(19)日本国特許庁(JP)　(12)特　許　公　報(B2)　(11)特許番号
特許第4726528号
(P4726528)
(45)発行日　平成23年7月20日(2011.7.20)　(24)登録日　平成23年4月22日(2011.4.22)

(51)Int.Cl.　F I
G06F 17/30 (2006.01)　G06F 17/30 320D
G06F 17/30 210D
G06F 17/30 170A

請求項の数 43　外国語出願　(全 26 頁)

(21)出願番号　特願2005-118051 (P2005-118051)
(22)出願日　平成17年4月15日 (2005.4.15)
(65)公開番号　特開2005-302042 (P2005-302042A)
(43)公開日　平成17年10月27日 (2005.10.27)
審査請求日　平成20年4月11日 (2008.4.11)
(31)優先権主張番号　10/825,894
(32)優先日　平成16年4月15日 (2004.4.15)
(33)優先権主張国　米国 (US)

(73)特許権者　500046438
マイクロソフト　コーポレーション
アメリカ合衆国　ワシントン州　９８０５２－６３９９　レッドモンド　ワン　マイクロソフト　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　ベンユー　チャン
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

最終頁に続く

(54)【発明の名称】マルチセンスクエリについての関連語提案

(57)【特許請求の範囲】
【請求項１】
関連語提案のためにプロセッサにより実行されるコンピュータ実行可能方法であって、
マルチセンスクエリを介して、検索結果を調べるステップであって、
前記マルチセンスクエリが、以下のステップ、すなわち、
サブミットした用語／フレーズに意味的に関連している用語／フレーズを判定するステップであって、意味関係が、意味を判定するために、用語／フレーズの内容を検索することにより、発見されるステップと、
出現度数（ＦＯＯ）のしきい値を構成するステップと、
前記構成されたしきい値に基づき、高ＦＯＯ、または、低ＦＯＯを履歴クエリに割り振るステップと、
検索エンジンに以前にサブミットされた１組の高出現度数（ＦＯＯ）履歴クエリに関連する検索結果から用語ベクトルを生成するステップと、
用語ベクトルの計算された類似性に応じて用語クラスタを生成するステップであって、前記計算された類似性であるｓｉｍ（q_j、q_k）は、以下の式、すなわち、

$$\mathrm{sim}(q_j, q_k) = \sum_{i=1}^{d} w_{ij} \cdot w_{ik};$$

に従って決定されるステップであって、
式中で、ｄは、ベクトル次数を表し、ｑは、クエリを表し、ｋは、次数インデックスであり、
ｉ番目のベクトルのｊ番目の用語についての重みｗは、以下の式、すなわち
$w_{ij} = TF_{ij} \times \log(N / DF_j)$
に従って計算され、
式中で、TF_{ij}は、用語頻度を表し、Ｎは、クエリ用語の合計数であり、DF_jは、ｉ番目のベクトルのｊ番目の用語を含む抽出済みのフィーチャレコード数である前記決定されるステップと
を含む前記検索結果を調べるステップと、
エンティティから用語／フレーズを受信したことに応答して、１つまたは複数の関連語提案を識別するために、前記用語クラスタ中の用語／フレーズを考慮して、前記マルチセンスクエリを介して前記用語／フレーズを評価するステップであって、前記識別することは、ＦＯＯと確信度値との組み合わせに基づいて行われることを特徴とする評価するステップと、
前記ＦＯＯと確信度値との組み合わせにより順序付けられる少なくとも１つの提案された用語リストを戻すステップであって、前記複数の提案された用語リストは、前記用語／フレーズが１以上の用語クラスタ内の用語に一致するときに、生成されるステップと
を含むことを特徴とするコンピュータ実行可能方法。

【請求項２】
前記エンティティは、コンピュータプログラムアプリケーションおよび／またはエンドユーザであることを特徴とする請求項１に記載の方法。

【請求項３】
クエリログから履歴クエリを収集するステップと、
前記履歴クエリのうちの高ＦＯＯを伴う履歴クエリを決定するステップと
をさらに含むことを特徴とする請求項１に記載の方法。

【請求項４】
前記用語クラスタを作成する前に、
前記用語ベクトルのサイズを縮小するステップと、
前記用語ベクトルを正規化するステップと
をさらに含むことを特徴とする請求項１に記載の方法。

【請求項５】
評価するステップは、
前記用語／フレーズと１つまたは複数の用語クラスタからの１つ（または複数）の用語／フレーズとの間のマッチを識別するステップと、
識別するステップに応答して、前記１つ（または複数）の用語／フレーズを含む１つ（または複数）の関連語提案を生成するステップと
をさらに含むことを特徴とする請求項１に記載の方法。

【請求項６】
前記１つ（または複数）の関連語提案は、前記１つ（または複数）の用語／フレーズのうちの用語／フレーズごとに、前記用語／フレーズが１組の調べられた履歴クエリ中で発生する回数を示す出現度数値をさらに含むことを特徴とする請求項５に記載の方法。

【請求項７】

前記用語クラスタを生成するステップは、
前記各高ＦＯＯ履歴クエリを前記検索エンジンに送信して前記検索結果を取得するステップと、
前記各高ＦＯＯ履歴クエリに対応する少なくとも１サブセットの検索結果からフィーチャを抽出するステップと、
前記フィーチャから用語頻度および逆用語頻度に応じて用語ベクトルを生成するステップと
をさらに含むことを特徴とする請求項１に記載の方法。
【請求項８】
前記フィーチャは、前記高ＦＯＯ履歴クエリ用語ごとにタイトル、簡単な説明、および／またはコンテキストを含むことを特徴とする請求項７に記載の方法。
【請求項９】
前記フィーチャを抽出するステップの検索結果のサブセットは、前記検索結果のうちの上位順位付けされた結果を含むことを特徴とする請求項７に記載の方法。
【請求項１０】
前記用語クラスタは、第１の組の用語クラスタであり、前記方法は、
前記用語／フレーズと前記複数の用語／フレーズとの間にマッチが存在しないことを判定するステップと、
前記判定するステップに応答して、
各用語ベクトルが、前記検索エンジンに以前にサブミットされた１組の低ＦＯＯ履歴クエリに関連する検索結果から生成される用語ベクトルの計算された類似性から第２の組の用語クラスタを作成するステップと、
１つまたは複数の関連語提案を識別するために前記第２の組の用語クラスタのうちの用語／フレーズを考慮して前記用語／フレーズを評価するステップと
をさらに含むことを特徴とする請求項１に記載の方法。
【請求項１１】
作成するステップは、
クエリログから調べられる履歴クエリから前記低ＦＯＯ履歴クエリを識別するステップと、
少なくとも１サブセットの前記各低ＦＯＯ履歴クエリを前記検索エンジンに送信して検索結果を取得するステップと、
少なくとも１サブセットの検索結果からフィーチャを抽出するステップと、
前記フィーチャから用語頻度および逆用語頻度に応じて前記用語ベクトルを生成するステップと
をさらに含むことを特徴とする請求項１０に記載の方法。
【請求項１２】
クラスタリングの後に、
前記用語／フレーズと高ＦＯＯ履歴クエリに基づいた前記第１の組の用語クラスタからの１つ（または複数）の用語／フレーズとの間にマッチが存在しないことを判定するステップと、
前記判定するステップに応答して、前記用語／フレーズと低ＦＯＯ履歴クエリに基づいた前記第２の組の１つまたは複数の用語クラスタからの１つ（または複数）の用語／フレーズとの間のマッチを識別するステップと、
識別するステップに応答して、前記１つ（または複数）の用語／フレーズを含む１つ（または複数）の関連語提案を生成するステップと
をさらに含むことを特徴とする請求項１１に記載の方法。
【請求項１３】
プロセッサに、以下の方法を実行させるためのコンピュータ実行可能命令を格納したコンピュータ読取り可能記憶媒体であって、前記方法は、
マルチセンスクエリを介して、検索結果を調べるステップであって、

前記マルチセンスクエリが、以下のステップ、すなわち、
サブミットした用語／フレーズに意味的に関連している用語／フレーズを判定するステップであって、意味関係が、意味を判定するために、用語／フレーズの内容を検索することにより、発見されるステップと、
出現度数（ＦＯＯ）のしきい値を構成するステップと、
前記構成されたしきい値に基づき、高ＦＯＯ、または、低ＦＯＯを履歴クエリに割り振るステップと、
検索エンジンに以前にサブミットされた１組の高出現度数（ＦＯＯ）履歴クエリに関連する検索結果から用語ベクトルを生成するステップと、
用語ベクトルの計算された類似性に応じて用語クラスタを生成するステップであって、前記計算された類似性であるｓｉｍ（q_j、q_k）は、以下の式、すなわち、

$$\mathrm{sim}(q_j, q_k) = \sum_{i=1}^{d} w_{ij} \cdot w_{ik};$$

に従って決定されるステップであって、
式中で、ｄは、ベクトル次数を表し、ｑは、クエリを表し、ｋは、次数インデックスであり、
ｉ番目のベクトルのｊ番目の用語についての重みｗは、以下の式、すなわち
$w_{ij} = TF_{ij} \times \log(N / DF_j)$
に従って計算され、
式中で、TF_{ij}は、用語頻度を表し、Ｎは、クエリ用語の合計数であり、DF_jは、ｉ番目のベクトルのｊ番目の用語を含む抽出済みのフィーチャレコード数である前記決定されるステップと
を含む前記検索結果を調べるステップと、
エンティティから用語／フレーズを受信したことに応答して、１つまたは複数の関連語提案を識別するために、前記用語クラスタ中の用語／フレーズを考慮して、前記マルチセンスクエリを介して前記用語／フレーズを評価するステップであって、前記識別することは、ＦＯＯと確信度値との組み合わせに基づいて行われることを特徴とする評価するステップと、
前記ＦＯＯと確信度値との組み合わせにより順序付けられる少なくとも１つの提案された用語リストを戻すステップであって、前記複数の提案された用語リストは、前記用語／フレーズが１以上の用語クラスタ内の用語に一致するときに、生成されるステップと
を含むことを特徴とするコンピュータ読取り可能記録媒体。
【請求項１４】
前記エンティティは、コンピュータプログラムアプリケーションおよび／またはエンドユーザであることを特徴とする請求項１３に記載のコンピュータ読取り可能記録媒体。
【請求項１５】
クエリログから履歴クエリ用語を収集するコンピュータ実行可能命令と、
前記履歴クエリ用語のうちの高ＦＯＯを伴う履歴クエリ用語を決定するコンピュータ実行可能命令と
をさらに含むことを特徴とする請求項１３に記載のコンピュータ読取り可能記録媒体。
【請求項１６】
前記用語クラスタを作成する前に、
前記用語ベクトルのサイズを縮小するコンピュータ実行可能命令と、
前記用語ベクトルを正規化するコンピュータ実行可能命令と
をさらに含むことを特徴とする請求項１３に記載のコンピュータ読取り可能記録媒体。

