしてそのまま出力する。

$$ST(j, w_i) = Sr(s, w_i)$$

【0054】

このように、総合ランクと、ユーザの Web ページについての関心の高さを表す指標であるユーザ特性値とを用いて、静的ランクを求めることにより、ユーザ個人の特性を考慮して Web ページにスコアを付与することが可能となる。

【0055】

この発明によるドキュメントスコア算出装置は、例えば検索エンジンや企業等が独自に構築している EC サイト内等に用いることができる。検索エンジンにこの発明によるドキュメントスコア算出装置が用いられる場合の処理の流れの例を簡単に説明する。

【0056】

ユーザが 20 代であり、女性であり、旅行が趣味であるとする。このユーザが、「イタリア」というワードで検索をする場合、その属性を考慮した検索結果を得たいときには、ユーザは、「イタリア @ 20 代 @ 女性 @ 旅行」というクエリを検索エンジンに送ればよい。この場合、「20 代」「女性」「旅行」という属性値が総合ランク計算部 30 に入力され、総合ランク計算部 30 は、「20 代女性旅行」というセグメントに属するユーザの集合の「イタリア」というワードを含む各 Web ページについての総合ランクを計算する。検索エンジンは、「イタリア」というワードを含む Web ページを総合ランクが高い順に並び替えて、総合ランクが高い順にユーザに提示する。もちろん、必要に応じて総合ランクではなく静的ランクに基づいて Web ページの並び替え・提示を行っても良い。このように、ユーザが自ら属性値を入力してもよい。

【0057】

また、以下に述べるように、ユーザの属性値を推定する属性値推定部 80 を、ドキュメントスコア算出装置又は検索エンジン等に設けて、この属性値推定部 80 が推定したユーザの属性値が総合ランク計算部 30 に入力されるようにしてもよい。

【0058】

属性値推定部 80 は、例えばユーザが用いるコンピュータの IP アドレスから、ユーザの「居住地」という属性の属性値「埼玉」「神奈川」「千葉」等を推定して、総合ランク計算部 80 に入力してもよい。これにより、ユーザの居住地に関連した Web ページに高いスコアが与えられる。これは、不動産情報提供サイト、学習塾・中学・高校受験情報提供サイト等のエリアに応じた情報を提供しているサイトに有効である。この場合であっても、ユーザが自分が住んでいる居住地以外の地域（例えば、東京）に関する情報を頻繁に閲覧している場合には、そのアクセスログにより生成されたユーザ特性値乃至ユーザ特性ベクトルにより、ユーザの居住地ではなく関心の高い地域に関連した Web ページにより高いスコアを与えることが可能である。

【0059】

また、属性値推定部 80 は、例えば「趣味」等の属性の属性値をアクセスログから推定してもよい。その推定の例を以下に述べる。旅行、スポーツ等のジャンルを予め選定し、各ジャンルに属する学習データとしての学習用 Web ページを用意する。各ジャンルの学習用 Web ページの本文を形態素解析し、各ジャンルに特徴的な各単語を TF・IDF 等の手法により抽出する。すなわち、各ジャンルの学習用 Web ページに特徴的であり、出現する頻度が高い単語を複数抽出する。そして、抽出された各ジャンルの特徴的な各単語からなるジャンルベクトルを、各ジャンルごとに求める。各ジャンルのジャンルベクトル

10

20

30

40

50

G_k ($k = 1, \dots, K$ (K は予め選択されたジャンルの総数))は、上記抽出された各ジャンルの特徴的な各単語が各ジャンルの学習用Webページに出現する出現頻度で構成される。ジャンルの総数が40であり、各ジャンルから特徴的な単語として100個が選択された場合には、各ジャンルのジャンルベクトル G_k は4000次元となり、4000個の各単語が各ジャンルの学習用Webページに出現する出現頻度で構成される。アクセスログ記憶部に記憶されたアクセスログを用いて、ユーザがアクセスしたWebページの本文を形態素解析し、ジャンルベクトルに対応するユーザ単語ベクトル U を生成する。ユーザ単語ベクトルは、ユーザがアクセスしたWebページに上記にて抽出された各ジャンルの特徴的な各単語が出現する出現頻度で構成される。各ジャンルベクトル G_k とユーザ単語ベクトル U との内積を求めて、内積値が一定値以上のジャンルベクトル G_k に対応するジャンルを、そのユーザの属性「趣味」の属性値とする。

10

【0060】

このように、属性値推定部80が属性値を推定することにより、属性値を入力するという負担をユーザにかけることなく、ユーザの属性を考慮したWebページへのスコア付与が可能となる。

