

(19) 日本国特許庁(JP)

(12) 特許公報(B2)

(11) 特許番号

特許第4714360号
(P4714360)

(45) 発行日 平成23年6月29日(2011.6.29)

(24) 登録日 平成23年4月1日(2011.4.1)

(51) Int.CI.

G06F 19/00 (2011.01)

F 1

G06F 19/00 110

請求項の数 6 (全 11 頁)

(21) 出願番号 特願2001-85101 (P2001-85101)
 (22) 出願日 平成13年3月23日 (2001.3.23)
 (65) 公開番号 特開2001-312683 (P2001-312683A)
 (43) 公開日 平成13年11月9日 (2001.11.9)
 審査請求日 平成20年3月21日 (2008.3.21)
 (31) 優先権主張番号 09/540976
 (32) 優先日 平成12年3月31日 (2000.3.31)
 (33) 優先権主張国 米国(US)

(73) 特許権者 596170170
 ゼロックス コーポレイション
 XEROX CORPORATION
 アメリカ合衆国、コネチカット州 068
 56、ノーウォーク、ピーオーボックス
 4505、グローバー・アヴェニュー 4
 5
 (74) 代理人 100059959
 弁理士 中村 稔
 (74) 代理人 100067013
 弁理士 大塚 文昭
 (74) 代理人 100082005
 弁理士 熊倉 賢男
 (74) 代理人 100065189
 弁理士 宍戸 嘉一

最終頁に続く

(54) 【発明の名称】ユーザトラフィックフローを予測する方法および装置

(57) 【特許請求の範囲】

【請求項 1】

ユーザトラフィックフローを予測する方法であって、
識別手段によって、入力された複数のドキュメントを含むドキュメントコレクションにおいて、複数の内容項目(コンテントアイテム)を含んでいる複数のドキュメントのハイパームディアリンクを識別するトポロジーマトリックスを決定することと、
分布決定手段によって、前記ドキュメント内の前記内容項目の分布を決定することと、
重み付け手段によって、前記ドキュメント内の各内容項目の出現頻度に基づいて、内容項目に重み付けを行い、重みマトリックスを決定することと、

比較手段によって、入力されたユーザのニーズを示す情報ニーズ項目を含む問い合わせベクトルを前記重みマトリックスに乗じて、関連性ベクトルを決定することと、
アソシエーション・ストレングス決定手段によって、前記関連性ベクトルを前記トポロジーマトリックスに乗じて、前記ドキュメントに関連付けられたハイパームディアリンクのためのアソシエーション・ストレングス・マトリックスを決定することと、
適用手段によって、入力された初期条件にアソシエーション・ストレングス・マトリックスを適用して、ユーザトラフィックフローを予測することと、

から成ることを特徴とする方法。

【請求項 2】

トラフィックフローをシミュレートするために前記アソシエーション・ストレングス・マトリックスを使用するネットワーク・フローモデルを、シミュレーション手段によって

前記初期条件に適用す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 3】

前記ネットワーク・フロー モデルは spreading activation アルゴリズムであることを特徴とする請求項 2 に記載の方法。

【請求項 4】

前記シミュレーションは、所定のステップ数だけ続けられる請求項 2 に記載の方法。

【請求項 5】

ユーザトラフィックフローを予測するシステムであって、

入力された複数のドキュメントを含むドキュメントコレクションにおいて、複数の内容項目（コンテンツアイテム）を含んでいる複数のドキュメントのハイパーメディアリンクを識別するトポロジーマトリックスを決定する識別手段と、 10

前記ドキュメント内の前記内容項目の分布を決定する分布決定手段と、

前記ドキュメント内の各内容項目の出現頻度に基づいて、内容項目に重み付けを行い、重みマトリックスを決定する重み付け手段と、

入力されたユーザのニーズを示す情報ニーズ項目を含む問い合わせベクトルを前記重みマトリックスに乗じて、関連性ベクトルを決定する比較手段と、

前記関連性ベクトルを前記トポロジーマトリックスに乗じて、前記ドキュメントに関連付けられたハイパーメディアリンクのためのアソシエーション・ストレングス・マトリックスを決定するアソシエーション・ストレングス決定手段と、

入力された初期条件にアソシエーション・ストレングス・マトリックスを適用して、ユーザトラフィックフローを予測する適用手段と、 20

から成ることを特徴とするシステム。

【請求項 6】

トラフィックフローをシミュレートするために前記アソシエーション・ストレングス・マトリックスを使用するネットワーク・フロー モデルを前記初期状態に適用するシミュレーション手段をさらに含む請求項 5 に記載のシステム。