【請求項１７】
マルチセンスクエリを介して、評価するステップは、
前記用語／フレーズと１つまたは複数の用語クラスタからの１つ（または複数）の用語／フレーズとの間のマッチを識別するコンピュータ実行可能命令と、
識別するステップに応答して、前記１つ（または複数）の用語／フレーズを含む１つ（または複数）の関連語提案を生成するコンピュータ実行可能命令と
をさらに含むことを特徴とする請求項１３に記載のコンピュータ読取り可能記録媒体。
【請求項１８】
前記１つ（または複数）の関連語提案は、前記１つ（または複数）の用語／フレーズのうちの用語／フレーズごとに、前記用語／フレーズが１組の調べられた履歴クエリ中で発生する回数を示す出現度数値をさらに含むことを特徴とする請求項１７に記載のコンピュータ読取り可能記録媒体。
【請求項１９】
前記用語クラスタを生成するステップは、
前記各高ＦＯＯ履歴クエリを前記検索エンジンに送信して前記検索結果を取得するコンピュータ実行可能命令と、
前記各高ＦＯＯ履歴クエリに対応する少なくとも１サブセットの検索結果からフィーチャを抽出するコンピュータ実行可能命令と、
前記フィーチャから用語頻度および逆用語頻度に応じて用語ベクトルを生成するコンピュータ実行可能命令と
をさらに含むことを特徴とする請求項１３に記載のコンピュータ読取り可能記録媒体。
【請求項２０】
前記フィーチャは、前記高ＦＯＯ履歴クエリ用語ごとにタイトル、簡単な説明、および／またはコンテキストを含むことを特徴とする請求項１９に記載のコンピュータ読取り可能記録媒体。
【請求項２１】
前記フィーチャを抽出するステップの検索結果のサブセットは、前記検索結果のうちの上位順位付けされた結果を含むことを特徴とする請求項１９に記載のコンピュータ読取り可能記録媒体。
【請求項２２】
前記用語クラスタは、第１の組の用語クラスタであり、前記コンピュータ実行可能命令は、
前記用語／フレーズと前記複数の用語／フレーズとの間にマッチが存在しないことを判定する命令と、
前記判定するステップに応答して、
各用語ベクトルが、前記検索エンジンに以前にサブミットされた１組の低ＦＯＯ履歴クエリに関連する検索結果から生成される用語ベクトルの計算された類似性から第２の組の用語クラスタを作成する命令と、
前記第２の組の用語クラスタのうちの用語／フレーズを考慮して前記用語／フレーズを評価し、１つまたは複数の関連語提案を識別する命令と
をさらに含むことを特徴とする請求項１３に記載のコンピュータ読取り可能記録媒体。
【請求項２３】
第２の組の用語クラスタを作成するステップは、
クエリログから調べられる履歴クエリから前記低ＦＯＯ履歴クエリを識別するコンピュータ実行可能命令と、
少なくとも１サブセットの前記各低ＦＯＯ履歴クエリを前記検索エンジンに送信して検索結果を取得するコンピュータ実行可能命令と、
少なくとも１サブセットの検索結果からフィーチャを抽出するコンピュータ実行可能命令と、
前記フィーチャから用語頻度および逆用語頻度に応じて前記用語ベクトルを生成するコ

ンピュータ実行可能命令と
をさらに含むことを特徴とする請求項２２に記載のコンピュータ読取り可能記録媒体。
【請求項２４】
クラスタリングの後に、
前記用語／フレーズと高ＦＯＯ履歴クエリに基づいた前記第１の組の用語クラスタからの１つ（または複数）の用語／フレーズとの間にマッチが存在しないことを判定するコンピュータ実行可能命令と、
前記判定するステップに応答して、前記用語／フレーズと低ＦＯＯ履歴クエリに基づいた前記第２の組の１つまたは複数の用語クラスタからの１つ（または複数）の用語／フレーズとの間のマッチを識別するコンピュータ実行可能命令と、
識別するステップに応答して、前記１つ（または複数）の用語／フレーズを含む１つ（または複数）の関連語提案を生成するコンピュータ実行可能命令と
をさらに含むことを特徴とする請求項２３に記載のコンピュータ読取り可能記録媒体。
【請求項２５】
プロセッサと、
前記プロセッサに結合されたメモリであって、前記プロセッサに、以下の方法を実行させるためのコンピュータ実行可能命令を格納した前記メモリとを備え、
前記方法は、
マルチセンスクエリを介して、検索結果を調べるステップであって、
前記マルチセンスクエリが、以下のステップ、すなわち、
サブミットした用語／フレーズに意味的に関連している用語／フレーズを判定するステップであって、意味関係が、意味を判定するために、用語／フレーズの内容を検索することにより、発見されるステップと、
出現度数（ＦＯＯ）のしきい値を構成するステップと、
前記構成されたしきい値に基づき、高ＦＯＯ、または、低ＦＯＯを履歴クエリに割り振るステップと、
検索エンジンに以前にサブミットされた１組の高出現度数（ＦＯＯ）履歴クエリに関連する検索結果から用語ベクトルを生成するステップと、
用語ベクトルの計算された類似性に応じて用語クラスタを生成するステップであって、前記計算された類似性であるｓｉｍ（q_j、q_k）は、以下の式、すなわち、

$$\mathrm{sim}(q_j, q_k) = \sum_{i=1}^{d} w_{ij} \cdot w_{ik};$$

に従って決定されるステップであって、
式中で、ｄは、ベクトル次数を表し、ｑは、クエリを表し、ｋは、次数インデックスであり、
ｉ番目のベクトルのｊ番目の用語についての重みｗは、以下の式、すなわち
$w_{ij} = TF_{ij} \times \log(N / DF_j)$
に従って計算され、
式中で、TF_{ij}は、用語頻度を表し、Ｎは、クエリ用語の合計数であり、DF_jは、ｉ番目のベクトルのｊ番目の用語を含む抽出済みのフィーチャレコード数である前記決定されるステップと
を含む前記検索結果を調べるステップと、
エンティティから用語／フレーズを受信したことに応答して、１つまたは複数の関連語提案を識別するために、前記用語クラスタ中の用語／フレーズを考慮して、前記マルチセンスクエリを介して前記用語／フレーズを評価するステップであって、前記識別すること

は、ＦＯＯと確信度値との組み合わせに基づいて行われることを特徴とする評価するステップと、
前記ＦＯＯと確信度値との組み合わせにより順序付けられる少なくとも１つの提案された用語リストを戻すステップであって、前記複数の提案された用語リストは、前記用語／フレーズが１以上の用語クラスタ内の用語に一致するときに、生成されるステップと
を含む、
ことを特徴とするコンピューティングデバイス。
【請求項２６】
前記エンティティは、コンピュータプログラムアプリケーションおよび／またはエンドユーザであることを特徴とする請求項２５に記載のコンピューティングデバイス。
【請求項２７】
クエリログから履歴クエリ用語を収集するコンピュータ実行可能命令と、
前記履歴クエリ用語のうちの高ＦＯＯを伴う履歴クエリ用語を決定するコンピュータ実行可能命令と
をさらに含むことを特徴とする請求項２５に記載のコンピューティングデバイス。
【請求項２８】
前記用語クラスタを作成する前に、
前記用語ベクトルのサイズを縮小するコンピュータ実行可能命令と、
前記用語ベクトルを正規化するコンピュータ実行可能命令と
をさらに含むことを特徴とする請求項２５に記載のコンピューティングデバイス。
【請求項２９】
評価するステップは、
前記用語／フレーズと１つまたは複数の用語クラスタからの１つ（または複数）の用語／フレーズとの間のマッチを識別するコンピュータ実行可能命令と、
識別するステップに応答して、前記１つ（または複数）の用語／フレーズを含む１つ（または複数）の関連語提案を生成するコンピュータ実行可能命令と
をさらに含むことを特徴とする請求項２５に記載のコンピューティングデバイス。
【請求項３０】
前記１つ（または複数）の関連語提案は、前記１つ（または複数）の用語／フレーズのうちの用語／フレーズごとに、前記用語／フレーズが１組の調べられた履歴クエリ中で発生する回数を示す出現度数値をさらに含むことを特徴とする請求項２９に記載のコンピューティングデバイス。
【請求項３１】
前記用語クラスタを生成するステップは、
前記各高ＦＯＯ履歴クエリを前記検索エンジンに送信して前記検索結果を取得するコンピュータ実行可能命令と、
前記各高ＦＯＯ履歴クエリに対応する少なくとも１サブセットの検索結果からフィーチャを抽出するコンピュータ実行可能命令と、
前記フィーチャから用語頻度および逆用語頻度に応じて用語ベクトルを生成するコンピュータ実行可能命令と
をさらに含むことを特徴とする請求項２５に記載のコンピューティングデバイス。
【請求項３２】
前記フィーチャは、前記高ＦＯＯ履歴クエリ用語ごとにタイトル、簡単な説明、および／またはコンテキストを含むことを特徴とする請求項３１に記載のコンピューティングデバイス。
【請求項３３】
前記フィーチャを抽出するステップの検索結果のサブセットは、前記検索結果のうちの上位順位付けされた結果を含むことを特徴とする請求項３１に記載のコンピューティングデバイス。
【請求項３４】

前記用語クラスタは、第１の組の用語クラスタであり、前記コンピュータ実行可能命令は、
前記用語／フレーズと前記複数の用語／フレーズとの間にマッチが存在しないことを判定する命令と、
前記判定するステップに応答して、
各用語ベクトルが、前記検索エンジンに以前にサブミットされた１組の低ＦＯＯ履歴クエリに関連する検索結果から生成される用語ベクトルの計算された類似性から第２の組の用語クラスタを作成する命令と、
前記第２の組の用語クラスタのうちの用語／フレーズを考慮して前記用語／フレーズを評価し、１つまたは複数の関連語提案を識別する命令と
をさらに含むことを特徴とする請求項２５に記載のコンピューティングデバイス。
【請求項３５】
第２の組の用語クラスタを作成するステップは、
クエリログから調べられる履歴クエリから前記低ＦＯＯ履歴クエリを識別するコンピュータ実行可能命令と、
少なくとも１サブセットの前記各低ＦＯＯ履歴クエリを前記検索エンジンに送信して検索結果を取得するコンピュータ実行可能命令と、
少なくとも１サブセットの検索結果からフィーチャを抽出するコンピュータ実行可能命令と、
前記フィーチャから用語頻度および逆用語頻度に応じて前記用語ベクトルを生成するコンピュータ実行可能命令と
をさらに含むことを特徴とする請求項３４に記載のコンピューティングデバイス。
【請求項３６】
クラスタリングの後に、
前記用語／フレーズと高ＦＯＯ履歴クエリに基づいた前記第１の組の用語クラスタからの１つ（または複数）の用語／フレーズとの間にマッチが存在しないことを判定するコンピュータ実行可能命令と、
前記判定するステップに応答して、前記用語／フレーズと低ＦＯＯ履歴クエリに基づいた前記第２の組の１つまたは複数の用語クラスタからの１つ（または複数）の用語／フレーズとの間のマッチを識別するコンピュータ実行可能命令と、
識別するステップに応答して、前記１つ（または複数）の用語／フレーズを含む１つ（または複数）の関連語提案を生成するコンピュータ実行可能命令と
をさらに含むことを特徴とする請求項３５に記載のコンピューティングデバイス。
【請求項３７】
プロセッサと、
マルチセンスクエリを介して、検索結果を調べる手段であって、
前記マルチセンスクエリが、以下の手段、すなわち、
サブミットした用語／フレーズに意味的に関連している用語／フレーズを判定するステップであって、意味関係が、意味を判定するために、用語／フレーズの内容を検索することにより、発見される手段と、
出現度数（ＦＯＯ）のしきい値を構成する手段と、
前記構成されたしきい値に基づき、高ＦＯＯ、または、低ＦＯＯを履歴クエリに割り振る手段と、
検索エンジンに以前にサブミットされた１組の高出現度数（ＦＯＯ）履歴クエリに関連する検索結果から用語ベクトルを生成する手段と、
用語ベクトルの計算された類似性に応じて用語クラスタを生成する手段であって、前記計算された類似性である$sim(q_j, q_k)$は、以下の式、すなわち、

