

(19) 日本国特許庁(JP)

(12) 特 許 公 報(B2)

(11) 特許番号

特許第4621680号
(P4621680)

(45) 発行日 平成23年1月26日 (2011. 1. 26)

(24) 登録日 平成22年11月5日 (2010. 11. 5)

(51) Int. Cl. F I
G 0 6 F 17/30 (2006. 01)
 G 0 6 F 17/30 1 1 0 C
 G 0 6 F 17/30 1 7 0 J
 G 0 6 F 17/30 1 7 0 A

請求項の数 42 (全 14 頁)

(21) 出願番号	特願2006-533349 (P2006-533349)	(73) 特許権者	305029922 グーグル・インク
(86) (22) 出願日	平成16年5月20日 (2004. 5. 20)		アメリカ合衆国、カリフォルニア州 94
(65) 公表番号	特表2006-529044 (P2006-529044A)		043、マウンテン・ビュー、アンフィシ
(43) 公表日	平成18年12月28日 (2006. 12. 28)		アトル・パークウェイ 1600
(86) 国際出願番号	PCT/US2004/016258	(74) 代理人	110000408 特許業務法人高橋・林アンドパートナーズ
(87) 国際公開番号	W02004/114158	(72) 発明者	クレイグ・ネビル・マニング アメリカ合衆国 ニューヨーク ニューヨ
(87) 国際公開日	平成16年12月29日 (2004. 12. 29)		ーク グリーンウィッチストリート #4
審査請求日	平成19年3月1日 (2007. 3. 1)		34 666
(31) 優先権主張番号	60/472, 445	審査官	岩間 直純
(32) 優先日	平成15年5月20日 (2003. 5. 20)		
(33) 優先権主張国	米国 (US)		
(31) 優先権主張番号	10/608, 270		
(32) 優先日	平成15年6月27日 (2003. 6. 27)		
(33) 優先権主張国	米国 (US)		

最終頁に続く

(54) 【発明の名称】 定義付けシステムおよび方法

(57) 【特許請求の範囲】

【請求項1】

定義(44)を供給するためのシステム(10)であって、

1つ以上のプロセッサと、

前記1つ以上のプロセッサによって実行されて、前記1つ以上のプロセッサに対して、定義付けすべきフレーズ(41)を受信させ、少なくとも一つの定義(44)を各々含む複数のドキュメント(42)をサーチエンジンによって実行されるインターネットのウェブ・クロウリングによって特定させ、前記定義を検索するために、前記定義を示す第1のHTMLタグセットに基づいて前記定義を特定すること、パラグラフ、表の行、改行またはリストにおける項目を示す第2のHTMLタグセットに基づいて前記定義を分離すること、または、強調文を示す第3のHTMLタグセットに基づいて見出し語を特定すること、
 のうちの1つ以上を含む解析で前記特定された複数のドキュメントを解析させ、前記定義(44)に対し前記フレーズ(41)をマッチングさせ、前記定義を含む前記特定されたドキュメントのリンク構造のランク付けに基づいて決定される順番で前記フレーズ(41)に対する前記定義(44)を提示させる命令を格納するメモリと、
 を含むサーバ(11)
 を具備することを特徴とするシステム。

【請求項2】

請求項1に記載のシステム(10)であって、

定義付けすべきフレーズ(41)を受信し、少なくとも一つの定義(44)を各々含む複

数のドキュメント(42)を特定し、前記定義(44)に対し前記フレーズ(41)をマッチングし、前記フレーズ(41)に対する定義(44)を提示することは、実質的に、リアルタイム、バッチモード、またはその組み合わせで行われることを特徴とするシステム。

【請求項3】

請求項1に記載のシステム(10)であって、前記ドキュメント(42)は、ウェブページ(22)であることを特徴とするシステム。

【請求項4】

請求項1に記載のシステム(10)であって、前記特定は、前記サーチエンジン(21)上で問合せを行うことを含むことを特徴とするシステム。 10

【請求項5】

請求項4に記載のシステム(10)であって、前記特定は、所定のフィールド内に所定のタームを含むドキュメント(42)を検索することを含むことを特徴とするシステム。

【請求項6】

請求項5に記載のシステム(10)であって、前記所定のタームは、用語集、定義(44)、および辞書のうちの一つを含むことを特徴とするシステム。

【請求項7】

請求項5に記載のシステム(10)であって、前記所定のフィールドは、タイトルフィールドであることを特徴とするシステム。 20

【請求項8】

請求項1に記載のシステム(10)であって、前記マッチングは、一つ以上の特定されたドキュメント(42)における前記フレーズ(41)の存在を決定することを含むことを特徴とするシステム。

【請求項9】

請求項8に記載のシステム(10)であって、前記マッチングは、一つ以上の特定されたドキュメント(42)における前記フレーズ(41)の非存在を決定することを特徴とするシステム。 30

【請求項10】

請求項8に記載のシステム(10)であって、前記フレーズ(41)の存在を決定することはさらに、前記フレーズ(41)が正確に一致するという決定を含むことを特徴とするシステム。

