

(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2006-190253

(P2006-190253A)

(43) 公開日 平成18年7月20日(2006.7.20)

| | | | | |
|-------------------|------------------|------------|------|-------------|
| (51) Int. Cl. | | F I | | テーマコード (参考) |
| G06F 12/00 | (2006.01) | G06F 12/00 | 546B | 5B082 |
| G06F 15/00 | (2006.01) | G06F 15/00 | 310A | 5B185 |

審査請求 未請求 請求項の数 20 O L (全 14 頁)

| | | | |
|--------------|------------------------------|----------|---------------------|
| (21) 出願番号 | 特願2005-326364 (P2005-326364) | (71) 出願人 | 500046438 |
| (22) 出願日 | 平成17年11月10日 (2005.11.10) | | マイクロソフト コーポレーション |
| (31) 優先権主張番号 | 11/027,661 | | アメリカ合衆国 ワシントン州 9805 |
| (32) 優先日 | 平成16年12月30日 (2004.12.30) | | 2-6399 レッドモンド ワン マイ |
| (33) 優先権主張国 | 米国 (US) | | クロソフト ウェイ |
| | | (74) 代理人 | 100077481 |
| | | | 弁理士 谷 義一 |
| | | (74) 代理人 | 100088915 |
| | | | 弁理士 阿部 和夫 |
| | | (72) 発明者 | マイケル エー. スターバード |
| | | | アメリカ合衆国 98052 ワシントン |
| | | | 州 レッドモンド ワン マイクロソフト |
| | | | ウェイ マイクロソフト コーポレーシ |
| | | | ョン内 |
| | | Fターム(参考) | 5B082 EA01 |
| | | | 5B185 AA01 |

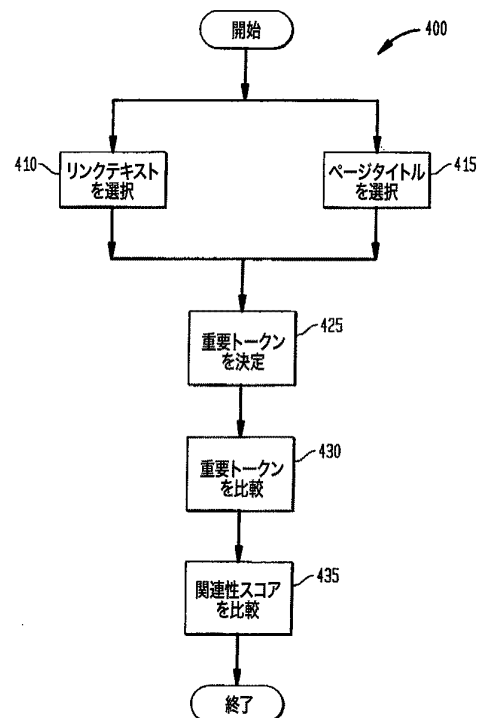
(54) 【発明の名称】 ウェブページのアスペクトを評価するための方法と装置

(57) 【要約】

【課題】ウェブページ上のハイパーテキストリンクを指し示すリンクテキストが、そのリンクによって指し示されるウェブページにどの程度対応しているかを評価するための方法と装置を提供すること。

【解決手段】一実施形態では、リンクテキストは、リンクテキストとページタイトルとを個々のトークンに解析し、これらのトークンを比較する等の方法によって、リンクにより指し示されるウェブページのタイトルと比較される。リンクテキストとページタイトルとがどの程度対応しているかは、トークン中の一致するパーセンテージとして表わすことができる。グラフィカル・ユーザ・インタフェース(GUI)を設置し、トークンの最低パーセンテージが一致していないときに視覚的表示を行うようにしてもよい。

【選択図】 図4



【特許請求の範囲】**【請求項 1】**

第一のウェブページに含まれるハイパーテキストリンクを評価するための自動化された方法であって、前記リンクはウェブリソースを指し示し、

(A) 前記リンクの特徴が前記ウェブリソースの特徴に十分に対応するか否かを判断する動作を有することを特徴とする方法。

【請求項 2】

請求項 1 に記載の方法であって、

前記ウェブリソースは第二のウェブページであり、前記ウェブリソースの特徴は前記第二のウェブページのタイトルであり、前記リンクの特徴は、前記第一のウェブページ上の前記リンクを指し示すテキストであることを特徴とする方法。

10

【請求項 3】

請求項 2 に記載の方法であって、

前記動作 (A) は、

(A1) 前記第一のウェブページ上の前記リンクを指し示す前記テキストを少なくとも一つのトークンを含む第一のトークンリストに解析する動作と、

(A2) 前記第二のウェブページの前記タイトルを少なくとも一つのトークンを含む第二のトークンリストに解析する動作と、

(A3) 前記第一のトークンリストを前記第二のトークンリストと比較する動作と、
を含むことを特徴とする方法。

20

【請求項 4】

請求項 3 に記載の方法であって、

前記動作 (A3) は、

前記第一のトークンリストから第一のトークンを選択する動作と、

前記第二のトークンリストから第二のトークンを選択する動作と、

前記第一と第二のトークンのいずれが大きい方のトークンで、いずれが小さい方のトークンかを判断する動作と、

前記大きい方のトークンの基準パーセンテージに相当する部分を特定する動作と、

前記基準パーセンテージが前記小さい方のトークンに含まれるか否かを判断する動作と

30

をさらに含むことを特徴とする方法。

【請求項 5】

請求項 3 に記載の方法であって、

前記動作 (A1) は、前記第一のトークンリスト内のトークンの各々を、非重要トークンの集合と比較することにより、前記第一のトークンリストから第一の重要トークンリストを決定する動作をさらに含み、前記動作 (A2) は、前記第二のトークンリスト内のトークンの各々を非重要トークンの集合と比較することにより、前記第二のトークンリストから第二の重要トークンリストを決定する動作をさらに含み、前記動作 (A3) は、前記第一の重要トークンリストを前記第二の重要トークンリストと比較する動作をさらに含むことを特徴とする方法。

