

(19) 日本国特許庁(JP)

(12) 特許公報(B2)

(11) 特許番号

特許第5133953号
(P5133953)

(45) 発行日 平成25年1月30日(2013.1.30)

(24) 登録日 平成24年11月16日(2012.11.16)

(51) Int.Cl. F I
G06F 17/30 (2006.01) G06F 17/30 350C
 G06F 17/30 419B

請求項の数 4 (全 15 頁)

(21) 出願番号	特願2009-212380 (P2009-212380)	(73) 特許権者	502192546 清華大学 中華人民共和国北京市海淀区清華大学 郵 編 100084
(22) 出願日	平成21年9月14日(2009.9.14)	(73) 特許権者	000155469 株式会社野村総合研究所 東京都千代田区丸の内一丁目6番5号
(65) 公開番号	特開2011-60228 (P2011-60228A)	(74) 代理人	100105924 弁理士 森下 賢樹
(43) 公開日	平成23年3月24日(2011.3.24)	(72) 発明者	李 春平 中華人民共和国北京市海淀区清華大学内
審査請求日	平成24年2月9日(2012.2.9)	(72) 発明者	高 松 中華人民共和国北京市海淀区清華大学内
(31) 優先権主張番号	200910092356.1	(72) 発明者	王 益▲びん▼ 中華人民共和国北京市海淀区清華大学内
(32) 優先日	平成21年9月11日(2009.9.11)		最終頁に続く
(33) 優先権主張国	中国 (CN)		

(54) 【発明の名称】 情報伝播を検知するウェブページ関係評価装置

(57) 【特許請求の範囲】

【請求項 1】

複数の判断対象のウェブページそれぞれを構成する情報を取得する取得部と、
 前記取得部によって取得された情報に基づいて、複数の判断対象のウェブページそれ
 ぞれの作成時刻を検出する作成時刻検出部と、

前記取得部によって取得された情報に基づいて、複数の判断対象のウェブページそれ
 ぞれについて、他の判断対象のウェブページそれぞれとの間の類似度を算出する類似度算出
 部と、

前記類似度算出部によって得られた複数の類似度のうちの、2個のウェブページ相互が
 関連していると判断するための基準値以上である類似度それぞれについて、その類似度を
 算出する際に用いられた2個の判断対象のウェブページを、相互に関連しているウェブペ
 ージであると特定する関連ウェブページ特定部と、

前記作成時刻検出部によって検出された作成時刻に基づいて、前記類似度算出部によっ
 て得られた各類似度について、その類似度を算出する際に用いられた2個の判断対象のウ
 ェブページの作成時刻の間に作成された他の判断対象のウェブページの数に応じた補正係
 数をその類似度に乗算して補正類似度を算出する類似度補正部を備え、

前記関連ウェブページ特定部は、前記類似度補正部によって得られた複数の補正類似度
 のうちの前記基準値以上である補正類似度それぞれについて、その補正類似度を算出する
 際に用いられた2個の判断対象のウェブページを、相互に関連しているウェブページであ
 ると特定するウェブページ関係評価装置。

【請求項2】

複数の判断対象のウェブページそれぞれを構成する情報を取得する取得部と、
前記取得部によって取得された情報に基づいて、複数の判断対象のウェブページそれぞれの作成時刻を検出する作成時刻検出部と、

前記取得部によって取得された情報に基づいて、複数の判断対象のウェブページそれぞれについて、他の判断対象のウェブページそれぞれとの間の類似度を算出する類似度算出部と、

前記類似度算出部によって得られた複数の類似度のうちの、2個のウェブページ相互が関連していると判断するための基準値以上である類似度それぞれについて、その類似度を算出する際に用いられた2個の判断対象のウェブページを、相互に関連しているウェブページであると特定する関連ウェブページ特定部と、

前記類似度算出部によって得られた各類似度を補正する類似度補正部を備え、
前記取得部は、複数の判断対象のウェブページそれぞれの単位期間あたりのアクセス数を取得し、

前記類似度補正部は、前記取得部によって取得された複数の判断対象のウェブページそれぞれの単位期間あたりのアクセス数に基づいて、前記類似度算出部によって得られた各類似度について、その類似度を算出する際に用いられた2個の判断対象のウェブページのうちの先に作成されたウェブページの単位期間あたりのアクセス数に応じた補正係数をその類似度に乗算して補正類似度を算出し、

前記関連ウェブページ特定部は、前記類似度補正部によって得られた複数の補正類似度のうちの前記基準値以上である補正類似度それぞれについて、その補正類似度を算出する際に用いられた2個の判断対象のウェブページを、相互に関連しているウェブページであると特定するウェブページ関係評価装置。

【請求項3】

複数の判断対象のウェブページそれぞれを構成する情報を取得する機能と、
取得した情報に基づいて、複数の判断対象のウェブページそれぞれの作成時刻を検出する機能と、

取得した情報に基づいて、複数の判断対象のウェブページそれぞれについて、他の判断対象のウェブページそれぞれとの間の類似度を算出する機能と、

算出した複数の類似度のうちの、2個のウェブページ相互が関連していると判断するための基準値以上である類似度それぞれについて、その類似度を算出する際に用いられた2個の判断対象のウェブページを、相互に関連しているウェブページであると特定する機能と、

検出された作成時刻に基づいて、得られた各類似度について、その類似度を算出する際に用いられた2個の判断対象のウェブページの作成時刻の間に作成された他の判断対象のウェブページの数に応じた補正係数をその類似度に乗算して補正類似度を算出する機能を備え、

前記相互に関連しているウェブページであると特定する機能は、得られた複数の補正類似度のうちの前記基準値以上である補正類似度それぞれについて、その補正類似度を算出する際に用いられた2個の判断対象のウェブページを、相互に関連しているウェブページであると特定することをコンピュータに実現させるためのプログラム。

