

(19) 日本国特許庁(JP)

(12) 特 許 公 報(B2)

(11) 特許番号

特許第5679993号
(P5679993)

(45) 発行日 平成27年3月4日(2015.3.4)

(24) 登録日 平成27年1月16日(2015.1.16)

(51) Int. Cl.	F I
G06F 17/30 (2006.01)	G06F 17/30 330Z
	G06F 17/30 330C

請求項の数 15 (全 22 頁)

(21) 出願番号	特願2011-545381 (P2011-545381)	(73) 特許権者	510330264
(86) (22) 出願日	平成22年1月12日 (2010.1.12)		アリババ・グループ・ホールディング・リ ミテッド
(65) 公表番号	特表2012-515379 (P2012-515379A)		ALIBABA GROUP HOLDI NG LIMITED
(43) 公表日	平成24年7月5日 (2012.7.5)		英国領、ケイマン諸島、グランド・ケイマ ン、ジョージ・タウン、ワン・キャピタル ・プレイス、フォース・フロア、ピー・オ ー、ボックス 847
(86) 国際出願番号	PCT/US2010/000075		
(87) 国際公開番号	W02010/080719	(74) 代理人	110000028
(87) 国際公開日	平成22年7月15日 (2010.7.15)		特許業務法人明成国際特許事務所
審査請求日	平成24年12月27日 (2012.12.27)	(72) 発明者	シエ・ユヘン
(31) 優先権主張番号	200910001555.7		中華人民共和国 ハンチョウ、ワーナー・ ロード、ウエスト・レイク・インターナシ ョナル・プラザ、10階、ナンバー391 最終頁に続く
(32) 優先日	平成21年1月12日 (2009.1.12)		
(33) 優先権主張国	中国 (CN)		
(31) 優先権主張番号	12/655,980		
(32) 優先日	平成22年1月11日 (2010.1.11)		
(33) 優先権主張国	米国 (US)		

(54) 【発明の名称】 クエリを実行する方法およびクエリシステム

(57) 【特許請求の範囲】

【請求項1】

クエリを実行する方法であって、
ユーザ入力を受信し、

前記ユーザ入力に少なくとも部分的に基づいて第1の組のクエリキーワードを決定し、
前記第1の組のクエリキーワードの少なくとも一部に基づいて、前記第1の組のクエリ
キーワードを用いて以前に生成されたクエリに応じた履歴クエリ結果および前記履歴クエリ
結果の選択頻度を含むユーザフィールドバックログを取得し、

前記ユーザフィールドバックログに少なくとも部分的に基づいて、第2の組のクエリキー
ワードを決定し、前記第2の組のクエリキーワードの決定は、前記ユーザフィールドバック
ログに少なくとも部分的に基づいて前記第1の組のクエリキーワードの1または複数の潜
在単語を選択することを含み、

前記第2の組のクエリキーワードの少なくとも一部に基づいてクエリを実行する、
ことを備え、

前記履歴クエリ結果は文脈に基づいて複数の分割単語に分割され、前記第2の組のクエ
リキーワードは前記複数の分割単語を含み、

前記第2の組のクエリキーワードは、前記複数の分割単語の内の少なくとも一部に隣接
する1または複数の単語、前記複数の分割単語の少なくとも一部分を含む1または複数の
単語、前記複数の分割単語の内の少なくとも1つの分割単語を含む1または複数の単語、
または、これらの組み合わせを含み、

10

20

前記第 2 の組のクエリキーワードは、前記第 1 の組のクエリキーワードに対応する前記 1 または複数の潜在単語を含み、前記 1 または複数の潜在単語は、前記第 1 の組のクエリキーワードに関する潜在的語義を有し、前記潜在的語義は、対で出現する単語、特定の順序で出現する複数の単語、別の単語を形成する単語または文字、あるいはこれらの任意の組合せを含む、方法。

【請求項 2】

請求項 1 に記載の方法はさらに、
クエリ結果を返信し、
前記クエリ結果を用いてユーザフィールドバックデータベースを更新する、
ことを備える、方法。

10

【請求項 3】

請求項 1 に記載の方法において、前記ユーザフィールドバックログは、前記第 1 の組のクエリキーワードの少なくとも一部をクエリ対象として用いて取得された履歴クエリ結果を含む、方法。

【請求項 4】

請求項 1 に記載の方法において、前記選択頻度は、前記履歴クエリ結果のクリック頻度、前記履歴クエリ結果の表示頻度、前記履歴クエリ結果の閲覧時間、および、前記履歴クエリ結果の重要度、の少なくともいずれか一つに基づいて決定される、方法。

【請求項 5】

請求項 1 に記載の方法において、前記第 2 の組のクエリキーワードの少なくとも一部に基づくクエリの実行は、前記第 2 の組のクエリキーワードの少なくとも一部を検索エンジンに送信することを含む、方法。

20

【請求項 6】

請求項 1 に記載の方法はさらに、前記ユーザ入力を生成したユーザのユーザ特性を取得することを備え、前記ユーザフィールドバックログは、前記ユーザ特性に少なくとも部分的に従って取得される、方法。

【請求項 7】

請求項 1 に記載の方法はさらに、前記ユーザ入力を生成したユーザのユーザ特性を取得することを備え、前記第 2 の組のクエリキーワードは、前記ユーザ特性に少なくとも部分的に従って取得される、方法。

30

【請求項 8】

クエリシステムであって、
ユーザ入力を受信するよう構成されたインターフェースと、
前記インターフェースに接続された 1 または複数のプロセッサであって、
前記ユーザ入力に少なくとも部分的に基づいて第 1 の組のクエリキーワードを決定し

、
前記第 1 の組のクエリキーワードの少なくとも一部に基づいて、前記第 1 の組のクエリキーワードを用いて以前に生成されたクエリに応じた履歴クエリ結果および前記履歴クエリ結果の選択頻度を含むユーザフィールドバックログを取得し、

前記ユーザフィールドバックログに少なくとも部分的に基づいて、第 2 の組のクエリキーワードを決定し、前記第 2 の組のクエリキーワードの決定は、前記ユーザフィールドバックログに少なくとも部分的に基づいて前記第 1 の組のクエリキーワードの 1 または複数の潜在単語を選択することを含み、

40

前記第 2 の組のクエリキーワードの少なくとも一部に基づいてクエリを実行するよう構成された、1 または複数のプロセッサと、

前記履歴クエリ結果は文脈に基づいて複数の分割単語に分割され、前記第 2 の組のクエリキーワードは前記複数の分割単語を含み、

前記第 2 の組のクエリキーワードは、前記複数の分割単語の内の少なくとも一部に隣接する 1 または複数の単語、前記複数の分割単語の少なくとも一部分を含む 1 または複数の単語、前記複数の分割単語の内の少なくとも 1 つの分割単語を含む 1 または複数の単語、

50

または、これらの組み合わせを含み、

前記第 2 の組のクエリキーワードは、前記第 1 の組のクエリキーワードに対応する前記 1 または複数の潜在単語を含み、前記 1 または複数の潜在単語は、前記第 1 の組のクエリキーワードに関する潜在的語義を有し、前記潜在的語義は、対で出現する単語、特定の順序で出現する複数の単語、別の単語を形成する単語または文字、あるいはこれらの任意の組合せを含み、

前記 1 または複数のプロセッサに接続され、前記プロセッサに命令を提供するよう構成された 1 または複数のメモリと、
を備える、システム。

【請求項 9】

請求項 8 に記載のシステムにおいて、前記プロセッサはさらに、
クエリ結果を返信し、
前記クエリ結果を用いてユーザフィードバックデータベースを更新するように構成されている、システム。

【請求項 10】

請求項 8 に記載のシステムにおいて、前記ユーザフィードバックログは、前記第 1 の組のクエリキーワードの少なくとも一部をクエリ対象として用いて取得された履歴クエリ結果を含む、システム。

【請求項 11】

請求項 8 に記載のシステムにおいて、前記選択頻度は、前記履歴クエリ結果のクリック頻度、前記履歴クエリ結果の表示頻度、前記履歴クエリ結果の閲覧時間、および、前記履歴クエリ結果の重要度、の少なくともいずれか一つに少なくとも部分的に基づいて決定される、システム。

【請求項 12】

請求項 8 に記載のシステムにおいて、前記第 2 の組のクエリキーワードの少なくとも一部に基づいてクエリの実行は、前記第 2 の組のクエリキーワードの少なくとも一部を検索エンジンに送信することを含む、システム。

【請求項 13】

請求項 8 に記載のシステムにおいて、前記プロセッサは、さらに、前記ユーザ入力を生成したユーザのユーザ特性を取得するよう構成され、前記ユーザフィードバックログは、前記ユーザ特性に少なくとも部分的に従って取得される、システム。