40

【請求項 6】

請求項 1 に記載の方法であって、

(B) 前記動作 (A) の決定の結果を、グラフィカル・ユーザ・インタフェース (GUI) 上で表示する動作をさらに有することを特徴とする方法。

【請求項 7】

請求項 6 に記載の方法であって、

前記動作 (B) は、前記リストの特徴が前記ウェブリソースの特徴に十分に対応しないと判断された場合、前記 GUI 上で視覚的表示を提供する動作をさらに含むことを特徴とする方法。

50

【請求項 8】

実行されると、ウェブリソースを指し示す、第一のウェブページに含まれるハイパーテキストリンクを評価する方法を遂行する命令が符号化されたコンピュータ読取可能媒体であって、前記方法は、

(A) 前記リンクの特徴が前記ウェブリソースの特徴に十分に対応するか否かを判断する動作を有することを特徴とするコンピュータ読取可能媒体。

【請求項 9】

請求項 8 に記載のコンピュータ読取可能媒体であって、

前記ウェブリソースは第二のウェブページであり、前記ウェブリソースの特徴は前記第二のウェブページのタイトルであり、前記リンクの特徴は前記第一のウェブページ上で前記リンクを指し示すテキストであることを特徴とするコンピュータ読取可能媒体。

10

【請求項 10】

請求項 9 に記載のコンピュータ読取可能媒体であって、

前記動作 (A) は、

(A1) 前記第一のウェブページ上の前記リンクを指し示す前記テキストを少なくとも 1 つのトークンを含む第一のトークンリストに解析する動作と、

(A2) 前記第二のウェブページの前記タイトルを少なくとも 1 つのトークンを含む第二のトークンリストに解析する動作と、

(A3) 前記第一のトークンリストを前記第二のトークンリストと比較する動作と、
をさらに含むことを特徴とする読取可能媒体。

20

【請求項 11】

請求項 10 に記載のコンピュータ読取可能媒体であって、

前記動作 (A3) は、

前記第一のトークンリストから第一のトークンを選択する動作と、

前記第二のトークンリストから第二のトークンを選択する動作と、

前記第一と第二のトークンのどちらが大きい方のトークンであり、どちらが短いほうのトークンかを判断する動作と、

前記大きい方のトークンの基準パーセンテージに相当する部分を特定する動作と、

前記基準パーセンテージが前記小さい方のトークンに含まれるか否かを判断する動作と

30

、
をさらに含むことを特徴とするコンピュータ読取可能媒体。

【請求項 12】

請求項 10 に記載のコンピュータ読取可能媒体であって、

前記動作 (A1) は、前記第一のトークンリスト内のトークンの各々を、非重要トークンの集合と比較することにより、前記第一のトークンリストから第一の重要トークンリストを決定する動作をさらに含み、前記動作 (A2) は、前記第二のトークンリスト内のトークンの各々を非重要トークンの集合と比較することにより、前記第二のトークンリストから第二の重要トークンリストを決定する動作をさらに含み、前記動作 (A3) は、前記第一の重要トークンリストを前記第二の重要トークンリストと比較する動作をさらに含むことを特徴とするコンピュータ読取可能媒体。

40

【請求項 13】

請求項 8 に記載のコンピュータ読取可能媒体であって、

(B) 前記動作 (A) の決定の結果をグラフィカル・ユーザ・インタフェース (GUI) に表示する動作をさらに含むことを特徴とするコンピュータ読取可能媒体。

【請求項 14】

請求項 13 に記載のコンピュータ読取可能媒体であって、

前記動作 (B) は、前記リンクの特徴が前記ウェブリソースの特徴に十分に対応していないと判断された場合、前記 GUI 上で視覚的表示を提供する動作をさらに含むことを特徴とするコンピュータ読取可能媒体。

【請求項 15】

50

ウェブリソースを指し示す、第一のウェブページに含まれるハイパーテキストリンクを評価するためのシステムであって、前記システムは、

前記リンクの特徴が前記ウェブリソースの特徴に十分対応するか否かを判断する判断コントローラを備えたことを特徴とするシステム。

【請求項 16】

請求項 15 に記載のシステムであって、

前記第一のウェブページ上で前記リンクを指し示す前記テキストを少なくとも 1 つのトークンを含む第一のトークンリストに解析するリンクテキスト解析コントローラと、

前記第二のウェブページ上の前記タイトルを少なくとも 1 つのトークンを含む第二のトークンリストに解析するページタイトル解析コントローラと、

前記第一のトークンリストを前記第二のトークンリストと比較する比較コントローラと、

をさらに備えたことを特徴とするシステム。

【請求項 17】

請求項 16 に記載のシステムであって、前記比較コントローラはさらに、

前記第一のトークンリストから第一のトークンを選択し、

前記第二のトークンリストから第二のトークンを選択し、

前記第一と第二のトークンのいずれが大きい方のトークンで、いずれが小さい方のトークンかを判断し、

前記大きい方のトークンの基準パーセンテージに相当する部分を特定し、

前記基準パーセンテージが前記小さい方のトークンに含まれるか否かを判断することを特徴とするシステム。

【請求項 18】

請求項 16 に記載のシステムであって、

前記リンクテキスト解析コントローラはさらに、前記第一のトークンリストのトークンの各々を非重要トークンの集合と比較することによって前記第一のトークンリストから第一の重要トークンリストを決定し、前記ページタイトル解析コントローラはさらに、前記第二のトークンリストのトークンの各々を非重要トークンリストと比較することによって前記第二のトークンリストから第二の重要トークンリストを決定し、前記比較コントローラはさらに、前記第一の重要トークンリストを前記第二の重要トークンリストと比較することを特徴とするシステム。