【0061】

ドキュメントスコア算出装置は、Webページのスコアとして、総合ランク計算部30が計算した総合ランクを最終的に出力しても良い。この場合、ドキュメントスコア算出装置は、静的ランク計算部40、ユーザ特性ベクトル記憶部50及びユーザ特性ベクトル計算部60を有していなくてもよい。

20

【0062】

ドキュメントスコア算出装置は、アクセスログ記憶部10に記憶されたアクセスログによるとアクセスされたWebページ(以下、リンク元Webページとする。)のリンク先のWebページであって、そのアクセスログによるとアクセスされていないWebページの総合ランク又は静的ランクとして、上記リンク元Webページの総合ランク又は静的ランクの A (A は0より大1以下の実数)倍の値を割り当てる割当部70を含んでいてもよい。これにより、最新ニュース記事等のアクセスログに存在しないWebページに対しても、スコアを付与することが可能となる。

【0063】

リーダーランク計算部20内に、属性値 z を有するユーザのWebページ w_i へのアクセスの周期性を分析する周期性分析部29を設けて、その分析結果に応じて閲覧頻度 $f(z, w_i)$ を変化させてもよい。例えば図7(a)のように、属性値 z を有するユーザのWebページ j へのアクセスの周期性に対してFFT等のフーリエ変換をした結果、狭い帯域幅を持つピークが抽出された場合、そのWebページ w_i についての閲覧頻度 $f(z, w_i)$ を下げる。また、例えば図7(b)のように、属性値 z を有するユーザのWebページ j へのアクセスの周期性に対してFFT等のフーリエ変換をした結果、広い帯域幅を持つピークが抽出された場合、そのWebページ w_i についての閲覧頻度 $f(z, w_i)$ を上げ下げしないか、上げる。ピークを含む帯域幅の広いか狭いかは、パワースペクトルが所定の値以上となる帯域の幅が所定の幅以上となるかどうかで例えば判断することができる。

30

40

【0064】

ニュース、ブログ等のユーザが定期的に閲覧するWebページがある場合、ユーザはそのWebページの内容そのものへの興味があるためではなく、そのWebページのニュースサイト及びブログの著者に興味があるために、定期的に閲覧している場合が多い。周期性分析部29を設けることにより、コンテンツそのものに興味を持たれているWebサイト以外のWebサイトへのスコアを下げるができる。

【0065】

これまで、スコアを付与する対象となるドキュメントがWebページである場合を例に挙げて説明して来たが、Webページ以外の、各ユーザのアクセスログを取ることが可能なドキュメントについても上述のWebページに対するスコアの付与と同様にスコアを付

50

与することが可能であるのは言うまでもない。

【0066】

リーダーランク計算部20は、閲覧頻度 $f(z, w_i)$ 、ユニークユーザ数 $u(z, w_i)$ 、アクセス日数 $d(z, w_i)$ 、滞在時間 $t(z, w_i)$ 、最後に閲覧してから
の間 $r(z, w_i)$ 、ドキュメントクリック数 $c(z, w_i)$ 、ページ到達ステップ数 $s(z, w_i)$ の少なくとも1つを求めて、これらを上記と同様に
して統合してリーダーランク $Pr(z, w_i)$ を求めばよい。例えば、リーダーランク $Pr(z, w_i)$ を求める際に、閲覧頻度 $f(z, w_i)$ 及びユニークユーザ数 $u(z, w_i)$ のみを用いてもよい。
この場合、ドキュメントスコア算出装置は、閲覧頻度計算部21、ユニークユーザ数計算部22、アクセス日数計算部23、滞在時間計算部24、閲覧時間計算部25、ドキュメント
クリック数計算部26、ページ到達ステップ数計算部27、スコア統合部28、周期性分析部29のうち使われない部を備えていなくてもよい。

10

【0067】

上述の構成をコンピュータによって実現する場合、ドキュメントスコア算出装置の各部が有する機能の処理内容はプログラムによって記述される。そして、このプログラムをコンピュータで実行することにより、上記各部の機能がコンピュータ上で実現される。

【0068】

すなわち、CPUがプログラムを逐次読み込んで実行することにより、リーダーランク計算部20、総合ランク計算部30、静的ランク計算部40、ユーザ特性ベクトル計算部60、割当部70、属性値推定部80、閲覧頻度計算部21、ユニークユーザ数計算部22、アクセス日数計算部23、滞在時間計算部24、閲覧時間計算部25、ドキュメント
クリック数計算部26、ページ到達ステップ数計算部27、スコア統合部28、周期性分析部29がそれぞれ実現される。また、補助記憶装置又はメモリが、アクセスログ記憶部10、ユーザ特性ベクトル記憶部50として機能する。