【請求項4】

複数の判断対象のウェブページそれぞれを構成する情報を取得する機能と、
取得した情報に基づいて、複数の判断対象のウェブページそれぞれの作成時刻を検出する機能と、

取得した情報に基づいて、複数の判断対象のウェブページそれぞれについて、他の判断対象のウェブページそれぞれとの間の類似度を算出する機能と、

算出した複数の類似度のうちの、2個のウェブページ相互が関連していると判断するための基準値以上である類似度それぞれについて、その類似度を算出する際に用いられた2個の判断対象のウェブページを、相互に関連しているウェブページであると特定する機能

10

20

30

40

50

と、

得られた各類似度を補正する機能を備え、

前記取得する機能は、複数の判断対象のウェブページそれぞれの単位期間あたりのアクセス数を取得し、

前記各類似度を補正する機能は、取得された複数の判断対象のウェブページそれぞれの単位期間あたりのアクセス数に基づいて、得られた各類似度について、その類似度を算出する際に用いられた2個の判断対象のウェブページのうちの先に作成されたウェブページの単位期間あたりのアクセス数に応じた補正係数をその類似度に乗算して補正類似度を算出し、

前記相互に関連しているウェブページであると特定する機能は、得られた複数の補正類似度のうちの前記基準値以上である補正類似度それぞれについて、その補正類似度を算出する際に用いられた2個の判断対象のウェブページを、相互に関連しているウェブページであると特定することをコンピュータに実現させるためのプログラム。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、複数のウェブページ相互の関係を評価することで、情報伝播を検知する技術に関する。

【背景技術】

【0002】

近年、企業は、ウェブページを用いて情報を発信しており、個人も、ウェブページ、ブログ、及びSNS(Social Network Service)を用いて情報を発信している。各情報は、他の情報の影響を受けて発信されることがある。例えば、ある企業が新製品を発売することをウェブページにおいて発表すると、そのウェブページを見たある個人がその情報とともにその新製品についての意見をブログにより公表することがある。このように、ある情報がウェブページを伝播して広がることがある。

【先行技術文献】

【特許文献】

【0003】

【特許文献1】特開2008-130044号公報

【発明の概要】

【発明が解決しようとする課題】

【0004】

ところで、ウェブページにおいて公表された情報の伝播を不快に感じることもある。例えば、不祥事が発生していないにもかかわらず、企業を中傷する情報が多数のウェブページに広がると、多数の人がその情報を知ってしまい、その企業の経済活動を阻害する可能性があるからである。そのような場合、その企業に属する人は、その情報の伝播経路を知りたいと考える。例えば、伝搬経路を構成している中核的なウェブサイトや大元の情報発信源がどこかを知ることができれば、定期的にそのサイトを読み、不利益な情報が流れていることをいち早く察知して早期に対策をとることができる。

【0005】

また、広告効果を狙ってウェブページ等で情報公開をすることがある。例えば企業が新製品を発売した事実を、自社ウェブサイトプレスリリースを掲載して発表した場合、ニュースサイトや著名なブログサイト等を経由して、一般の人々はその事実を目にする。このような場合、どのような経路を経て情報が伝播したかを、その企業に属する人は知りたいと考える。例えば、より多くの人に直接的に情報を伝播させたり、他の著名ブログに影響を与えて間接的に情報を伝播させたりするニュースサイトや著名ブログを特定することができれば、それらのサイトにフォーカスをあててより効果的な広告を行うことが出来る。

【0006】

10

20

30

40

50

しかしながら、現在、ウェブページにより広がった情報の伝播経路を精度よく特定する技術は存在しない。つまり、関連するウェブページ相互の関係を精度よく特定する技術は存在しない。

【0007】

本発明は、関連するウェブページ相互の関係を精度よく特定するウェブページ関係評価装置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】

【0008】

上記課題を解決し上記目的を達成するために、本発明のウェブページ関係評価装置は、複数の判断対象のウェブページそれぞれを構成する情報を取得する取得部と、前記取得部によって取得された情報に基づいて、複数の判断対象のウェブページそれぞれの作成時刻を検出する作成時刻検出部と、前記取得部によって取得された情報に基づいて、複数の判断対象のウェブページそれぞれについて、他の判断対象のウェブページそれぞれとの間の類似度を算出する類似度算出部と、前記類似度算出部によって得られた複数の類似度のうちの、2個のウェブページ相互が関連していると判断するための基準値以上である類似度それぞれについて、その類似度を算出する際に用いられた2個の判断対象のウェブページを、相互に関連しているウェブページであると特定する関連ウェブページ特定部とを備える。

10

【0009】

また、本発明のウェブページ関係評価装置の各構成要件の機能をコンピュータに実現させるためのプログラムも、本発明の一態様である。

20

【発明の効果】

【0010】

本発明は、関連するウェブページ相互を精度よく特定するウェブページ関係評価装置を提供することができる。

【図面の簡単な説明】

【0011】

【図1】本実施の形態のウェブページ分析システムの構成図である。

【図2】本実施の形態のウェブページ分析装置の構成図である。

【図3】本実施の形態のウェブページ分析システムの動作の概要を説明するための図である。

30

【図4】図2の検索キーワード作成装置が行う検索キーワードを作成する動作の各ステップを示すフローチャートである。

【図5】図2のウェブページ関係評価装置が行う複数のウェブページ相互の関係を評価する動作の各ステップを示すフローチャートである。

【図6(A)】作成時刻の順に並べられた4個の判断対象のウェブページと、各判断対象のウェブページ間の類似度とを示す図である。

【図6(B)】図6(A)の4個の判断対象のウェブページの関連しているウェブページ相互を線で結ぶことにより作成されたネットワーク図である。

【図7】図2のサイト評価装置が行うウェブサイトを評価する動作の各ステップを示すフローチャートである。

40

【図8】ウェブサイトを説明するための図である。

【発明を実施するための形態】

【0012】

以下に、本発明を実施するための形態を図面を参照して説明する。

【0013】

先ず、本実施の形態のウェブページ分析システムの構成を図1を用いて説明する。図1は、本実施の形態のウェブページ分析システムの構成図である。本実施の形態のウェブページ分析システムは、ウェブページを分析するシステムであって、図1に示すように、ウェブページ分析装置100と、クライアント端末装置200と、検索装置300と、通信