【請求項11】

請求項8に記載のシステム(10)であって、前記マッチングは、前記フレーズ(41)の修正を含むことを特徴とするシステム。

【請求項12】

請求項11に記載のシステム(10)であって、前記フレーズ(41)の修正は、前記フレーズ(41)の標準的な形態の決定を含むことを特徴とするシステム。 40

【請求項13】

請求項1に記載のシステム(10)であって、前記マッチングはさらに、前記フレーズ(41)について対応する定義(44)を検索することを含むことを特徴とするシステム。

【請求項14】

請求項1に記載のシステム(10)であって、前記ランク付けは、少なくとも部分的に、前記定義(44)に対応するドキュメント(42)の前記リンク構造に基づくスコアに基づくことを特徴とするシステム。

【請求項15】

50

請求項 1 に記載のシステム (1 0) であって、
前記提示はさらに、前記定義 (4 4) を処理することを含むことを特徴とするシステム。

【請求項 1 6】

請求項 1 に記載のシステム (1 0) であって、
前記フレーズ (4 1) に対して前記定義 (4 4) を提示することは、該フレーズ (4 1) を、実質的に最も一般的に大文字化して提示することを含むことを特徴とするシステム。

【請求項 1 7】

請求項 1 6 に記載のシステム (1 0) はさらに、前記フレーズ (4 1) についてそれほど一般的でない形態の提示を含むことを特徴とするシステム。

【請求項 1 8】

請求項 1 に記載のシステム (1 0) はさらに、前記ドキュメント (4 2) に存在する前記フレーズ (4 1) のスーパーstring の決定を含むことを特徴とするシステム。

【請求項 1 9】

請求項 1 8 に記載のシステム (1 0) はさらに、前記決定されたスーパーstring のうちの少なくともいくつかについての提示を含むことを特徴とするシステム。

【請求項 2 0】

請求項 1 9 に記載のシステム (1 0) であって、
前記提示されたスーパーstring の少なくとも一つは、関連するフレーズ (4 1) または提案された問合せのうちのいずれか一つとして提示されることを特徴とするシステム。

【請求項 2 1】

コンピュータシステムで実行されて、定義 (4 4) を供給するための方法 (4 0) であって、

定義付けすべきフレーズ (4 1) を前記コンピュータシステムに関連するネットワークインターフェースによって受信し、

少なくとも一つの定義 (4 4) を各々含む複数のドキュメント (4 2) を前記コンピュータシステムに関連する一つ以上のプロセッサに関連するサーチエンジンによって実行されるインターネットのウェブ・クロウリングによって特定し、

前記定義を検索するために、前記定義を示す第 1 の HTML タグセットに基づいて前記定義を特定すること、パラグラフ、表の行、改行またはリストにおける項目を示す第 2 の HTML タグセットに基づいて前記定義を分離すること、または、強調文を示す第 3 の HTML タグセットに基づいて見出し語を特定すること、のうちの一つ以上を含む解析で前記

特定された複数のドキュメントを解析し、
前記定義 (4 4) に対し前記フレーズ (4 1) を前記コンピュータシステムに関連する一つ以上のプロセッサによってマッチングし、

前記定義を含む前記特定されたドキュメントのリンク構造のランク付けに基づいて決定される順番で前記フレーズ (4 1) に対する定義 (4 4) を前記コンピュータシステムに関連する一つ以上のプロセッサによって提示することを具備する方法 (4 0) 。

【請求項 2 2】

請求項 2 1 に記載の方法 (4 0) であって、
定義付けすべきフレーズ (4 1) を受信し、少なくとも一つの定義 (4 4) を各々含む複数のドキュメント (4 2) を特定し、前記定義 (4 4) に対し前記フレーズ (4 1) をマッチングし、前記フレーズ (4 1) に対する定義 (4 4) を提示することは、実質的に、リアルタイム、バッチモード、またはその組み合わせで行われることを特徴とする方法。

【請求項 2 3】

請求項 2 1 に記載の方法 (4 0) であって、
前記ドキュメント (4 2) は、ウェブページ (2 2) であることを特徴とする方法。

【請求項 2 4】

請求項 2 1 に記載の方法 (4 0) であって、
前記特定は、前記サーチエンジン (2 1) 上で問合せを行うことを含むことを特徴とする方法。

10

20

30

40

50

【請求項 25】

請求項 21 に記載の方法 (40) であって、
前記特定は、所定のフィールド内に所定のタームを含むドキュメント (42) を検索することを含むことを特徴とする方法。

【請求項 26】

請求項 25 に記載の方法 (40) であって、
前記所定のタームは、用語集、定義 (44)、および辞書のうちの一つを含むことを特徴とする方法。

【請求項 27】

請求項 25 に記載の方法 (40) であって、
所定のフィールドは、タイトルフィールドであることを特徴とする方法。

10

【請求項 28】

請求項 21 に記載の方法 (40) であって、
前記マッチングは、一つ以上の決定されたドキュメント (42) における前記フレーズ (41) の存在を決定することを含むことを特徴とする方法。