【請求項 14】

請求項 8 に記載のシステムにおいて、前記プロセッサは、さらに、前記ユーザ入力を生成したユーザのユーザ特性を取得するよう構成され、前記第 2 の組のクエリキーワードは、前記ユーザ特性に少なくとも部分的に従って取得される、システム。

【請求項 15】

請求項 1 に記載の方法において、
前記第 2 の組のクエリキーワードを決定することは、IP アドレス、ユーザ側ブラウザ履歴データ、ユーザ側クッキーデータ、ユーザウェブ登録情報、またはこれらの任意の組合せを解析することを含む、方法。

【発明の詳細な説明】

【技術分野】

【0001】

本願は、データ処理技術に関し、特に、単語マッチングおよび情報クエリのための方法および装置に関する。

【背景技術】

【0002】

潜在的語義とは、通常、単語または句の潜在的な意味を指し、通常、1 または複数の単語または句で表現されうる。なお、以下では中文を C 1 ~ C 4 2 の符号で記し、中文と符号との対応関係は末尾の表 1 に示す。例えば、「C 1」(f r i d g e : 冷蔵庫の短縮語

10

20

30

40

50

)の潜在的語義は、一般に、「C2」(refrigerator:冷蔵庫)を指し、「C3」(cotton slippers:綿スリッパ)の潜在的語義は、一般に、「C4」(all-cotton slippers:純綿スリッパ)を指すなどである。

【0003】

潜在的意味の自動検索に関して多くの研究が行われており、それらのほとんどは、単語の共出現および関係性を用いて類義語を見つけようとするものである。一部の既存の技術は、同義語を用いて、単語間の関係を決定する。しかしながら、手作業で分類されたコーパスに基づいて得られる見出し語の数は限られており、同義語の自動発見の効果を保証することは難しい場合がある。

【0004】

検索エンジンのインデックスモードは、通常、単独語検索(separate word search)、単語分割インデックス方式(word partitioning indexing)、および、ハイブリッドインデックス方式を含む。単独語インデックス方式の技術を用いる場合、通常、ファイル内での単独語間の距離が算出される必要がある。したがって、効率が悪いことが多く、精度が低いことが多い。この問題は、単語間に自然な単語区切り文字(例えば、スペース)を持たない言語(中国語など)で特に顕著である。例えば、「C5」(pesticides:農薬)、「C6」(Shen Nong pharmaceuticals:神衣薬品)、および、「C7」(Shen Nong pesticides factory:神衣農薬工場)の間の差異は、単独語インデックス方式を用いると容易に区別できない。逆に、単語分割検索技術は、より高い精度を有すると共に高速であるが、再現率は低いことが多い。例えば、その技術で「C1」(fridge)を検索すると、「C1」(fridge)の結果のみを見出すことが可能であり、「C2」(refrigerator)の結果を見出すことはできない。単独語インデックス方式と単語分割インデックス方式とを組み合わせたハイブリッドインデックス方法では、通常、最初に単語分割インデックス方式の技術に従ってクエリを行い、その後、単独語インデックス方式の技術に従ってクエリを行う。例えば、「C8」(glass bottle:ガラス瓶)をクエリすると、単語分割インデックス方式に従って「C8」(glass bottle)が見つかり、その後、単独語インデックス方式に従って別の結果が見つかる。これは、上述の2つの方法の欠点を補うが、単独語インデックス方式に従って「C9」(glass bottles:ガラス瓶(複数))が見つけれられ、検索エンジンは、「C9」(glass bottles)と「C10」(causing bottle neck:瓶首を作る)との区別ができないため、精度に影響がある。したがって、より効果的な検索技術が求められている。

【図面の簡単な説明】

【0005】

【図1A】クエリ処理システムの一実施形態を示すブロック図。

【図1B】クエリを実行するための処理の一実施形態を示すフローチャート。

【図2】クエリ処理システムの一実施形態を示すブロック図。

【図3】マッチングモジュールの一実施形態を示すブロック図。

【図4】第1クエリキーワード取得モジュールの一実施形態を示すブロック図。

【図5】単語マッチングのための処理の一実施形態を示すフローチャート。

【図6】単語マッチングのための装置の一実施形態を示すブロック図。

【発明を実施するための形態】

【0006】

本発明は、処理、装置、システム、物質の組成、コンピュータ読み取り可能な格納媒体上に具現化されたコンピュータプログラム製品、および/または、プロセッサ(プロセッサに接続されたメモリに格納および/またはそのメモリによって提供される命令を実行するよう構成されたプロセッサ)を含め、様々な形態で実装されうる。本明細書では、これらの実装または本発明が取りうる任意の他の形態を、技術と呼ぶ。一般に、開示された処理の工程の順序は、本発明の範囲内で変更されてもよい。特に言及しない限り、タスクを

10

20

30

40

50

実行するよう構成されるものとして記載されたプロセッサまたはメモリなどの構成要素は、ある時間にタスクを実行するよう一時的に構成された一般的な構成要素として、または、タスクを実行するよう製造された特定の構成要素として実装されてよい。本明細書では、「プロセッサ」という用語は、1または複数のデバイス、回路、および/または、コンピュータプログラム命令などのデータを処理するよう構成された処理コアを指すものとする。

【0007】

以下では、本発明の原理を示す図面を参照しつつ、本発明の1または複数の実施形態の詳細な説明を行う。本発明は、かかる実施形態に関連して説明されているが、どの実施形態にも限定されない。本発明の範囲は、特許請求の範囲によってのみ限定されるものであり、多くの代替物、変形物、および、等価物を含む。以下の説明では、本発明の完全な理解を提供するために、多くの具体的な詳細事項が記載されている。これらの詳細事項は、例示を目的としたものであり、本発明は、これらの具体的な詳細事項の一部または全てがなくとも特許請求の範囲に従って実施可能である。簡単のために、本発明に関連する技術分野で周知の技術事項については、本発明が必要以上にわかりにくくならないように、詳細には説明していない。

【0008】

図1Aは、クエリ処理システムの一実施形態を示すブロック図である。図の例において、クエリ処理システム150は、クエリ処理サーバ1512を備えており、クエリ処理サーバ152は、いくつかの実施形態において、1または複数のメモリに接続された1または複数のプロセッサを有する1または複数のデバイスを備える。クエリ処理サーバ152は、ユーザ入力を受信するよう構成された1または複数のインターフェースを備える。ユーザ入力は、例えば、ネットワーク155（インターネットまたはその他の通信ネットワークであってよい）を介してクライアントデバイス（154など）によって送信されたユーザ入力である。通信インターフェースの例としては、ポート、ケーブル、有線または無線ネットワークインターフェースカードなどの外部接続、および、通信バスなどの内部接続が挙げられるが、これらに限定されない。いくつかの実施形態では、クライアントデバイスは、ネットワーク機能を有するコンピュータまたは携帯デバイスなどのコンピュータデバイスである。クエリ処理サーバ152は、ユーザフィードバックログを格納するユーザフィードバックデータベース156にアクセスできる。

【0009】

後に詳述するように、ユーザ入力を受信すると、クエリ処理サーバは、ユーザ入力を処理して、ユーザ入力およびユーザフィードバックログに基づいてクエリキーワードを生成する。いくつかの実施形態では、クエリキーワードは、クエリキーワードに基づいて検索を実行し検索結果を返す検索エンジン158に送信される。いくつかの実施形態では、検索エンジンは、データベースに格納された特定のコンテンツ（電子商取引ウェブサイト上の製品情報など）を検索するよう構成される。いくつかの実施形態では、検索エンジンは、インターネット上のコンテンツを検索する汎用の検索エンジンとして構成される。キーワード検索エンジンおよびクエリ処理サーバの機能は、いくつかの実施形態では統合されており、その場合、クエリ処理サーバは、キーワードを用いて検索を行い、直接的に結果を返す。検索結果、および、検索結果に対するユーザの応答（ユーザ選択、クリックスルーなど）は、ユーザフィードバックデータベース内で更新される。

【0010】

図1Bは、クエリを実行するための処理の一実施形態を示すフローチャートである。処理100は、システム（150など）上で実行されてよい。工程102において、複数のシンボルを含むユーザ入力を受信される。いくつかの実施形態では、ユーザ入力は、クライアントデバイス（154など）からウェブブラウザまたはその他のユーザインターフェースツールによってクエリ処理サーバに送信される。本明細書で用いられているように、シンボルとは、文字（character or letter）を指す。複数の文字（charactersまたはletters）が単語を形成する。