50

ネットワーク４００と、サイトアクセス数推定装置５００とを有する。

【００１４】

ウェブページ分析装置１００は、通信ネットワーク４００において開示されているウェブページを分析する装置である。ウェブページ分析装置１００の構成の詳細は図２を用いて後述する。クライアント端末装置２００は、ユーザによって使用される装置であって、検索キーワードや、検索キーワードに関連するウェブページを検索させる指示（以下、「検索指示」という。）等をユーザから受け付ける。クライアント端末装置２００は、受け付けた指示等をウェブページ分析装置１００又は検索装置３００に送信し、ウェブページ分析装置１００及び検索装置３００からそれらによって得られた結果を受信する。図１に示すように、クライアント端末装置２００には、表示装置２５０が接続されており、クライアント端末装置２００が受信する情報は、表示装置２５０によって表示される。

10

【００１５】

検索装置３００は、通信ネットワーク４００において開示されているウェブページのなかから、検索キーワードに関連するウェブページを検索する装置である。通信ネットワーク４００は、インターネット等のデータを通信するためのネットワークである。サイトアクセス数推定装置５００は、ウェブサイトのアクセス数を推定する装置である。ウェブページ分析装置１００、クライアント端末装置２００、検索装置３００、及びサイトアクセス数推定装置５００は、通信ネットワーク４００に接続されており、相互に通信可能である。

【００１６】

次に、ウェブページ分析装置１００の構成を図２を用いて説明する。図２は、ウェブページ分析装置１００の構成図である。ウェブページ分析装置１００は、上述したように、通信ネットワーク４００において開示されているウェブページを分析する装置であって、図２に示すように、通信装置１０と、検索キーワード作成装置２０と、ウェブページ関係評価装置３０と、サイト評価装置４０とを有する。

20

【００１７】

通信装置１０は、データを通信する。検索キーワード作成装置２０は、ユーザによって選択されたキーワードに基づいて、適宜の変換処理を施したより適切な検索キーワードを作成する。

【００１８】

ウェブページ関係評価装置３０は、検索キーワード作成装置２０が作成した検索キーワードに基づいて検索装置３００によって検索された複数のウェブページ相互の関係を評価する装置である。ウェブページ関係評価装置３０は、図２に示すように、取得部１と、作成時刻検出部２と、類似度算出部３と、類似度補正部４と、関連ウェブページ特定部５と、ネットワーク図作成部６と、出力部７とを有する。

30

【００１９】

取得部１は、検索装置３００によって検索された複数のウェブページそれぞれを構成する情報を取得する。検索装置３００によって検索されたウェブページは、ウェブページ関係評価装置３０によって評価されるウェブページであるので、以下では、検索装置３００によって検索されたウェブページを「判断対象のウェブページ」という。作成時刻検出部２は、取得部１によって取得された情報に基づいて、複数の判断対象のウェブページそれぞれの作成時刻を検出する。

40

【００２０】

類似度算出部３は、取得部１によって取得された情報に基づいて、複数の判断対象のウェブページそれぞれについて、他の判断対象のウェブページそれぞれとの間の類似度を算出する。具体的には、類似度算出部３は、ある２個の判断対象のウェブページの類似度を算出する際、それら２個の判断対象のウェブページそれぞれを構成する情報の形態素を分析し、TF-IDF (Term Frequency - Inverse Document Frequency) 法及びベクトル空間法を用いて、それら２個の判断対象のウェブページの類似度を算出する。なお本方法以外にも、ウェブページ類似度の算出には一般

50

に広く用いられている文書類似度算出法のいずれを使ってもよく、たとえばTF-IDF法のかわりにN-Gram法を用いても良い。

【0021】

類似度補正部4は、類似度算出部3によって算出された類似度を補正する。以下では、類似度補正部4によって得られた値を「補正類似度」という。なお、類似度補正部4は、複数の方法により類似度を補正するので、各方法については後に類似度補正部4の動作を説明する際に説明する。

【0022】

関連ウェブページ特定部5は、類似度補正部4によって得られた複数の補正類似度のうちの、2個のウェブページ相互が関連していると判断するための基準値以上である補正類似度を検出する。そして、関連ウェブページ特定部5は、検出した各補正類似度について、その補正類似度を算出する際に用いられた2個の判断対象のウェブページを、相互に関連しているウェブページであると特定する。

10

【0023】

ネットワーク図作成部6は、関連ウェブページ特定部5によって特定された2個のウェブページの組それぞれについて、その組を構成する2個のウェブページを、作成時刻検出部2によって検出された作成時刻の順に仮想的に並べて仮想的に線で結び、関連しているウェブページのネットワーク図を作成する。つまり、ネットワーク図作成部6は、関連しているウェブページを作成時刻の順にツリー状につなげてネットワーク図を作成する。

【0024】

出力部7は、ネットワーク図作成部6によって作成されたネットワーク図を通信装置10に出力する。

20

【0025】

サイト評価装置40は、通信ネットワーク400において開示されているウェブサイトを評価する装置である。

【0026】

次に、本実施の形態のウェブページ分析システムの動作を説明する。図3は、本実施の形態のウェブページ分析システムの動作の概要を説明するための図である。本実施の形態のウェブページ分析システムの動作は、図3に示すように、絞り込み検索を行わせるための検索キーワードを作成する動作(フェーズF1)と、複数のウェブページ相互の関係を評価する動作(フェーズF2)と、ウェブサイトを評価する動作(フェーズF3)とに大別される。そのため以下では、本実施の形態のウェブページ分析システムの動作をフェーズ毎に説明する。

