

(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2012-248119

(P2012-248119A)

(43) 公開日 平成24年12月13日(2012.12.13)

(51) Int.Cl.		F 1	テーマコード (参考)
G06F 17/30	(2006.01)	G06F 17/30	220Z
G06F 13/00	(2006.01)	G06F 13/00	540R
		G06F 17/30	340B
		G06F 17/30	419B

審査請求 未請求 請求項の数 10 O L (全 18 頁)

(21) 出願番号	特願2011-121056 (P2011-121056)	(71) 出願人	000005223
(22) 出願日	平成23年5月30日 (2011.5.30)		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	100070150
			弁理士 伊東 忠彦
		(74) 代理人	100146776
			弁理士 山口 昭則
		(72) 発明者	黒田 紗也子
			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
		(72) 発明者	友野 一生
			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
		Fターム(参考)	5B084 AA02 AA12 AB30 AB36 AB39 BB15 CD07 CE03 CE12 CF12

(54) 【発明の名称】 キーワード抽出及びウェブコンテンツアクセス情報の収集装置

(57) 【要約】

【課題】

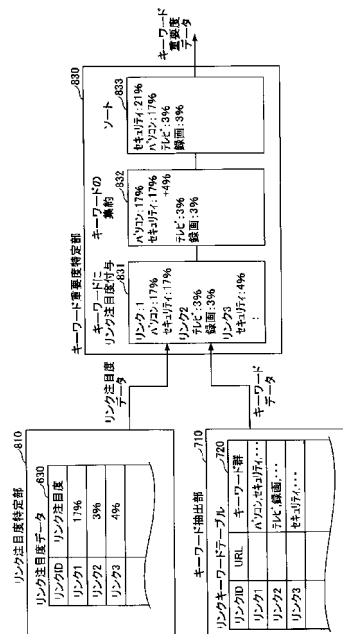
ユーザの嗜好に適合したキーワードの抽出及びウェブコンテンツアクセス情報を収集する。

【解決手段】

ウェブページを閲覧するユーザに関して、ウェブページ内の各リンクに対するクリック回数をカウントする。別途ウェブページから、各リンクに対する、クリックの比率の予測値を算出する。クリック回数と、クリックの比率の予測値とに基づいて、ユーザに関するリンクの注目度、及びウェブページに関連するキーワードの重要度を特定する。

【選択図】 図 8 A

キーワードの重要度を特定する手法の例を示す図



【特許請求の範囲】

【請求項 1】

ウェブページの各リンクに対するユーザの指示回数をカウントする指示回数カウント部と、

前記ウェブページから、前記各リンクに対する、指示の比率の予測値を算出する指示比率予測値算出部と、

前記指示回数と、前記指示の比率の予測値とに基づいて、前記各リンクの注目度を特定するリンク注目度特定部と、

を有することを特徴とするサーバ装置。

【請求項 2】

10

前記各リンクに関連する 1 つ以上のキーワードを抽出するキーワード抽出部と、

前記各リンクに対応する、前記注目度と、前記各キーワードとに基づいて、前記各キーワードの重要度を特定するキーワード重要度特定部と、

を更に有することを特徴とする請求項 1 に記載のサーバ装置。

【請求項 3】

前記リンク注目度特定部は、

【数 4】

$$F_m = \frac{n_m}{\sum_{k=1}^M n_k} - \beta \alpha_m \quad (1)$$

20

ここで、

F_m は、全てのリンクの数を M 個とした場合の m 番目のリンクの注目度、

n_m は、前記 m 番目のリンクの指示回数、

α_m は、前記 m 番目のリンクの指示の比率の予測値、

β は $0 < \beta < 1$ を満足する定数、

30

によってリンクの注目度を算出することを特徴とする請求項 1 又は 2 に記載のサーバ装置。

【請求項 4】

前記指示の比率の予測値は、前記各リンクの、面積、位置、及び背景とのコントラスト比のうち少なくともいずれか一つに基づき決定される、請求項 1 乃至 3 のうちいずれか 1 項に記載のサーバ装置。

【請求項 5】

前記各キーワードは、前記リンクを構成する文字列を含む情報から、及び / 又は、前記リンクが指し示すコンテンツから、抽出することを特徴とする請求項 1 乃至 4 のうちいずれか 1 項に記載のサーバ装置。

40

【請求項 6】

複数のウェブページの各々の指示回数に応じてリンクの注目度を重み付けすることによって、前記複数のウェブページに存在する前記リンクの各々の注目度を正規化するリンク注目度正規化部、

を更に有することを特徴とする請求項 1 乃至 5 のうちいずれか 1 項記載のサーバ装置。

【請求項 7】

コンピュータにより実行される方法であって、

ウェブページの各リンクに対するユーザの指示回数をカウントし、

前記ウェブページから、前記各リンクに対する、指示の比率の予測値を算出し、

前記指示回数と、前記指示の比率の予測値とに基づいて、前記各リンクの注目度を特定

50

することを特徴とする方法。

【請求項 8】

前記各リンクに関連する 1 つ以上のキーワードを抽出し、

前記各リンクに対応する、前記注目度と、前記各キーワードとに基づいて、前記各キーワードの重要度を特定することを特徴とする請求項 7 に記載の方法。

【請求項 9】

ウェブページの各リンクに対するユーザの指示回数をカウントし、

前記ウェブページから、前記各リンクに対する、指示の比率の予測値を算出し、

前記指示回数と、前記指示の比率の予測値とに基づいて、前記各リンクの注目度を特定する、

処理をコンピュータに実行させるプログラム。

【請求項 10】

前記各リンクに関連する 1 つ以上のキーワードを抽出し、

前記各リンクに対応する、前記注目度と、前記各キーワードとに基づいて、前記各キーワードの重要度を特定する、

処理を更にコンピュータに実行させることを特徴とする請求項 9 に記載のプログラム。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、ユーザのウェブアクセスを基にした、キーワードの抽出及びウェブコンテンツアクセス情報の収集のための装置、方法、及びプログラムに関する。

【背景技術】

【0002】

ウェブページにより提供されるコンテンツは多種多様である。このため、ユーザの嗜好に適合したコンテンツを選択し適切に提供することが望まれている。このためには、ユーザの嗜好に合致したコンテンツやキーワードを把握することが重要である。また、同じコンテンツを提供する場合であっても、ユーザの嗜好に合ったウェブページのデザインを構築することも重要である。

【0003】

このため、クライアントが 1 つのコンテンツを実際に見ている状況を測定し、各コンテンツの利用率を算出する技術が知られている（特許文献 1）。これは、クライアントのブラウザのサイズと、コンテンツのサイズと、クライアントでのクリック回数とから利用率が計算される。あるいは、クライアントのブラウザでのスクロールバーの移動量と、ブラウザのサイズとから利用率が計算される。