20

【0069】

ドキュメントスコア算出装置の各部として機能するCPUは、メモリ又は補助記憶装置から読み込み込んだデータに対して処理を行い、処理を行った後のデータをメモリ又は補助記憶装置に格納する。すなわち、メモリ又は補助記憶装置を介して、ドキュメントスコア算出装置の各部間でデータがやり取りされる。

【0070】

この処理内容を記述したプログラムは、コンピュータで読み取り可能な記録媒体に記録しておくことができる。コンピュータで読み取り可能な記録媒体としては、例えば、磁気記録装置、光ディスク、光磁気記録媒体、半導体メモリ等のようなものでもよいが、具体的には、例えば、磁気記録装置として、ハードディスク装置、フレキシブルディスク、磁気テープ等を、光ディスクとして、DVD(Digital Versatile Disc)、DVD-RAM(Random Access Memory)、CD-ROM(Compact Disc Read Only Memory)、CD-R(Recordable)/RW(ReWritable)等を、光磁気記録媒体として、MO(Magneto-Optical disc)等を、半導体メモリとしてEEP-ROM(Electronically Erasable and Programmable-Read Only Memory)等を用いることができる。

30

【0071】

また、このプログラムの流通は、例えば、そのプログラムを記録したDVD、CD-ROM等の可搬型記録媒体を販売、譲渡、貸与等することによって行う。さらに、このプログラムをサーバコンピュータの記憶装置に格納しておき、ネットワークを介して、サーバコンピュータから他のコンピュータにそのプログラムを転送することにより、このプログラムを流通させる構成としてもよい。

40

【0072】

また、上述した実施形態とは別の実行形態として、コンピュータが可搬型記録媒体から直接このプログラムを読み取り、そのプログラムに従った処理を実行することとしてもよく、さらに、このコンピュータにサーバコンピュータからプログラムが転送されるたびに、逐次、受け取ったプログラムに従った処理を実行することとしてもよい。また、サーバ

50

コンピュータから、このコンピュータへのプログラムの転送は行わず、その実行指示と結果取得のみによって処理機能を実現する、いわゆる A S P (Application Service Provider) 型のサービスによって、上述の処理を実行する構成としてもよい。なお、本形態におけるプログラムには、電子計算機による処理の用に供する情報であってプログラムに準ずるもの(コンピュータに対する直接の指令ではないがコンピュータの処理を基底する性質を有するデータ等)を含むものとする。

【0073】

また、この形態では、コンピュータ上で所定のプログラムを実行させることにより、本装置を構成することとしたが、これらの処理内容の少なくとも一部をハードウェア的に実現することとしてもよい。

10

【0074】

また、上述の各種の処理は、記載に従って時系列に実行されるのみならず、処理を実行する装置の処理能力あるいは必要に応じて並列的にあるいは個別に実行されてもよい。

その他、本発明の趣旨を逸脱しない範囲で適宜変更が可能であることはいうまでもない。

【図面の簡単な説明】

【0075】

【図1】ドキュメントスコア算出装置の例の機能ブロック図。

【図2】ドキュメントスコア算出方法の例を示す図。

【図3】リーダーランク計算部20の例の機能ブロック図。

20

【図4】閲覧頻度を説明するための図。

【図5】閲覧時間を説明するための図。

【図6】ページ到達ステップ数を説明するための図。

【図7】周期性分析部を説明するための図。

【符号の説明】

【0076】

- 10 アクセスログ記憶部
- 20 リーダーランク計算部
- 21 閲覧頻度計算部
- 22 ユニークユーザ数計算部
- 30 総合ランク計算部
- 40 静的ランク計算部
- 50 ユーザ特性ベクトル記憶部
- 60 ユーザ特性ベクトル計算部
- 70 割当部