【請求項 19】

請求項 15 に記載のシステムであって、

前記決定の結果をグラフィカル・ユーザ・インタフェース (GUI) に表示する表示コントローラをさらに備えたことを特徴とするシステム。

【請求項 20】

請求項 19 に記載のシステムであって、

前記表示コントローラは、前記リンクの特徴が前記ウェブリソースの特徴に十分に対応していないと判断されると、前記 GUI 上で視覚的表示を提供することを特徴とするシステム。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、コンピュータソフトウェアに関し、より詳しくは、ウェブページのアスペクトの評価に使用できるソフトウェアに関する。

【背景技術】

【0002】

多くの人々がインターネットを用いてワールド・ワイド・ウェブ (「ウェブ」) を利用する。ウェブ環境において、サーバコンピュータはクライアントコンピュータによって要求される情報をウェブページの形で提供する。ウェブページには、他の情報に加え、ハイ

10

20

30

40

50

パーテキスト・マークアップ・ランゲージ (HTML) またはエクステンシブル・マークアップ・ランゲージ (XML) 等のマークアップ言語で提供される命令の集合、つまり「タグ」が含まれる。クライアントコンピュータ上で実行されるブラウザプログラムが、ページに含まれるタグを受け取り、処理し、ユーザのためのディスプレイを作る。タグは、例えば、ページの要素の表示態様を定義するかもしれない。

【0003】

タグはまた、ハイパーテキストリンク (以下、「リンク」という) も定義する場合がある。リンクは、ユニフォーム・リソース・ロケータ (URL) を介して、別のウェブページ等、他のウェブリソースを特定する。リンクは、英数字 (「リンクテキスト」) によってウェブページ上に表示されるかもしれない。リンクテキストは通常、そのリンクがユーザによって容易に認識されるようにウェブページ上に提示される。例えば、多くのリンクは太字または下線付きのテキストでページ上に表示される。ユーザは、例えば、それを「クリックする」(例えば、マウスでカーソルをリンクの上に移動させ、マウスのボタンを押す) ことによってリンクを呼び出すことができる。リンクをクリックすると、そのリンクによって定義される URL のウェブリソースにアクセスするよう求めるリクエストがサーバコンピュータに発行される。

10

【発明の開示】

【発明が解決しようとする課題】

【0004】

論理的に関係のあるウェブページのグループは一般に、ウェブサイトと呼ばれる。一部のウェブサイトは、保守するのが面倒な場合がある。例えば、ウェブページ上のリンクによって定義される URL は、特定のウェブリソースの URL が変更される可能性があるため、時間がたつと使われなくなってしまうかもしれない。またウェブリソースが削除されているかもしれない。ウェブサイトの保守を支援するために、管理者やその他のユーザがウェブサイトのページに掲載されるリンクを管理できるような、さまざまな自動ツールが出現している。これらのツールは、例えば、サイトのページに掲載されるリンクが現在存在する URL を定義しているか否かをユーザが判断するのを助ける。また、このツールは、ユーザがサイト内のリンクの配置を見ることができるよう、グラフィカル・ユーザ・インタフェース (GUI) も提供する。

20

【課題を解決するための手段】

30

【0005】

一態様によれば、第一のウェブページに掲載されているハイパーテキストリンクを評価するための自動化された方法が提供され、このリンクはあるウェブリソースを指し示す。この自動化された方法は、リンクの特徴がそのウェブリソースの特徴に十分に対応しているか否かを判断するステップを含む。

【0006】

別の態様によれば、実行されると、あるウェブリソースを指し示す、第一のウェブページに掲載されるハイパーテキストリンクの評価方法を遂行する命令が符号化されたコンピュータ読取可能媒体が提供される。この方法は、そのリンクの特徴がそのウェブリソースの特徴に十分に対応しているか否かを判断するステップを含む。

40

【0007】

さらに別の態様によれば、あるウェブリソースを指し示す、第一のウェブページに掲載されるハイパーテキストリンクを評価するためのシステムが提供される。システムは、そのリンクの特徴がそのウェブリソースの特徴に十分に対応しているか否かを判断する判断コントローラを備える。

【0008】

添付の図面は、縮尺比に従って拡大縮小したものではない。図中、複数の図面に示されている同一のコンポーネントには同様の参照番号が付与されている。すべての図面のすべてのコンポーネントに名前が付けられているとは限らない。

【発明を実施するための最良の形態】

50

【0009】

ウェブページ上のリンクが、そこにリソースが確かに存在するURLを定義しているか否かを判断するために使用できる数多くのユーティリティはあるが、リンクによって定義されるURLに存在するリソース（例えば、ウェブページ）がそのページ上にあるリンクテキストに十分に対応しているか否かを判断するユーティリティは皆無である。つまり、リンクテキストをそのリンクが実際に指し示しているリソースと比較し、そのリンクが、それが指し示そうとしているリソースを指し示しているか否かを判断するユーティリティはない。