【0004】

また、ウェブブラウザに表示されているウェブコンテンツの該当個所から、特定の操作の対象となったテキストを取得する技術が知られている（特許文献 2）。この技術において、取得したテキストは、当該テキストを利用する他のシステムに送信される。そして、例えば、ウェブ視聴率調査を行うシステムや検索エンジンが、この取得されたテキストを受信し、ユーザが興味を持った対象に関する情報を提供するために利用する。

【0005】

また、ユーザが WWW 上の任意のページを見ている場合において、ユーザが表示した範囲やマウスでなぞったウェブページ上の一部領域を特定し、そこに含まれるテキスト情報を検索キーとして抽出する技術が知られている（特許文献 3）。この抽出したテキスト情報を情報処理サーバ装置へ送信する。情報処理サーバ装置は、注目テキスト抽出部からの受信情報と嗜好情報データベースに事前に格納されたユーザの嗜好情報を用いて、広告の検索を行う。その検索結果を用いて広告表示のためのページを作成する。作成された広告表示のページが広告提供ブラウザの広告表示部へ送信され、広告が表示される。

【0006】

また、ユーザが閲覧したコンテンツから、当該ユーザの閲覧したキーワード（ユーザー

10

20

30

40

50

キーワード)を記録する技術が知られている(特許文献4)。この技術は、ユーザーキーワード毎の評価点をキーワードデータベースに蓄積し、このキーワードデータベース上のユーザーキーワードの評価点に基づき、評価点の高い順にコンテンツを個別ユーザのコンピュータ端末に表示している。

【先行技術文献】

【特許文献】

【0007】

【特許文献1】国際公開第2004/097654号公報

【特許文献2】特開2001-188792号公報

【特許文献3】特開2004-348241号公報

【特許文献4】特開2003-308339号公報

【発明の概要】

【発明が解決しようとする課題】

【0008】

しかしながら、ユーザは、ウェブページのレイアウト等に基づくクリックのしやすさや、ウェブページの見栄えなど、ユーザの嗜好以外の要因により、ウェブページをアクセスする場合も多い。従って、上記の技術においては、ユーザの嗜好に合ったキーワード抽出等の技術を十分に提供しているとは言えない。

【0009】

本装置は、このような技術的課題を解決するためになされたものであって、ユーザの嗜好に適合したキーワードの抽出及びウェブコンテンツアクセス情報を収集し、これらの情報の提供を行う装置、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】

【0010】

本装置の一観点によれば、ウェブページの各リンクに対するユーザの指示回数をカウントする指示回数カウント部と、前記ウェブページから、前記各リンクに対する、指示の比率の予測値を算出する指示比率予測値算出部と、前記指示回数と、前記指示の比率の予測値とに基づいて、前記各リンクの注目度を特定するリンク注目度特定部と、を有することを特徴とするサーバ装置が提供される。

【発明の効果】

【0011】

本装置の一観点によれば、ユーザの嗜好に適合したキーワードの抽出及びウェブコンテンツアクセス情報を収集し、これらの情報の提供を行うことができる。

【図面の簡単な説明】

【0012】

【図1】本発明の一実施形態のシステムの構成を示す図である。

【図2】サーバ装置のハードウェア構成を示す図である。

【図3】ウェブページのリンク毎のクリック回数・クリック比率の例を示す図である。

【図4】ウェブページを分割しクリック比率の予測値を算出する例を示す図である。

【図5】リンク毎のクリック比率の予測値の算出結果の例を示す図である。

【図6】リンクの注目度を特定する手法の例を示す図である。

【図7】キーワードを抽出する手法の例を示す図である。

【図8A】キーワードの重要度を特定する手法の例を示す図である。

【図8B】複数のページを扱う場合のリンク注目度の正規化の例を示す図である。

【図9A】ウェブページレイアウトの変更の処理の例を示す図である

【図9B】ウェブページレイアウトの変更の結果の例を示す図である。

【図10】各実施形態における機能的ブロック図の一例を示す図である。

【図11】一実施形態としての方法のフローを示す図である。

【発明を実施するための形態】

【0013】

10

20

30

40

50

以下、図面を参照しながら、本発明の実施形態について説明する。なお、以下の各実施形態においては、例示として「クリック」の語を用いて説明するが、「クリック」は、「指示」の一例である。「指示」とは、例えば、マウスの場合には、「クリック」を意味し、タッチパネルの場合には、「タップ」と呼ばれる場合がある。あるいは、ウェブページにおいて、タブキーを使って、フォーカスをリンクやボタンに移動させ、所望のリンクにフォーカスが移動した状態で、Enterキーを押下すれば、マウスによるクリックと同様の処理が行われる。したがって、本発明は、マウスによるクリックに限られるものではなく、ユーザーインターフェースを用いたあらゆる「指示」を対象とする点に留意すべきである。

【0014】

図1は、本発明の一実施形態のシステム100の構成を示している。システム100は、記憶装置A122が接続されたウェブサーバ装置120、記憶装置B142が接続されたウェブコンテンツ管理サーバ装置140、クライアント131、クライアント132、及びネットワーク110を有する。また、ウェブサーバ装置120、ウェブコンテンツ管理サーバ装置130、クライアント131及びクライアント132は、それぞれ、回線111乃至114で、ネットワーク110と接続されている。

【0015】

ウェブサーバ装置120は、記憶装置A122に記憶されたウェブコンテンツ124を用いて、ウェブページを、ネットワーク110を介してクライアント131又は132に提供する。また、ウェブサーバ装置120は、ウェブコンテンツへのアクセスに関する種々のイベントをアクセスログ126に記憶する。このアクセスログ126は、ウェブコンテンツ管理サーバ装置140等に提供されてもよい。

【0016】

ウェブコンテンツ管理サーバ装置140は、ウェブサーバ装置120が公開しているウェブコンテンツ124の中にある複数のリンクの注目度を算出し、リンク注目度データ144として、記憶装置B142に記憶してもよい。また、ウェブコンテンツ管理サーバ装置140は、ウェブコンテンツ124に存在するキーワードや、ウェブコンテンツ内のリンクで示されたコンテンツ内のキーワードの重要度を算出し、キーワード重要度データ145として、記憶装置B142内に記憶してもよい。ウェブコンテンツ管理サーバ装置140は、上述の算出を行うために、アクセスログ126の情報などを利用してもよい。また、記憶装置142内には、ウェブコンテンツ124のリンクのクリック比率予測値146や、それぞれのユーザにカスタマイズされたウェブコンテンツのレイアウトをウェブレイアウトデータ147として記憶してもよい。上述の各種データ及び処理の詳細に関しては後述する。