【請求項 29】

請求項 28 に記載の方法 (40) であって、
前記マッチングは、一つ以上の決定されたドキュメント (42) 内におけるフレーズ (41) の非存在を決定することを含むことを特徴とする方法。

【請求項 30】

請求項 28 に記載の方法 (40) であって、
前記フレーズ (41) の存在を決定することはさらに、前記フレーズ (41) が正確に一致するという決定を含むことを特徴とする方法。

20

【請求項 31】

請求項 28 に記載の方法 (40) であって、
前記マッチングは、前記フレーズ (41) の修正を含むことを特徴とする方法。

【請求項 32】

請求項 31 に記載の方法 (40) であって、
前記フレーズ (41) の修正は、前記フレーズ (41) の標準的な形態の決定を含むことを特徴とする方法。

30

【請求項 33】

請求項 21 に記載の方法 (40) であって、
前記マッチングはさらに、前記フレーズ (41) に対応する定義 (44) を検索することを含むことを特徴とする方法。

【請求項 34】

請求項 21 に記載の方法 (40) であって、
前記ランク付けは、少なくとも部分的に、前記定義 (44) に対応するドキュメント (42) の前記リンク構造に基づくスコアに基づくことを特徴とする方法。

【請求項 35】

請求項 21 に記載の方法 (40) であって、
前記提示はさらに、前記定義 (44) を処理することを含むことを特徴とする方法。

40

【請求項 36】

請求項 21 に記載の方法 (40) であって、
前記フレーズ (41) に対して前記定義 (44) を提示することは、該フレーズ (41) を実質的に最も一般的に大文字化して提示することを含むことを特徴とする方法。

【請求項 37】

請求項 36 に記載の方法 (40) はさらに、前記フレーズ (41) についてそれほど一般的でない形態での提示を含むことを特徴とする方法。

【請求項 38】

請求項 21 に記載の方法 (40) はさらに、前記ドキュメント (42) に存在する前記フ

50

フレーズ(41)のスーパーstringの決定を含むことを特徴とする方法。

【請求項39】

請求項38に記載の方法(40)はさらに、前記決定されたスーパーstringのうちの少なくともいくつかについての提示を含むことを特徴とする方法。

【請求項40】

請求項39に記載の方法(40)であって、

前記提示されたスーパーstringの少なくとも一つは、関連するフレーズ(41)または提案された問合せのうちのいずれか一つとして提示されることを特徴とする方法。

【請求項41】

請求項21乃至請求項40による方法(40)を1つ以上のプロセッサに実行させる、前記1つ以上のプロセッサによって実行されて前記1つ以上のプロセッサによって実行可能な命令を格納するコンピュータで読み取り可能な記憶媒体。

10

【請求項42】

定義(44)を供給するための装置(30)であって、

定義付けすべきフレーズ(41)を受信する手段と、

少なくとも一つの定義(44)を各々含む複数のドキュメント(42)をサーチエンジンによって実行されるインターネットのウェブ・クロウリングによって特定する手段と、

前記定義を検索するために、前記定義を示す第1のHTMLタグセットに基づいて前記定義を特定する手段、パラグラフ、表の行、改行またはリストにおける項目を示す第2のHTMLタグセットに基づいて前記定義を分離する手段、または、強調文を示す第3のHTMLタグセットに基づいて見出し語を特定する手段、のうちの1つ以上を含む、前記特定された複数のドキュメントを解析する手段と、

20

前記定義(44)に対し前記フレーズ(41)をマッチングする手段と、

前記定義を含む前記特定されたドキュメントのリンク構造のランク付けに基づいて決定される順番で前記フレーズ(41)に対する定義(44)を提示する手段と、

を具備することを特徴とする装置。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、一般に、定義付けを行うことに関連し、特に、定義付けを行うためのシステムおよび方法に関連する。

30

【背景技術】

【0002】

定義付けを行うためのシステムと方法については文献がある。インターネットで利用可能なコンテンツは膨大な量が存在する。このコンテンツには、用語集または定義の形式で構成されるものがある。本願明細書において記載されるシステムおよび方法により、フレーズに対する定義付けを迅速かつ効率的に行えるように、これらの利用可能なリソースを活用することが可能となる。ここで「フレーズ」とは、単語、語句、または定義付けすることが可能な他の語義単位をさす。

【発明の開示】

40

【発明が解決しようとする課題】

【0003】

本実施例は、定義付けを行うためのシステム及び方法を提供するものである。

【課題を解決するための手段】

【0004】

定義付けすべきフレーズを受信する。各々少なくとも一つの定義を含む一つ以上のドキュメントを決定する。これらの定義のうちの少なくとも一つに対し、フレーズをマッチングさせる。該フレーズに対し一つ以上の定義が提示される。