【0010】

したがって、本発明の一実施形態では、リンクテキストがそのリンクによって指し示されたウェブページにどの程度対応しているかを評価するための、自動化された方法を提供する。一実施形態では、リンクテキストは、そのリンクによって指し示されたウェブページのタイトルと比較される。一実施形態では、リンクテキストとページタイトルの各々が個々の「トークン」に解析され、これらのトークンを比較して、そのリンクテキストとページタイトルがどの程度対応しているかを判断する。一実施形態では、リンクテキスト内に見られる個々のトークンの各々を、ページタイトルに見られる各トークンと第一のアルゴリズムにしたがって比較し、それらが一致するか否かを判断する。一実施形態では、その後、リンクテキストとページタイトルの間の関連性を、リンクテキストまたはタイトルページのいずれかの一方の全トークンの内、他方のリストのトークンと一致する割合として表現してもよい。

10

20

【0011】

本発明の実施形態は、例えば、ウェブページに掲載されるリンクの全体的な有効性を判断する、自動化されたユーティリティによって用いられてもよい。例えば、実施形態は、ウェブページに掲載されているリンクが有効な、または現存するURLを定義しているか否かだけでなく、各リンクが、それが指し示そうとしているリソースを指し示しているかを評価するユーティリティによって用いられるかもしれない。この評価の結果は、グラフィカル・ユーザ・インタフェース（GUI）を介してユーザに提示されるかもしれない。このため、ユーザはあるページに掲載されたリンクの全体的有効性をより効果的に判断することができる。しかしながら、本発明は、その態様に多くの応用の可能性があることから、これらの使用に限定されないものと認識すべきである。例えば、本発明の態様はブラウザプログラムによって採用されるかもしれず、リンクが指し示そうとしているページを明らかに指し示していないリンクについてユーザに警告する役割を果たすかもしれない。

30

【0012】

本発明の各種の態様は、図1に示されているようなコンピュータシステムの一例100に代表される、1つまたは複数のコンピュータシステムによって実現されるかもしれない。コンピュータシステム100は、入力装置102、出力装置101、プロセッサ103、メモリシステム104、記憶装置106を備え、これらはすべて直接または、相互接続機構105を介して連結されており、相互接続機構は、1つまたは複数のバス、スイッチおよび/またはネットワークからなる。入力装置102は、ユーザまたは機械（例えば、人間のオペレータまたは電話の受話器）からの入力を受け、出力装置101は、ユーザまたは機械（例えば、液晶ディスプレイ）に情報を表示または伝送する。プロセッサ103は一般に、オペレーティングシステム（例えば、Microsoft Windows（登録商標）ファミリーのオペレーティングシステムまたはその他の適当なオペレーティングシステム）と呼ばれる、他のコンピュータプログラムの実行を制御するコンピュータプログラムを動作させ、スケジューリング、入力/出力その他の装置の制御、アカウントティング、編集、記憶域割当、データ管理、メモリ管理、通信およびデータフロー管理を提供する。プロセッサとオペレーティングシステムは、集合的に、他のコンピュータプログラミング言語で書かれたアプリケーションプログラムを持つコンピュータプラットフォームを画定する。

40

50

【0013】

プロセッサ103はまた、1つまたは複数のコンピュータプログラムを実行し、各種の機能を実現することもできる。これらのコンピュータプログラムは、手続き型プログラミング言語、オブジェクト指向型プログラミング言語、マクロ言語またはこれらの組み合わせをはじめとする、いずれのタイプのコンピュータプログラミング言語で書かれていてもよい。これらのコンピュータプログラムは、記憶システム106に記憶できる。記憶システム106は、揮発性または不揮発性媒体上に情報を保持してもよく、固定されていても取り外し可能でもよい。記憶システム106を図2にさらに詳しく示す。

【0014】

記憶システム106は通常、コンピュータ読取可能およびコンピュータ書込み可能不揮発性記録媒体201を含み、その上に、コンピュータプログラムまたはそのプログラムによって使用される情報を画定する信号が記憶される。媒体は、例えば、ディスクまたはフラッシュメモリでもよい。一般に、動作中、プロセッサ103により、データは不揮発性記録媒体201から、プロセッサ103による情報アクセスが媒体201より高速となる揮発性メモリ202（例えば、ランダムアクセスメモリ、つまりRAM）に読み出される。このメモリ202は、図2に示すように、記憶システム106の中、あるいは図1に示すようにメモリシステム104の中に設置できる。プロセッサ103は一般に、集積回路メモリ104、202内のデータを操作し、次に、処理が完了した後で、このデータを媒体201に複写する。媒体201と集積回路メモリ素子104、202との間のデータの移動を管理するものとして知られている機構は各種あり、本発明はそれに限定されない。本発明は、特定のメモリシステム104または記憶システム106にも限定されない。

【0015】

前述のように、本発明の一実施形態は、ウェブページ上のリンクを特徴付けるテキストがそのリンクによって指し示されるリソースにどの程度対応しているかを評価するための、コンピュータシステム100によって実行されるかもしれない自動化された方法を提供する。本発明の実施形態によって評価されるリンクを含むウェブページの例を図3A、3Bに示す。具体的には、図3Aは、ウェブページ302を提示するブラウザインタフェース301を示し、図3Bは、ウェブページ304を提示するブラウザインタフェース303を示す。

【0016】

ウェブページ302は、グラフィクス、テキストおよびリンク305、310、315、320等、ウェブページに一般的に見られる各種要素を含む。ウェブページ302はまた、メニュー部330を含み、これには、“Developer Tools”と題するリンク331を含む多数の追加リンクが表示されている。ユーザがリンク331を呼び出すと（例えば、カーソルをリンク331の上に移動させ、マウスのボタンを押すか、「エンター」キーを打つことによる）、ブラウザはウェブページ304へのアクセスのリクエストを発行するかもしれない。