【 0 0 1 1 】

工程 1 0 4 において、第 1 の組のクエリキーワードが、ユーザ入力に基づいて決定される。第 1 の組のクエリキーワードは、1 または複数の単語を含んでよい。いくつかの実施形態では、ユーザ入力は、第 1 の組のクエリキーワードとして直接的に利用される。いくつかの実施形態では、ユーザ入力は分割され、分割された単語が、第 1 の組のクエリキーワードとして利用される。いくつかの場合、ユーザ入力は、文字 (l e t t e r) ベースの言語 (英語など) であり、単語は、スペースによって必然的に区切られる。したがって、ユーザ入力は、文脈ベースの処理を必要とせず、第 1 の組のクエリキーワードは、ユーザ入力内の単語および / または語幹入力単語を含む。一部の例では、漢字ベースの言語 (中国語など) が、ユーザ入力言語として用いられる。かかる言語には特定の区切り文字がないため、入力シンボルは、文脈に基づいて単語に分割され、第 1 の組のクエリキーワードは、分割された単語を含む。いくつかの実施形態では、第 1 の組のクエリキーワードは、さらに、ユーザ入力内の文字 (c h a r a c t e r) を含み、それらの文字は分割することなしに利用可能である。

10

【 0 0 1 2 】

工程 1 0 6 において、ユーザフィールドバックログが、第 1 の組のクエリキーワードに基づいて取得される。ユーザフィールドバックログは、いくつかの実施形態では、データベース (1 5 6 など) から取得され、第 1 の組のクエリキーワードをクエリの対象 (または、目標) として用いた履歴クエリ結果と、ユーザによる選択頻度 (クリックスルー率とも呼ぶ) とを含む。様々な実施形態において、選択頻度は、以前のクエリ結果に対する選択頻度および / または以前の結果のコンテンツに対する選択頻度を含みうる。ユーザフィールドバックログは、検索エンジンによって収集される。いくつかの実施形態では、ユーザによるキーワード入力、過去のクエリ結果 (例えば、ウェブページファイル識別子 / ユニバーサルリソースロケータ (U R L))、過去のクエリ結果に対するクリック頻度、過去のクエリ結果の表示頻度などが、検索エンジンによって収集されて格納される。例えば、第 1 の組のクエリキーワードが、単語「western medicine : 西洋医学」すなわち「C 1 1」を含むとすると、ユーザフィールドバックログは、ユーザが以前に単語「western medicine」すなわち「C 1 1」を用いてクエリを行った時に返された結果のすべてまたは一部の記録を含む。

20

【 0 0 1 3 】

工程 1 0 8 において、ユーザフィールドバックログに基づいて第 1 の組のクエリキーワードに一致する第 2 の組のクエリキーワードが決定される。いくつかの実施形態では、第 2 の組のクエリキーワードは、ユーザフィールドバックログに基づいて導出された潜在的語義を持つ単語を含む。潜在的語義を確立するために、以前の記録のすべてまたは一部 (ランダムに選択された部分など) が利用されてよい。潜在的意味を有する潜在単語の自動発見とは、元の単語と関連する意味または同様の意味を有する 1 または複数の単語 (または、句) を見つけることを指す。ユーザフィールドバックログに基づいて、クエリ単語と、ユーザの意図を具体化する履歴クエリ結果との間の潜在的語義の関係を、高い信頼性で自動的に見つけることができる。潜在的語義の関係は、次いで、検索エンジンの精度およびインテリジェンスを改善するために用いられる。したがって、ユーザフィールドバックログは、第 1 の組のクエリキーワードを検索対象として用いている以前の全履歴クエリ結果と、履歴クエリ結果に対するユーザによる以前の全クリック選択頻度とを含みうる。

30

40

【 0 0 1 4 】

いくつかの実施形態では、第 2 の組のクエリキーワードを決定する工程は、第 1 の組のキーワードの履歴クエリ結果のコンテンツを取得する工程と、履歴クエリ結果のコンテンツを分割して 1 組の分割単語を取得する工程と、分割単語の選択頻度に基づいて第 2 の組のクエリキーワードを決定する工程と、を含む。

【 0 0 1 5 】

工程 1 1 0 において、第 2 の組のクエリキーワードの少なくとも一部を用いてクエリが実行され、クエリ結果が取得される。いくつかの実施形態では、クエリキーワードは、ク

50

エリ結果を取得するために、検索エンジンに入力される。

【0016】

工程112において、第2の組のクエリキーワードに基づいたクエリ結果を用いて、ユーザフィールドバックログの取得先であるユーザフィールドバックデータベースが更新される。

【0017】

様々な実施形態において、第1の組のクエリキーワードは、以下のタイプの内の1または複数に属する：

【0018】

タイプ1：分割単語に隣接する単語。説明のため、以下では、かかる場合の選択頻度に関連する統計結果をP1とする。

【0019】

タイプ2：分割単語の一部分を含む単語。以下では、かかる場合の選択頻度に関連する統計結果をP2とする。

【0020】

タイプ3：分割単語を含む単語。以下では、かかる場合の選択頻度に関連する統計結果をP3とする。

【0021】

ユーザフィールドバックログは、履歴クエリ結果、クリック頻度、および、履歴クエリ結果の表示頻度など、クエリ単語に対応する情報を記録するよう適合される。例えば、クエリ結果は、ウェブページであってよい。与えられたクエリ単語について、結果として生じるウェブページが高い選択頻度を有していれば、クエリ単語との関連性が高いと見なされる。単語の潜在的語義とは、元の単語の同義語、類義語、または、部分的な同義語である別の単語を意味するものであり、一例では、「C8」と「C9」（「glass bottle」と「glass bottles」）、別の例では、「C12」と「C13」と「C14」（「double bed：ダブルベッド」と「single bed：シングルベッド」と「mattress bed：マットレスベッド」）であり、後者の例では、すべての単語が、人の寝る場所である「C15」（「bed：ベッド」）という潜在的語義を有しているが、「C16」（「flower bed：花壇」）は、「C15」（「bed」）という潜在的語義を有していない。

【0022】

本願の様々な実施形態では、少なくとも3タイプの潜在的語義が用いられる。第1のタイプの潜在的語義は、対で現れることが多い単語であり、例えば、「Motorola：モトローラ」および「company：社」（「C17」および「C18」）、「Motorola」および「mobile phone：携帯電話」（「C17」および「C19」）である。かかる関係は、2つの単語の間の相関関係を示す。換言すれば、単語分割後のいくつかの単語は、クエリ単語に（少なくとも意味において）隣接すると見なされる。単語の第2のタイプの潜在的語義は、特定の順序で現れる複数の単語を含む。例えば、クエリキーワード「glass bottle」は、対応する潜在単語「glass」および「bottle」を有しており、「C8」は、潜在的語義単語「C20」および「C21」を有しており、「C22」（beauty）は、対応する潜在単語「C23」（beautiful：美しい）、「C24」（woman：女性）を有する。つまり、分割単語は、単語分割後のクエリ単語の少なくとも一部分を含む。第3のタイプの潜在的語義は、別の単語を形成する単語または文字である。例えば、「C25」（shrimp：エビ）および「C26」（prawns：テナガエビ）、「C27」（alcohol：酒）および「C28」（beer：ビール）であり、すなわち、分割単語は、クエリキーワードを含む。クリック頻度などのユーザフィールドバックを用いて自動的に見出された潜在的語義は、通常、ユーザによって入力された検索キーワードの意図を表し、検索エンジンの精度を改善するために利用可能である。例えば、ユーザが、「C15」（「bed」）を検索する場合、ユーザの実際の意図は、通常、「C16」（「flower bed」

10

20

30

40

50

)などの一般的な支持構造、または、ある種の機械「C30」(「lathe:旋盤」)ではなく、就寝のためのベッド、例えば、「C13」(「single bed」)、「C12」(「double bed」)、および、「C29」(「wooden bed:木製ベッド」)である。表示された検索結果の選択などのユーザフィードバックに基づいて、「C13」(「single bed」)、「C12」(「double bed」)、および、「C29」(「wooden bed」)が、「C15」(bed for sleeping:就寝用のベッド)という潜在的語義を持ち、「C16」(「flower bed」)が、「C15」という潜在的語義を持たないことを決定できる。

【0023】

いくつかの実施形態では、第1の組のクエリキーワードが、工程102および104に従って取得され、履歴クエリ結果(ウェブページURL、ファイルIDなど)、クリック頻度、および、履歴クエリ結果の表示頻度、もしくは、これらの組み合わせが、工程106に従って取得される。単語分割が、第1の組のクエリキーワードに対して実行される。第1の組のクエリキーワードが複数の単語を含む場合、クエリ単語のユーザフィードバックログに含まれる対応する履歴クエリ結果および関連情報が、各分割単語に対応する各ユーザフィードバックログに追加される。したがって、各分割単語は、自身の対応する履歴クエリ結果を有する。処理の結果として、ユーザフィードバックログの各クエリは、独立した分割単語になる。P1、P2、P3に関連する上述の処理は、それぞれ、各分割単語または分割単語の部分に対して実行され、履歴クエリ結果の選択は、総クエリ回数、クリック回数、履歴クエリ結果の公開回数、または、それらの組み合わせの情報に従って決定されてよい。分割単語に対応する履歴クエリ結果は、それぞれ処理される。分割単語に完全にマッチする文字列すべてが、ユーザフィードバックログの履歴クエリ結果内で見出され(ここで、完全に一致することとは、分割単語が文字列の部分列であることを意味する)、文字列の長さは、分割単語を含む文の長さ、または、分割単語を含む文の長さのM倍であってよい(ただし、Mは2以上の任意の整数であってよい)。分割単語のタイプに応じて、対応する統計結果P1、P2、または、P3が、各分割単語について算出される。説明のために、以下の実施形態では、ファイルをクエリ結果として用いるが、URLまたはその他の適切な検索オブジェクトなど、他の結果を用いてもよい。実装例において、クエリ結果の選択頻度および/またはクエリ結果のコンテンツの選択頻度が考慮されてよい。