30

【0027】

(検索キーワードを作成する動作)

まず、絞り込み検索を行わせるための検索キーワードを作成する動作(フェーズF1)を説明する。

【0028】

ユーザは、検索キーワードと、その検索キーワードに関連するウェブページを検索させる指示(検索指示)をクライアント端末装置200に入力する。クライアント端末装置200は、入力された検索キーワード及び検索指示を受け付け、それらを通信ネットワーク400を介して検索装置300に送信する。検索装置300は、クライアント端末装置200からの検索キーワード及び検索指示を受信し、通信ネットワーク400において開示されているウェブページのなかから、受信した検索キーワードに関連するウェブページを検索する。説明の便宜上、検索装置300は、複数のウェブページを検索したと仮定する。検索装置300は、検索した複数のウェブページのタイトルを通信ネットワーク400を介してクライアント端末装置200に送信する。

40

【0029】

クライアント端末装置200は、検索装置300からの複数のウェブページのタイトル及び本文内容概要を受信し、それらを検索装置300によって決められた順にしたがって

50

表示装置 250 に表示させる。ユーザは、表示装置 250 によって表示された各ウェブページのタイトルを見て、いずれかのウェブページを選択し、選択したウェブページを取得させる指示をクライアント端末装置 200 に行う。

【0030】

クライアント端末装置 200 は、ユーザによって選択されたウェブページを取得して表示装置 250 に表示させる。ユーザは、表示装置 250 によって表示されたウェブページがより適切な検索キーワードを作成させる題材として妥当なウェブページであると判断した場合、その判断結果をクライアント端末装置 200 に入力する。

【0031】

クライアント端末装置 200 は、そのようなウェブページを特定する情報を、通信ネットワーク 400 を介してウェブページ分析装置 100 に送信する。ウェブページを特定する情報は、例えばそのウェブページの URL (Uniform Resource Locator) である。上述したようにして、ユーザは、より適切な検索キーワードを作成させるためのウェブページを複数個選択し、クライアント端末装置 200 は、ユーザによって選択された複数のウェブページを特定する情報を受け付けてウェブページ分析装置 100 に送信する。また、クライアント端末装置 200 は、ユーザによって入力された検索キーワードもウェブページ分析装置 100 に送信する。

10

【0032】

ウェブページ分析装置 100 では、通信装置 10 が、クライアント端末装置 200 からの、より適切な検索キーワードを作成させるための複数のウェブページを特定する情報と、ユーザによって入力された検索キーワードとを受信する。

20

【0033】

以下の動作は、図 4 を用いて説明する。図 4 は、検索キーワード作成装置 20 が行う検索キーワードを作成する動作の各ステップを示すフローチャートである。

【0034】

検索キーワード作成装置 20 は、通信装置 10 によって受信された、より適切な検索キーワードを作成させるための複数のウェブページを特定する情報と、ユーザによって入力された検索キーワードとを取得する (S1)。次に、検索キーワード作成装置 20 は、より適切な検索キーワードを作成させるための複数のウェブページそれぞれを構成する情報を、通信ネットワーク 400 を介して取得する (S2)。

30

【0035】

検索キーワード作成装置 20 は、取得した情報に基づいて、複数のウェブページに関連するキーワードを特定する。具体的には、検索キーワード作成装置 20 は、取得した情報に基づいて、複数のウェブページそれぞれを構成する形態素を分析し、TF-IDF 法を用いることによって得られる値が大きい方から所定の数のキーワードを、複数のウェブページに関連するキーワードとして特定する (S3)。そして、検索キーワード作成装置 20 は、特定したキーワードを、ユーザによって入力された検索キーワードに加えて、絞り込み検索を行わせるための新たな検索キーワードを作成する (S4)。本手法は実現方法の一例であり、たとえばこの代わりに Key-Graph 法を用いて検索キーワードを作成しても良い。

40

【0036】

そうすると、通信装置 10 は、検索キーワード作成装置 20 によって作成された新たな検索キーワードを、通信ネットワーク 400 を介して検索装置 300 に送信する。検索装置 300 は、ウェブページ分析装置 100 の通信装置 10 からの新たな検索キーワードを受信し、通信ネットワーク 400 において開示されているウェブページのなかから、新たな検索キーワードに関連するウェブページを検索する。ここでも、説明の便宜上、検索装置 300 は、複数のウェブページを検索したと仮定する。検索装置 300 は、検索した複数のウェブページそれぞれを構成する情報を、通信ネットワーク 400 を介してウェブページ分析装置 100 に送信する。

【0037】

50

(複数のウェブページ相互の関係を評価する動作)

次に、複数のウェブページ相互の関係を評価する動作(フェーズF2)を図5を用いて説明する。図5は、ウェブページ関係評価装置30が行う複数のウェブページ相互の関係を評価する動作の各ステップを示すフローチャートである。

【0038】

ウェブページ分析装置100の通信装置10は、フェーズF1の最後のステップにおいて検索装置300によって検索された複数のウェブページ(判断対象のウェブページ)を構成する情報を受信する。ウェブページ関係評価装置30では、取得部1が、通信装置10によって受信された複数の判断対象のウェブページそれぞれを構成する情報を取得する(S11)。作成時刻検出部2は、取得部1によって取得された情報に基づいて、複数の判断対象のウェブページそれぞれの作成時刻を検出する(S12)。

10

【0039】

類似度算出部3は、取得部1によって取得された情報に基づいて、複数の判断対象のウェブページそれぞれについて、他の判断対象のウェブページそれぞれとの間の類似度を算出する(S13)。