$$\mathrm{sim}(q_j, q_k) = \sum_{i=1}^{d} w_{ij} \cdot w_{ik};$$

に従って決定される手段であって、
式中で、ｄは、ベクトル次数を表し、ｑは、クエリを表し、ｋは、次数インデックスであり、
ｉ番目のベクトルのｊ番目の用語についての重みｗは、以下の式、すなわち
$w_{ij} = TF_{ij} \times \log(N / DF_j)$
に従って計算され、
式中で、TF_{ij}は、用語頻度を表し、Ｎは、クエリ用語の合計数であり、DF_jは、ｉ番目のベクトルのｊ番目の用語を含む抽出済みのフィーチャレコード数である前記決定される手段と
を含む前記検索結果を調べる手段と、
エンティティから用語／フレーズを受信したことに応答して、１つまたは複数の関連語提案を識別するために、前記用語クラスタ中の用語／フレーズを考慮して、前記マルチセンスクエリを介して前記用語／フレーズを評価する手段であって、前記識別することは、ＦＯＯと確信度値との組み合わせに基づいて行われることを特徴とする評価する手段と、
前記ＦＯＯと確信度値との組み合わせにより順序付けられる少なくとも１つの提案された用語リストを戻す手段であって、前記複数の提案された用語リストは、前記用語／フレーズが１以上の用語クラスタ内の用語に一致するときに、生成される手段と
を含む、
ことを特徴とするコンピューティングデバイス。

【請求項３８】
前記エンティティは、コンピュータプログラムアプリケーションおよび／またはエンドユーザであることを特徴とする請求項３７に記載のコンピューティングデバイス。

【請求項３９】
クエリログから履歴クエリ用語を収集する収集手段と、
前記履歴クエリ用語のうちの高ＦＯＯを伴う履歴クエリ用語を決定する決定手段と
をさらに備えることを特徴とする請求項３７に記載のコンピューティングデバイス。

【請求項４０】
前記評価手段は、
前記用語／フレーズと１つまたは複数の用語クラスタからの１つ（または複数）の用語／フレーズとの間のマッチを識別する識別手段と、
識別するステップに応答して、前記１つ（または複数）の用語／フレーズを含む１つ（または複数）の関連語提案を生成する生成手段と
をさらに備えることを特徴とする請求項３７に記載のコンピューティングデバイス。

【請求項４１】
前記用語クラスタを生成する前記生成手段は、
前記各高ＦＯＯ履歴クエリを前記検索エンジンに送信して前記検索結果を取得する送信手段と、
前記各高ＦＯＯ履歴クエリに対応する少なくとも１サブセットの検索結果からフィーチャを抽出する抽出手段と、
前記フィーチャから用語ベクトルを生成する生成手段と
をさらに備えることを特徴とする請求項３７に記載のコンピューティングデバイス。

【請求項４２】
前記用語クラスタは、第１の組の用語クラスタであり、前記コンピューティングデバイ

スは、
前記用語／フレーズと前記複数の用語／フレーズとの間にマッチが存在しないことを判定する判定手段と、
前記判定するステップに応答して、
各用語ベクトルが、前記検索エンジンに以前にサブミットされた１組の低ＦＯＯ履歴クエリに関連する検索結果から生成される用語ベクトルの計算された類似性から第２の組の用語クラスタを作成する作成手段と、
前記第２の組の用語クラスタのうちの用語／フレーズを考慮して前記用語／フレーズを評価し、１つまたは複数の関連語提案を識別する評価手段と
をさらに備えることを特徴とする請求項３７に記載のコンピューティングデバイス。
【請求項４３】
前記用語／フレーズと高ＦＯＯ履歴クエリに基づいた前記第１の組の用語クラスタからの１つ（または複数）の用語／フレーズとの間にマッチが存在しないことを計算する計算手段と、
前記計算するステップに応答して、前記用語／フレーズと低ＦＯＯ履歴クエリに基づいた前記第２の組の１つまたは複数の用語クラスタからの１つ（または複数）の用語／フレーズとの間のマッチを識別する識別手段と、
識別するステップに応答して、前記１つ（または複数）の用語／フレーズを含む１つ（または複数）の関連語提案を生成する生成手段と
をさらに備えることを特徴とする請求項４２に記載のコンピューティングデバイス。
【発明の詳細な説明】
【技術分野】
【０００１】
本発明のシステムおよび方法は、データマイニングに関する。
【背景技術】
【０００２】
キーワードまたはフレーズは、ＷＷＷ（ワールドワイドウェブ）上の関連したウェブページ／サイトを求めて検索する際におけるウェブサーファ（Ｗｅｂ　ｓｕｒｆｅｒ）が検索エンジンにサブミット（ｓｕｂｍｉｔ）する単語または１組の用語である。検索エンジンは、そのページ／サイト上に現れるこれらのキーワードおよびキーワードフレーズに基づいてウェブサイトの関連性を決定する。ウェブサイトトラフィックのかなりの割合は、検索エンジンの使用からもたらされるので、ウェブサイトプロモータは、適切なキーワード／フレーズの選択が、サイトトラフィックを増加させて所望のサイトを見せるために不可欠であることを知っている。検索エンジン結果最適化（ｓｅａｒｃｈ　ｅｎｇｉｎｅ　ｒｅｓｕｌｔ　ｏｐｔｉｍｉｚａｔｉｏｎ）のためにウェブサイトに関連するキーワードを識別する技法は、例えばウェブサイトのコンテンツ、および関連したキーワードを識別する目的の人間による評価を含んでいる。この評価には、キーワード人気ツール（ｋｅｙｗｏｒｄ　ｐｏｐｕｌａｒｉｔｙ　ｔｏｏｌ）の使用を含むこともある。かかるツールは、どれだけ多くの人々が、検索エンジンに対するキーワードを含めて、特定のキーワードまたはフレーズをサブミットしたかを決定する。そのウェブサイトに関連しており、検索クエリを生成するに際して非常に頻繁に使用されると決定されたキーワードが、一般にそのウェブサイトに関して検索エンジン結果最適化のために選択される。
【０００３】
そのウェブサイトの検索エンジン結果最適化のために１組のキーワードを識別した後に、プロモータは、その検索エンジン結果中で（他のウェブサイトの検索エンジン結果の表示位置に比べて）さらに高い位置にウェブサイトを進めたいと望むこともある。この目的を達成するために、このプロモータは、この１つ（または複数）のキーワードに値を付けて、ウェブサーファがこの１つ（または複数）のキーワードに関連するこのプロモータのリストをクリックするたびに、このプロモータがいくら支払うかを指し示す。換言すれば、キーワードの入札が、ペイパークリック入札（ｐａｙ－ｐｅｒ－ｃｌｉｃｋ　ｂｉｄ）

になっている。この同じキーワードについての他の入札に比べてこのキーワード入札の金額が多額になればなるほど、この検索エンジンは、このキーワードに基づいて検索結果中でより高い位置に（重要性に関してより目立つように）この関連するウェブサイトを表示することになる。

【発明の開示】

【発明が解決しようとする課題】

【０００４】

以上に鑑みて、ウェブサイトコンテンツに関連したキーワードをより良好に識別するシステムおよび方法があれば、ウェブサイトプロモータによって歓迎されるはずである。これにより、これらのプロモータは、ユーザの好ましい用語を入札することができるようになる。これらのシステムおよび方法が、人間がウェブサイトコンテンツを評価して、検索エンジン最適化およびキーワード入札のための関連したキーワードを識別する必要性と無関係になるとすれば、それは理想的である。

【課題を解決するための手段】

【０００５】

関連語提案（ｒｅｌａｔｅｄ　ｔｅｒｍ　ｓｕｇｇｅｓｔｉｏｎ）についてのシステムおよび方法について説明している。一態様においては、用語クラスタ（ｔｅｒｍ　ｃｌｕｓｔｅｒ）が、用語ベクトル（ｔｅｒｍ　ｖｅｃｔｏｒ）の計算された類似性に応じて生成される。検索結果から生成されている各用語ベクトルは、検索エンジンに以前にサブミットされた１組の高ＦＯＯ（ｆｒｅｑｕｅｎｃｙ　ｏｆ　ｏｃｃｕｒｒｅｎｃｅ出現度数）履歴クエリに関連したものである。あるエンティティから用語／フレーズを受信するのに応答して、この用語／フレーズを、用語クラスタ中の用語／フレーズに鑑みて評価して、１つまたは複数の関連語提案を識別する。

【０００６】

図面中において、コンポーネント参照番号のいちばん左側の桁は、そのコンポーネントが最初に現れる特定の図面を識別している。

【発明を実施するための最良の形態】

【０００７】

概要

関連した用語／フレーズを提案する最も簡単な方法は、サブストリングマッチングアプローチ（ｓｕｂｓｔｒｉｎｇ　ｍａｔｃｈｉｎｇ　ａｐｐｒｏａｃｈ）を使用することであると思われ、このアプローチは、１つの用語／フレーズが、別の用語／フレーズの単語の一部または全部を含むときに２つの用語／フレーズが関連していると判断するものである。しかし、この技法は、実質的には制限されている。関連語は、共通の単語を含む必要がないので、この方法は、多くの意味的に関連した用語を無視してしまうこともある。例えば、履物会社が「ｓｈｏｅ（靴）」についての関連語を知りたいと思うことについて考察してみる。従来のマッチングアプローチが使用される場合には、「ｗｏｍｅｎ’ｓ　ｓｈｏｅｓ（婦人靴）」、「ｄｉｓｃｏｕｎｔ　ｓｈｏｅｓ（ディスカウント靴）」などだけしか提案されないことになる。しかし、「ｓｎｅａｋｅｒｓ（スニーカ）」、「ｈｉｋｉｎｇ　ｂｏｏｔｓ（ハイキングブーツ）」、「Ｎｉｋｅ（ナイキ）」などの他の多くの関連語も存在する。