【0005】

さらに別の実施例によれば、分散情報記憶装置から定義を決定する。一つ以上のドキュメ

50

ントを特定する。ドキュメントは各々、分散情報処理装置内に維持され、対応するフレーズ用の定義を含む。特定されたドキュメント各々に関する情報が格納される。定義を検索すべきフレーズを、特定したドキュメント各々に格納された情報と突き合わせる。特定したドキュメントを分散情報処理装置から取り込み、一致する一以上の定義が返される。一致する定義が各々提示される。

【0006】

本発明のさらに別の実施例は、以下の詳細な説明により、直ちに当業者にとって明らかなものとなる。ここで、本発明の実施例は、本発明を実施するために考察される最良の形態を例示することにより説明される。当然のことながら、本発明については、これとは別の異なる実施例が可能であり、その詳細のいくつかは、すべて本発明の趣旨および範囲から逸脱することなく、さまざまな種類の自明な観点において変更可能である。従って、図面および詳細な説明は、本質的に例示されているとみなされるのであって制限的にみなされるものではない。

【発明の効果】

【0007】

本実施例は、定義付けを行うためのシステム及び方法を提供するものである。

【発明を実施するための最良の形態】

【0008】

本特許出願は、米国特許法119条(e)項に基づき、2003年5月20日付で出願された米国仮特許出願番号60/472,445、発明の名称「定義付けシステムおよび方法」、及び米国特許法120条に基づき、2003年6月27日付で出願された米国特許出願番号10/608,270、発明の名称「定義付けシステムおよび方法」を基礎として優先権を主張し、それらの開示内容は参照としてここに引用される。

【0009】

図1は、本発明に従い、定義付けを行うシステム10を示すブロック図である。当業者であれば理解できるように、複数の個別クライアント12は、例えばインターネット、または他の通信ネットワーク形態の相互接続ネットワーク13を介して、サーバ11と通信可能にインターフェースをとっている。個々のクライアント12は、それぞれのクライアント12を介してウェブコンテンツのリクエストや他の動作を行うユーザ19により操作される。

【0010】

一般に、各クライアント12は、例えば相互接続ネットワーク13のようなネットワークに接続可能であり、アプリケーションプログラムと対話可能であれば、いかなる形態の演算プラットフォームでもよい。当業者であれば理解できるように、個々のクライアントの例としては、パーソナル・コンピュータ、情報端末、「スマート」フォンおよびページャ、軽量クライアント、ワークステーション、アプリケーションサーバとインターフェースをとるダム端末、そしてこれらをさまざまに配置したものや構成したものが含まれ、これらに限定されるものではない。当業者であれば理解できるように、相互接続ネットワーク13は、ネットワーク相互接続可能コンポーネントについての多様なトポロジ、構成、配置を含み、これらは、企業、広域およびローカルエリアネットワークと相互作用的に連動するよう構成されており、また、限定的ではないが、従来からある有線、無線、衛星、光学および等価ネットワーク技術を含む。

【0011】

ウェブコンテンツをやりとりするため、そして特に、検索を行うために、各クライアント12はウェブ・ブラウザ18(「ウェブ・ブラウザ」)を実行する。これは、グラフィカル・ユーザ・インターフェースを実現するものであり、これを介して検索問合せが、サーバ11上で動作しているウェブサーバ20に送信される。これはさらに図2を参照して以下に説明する。各検索問合せは、潜在的にウェブサーバ20を介して読み出し可能な情報を、一般的にウェブコンテンツの形態で表現または特定するものである。さらに該検索問合せは、図3を参照して後述するように、定義を検索すべきフレーズを含めることができ

10

20

30

40

50

る。この検索問合せは、キーワード等の、通常は条件として表現される特徴、および、言語、文字コードなどの属性を提供し、これにより、サーバ11で実行されるサーチエンジン21がウェブページを特定し返送できるようにする。条件および属性はメタデータの形式であり、データ記述データを構成する。当業者であれば理解できるように、検索問合せ、検索問合せの特徴、およびメタデータを、他のスタイル、形式または定義で実現してもよい。

【0012】

ウェブページは、ウェブ・ブラウザ18に返送されて表示される。これは通常、ウェブコンテンツのタイトル、ハイパーリンク、および、ウェブページから取り出したテキストの抜粋などの他の説明情報の形態をとる。ユーザは、グラフィカル・ユーザ・インターフェース上のウェブページを閲覧またはそれにアクセスすることが可能であり、タイプされたテキスト、クリック、またはその両方の形式で、選択や応答を入力できる。サーバ11は、ウェブコンテンツ22がその内部に支持されている付属の記憶装置15を保持する。ウェブコンテンツ22はまた、相互接続ネットワーク13を介して直接的または間接的に相互接続され、好ましくは各クライアント12からアクセス可能な他のウェブサーバ(図示せず)上に遠隔的に保持されるようにしてもよい。