【0040】

類似度補正部4は、類似度算出部3によって算出された類似度を補正する(S14)。類似度補正部4の動作を、複数の判断対象のウェブページのうちの2個の判断対象のウェブページを取り上げて具体的に説明する。説明の便宜上、2個の判断対象のウェブページをウェブページX及びウェブページYと呼び、ウェブページXがウェブページYより先に作成されたと仮定する。また、類似度算出部3によって算出されたウェブページXとウェブページYとの類似度を「類似度S」という。

20

【0041】

このとき、類似度補正部4は、取得部1によって取得された情報に基づいて、ウェブページYがウェブページXにリンクしているか否かを判断する。ウェブページYがウェブページXにリンクしていると判断した場合、類似度補正部4は、類似度Sに第1の補正係数C1を付与する。ウェブページYがウェブページXにリンクしていないと判断した場合、類似度補正部4は、類似度Sには第1の補正係数C1を付与しない。

【0042】

また、類似度補正部4は、取得部1によって取得された情報に基づいて、ウェブページX及びウェブページYがともに他のウェブページにリンクしているか否かを判断する。2個のウェブページがともに他のウェブページにリンクしていると判断した場合、類似度補正部4は、類似度Sに第2の補正係数C2を付与する。2個のウェブページの少なくとも一方が他のウェブページにリンクしていないと判断した場合、類似度補正部4は、類似度Sには第2の補正係数C2を付与しない。

30

【0043】

また、類似度補正部4は、作成時刻検出部2によって検出された作成時刻に基づいて、ウェブページXの作成時刻とウェブページYの作成時刻との差を判断し、その差に応じた第3の補正係数C3を類似度Sに付与する。例えば、類似度補正部4は、以下に示す式(1)により特定される第3の補正係数C3を類似度Sに対して付与する。

40

【0044】

$$C3 = \exp(-x \times (s_p / t_p)) \dots (1)$$

は、実験により決定される定数であり、 s_p は、ウェブページXの作成時刻とウェブページYの作成時刻との差であり、 t_p は、全ての判断対象のウェブページのうち最初に作成されたウェブページの作成時刻と最後に作成されたウェブページの作成時刻との差である。

【0045】

また、類似度補正部4は、作成時刻検出部2によって検出された作成時刻に基づいて、ウェブページXの作成時刻とウェブページYの作成時刻との間に作成された他の判断対象のウェブページの数を検出する。類似度補正部4は、検出した数に応じた第4の補正係数

50

C 4 を類似度 S に付与する。例えば、類似度補正部 4 は、以下に示す式 (2) により特定される第 4 の補正係数 C 4 を類似度 S に対して付与する。

【 0 0 4 6 】

$$C 4 = \exp (- \quad \times (s n / t n) \quad \dots (2)$$

は、実験により決定される定数であり、s n は、全ての判断対象のウェブページのうち、ウェブページ X の作成時刻とウェブページ Y の作成時刻との間に作成されたウェブページの数であり、t n は、全ての判断対象のウェブページの数である。

【 0 0 4 7 】

また、類似度補正部 4 は、ウェブページ X の単位期間あたりのアクセス数に応じた第 5 の補正係数 C 5 を類似度算出部 3 によって算出された類似度 S に対して付与する。この場合、類似度補正部 4 は、先ず、ウェブページ X 及びウェブページ Y を含む全ての判断対象のウェブページそれぞれの単位期間あたりのアクセス数を取得部 1 に取得させる。そして、類似度補正部 4 は、以下に示す式 (3) により特定される第 5 の補正係数 C 5 を類似度 S に対して付与する。

【 0 0 4 8 】

$$C 5 = \exp (- \quad \times (n / N) \quad \dots (3)$$

は、実験により決定される定数であり、n は、ウェブページ X の単位期間あたりのアクセス数であり、N は、ウェブページ X 及びウェブページ Y を含む複数の判断対象のウェブページそれぞれの単位期間あたりのアクセス数の最大値である。

【 0 0 4 9 】

そして、類似度補正部 4 は、上述した全部の補正係数 (第 1 の補正係数 C 1 ~ 第 5 の補正係数 C 5) を加算し、加算することによって得た加算補正係数を類似度 S に乗算して類似度 S を補正する。

【 0 0 5 0 】

ここまでは、類似度補正部 4 が類似度算出部 3 によって算出されたウェブページ X とウェブページ Y との類似度 S を補正する動作を説明した。類似度補正部 4 は、その動作と同じ動作を類似度算出部 3 によって算出された全ての類似度に対して行い、それにより類似度算出部 3 によって算出された全ての類似度を補正し、全ての類似度について補正類似度を算出する (S 1 4) 。

【 0 0 5 1 】

関連ウェブページ特定部 5 は、類似度補正部 4 によって得られた複数の補正類似度のうちの、2 個のウェブページ相互が関連していると判断するための基準値以上である補正類似度を検出する。そして、関連ウェブページ特定部 5 は、検出した各補正類似度について、その補正類似度を算出する際に用いられた 2 個の判断対象のウェブページを、相互に関連しているウェブページであると特定する (S 1 5) 。

【 0 0 5 2 】

ネットワーク図作成部 6 は、関連ウェブページ特定部 5 によって特定された 2 個のウェブページの組それぞれについて、その組を構成する 2 個のウェブページを、作成時刻検出部 2 によって検出された作成時刻の順に仮想的に並べて仮想的に線で結び、関連しているウェブページのネットワーク図を作成する (S 1 6) 。