【0017】

なお、ウェブサーバ装置120及びウェブコンテンツ管理サーバ装置140は、必ずしも別個のサーバ装置である必要はない。上述の種々の機能を1つのサーバ装置にインプリメントしてもよい。また、本明細書において、ウェブサーバ装置120に存在するとして説明する機能やデータの一部をウェブコンテンツ管理サーバ装置140に存在させてもよい。その逆も同様である。また、データについては、そのコピーを双方のサーバが保持してもよい。

【0018】

また、クライアント131は、パーソナルコンピュータ、携帯電話、多機能電話、タブレットコンピュータなど、ネットワークに接続する機能を有する装置が含まれる。

【0019】

図2は、本実施形態のサーバ装置120及び140のハードウェア構成200を示している。サーバ装置は、CPU210、CD-ROM等の機械読み取り可能な媒体225を読み書きできるドライブ220、ハードディスクなどの動的記憶装置230、RAM232、ROM234、I/O236、及び通信装置238を含む。

【0020】

10

20

30

40

50

以下に説明する各種の実施形態の機能は、適切なプログラムがCPU 210に読み込まれ実行されることにより仮想的に実現されてもよい。

【0021】

また、プログラムに係る本発明は、機械読み取り可能な記録媒体に格納されてもよい。機械読み取り可能な記録媒体には、磁気記録媒体、光ディスク、光磁気記録媒体、半導体メモリなどがある。磁気記録媒体には、HDD、フレキシブルディスク(FD)、磁気テープ(MT)などがある。光ディスクには、DVD(Digital Versatile Disc)、DVD-RAM、CD-ROM(Compact Disc-Read Only Memory)、CD-R(Recordable)/RW(Rewritable)などがある。また、光磁気記録媒体には、MO(Magneto-Optical disk)などがある。

10

【実施例】

【0022】

<リンクの注目度の特定>

図3は、ウェブページのリンク、リンク毎の計測されたクリック回数、及びクリック比率の例を示している。図3(a)は、ウェブページ300のレイアウトを示しており、リンク1乃至リンク8が各々の領域に貼り付けられている。例えば、ユーザがリンク3の領域303をクリックすると、ブラウザは、リンク3(例えばハイパーリンク)に割り当てられたURLのページをディスプレイに表示する。なお、リンクは、単にリンクに埋め込まれたURLを表示させる以外に、スクリプトなどを起動させて、動画の表示や所望のアプリケーションを起動させるものであってもよい。

20

【0023】

図3(b)は、ウェブページ内の各々の領域のクリック回数を示している。リンク3の領域313は、クリック回数が12233回であることを示している。このクリックの回数は、ウェブページを閲覧したすべてのユーザに対するクリック回数、特定のユーザ集合(例えば男性のユーザ)のクリック回数、あるいは、特定のユーザのクリック回数などが取得できる。特定のリンクがクリックされた情報は、ウェブページ300及びそれぞれのリンク先のウェブページのアクセスログを解析することにより取得できる。あるいは、ウェブページ300に、各リンクをクリックした際のイベントを収集するスクリプトを記述しておき、そのイベント情報を例えばウェブコンテンツ管理サーバ装置140に送信するようにしてもよい。ユーザを特定するためには、そのページを表示させるためにユーザIDとパスワードを入力させることにより実現できる。またユーザ層(集合を構成する1以上のユーザ)を特定するためには、ユーザIDとパスワードの登録時に必要な情報(性別、生年月日、パソコン使用経験年数、業種等)の入力をユーザに促すことにより行える。また、クライアントマシン上に置かれるクッキー等を用いることにより、ユーザIDとパスワードを一旦入力すれば、その後においても、ID及びパスワードの入力を省略しても個人を特定することは一定の限度で可能である。

30

【0024】

図3(c)は、ウェブページの各領域のクリックの比率を示している。例えば、ウェブページ全体のクリック回数に対するリンク3に該当する領域323のクリックの比率は、21%である。このクリックの比率は、リンク3の領域のクリック回数である12233回を、ウェブページ全体のクリック回数で割ることにより算出できる。各領域のクリックの比率を合計すると100%となるようにしてもよい。なお、図3の例では、ページ全体のクリック回数を基に各領域のクリック回数の比を算出したが、これに限られるものではない。例えば、特定のユーザに注目し、その特定の1ユーザのクリック回数を他のユーザのクリック回数よりも大きな重みを付けて各領域のクリック比率を算出してよい。

40

【0025】

図4は、ウェブページを分割しクリック比率の予測値を算出するための例を示している。図4(a)は、ウェブページを35等分に分割した例を示している。ウェブページを分割する理由は、クリックの比率の予測値の算出を単純化するためである。分割の個数や、

50

分割の方法は、任意に変更してもよい。本明細書では、この分割された各区域を「スクエア」と呼ぶ（なお、ここではスクエアと呼ぶが、正方形である必要はない）。これに対して、各リンクを構成している区域を「領域」（例えば領域403）と呼び、上記スクエア410と区別することとする。

【0026】

図4(b)は、フィッツのUIの法則を用いて、クリックのしやすさをパーセンテージで表現したものである。フィッツのUIの法則とは、マウスポインタの現在位置から遠く、サイズも小さいオブジェクトは、一般にクリックしにくくなるという法則を言う。図4(b)の例では、ページ中央やや上のスクエア410が一番クリックしやすいものと設定し、このスクエア410に100%の値を付与している。そして、スクエア410から遠くなるに従って、マウスのポインタの移動距離に応じて各スクエアに付与する値を減少させている。これらのクリックのしやすさに係る値は、予め固定値を設定しておいてもよい。

10

【0027】

図4(c)は、ページの背景色とスクエアの色のコントラスト比が高いスクエアに、高いクリック比率の予測値を付与している。この理由は、コントラスト比が高いほどユーザに目立ちやすいため、ユーザのクリックを誘発しやすいという人間の行動パターンに基づくものである。

【0028】

例えば、背景色を白とした場合、スクエアの色が白(RGB: 0 x F F F)のときを0%とし、スクエアの色が黒(RGB: 0 x 0 0 0)を100%とする。そして、例えば、各スクエアの一番広い部分を占める色をそのスクエアの色として採用する。例えばスクエアの色が0 x 8 0 0であれば、50%の値をそのスクエアに付与してもよい。