【発明の詳細な説明】

【0001】

【産業上の利用分野】

本発明は、ハイパーメディアリンクでリンクされたドキュメントのコレクションの分析と設計の分野、詳細には、観察された使用情報に頼らずに上記コレクション内のユーザトラフィックフローを予測する方法および装置に関するものである。 30

【0002】

【従来の技術】

World Wide Webのようなハイパーテキストリンクでリンクされたドキュメントのユーザは、一般にハイパーテキストリンクを選択し、ドキュメントからドキュメントへナビゲートして情報を探し回る。テキストの抜粋のような一片の情報は、一般に各ハイパーテキストリンクに対応づけられている。テキストの抜粋はリンクの他端でドキュメントの内容に関する情報をユーザに与える。リンクがユーザの情報ニーズに関係のあるドキュメントへユーザを導けば、ユーザはもう少しのところで彼の情報ニーズを満すので、ユーザが情報を探し回り続ける時間の長さは短くなる。しかし、もしリンクが関係のないドキュメントへユーザを導けば、ユーザは情報を探し回り続けるであろう。 40

【0003】

ハイパーメディアリンクでリンクされたドキュメントのコレクションの構造的連係トポロジーは、ハイウェイシステムと似ている。ハイウェイシステムでは、旅行者は一定の起点から出発して、希望する行き先へ達するためハイウェイシステムの道路に沿って移動する。途中で、旅行者は希望する行き先へ達するため通らなければならない道路を指示する標識を見ることができる。たとえば、自宅から地方空港へ行きたいと思っている旅行者は、「国際空港」の文字のある標識、あるいは航空機の絵のある標識を見るまで道路に沿って移動するであろう。どちらの標識も、空港に行くにはどのハイウェイランプを通るべきか 50

に関する情報を旅行者に与えることができよう。もし標識が存在しなかったり、不明確であれば、旅行者はおそらく行き先を見つけることができないであろう。

【0004】

同様に、ウェブ上のユーザは、あるウェブページから出発し、彼の情報ニーズを満たす別のウェブページへユーザを導くことができるよう見えるかどうかに基づいて、リンクを選択するであろう。リンクは彼の行き先、すなわち情報ニーズへユーザを導くことができる道路と似ている。それらのリンクがどのようにユーザを希望する行き先へうまく導くかは、ユーザの目的、ユーザの行動、およびウェブサイトの設計の複雑な対話によって決まる。

【0005】

ユーザがどのようにウェブと対話するかを知りたい設計者や研究者は、それらの複雑な対話について仮説を立てる。それらの仮説を迅速にかつ効率的に評価するため対話の複雑さを処理するツールを作り出す必要がある。それらの仮説を評価する既存の手法には、ウェブログファイル(Web log file)のような使用データから情報を抽出する方法と、ユニークユーザの数、ページ閲覧の数、読み出し時間、セッションリンク、およびユーザ経路のような計量(metrics)を適用する方法がある。それらの手法の信頼性の度合いは、使用したいいろいろな発見的手法(heuristics)によって大きく違ってくる。たとえば、たいていの既存のウェブルog・ファイル分析プログラムは、単にユーザがそこにいたという簡単な記述的統計データを提供するだけなので、ユーザとウェブの対話の見通しはほとんどない。

【0006】

【発明が解決しようとする課題】

既存手法の1つの欠点は、予測を実行するためユーザの過去の行動を収集する必要があることである。既存手法のもう1つの欠点は、ハイパーリンクでリンクされたドキュメントに含まれる内容を分析していないことである。したがって、予測を実行するためにユーザ対話情報を収集する必要がなく、かつドキュメントの内容を考慮に入れるやり方で、ハイパーメディアリンクでリンクされたドキュメントのコレクション内のユーザトラフィックフローを予測する装置および方法が求められている。

【0007】

【課題を解決するための手段】

本発明の実施例は、ハイパーメディアリンクの「アソシエーション・ストレングス(association strength: 関連付け又は連想の強さ)」を決定することによって、ハイパーメディア・ドキュメントのコレクション内のユーザトラフィックフローを予測する装置および方法を提供する。概念的には、「アソシエーション・ストレングス」は、ユーザが特定のハイパーメディアリンクをフローダウンする確率の尺度である。本発明の装置および方法は、ユーザがドキュメントの内容を考慮に入れるので、予測を実行するためにユーザの対話情報を収集する必要がない。本発明の実施例は、ユーザの情報ニーズとドキュメントに含まれている内容項目(コンテンツアイテム)に基づいて、ドキュメントコレクション内のハイパーメディアリンクのアソシエーション・ストレングスを決定する装置および方法を含んでいる。装置はユーザの情報ニーズと関連があるかもしれない内容項目を含む複数のドキュメントのコレクションの中からハイパーメディア連係構造を識別する。装置はドキュメント・コレクション内の内容項目の分布を決定する。装置は入力として情報項目を受け取り、その情報項目と内容項目を比較する。その比較に応答して、装置はハイパーメディアリンクにアソシエーション・ストレングスを割り当てる。装置は、さらに、ハイパーメディアリンクのアソシエーション・ストレングスを初期条件に適用することによってユーザトラフィックフローを予測するネットワークフローモデルを使用している。