【 0 0 5 3 】

ここで、関連ウェブページ特定部 5 及びネットワーク図作成部 6 の動作の具体例を図 6 (A) 及び図 6 (B) を用いて説明する。図 6 (A) は、作成時刻の順に並べられた 4 個の判断対象のウェブページと、各判断対象のウェブページ間の類似度とを示す図である。図 6 (B) は、図 6 (A) の 4 個の判断対象のウェブページの関連しているウェブページ相互を線で結ぶことにより作成されたネットワーク図である。図 6 (A) 及び図 6 (B) における横軸 t は時間軸である。

【 0 0 5 4 】

説明の便宜上、図 6 (A) 及び図 6 (B) に示される 4 個の判断対象のウェブページは、ウェブページ A、ウェブページ B、ウェブページ C、及びウェブページ D であると仮定

10

20

30

40

50

する。図6(A)及び図6(B)における丸はウェブページである。また、図6(A)及び図6(B)における4個のウェブページのうちで、ウェブページAが最先に作成され、その次にウェブページBが作成され、その次にウェブページCが作成され、最後にウェブページDが作成されたと仮定する。

【0055】

また、図6(A)に示すように、ウェブページAとウェブページBとの補正類似度が0.8であり、ウェブページAとウェブページCとの補正類似度が0.8であり、ウェブページAとウェブページDとの補正類似度が0.3であり、ウェブページBとウェブページCとの補正類似度が0.2であり、ウェブページBとウェブページDとの補正類似度が0.3であり、ウェブページCとウェブページDとの補正類似度が0.6であると仮定する。更に、2個のウェブページ相互が関連していると判断するための基準値が0.5であると仮定する。

10

【0056】

このとき、関連ウェブページ特定部5は、ウェブページAとウェブページB、ウェブページAとウェブページC、ウェブページCとウェブページDそれぞれを、相互に関連しているウェブページであると特定する。ネットワーク図作成部6は、ウェブページAとウェブページB、ウェブページAとウェブページC、ウェブページCとウェブページDそれぞれを線で結び、図6(B)に示すネットワーク図を作成する。

【0057】

出力部7は、ネットワーク図作成部6によって作成されたウェブページのネットワーク図を通信装置10に出力する。

20

【0058】

通信装置10は、ウェブページ関係評価装置30からのウェブページのネットワーク図を、通信ネットワーク400を介してクライアント端末装置200に出力する。クライアント端末装置200は、ウェブページ関係評価装置30によって得られたウェブページのネットワーク図を受信して表示装置250に表示させる。

【0059】

これにより、ユーザは、自らが決定した検索キーワードと、その検索キーワードに関連するウェブページのうちの自らが選択したウェブページとに関連するウェブページについて、作成時刻の順に関連付けられたネットワークを知ることができる。つまり、ユーザは、自らが決定した検索キーワードに関連する複数のウェブページ相互の関連性を知ることができる。

30

【0060】

なお、上記の説明では、類似度補正部4は、第1の補正係数C1~第5の補正係数C5の全部の補正係数を加算し、加算することによって得た加算補正係数を類似度Sに乘算して類似度Sを補正する。しかしながら、類似度補正部4は、第1の補正係数C1~第5の補正係数C5の一部の補正係数を加算して加算補正係数を取得し、その加算補正係数を類似度Sに乘算して類似度Sを補正してもよい。また、類似度補正部4は、第1の補正係数C1~第5の補正係数C5の1個の補正係数を類似度Sに乘算して類似度Sを補正してもよい。

40

【0061】

また、上記の説明では、関連ウェブページ特定部5は、類似度補正部4によって得られた複数の補正類似度のうちの2個のウェブページ相互が関連していると判断するための基準値以上である補正類似度を検出する。しかしながら、関連ウェブページ特定部5は、類似度算出部3によって得られた(補正前の)複数の類似度のうちの上記基準値以上である類似度を検出してもよい。その場合、関連ウェブページ特定部5は、検出した各類似度について、その類似度を算出する際に用いられた2個の判断対象のウェブページを、相互に関連しているウェブページであると特定する。

【0062】

(ウェブサイトを評価する動作)

50

次に、ウェブサイトの評価する動作（フェーズF3）を図7を用いて説明する。図7は、サイト評価装置40が行うウェブサイトの評価する動作の各ステップを示すフローチャートである。

【0063】

サイト評価装置40は、通信装置10によって受信された全ての判断対象のウェブページをドメイン毎に分類し、それによって各ウェブサイト特定する（S21）。ウェブサイトは、判断対象のウェブページのうちの共通するドメインを有するページの集合である。図8を用いてウェブサイトを説明する。図8は、ウェブサイトを説明するための図である。図8では、ウェブページH、ウェブページJ、ウェブページK、及びウェブページMは、同じドメインを有するウェブページであって、これらの集合がウェブサイトである。

10

【0064】

したがって、図8では、ウェブページH、ウェブページJ、ウェブページK、及びウェブページMは、ウェブサイトGの下位のウェブページである。また、ウェブページH、ウェブページJ、ウェブページK、及びウェブページMのランクは、それぞれh、j、k、及びmである。ランクは、検索装置300によって決定される値である。例えば、一般的な検索エンジンを検索装置として用いた場合、ランクは検索結果の表示順位である。

【0065】

次に、サイト評価装置40は、通信装置10を介してサイトアクセス数推定装置500を用い、各ウェブサイトへのアクセス数を取得する（図7のS22）。図8の例ではウェブサイトGのアクセス数gを取得する。

20

【0066】

次に、サイト評価装置40は、通信装置10を介して、ウェブサイト毎に、そのウェブサイトのドメインと同じドメインを有するウェブページのランクを取得する（図7のS23）。図8の例では、サイト評価装置40は、ウェブページH、ウェブページJ、ウェブページK、及びウェブページMそれぞれのランクを取得する。

【0067】

次に、サイト評価装置40は、ウェブサイト毎に、そのウェブサイトへのアクセス数とウェブサイトにも所属するウェブページのランクを用い、任意の方法を用いて、そのウェブサイトの重要度を算出する（図7のS24）。