【0013】

サーチエンジン21は、高品質ウェブページを提供するために、好ましくは、検索問合せ条件に最もマッチするウェブコンテンツ22を特定する。これは例えば、S・プランとL・ページの「ハイパーテキスト対応大規模ウェブサーチエンジンの分析(The Anatomy of a Large-Scale Hypertextual Web Search Engine)」(1988年)、およびL・ページに対して2001年9月4日付で発行された米国特許6,285,999号に記載されており、これらの開示は参照により本明細書に引用される。図2を用いてさらに後述するように、マッチするウェブコンテンツ22を特定する場合、サーチエンジン21は、潜在的に検索可能なウェブコンテンツを表している情報特性に基づき動作する。ここで、ウェブサーバ20およびサーチエンジン21を含むサーバ20により提供される機能性は、ユニプロセッシング環境に加え、疎結合または密結合の、分散または並列演算構成によって提供されてもよい。

【0014】

サーバ11およびクライアント12を含む個々のコンピュータシステムは、汎用目的の、プログラムされたデジタル演算処理装置を備え、これは中央演算処理装置(それぞれプロセッサ13および16)、ランダム・アクセス・メモリ(それぞれメモリ14および17)、ハードディスクドライブまたはCD-ROMドライブなどの不揮発性二次記憶装置15、ネットワークまたは無線インターフェース、および、キーボードやディスプレイなどのユーザ・インターフェース手段を含む周辺装置から成る。ソフトウェア・プログラムを含むプログラムコードおよびデータは、CPUによる実行および処理のためにRAM内にロードされ、表示、出力、送信または格納の対象となる結果が生成される。ウェブ・ブラウザ18は、マイクロソフト社(レッドモンド、ワシントン)によりライセンスされているInternet Explorer(インターネット・エクスプローラ)、ネットスケープ社(マウンテンビュー、カリフォルニア)によりライセンスされているNavigator(ナビゲーター)などのHTTP互換のウェブ・ブラウザ、または、モジラもしくはジャバスクリプトが使用可能なブラウザであり、これらは周知技術である。

【0015】

図2は、図1のシステム10において使用する、コンピュータシステム30を示すブロック図である。コンピュータシステム30は、プロセッサ31と、コンピュータ・モニタまたは液晶ダイオード(LCD)ディスプレイのようなビジュアル・ディスプレイ32を有し、これは従来技術において周知である。コンピュータシステム30は、ウェブ・ブラウザ18(図1に示す)を実行し、これにより、グラフィカル・ユーザ・インターフェース37が実現される。検索された定義を含むビジュアル・ウェブコンテンツは、グラフィカル・ユーザ・インターフェース37上に定められた表示領域内に出力され、一方、ユーザ

10

20

30

40

50

入力は、表示領域および指定されたユーザ入力領域の両方に入力されるのが一般的である。テキストによるユーザ入力は、キーボード33を介して受信される。線形、非テキストの入力は、マウス、トラックボール、トラックパッドまたは矢印キーなどのポインティング・デバイス34を介して受信される。同様に、音声およびサウンドベースの入力は、マイクロホン35を介して受信される。ビジュアル出力は、ビジュアル・ディスプレイ32上のグラフィカル・ユーザ・インターフェース37を介して表示され、一方、音声出力はスピーカ36上で行われる。当業者であれば理解できるように、プロセッサ31、ビジュアル・ディスプレイ32、および入出力装置を含むコンピュータの構成要素については他の形態を使用してもよい。

【0016】

ここで、図3を参照して本発明の一実施例を説明する。図3は、本発明に従い定義付けを行う方法を示すフロー図である。この方法は、一連の処理動作またはステップとして説明され、例えば、図1に示すシステムまたはそれに相当する構成要素により実行することが可能である。

【0017】

まず、定義を検索すべきフレーズが提供される(ブロック310)。例えば、このフレーズは、ユーザリクエストまたは問合せ、あるいは他のなんらかの手段により提供される。フレーズを提供するシステムの一例としては、<http://labs.google.com/glossary>により特定されるURLに位置するもので、そのコンテンツは参照として本明細書に引用する。また、このフレーズのスペルは、より一貫した定義結果をもたらすために、一般的な原形に、必要に応じて修正または標準化することが可能である。

【0018】

定義を含むドキュメントが決定される(ブロック320)。これらのドキュメントが決定するために、いくつかの方法を用いてもよい。例えば、この種のドキュメントは、リアルタイムまたはバッチ処理モードのいずれかにおいて、サーチエンジンにより行われるウェブ・クロウリングまたはスパイダリングの間に決定してもよい。ドキュメントが定義を含んでいると決定されると、このドキュメント(またはドキュメントのURLなど、ドキュメントに関する情報)が、将来の使用に備えて格納または記憶される。定義用の「正式な」供給元、例えば、<http://www.dictionary.com>等のウェブサイトに対応するドキュメントを使用することも可能である。