【0008】

【発明の実施の形態】

本発明は、ハイパーメディアリンクのアソシエーション・ストレングスを決定することによって、ハイパーメディアドキュメントのコレクション内のユーザトラフィックフローを予測する装置及び方法を提供する。装置は、ドキュメントの内容と装置に入力されたユー

10

20

30

40

50

ザ情報ニーズ項目とがどのように関連しているかを分析するので、予測を実行するためにユーザの対話情報を収集する必要がない。

【0009】

図1は、ハイパーテディアリンクでリンクされたドキュメントのコレクションの構造的連係と内容(content)を示すブロック図である。ドキュメントP0、P1、P2、P3、P4、P5、P6は、それぞれ、102、104、106、108、110、112、114として索引が付けられている。図示のように、ドキュメントP0～P6は、ハイパーテディアリンク120, 122, 124, 126, 128, 130, 132によってリンクされている。ハイパーテディアリンクは、あるドキュメントから別のドキュメントへリンクされたハイパーテキストリンクを含む、どんな形式のものでもよい。P0～P6(102～114)内に示したドキュメントの種類の例はウェブサイトである。図示のように、内容項目144～154はドキュメントP0～P6内にある。これらのハイパーテディアリンクに対応づけられたドキュメントの内容は、通例、テキストの抜粋または図形のようなある種のプロキシマル(近称)キュー(proximal cue)によってユーザへ提示される。情報を探し回っているウェブのユーザはこれらのプロキシマルキューを用いて、リンクの他端でドキュメントの末端内容(distal content)を処理する。これらのハイパーテディアリンクのアソシエーション・ストレンジスは、「情報の手がかり」とも呼ばれており、ウェブリンクまたは内容ソースを表すアイコンのようなプロキシマルキューから得られる情報ソースの価値、コスト、あるいはアクセス経路の不完全な、主観的な知覚である。概念的には、アソシエーション・ストレンジスはユーザが特定のハイパーテディアリンクをフローダウンする確率の尺度である。10

【0010】

図2は、本発明の実施例のハイパーテディアリンクのアソシエーション・ストレンジスを決定する方法において実行されるステップを示すフローチャート200である。最初に、ステップ202において、複数のドキュメントP0～P6(102～114)のハイパーテディアリンク120～132を識別する。

【0011】

ハイパーテディアリンクを識別した後、ステップ204において、ドキュメントコレクション内の内容項目の分布を決定する。“Foundations of Statistical National Language Processing”, C. Manning and H. Schuetze, 1999, MIT Press, p. 542に述べられているように、標準情報検索手法(たとえばTF. IDF)によって内容項目の分布を決定して、ドキュメントコレクション内の出現頻度によって内容項目に重みを付けることができる。他のいろいろな加重法を使用することができる。30

【0012】

次に、ステップ206において、情報ニーズ項目とドキュメントコレクション内の内容項目とを比較する。情報ニーズ項目は、ユーザがそのドキュメントコレクションの中で見つけたいと思っている情報を示す。この情報ニーズ項目は、ユーザの情報ニーズに関連のあるキーワードのリストが入っている問合せとして表すことができる。キーワードのリストは同意語または関連語を含むように拡張することができる。

【0013】

ステップ208において、情報ニーズと内容項目との比較の結果と、内容項目の分布と一緒に用いて、ハイパーテディアリンクのアソシエーション・ストレンジスを決定する。

【0014】

図3は、図2で説明した方法のステップをより詳細に記述するフローチャート300である。最初に、ステップ302において、コレクション内のドキュメントに索引を付ける。たとえば、図1に示した7つのドキュメントをドキュメントP0～P6として索引を付ける。次にステップ304において、コレクション内のドキュメント間のハイパーテディアリンクを識別する。ハイパーテディアリンクとドキュメントの情報は、ハイパーテディアリンクでリンクされたドキュメントコレクションのトポロジーを表す(ドキュメント×ドキュメント)隣接マトリックスとして表すことができるグラフを作る。そのようなトポロ4050