20

【0029】

図5は、リンク毎のクリック比率の予測値の算出結果の例を示している。この例では、図4(b)及び図4(c)において求めたクリックのしやすさの度合いをスクエア毎に掛け合わせる。次に、各リンクの領域に重なるスクエアの数値を合計し、各リンクの領域の仮のクリック比率の予測値を求める。リンクの領域とスクエアとが完全に重ならない場合は、リンクの領域とスクエアとが重なる部分と重ならない部分の比率で、スクエアに付与された値を按分して、重なった部分のスクエア領域の値を算出し、リンクの領域の値に加算するべき値としてもよい。そして、各リンクの領域の値を合計した数字が100%となるように、各リンクの領域の値に定数をかけ正規化する。この結果が図5に示されている。図5は、ウェブページのコントラストの分布やマウスカーソルの移動のしやすさに基づく、クリック比率の予測値を表す一例である。この予測値は、リンクに係るキーワードによるユーザの嗜好に基づかないクリック比率の予測値ととらえることができる。なお、上述の実施形態では、ウェブページのコントラストの分布やマウスカーソルの移動のしやすさに基づく、クリック比率の予測値を算出したが、これに限られるものではない。

30

【0030】

ウェブページを見たユーザは、ウェブページ上に表示された言葉や意味のある絵などから、ユーザの嗜好に基づいて、リンクをクリックする場合と、ウェブページのコントラストの分布やマウスカーソルの移動のしやすさに基づき、リンクをクリックする場合とがある。計測されたクリック回数から、ユーザの嗜好を抽出するためには、ウェブページのコントラストの分布やマウスカーソルの移動のしやすさ等による、ユーザの嗜好に基づかないクリックの計測値の影響を可能な限り取り除くことが重要である。

40

【0031】

したがって、例えば、図3(c)の実際のクリック比率から、図5のクリック比率の予測値を減算することにより、真にユーザの嗜好に基づくクリック比率により近い、リンクの注目度の値を算出することができる。本実施形態では、図3(c)の実際のクリック比率から図5のクリック比率の予測値を差し引くことにより、ユーザの嗜好に基づくクリック比率を推定し算出している。この値を以下、リンクの注目度と呼ぶ。リンクの注目度は

50

、ユーザの嗜好に基づいて、ユーザがそのリンクを注目している度合いの指標として利用できる。リンクの注目度は、下式によって算出してもよい。

【0032】

【数1】

$$F_m = \frac{n_m}{\sum_{k=1}^M n_k} - \beta \alpha_m \quad (1)$$

10

【0033】

ここで、 F_m は、全てのリンクの数を M 個とした場合の m 番目のリンクの注目度、 n_m は、前記 m 番目のリンクのクリック回数、 α_m は、前記 m 番目のリンクのクリックの比率の予測値である。 β は $0 < \beta < 1$ を満足する定数である。

【0034】

なお、上記実施形態では、単純に減算を行ったが ($\beta = 1$)、 β の値を変更させて、クリック比率の予測の値に、必要な重みをかけて減算を行ってもよい。あるいは、その他の計算方法を適宜用いてもよい。また、 α_m の値は、アクセスしたユーザ全体に対する各リンクの注目度を算出する場合と、特定の一人のユーザに対するリンクの注目度を算出する場合とで異ならせてもよい。

20

【0035】

図6は、各リンクに対する実際のアクセス比率と、アクセス比率の予測値からリンクの注目度を算出する手法の一例を示している。図6(a)は、図3において説明した各リンクに対して計測された実際のクリック比率を示している。図6(b)は、図5において説明したクリック比率の予測値を示している。図6(c)は、各リンクに対して、図6(a)の値から図6(b)の値を減算した値を採用した場合のリンクの注目度を示している。算術の仕方は、上述の式(1)を用いてもよく、あるいは他の算術を採用してもよい。

【0036】

たとえばリンク1に対応する図6(a)の領域611の実際のクリック比率は、20%であり、図6(b)に示すようにクリック比率の予測値は、3%である。これらの数値を用いて、図6(c)に示すように領域631のリンクの重要度は17%となり、ウェブページ630の中で一番高い注目度を示していることが分かる。これに対して、リンク3に対応する領域613は実際のクリック比率が30%と高い数値であった。しかしながら、領域623のクリック比率の予測値が26%であったため、領域633のリンク注目度は、4%と特定されたことを示している。

30

【0037】

図6(c)に示されたリンク注目度に基づいて、そのリンクが指し示すコンテンツなどの注目度を認識することができる。また、上述したように、母集団を特定することにより (例えば男性のユーザ)、特定のユーザ層に関するリンクの注目度を知ることができる。また、特定の1ユーザに対して計測することにより、特定の1ユーザのリンク注目度を特定することができる。

40

【0038】

この特定されたリンク注目度に基づいて、注目されているリンクは、よりクリックしやすい位置に移動させてもよい。また、特定のユーザのためのレイアウト(マイページ)を設定することも可能である。レイアウトの変更の詳細については後述する。

【0039】

<キーワードの重要度の特定>

図7は、キーワードを抽出する手法の一例を示した図である。表示しているウェブペー

50

ジにリンク1(701)が表示されている。以下、このリンク1(701)に関するキーワード群をリンクキーワードテーブル720に格納する処理を説明する。リンク1(701)は、ウェブページ700上に「パソコンのセキュリティ対策」と表されており、ハイパーリンクのURL情報として「http://ABC.def.hg・・・」が埋め込まれている。このハイパーリンクをクリックすると、上記URLのリンク先のページ730がディスプレイ上に表示される。ウェブコンテンツ管理サーバ装置又はキーワード抽出部710は、リンクキーワードテーブルのリンクID欄(721)の「リンク1」に対応するURLのエントリ欄(722)に、上記URL「http://ABC.def.hg・・・」を記入する。そして、キーワード抽出部710は、リンク1(701)のテキスト情報、オルト属性を含む情報から、キーワードを抽出して、キーワード群のエントリ欄(723)に格納してもよい。なお、リンク1が写真など、テキスト以外の情報で構成されている場合には、オルト属性などの情報を用いてもよい。また、キーワード抽出部710は、リンク先のページ730のテキスト情報731やメタタグ情報732を解析し、キーワードを抽出し、キーワード群のエントリ欄(723)に格納してもよい。

10

20

30

40

50

【0040】

なお、キーワードを抽出するためには、日本語の場合には、形態素解析を行い、活用変換や不要語リストによるキーワード候補の絞り込み等のアルゴリズムを使用すればよい。

【0041】

以上のようにして、ウェブページ700の各リンクに対応するキーワード群が、リンクキーワードテーブル720に蓄積される。なお、このリンクキーワードテーブルは、後述するキーワード重要度特定部によって利用され各キーワードに重要度が付与されることになる。