【0017】

ウェブページ304を図3Bに示す。ウェブページ304はウェブページ302と多くの点で似ている。例えば、ウェブページ304はリンク305と310を含み、これらはウェブページ302においても提示されている。ウェブページ304にはまた、リンク340、342、344等も掲載されている。ウェブページ304には、インタフェース303の最上部に表示されるテキスト“MSDN Home Page”によって表わされるタイトル350が含まれている。

【0018】

ウェブページに含まれるリンクを評価するための技術の一例を、図4-6を参照しながら以下に説明する。図4-6の各々は、この技術を段階的に、より詳しく説明するフローチャートである。図4は、技術全体を示すフローチャートである。図5は、リンクテキストとページタイトルに見られる個々のトークンを比較する動作を詳細に説明するフローチャートである。最後に、図6はその比較をさらに詳細に説明するフローチャートである。

【0019】

まず図4において、プロセス400の初めに、動作410と415が開始される。動作410において、リンクテキストが評価のために選択される。これは、例えばリンクテキストをメモリ内に読み込むこと等、どのような適当な方法でも実行できる。一実施形態において、動作410の結果、リンクテキストを構成するトークン（つまり、個々の単語または文字ストリング）の集合である「トークンリスト」が作成される。一実施形態において、リスト内の各トークンは、「blank」または「space」文字によって分離され、または境界を付けられる。リンク331（図3A）の例を使うと、リンクテキスト“Developer Tools”から、動作410の結果、“Developer”と“Tools”というトークンを含むトークンリストが作成される。

10

【0020】

動作415において、プロセスはリンクにより指し示されるページのタイトルを判断しようとする。これはまた、例えば指し示されたページにアクセスするリクエストを発行する等、どのような適当な方法でも実行できる。動作410と同様に、動作420の結果、トークンリストが作成される。ページ304（つまり、ユーザがリンク331を呼び出したときに表示されるページ）のタイトル350（図3B）の例を使うと、ページタイトルは“MSDN Home Page”で、動作420の結果、“MSDN”、“Home”、“Page”のトークンを含むトークンリストが作成される。

【0021】

動作410と動作420が完了すると、プロセスは動作425に進み、ここで、各トークンリスト内の「重要トークン」が判断される。一実施形態において、各リストの重要トークンは、わかっている非重要トークンを除去することによって判断される。非重要トークンは、例えば、トークンリストを比較する上で有益性が低いとわかる単語である。つまり、非重要トークンがリンクテキストトークンリストとページタイトルトークンリストの両方に見られた場合でも、非重要トークンがトークンリスト間の一致を発生させるという事実は、リンクテキストトークンリストがページタイトルトークンリストに対応するか否かを判断する上で有益ではない。例えば、非重要トークンには、“the”、“and”および/またはその他の単語または文字の集合等がある。

20

【0022】

一実施形態において、非重要トークンを、実行中にプロセス400によってアクセスされるデータストラクチャに格納させることができる。一実施形態では、データストラクチャは設定可能で、ユーザがその中に表示されている非重要トークンの集合の追加、削除、あるいは変更を行うことができるようにしてもよい。非重要トークンの集合を設定する能力は、例えば、英語以外の言語のトークンで使用するためにリストを適応させるのに有益かもしれない。例えば、ユーザはフランス語のウェブサイトに表示されるリンクに対応するリンクテキストを評価するために、一般的なフランス語の代名詞の集合をリストに追加することができる。

30

【0023】

一実施形態において、動作425はまた、各トークンリストから特定の文字を除去することが含まれる。例えば、ピリオド、セミコロン、ハイフオン、アンパサンドおよび/またはその他の文字を各トークンリストから除去し、これら2つのリストをより効果的に比較しやすくすることができる。

40

【0024】

動作425が完了すると、プロセスは動作430に進み、ここで、重要トークンのリストが比較される。重要トークンのリストを比較する技術の一例を図5に示す。図5のプロセスにおいて、2つのトークンリストのうち短いほうを最初に選び、次に短いほうのリストの中の各トークンを長いほうのリストの各トークンと順番に比較する。

【0025】

プロセス500が始まると、プロセスは動作510に進み、ここで2つのトークンリストのうちの短いほうが判断される。これは、どのような適当な方法でも実行できる。例え

50

ば、一実施形態において、これは、トークンリストのどちらのトークンの数が少ないかを判断することによって行われる。別の実施形態では、どちらのトークンリストの文字が少ないかを判断することによって行われる。本発明は特定の実施例に限定されない。

【0026】

動作510が完了するとプロセスは動作515に進み、ここで、短いほうのリストからトークンが選択され(動作510で決定)、長いほうのリストのトークンと比較される。これは、どのような適当な方法でも実行できる。例えば、トークンはトークンリストからランダムに選択してもよい。

【0027】

動作515が完了すると、プロセスは動作520に進み、ここで、長いほうのリストから比較のための最初のトークンが選択される。動作515における選択と同様に、これは、どのような適当な方法でも実行できる。

【0028】

動作520が完了すると、プロセスは動作525に進み、ここで、短いほうのリストから選択されたトークンが、長いほうのリストから選択されたトークンと比較され、トークンが一致するか否かが判断される。動作525を実行するための技術の例を図6に示す。図6のプロセスを、2つのトークンの例、“referral”と“refers”の比較に関して以下に説明する。

【0029】

プロセス600が開始すると、プロセスは動作610に進み、ここで、2つのトークンのうち、大きい方と小さい方が判断される。これは、どのような適当な方法でも実行できる。例えば、文字数の少ないトークンを小さいトークンと判断し、文字数の多いトークンを大きいトークンと判断する。一実施形態において、トークンの文字数が同じであれば、大きい方のトークンと小さい方のトークンは、ランダムな順番で判断される。図の例では、プロセスは、大きい方のトークンが“referral”、小さい方のトークンが“refers”であると判断する。