【0068】

最後に、サイト評価装置40は、重要度が大きい順に全てのウェブサイトに順位を付与する（図7のS25）。

30

【0069】

通信装置10は、サイト評価装置40によって得られた各ウェブサイトのタイトルをその順位及びその重要度とともに、通信ネットワーク400を介してクライアント端末装置200に送信する。クライアント端末装置200は、サイト評価装置40によって得られた各ウェブサイトのタイトル、順位、及び重要度を受信し、各ウェブサイトのタイトルを順位にしたがって表示装置250に表示させる。その際、クライアント端末装置200は、各ウェブサイトの重要度もタイトルと組にして表示装置250に表示させる。

【0070】

上述したように、本実施の形態では、ウェブページ関係評価装置30は、複数の判断対象のウェブページを、作成時刻の順に仮想的に並べ、それらの内で関連しているウェブページを仮想的に線で結び、それによりツリー状のウェブページのネットワーク図を作成する。作成されたネットワーク図において、時系列順の上位にあるサイトは情報に敏感であることを意味し、関連するサイトを時系列順の下位に多く持つサイトは情報を他のウェブサイトへ伝播する能力が高いことを意味する。また、サイト評価装置40は、各ウェブサイトの重要度を算出する。重要度にもとづき順位付けされたウェブサイトの一覧の上位にあるサイトは、アクセス数や検索装置でのランクからの観点から情報の伝播に高い影響力があることを意味する。

40

【0071】

50

本発明装置に対し、例えば企業を中傷する事実無根の情報が伝播した事例に関連したキーワードを入力して上述の分析結果を得ることで、当該事例で情報伝播に大きな役割をはたしたウェブサイトを見つけることができる。当該サイトへの自粛要請を行うことで再発防止策の検討につなげることができる。また例えば、メーカーが自社製品発売の発表における広告効果分析を行いたい場合、複数の自社製品名を本発明装置に入力して上述の分析結果を得る。得られた分析結果のうち、多くの自社製品分析事例において高い情報伝播能力をもつとの結果が得られる傾向のあるサイトは、今後も自社製品発売の発表における情報伝播に大きな役割を持つと考えられる。このサイトに働きかけを行うことでより効果的な広告効果を得ることができる。

【 0 0 7 2 】

なお、ウェブページ分析装置 1 0 0 を構成する各装置の機能は、例えばコンピュータの CPU 及びメモリ等のハードウェアと、コンピュータプログラムとが協働することによって実現される。しかしながら、上記各機能は、専用の回路により実現される等、どのような形態により実現されてもよい。

【 0 0 7 3 】

また、ウェブページ関係評価装置 3 0 は、作成時刻検出部 2、類似度算出部 3、類似度補正部 4、関連ウェブページ特定部 5、及びネットワーク図作成部 6 それぞれに対応する専用のメモリを有していてもよいし、それら各構成部に共通のメモリを有していてもよい。それら各構成部は、処理するためのデータ及び処理結果を専用のメモリ又は共通のメモリに格納してもよい。

【符号の説明】

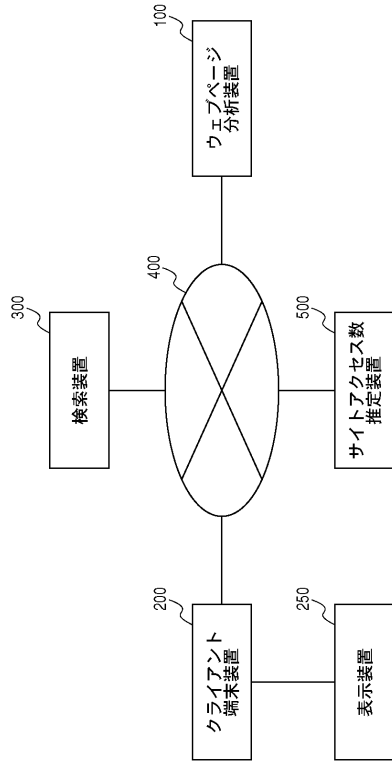
【 0 0 7 4 】

1 0 0 ウェブページ分析装置、 2 0 0 クライアント端末装置、 3 0 0 検索装置、 4 0 0 通信ネットワーク、 5 0 0 サイトアクセス数推定装置、 2 5 0 表示装置、 1 0 通信装置、 2 0 検索キーワード作成装置、 3 0 ウェブページ関係評価装置、 4 0 サイト評価装置、 1 取得部、 2 作成時刻検出部、 3 類似度算出部、 4 類似度補正部、 5 関連ウェブページ特定部、 6 ネットワーク図作成部、 7 出力部。