【0042】

図8Aは、キーワードの重要度を特定する手法の例を示している。リンク注目度特定部810によって、リンクID毎のリンク注目度が特定され、リンク注目度データが出力される。リンク注目度の特定の一例は、図6を用いて既に説明した。

【0043】

また、キーワード抽出部710は、各リンクのキーワードを抽出する。キーワード抽出部710の処理の一例は、図7を用いて既に説明した。

【0044】

キーワード重要度特定部830には、リンク注目度特定部810からのリンク注目度データ630と、キーワード抽出部710から、リンクキーワードテーブル720のデータが入力される。

【0045】

キーワード特定部830は、まず、リンクIDを用いて、各キーワードにリンク注目度を対応づけ、キーワードにリンク注目度を付与する処理を行う(831)。次に、同じキーワードに関するリンク注目度を合算することにより、キーワードの集約を行う(832)。次に、対応付けられたリンク注目度の合算値が大きい順にキーワードをソートする(833)。

【0046】

以上の処理によって、キーワードが重要度の順に並んだ、キーワード重要度データが生成され、出力される。そして、このデータは、後述するレイアウト変更部で利用されてもよい。なお、キーワード重要度データは、特定の1ユーザ、特定の集合に属するユーザ(例えば男性)、あるいは、ウェブページをアクセスした全てのユーザを対象としたキーワード重要度データが取得できる。なお、リンク注目度データについても、同様に、種々のユーザに対してリンク注目度データが取得できることは言うまでもない。

【0047】

図8Bは、複数のページを扱う場合のリンク注目度の正規化の例を示している。この図は、ウェブページAのリンク注目度データ631と、ウェブページBのリンク注目度データ632とを統合して、リンク注目度データ633を作成する際の例を示している。ウエ

ブページ A の総クリック回数が 1000 回であり、ウェブページ B の総クリック回数が 100 回であった場合を想定する。この場合には、ウェブページ A のリンク注目度に合わせるため、ウェブページ B のリンク注目度に、ウェブページ A とウェブページ B の総クリック回数の比 $100 / 1000$ をかけて、正規化を行ってもよい。この操作を行うことにより、ウェブページ A とウェブページ B との両者のリンク注目度データを統合して後の処理に利用することもできる。

【0048】

<ウェブページレイアウトの変更>

次に、上述のリンク注目度データ及びキーワード重要度データを用いて、ウェブページのレイアウトを変更する実施例について、図を用いて説明する。なお、ウェブページのレイアウトの変更においては、特定のユーザのウェブページのアクセスの際に、以下に説明する処理がリアルタイムで実行される必要はない。すなわち、ウェブコンテンツ管理サーバ装置は、ログ情報やイベント情報を利用して、順次処理を行い、後の利用のためにウェブページのレイアウト情報の処理結果を蓄積しておけばよい。

10

【0049】

図 9 A は、特定のユーザに係るウェブページレイアウトの変更に関する処理の例を示している。既に述べたように、リンク注目度特定部 810、及びキーワード重要度特定部 830 は、種々のユーザに関連する情報を出力することができる。図 9 A における実施形態では、リンク注目度特定部 810 は、特定ユーザのリンク注目度データ 941、及び集合に属する複数のユーザのリンク注目度データ 942 をレイアウト変更部 950 に出力してもよい。また、キーワード重要度特定部 830 は、特定のユーザのキーワード重要度データ 943、及び集合に属する複数のユーザのキーワード重要度データ 944 をレイアウト変更部 950 に出力してもよい。なお、図 9 A では、リンク注目度特定部 810 から 2 本のデータ線 941 及び 942 が存在しているが、実際には、1 つの線であってもよい。二つのデータを直列に伝送すればよいからである。この点は、キーワード重要度特定部 830 についても同様である。

20

【0050】

レイアウト変更部 950 は、これらの入力をスイッチ機能を有する SW1 乃至 SW4 で受け取り、例えば特定のユーザによる設定入力を受け付けることにより、必要な入力だけを選定してもよい。例えば、SW1 及び SW3 をオフすることにより、特定のユーザに係るデータを考慮しないようにすることができる。この場合には、特定のユーザのウェブページのレイアウトは、集合に属する複数のユーザの嗜好を反映したレイアウトにすることができる。逆に、SW2 及び SW4 をオフにすれば、特定のユーザに係る嗜好を強く反映したウェブページのレイアウトにすることが可能である。

30

【0051】

次に、各入力データ 941 乃至 944 は、閾値の設定部 951 で、例えばユーザの入力設定に基づいて、各入力データの閾値を設定してもよい。この閾値は、特定の値以上の注目度を有するリンク注目度データに限定し、または、特定の値以上の重要度を有するキーワードだけに限定して、後のレイアウト変更処理に利用するために用いてもよい。あるいは、特定のユーザに係る情報については、ウェブページの閲覧回数、クリック回数などが、一定値以上であるかをチェックする閾値として利用してもよい（この場合には、クリック回数を閾値の設定部 951 に与えればよい（不図示））。そして、これらの値が一定値以下の場合には、データ母数が少ないことを意味する。この場合には、取得されたデータを利用せず、レイアウトを変更しないよう設定してもよい。

40

【0052】

その後、各入力データは、処理順序設定部 952 により、例えばユーザの入力設定に基づいて、各入力データの処理順序を定めてもよい。例えば、集合に属するユーザのデータを先に処理し、その後特定のユーザのデータを処理する設定とした場合を想定する。この場合には、特定のユーザのデータの処理が後になされるため、特定のユーザの嗜好が、レイアウトに対してより強く反映されることになる。

50

【 0 0 5 3 】

なお、以上の設定は、ユーザが行わない場合は、デフォルト値を設定しておいてもよい。また、ウェブページの管理を行うオペレータが設定してもよい。なお、レイアウト変更 9 5 3 の処理においては、リンクキーワードテーブルに保存された、リンクとキーワードとの関係を利用してよい。

【 0 0 5 4 】

また、リンクをストックしておき、ストックされたリンクも含めて、ユーザの嗜好に合うページを選定しウェブページを構築してもよい。また、検索エンジンに、重要度の高いキーワードを検索させて、ウェブページ上の領域に検索結果を表示させてもよい。

【 0 0 5 5 】

そして、このレイアウトのデータは、特定のユーザのユーザ ID (または、ユーザ集合 ID)、特定のユーザ (または、特定のユーザ集合) の嗜好するキーワードなどと共に保存しておいてもよい (9 6 0)。