30

【図 1】

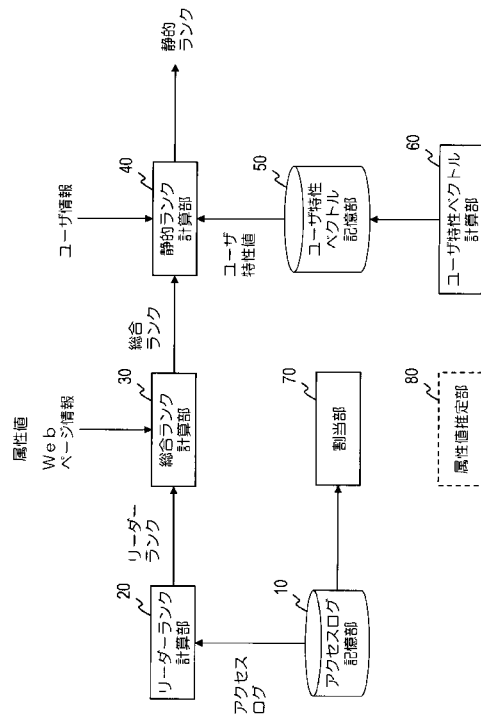


図 1

【図 2】

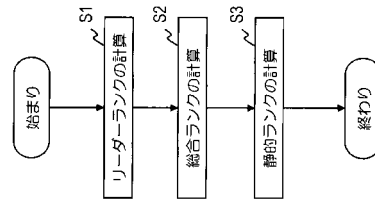


図 2

【図 3】

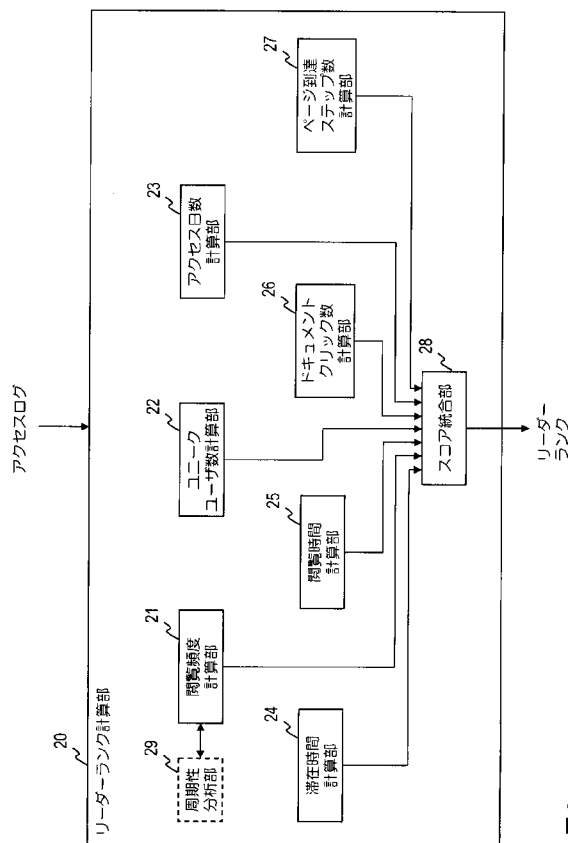


図 3

【図 4】

time	IPアドレス	URL	Ref	ユーザID
2006-06-20-15-06	210.131.43	http://www.ntt.com	http://www.yahoo.co.jp/search?q=NE%99%8B2%	12N3456
2006-06-20-15-05	210.131.43	http://www.yahoo.co.jp/search?q=NE%99%8B2%	http://www.sony.jp/products/ic-recorder/	12N3456
2006-06-20-15-01	210.131.43	http://www.sony.jp/products/ic-recorder/	http://www.yahoo.co.jp/search?q=NE%99%8B2%	12N3456
2006-06-20-15-59	210.131.43	http://www.yahoo.co.jp	-	12N3456
2006-06-20-15-55	203.30.11	http://www.google.co.jp/search?q=NE%99%8B2%	http://www.yamaha.co.jp/soundsketcher/	54S1222
2006-06-20-15-50	203.30.11	http://www.yamaha.co.jp/soundsketcher/	http://www.google.co.jp/search?q=NE%99%8B2%	54S1222
2006-06-20-15-43	203.30.11	http://www.google.co.jp/search?q=NE%99%8B2%	http://www.sony.jp/products/ic-recorder/	54S1222
2006-06-20-15-40	203.30.11	http://www.sony.jp/products/ic-recorder/	http://www.google.co.jp/search?q=NE%99%8B2%	54S1222
2006-06-20-15-35	203.30.11	http://www.google.co.jp/search?q=NE%99%8B2%	http://www.microsoft.com/	54S1222
2006-06-20-15-32	203.30.11	http://www.microsoft.com/	http://www.google.co.jp/search?q=NE%99%8B2%	54S1222
2006-06-20-15-30	203.30.11	http://www.google.co.jp	-	54S1222