【０００８】

マルチセンスクエリについての関連語を提案するための以下のシステムおよび方法は、従来のサブストリングマッチング技法のこれらの制限に対処している。この目的を達成するためにこれらのシステムおよび方法では、エンドユーザ（例えば、ウェブサイトプロモータ、広告主など）がサブミットした用語／フレーズに意味的に関連している用語／フレーズについての情報を得るために検索エンジン結果が調べられる。この意味的な関係については、用語／フレーズの意味を考慮することができる用語／フレーズを取り巻くコンテキスト（例えば、テキストなど）をこれらの検索エンジン結果から調べることによって構築することができる。より詳細には、１組のクエリ用語を履歴クエリログから集めて、そ

れらの出現度数（ＦＯＯ）がカウントされる。これらのクエリ用語は、１つずつその検索エンジンにサブミットされる。一実装形態においては、このサブミットされた履歴クエリログ用語は、他の履歴クエリログ用語の出現度数に比べて相対的に高い出現度数を有している。

【０００９】

これらのサブミットされた各クエリを受信するのに応答して、この検索エンジンは、このサブミットされたクエリの周囲のＵＲＬ、結果タイトル、各結果の簡単な説明および／またはコンテキストを含めて、検索結果の順位付けされたリストを返す。検索エンジン結果が受信された後に、これらのシステムおよび方法は、これらの返された検索結果（例えば、１つまたは複数の上位に順位付けされた結果）のうちの選択された結果から１組のフィーチャ（キーワードおよび知られているＴＦＩＤＦ技法を使用して計算される対応する重み）を抽出する。これらのサブミットされた検索クエリからの対応する検索エンジン結果のフィーチャを抽出した後に、これらの抽出されたフィーチャが正規化される。これらの正規化されたフィーチャを使用して、各サブミットされたクエリが表され、またこれらの正規化されたフィーチャをテキストクラスタリングアルゴリズム（ｔｅｘｔ　ｃｌｕｓｔｅｒｉｎｇ　ａｌｇｏｒｉｔｈｍ）中で使用して、サブミットされたクエリ用語がクラスタにグループ化される。

【００１０】

そのエンドユーザからこの用語／フレーズを受信するのに応答して、この用語／フレーズは、これらの用語クラスタ中の各用語／フレーズと比較される。これらの用語クラスタは互いにコンテキスト的に関連した用語を含んでいるので、この用語／フレーズがこれらクラスタ内の用語と比較されるときに、この用語／フレーズは、任意の複数の関連したコンテキストまたは「意味」に鑑みて評価される。一実装形態においては、用語／フレーズがクラスタからの用語とマッチングする場合、このクラスタは、提案される用語リストの形でそのエンドユーザに返される。この提案される用語リストは、この用語／フレーズと、各用語／フレーズ間の類似性評価値（確信度値）と、各用語／フレーズの出現度数（ＦＯＯ）とに意味的および／またはコンテキスト的に関連づけて決定される用語／フレーズを含んでいる。この返されたリストは、ＦＯＯおよび確信度値の組合せによって順序付けられる。この用語／フレーズが、複数の用語クラスタ中の用語とマッチングする場合には、複数の提案される用語リストが生成される。これらのリストは、これらのクラスタサイズによって順序付けられる。また各リスト内の用語は、ＦＯＯおよび確信度値の組合せによって順序付けられる。マッチングクラスタが識別されない場合には、このクエリ用語はさらに、低ＦＯＯのクエリ用語から生成される拡張クラスタに対してマッチングが行われる。

【００１１】

一実装形態においては、低ＦＯＯを伴うクエリ用語は、高出現度数の履歴クエリログ用語から生成される用語クラスタについて分類機構（ｃｌａｓｓｉｆｉｅｒ）（例えば、Ｋ個の最近接分類機構）をトレーニングすることによってクラスタ化される。低出現度数を有すると決定された履歴クエリ用語が、１つずつこの検索エンジンに対してサブミットされる。次いで、フィーチャが、この返された検索結果のうちの選択された結果（例えば、第１の上位に順位付けされたウェブページなど）から抽出される。この抽出されたフィーチャは正規化され、これを使用して低ＦＯＯを伴うクエリ用語が表される。次いで、これらのクエリ用語を既存のクラスタに分類して、このトレーニングされた分類機構に基づいて拡張クラスタが生成される。次いでこのエンドユーザがサブミットした用語／フレーズをこれらの拡張クラスタに鑑みて評価して、提案される用語リストを識別しこのエンドユーザに返す。

【００１２】

次に、マルチセンスクエリについての関連した用語／キーワードを提案するためのこれらおよび他の態様のシステムおよび方法について、より詳細に説明する。

【００１３】

例示のシステム

図面を参照すると、マルチセンスクエリについての関連語を提案するためのシステムおよび方法が、適切なコンピューティング環境において実装されるものとして説明され示されている。図面では、同様な参照番号は、同様なエレメントを示している。必要という訳ではないが、本発明は、パーソナルコンピュータによって実行されるコンピュータ実行可能命令（プログラムモジュール）の一般的な文脈で説明される。プログラムモジュールは一般に、特定のタスクを実施し、または特定の抽象データ型を実装するルーチン、プログラム、オブジェクト、コンポーネント、データ構造などを含んでいる。システムおよび方法は、前述の文脈で説明されるが、以降に説明する動作およびオペレーションは、ハードウェアで実装することもできる。

【００１４】

図１は、マルチセンスクエリについての関連語を提案するための例示のシステム１００を示している。この実装形態においては、システム１００は、ネットワーク１０４を横切ってクライアントコンピューティングデバイス１０６に結合されたＥＶＳ（ｅｄｉｔｏｒｉａｌ　ｖｅｒｉｆｉｃａｔｉｏｎ　ｓｅｒｖｅｒ編集検証サーバ）１０２を含んでいる。例えばクライアントコンピューティングデバイス１０６から、またはＥＶＳ１０２上で実行される別のアプリケーション（図示せず）から用語／フレーズ１０８を受信するのに応答して、ＥＶＳ１０２は、提案される用語リスト１１０を生成し、このクライアントコンピューティングデバイス１０６に伝えて、これによりエンドユーザは、この用語／フレーズに実際に入札するのに先立ってこの用語／フレーズ１０８に意味的／コンテキスト的に関連する１組の用語を評価できるようになる。ネットワーク１０４は、オフィス、企業規模のコンピュータネットワーク、イントラネット、およびインターネットにおいて一般的になっている通信環境など、ＬＡＮ（ローカルエリアネットワーク）通信環境、および一般的なＷＡＮ（ワイドエリアネットワーク）通信環境の任意の組合せを含むことができる。システム１００が、クライアントコンピューティングデバイス１０６を含むとき、このクライアントコンピューティングデバイスは、パーソナルコンピュータ、ラップトップ、サーバ、モバイルコンピューティングデバイス（例えば、セルラ電話、携帯情報端末（ｐｅｒｓｏｎａｌ　ｄｉｇｉｔａｌ　ａｓｓｉｓｔａｎｔ）、またはハンドヘルドコンピュータ）など、どのようなタイプのコンピューティングデバイスでもよい。

【００１５】

提案される用語リスト１１０は、例えば、この用語／フレーズ１０８と、各用語／フレーズ１０８間の類似性評価値（確信度値）と、各用語／フレーズ出現度数（ＦＯＯ）（履歴クエリログにおける頻度）とに関連していると決定された用語／フレーズを含んでいる。関連した用語／フレーズを識別し、類似性評価値を生成し、ＦＯＯ値を生成するための技法については、キーワードマイニング（ｋｅｙｗｏｒｄ　ｍｉｎｉｎｇ）、フィーチャ抽出、および用語クラスタリングとタイトルが付けられたセクションに関連して以下で非常に詳細に説明している。

【００１６】

表１は、「ｍａｉｌ（メール）」の用語／フレーズ１０８に関連していると決定された用語の例示の提案される用語リスト１１０を示している。用語／フレーズ１０８に関連した用語は、この実施例においては、提案される用語とタイトルが付けられたカラム１中に示される。

【００１７】

【表１】

表１

入札用語「ＭＡＩＬ」についての例示の提案される用語リスト

提案される用語	類似性	頻度	〈コンテキスト〉
hotmail	0.246142	93161	オンライン電子メール関連
yahoo	0.0719463	165722	
mail.com	0.352664	1455	
yahoo mail	0.0720606	39376	
www.mail.com	0.35367	711	
email.com	0.484197	225	
www.hot	0.186565	1579	
www.msn.com	0.189117	1069	
mail.yahoo.com	0.0968248	4481	
free email	0.130611	1189	
www.aolmail.com	0.150844	654	
check mail	0.221989	66	
check email	0.184565	59	
msn passport	0.12222	55	
www.webmail.aol.com	0.0800538	108	
webmail.yahoo.com	0.08789	71	
free email account	0.0836481	65	
提案される用語	類似性	頻度	
mail	1	2191	従来メール関連
usps	0.205141	4316	
usps.com	0.173754	779	
united parcel service	0.120837	941	
postal rates	0.250423	76	
stamps	0.156702	202	
stamp collecting	0.143618	152	
state abbreviations	0.104614	300	
postal	0.185255	66	
postage	0.180112	55	
postage rates	0.172722	51	
usps zip codes	0.138821	78	
us postmaster	0.109844	58	

【００１８】

表１を参照すると、この提案される用語リスト中の用語は、用語類似性値（「類似性」とタイトルの付けられたカラム２を参照）と出現度数スコア（「頻度」とタイトルが付けられたカラム３を参照）にマッピングされることに留意されたい。「用語クラスタリング」とタイトルが付けられたセクション中において以下で説明しているように計算された各用語の類似性値は、対応する提案される用語（カラム１）とこの実施例においては「ｍａｉｌ」となる用語／フレーズ１０８との間の類似性評価値を提供する。各頻度値またはス

コアは、この提案される用語が履歴クエリログ中に現れる回数を指し示す。この提案される用語リストは、用語の類似性、および／またはビジネスゴールに応じた出現度数スコアに応じてソートされる。

【００１９】

所与の任意の用語／フレーズ１０８（例えば、ｍａｉｌなど）は、その内部で入札用語を使用することができる複数のコンテキストを有することもある。これを明らかにするために、ＳＴＳモジュール１１２は、どの提案される用語の提案される用語リスト１１０が用語／フレーズ１０８の複数のコンテキストのうちのどれに対応するかについての指示を提供する。例えば、表１を参照すると、「ｍａｉｌ」の用語／フレーズ１０８は、２つのコンテキスト、すなわち（１）従来のオフラインメール、および（２）オンライン電子メールを有する。関連語の各リストが、これらの２つの各入札用語コンテキストについて示されることに留意されたい。