【0024】

いくつかの実施形態では、第1の組のクエリキーワードが、工程102および104に従ってユーザ入力に基づいて取得され、履歴クエリ結果(ウェブページURL、ファイルIDなど)、クリック頻度、および、履歴クエリ結果の表示頻度、もしくは、これらの組み合わせが、工程106に従ってユーザフィードバックログから取得される。一方、クエリ辞書が、随機的に事前構成されてもよく、異なる検索単語に対応する情報(履歴クエリ結果(ウェブページURL、ファイルIDなど)、クリック頻度、および、履歴クエリ結果の表示頻度、もしくは、これらの組み合わせ)が、事前に入力されてよい。このように、第1の組のクエリキーワードが入力された時に、第2の組のクエリキーワードは、クエリ辞書を用いて迅速かつ容易に取得されうる。すなわち、以前のユーザフィードバックログのコンテンツは、クエリのために事前に格納されており、クエリ辞書は、新しいユーザフィードバックログに従って更新されてもよいが、ユーザフィードバックログは、第1のクエリキーワードが入力された後に呼び出されてもよい。

【0025】

異なるタイプの分割単語の処理について説明する。

【0026】

タイプ1:分割単語が第1の組のクエリキーワードに隣接する。第1の組のクエリキーワードが文字列の分割単語を含む場合、例えば、第1のクエリキーワードが「C22」であり、ユーザフィードバックログ内の履歴クエリ結果が、コンテンツ「C31、C32」(Chinese|ancient|beauty|Xishi|named|Yigu

10

20

30

40

50

an | , born | during | Chunqiu | Zhanguo | period) を含むファイルである場合である(ここで、「|」は、単語分割が起きる位置を示す)。ここで、クエリ結果文字列内でクエリキーワードに直接隣接する分割単語の出現回数(T)に、クエリ結果のクリック頻度および/または表示頻度を掛けて、その積を重み付け係数(回数重み(1)と呼ぶ)として、全クエリ結果のカウントP1に加える。P1は、各単語が第1のクエリキーワードの前または後ろに出現する回数の重み(1)を含む。例えば、この例では、ファイルの重みが0.5である場合、P1における「C33」(ancient)および「C34」(Xishi)に対応する結果(ここで、2つの単語はクエリキーワードに直接隣接し、T=1である)は、それに応じて0.5だけ増大される。

【0027】

10

タイプ2: 分割単語が第1の組のクエリキーワードの部分を含む。第1の組のクエリキーワードが、複数の隣接する単語分割結果に含まれる場合である。例えば、第1のクエリキーワードが「C22」(beauty)であり、ユーザフィードバックログ内の履歴クエリ結果が、分割されたコンテンツ「C35」(Xishi | is | a | beautiful | woman |)を有する場合である。この時、第1のクエリキーワードを含む分割単語の出現回数に、クエリ結果のクリック頻度および/または表示頻度を掛けて、その積を重み付け係数(回数重み(2)と呼ぶ)として、全クエリ結果のカウントP2に加える。P2は、第1の組のクエリキーワードの複数の分割単語が同じ順序で出現する回数の重み(2)を含む。したがって、この例では、ファイルの重みが0.3の場合、P2における「C36」に対応する結果は、0.3だけ増大される。

20

【0028】

タイプ3: 分割単語が第1の組のクエリキーワードを含む。第1のクエリキーワードが分割単語の部分列である場合、例えば、クエリ単語が「C1」(fridge)であり、ユーザフィードバックログ内の履歴クエリ結果が、分割されたコンテンツ「C37」(refrigerator | air conditioning | maintenance | and | repair)を有する場合。この時、第1のクエリキーワードを含む分割単語の出現回数に、クエリ結果のクリック頻度および/または表示頻度を掛けて、その積を重み付け係数(回数重み(3)と呼ぶ)として、全クエリ結果のカウントP3に加える。なお、P3は、クエリ結果内に第1のクエリキーワードを含む分割単語の出現回数の重み(3)である。したがって、この例では、ファイルの重みが0.8の場合、P3における「C2」(refrigerator)に対応する結果は、0.8だけ増大される。

30

【0029】

処理工程は、ユーザフィードバックログ内の履歴クエリ結果の分割単語がすべて処理されるまで繰り返される。分割単語がP1に従って出現する回数の加重和に基づいて、回数重みの合計が第1の閾値よりも大きい分割単語を、クエリ単語の第1の潜在的語義関係とする。同様に、分割単語がP2およびP3で出現する回数の加重和に従って、回数重みの合計がそれぞれ第2の閾値および第3の閾値よりも大きい分割単語を、単語の第2の潜在的語義関係および第3の潜在的語義関係とする。

【0030】

様々な実施形態において、潜在的語義の内の1または複数が処理される。第1、第2、および、第3の閾値は、固定の閾値であってもよいし、クエリ単語の全クエリ結果に従って動的に調整されてもよい。例えば、一致した文字列を含むすべてのクエリ結果ファイルに加重和を実行し、次いで、得られた結果に係数を掛ける。なお、係数は、クエリ結果に従って動的に設定されてよい。閾値を設定することにより、クエリ単語の一部の潜在的語義を有する単語は、無条件のフィードバックになるのではなく、選択的に決定される。

40

【0031】

いくつかの実施形態において、第1の組のクエリキーワードに一致する第2の組のクエリキーワードが、履歴クエリ結果および選択頻度に従って決定される時に、選択頻度が所定の閾値よりも大きい必要があってよく、選択頻度は、履歴クエリ結果に対するユーザの選択頻度であってよく、また、履歴クエリ結果のコンテンツに対するユーザの選択頻度で

50

あってもよい。ファイルまたはそのコンテンツのクリック頻度および/または表示頻度を重み係数とする。その係数は、クリック頻度および表示頻度の一方または両方であってよく、係数の値は、クリックおよび表示頻度と線形または非線形の関係性を有しうる。例えば、両方の頻度が所定の閾値よりも大きい場合に係数は1に設定され、そうでない場合に係数は0に設定されるか、もしくは、最大のクリック頻度および表示頻度を有する係数は1であり、他の係数は、最大値で割ることによって[0, 1]に正規化される。選択頻度を選択する目的は、選択頻度を用いて潜在的語義を見つけることである。低い選択頻度を有する一部の情報は、所定の閾値によってフィルタリングされるため、それにより、潜在的語義の発見速度を向上させると同時に、いくつかの情報の干渉を避けることができる。

【0032】

様々な実施形態において、選択頻度は、以下の内の1または複数を含む：履歴クエリ結果のクリック頻度、履歴クエリ結果の表示頻度、履歴クエリ結果の閲覧時間 (reading time)、および、履歴クエリ結果の重要度。ファイルのクリック頻度および表示頻度(または、それらの内の一方)は、重み付け係数と見なされ、その係数は、ファイルの他の情報、例えば、ユーザがそのファイルを読むのに費やした時間、重要度、クリック頻度、および、表示頻度の組み合わせであってもよい。

【0033】

実施例において、クエリ単語とその潜在的語義との間の関係は、交換可能である。例えば、「C38」(glass | bottles)は、「C8」(glass bottle)の潜在的語義であり、「C8」(glass bottle)は、「C38」(glass | bottles)の潜在的語義であり、また、「C2」(refrigerator)は、「C1」(fridge)の潜在的語義であり、「C1」(fridge)は、「C2」(refrigerator)の潜在的語義である。

【0034】

工程104において第1の組のクエリキーワードを決定するために、様々な技術を用いることができる。1つの技術は、ユーザによって入力された情報コンテンツに対して、単語分割を最初に実行し、分割された単語を用いて、クエリキーワードを取得することを含む。別の技術では、情報コンテンツは文字に分解され、各文字がクエリキーワードとして利用される。異なるキーワード決定方法が、同時に実行されてもよいし、組み合わせられてもよい。組み合わせ技術では、ユーザが入力したクエリ単語に対して、最初に単語分割が実行される。分割された単語を用いて、単語分割結果のクエリが実行される。クエリ単語の分割単語の潜在的意味を用いて、さらなるクエリが実行され、最終的に単独語のクエリが実行されてよい。単語分割結果のクエリとは、相関のある結果が、クエリ単語の単語分割結果に従って単語分割インデックスからクエリされることを指しており、単独語のクエリとは、単独語インデックス化から結果がクエリされることを指しており、潜在的意味のクエリとは、クエリ単語の潜在的語義を用いてクエリ結果が取得されることを指す。