【0030】

動作610が完了すると、プロセスは動作615に進み、ここで、大きい方のトークンの中で、大きい方のトークンの少なくとも「基準パーセンテージ」を構成するテキストが判断される。一実施形態において、基準パーセンテージは、大きい方のトークンのテキストのうち、小さい方のトークンとの比較に使用される部分である。一実施形態において、この部分は、大きい方のトークンの文字総数を明確にし、次に、トークン内の最初の文字から始まり、基準パーセンテージを満たす、またはこれを超える文字数を特定することによって決められる。図の例において、基準パーセンテージは60%であれば、大きい方のトークン、“referral”のうち基準パーセンテージにあたるテキストは“refer”である(つまり、“referral”の8文字のうちの5文字、すなわちテキストの62.5%である)。

【0031】

一実施形態において、基準パーセンテージは具体的な構成の必要性に適合するように(例えば、ユーザによって)設定可能である。例えば、GUIを設置し、ユーザが具体的な構成に適合するように基準パーセンテージを変更できるようにしてもよい。

【0032】

動作615が完了すると、プロセスは動作620に進み、ここで、動作615において特定されたテキストと小さい方のトークンの比較が行われる。一実施形態において、比較により、動作615において特定されるテキストが小さい方のトークンに含まれているか否かが判断される。図の例において、プロセスは、“refer”(動作615で判断される)が“refers”に含まれているか判断する。しかしながら、この比較は、本発明がこの点で限定されないことから、どのような適当な方法でも実行できる。

【0033】

動作620が完了すると、プロセス600は終了し、全体的プロセスはプロセス500

10

20

30

40

50

(図5)に戻る。つまり、図6のプロセスは動作525を実行するための技術の一例であるため、全体的プロセスは図5の動作525に戻る。

【0034】

動作525が完了すると、プロセスは動作530に進み、ここで、一致が見られるか否かが判断される。一実施形態では、動作620(図6)において、動作615で特定されたテキストが小さい方のトークンに含まれると判断されると一致が見られる。一致が見られると、プロセスは動作535に進み、ここで、一致の表示が記録される。表示は、例えば、メモリ内に記録される。

【0035】

一致が見られない場合、プロセスは動作545に進み、ここで、大きい方のトークンリストにさらに別のトークンが存在するか否かが判断される。大きい方のトークンリストにさらに別のトークンが存在すると判断されると、プロセスは動作520に戻り、大きい方のリスト内の次のトークンが選択される。このように、プロセスは小さい方のリストの中の各トークンと大きい方のリスト内のすべてのトークンとを比較する。

10

【0036】

動作545において、大きい方のリストにさらに別のトークンは存在しないと判断されると、プロセスは動作550に進み、短いほうのリストの中のトークンと大きい方のリストのトークンのいずれとも一致が見られないことが表示される。

【0037】

動作535と動作550のいずれかが完了したところで、プロセスは動作540に進み、ここで、小さい方のリストにさらに別のトークンが存在するか否かが判断される。存在しなければ、プロセスは終了する。短いほうのリストにさらに別のトークンが存在する場合、プロセスは動作515に戻り、短いほうのリストの次のトークンが比較のために選択される。このように、プロセスは短いほうのリストのすべてのトークンについての比較を繰り返す。

20

【0038】

小さい方のトークンリストと大きい方のトークンリストの両方のトークンがすべて比較されると、プロセス500は終了し、プロセス全体がプロセス400(図4)に戻る。つまり、図5のプロセスは、動作430を実行するための技術の一例であるため、プロセス全体は図4の動作430に戻る。

30

【0039】

動作430が完了すると、プロセス400は動作435に進み、ここで、関連性スコアを計算し、リンクテキストとページタイトルとがどの程度対応するかを画定する。一実施形態において、関連性スコアは、一致する重要トークンの数(動作620で判断)を短いほうのトークンリストの重要トークンの総数で割り、その商に100%を乗じて算出される。しかしながら、2つのトークンリストがどの程度対応しているかは、本発明がこの点で限定されないことから、どのような適当な方法でも判断できる。

【0040】

動作435が完了すると、プロセス400は終了する。

【0041】

一実施形態において、最低の関連性スコアは、2つのトークンリストが十分に対応しているか否かを決定してもよい。例えば、最低関連性スコアを70%と設定し、2つのトークンリストが「一致」するのに対応しなければならない程度を定義し、これによって(トークンリストによって表わされる)リンクテキストとページタイトルとが一致するか否かを判断してもよい。

40

【0042】

一実施形態において、前述の基準パーセンテージと同様に、トークンリスト間の十分な対応を画定する最低関連性スコアは、特定の構成の必要性に適合するように(例えば、ユーザによって)設定可能であってもよい。例えば、GUIを設置し、ユーザが具体的な構成に適合するような最低関連性スコアをカスタマイズできるようにしてもよい。

50

【0043】

一致しないトークンリストがユーザに明示されてもよい。例えば、GUIは、リンクテキストとページタイトルとを表わすトークンリストが一致しないことをユーザに視覚的に明示するかもしれない。図7に示されるGUIの例700は、ウェブページ302（図3A）に含まれるリンクと各々によって指し示されるページのタイトルとの間の比較の結果を提供する。

【0044】

GUI700は、部分701、702を有する。部分702はグリッドディスプレイを提供し、その各列においてリンクに関連する具体的な情報が表示される。例えば、列702Aにはリンクテキストが示され、列702Bにはそのリンクによって指し示されるページのタイトルが示される。