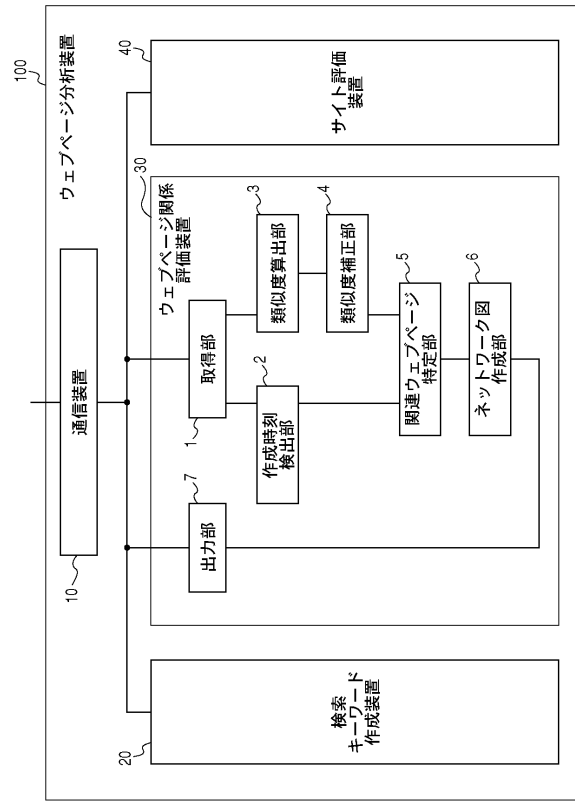
10

20

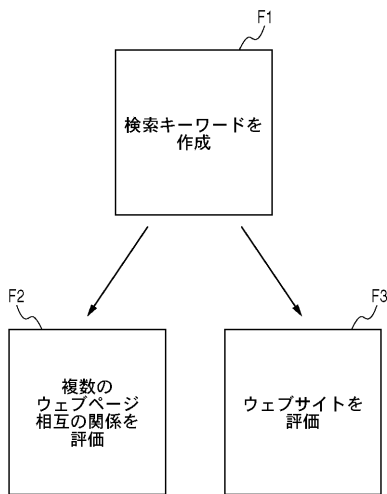
【図1】



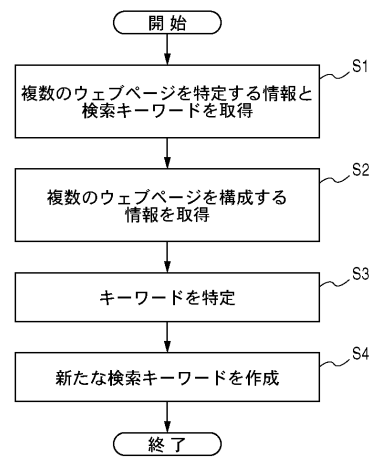
【図2】



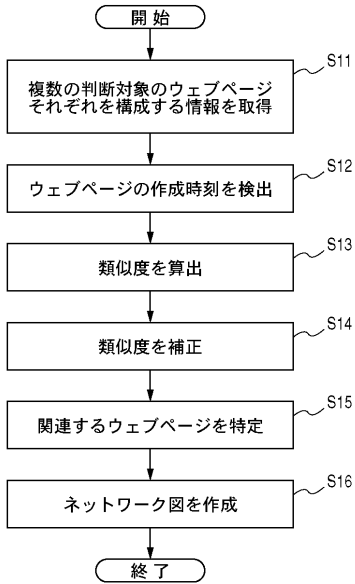
【図3】



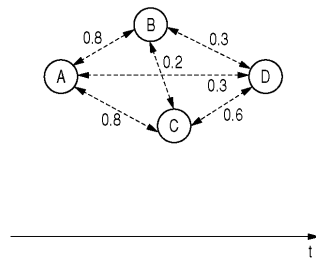
【図4】



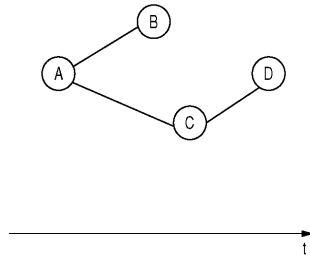
【図5】



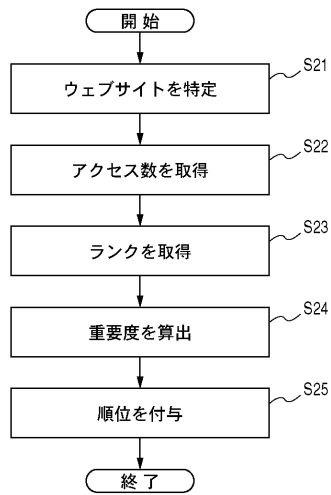
【図6(A)】



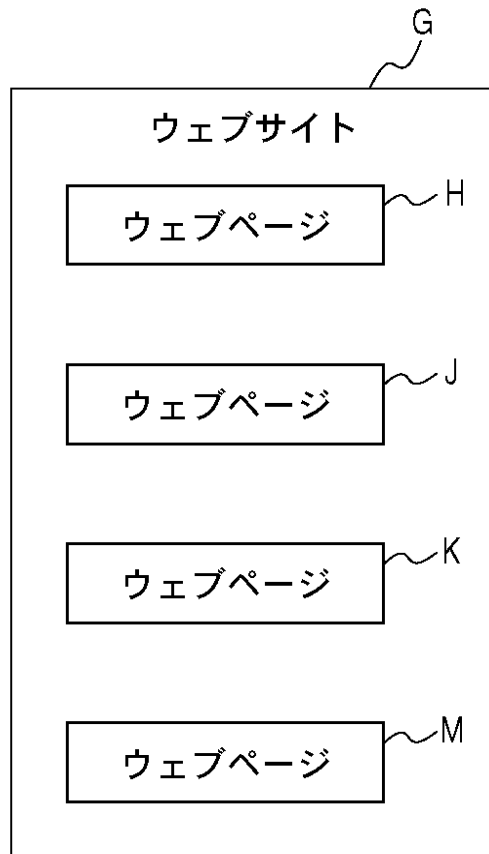
【図6(B)】



【図7】



【図8】



フロントページの続き

- (72)発明者 顧 明
中華人民共和国北京市海淀区清華大学内
- (72)発明者 古川 和年
東京都千代田区丸の内一丁目6番5号 株式会社野村総合研究所内
- (72)発明者 阿部 昌平
東京都千代田区丸の内一丁目6番5号 株式会社野村総合研究所内
- (72)発明者 犬塚 裕介
東京都千代田区丸の内一丁目6番5号 株式会社野村総合研究所内

審査官 伊知地 和之

- (56)参考文献 特開2004-021665(JP,A)
特開平11-053387(JP,A)
特開2001-147936(JP,A)
特開2009-151373(JP,A)
特開2005-122295(JP,A)

- (58)調査した分野(Int.Cl., DB名)
G06F 17/30
CSDB(日本国特許庁)