【0035】

第1の潜在的語義を有する単語については、「クエリ単語 + 第1の潜在的語義を有する単語」によって、相関する結果がクエリされる。クエリ単語が「C17」(Motorola)である場合、第1の潜在的語義を有する対応する単語は、「C39」(Motorola company)および「C40」(Motorola mobile phone)であり、ここで、「C17」(Motorola)の第1のタイプの潜在的語義は、「C18」(company)および「C19」(mobile phone)である。第2のタイプの潜在的語義を有する単語については、第2の潜在的語義を有する「隣接するクエリ単語」を用いてクエリ結果が取得される。例えば、「C8」(glass)の第2の潜在的語義を有する対応する単語は、「C38」(glass | bottles)である。第3のタイプの潜在的語義を有する単語については、第3の潜在的語義を有する単語によってクエリ結果が取得される。例えば、「C2」(refrigerator)がクエリされた場合、第3の潜在的語義を有する単語は、「C1」(fridge)である。

10

20

30

40

50

【 0 0 3 6 】

クエリ単語とファイルとの間の相関度が、潜在的語義クエリのクエリ結果に基づいて算出される場合、それは、単独語クエリで得られた結果の相関度よりも高いことが好ましい。相関度の値は、（相関度およびウェブページの重要度などに従った）クエリ結果の順序付けに影響する。

【 0 0 3 7 】

いくつかの実施形態において、第1の組のクエリキーワードが工程104で取得される時に、第1の組のクエリキーワードを入力するユーザのユーザ特性も取得されてよい。すなわち、ユーザが第1の組のクエリキーワードを入力する時に、ユーザのユーザ特性が取得されてよい。

10

【 0 0 3 8 】

したがって、ユーザフィードバックログが工程106で取得される時に、ユーザフィードバックログは、ユーザ特性に従って取得されてもよい。

【 0 0 3 9 】

もしくは、ユーザフィードバックログが取得される時に、取得されるユーザフィードバックログは、第1の組のクエリキーワード内の単語を対象とする履歴クエリ結果と、履歴クエリ結果に対するユーザの選択頻度とを含み、これらの履歴クエリ結果は、ユーザ特性を含む。

【 0 0 4 0 】

もしくは、ユーザフィードバックログに従って第2の組のクエリキーワードを決定する時に、第2の組のクエリキーワードは、ユーザ特性に従って決定されてもよい。

20

【 0 0 4 1 】

すなわち、第2の組のクエリキーワードがユーザフィードバックログに従ってマッチングされる時に、第1の組のクエリキーワードを入力したユーザのユーザ特性に従って、異なる第2の組のクエリキーワードが決定されてもよい。ユーザ特性を用いてユーザフィードバックログを選択することにより、第1の組のクエリキーワードの潜在的語義をさらに見つけることが好ましい。例えば、上述の実施形態によると、ユーザが、「C15」を検索する場合、ユーザの実際の意図は、「C41」または「C30」(lathe)などの機械設備ではなく、ほとんどの場合、就寝のためのベッド、例えば、「C13」(single bed)、「C12」(double bed)、および、「C29」(wooden bed)である。この時、前出の検索語が「C15」(bed)の潜在的語義を有し、潜在的語義が「C41」(lathe)などを含まないことが、就寝用ベッドに関するコンテンツのユーザ選択など、ユーザのフィードバックに基づいて推測されてよい。しかしながら、同じクエリキーワード「C15」(bed)について、ユーザが機械設備の分野の技術者である場合には、潜在的語義は、「C13」(single bed)、「C12」(double bed)、および、「C29」(wooden bed)などではなく、「C41」(lathe)であることが好ましい。この実施形態では、「機械設備の分野の技術者」がユーザ特性となり、単語の潜在的語義をより良好に見つけるために、ユーザフィードバックログを分類するために用いられる。

30

【 0 0 4 2 】

別の例では、入力された第1の組のクエリキーワードは「C42」(apple:アップル)であり、ユーザ特性がコンピュータ関係の労働者である場合には、コンピュータのカテゴリの第2の組のクエリキーワードが、マッチングセットとして利用され、ユーザ特性が食品産業の労働者または農業労働者である場合には、果物のカテゴリの第2のキーワードがマッチングセットとして利用される。具体的な実施の際には、ユーザ特性は、ユーザの地域(例えば、国、地域、および、町)、ユーザが頻繁にブラウズしていた以前のウェブページ、ユーザが最近ブラウズしたウェブページ、ユーザが入力した以前の検索キーワード、ユーザの性別、年齢、職業、および、好みなどを含んでよい。ユーザ特性の解析および分類のために、IPアドレス解析、ユーザ側ブラウザ履歴データ解析、ユーザ側COOKIEデータ解析、および、ユーザウェブ登録情報の解析などの技術的手段が、必要

40

50

に応じて利用されてよく、これは、当業者にとって容易に理解されうる。

【0043】

同じ発明概念に基づいて、本願は、さらに、単語マッチングのための方法および装置、ならびに、情報クエリのための装置を提供する。単語マッチングのための方法および装置、情報クエリのための装置、ならびに、情報クエリのための方法は、同じ発明概念に基づいているため、それらは同様の原理を有する。したがって、単語マッチングのための方法および装置ならびに情報クエリのための装置の実施例では、情報クエリのための方法の実施例を参照することができるため、本明細書では、繰り返しの部分は省略する。

【0044】

クエリ処理システムの一実施形態の実装について以下で説明する。システムは、複数のモジュールを備え、モジュールは、サブモジュールまたはユニットを備える。これらのモジュール/サブモジュール/ユニットは、1または複数のプロセッサ上で実行されるソフトウェアコンポーネントとして、特定の機能を実行するよう設計されたプログラム可能論理デバイスおよび/または特定用途向け集積回路などのハードウェアとして、もしくは、それらの組み合わせとして実装することができる。いくつかの実施形態において、モジュールは、コンピュータデバイス（パーソナルコンピュータ、サーバ、ネットワーク装置など）に本発明の実施形態に記載された方法を実行させるための複数の命令など、不揮発性記憶媒体（光学ディスク、フラッシュ記憶装置、携帯用ハードディスクなど）に格納することができるソフトウェア製品の形態で具現化されてよい。モジュール/サブモジュール/ユニットは、単一のデバイス上に実装されてもよいし、複数のデバイスにわたって分散されてもよい。モジュールの機能は、互いに統合されてもよいし、複数のサブモジュールにさらに分割されてもよい。

【0045】

図2は、クエリ処理システムの一実施形態を示すブロック図である。図に示すように、システムは、ユーザ入力に基づいて第1の組のクエリキーワードを取得するための第1クエリキーワード取得モジュール201と、第1の組のクエリキーワードのユーザフィードバックログを取得するためのユーザフィードバックログ取得モジュール202と、ユーザフィードバックログに従って第1の組のクエリキーワードに一致する第2の組のクエリキーワードを決定するためのマッチングモジュール203と、第2の組のクエリキーワードを対象とするクエリ結果をフィードバックするためのクエリ結果フィードバックモジュール204と、を備える。

【0046】

実装例において、ユーザフィードバックログ取得モジュールは、さらに、ユーザフィードバックログを取得するよう適合されてよく、ユーザフィードバックログは、第1の組のクエリキーワードに含まれる単語を検索対象とした以前の全履歴クエリ結果と、ユーザによる履歴クエリ結果の以前の全選択頻度とを含む。

【0047】

マッチングモジュールは、さらに、ユーザフィードバックログ内の履歴クエリ結果および選択頻度に従って、第1の組のクエリキーワードにマッチする第2の組のクエリキーワードを決定するよう適合されてよい。

【0048】

実装例において、ユーザフィードバックログ取得モジュールは、さらに、選択頻度として、履歴クエリ結果に対する選択頻度および/または履歴クエリ結果のコンテンツに対する選択頻度を取得するよう適合されてよい。

【0049】

図3は、マッチングモジュールの一実施形態を示すブロック図である。図に示すように、マッチングモジュールは、第1の組のキーワードに対応する履歴クエリ結果のコンテンツを取得するためのコンテンツ取得ユニット2031と、履歴クエリ結果のコンテンツに対して単語分割を実行して単語分割後の単語を取得するための単語分割ユニット2032と、単語分割後の単語の選択頻度に従って、第1の組のクエリキーワードにマッチする第