【００２０】

さらに、任意の用語／フレーズ１０８についての提案される用語は、入札用語の同義語以上のものとすることができる。例えば、表１を参照すると、この提案される用語「ｕｓｐｓ」は、メールを扱う組織についての頭字語（ａｃｒｏｎｙｍ）であるが、入札用語「ｍａｉｌ（メール）」についての同義語ではない。しかし、「ｕｓｐｓ」は、「ｍａｉｌ」入札用語に非常に関連した用語でもあり、したがってこの提案される用語リスト１１０中に示される。一実装形態においては、ＳＴＳモジュール１１２は、関連付けルール、すなわちｉｔｒ（Ｔ）→ｉｔｒ（Ｒ）に応じて関連語Ｒ（例えば「ｕｓｐｓ」）とターゲット用語Ｔ（例えば「ｍａｉｌ」）の間の関係を決定し、式中で「ｉｔｒ」は、「関心がある（ｉｎｔｅｒｅｓｔｅｄ　ｉｎ）」を表している。ユーザ（広告主、および／またはウェブサイトプロモータなど）が、Ｒに関心がある場合、このユーザは、Ｔにも関心があることになる。

【００２１】

ＥＶＳ１０２は、提案される用語リスト１１０を生成するいくつかのコンピュータプログラムモジュールを含んでいる。このコンピュータプログラムモジュールは、例えば、ＳＴＳ（ｓｅａｒｃｈ　ｔｅｒｍ　ｓｕｇｇｅｓｔｉｏｎ検索用語提案）モジュール１１２および分類モジュール１１４を含んでいる。ＳＴＳモジュール１１２は、クエリログ１１８から１組の履歴クエリ１１６を検索する。この履歴クエリは、検索エンジンに対して以前にサブミットされた検索クエリ用語を含んでいる。ＳＴＳモジュール１１２は、出現度数に応じて履歴クエリ１１６を評価して、高出現度数（ＦＯＯ）検索用語１２０と相対的により低い出現度数検索用語１２２を識別する。この実装形態においては、構成可能なしきい値を使用して、履歴クエリが比較的より高い出現度数を有するか、それとも低い出現度数を有するかを判定する。例えば、少なくともしきい値回数だけ現れる履歴クエリ１１６中の検索クエリ用語は、高出現度数を有していると言われる。同様に、しきい値回数よりも少なくしか現れない履歴クエリ１１６中の検索クエリ用語は、低出現度数を有していると言われる。図では、かかるしきい値は、「他のデータ」１２４の各部分として示される。

【００２２】

キーワードマイニングおよびフィーチャ抽出

ＳＴＳモジュール１１２は、各クエリ（検索クエリ１２８）を１つずつ検索エンジン１２６にサブミットすることにより、意味的／コンテキスト的な意味の高出現度数クエリ用語１２０を調べる。検索クエリ１２８を受信するのに応答して、検索エンジン１２６は、検索結果１３０中の（番号を構成可能な）順位付けされたリストをＳＴＳモジュール１１２に返す。この順位付けされたリストは、サブミットされた検索クエリ１２８に関連したＵＲＬ、結果タイトル、簡単な説明および／またはクエリ用語のコンテキストを含んでいる。この順位付けされたリストは、検索結果１３２に記憶される。かかる検索結果の取出しは検索クエリ１２８ごとに行われる。

【００２３】

ＳＴＳモジュール１１２は、ウェブページのＨＴＭＬ（ハイパーテキストマークアップ言語）を解析して各取り出された検索結果１３２からクエリ用語１２０ごとにそのＵＲＬ、結果タイトル、簡単な説明および／またはそのクエリ用語のコンテキストを抽出する。このＵＲＬ、結果タイトル、簡単な説明および／またはそのクエリ用語のコンテキスト、ならびにこの取り出された検索結果１３２を取得するために使用される検索クエリ１２８は、抽出されたフィーチャ１３４の各レコードにＳＴＳモジュール１１２によって記憶される。

【００２４】

高出現度数クエリ用語１２０についての検索結果１３０を解析した後に、ＳＴＳモジュール１１２は、抽出されたフィーチャ１３４上でテキスト前処理オペレーションを実施して、この抽出されたフィーチャから言語トークンを生成（トークン化）して個別のキーワードに入れる。これらのトークンのサイズを縮小するために、ＳＴＳモジュール１１２は、どのようなストップ語句（ｓｔｏｐ－ｗｏｒｄ）（例えば、「ｔｈｅ」、「ａ」、「ｉｓ」など）も除去し、共通の接尾辞を除去して、例えば知られているポーターステミングアルゴリズム（Ｐｏｒｔｅｒ　ｓｔｅｍｍｉｎｇ　ａｌｇｏｒｉｔｈｍ）を使用してこれらのキーワードを正規化する。ＳＴＳモジュール１１２は、この結果得られる抽出されたフィーチャ１３４を１つまたは複数の用語ベクトル１３６に構成する。

【００２５】

各用語ベクトル１３６は、ＴＦＩＤＦ（ｔｅｒｍ　ｆｒｅｑｕｅｎｃｙ　ａｎｄ　ｉｎｖｅｒｔｅｄ　ｄｏｃｕｍｅｎｔ　ｆｒｅｑｕｅｎｃｙ用語頻度および逆ドキュメント頻度）スコアに基づいたサイズを有する。ｉ番目のベクトルのｊ番目のキーワードについての重みは、次式のように計算される。

【００２６】

$w_{ij} = TF_{ij} \times \log(N / DF_j)$

式中で、TF_{ij}は用語頻度（ｉ番目のレコードにおけるキーワードｊの出現回数）を表し、Ｎは、クエリ用語の合計数であり、DF_jは、キーワードｊを含むレコード数である。

【００２７】

用語クラスタリング

ＳＴＳモジュール１１２は、類似の用語をグループ化して用語ベクトル１３６から用語クラスタ１３８を生成する。この目的を達成するために、この実装形態においては、各用語のベクトル表現を仮定して、コサインファンクションを使用して１対の用語の間の類似性を評価する（これらのベクトルが正規化されていることを想起されたい）。

【００２８】

【数１】

$$\mathrm{sim}(q_j, q_k) = \sum_{i=1}^{d} W_{ij} \cdot W_{ik}$$

【００２９】

したがって、２つの用語間の距離（類似性評価値）は、次式のように定義される。

【００３０】

$dist(q_j, q_k) = 1 - sim(q_j, q_k)$

かかる類似性評価値は、「他のデータ」１２４の各部分として示されている。かかる例示の類似性値は、表１の例示の提案される用語リスト１１０に示されている。

【００３１】

ＳＴＳモジュール１１２は、この計算された１つ（または複数）の類似性評価値を使用して、キーワードベクトル１３４によって表される用語をクラスタ／グループ化して１つ（または複数）の用語クラスタ１３８に入れる。より詳細には、この実装形態において、ＳＴＳモジュール１１２は、知られているＤＢＳＣＡＮ（ｄｅｎｓｉｔｙ－ｂａｓｅｄ　ｃｌｕｓｔｅｒｉｎｇ　ａｌｇｏｒｉｔｈｍ密度ベースのクラスタリングアルゴリズム）

を使用して、１つ（または複数）の用語クラスタ１３８を生成する。ＤＢＳＣＡＮは、２つのパラメータ、すなわちＥｐｓおよびＭｉｎＰｔｓを使用する。Ｅｐｓは、クラスタ１３８中のポイント間の最大距離を表す。ここでは、その尾部がその起点に移動されるときに各ベクトルはそのベクトルの頭部のポイントによって表現することができるので、ポイントは、ベクトルと等価である。ＭｉｎＰｔｓは、クラスタ１３８中の最小ポイント数を表す。クラスタ１３８を生成するためには、ＤＢＳＣＡＮは、任意ポイントｐを用いて開始され、ＥｐｓおよびＭｉｎＰｔｓに関してｐからの密度の影響が及ぶすべてのポイントを取り出す。ｐがコアポイントである場合、このプロシージャは、ＥｐｓおよびＭｉｎＰｔｓに関するクラスタ１３８をもたらす。ｐが境界ポイントである場合には、ｐからの密度の影響が及ぶポイントは存在せず、ＤＢＳＣＡＮは、次のポイントを訪れることになる。

【００３２】

用語マッチング

エンドユーザ（例えば、広告主、ウェブサイトプロモータなど）から用語／フレーズ１０８を受信するのに応答して、ＳＴＳモジュール１１２は、用語／フレーズ１０８を用語クラスタ１３８中の各用語／フレーズと比較する。用語クラスタ１３８は、互いにコンテキスト的に関連した用語を含んでいるので、用語／フレーズ１０８は、複数の関連した履歴コンテキスト、または「意味」に鑑みて評価される。一実装形態においては、用語／フレーズ１０８がクラスタ１３８からの用語／フレーズとマッチングするとＳＴＳモジュール１１２が判定した場合、検索用語提案モジュール１１２は、クラスタ１３８から提案される用語リスト１１０を生成する。この実装形態においては、マッチは、厳密なマッチでもよく、また単数形／複数形、スペルの誤り、句読点マークなどわずかな数の変形を伴うマッチでもよい。この返されたリストは、ＦＯＯと確信度値の組合せによって順序づけされる。

【００３３】

用語／フレーズ１０８が複数の用語クラスタ１３８中の用語とマッチしているとＳＴＳモジュール１１２が判定する場合、検索用語提案モジュール１１２は、用語クラスタ１３８の複数のクラスタ中の用語から複数の提案される用語リスト１１０を生成する。このリストは、そのクラスタサイズによって順序付けられる。また、各リスト内のこれらの用語は、ＦＯＯと確信度値の組合せによって順序付けられる。

【００３４】

低ＦＯＯ用語の分類

高出現度数（ＦＯＯ）クエリ用語１２０から生成された用語クラスタ１３８が、エンドユーザの入力用語／フレーズ１０８に対する同じ用語を含んでいないときには、分類モジュール１１４は、提案される用語リスト１１０を生成する。この目的を達成するために、分類モジュール１１４は、高出現度数（ＦＯＯ）クエリログ用語１２０から生成された用語クラスタ１３８からトレーニングされた分類機構１４０を生成する。用語クラスタ１３８中の用語は、すでに分類オペレーションにとって適切なベクトル空間モデル中の対応するキーワードベクトルを有している。さらにストップ語句除去、およびワードステミング（ｗｏｒｄ　ｓｔｅｍｍｉｎｇ）（接尾辞除去）が、（クラスタ１３８が基づいている）用語ベクトル１３６のサイズを縮小している。一実装形態においては、追加のサイズ縮小技法、例えばフィーチャ選択または再パラメータ化を使用することができる。