【0019】

本発明の一実施例において、定義を含むドキュメントは、インターネットのサーチエンジンを介して問合せを行うことにより、実質的にリアルタイムで位置が特定される。他の実施例においては、例えば、問合せを受信する前にオフラインで、定義を含むドキュメントを取り込み、情報解析し、インデックス付けを行うことにより、このドキュメントの位置を実質的にバッチ処理モードで特定する。また、定義を含むドキュメントを特定するためにバッチ処理を提供し、かつ、各問合せに回答した定義を取り込み、重複を失くし、クリーンアップするためにリアルタイム処理を用いることにより、バッチ処理とリアルタイム処理の組み合わせを利用することも可能である。

【0020】

この問合せは、「用語集」、「定義」、「辞書」等の定義について、その変形や標準化したものだけでなく、それらの存在を示す可能性が高いタームを検索する。この検索は、ドキュメントのテキスト全体に亘って行ってもよいし、あるいは、タイトルフィールド、他のメタデータを含むフィールド等、ドキュメント内のある部分やフィールドに限定してもよい。また、ドキュメントの構造、すなわち、HTMLドキュメントのタグ付け特性を、問合せをどのように構築するか決定に関連させてもよい。本発明の一実施例において、ウェブページのタイトルで「用語集」、「定義」、「辞書」を検索することを利用して、関連するドキュメントを決定する。情報検索技術において、通常の知識を有する者により理解できるように、上記の方法は、定義を含むドキュメントを決定するために、さまざまなやり方や、数多くの他の方法に組み合わせてもよい。

10

20

30

40

50

【 0 0 2 1 】

定義を検索すべきフレーズを、このように決定されたドキュメントに対して突き合せ、定義を返してもらう（ブロック 3 3 0）。このステップ（ブロック 3 3 0）で決定されるドキュメントは、検索対象のフレーズと、このフレーズに対応する定義の存在を特定するために解析される。例えば、定義含有ドキュメントは、「見出し語」、または辞書形式で調べることが可能な単語で編成してもよい。見出し語を特定する、およびまたは定義を特定するための方法にはさまざまなものがある。本発明の一実施例において、別々のドキュメントを解析して、見出し語を特定し、およびまたは定義を返すために、以下の方法のうちの少なくとも一つが使用される。

【 0 0 2 2 】

・ ページが、定義リストを指定するために使用する HTML タグである、<dl>、<dt>および<dd>を使用する場合、定義を特定するために HTML によるマークアップに依拠する。すなわち、

定義リスト例として、

```
<dl>
```

```
<dt>Headword 1
```

```
<dd>This
```

```
is the definition of Headword 1
```

```
<dt>Headword 2
```

```
<dd>This
```

```
is the definition of Headword 2
```

```
<dt>Headword 3
```

```
<dd>This
```

```
is the definition of Headword 3
```

```
</dl>
```

となる。

・ <p>、<tr>、および
のような HTML タグは、連続した定義の間でセパレータとして処理される。

・ 空白または句読点（. , : - ）は定義の開始時に除去する。

・ 見出し語は、見出し語が HTML タグ、、、、<code> または、によって囲まれるという事実により特定される。

・ 見出し語で始まっていないラインは、削除される。

・ ドキュメントまたはページ内で、定義の数が N（例えば N = 5）より少ない場合、ドキュメント内またはページ内のすべての定義は廃棄される。

【 0 0 2 3 】

構文分析ツールは、全ての見出し語および定義を特定する際に完全である必要はない。実施例の一つにおいて、定義ドキュメント決定ステップ（ブロック 3 2 0）において決定される、定義含有ドキュメントは非常に数が多いので、構文分析ツールは、完全性よりは精度を重要視する。換言すれば、構文分析ツールは、正しくない可能性があるエントリを保持しておくというよりはむしろ、たとえそれが誤りであるとしても、捨てる方向へ判断するようになっている。というのは、満足な結果を供給するには十分過ぎるほどの定義があるからである。同様に、さらに別の実施例においては、構文分析ツールは、重複する、または単に他のエントリの累積に過ぎないエントリについてはその重複を取り除く。

【 0 0 2 4 】

その後、返された定義の少なくとも一つが提供される（ブロック 3 4 0）。実施例の一つにおいて、上記で引用した米国特許 6, 285, 999 号で開示されている方法により、定義が検索されたドキュメントの Page Rank（登録商標）に従い、返された定義がランク付けされる。この検索された定義はまた、例えば、以下のステップのうち少なくとも一つを実行することにより、処理され提示される。

【 0 0 2 5 】

10

20

30

40

50

- ・以下のものを取り除く：
 - 全てのHTMLマークアップ、
 - 見出し語および定義双方における最初と最後の空白、
 - 見出し語におけるすべての句読点：(. : ; ! ? -)、
 - 見出し語および定義内の最初の非アルファベットかつ非括弧の全て、
 - 見出し語内の最後の非英数字および非括弧の全て。
- ・以下の場合、定義を廃棄する：
 - 定義が"see"で始まる場合。
 - 定義が、すでに検索されたものの複製である場合。
 - ・定義の最初の文字は大文字にする。