10

20

30

40

50

2の組のクエリキーワードを決定するためのマッチングユニット2033と、を備える。

【0050】

実装例において、単語分割ユニットは、さらに、単語分割後に、以下のタイプの内の1または複数のタイプの単語を取得するよう適合されてよい。単語分割後に第1の組のクエリキーワードに隣接する単語、単語分割後に第1の組のクエリキーワードを含む単語、および、単語分割後に第1の組のクエリキーワードの構成部分を含む単語。

【0051】

実装例において、マッチングモジュールは、さらに、履歴クエリ結果および選択頻度に従って、第1の組のクエリキーワードにマッチする第2の組のクエリキーワードを決定するよう適合されてよく、選択頻度は、所定の閾値よりも大きい。

10

【0052】

図4は、第1クエリキーワード取得モジュールの一実施形態を示すブロック図である。図に示すように、第1クエリキーワード取得モジュールは、ユーザが入力する情報コンテンツを取得するための情報コンテンツ取得ユニット2011と、情報コンテンツに対して単語分割を実行した後に単語分割後の単語を取得する、および/または、情報コンテンツを文字に分解するための単語分割/分解ユニット2012と、単語分割後の単語および/または文字を第1の組のクエリキーワードとして選択するための第1クエリキーワード決定ユニット2013と、を備えてよい。

【0053】

いくつかの実施形態において、ユーザフィールドバックログ取得モジュールは、さらに、選択頻度として、以下の内の1つまたは組み合わせを取得するよう適合されてよい。履歴クエリ結果のクリック頻度、履歴クエリ結果の表示頻度、履歴クエリ結果の閲覧時間、および、履歴クエリ結果の重要度。

20

【0054】

いくつかの実施形態において、第1クエリキーワード取得モジュールは、さらに、ユーザが第1の組のクエリキーワードを入力する時にユーザのユーザ特性を取得するよう適合されてよく、ユーザフィールドバックログ取得モジュールは、さらに、ユーザ特性に従ってユーザフィールドバックログを取得するよう適合されてよい。

【0055】

いくつかの実施形態において、第1クエリキーワード取得モジュールは、さらに、ユーザが第1の組のクエリキーワードを入力する時にユーザのユーザ特性を取得するよう適合されてよい。

30

【0056】

ユーザフィールドバックログ取得モジュールは、さらに、ユーザフィールドバックログを取得する時に、マッチング対象単語を対象とした履歴クエリ結果と、履歴クエリ結果に対するユーザの選択頻度とを含むユーザフィールドバックログを取得するよう適合されてよく、履歴クエリ結果はユーザ特性を含む。

【0057】

いくつかの実施形態において、第1クエリキーワード取得モジュールは、さらに、ユーザが第1の組のクエリキーワードを入力する時にユーザのユーザ特性を取得するよう適合されてもよい。マッチングモジュールは、さらに、ユーザフィールドバックログに従って第2の組のクエリキーワードを決定する時に、ユーザ特性に従って第2の組のクエリキーワードを決定するよう適合されてよい。

40

【0058】

図5は、単語マッチングのための処理の一実施形態を示すフローチャートである。図に示すように、その処理は、単語マッチングが実行される際に以下の工程を含んでよい。第1の組のクエリキーワードなど、マッチング対象単語を取得する工程501、マッチング対象単語に従ってユーザフィールドバックログを取得する工程であって、ユーザフィールドバックログは、マッチング対象単語を対象とした以前の全履歴クエリ結果と、履歴クエリ結果に対するユーザの以前の全選択頻度とを含む工程502、および、履歴クエリ結果およ

50

び選択頻度に従って、マッチング対象単語にマッチする単語を決定する工程503。

【0059】

いくつかの実施形態において、選択頻度は、履歴クエリ結果に対する選択頻度および/または履歴クエリ結果のコンテンツに対する選択頻度を含む。

【0060】

いくつかの実施形態において、履歴クエリ結果のコンテンツに対する選択頻度に従って、マッチング対象単語にマッチする単語を決定する工程は：マッチング対象単語の履歴クエリ結果のコンテンツを取得する工程と；履歴クエリ結果のコンテンツに対して単語分割を実行し、単語分割後の単語を取得する工程と；単語分割後の単語の選択頻度に従って、マッチング対象単語にマッチする単語を決定する工程と、を備える。

10

【0061】

様々な実施形態において、単語分割後の単語とは、以下タイプの内の1または複数のタイプの単語を指す：単語分割後にマッチング対象単語に隣接する単語；単語分割後にマッチング対象単語を含む単語；および、単語分割後にマッチング対象単語の構成部分を含む単語。

【0062】

いくつかの実施形態において、マッチング対象単語にマッチする単語が、履歴クエリ結果および選択頻度に従って決定される場合、選択頻度は、所定の閾値よりも大きい。

【0063】

マッチング対象キーワードを取得する工程は、以下の工程を含んでよい。ユーザが入力した情報コンテンツを取得する工程、情報コンテンツに対して単語分割を実行した後に単語分割後の単語を取得する、および/または、情報コンテンツを文字に分解する工程、および、単語分割後の単語および/または文字を、マッチング対象単語として選択する工程。

20

【0064】

いくつかの実施形態において、選択頻度は、以下の内の1または組み合わせを含んでよい。履歴クエリ結果のクリック頻度、履歴クエリ結果の表示頻度、履歴クエリ結果の閲覧時間、および、履歴クエリ結果の重要度。

【0065】

いくつかの実施形態において、方法は、さらに、ユーザがマッチング対象単語を入力する時にユーザのユーザ特性を取得する工程と、ユーザフィールドバックログを取得する時にユーザ特性に従ってユーザフィールドバックログを取得する工程と、を備えてよい。

30

【0066】

いくつかの実施形態において、方法は、ユーザがマッチング対象単語を入力する時にユーザのユーザ特性を取得する工程と、ユーザフィールドバックログを取得する工程とを備えてよく、取得されるユーザフィールドバックログは、マッチング対象単語を対象とした履歴クエリ結果と、履歴クエリ結果に対するユーザの選択頻度とを含み、履歴クエリ結果はユーザ特性を含んでも良い。

【0067】

いくつかの実施形態において、方法は、さらに、ユーザがマッチング対象単語を入力する時にユーザのユーザ特性を取得する工程と、ユーザフィールドバックログに従って、マッチング対象単語にマッチする単語を決定する時に、ユーザ特性に従って、マッチング対象単語にマッチする単語を決定する工程と、を備えてよい。

40

【0068】

図6は、単語マッチングのための装置の一実施形態を示すブロック図である。図に示すように、装置は、マッチング対象単語を取得するためのマッチング対象単語取得モジュール601と、マッチング対象単語に従ってユーザフィールドバックログを取得するためのユーザフィールドバックログ取得モジュール602と、ユーザフィールドバックログに従って、マッチング対象単語にマッチする単語を決定するためのマッチングモジュール603と、を備える。

50

【 0 0 6 9 】

いくつかの実施形態において、ユーザフィールドバックログ取得モジュールは、さらに、マッチング対象単語を対象とした以前の全履歴クエリ結果と、履歴クエリ結果に対するユーザの以前の全選択頻度とを含むユーザフィールドバックログを取得するよう適合されてよく、マッチングモジュールは、さらに、ユーザフィールドバックログ内の履歴クエリ結果および選択頻度に従って、マッチング対象単語にマッチする単語を決定するよう適合されてよい。

【 0 0 7 0 】

ユーザフィールドバックログ取得モジュールは、さらに、選択頻度として、履歴クエリ結果に対する選択頻度および/または履歴クエリ結果のコンテンツに対する選択頻度を取得するよう適合されてよい。

10

【 0 0 7 1 】

いくつかの実施形態において、マッチングモジュールは：マッチング対象単語の履歴クエリ結果のコンテンツを取得するためのコンテンツ取得ユニットと、履歴クエリ結果のコンテンツに対して単語分割を実行して単語分割後の単語を取得するための単語分割ユニットと、単語分割後の単語の選択頻度に従って、マッチング対象単語にマッチする単語を決定するためのマッチングユニットと、を備える。

【 0 0 7 2 】

単語分割ユニットは、さらに、単語分割後に、以下のモードまたはそれらを組み合わせたモードの単語を取得するよう適合されてよい。単語分割後にマッチング対象単語に隣接する単語、単語分割後にマッチング対象単語を含む単語、および、単語分割後にマッチング対象単語の構成部分を含む単語。

20

【 0 0 7 3 】

マッチングモジュールは、さらに、履歴クエリ結果および選択頻度に従って、マッチング対象単語にマッチする単語を決定するよう適合されてよく、選択頻度は、所定の閾値よりも大きくても良い。

【 0 0 7 4 】

マッチング対象単語取得モジュールは、ユーザが入力する情報コンテンツを取得するための情報コンテンツ取得ユニットと、情報コンテンツに対して単語分割を実行した後に単語分割後の単語を取得する、および/または、情報コンテンツを文字に分解するための単語分割/分解ユニットと、単語分割後の単語および/または文字をマッチング対象単語として選択するためのマッチング対象単語決定ユニットと、を備えてよい。