ジー・マトリックス T の例を図 6 (A) に示す。マトリックス T の行と列は、ドキュメント ID (0 ~ 6) によって索引が付けられる。“ 1 ” はあるドキュメントから別のドキュメントへアウトリリンクが存在することを示す。たとえば、第 1 行、第 2 列にある“ 1 ” は、ドキュメント P 0 (1 0 2) からドキュメント P 1 (1 0 4) へアウトリリンク 1 2 0 があることを示す。同様に、トポロジー・マトリックス T (6 0 2) 内の他の“ 1 ” の値もアウトリリンクを示す。

【 0 0 1 5 】

ステップ 3 0 6において、図 6 (B) に示すように、ドキュメント内のユニーク内容項目に索引を付ける。たとえば、図 1 には、8 個のユニーク内容項目：“ Java ” 1 4 0 (ドキュメント P 1 、 P 2 、 P 3 、および P 5 に含まれている) 、“ API ” 1 4 2 (ドキュメント P 3 に含まれている) 、“ Sun ” 1 4 4 (ドキュメント P 0 に含まれている) 、“ Home ” 1 4 6 (ドキュメント P 0 と P 4 に含まれている) 、“ coffee ” 1 4 8 (ドキュメント P 5 に含まれている) 、“ support ” 1 5 0 (ドキュメント P 2 に含まれている) 、“ Petes ” 1 5 2 (ドキュメント P 4 に含まれている) 、“ Tea ” 1 5 4 (ドキュメント P 6 に含まれている) が存在する。これら 8 個の内容項目は、0: Java 、 1: API 、 2: Sun 、 3: Home 、 4: Coffee 、 5: Support 、 6: Petes 、 7: Tea のように索引が付けられる。図 6 (B) には、これらの索引が付けられた項目と、対応するユニーク内容項目番号と一緒に示してある。

【 0 0 1 6 】

次に、ステップ 3 0 8において、内容項目の出現頻度を反映させるため、ユニーク内容項目索引番号を使用して、ワード × ドキュメント・マトリックス W を生成する。図 6 (C) に示したマトリックス 6 0 6 はそのようなマトリックスの例である。マトリックス 6 0 6 は、ドキュメント P 0 ~ P 6 のコレクション内の内容項目の分布を反映しており、重みマトリックス W とも呼ばれる。たとえば、マトリックス 6 0 6 の行 1 は、列 2 , 3 , 4 , 6 内の“ 1 ” 値の配置によって、ワード “ Java ” がドキュメント P 1 、 P 2 、 P 3 、および P 5 に出現することを示す。図 2 の検討に関連して述べたように、標準情報検索手法 (たとえば、TF. IDF) によって内容項目の分布を決定して、ドキュメントコレクション内の出現頻度によって内容項目に重みを付けることができる。また、他のいろいろな加重法を使用することができる。

【 0 0 1 7 】

次にステップ 3 1 0において、ユーザ情報ニーズを内容項目のリストとして表すために問合せベクトルを生成する。問合せベクトル 6 1 0 の例を図 6 (D) に示す。問合せベクトル 6 1 0 の n 番目の成分は n 番目の索引付きワードに対応する。たとえば、問合せベクトル 6 1 0 は、ワード “ Java ” とワード “ API ” (情報ニーズを表すキーワード群の中の 2 つの項目である) を表す最初の 2 つの行に、“ 1 ” の値をもっている。情報ニーズを表すため使用できるキーワードのリストは、同意語または関連語を含めるように拡張することができる。

【 0 0 1 8 】

次にステップ 3 1 2において、ユーザ情報ニーズにおおよそ一致するドキュメントのリストを得るために、次式に従って、問合せベクトル Q (6 1 0) に重みマトリックス W (6 0 6) を乗じる。

$$R = W^T Q \quad (1)$$

【 0 0 1 9 】

ドキュメント P 0 ~ P 6 のリストは関連性 (relevance (適合性)) ベクトル R として表現される。関連性ベクトルに含まれている値は、どのドキュメントが最も関連性があるとみなされるかを示す。たとえば、問合せベクトル Q (6 1 0) と重みマトリックス W (6 0 6) を使用した場合、得られる関連性ベクトル R = [0 1 1 2 0 1 0] である。この例の場合、7 つのドキュメント P 0 ~ P 6 が存在し、4 番目のドキュメント (P 3 に対応する) は関連性値 2 を有し、これは他の値より高いので、最も関連性が高い。次にステップ 3 1 4において、関連性ベクトル R にトポロジー・マトリックス T (6 0 2)