【００３５】

この実装形態においては、クラスが知られていないクエリ用語１２０を分類するために、分類モジュール１１４は、ｋ－最近接分類機構アルゴリズム（ｋ－Ｎｅａｒｅｓｔ　Ｎｅｉｇｈｂｏｒ　ｃｌａｓｓｉｆｉｅｒ　ａｌｇｏｒｉｔｈｍ）を使用して、それらの対応するフィーチャベクトルを用いてすべてのクラスが知られているクエリ用語１２０中のｋ個の最も類似した近接用語を見出し、この近接用語のクラスラベルの重み付けされた大多数を使用して、この新しいクエリ用語のクラスを予測する。ここで、すでに用語クラスタ１３８中に存在する各クエリ用語にはそれらの対応するクラスタラベルと同じラベルが

割り当てられるが、各クラスタ１３８は、簡単な連続番号でラベル付けされる。これらの近接用語は、Ｘに対する各近接用語の類似性を使用して重み付けされ、ここで、類似性は、ユークリッド距離（Ｅｕｃｌｉｄｅａｎ　ｄｉｓｔａｎｃｅ）、または２つのベクトル間のコサイン値によって評価される。このコサイン類似性は、以下のように表される。

【００３６】

【数２】

$$\mathrm{sim}(X, D_j) = \frac{\sum_{t_i \in (X \cap D_j)} x_i \cdot d_{ij}}{\|X\|_2 \cdot \|D_j\|_2}$$

【００３７】

式中で、Ｘは、テスト用語、すなわちベクトルとして表現される分類すべきクエリ用語である。D_jは、ｊ番目のトレーニング用語である。t_iは、ＸとＤ$_j$によって共有される単語である。x_iは、Ｘ中のキーワードt_iの重みである。d_{ij}は、Ｄｊ中のキーワードt_iの重みである。

【００３８】

【数３】

$$\|X\|_2 = \sqrt{x_1^2 + x_2^2 + x_3^2}$$

【００３９】

は、Ｘのノルム（ｎｏｒｍ）であり、｜｜Ｄ$_j$｜｜$_2$はＤ$_j$のノルムである。したがって、このテスト用語Ｘのクラスラベルは、すべての近接用語のうちの重み付けされた大多数の近接用語のクラスラベルである。

【００４０】

【数４】

$$\mathrm{label}(X) = \arg\max_{l_i} \left(\sum_{\text{すべての}D_j\text{、ただし}\mathrm{label}(D_j)=l_i} \mathrm{sim}(X, D_i) \right)$$

【００４１】

別の実装形態においては、最近接分類技法以外の（例えば、回帰モデル（ｒｅｇｒｅｓｓｉｏｎ　ｍｏｄｅｌ）、ベイジアン分類機構（Ｂｅｙｅｓｉａｎ　ｃｌａｓｓｉｆｉｅｒ）、判断木、ニューラルネットワーク、およびサポートベクトルマシンを含めて）異なる統計的分類および機械学習技法を使用して、トレーニングされた分類機構１４０が生成される。

【００４２】

分類モジュール１１４は、（各検索クエリ１２８を介して）１つずつ、低出現度数（ＦＯＯ）クエリ用語１２２を検索エンジン１２６にサブミットする。特定の検索クエリ１２８に関連する検索結果１３０を受信し、すでに説明した技法を使用するのに応答して、分類モジュール１１４は、この検索結果１３０によって識別される１つまたは複数の取り出された検索結果１３２からフィーチャを抽出する（フィーチャ１３４を抽出している）。この実装形態においては、フィーチャは、第１の上位順位付けされた検索結果１３２から抽出される。取り出され解析された検索結果１３２ごとに、分類モジュール１１４は、抽出済みのフィーチャ１３４の各レコード中の以下の情報、すなわちこのクエリ用語のＵＲＬ、結果タイトル、簡単な説明および／またはコンテキスト、ならびにこの取り出された検索結果１３２を取得するために使用される検索クエリ１２８を記憶する。次に、分類モジュール１１４は、トークン化を行い、サイズを縮小し、低ＦＯＯクエリ用語１２２から導き出された抽出済みのフィーチャを正規化して用語ベクトル１３６を生成する。次いで、分類モジュール１１４は、これらのクエリ用語をクラスタ化して各組のクラスタ１３８に入れる。このクラスタリングオペレーションは、（高ＦＯＯクエリ用語１２０から生成

された）トレーニングされた分類機構１４０を使用して実施される。

【００４３】

分類モジュール１１４は、（低ＦＯＯクエリ用語１２２に基づいて生成される）これらの拡張された用語クラスタに鑑みてエンドユーザがサブミットした用語／フレーズ１０８を評価して、１つまたは複数の提案される用語リスト１１０を識別しこのエンドユーザに対して返す。かかる例示のプロシージャは、用語マッチングの項にて説明しており、以下のセクションにおいても説明する。

【００４４】

例示のプロシージャ

図２は、マルチセンスクエリについての関連語を提案するための例示のプロシージャ２００を示している。説明の目的で、このプロシージャのオペレーションについて、図１のコンポーネントに関連して考察する。（すべての参照番号は、このコンポーネントが最初に導入されている図面の番号から開始される）。ブロック２０２において、検索用語提案（ＳＴＳ）モジュール１１２（図１）は、クエリログ１２０から履歴クエリ用語１１６を収集する。ＳＴＳモジュール１１２は、出現度数に応じて履歴クエリ１１６を構成する。ブロック２０４において、ＳＴＳモジュール１１２は、高出現度数クエリ用語１２０を検索エンジン１２６に送信し、対応する検索結果１３０を受信する。ブロック２０６において、ＳＴＳモジュール１１２は、各検索結果１３０から断片記述を抽出し、これらの断片記述（抽出済みのフィーチャ１３４）を一緒にマージして用語ベクトル１３６を形成する。各用語ベクトル１３６は、高出現度数クエリ用語１２０ごとに生成される。

【００４５】

ブロック２０８において、ＳＴＳモジュール１１２は、クラスタリングアルゴリズムを適用して、用語ベクトル１３６に基づいてほぼ同様な用語を用語クラスタ１３８にグループ分けする。ブロック２１０において、エンドユーザから用語／フレーズ１０８を受信するのに応答して、ＳＴＳモジュール１１２は、用語／フレーズ１０８とほぼ同様であると判定された用語クラスタ１３８からの任意のキーワード／キーフレーズから、提案される用語リスト１１０を生成する。ブロック２１２において、ＳＴＳモジュール１１２は、キーワードクラスタ１３８からのどのようなキーワード／キーフレーズも、用語／フレーズ１０８とほぼ同様であるかどうかを判定する。同様である場合、このプロシージャは、ブロック２１４から継続され、このブロックにおいて、ＳＴＳモジュール１１２は、この提案される用語リスト１１０をこのエンドユーザに送信する。そうでない場合には、このプロシージャは、ページ上のリファレンス「Ａ」で示すように図３のブロック３０２から継続される。

【００４６】

図３は、マルチセンスクエリについての関連語を提案するための例示のプロシージャ３００を示している。プロシージャ３００のオペレーションは、図２のプロシージャ２００のオペレーションに基づいている。説明するために、このプロシージャのオペレーションでは、図１のコンポーネントに関連して説明している。（すべての参照番号は、そのコンポーネントが最初に導入される図面番号を用いて開始される）。ブロック３０２において、ＳＴＳモジュール１１２は、用語クラスタ１３８から分類機構１４０を生成し、それはこの時点では高出現度数クエリ用語１２０に基づいている。ブロック３０４において、ＳＴＳモジュール１１２は、低出現度数クエリ用語１２２を１つずつ検索エンジン１２６に送信し、対応する検索結果１３０を受信する。ブロック３０６において、ＳＴＳモジュール１１２は、検索結果１３０から断片記述を抽出し（フィーチャ１３４を抽出しており）、そこから用語ベクトル１３６を生成する。ブロック３０８において、ＳＴＳモジュール１１２は、トレーニングされた分類機構１４０に鑑みて低出現度数クエリ用語１２２から生成された用語ベクトル１３６を分類して、この低出現度数クエリ用語に基づいた各用語クラスタ１３８を生成する。

【００４７】

ブロック３１０において、ＳＴＳモジュール１１２は、用語／フレーズ１０８とほぼ同

様であると判定されている低出現度数クエリ用語１２２に基づいた用語クラスタ１３８からのキーワード／キーフレーズから、提案される用語リスト１１０を生成する。ブロック３１２において、ＳＴＳモジュール１１２は、この提案される用語リスト１１０をこのエンドユーザに送信する。

【００４８】

例示の動作環境

図４は、マルチセンスクエリについての関連語を提案するための図１のシステム１００と図２および３の方法を完全にまたは部分的に実装することができる適切なコンピューティング環境４００の一実施例を示している。例示のコンピューティング環境４００は、適切なコンピューティング環境の一実施例にすぎず、本明細書中に説明されるシステムおよび方法の用途または機能の範囲についてのどのような限定をも示唆することを意図してはいない。また、コンピューティング環境４００は、コンピューティング環境４００に示すコンポーネントのうちのどの１つまたは組合せにも関連したどのような依存性または必要性を有するものとも解釈すべきではない。

【００４９】

本明細書中で説明している方法およびシステムは、他の多数の汎用または専用のコンピューティングシステム環境またはコンピューティングシステム構成を用いて動作することが可能である。使用するのに適したものとすることができるよく知られているコンピューティングシステム、コンピューティング環境、および／またはコンピューティング構成の実施例には、それだけには限定されないが、パーソナルコンピュータ、サーバコンピュータ、マルチプロセッサシステム、マイクロプロセッサベースのシステム、ネットワークＰＣ、ミニコンピュータ、メインフレームコンピュータ、任意の上記システムまたはデバイスを含む分散コンピューティング環境などが含まれる。このフレームワークのコンパクトバージョンまたはサブセットバージョンはまた、ハンドヘルドコンピュータや他のコンピューティングデバイスなど限られたリソースのクライアント中で実装することもできる。本発明は、分散コンピューティング環境中でも実施され、ここでタスクは、通信ネットワークを介してリンクされるリモート処理デバイスによって実施される。分散コンピューティング環境においては、プログラムモジュールは、ローカルメモリストレージデバイスにも、リモートメモリストレージデバイスにも配置することができる。