10

【0045】

図7の実施例では、ウェブページ上のリンクを示すテキストと一致すると思われるページタイトルが視覚的に表示される。例えば、行705にはリンク331（図3A）を示すテキスト710と、リンク331が指し示すウェブページ304上のタイトル715（つまり、図3Bのタイトル350）が表示される。行705はタイトル715を太字で示し、そのタイトルがリンクテキスト710と一致しないと思われることを視覚的に表示する。

【0046】

上記の技術を使い、管理者またはその他のユーザは、ウェブサイトによって提供されるリンクをより効果的に保守できる。例えば、リンクを指し示すテキストが、そのリンクが指し示すページのタイトルと一致しないとの警告を受けると（例えば、GUI700による）、ユーザはそのリンクをより詳しく検討し、そのリンクが正しいページを指し示しているか判断することができる。その結果、ユーザは、（従来のツールのように）使われなくなっているリンクを特定するだけでなく、無効なリソースを指し示すリンクをより効率的にアップデートすることができる。

20

【0047】

しかしながら、他の多数の応用が可能であるため、本発明はこの実装に限定されないと理解すべきである。例えば、本発明は、管理者によってウェブサイトの保守に使用されなければならないわけではない。その代わりに、本発明の実施形態は、ウェブページに含まれるリンクを調査し、これらのリンクが、それが指し示そうとしている文書を指し示しているか否かを判断するブラウザプログラムの中に取り入れることもできる。ブラウザは、リンクが指し示そうとしているページのタイトルと一致しないリンクテキストを視覚的に表示し、および/またはユーザが、指し示されているページにアクセスしないように阻止することもできる。このように、本発明の実施形態は、ユーザが悪意の、有害な、またはその他の望ましくないコンテンツを避けるのを助けるのに有益である。

30

【0048】

別の例として、図4-6について上述した比較技術を使って、リンクテキストとページタイトルとの一致を判断しなければならないわけではない。例えば、アルゴリズムを使い、ページタイトルのクエリストリングとの関連性を判断してもよい。（検索エンジンのように）ストリングをウェブページのコンテンツと一致させることによってクエリストリングとの関係する一致を判断するのではなく、ストリングをページタイトルに一致させてもよい。さらに、一致を、前述の関連性スコアを使う等によって、クエリストリングとの関連性の順序で分別してもよい。

40

【0049】

以上の説明から、本発明の実施形態の態様は、1つまたは複数のコンピュータプログラムおよび/またはハードウェア、ファームウェアまたはそれらの組み合わせによっても実現できる。例えば、ある実施形態の各種の要素を個々に、あるいは組み合わせて、プロセッサによってアクセス、実行される命令が記憶されるコンピュータ読取可能媒体を含むコンピュータプログラム製品として実現することもできる。コンピュータによって実行され

50

る場合、命令はコンピュータに実施形態の各種の態様を実現するよう指示する。

【0050】

本発明の少なくとも一実施形態のいくつかの態様について説明したが、当業者にとって、さまざまな変更、改変、改良は容易に着想できると理解されるはずである。このような変更、改変、改良も本明細書の一部とされ、本発明の精神と範囲に含まれるものとする。したがって、上記の説明と図面は例を示しているのにすぎない。

【図面の簡単な説明】

【0051】

【図1】本発明の実施形態を実現できるコンピュータシステムの一例のブロック図である。

【図2】本発明の実施形態を含むプログラムされた命令を記憶することのできるコンピュータメモリの一例のブロック図である。

【図3A】ユーザに対してウェブページを提示するためのブラウザインタフェースの一例を示す図である。

【図3B】ユーザに対してウェブページを提示するためのブラウザインタフェースの一例を示す図である。

【図4】本発明の一実施形態による、第一と第二のトークンストリングがどの程度対応するかを判断するプロセスの一例を示すフローチャートである。

【図5】本発明の一実施形態による、第一と第二のトークンストリング内のトークンを比較するプロセスの一例を示すフローチャートである。

【図6】本発明の一実施形態による、具体的なトークンを比較するプロセスの一例を示すフローチャートである。

【図7】本発明の一実施形態による、第一と第二のトークンストリングがどの程度対応するかを表示することのできるグラフィカル・ユーザ・インタフェース(GUI)の一例を示す図である。