10

20

30

40

50

を乗じて、コレクション内のドキュメントを結んでいる各リンクのアソシエーション・ストレンジスを決定する。

【0020】

この例では、アソシエーション・ストレンジスは以下のように決定される。関連性ベクトル R 内の項目の指数は実際上および理論上の理由（たとえばLuceの選択理論）で選ばれる。もしTF-IDFの加重法を早期に使用したならば、これは必要ないことに留意されたい。その結果は、関連性ベクトル $R^T = [1 \ 2.718 \ 2.718 \ 7.389 \ 1 \ 2.718 \ 1]$ である。この関連性ベクトル R はトポロジー・マトリックスの各行 T_i に適用される。トポロジー・マトリックスの各行 T_i は1つのドキュメントに対応しており、そのドキュメントから出てくるリンクを指定する。言い替えると、ノードから出てくるハイパーテディアリンクはトポロジー・マトリックス T の行によって表される。ドキュメント P_1 の場合のアウトリンクは、トポロジー・マトリックス T の行と列を入れ替え、図7(A)に $T^T i$ として示した第1列 702 を選ぶことによって得られる。関連性ベクトル内の対応する要素 R_i が乗じられたこの行からの要素 T_{ij} は、アソシエーション・ストレンジスすなわち「手がかり」 S_{ij} を与える。 T の各行を取り出し、1要素づつ関連性ベクトル R を乗じることによって、アソシエーション・ストレンジス・ベクトル S_i^* が得られる。たとえば、もし関連性ベクトル R を要素ごとにアウトリンク・ベクトルに乗じれば、 P_1 にリンクされるドキュメントの関連性が得られる。すなわち、

$$X = T^T i R \quad X^T = [0 \ 0 \ 2.718 \ 7.389 \ 0 \ 2.718 \ 0] \quad (2)$$

【0021】

すべてのリンクの合計が1になるように、アソシエーション・ストレンジスを比例化することができる。アソシエーション・ストレンジスを比例化する理由は、リンクを通って流れるユーザの量を一定のままにすべきであるからである。たとえば、図7(B)に示すように、各アウトリンクのアソシエーション・ストレンジス S_i (704) はすべてのアウトリンクからのアソシエーション・ストレンジスの和に比例すべきである。全てのドキュメントとアウトリンクについて上記の計算を繰返すことによって、完全なドキュメント×ドキュメント・アソシエーション・ストレンジス・マトリックス S (706) を生成することができる。図7(C)を参照されたい。

【0022】

得られたアソシエーション・ストレンジス・マトリックス S (706) は、その後、ドキュメントコレクションの異なるリンク、すなわち任意のハイパーテディアリンクをフローダウンするユーザをシミュレートするのに使用することができる。アソシエーション・ストレンジス・マトリックス S (706) は、各ハイパーテディアリンクに対応づけられたアソシエーション・ストレンジスのネットワークを指定し、かつ、各リンクをフローダウンするユーザ同業者の総数を示す。

【0023】

ハイパーテディアリンクのアソシエーション・ストレンジスを決定した後、トラフィックフローを予測する方法を使用することができる。図4は、本発明の実施例に係るトラフィックフローを予測する方法のフローチャート400を示す。ステップ402において、本方法に使用する初期条件を決定する。一般に、初期条件は入口点たとえばドキュメントまたはウェブページを指定する。ユーザはその入口点から彼の情報ニーズを満たすためリンクの選択を開始する。それらの初期条件は同じ4つの別々の入口点から出発する複数のユーザを含むことがある。初期条件は入口ベクトル E として表すことができる。たとえば、入口ベクトル $E = [0 \ 1 \ 0 \ 0 \ 0 \ 0 \ 0]$ は、初期条件がドキュメント P_1 であることを指定する。次にステップ404において、ハイパーテディアリンクのアソシエーション・ストレンジスを得る。次にステップ406において、そのアソシエーション・ストレンジスを用いて、ネットワークフローモデルを初期条件に適用する。前記ステップ406において、任意の伝統的なネットワークフローモデルを使用することができる。たとえば、米国特許第5,835,905号（発明の名称“System for Predicting Documents Relevant to Focus Documents by Spreading Activation Through Network Representations”）

10

20

30

40

50

tions of a Linked Collection of Documents")に論じられているように、Spreading Activationアルゴリズムを使用することができる。