【００５０】

図４を参照すると、マルチセンスクエリについての関連語を提案するための例示のシステムが、コンピュータ４１０の形態の汎用コンピューティングデバイスを含んでいる。コンピュータ４１０の以下に説明している態様は、ＰＳＳサーバ１０２（図１）および／またはクライアントコンピューティングデバイス１０６の例示の実装形態である。コンピュータ４１０のコンポーネントは、それだけには限定されないが、１つ（または複数）の処理装置４２０、システムメモリ４３０、およびこのシステムメモリを含めて様々なシステムコンポーネントを処理装置４２０に結合するシステムバス４２１を含むことができる。システムバス４２１は、メモリバスまたはメモリコントローラ、ペリフェラルバス、および様々なバスアーキテクチャのうちのどれかを使用したローカルバスを含めて、いくつかのタイプのバス構造のうちのどれにすることもできる。実施例として、限定するものではないが、かかるアーキテクチャには、ＩＳＡ（Ｉｎｄｕｓｔｒｙ Ｓｔａｎｄａｒｄ Ａｒｃｈｉｔｅｃｔｕｒｅ業界標準アーキテクチャ）バス、ＭＣＡ（Ｍｉｃｒｏ Ｃｈａｎｎｅｌ Ａｒｃｈｉｔｅｃｔｕｒｅマイクロチャネルアーキテクチャ）バス、ＥＩＳＡ（Ｅｎｈａｎｃｅｄ ＩＳＡ拡張ＩＳＡ）バス、ＶＥＳＡ（Ｖｉｄｅｏ Ｅｌｅｃｔｒｏｎｉｃｓ Ｓｔａｎｄａｒｄｓ Ａｓｓｏｃｉａｔｉｏｎビデオエレクトロニクス規格協会）ローカルバス、およびメザニンバスとしても知られているＰＣＩ（Ｐｅｒｉｐｈｅｒａｌ Ｃｏｍｐｏｎｅｎｔ Ｉｎｔｅｒｃｏｎｎｅｃｔ）バスが含まれる。

【００５１】

コンピュータ４１０は、一般的に様々なコンピュータ読取り可能媒体を含んでいる。コンピュータ読取り可能媒体は、コンピュータ４１０からアクセスすることができる使用可

能な任意の媒体とすることができ、揮発性媒体も不揮発性媒体も、着脱可能媒体も着脱不能媒体も共に含んでいる。実施例として限定するものではないが、コンピュータ読取り可能媒体は、コンピュータストレージ媒体および通信媒体を含むことができる。コンピュータストレージ媒体は、コンピュータ読取り可能命令、データ構造、プログラムモジュール、他のデータなどの情報のストレージのための任意の方法または技術で実装される揮発性媒体および不揮発性媒体、着脱可能媒体および着脱不能媒体を含んでいる。コンピュータストレージ媒体には、それだけには限定されないが、ＲＡＭ、ＲＯＭ、ＥＥＰＲＯＭ、フラッシュメモリまたは他のメモリ技術、ＣＤ－ＲＯＭ、ＤＶＤ（デジタル多用途ディスク）または他の光ディスクストレージ、磁気カセット、磁気テープ、磁気ディスクストレージまたは他の磁気ストレージデバイス、あるいは所望の情報を記憶するために使用することができ、コンピュータ４１０からアクセスすることができる他の任意の媒体が含まれる。

【００５２】

通信媒体は、一般的に搬送波や他の搬送メカニズムなどの被変調データ信号の形のコンピュータ読取り可能命令、データ構造、プログラムモジュールまたは他のデータを実施し、任意の情報配信媒体を含んでいる。用語「被変調データ信号」は、信号に情報を符号化するようにその１つまたは複数の特性が設定または変更された信号を意味する。実施例として、限定するものではないが、通信媒体は、有線ネットワークや直接配線接続などの有線媒体と、音響、ＲＦ、赤外線、他の無線媒体などの無線媒体とを含んでいる。以上のうちの任意の組合せもまた、コンピュータ読取り可能媒体の範囲内に含められるべきである。

【００５３】

システムメモリ４３０は、ＲＯＭ（読取り専用メモリ）４３１やＲＡＭ（ランダムアクセスメモリ）４３２など、揮発性および／または不揮発性のメモリの形態のコンピュータストレージ媒体を含んでいる。起動中などコンピュータ４１０内のエレメント間で情報を転送する助けをする基本ルーチンを含むＢＩＯＳ（基本入出力システム）４３３は、一般的にＲＯＭ４３１に記憶される。ＲＡＭ４３２は、処理装置４２０にとって直接にアクセス可能な、または処理装置４２０によって現在動作させられている、あるいはその両方が行われるデータおよび／またはプログラムモジュールを一般的に含んでいる。実施例として、限定するものではないが、図４は、オペレーティングシステム４３４、アプリケーションプログラム４３５、他のプログラムモジュール４３６、およびプログラムデータ４３７を示している。一実装形態においては、コンピュータ４１０は、ＰＳＳサーバ１０２である。このシナリオにおいては、アプリケーションプログラム４３５は、検索用語提案モジュール１１２、および分類モジュール１１４を含んでいる。この同じシナリオにおいて、プログラムデータ４３７は、用語／フレーズ１０８、提案される用語リスト１１０、履歴クエリ１１６、検索クエリ１２８、検索結果１３０、複数の検索結果１３２、抽出済みのフィーチャ１３４、用語ベクトル１３６、キーワードクラスタ１３８、トレーニングされる分類機構１４０、および他のデータ１２４を含む。

【００５４】

コンピュータ４１０は、他の着脱可能／着脱不能な、揮発性／不揮発性のコンピュータストレージ媒体を含むこともできる。実施例にすぎないが、図４は、着脱不能な不揮発性磁気媒体から情報を読み取りまたはそれに情報を書き込むハードディスクドライブ４４１、着脱可能な不揮発性磁気ディスク４５２から情報を読み取りまたはそれに情報を書き込む磁気ディスクドライブ４５１、およびＣＤ－ＲＯＭや他の光媒体など着脱可能な不揮発性光ディスク４５６から情報を読み取りまたはそれに情報を書き込む光ディスクドライブ４５５を示している。例示の動作環境において使用することができる他の着脱可能／着脱不能な揮発性／不揮発性のコンピュータストレージ媒体には、それだけには限定されないが、磁気テープカセット、フラッシュメモリカード、デジタル多用途ディスク、デジタルビデオテープ、ソリッドステートＲＡＭ、ソリッドステートＲＯＭなどが含まれる。ハードディスクドライブ４４１は、一般的にインターフェース４４０など着脱不能メモリイン

ターフェースを介してシステムバス４２１に接続され、磁気ディスクドライブ４５１および光ディスクドライブ４５５は、一般的にインターフェース４５０など着脱可能なメモリインターフェースによってシステムバス４２１に接続される。

【００５５】

前述され図４に示されるこれらのドライブおよびその関連するコンピュータストレージ媒体は、コンピュータ４１０についてのコンピュータ読取り可能命令、データ構造、プログラムモジュールおよび他のデータの記憶域を提供する。図４において、例えばハードディスクドライブ４４１は、オペレーティングシステム４４４、アプリケーションプログラム４４５、他のプログラムモジュール４４６、およびプログラムデータ４４７を記憶するものとして示されている。これらのコンポーネントは、オペレーティングシステム４３４、アプリケーションプログラム４３５、他のプログラムモジュール４３６、およびプログラムデータ４３７と同じとすることもでき、また異なるものとすることもできることに留意されたい。オペレーティングシステム４４４、アプリケーションプログラム４４５、他のプログラムモジュール４４６、およびプログラムデータ４４７には、これらが少なくとも異なるコピーであることを示すために、ここでは異なる番号が付与されている。

【００５６】

ユーザは、キーボード４６２や、一般にマウス、トラックボールまたはタッチパッドと呼ばれるポインティングデバイス４６１などの入力デバイスを介してコンピュータ４１０にコマンドおよび情報を入力することができる。他の入力デバイス（図示せず）は、マイクロフォン、ジョイスティック、ゲームパッド、衛星パラボラアンテナ、スキャナなどを含むことができる。これらおよび他の入力デバイスは、このシステムバス４２１に結合されるユーザ入力インターフェース４６０を介して処理装置４２０にしばしば接続されるが、これは、パラレルポート、ゲームポート、ＵＳＢ（ユニバーサルシリアルバス）など他のインターフェースおよびバス構造によって接続することもできる。

【００５７】

モニタ４９１または他のタイプのディスプレイデバイスもまた、ビデオインターフェース４９０などのインターフェースを介してシステムバス４２１に接続される。このモニタに追加して、コンピュータは、スピーカ４９７やプリンタ４９６などの他のペリフェラル出力デバイスを含むこともでき、これらは、出力ペリフェラルインターフェース４９５を介して接続することができる。

【００５８】

コンピュータ４１０は、リモートコンピュータ４８０など１つまたは複数のリモートコンピュータに対する論理接続を使用してネットワーク環境中で動作する。リモートコンピュータ４８０は、パーソナルコンピュータ、サーバ、ルータ、ネットワークＰＣ、ピアデバイス、または他の共通ネットワークノードとすることができ、またその特定の実装形態に応じて、コンピュータ４１０に関連した前述のエレメントの多くまたはすべてを含むこともできるが、メモリストレージデバイス４８１だけしか図４には示されていない。図４に示す論理接続は、ＬＡＮ（ローカルエリアネットワーク）４７１およびＷＡＮ（ワイドエリアネットワーク）４７３を含んでいるが、他のネットワークを含むこともできる。かかるネットワーキング環境は、オフィス、企業規模のコンピュータネットワーク、イントラネットおよびインターネットにおいては、一般的なものである。

【００５９】

ＬＡＮネットワーキング環境中で使用する際には、コンピュータ４１０は、ネットワークインターフェースまたはアダプタ４７０を介してＬＡＮ４７１に接続される。ＷＡＮネットワーキング環境中で使用する際には、コンピュータ４１０は、一般的にインターネットなどのＷＡＮ４７３上で通信を確立するためのモデム４７２または他の手段を含んでいる。モデム４７２は、内蔵でも外付けでもよいが、ユーザ入力インターフェース４６０または他の適切なメカニズムを介してシステムバス４２１に接続することができる。ネットワーク環境においては、コンピュータ４１０に関連して示されるプログラムモジュール、またはその一部分は、リモートメモリストレージデバイスに記憶することができる。実施

例として、限定するものではないが、図４は、リモートアプリケーションプログラム４８５をメモリデバイス４８１上に存在するものとして示している。図に示すこのネットワーク接続は、例示的なものであり、コンピュータ間で通信リンクを確立する他の手段を使用することもできる。

【００６０】

結論

マルチセンスクエリについての関連語を提案するためのシステムおよび方法について、構造的な特徴および／または方法的なオペレーションもしくは動作に特有の言語で説明してきたが、添付の特許請求の範囲中で定義される実装形態は、必ずしも説明したこの特定の特徴または動作だけに限定されるとは限らないことが理解されよう。したがって、この特定の特徴および動作は、この請求される内容を実装する例示の形態として開示されている。

【図面の簡単な説明】

【００６１】

【図１】マルチセンスクエリについての関連語を提案するためのシステムの一例を示す図である。

【図２】マルチセンスクエリについての関連語を提案するためのプロシージャの一例を示す図である。

【図３】マルチセンスクエリについての関連語を提案するための、図２のオペレーションに基づいたオペレーションであるプロシージャの一例を示す図である。

【図４】マルチセンスクエリについての関連語を提案するための、本発明によるシステム、装置および方法を完全にまたは部分的に実装することができる適切なコンピューティング環境の一例を示す図である。