30

【 0 0 7 5 】

ユーザフィールドバックログ取得モジュールは、さらに、選択頻度として、以下の内の1つまたは組み合わせを取得するよう適合されてよい。履歴クエリ結果のクリック頻度、履歴クエリ結果の表示頻度、履歴クエリ結果の閲覧時間、および、履歴クエリ結果の重要度。

【 0 0 7 6 】

いくつかの実施形態において、マッチング対象単語取得モジュールは、さらに、ユーザがマッチング対象単語を入力する時にユーザのユーザ特性を取得するよう適合され、ユーザフィールドバックログ取得モジュールは、さらに、ユーザ特性に従ってユーザフィールドバックログを取得するよう適合されても良い。

40

【 0 0 7 7 】

いくつかの実施形態において、マッチング対象単語取得モジュールは、さらに、ユーザがマッチング対象単語を入力する時にユーザのユーザ特性を取得するよう適合されてよく、ユーザフィールドバックログ取得モジュールは、さらに、ユーザフィールドバックログを取得する時に、マッチング対象単語を対象とした履歴クエリ結果と、履歴クエリ結果に対するユーザの選択頻度とを含むユーザフィールドバックログを取得するよう適合されてよく、履歴クエリ結果はユーザ特性を含んでも良い。

【 0 0 7 8 】

50

いくつかの実施形態において、マッチング対象単語取得モジュールは、さらに、ユーザがマッチング対象単語を入力する時にユーザのユーザ特性を取得するよう適合されてよく、マッチングモジュールは、さらに、ユーザフィードバックログに従って、マッチング対象単語にマッチする単語を決定する時に、ユーザ特性に従って、マッチング対象単語にマッチする単語を決定するよう適合されてよい。

【0079】

上述の実施形態によると、ユーザフィードバックログの解析に基づいて、単語の潜在的意味を自動的に見つけることができ、単語間の隠れた関係性を正確に見つけられることがわかる。さらに、自動的に見つけられた単語の潜在的意味と、クエリされる単語の関連する意味とを用いて、検索エンジンの効果を改善することができる。さらに、クエリ単語の潜在的語義が自動的に見つけられる時に、単語分割結果だけではなく、検索文字列内でクエリ単語に直接隣接する単語の単語頻度に従っても、同様の効果を得ることができる。したがって、本願の一実施形態によると、自動的に見つけられた単語の潜在的語義により、検索エンジンの性能を改善することができる。従来技術と比べて、検索精度および効率を改善することができる。

【0080】

説明しやすいように、上記のシステムについては、機能に従って様々なモジュールまたはユニットに分けて、それぞれを説明してきた。しかしながら、本発明の実施の際には、各モジュールまたはユニットの機能は、1または複数のソフトウェアおよび/またはハードウェアで実現される。

【0081】

当業者であれば、本願の実施形態は、方法、システム、または、コンピュータプログラム製品として提供されることを理解できる。したがって、本願は、完全にハードウェア的な実施形態、完全にソフトウェア的な実施形態、または、その組み合わせの形態を取りうる。さらに、本願は、コンピュータ利用可能なプログラムコードを含む1または複数のコンピュータ利用可能な記憶媒体（磁気ディスクストレージ、CD-ROM、および、光学ストレージを含むが、これらに限定されない）に実装されたコンピュータプログラム製品の形態を取ってもよい。

【0082】

本願は、本願の実施形態に従った方法、装置（システム）、および、コンピュータプログラム製品のフローチャートおよび/またはブロック図を参照して説明されている。フローチャートおよび/またはブロック図内の各フローおよび/またはブロック、ならびに、フローチャートおよび/またはブロック図内のフローおよび/またはブロックの組み合わせは、コンピュータプログラム命令によって実現されることを理解されたい。かかるコンピュータプログラム命令は、マシンを生成するために、汎用コンピュータ、専用コンピュータ、内蔵プロセッサ、または、他のプログラム可能なデータ処理デバイスのプロセッサに供給され、その結果、コンピュータまたはその他のプログラム可能なデータ処理デバイスのプロセッサによって実行された命令が、フローチャート内の1または複数のフローおよび/またはブロック図内の1または複数のブロックに記載された機能を実現するためのデバイスを生成しうる。

【0083】

かかるコンピュータプログラム命令は、コンピュータまたはその他のプログラム可能なデータ処理デバイスが特定のモードで機能するよう導くことができるコンピュータ読み取り可能な記憶装置に格納されてもよく、それにより、コンピュータ読み取り可能な記憶装置に格納された命令は、命令装置を含む製品を生成することが可能であり、命令装置は、フローチャート内の1または複数のフローおよびブロック図内の1または複数のブロックに記載された機能を実現することができる。

【0084】

かかるコンピュータプログラム命令は、コンピュータまたはその他のプログラム可能なデータ処理デバイスにロードされてもよく、そうすれば、一連の動作処理をコンピュータ

10

20

30

40

50

またはその他のプログラム可能なデバイスで実行して、コンピュータで処理を実現することが可能になり、それにより、コンピュータまたはその他のプログラム可能なデバイスで実行された命令は、フローチャート内の1または複数のフローおよび/またはブロック図内の1または複数のブロックに記載された機能を実現するための処理を提供しうる。

【0085】

本願の好ましい実施形態について上述したが、当業者であれば、基本的な創造概念の教示内で他の変更例および変形例を実現することが可能である。したがって、好ましい実施形態と、これらの変更例および変形例はすべて、添付の特許請求の範囲に含まれるものである。

【0086】

当業者は、本発明の精神および範囲から逸脱することなく、本発明に様々な変形および変更を施すことができることを理解されたい。したがって、本発明に対するこれらの変形例および変更例が、本発明の請求項およびその等価物の範囲に属している場合には、本発明は、これらの変形例および変更例を含むものである。

適用例1：クエリを実行する方法であって、ユーザ入力を受信し、前記ユーザ入力に少なくとも部分的に基づいて第1の組のクエリキーワードを決定し、前記第1の組のクエリキーワードの少なくとも一部に基づいて、履歴クエリ結果を含むユーザフィードバックログを取得し、前記ユーザフィードバックログに基づいて、第2の組のクエリキーワードを決定し、前記第2の組のクエリキーワードの少なくとも一部に基づいてクエリを実行する、ことを備える、方法。

適用例2：適用例1に記載の方法はさらに、クエリ結果を返信し、前記クエリ結果を用いてユーザフィードバックデータベースを更新する、ことを備える、方法。

適用例3：適用例1に記載の方法において、前記ユーザ入力は、文脈に基づいて複数の分割単語に分割され、前記第1の組のクエリキーワードは、前記複数の分割単語を含む、方法。

適用例4：適用例3に記載の方法において、前記第1の組のクエリキーワードは、前記複数の分割単語の内の少なくとも一部に隣接する1または複数の単語、前記複数の分割単語の少なくとも一部分を含む1または複数の単語、前記複数の分割単語の内の少なくとも1つの分割単語を含む1または複数の単語、または、これらの組み合わせを含む、方法。

適用例5：適用例1に記載の方法において、前記ユーザフィードバックログは、前記第1の組のクエリキーワードの少なくとも一部をクエリ対象として用いて取得された履歴クエリ結果を含む、方法。

適用例6：適用例1に記載の方法において、前記ユーザフィードバックログは、選択頻度を含む、方法。

適用例7：適用例6に記載の方法において、前記選択頻度は、前記履歴クエリ結果のクリック頻度、前記履歴クエリ結果の表示頻度、前記履歴クエリ結果の閲覧時間、前記履歴クエリ結果の重要度、または、これらの組み合わせ、に少なくとも部分的に基づいて決定される、方法。

適用例8：適用例1に記載の方法において、前記第2の組のクエリキーワードは、前記第1の組のクエリキーワードに対応する1または複数の潜在単語を含み、前記潜在単語は、前記第1の組のクエリキーワードに関する潜在的意味を有する、方法。

適用例9：適用例1に記載の方法において、前記第2の組のクエリキーワードの少なくとも一部に基づくクエリの実行は、前記第2の組のクエリキーワードの少なくとも一部を検索エンジンに送信することを含む、方法。

適用例10：適用例1に記載の方法はさらに、前記ユーザ入力を生成したユーザのユーザ特性を取得することを備え、前記ユーザフィードバックログは、前記ユーザ特性に少なくとも部分的に従って取得される、方法。

適用例11：適用例1に記載の方法はさらに、前記ユーザ入力を生成したユーザのユーザ特性を取得することを備え、前記第2の組のクエリキーワードは、前記ユーザ特性に少なくとも部分的に従って取得される、方法。