【0024】

たとえば、Spreading Activationアルゴリズムを初期条件に適用することによって、入口ベクトル E から $A(1) = E^T = [0 \ 1 \ 0 \ 0 \ 0 \ 0]$ が得られる。前記 $A(1)$ をアソシエーション・ストレンジス・マトリックス $S(706)$ によってポンプして、図7(D)に示すように、 $A(2)708$ を得ることができる。ステップ406においてネットワークフロー・モデルを適用した後、ステップ408において、シミュレーションを続けるかどうかに関して決定をする。この決定は、所定のステップ数、ハイパー・メディアリンクを選択し続けるユーザの割合、あるいはユーザの総数を所定のしきい値と比較することを含む、多くの要因を基礎にすることができる。ハイパー・メディアリンクを選択し続けるユーザの割合は、P. Pirolli and J. E. Pitkow, "Distributions of surfers' paths through the World Wide Web; Empirical characterization", 1999, World Wide Web(2): pp. 29-45、およびHuberman, B. A., P. Pirolli, J. Pitkow, R. Lukose, "Strong regularities in World Wide Web surfing", 1999, Science 280: pp. 95-97に記載されているように、「サーフィングの法則」として知られる関数 (L) によって決定することができる。たとえば、Spreading Activationは、式 $A(t) = SA(t-1)$ に従って、多数の繰返し $t = 1 \dots n$ を重ねることができる。

10

【0025】

もしステップ408の決定がシミュレーションを続けるべきであると指示すれば、処理はステップ406においてもう1回繰返しを続ける。もしシミュレーションを続けるべきでないと指示すれば、シミュレーションは完全であり、ステップ410において終了する。上に述べたSpreading Activationの例の場合、シミュレーションの結果は、 $A(n) = SA(n-1)$ である。

20

【0026】

図5は、本発明の実施例においてハイパー・メディアリンクのアソシエーション・ストレンジスを決定する装置のブロック図である。装置500は識別構成要素502を備えている。識別構成要素502は、ハイパー・メディアリンクでリンクされたドキュメントコレクション入力508に応答し、複数のドキュメントのハイパー・メディアリンクを識別する。識別構成要素502は図2のステップ202と図3のステップ304を実行するのに使用することができる。識別されたリンク510は、ドキュメント内の内容項目の分布を決定し、一組の重み付き内容項目512を与える分布構成要素504による作用を受ける。分布構成要素504はステップ204とステップ308を実行するのに使用することができる。重み付き内容項目512は、その後、比較構成要素506によって情報ニーズ項目514と比較され、ハイパー・メディアリンクに対応づけられたアソシエーション・ストレンジス516が生成される。比較構成要素506はステップ206とステップ312～314を実行するのに使用することができる。ドキュメントは上に述べたように複数の内容項目を含んでいる。アソシエーション・ストレンジス516はユーザトラフィックフローをシミュレートする装置において使用することができる。

30

【0027】

本発明の実施例においてハイパー・メディアリンクでリンクされた複数のドキュメントにおけるトラフィックフローをシミュレートする装置は、アソシエーション・ストレンジス入力516と初期条件520に応答して、予測ユーザ経路522を生成するシミュレーション構成要素518を備えている。シミュレーション構成要素518はステップ406～410を実行するのに使用することができる。シミュレーション構成要素518はネットワーク・フロー・モデルを初期条件に適用する。ネットワーク・フロー・モデルは情報ニーズ514を表す複数のアソシエーション・ストレンジス516に応答して、トラフィックフローをシミュレートする。初期条件はシミュレーションの開始時における一人またはそれ以上のユーザの入口点または状態を表す。たとえば、初期条件は入口ウェブページであってもよい。もし多数のユーザをモデル化することになっていれば、初期条件は複数のユーザ

40

50

の入口点であってもよい。

【0028】

ネットワーク・フロー モデルは、任意の適当な伝統的なネットワーク・フロー モデル、たとえば前に述べた spreading activation アルゴリズムであってもよい。シミュレーション構成要素は連続して繰返し動作し、特定の条件が満たされると停止する。たとえば、シミュレーションは所定の数のステップすなわちリンクトラバーサル(link traversal)にわたって続くことがある。シミュレーションは、続行するユーザの割合が前に述べた閾数(L)によって決定される望ましいレベルを越えるまで、またはそのレベルに達するまで、またはそのレベル以下に落ちるまで続けることができる。代わりに、続行するユーザの総数が所定のしきい値以下になるまでシミュレーションを続けることができる。