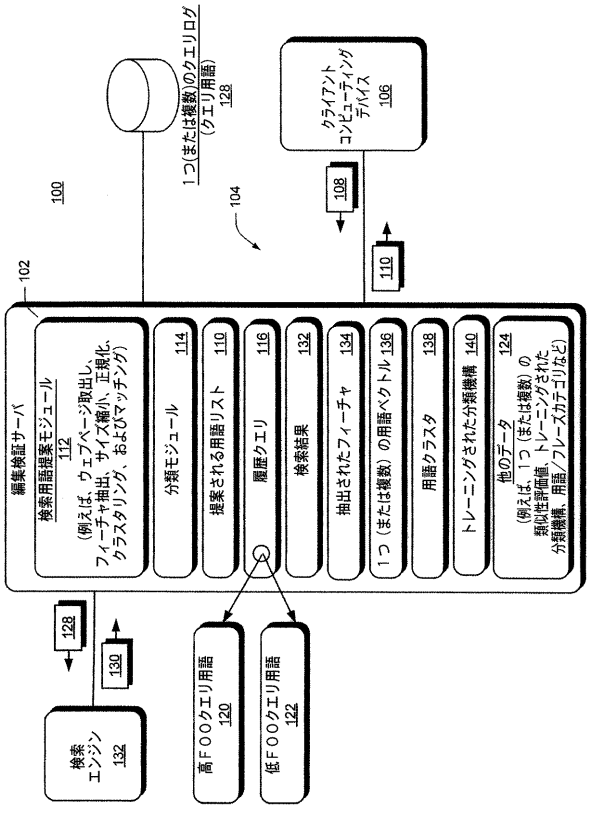
【符号の説明】

【００６２】

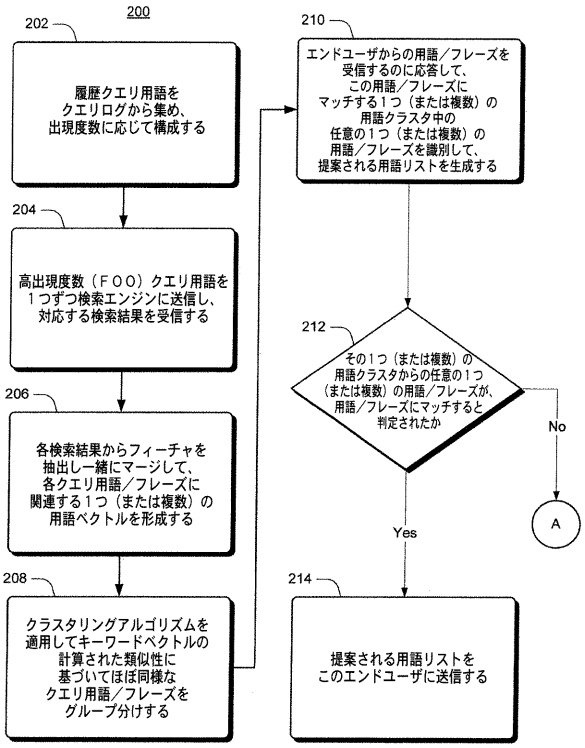
１０２　編集検証サーバ
１０６　クライアントコンピューティングデバイス
１１０　提案される用語リスト
１１２　検索用語提案モジュール
１１４　分類モジュール
１１６　履歴クエリ
１２０　高ＦＯＯクエリ用語
１２２　低ＦＯＯクエリ用語
１２４　他のデータ
１３２　検索エンジン
１２８　１つ（または複数）のクエリログ
１３２　検索結果
１３４　抽出されたフィーチャ
１３６　１つ（または複数）の用語ベクトル
１３８　用語クラスタ
１４０　トレーニングされた分類機構
４２０　１つ（または複数）の処理装置
４２１　システムバス
４３０　システムメモリ
４３４　オペレーティングシステム
４３５　アプリケーションプログラム
４３６　他のプログラムモジュール
４３７　プログラムデータ
４４０　着脱不能な不揮発性メモリインターフェース
４４４　オペレーティングシステム

４４５　アプリケーションプログラム
４４６　他のプログラムモジュール
４４７　プログラムデータ
４５０　着脱可能な不揮発性メモリインターフェース
４６０　ユーザ入力インターフェース
４６１　マウス
４６２　キーボード
４７０　ネットワークインターフェース
４７１　ローカルエリアネットワーク
４７２　モデム
４７３　ワイドエリアネットワーク
４８０　リモートコンピュータ
４８５　リモートアプリケーションプログラム
４９０　ビデオインターフェース
４９１　モニタ
４９４　入力ペリフェラルインターフェース
４９５　出力ペリフェラルインターフェース
４９６　プリンタ
４９７　スピーカ

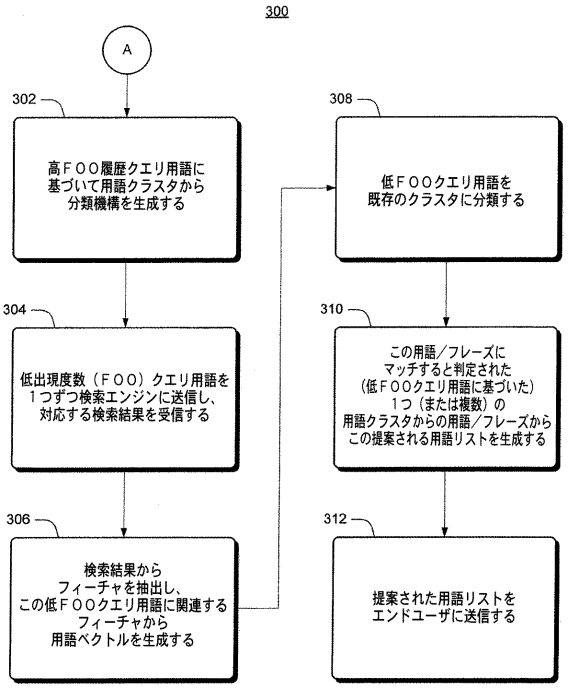
【図１】



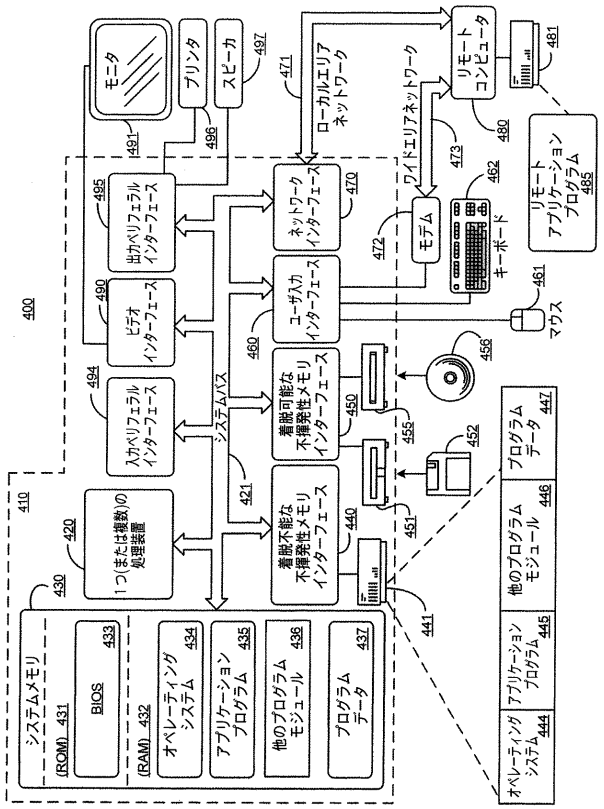
【図２】



【図３】


300
A
302
高ＦＯＯ履歴クエリ用語に基づいて用語クラスタから分類機構を生成する
304
低出現度数（ＦＯＯ）クエリ用語を１つずつ検索エンジンに送信し、対応する検索結果を受信する
306
検索結果からフィーチャを抽出し、この低ＦＯＯクエリ用語に関連するフィーチャから用語ベクトルを生成する
308
低ＦＯＯクエリ用語を既存のクラスタに分類する
310
この用語／フレーズにマッチすると判定された（低ＦＯＯクエリ用語に基づいた）１つ（または複数）の用語クラスタからの用語／フレーズからこの提案される用語リストを生成する
312
提案された用語リストをエンドユーザに送信する

【図４】


400
410
システムメモリ 430
ROM 431
BIOS 433
RAM 432
オペレーティングシステム 434
アプリケーションプログラム 435
他のプログラムモジュール 436
プログラムデータ 437
1つまたは複数の処理装置 420
システムバス 421
ビデオインターフェース 490
出力ペリフェラルインターフェース 495
入力ペリフェラルインターフェース 494
着脱可能な不揮発性メモリインターフェース 450
着脱不能な不揮発性メモリインターフェース 440
ユーザ入力インターフェース 460
ネットワークインターフェース 470
モニタ 491
プリンタ 496
スピーカ 497
ローカルエリアネットワーク 471
ワイドエリアネットワーク 473
モデム 472
キーボード 462
マウス 461
リモートコンピュータ 480
リモートアプリケーションプログラム 485
オペレーティングシステム 444
アプリケーションプログラム 445
他のプログラムモジュール 446
プログラムデータ 447

フロントページの続き

(72)発明者 ホアージュン チェン
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
(72)発明者 リー リー
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
(72)発明者 タレック ナジム
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
(72)発明者 ウェイーイェン マ
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
(72)発明者 イェン リー
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
(72)発明者 チェン ツェン
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

審査官 鈴木 和樹

(56)参考文献 特開２００３－２３３６８４（ＪＰ，Ａ）
特表２００２－５１８７４８（ＪＰ，Ａ）
Vijay V. Raghavan、外１名, ON THE REUSE OF PAST OPTIMAL QUERIES, Proceedings of the 18th annual international ACM SIGIR conference on Research and development in information retrieval(SIGIR'95)[online], １９９５年, ｐ．３４４－３５０, [DL from ACM Digital Library]
Myoung-Cheol Kim、外１名, A comparison of collocation-based similarity measures in query expansion, Information Processing and Management[online], １９９９年, 第３５巻, 第１号, ｐ．１９－３０, [DL from Science Direct]
Yonggang Qiu、外１名, Concept Based Query Expansion, Proceedings of the 16th annual international ACM SIGIR conference on Research and development in information retrieval(SIGIR'93)[online], １９９３年, ｐ．１６０－１６９, [DL from ACM Digital Library]

(58)調査した分野(Int.Cl., ＤＢ名)
Ｇ０６Ｆ １７／３０
ＴＨＥ ＡＣＭ ＤＩＧＩＴＡＬ ＬＩＢＲＡＲＹ
Ｓｃｉｅｎｃｅ Ｄｉｒｅｃｔ