【 0 0 5 6 】

図 9 B は、上述のレイアウト変更の処理により図 3 (a) に示されたウェブページのレイアウトを変更したウェブページを示す。例えば、リンク 1 は、図 6 (c) に示すように、1 7 % という高いリンク注目度を有しており、これに対してリンク 3 は、図 6 (c) に示すように、4 % と低いリンク注目度を有していた。その結果、図 9 B では、領域 9 0 1 と領域 9 0 3 のリンクが入れ替わり、領域 9 0 3 にリンク 1 が位置するよう、ウェブページのレイアウトが変更されている。領域 9 0 3 は、図 6 (b) に示すように、クリック比率の予測値が高い領域であり、クリックしやすい領域である。この領域 9 0 3 に、ウェブページで最もリンク注目度の高いリンク 1 を配置したことになる。このレイアウトの変更により、ユーザにとってより使いやすいマイウェブページが構築できたことになる。そして、このマイウェブページを、対応するユーザに提供すればよい。

【 0 0 5 7 】

なお、レイアウトを固定したい領域がある場合には、レイアウト変更部において、レイアウトが変更できないように、その領域を指定しておけばよい。

< システム全体の構成 >

図 1 0 は、以上説明した各実施形態における機能的ブロック図の一例を示している。ウェブサーバ装置 1 2 0 からのログ情報やイベント情報は、ウェブコンテンツサーバ装置 1 4 0 のクリック回数カウント部 1 0 1 0 に入力され、各リンクに対するクリック回数がカウントされる。このカウント回数はリンク注目度特定部 8 1 0 に送られる。

【 0 0 5 8 】

クリック比率予測値算出部 1 0 2 0 は、レイアウト変更部 9 5 0 からのレイアウト情報を用いて、各リンクのクリック比率の予測値を算出する。この予測値は、リンク注目度特定部 8 1 0 に送られる。

【 0 0 5 9 】

リンク注目度特定部 8 1 0 は、これらの入力を受けて、各リンクの注目度を特定する。そして、リンク注目度特定部 8 1 0 で特定されたリンク注目度データが出力され、レイアウト変更部 9 5 0 に入力される。なお、このリンク注目度データは、ウェブページを構築する際の重要なデータとして、ウェブデザインなどに利用してもよい。

【 0 0 6 0 】

なお、リンク注目度正規化部 8 1 1 は、例えば複数のウェブページから得られたリンク注目度データを統合する処理を行う。具体的には、例えば、上述したように図 8 B に示した処理を行い、各ウェブページの総クリック数を用いて、リンク注目度データの値を調整し、複数のウェブページに存在するリンクのリンク注目度を統合する。この統合されたデータは、リンク注目度特定部 8 1 0 に返される。

【 0 0 6 1 】

キーワード抽出部 7 1 0 は、ウェブレイアウト情報や、実際のウェブページのデータを用いて図 7 に示した処理を実行することにより、各リンクのキーワードを抽出し、リンク

10

20

30

40

50

キーワードテーブル 720 に保存する。

【0062】

キーワード重要度特定部 830 は、リンクキーワードテーブル 720 の情報と、リンク注目度データとを利用して、各キーワードの重要度を特定する。この処理の一例については、図 8A を用いて説明した。

【0063】

レイアウト変更部 950 は、リンク注目度データ、キーワード重要度データ、リンクキーワードテーブル 720 の情報等を用いて、ウェブページのレイアウトを変更する。この処理については、図 9A を用いて一例を示した。レイアウト変更部 950 には、レイアウト情報等を格納したウェブレイアウトデータ (1) 及びウェブレイアウトデータ (2) を保存しておいてもよい。

10

【0064】

レイアウト変更部 950 で変更されたウェブページのレイアウト情報は、ウェブサーバ装置 120、クリック比率予測値算出部 1020、キーワード抽出部 710 等で利用される。

【0065】

図 11 は、一実施形態としての方法のフローを示している。なお、添付の請求項に係る方法は、矛盾のない限り、各ステップの順序を入れ替えて実施することができる。また、ステップの順序を入れ替えた方法も、添付の請求項に係る方法の発明の技術的範囲に属することは言うまでもない。

20

【0066】

ステップ 1102 において、ウェブページの各リンクに対するユーザのクリック回数をカウントする。

【0067】

ステップ 1104 において、各リンクに対するクリックの比率の予測値を算出する。この算出に当たっては、ウェブページのコントラストの分布やマウスカーソルの移動のしやすさに基づいて算出を行ってもよい。

【0068】

ステップ 1106 において、各リンクの注目度を特定する。ウェブページの各リンクの実際のクリック数とクリックの比率の予測値から、この注目度を特定してもよい。そして、この各リンクの注目度は、外部に出力してもよい。

30

【0069】

ステップ 1108 において、各リンクに対するキーワードの抽出を行ってもよい。キーワードは、リンクのテキスト情報、オルト属性から、抽出してもよい。またリンク先のページのテキスト情報やメタタグ情報を解析し、キーワードを抽出してもよい。

【0070】

ステップ 1110 において、各キーワードの重要度が特定される。このキーワードの重要度の情報は、外部に出力してもよい。

【0071】

ステップ 1112 において、ウェブページのレイアウトの変更を行ってもよい。レイアウトの変更は、各リンクの注目度及び各キーワードの重要度のうちの少なくともいずれか一方に基づいて、実行されてもよい。

40

【0072】

なお、本実施例で説明したリンク注目度及びキーワード重要度の取得、並びに、ウェブページのレイアウトの変更に係る各処理に関しては、必ずしもリアルタイム処理が要求されるわけではない。すなわち、ウェブページのアクセスに伴うログ情報やイベント情報を含む情報を用いて、ウェブコンテンツ管理サーバ装置 140 は、必要なタイミングやインターバルで処理を行い、その処理結果を保存しておけばよい。そして、リンク注目度データ、キーワード重要度データ、ウェブページレイアウト情報等を必要なタイミングで、記憶装置から読み出し、利用すればよい。

50

【 0 0 7 3 】

以上、各実施形態について説明したが、それぞれの実施形態は、排他的なものではなく、各実施形態の構成要素を取捨選択し、必要に応じて適宜組み合わせることができる点に留意すべきである。

【 0 0 7 4 】

以上の実施形態に関し、更に以下の付記を開示する。

(付記 1)

ウェブページの各リンクに対するユーザの指示回数をカウントする指示回数カウント部と、

前記ウェブページから、前記各リンクに対する、指示の比率の予測値を算出する指示比率予測値算出部と、

前記指示回数と、前記指示の比率の予測値とに基づいて、前記各リンクの注目度を特定するリンク注目度特定部と、

を有することを特徴とするサーバ装置。

(付記 2)