10

【図面の簡単な説明】

【図1】ハイパーメディアリンクでリンクされたドキュメントのコレクションの構造的な連係および内容を示すブロック図である。

【図2】本発明の実施例においてハイパーメディアリンクのアソシエーション・ストレンジスを決定する方法の中で実行されるステップを示すフローチャートである。

【図3】本発明の実施例においてハイパーメディアリンクのアソシエーション・ストレンジスを決定する方法の中で実行されるステップを示すフローチャートである。

【図4】本発明の実施例においてトラフィックフローを予測する方法の中で実行されるステップを示すフローチャートである。

【図5】本発明の実施例においてユーザトラフィックフローを予測する装置を示すプロック図である。

20

【図6】本発明の実施例において使用される典型的なマトリックスである。

【図7】本発明の実施例において使用される典型的なマトリックスである。

【符号の説明】

P 0 ~ P 6 ドキュメント

1 0 2 ~ 1 1 4 索引付きドキュメント

1 2 0 ~ 1 3 2 ハイパーメディアリンク

1 1 4 ~ 1 5 4 内容項目

5 0 0 アソシエーション・ストレンジスを決定する装置

5 0 2 識別構成要素

30

5 0 4 分布構成要素

5 0 6 比較構成要素

5 0 8 ドキュメントコレクション入力

5 1 0 識別されたリンク

5 1 2 一組の重み付き内容項目

5 1 4 情報ニーズ項目

5 1 6 アソシエーション・ストレンジス

5 1 8 シミュレーション構成要素

5 2 0 初期条件

5 2 2 予測ユーザ経路

40

6 0 2 トポロジーマトリックス T

6 0 6 重みマトリックス W

6 0 8 索引付きドキュメント内のユニーク内容項目

6 1 0 問合せベクトル Q

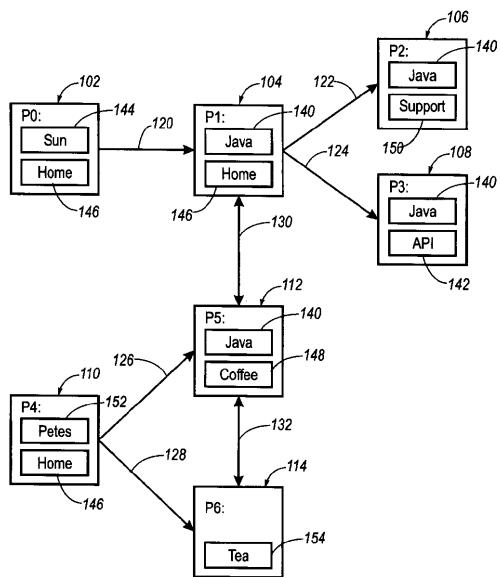
7 0 2 トポロジー・マトリックスの第 1 列

7 0 4 各アウトリンクのアソシエーション・ストレンジス S_i

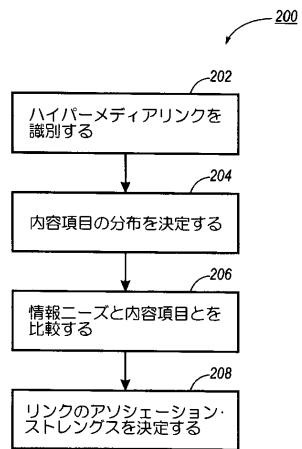
7 0 6 アソシエーション・ストレンジス・マトリックス S

7 0 8 A(1) をマトリックス S を通してポンプして得られた A(2)