10

20

30

40

50

適用例 1 2 : クエリシステムであって、ユーザ入力を受信するよう構成されたインターフェースと、前記インターフェースに接続された 1 または複数のプロセッサであって、前記ユーザ入力に少なくとも部分的に基づいて第 1 の組のクエリキーワードを決定し、前記第 1 の組のクエリキーワードの少なくとも一部に基づいて、履歴クエリ結果を含むユーザフィードバックログを取得し、前記ユーザフィードバックログに基づいて、第 2 の組のクエリキーワードを決定し、前記第 2 の組のクエリキーワードの少なくとも一部に基づいてクエリを実行するよう構成された、1 または複数のプロセッサと、

前記 1 または複数のプロセッサに接続され、前記プロセッサに命令を提供するよう構成された 1 または複数のメモリと、を備える、システム。

適用例 1 3 : 適用例 1 2 に記載のシステムにおいて、前記プロセッサは、クエリ結果を返信し、前記クエリ結果を用いてユーザフィードバックデータベースを更新するよう構成されている、システム。

10

適用例 1 4 : 適用例 1 2 に記載のシステムにおいて、前記ユーザ入力は、文脈に基づいて複数の分割単語に分割され、前記第 1 の組のクエリキーワードは、前記複数の分割単語を含む、システム。

適用例 1 5 : 適用例 1 4 に記載のシステムにおいて、前記第 1 の組のクエリキーワードは、前記複数の分割単語の内の少なくとも一部に隣接する 1 または複数の単語、前記複数の分割単語の少なくとも一部分を含む 1 または複数の単語、前記複数の分割単語の内の少なくとも 1 つの分割単語を含む 1 または複数の単語、または、これらの組み合わせを含む、システム。

20

適用例 1 6 : 適用例 1 2 に記載のシステムにおいて、前記ユーザフィードバックログは、前記第 1 の組のクエリキーワードの少なくとも一部をクエリ対象として用いて取得された履歴クエリ結果を含む、システム。

適用例 1 7 : 適用例 1 2 に記載のシステムにおいて、前記ユーザフィードバックログは、選択頻度を含む、システム。

適用例 1 8 : 適用例 1 7 に記載のシステムにおいて、前記選択頻度は、前記履歴クエリ結果のクリック頻度、前記履歴クエリ結果の表示頻度、前記履歴クエリ結果の閲覧時間、前記履歴クエリ結果の重要度、または、これらの組み合わせ、に少なくとも部分的に基づいて決定される、システム。

適用例 1 9 : 適用例 1 2 に記載のシステムにおいて、前記第 2 の組のクエリキーワードは、前記第 1 の組のクエリキーワードに対応する 1 または複数の潜在単語を含み、前記潜在単語は、前記第 1 の組のクエリキーワードに関する潜在的意味を有する、システム。

30

適用例 2 0 : 適用例 1 2 に記載のシステムにおいて、前記第 2 の組のクエリキーワードの少なくとも一部に基づいてクエリの実行は、前記第 2 の組のクエリキーワードの少なくとも一部を検索エンジンに送信することを含む、システム。

適用例 2 1 : 適用例 1 2 に記載のシステムにおいて、前記プロセッサは、さらに、前記ユーザ入力を生成したユーザのユーザ特性を取得するよう構成され、前記ユーザフィードバックログは、前記ユーザ特性に少なくとも部分的に従って取得される、システム。

適用例 2 2 : 適用例 1 2 に記載のシステムにおいて、前記プロセッサは、さらに、前記ユーザ入力を生成したユーザのユーザ特性を取得するよう構成され、前記第 2 の組のクエリキーワードは、前記ユーザ特性に少なくとも部分的に従って取得される、システム。

40

【 0 0 8 7 】

【表 1】

C 1	冰箱	C 2 2	美女	
C 2	电冰箱	C 2 3	美丽的	
C 3	棉拖	C 2 4	女人	
C 4	全棉拖鞋	C 2 5	虾	
C 5	农药	C 2 6	对虾	
C 6	神农药厂	C 2 7	酒	
C 7	神农农药厂	C 2 8	啤酒	
C 8	玻璃瓶	C 2 9	木板床	10
C 9	玻璃瓶子	C 3 0	车床	
C 1 0	生产玻璃瓶颈在于	C 3 1	中国 古代 美女 西施 名 夷光	
C 1 1	西药	C 3 2	春秋 战国 时期 出生	
C 1 2	双人床	C 3 3	古代	
C 1 3	单人床	C 3 4	西施	
C 1 4	弹簧床	C 3 5	西施 是个 美丽的 女人	
C 1 5	床	C 3 6	美丽的 女人	
C 1 6	花床	C 3 7	电冰箱 空调器 原理 与 维修	20
C 1 7	摩托罗拉	C 3 8	玻璃 瓶子	
C 1 8	公司	C 3 9	摩托罗拉公司	
C 1 9	手机	C 4 0	摩托罗拉手机	
C 2 0	玻璃	C 4 1	机床	
C 2 1	瓶子	C 4 2	苹果	

【図1A】

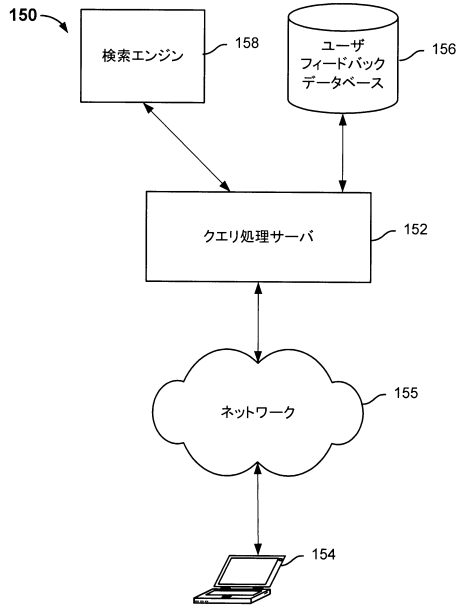


FIG. 1A

【図1B】

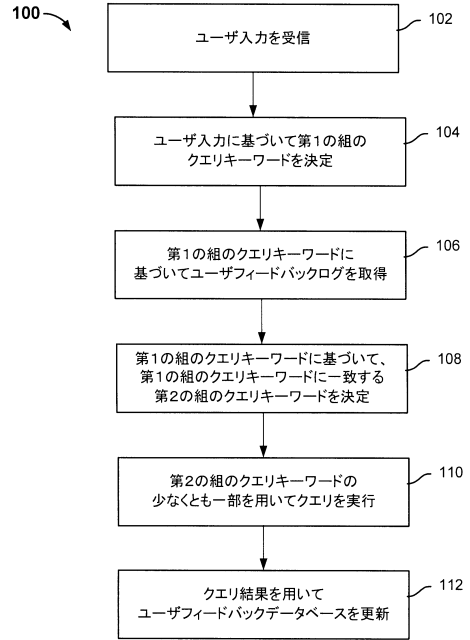


FIG. 1B

【図2】

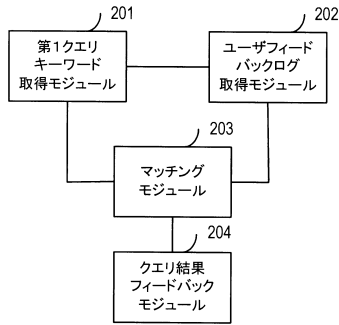


FIG. 2

【図4】

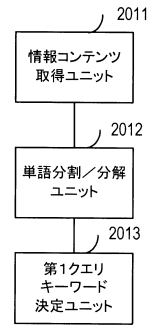


FIG. 4

【図3】

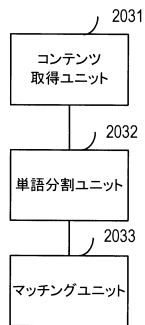


FIG. 3

【図5】

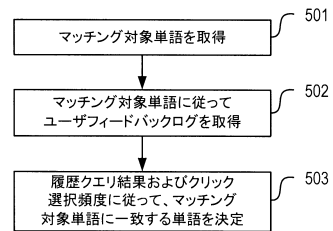


FIG. 5

【図6】

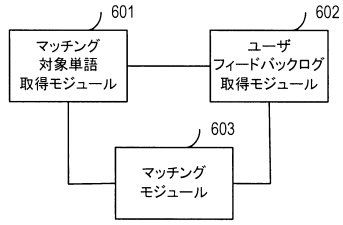


FIG. 6

フロントページの続き

(72)発明者 オウ・ウェンウ

中華人民共和国 ハンチョウ, ワーナー・ロード, ウエスト・レイク・インターナショナル・プラ
ザ, 10階, ナンバー391

審査官 鹿野 博嗣

(56)参考文献 特開2008-176746(JP, A)

特開2002-092033(JP, A)

特開2005-302042(JP, A)

特表2009-528636(JP, A)

特開平07-219957(JP, A)

特開2005-031949(JP, A)

特開2008-293211(JP, A)

(58)調査した分野(Int.Cl., DB名)

G06F 17/30