前記各リンクに関連する 1 つ以上のキーワードを抽出するキーワード抽出部と、

前記各リンクに対応する、前記注目度と、前記各キーワードとに基づいて、前記各キーワードの重要度を特定するキーワード重要度特定部と、

を更に有することを特徴とする付記 1 に記載のサーバ装置。

(付記 3)

前記リンク注目度特定部は、

【 数 2 】

$$F_m = \frac{n_m}{\sum_{k=1}^M n_k} - \beta \alpha_m \quad (1)$$

ここで、

F_m は、全てのリンクの数を M 個とした場合の m 番目のリンクの注目度、

n_m は、前記 m 番目のリンクの指示回数、

α_m は、前記 m 番目のリンクの指示の比率の予測値、

β は $0 < \beta < 1$ を満足する定数、

によってリンクの注目度を算出することを特徴とする付記 1 又は 2 に記載のサーバ装置。

(付記 4)

前記指示の比率の予測値は、前記各リンクの、面積、位置、及び背景とのコントラスト比のうち少なくともいずれか一つに基づき決定される、付記 1 乃至 3 のうちいずれか 1 項に記載のサーバ装置。

(付記 5)

前記各キーワードは、前記リンクを構成する文字列を含む情報から、及び / 又は、前記リンクが指し示すコンテンツから、抽出することを特徴とする付記 1 乃至 4 のうちいずれか 1 項に記載のサーバ装置。

(付記 6)

複数のウェブページの各々の指示回数に応じてリンクの注目度を重み付けすることによって、前記複数のウェブページに存在する前記リンクの各々の注目度を正規化するリンク注目度正規化部、

を更に有することを特徴とする付記 1 乃至 5 のうちいずれか 1 項記載のサーバ装置。

(付記 7)

10

20

30

40

50

前記ユーザは、全てのユーザ、女性のユーザ、男性のユーザ、特定の年齢層のユーザ、特定の時間帯に前記ウェブページにアクセスするユーザのうちのいずれか一つである、付記 1 乃至 6 のうちいずれか 1 項記載のサーバ装置。

(付記 8)

コンピュータにより実行される方法であって、
ウェブページの各リンクに対するユーザの指示回数をカウントし、
前記ウェブページから、前記各リンクに対する、指示の比率の予測値を算出し、
前記指示回数と、前記指示の比率の予測値とに基づいて、前記各リンクの注目度を特定することを特徴とする方法。

(付記 9)

前記各リンクに関連する 1 つ以上のキーワードを抽出し、
前記各リンクに対応する、前記注目度と、前記各キーワードとに基づいて、前記各キーワードの重要度を特定することを特徴とする付記 8 に記載の方法。

(付記 10)

前記リンク注目度を特定する処理は、

【数 3】

$$F_m = \frac{n_m}{\sum_{k=1}^M n_k} - \beta \alpha_m \quad (1)$$

10

20

ここで、

F_m は、全てのリンクの数を M 個とした場合の m 番目のリンクの注目度、

n_m は、前記 m 番目のリンクの指示回数、

α_m は、前記 m 番目のリンクの指示の比率の予測値、

β は $0 < \beta < 1$ を満足する定数、

によってリンクの注目度を算出することを特徴とする付記 8 又は 9 に記載の方法。

30

(付記 11)

前記指示の比率の予測値は、前記各リンクの、面積、位置、及び背景とのコントラスト比のうち少なくともいずれか一つに基づき決定される、付記 8 乃至 10 のうちいずれか 1 項に記載の方法。

(付記 12)

前記各キーワードは、前記リンクを構成する文字列を含む情報から、及び / 又は、前記リンクが指し示すコンテンツから、抽出することを特徴とする付記 8 乃至 11 のうちいずれか 1 項に記載の方法。

(付記 13)

複数のウェブページの各々の指示回数に応じてリンクの注目度を重み付けすることによって、前記複数のウェブページに存在する前記リンクの各々の注目度を正規化することを特徴とする付記 8 乃至 12 のうちいずれか 1 項記載の方法。

40

(付記 14)

前記ユーザは、全てのユーザ、女性のユーザ、男性のユーザ、特定の年齢層のユーザ、特定の時間帯に前記ウェブページにアクセスするユーザのうちのいずれか一つである、付記 8 乃至 13 のうちいずれか 1 項記載の方法。

(付記 15)

ウェブページの各リンクに対するユーザの指示回数をカウントし、

前記ウェブページから、前記各リンクに対する、指示の比率の予測値を算出し、

前記指示回数と、前記指示の比率の予測値とに基づいて、前記各リンクの注目度を特定

50

する、

処理をコンピュータに実行させるプログラム。

(付記 16)

前記各リンクに関連する 1 つ以上のキーワードを抽出し、

前記各リンクに対応する、前記注目度と、前記各キーワードとに基づいて、前記各キーワードの重要度を特定する、

処理を更にコンピュータに実行させることを特徴とする付記 15 に記載のプログラム。

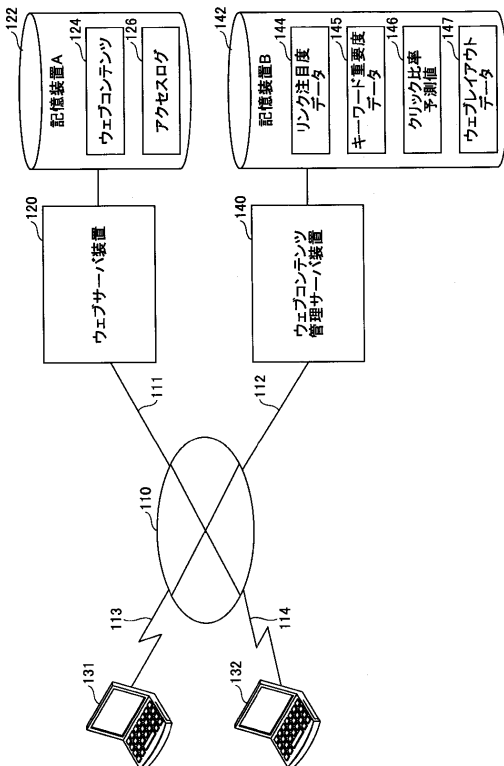
【符号の説明】

【0075】

- 110 ネットワーク
- 120 ウェブサーバ装置
- 122 記憶装置 A
- 124 ウェブコンテンツ
- 126 アクセスログ
- 140 ウェブコンテンツ管理サーバ装置
- 142 記憶装置 B
- 144 リンク注目度
- 145 キーワード重要度
- 146 クリック比率
- 147 ウェブレイアウト