10

【0026】

実施例の一つにおいて、フレーズについて前方完全一致するフレーズの定義のみが提示される。しかし、本発明の他の実施例においては、条件の緩い一致形態でも可能である。その他の情報もまた決定され提示される。本発明の一実施例において、フレーズのスーパーstringは、表形式にされ、問合せをさらに改善したもの、または関連するフレーズとして提示される。スーパーstringとは、フレーズ（またはフレーズの一般変形または標準化バージョン）を含むstringである。ワードやフレーズの一般変形または標準化バージョンを決定する方法は、例えば、2003年3月3日に出願された、係属中の米国特許出願番号10/377,117、代理人登録番号GP-091-00-US、発明者アミット・シングハル(Amit Singhal)らによる名称「クエリ意味情報に基づいて改良された検索クエリ(SEARCH QUERIES IMPROVED BASED ON QUERY SEMANTIC INFORMATION)」という発明において説明されており、その開示内容は本明細書において引用する。例えば、トップM個のスーパーstringをリストしてもよい。同様に、フレーズは、そのフレーズについて最も一般的な大文字化をするなど、処理済みの形式で表示してもよい。例えば、[pocket pc]または[pocket pcs]についてのユーザ問合せは、「Pocket PC」として表示される。というのは、これがこの定義において、最も一般的な形式およびまたは大文字化だからである。

20

【0027】

当業者により理解されるように、図3を用いて説明された上述のステップは、リストされた順序で実行される必要はなく、ステップを追加してもよいし、あるいは取り除いてもよい。

30

【0028】

本明細書において使用されているように、「ドキュメント」は、広義に解釈すべきであり、機械で読み取り可能なまたは記憶可能なワークプロダクトを含む。ドキュメントは、ファイル、ファイルの組合せ、他のファイル等への埋込みリンクを有する一つ以上のファイル等であってもよい。このファイルは、例えばテキスト、音声、画像、映像等のいかなるタイプのものであってもよい。インターネットのコンテキストにおいて一般的なドキュメントは、当業において周知のようにウェブページである。

【0029】

本発明のさらに別の態様によれば、定義が全く見つからない（または、定義が、当初提供されたフレーズに適切に一致するかどうかについて疑義がある場合などに、定義が提示用に選択されない）場合、元のフレーズに関連するタームまたはフレーズ、フレーズに関連するようであると見なされるタームまたはフレーズ、関心が高いと思われる（例えば、元のフレーズを入力したユーザの強い関心を引く）タームまたはフレーズのセット、または、定義が戻されたタームまたはフレーズの「無作為な」またはさまざまな要素を含んだセットを提供してもよい。このようなタームは、例えば、定義付けされるタームの種類に関するガイドをユーザに与えるため、またはユーザが楽しめるように提供してもよい。

40

【0030】

図4は、例えば、図3の方法により提供される定義を示す画面例400である。フレーズ「rdbms」に対する用語検索は、概ね図示されるように提供される。

50

【 0 0 3 1 】

図 5 は、例えば、図 3 の方法により提供される別の定義を示す画面例 5 0 0 である。フレーズ「pocket pc」に対する用語検索は、概ね図示されるように提供される。

【 0 0 3 2 】

図 6 は、例えば、図 3 の方法により提供されるさらに別の定義を示す画面例 6 0 0 である。フレーズ「pocket pcs」に対する用語検索は、概ね図示されるように提供される。

【 0 0 3 3 】

本発明については、その実施例を参照することにより具体的に示し説明がなされたが、形態および詳細における前述および他の変更は、本発明の意図と要旨の範囲から逸脱することなく、その範囲において行うことができると当業者により理解される。

10

【産業上の利用可能性】

【 0 0 3 4 】

本実施例は、定義付けを行うためのシステム及び方法を提供するものである。

【図面の簡単な説明】

【 0 0 3 5 】

【図 1】本発明に従い、定義付けを行うシステムを示すブロック図である。

【図 2】図 1 のシステムにおいて使用されるコンピュータシステムを示すブロック図である。

【図 3】本発明に従い、定義付けを行う方法を示すフロー図である。

【図 4】図 3 の方法により提供される定義を、一例として示す画面の写しである。

20

【図 5】図 3 の方法により提供される別の定義を、一例として示す画面の写しである。

【図 6】図 3 の方法により提供されるさらに別の定義を、一例として示す画面の写しである。

【符号の説明】

【 0 0 3 6 】

- 1 1 サーバ
- 1 2 クライアント
- 1 3 相互接続ネットワーク
- 1 4 メモリ
- 1 5 記憶装置
- 1 6 プロセッサ
- 1 7 メモリ
- 1 8 ブラウザ
- 1 9 ユーザ
- 2 0 ウェブサーバ
- 2 1 サーチエンジン
- 2 2 ウェブコンテンツ
- 4 0 スタート
- 4 1 定義付けすべきフレーズを決定
- 4 2 定義を有するドキュメントを決定
- 4 3 フレーズと決定されたドキュメントとをマッチング
- 4 4 検索された定義を提供

30

40

【 5 】

Use of this site is subject to express terms of use. By continuing past this page, you agree to abide by these terms.
 (Show details) (Show details)