【符号の説明】

【0052】

302、304 ウェブページ

305、310、315、320、331、340、342、344 リンク

330 メニュー部

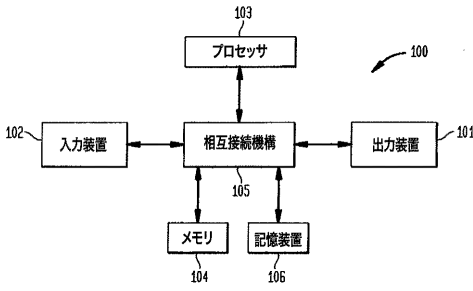
350 タイトル

10

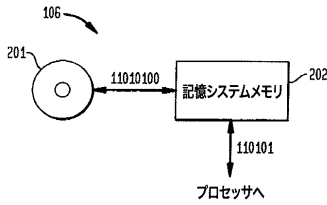
20

30

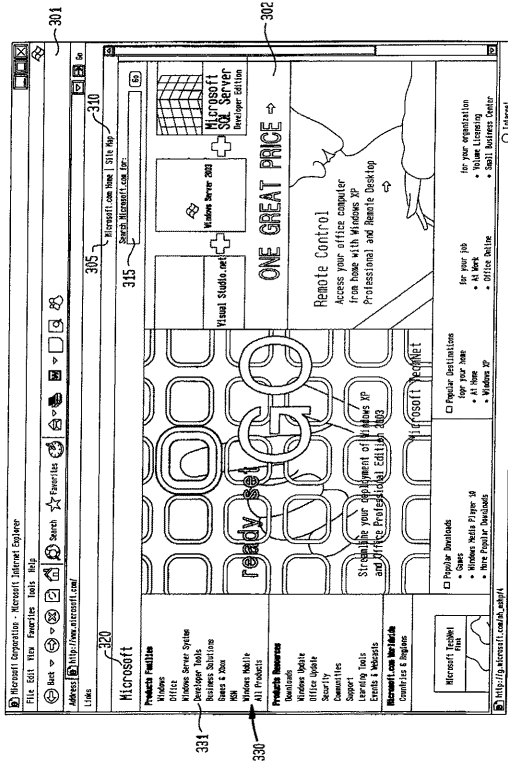
【図 1】



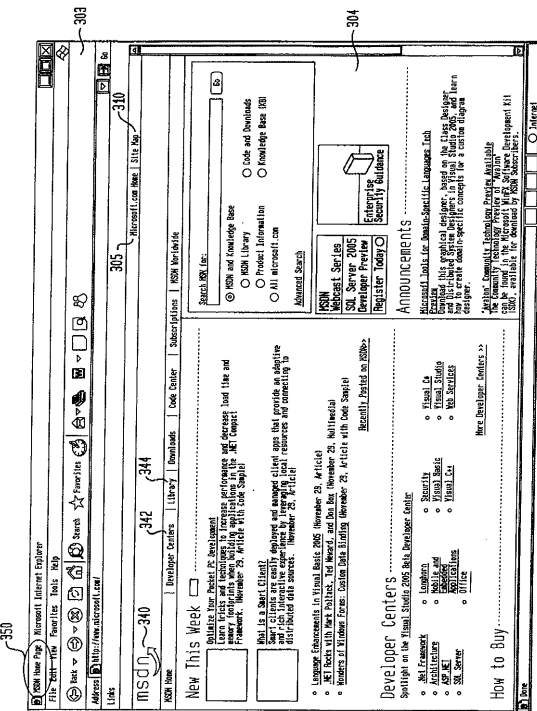
【図 2】



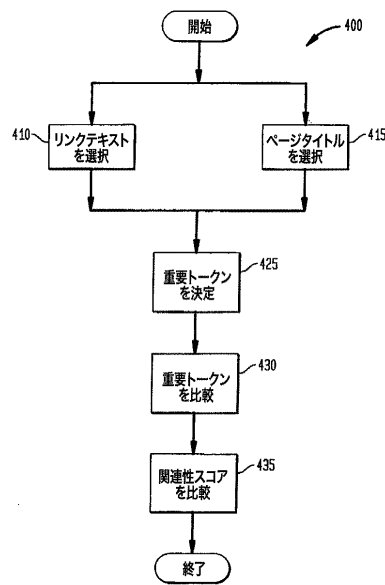
【図 3 A】



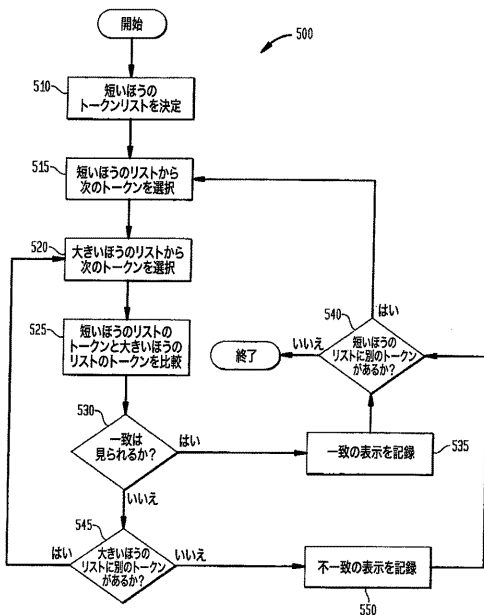
【図 3 B】



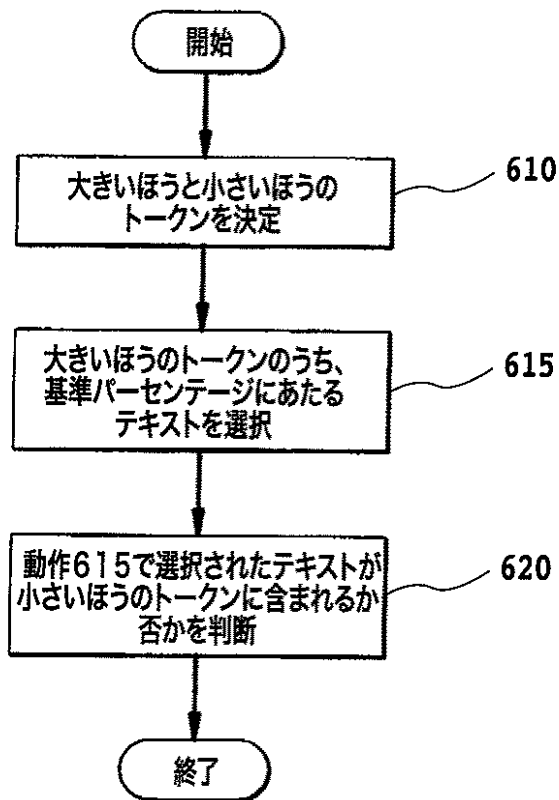
【図 4】



【 図 5 】



【 図 6 】



【 図 7 】