図 4

【 図 5 】

time	IPアドレス	URL	Ref	データ量	ユーザID
2006-06-20-16-06	210.131.4.3	http://www.ntt.com	http://www.yahoo.co.jp/search?q=NE9%8B2%		12N3456
2006-06-20-16-05	210.131.4.3	http://www.yahoo.co.jp/search?q=NE9%8B2%	http://www.sony.jp/products/ic-recorder/		12N3456
2006-06-20-16-01	210.131.4.3	http://www.sony.jp/products/ic-recorder/	http://www.yahoo.co.jp/search?q=NE9%8B2%		12N3456
2006-06-20-15-59	210.131.4.3	http://www.yahoo.co.jp	-		12N3456
2006-06-20-15-55	203.30.1.1	http://www.yahoo.co.jp/search?q=NE9%8B2%	http://www.yahoo.co.jp/soundsketcher/		54S1222
2006-06-20-15-50	203.30.1.1	http://www.yahoo.co.jp/soundsketcher/	http://www.google.co.jp/search?q=NE9%8B2%		54S1222
2006-06-20-15-48	203.30.1.1	http://www.google.co.jp/search?q=NE9%8B2%	http://www.sony.jp/products/ic-recorder/		54S1222
2006-06-20-15-40	203.30.1.1	http://www.sony.jp/products/ic-recorder/	http://www.google.co.jp/search?q=NE9%8B2%	100KB	54S1222
2006-06-20-15-35	203.30.1.1	http://www.google.co.jp/search?q=NE9%8B2%	http://www.microsoft.com/		54S1222
2006-06-20-15-32	203.30.1.1	http://www.microsoft.com/	http://www.google.co.jp/search?q=NE9%8B2%	30KB	54S1222
2006-06-20-15-30	203.30.1.1	http://www.google.co.jp	-		54S1222

図 5

【 図 7 】

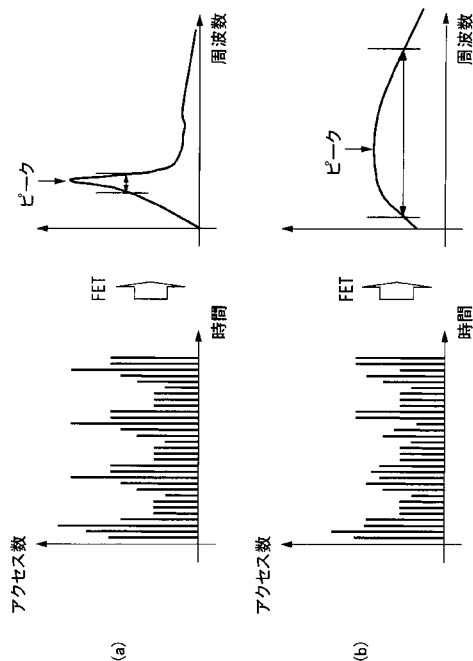


図 7

【 図 6 】

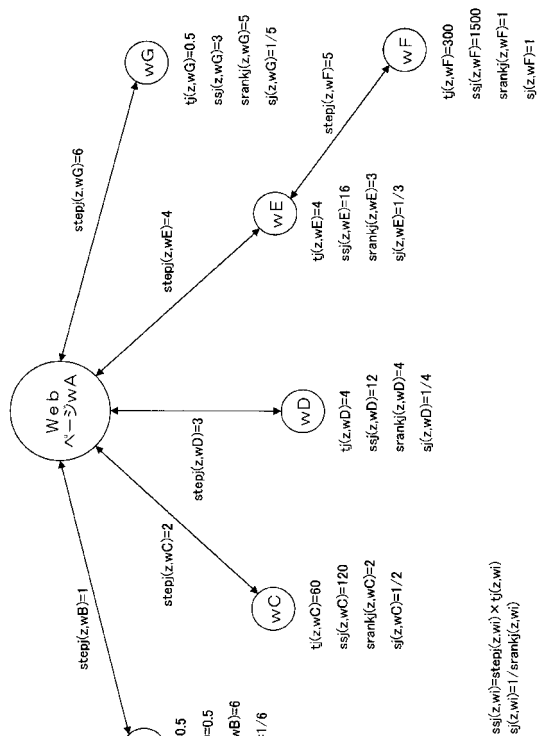


図 6

フロントページの続き

- (72)発明者 片淵 典史
東京都千代田区内幸町一丁目1番6号 エヌ・ティ・ティ・コミュニケーションズ株式会社内
- (72)発明者 中山 彰
東京都千代田区内幸町一丁目1番6号 エヌ・ティ・ティ・コミュニケーションズ株式会社内
- (72)発明者 石崎 貞大
東京都千代田区内幸町一丁目1番6号 エヌ・ティ・ティ・コミュニケーションズ株式会社内
- Fターム(参考) 5B075 PQ75 PR03