Google
© 2006 Google

Look up product prices at [Amazon.com](#) or [eBay.com](#)

Search | [Images](#) | [Maps](#) | [Gmail](#) | [Google Talk](#) | [Google Reader](#) | [Google News](#) | [Google Scholar](#) | [Google Books](#) | [Google Finance](#) | [Google Earth](#) | [Google AdSense](#) | [Google Analytics](#) | [Google AdWords](#) | [Google Maps API](#) | [Google Webmaster Tools](#) | [Google Analytics](#) | [Google AdSense](#) | [Google Analytics](#) | [Google AdWords](#) | [Google Maps API](#) | [Google Webmaster Tools](#)

Definition of Pocket PC from this page

- A computer that fits in the palm of your hand which runs the latest Windows CE Operating System 2.0. Includes Pocket Word, Pocket Excel, Calendar, Contacts and Tasks as well as other applications. For more information read Pocket PC Article and the PC Companion feature Companion <http://www.microsoft.com/hardware/pccompanion/>
- Latest Windows CE environment for PDA. The communicating organizer is a true pocket computer with integrated software such as Pocket Word, Pocket Excel, Microsoft Money, Pocket Internet Explorer and a Multimedia reader (MP3, WAV) with the possibility of synchronizing with your PC (contacts, agenda, tasks, etc.) <http://www.microsoft.com/hardware/pccompanion/>
- Microsoft's latest operating system for handheld or PDA devices. There are currently several hardware manufacturers selling devices that come with Pocket PC. Compare (PDA) - PC Companion 5.6 & 5.0 and Cereus Companion 5.1.1b. Pocket PCs come with Microsoft Reader pre-installed <http://www.microsoft.com/hardware/pccompanion/>

500

【 6 】

Use of this site is subject to express terms of use. By continuing past this page, you agree to abide by these terms.
 (Show details) (Show details)

Google
© 2006 Google

Look up product prices at [Amazon.com](#) or [eBay.com](#)

Search | [Images](#) | [Maps](#) | [Gmail](#) | [Google Talk](#) | [Google Reader](#) | [Google News](#) | [Google Scholar](#) | [Google Books](#) | [Google Finance](#) | [Google Earth](#) | [Google AdSense](#) | [Google Analytics](#) | [Google AdWords](#) | [Google Maps API](#) | [Google Webmaster Tools](#) | [Google Analytics](#) | [Google AdSense](#) | [Google Analytics](#) | [Google AdWords](#) | [Google Maps API](#) | [Google Webmaster Tools](#)

Definition of Pocket PC from this page

- A computer that fits in the palm of your hand which runs the latest Windows CE Operating System 2.0. Includes Pocket Word, Pocket Excel, Calendar, Contacts and Tasks as well as other applications. For more information read Pocket PC Article and the PC Companion feature Companion <http://www.microsoft.com/hardware/pccompanion/>
- Latest Windows CE environment for PDA. The communicating organizer is a true pocket computer with integrated software such as Pocket Word, Pocket Excel, Microsoft Money, Pocket Internet Explorer and a Multimedia reader (MP3, WAV) with the possibility of synchronizing with your PC (contacts, agenda, tasks, etc.) <http://www.microsoft.com/hardware/pccompanion/>
- Microsoft's latest operating system for handheld or PDA devices. There are currently several hardware manufacturers selling devices that come with Pocket PC. Compare (PDA) - PC Companion 5.6 & 5.0 and Cereus Companion 5.1.1b. Pocket PCs come with Microsoft Reader pre-installed <http://www.microsoft.com/hardware/pccompanion/>

600

フロントページの続き

(56)参考文献 特開2002-117058(JP,A)

MARTIN MAURER, Re: How to get a site listed in Google Glossary???, Newsgroups: google.public.labs.glossary [online], 2003年 1月30日

PFEIFER U, Searching structured documents with the enhanced retrieval functionality of free WAIS-sf and Sfgate, COMPUTER NETWORKS AND ISDN SYSTEMS, NL, NORTH HOLLAND PUBLISHING, 1995年 4月 1日, V27 N6, P1027-1036

YVAN CLOUTIER, Google Glossary, Newsgroups: google.public.labs.glossary [online], 2002年 5月24日

MIHALCEA R, AutoASC - A SYSTEM FOR AUTOMATIC ACQUISITION OF SENSE TAGGED CORPORA, INTERNATIONAL JOURNAL OF PATTERN RECOGNITION AND ARTIFICIAL INTELLIGENCE, SI, WORLD SCIENTIFIC PUBLISHING COMPAGNY, 2000年 2月, V14 N1, P3-17

YUWONO B, Search and Ranking Algorithms for locating Resources on the World Wide Web, PROCEEDINGS OF THE TWELFTH INTERNATIONAL CONFERENCE ON NEW ORLEANS, 米国, IEEE COMPUT. SOC, 1996年 2月26日, P164-171

(58)調査した分野(Int.Cl., DB名)

G06F 17/30