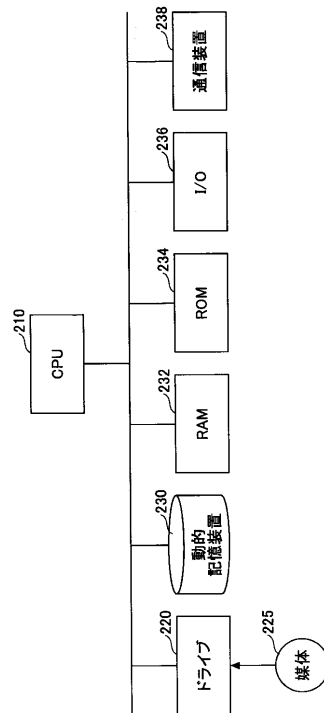
【図 1】

本発明の一実施形態のシステムの構成を示す図



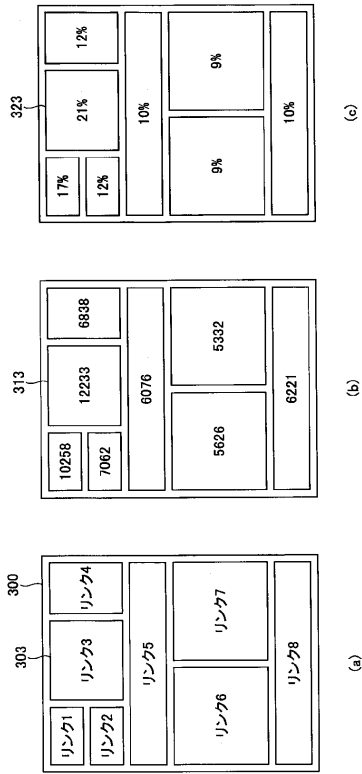
【図 2】

サーバ装置のハードウェア構成を示す図



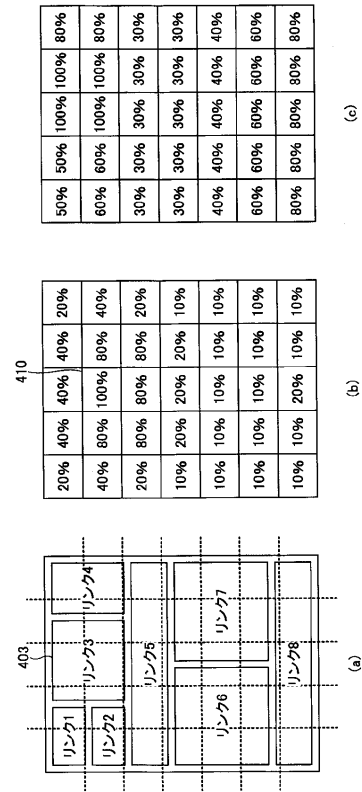
【図3】

ウェブページのリンク毎のクリック回数・クリック比率の例を示す図



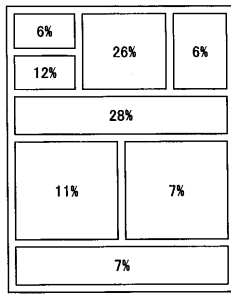
【図4】

ウェブページを分割しクリック比率の予測値を算出する例を示す図



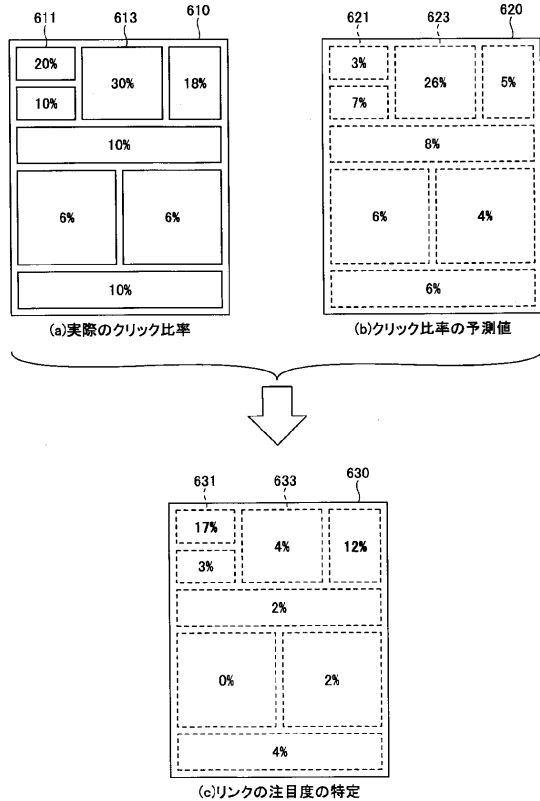
【図5】

リンク毎のクリック比率の予測値の算出結果の例を示す図



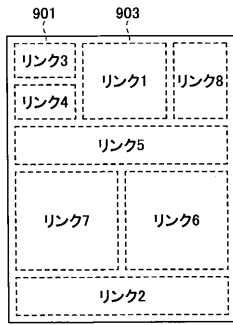
【図6】

リンクの注目度を特定する手法の例を示す図



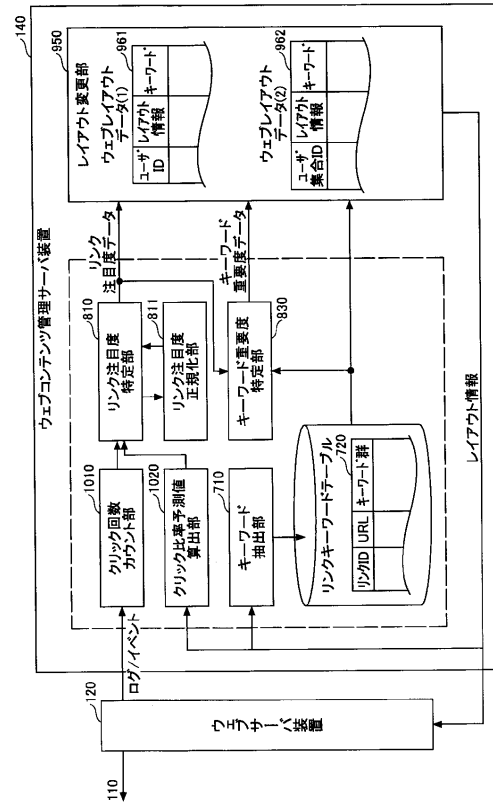
【 図 9 B 】

ウェブページレイアウトの変更の結果の例を示す図



【 図 1 0 】

各実施形態における機能的ブロック図の一例を示す図



【 図 1 1 】

一実施形態としての方法のフローを示す図