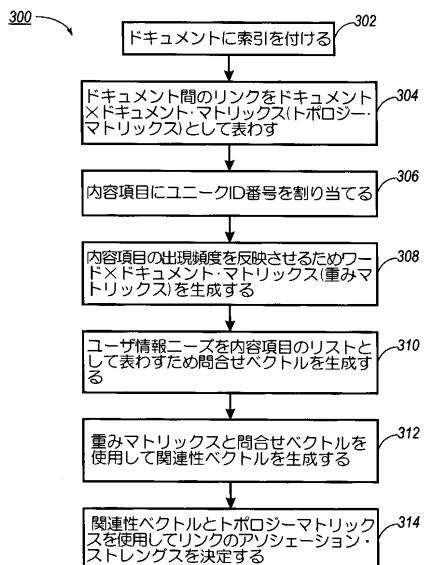
【図1】



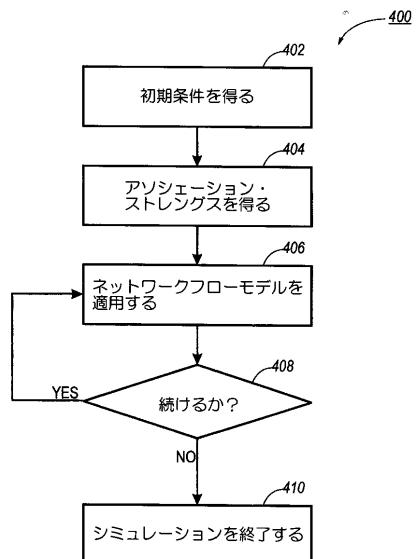
【図2】



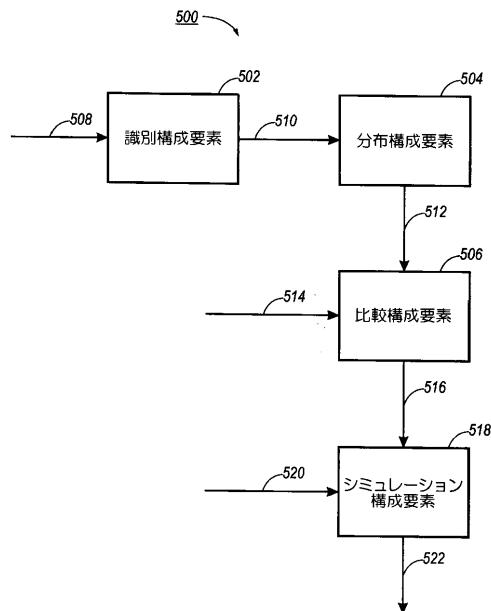
【図3】



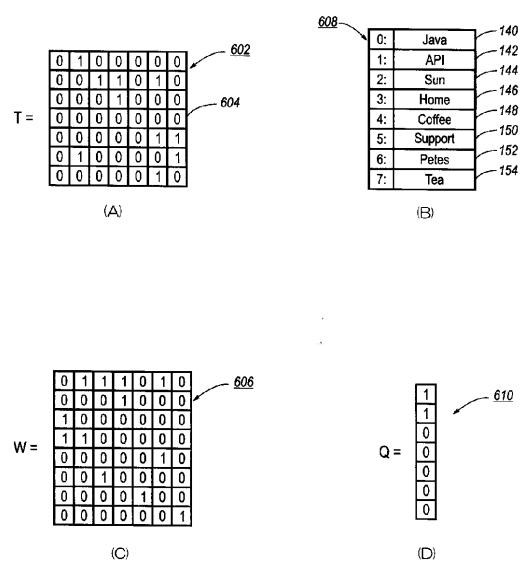
【図4】



【図5】



【図6】



【図7】

$$T_i^T = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \\ 1 \\ 0 \end{bmatrix} \quad S_i = \frac{X}{\sum X} = \begin{bmatrix} 0 \\ 0 \\ 0.212 \\ 0.576 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

(A) (B)

$$S = \begin{array}{|c|c|c|c|c|c|c|} \hline 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \hline 1 & 0 & 0 & 0 & 0 & 0.731 & 0 \\ \hline 0 & 0.212 & 0 & 0 & 0 & 0 & 0 \\ \hline 0 & 0.576 & 1 & 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \hline 0 & 0.212 & 0 & 0 & 0 & 0 & 0 \\ \hline 0 & 1 & 0 & 0.731 & 0 & 0 & 1 \\ \hline 0 & 1 & 0 & 0.269 & 0 & 0.269 & 0 \\ \hline \end{array}$$

(C)

$$A(2) = S A(1) = \begin{bmatrix} 0 \\ 0 \\ 0.212 \\ 0.576 \\ 0 \\ 0.212 \\ 1 \\ 1 \end{bmatrix}$$

(D)

フロントページの続き

(74)代理人 100096194
弁理士 竹内 英人

(74)代理人 100074228
弁理士 今城 俊夫

(74)代理人 100084009
弁理士 小川 信夫

(74)代理人 100082821
弁理士 村社 厚夫

(74)代理人 100086771
弁理士 西島 孝喜

(74)代理人 100084663
弁理士 箱田 篤

(72)発明者 ピーター エル ピロリー
アメリカ合衆国 カリフォルニア州 94116 サンフランシスコ スロート ブールヴァード
2958

(72)発明者 エド エイチ チー
アメリカ合衆国 カリフォルニア州 94301 パロ アルト ユニヴァーシティ アベニュー
488 - #412

(72)発明者 ジェイムズ イー ピットコウ
アメリカ合衆国 カリフォルニア州 94306 パロ アルト エルズワース プレイス 74
2

審査官 工藤 嘉晃

(56)参考文献 特開平10-027125(JP,A)

佐野 綾一 他,部分グラフを基本単位としたWeb文書検索:単語の出現密度分布の適用,情報処理学会研究報告,日本,社団法人情報処理学会,1999年7月23日,第99巻第61号,第79頁~第84頁

(58)調査した分野(Int.Cl., DB名)

G06F 19/00
JSTPlus(JDreamII)