(19)日本国特許庁(JP)　(12)特　許　公　報(B2)　(11)特許番号
特許第4838529号
(P4838529)
(45)発行日　平成23年12月14日(2011.12.14)　(24)登録日　平成23年10月7日(2011.10.7)

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

請求項の数 22　外国語出願　(全 30 頁)

(21)出願番号　特願2005-118052(P2005-118052)
(22)出願日　平成17年4月15日(2005.4.15)
(65)公開番号　特開2005-302043(P2005-302043A)
(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15日(2008.4.15)
(31)優先権主張番号　10/826,159
(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】検索語提案のためのマルチ型データオブジェクトの強化されたクラスタリング

(57)【特許請求の範囲】
【請求項１】
コンピューティングデバイスにより実行される方法であって、
第１の型の少なくとも１つのオブジェクト、および前記第１の型とは異なる第２の型の少なくとも１つのオブジェクトを含むマルチ型データオブジェクトの間の関係を処理ユニットが決定するステップであって、前記関係は、前記マルチ型データオブジェクトの間のレイヤ内関係か、レイヤ間関係かのうち少なくとも１である、ステップと、
前記関係に基づいて前記マルチ型データオブジェクトを繰り返しクラスタリングして強化されたクラスタを前記処理ユニットが生成するステップと、
ユーザから送信される語に関係のある、提案される検索語のリストを、前記強化されたクラスタを利用して前記処理ユニットが生成するステップであって、前記検索語は、前記ユーザから送信される語の受信に応答して生成される、ステップと、
ネットワークインターフェースを介して、前記提案される検索語のリストをユーザへ前記処理ユニットが送信するステップと、
次の

【数１】

$$\begin{cases} a(X) = \beta L_X{}^T h(X) + (1-\beta) L_{XY} i(Y) \\ h(X) = \beta L_X a(X) + (1-\beta) L_{XY} i(Y) \\ i(X) = a(X) + h(X) \\ \\ a(Y) = \gamma L_Y{}^T h(Y) + (1-\gamma) L_{YX} i(X) \\ h(Y) = \gamma L_Y a(Y) + (1-\gamma) L_{YX} i(X) \\ i(Y) = a(Y) + h(Y) \end{cases}$$

を使用してノードの権威スコアおよびハブスコアを更新することによって、オブジェクト型内および異なるオブジェクト型の間での前記マルチ型データオブジェクトの個々のオブジェクト重要度の相互強化を前記処理ユニットが行うステップであって、Ｘ＝｛x_1，x_2，・・・ x_m｝およびＹ＝｛y_1，y_2，・・・ y_n｝はR_X、R_Y、R_{XY}、およびR_{YX}の関係を有する異質のオブジェクト型のそれぞれのオブジェクトの集合を表現し、方向性が考慮される場合、L_XおよびL_Yはそれぞれ集合ＸおよびＹ内の関係を識別するリンクの隣接する行列を表現し、L_{XY}およびL_{YX}はＸの中のオブジェクトからＹの中のオブジェクトへの関係を識別するリンクの隣接する行列を表現し、ａ（Ｘ）およびｈ（Ｘ）はそれぞれＸ内のノードの権威スコアおよびハブスコアであり、ａ（Ｙ）およびｈ（Ｙ）は、Ｙ内のノードの権威スコアおよびハブスコアを表し、ｉ（Ｘ）およびｉ（Ｙ）はそれぞれＸおよびＹの中のノードの重要度を表し、βおよびγは異なる関係に由来するリンクの影響を調整するための重みパラメータである、ステップと

を備えたことを特徴とする方法。

【請求項２】

前記レイヤ間関係は、コンテントに関係のある情報、関連トピックに対するユーザの関心、および関連Ｗｅｂページに対するユーザの関心のうち少なくとも１つを含むことを特徴とする請求項１に記載の方法。

【請求項３】

前記レイヤ内関係は、クエリ改良、推奨されるＷｅｂページ、およびそれぞれのユーザの間の関係のうち少なくとも１つを含むことを特徴とする請求項１に記載の方法。

【請求項４】

前記マルチ型データオブジェクトの各々は、選択されたＷｅｂページ型およびユーザ情報型のうち少なくとも１つに関係のあることを特徴とする請求項１に記載の方法。

【請求項５】

前記レイヤ内関係は、前記マルチ型データオブジェクトの関連付けるオブジェクトに対する重要度を示すための第１の重み付けの仕組み、および前記第１の重み付けの仕組みとは異なる第２の重み付けの仕組みを含むことを特徴とする請求項１に記載の方法。

【請求項６】

前記識別するステップおよび繰り返しクラスタリングするステップは、検索語提案のために実行されることを特徴とする請求項１に記載の方法。

【請求項７】

前記繰り返しクラスタリングするステップは、第１の反復のクラスタリング結果を前記マルチ型データオブジェクトのすべての関係のあるデータオブジェクトに前記処理ユニットが伝播させるステップを含み、前記関係のあるデータオブジェクトの少なくとも２つは異質のデータの型であり、前記クラスタリング結果により形成される特徴空間は、強化されたクラスタリング操作の第２の反復において前記マルチ型データオブジェクトのそれぞれのオブジェクトのクラスタリングを高めるために使用されることを特徴とする請求項１に記載の方法。

【請求項８】

前記繰り返しクラスタリングするステップは、前記マルチ型データオブジェクトの個々

のオブジェクトの間の類似度を前記処理ユニットが決定するステップを含み、前記類似度はオブジェクト間およびオブジェクト内のコンテンツの類似度、ならびに前記識別された関係の少なくとも１つの間の類似度の少なくとも１つの関数であることを特徴とする請求項１に記載の方法。

【請求項９】

前記繰り返しクラスタリングするステップは、前記マルチ型データオブジェクトの関係のあるオブジェクトをマージして、前記関係のあるオブジェクトの特徴空間の次元を前記処理ユニットが縮小するステップを含むことを特徴とする請求項１に記載の方法。

【請求項１０】

オブジェクト型内および異なるオブジェクト型の間の前記マルチ型データオブジェクトの個々のオブジェクト重要度の相互強化を前記処理ユニットが行うステップ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】

前記送信される語と前記強化されたクラスタの中のオブジェクトの特徴空間とを比較して、前記提案される検索語を前記処理ユニットが識別するステップ

をさらに備えたことを特徴とする請求項１に記載の方法。

【請求項１２】

コンピュータに、

第１の型の少なくとも１つのオブジェクト、および前記第１の型とは異なる第２の型の少なくとも１つのオブジェクトを含む、マルチ型データオブジェクトの間のレイヤ内およびレイヤ間関係の少なくとも１つを決定する手順と、

前記少なくとも１つの関係によって前記マルチ型データオブジェクトを繰り返しクラスタリングすることにより強化されたクラスタ群を生成する手順と、

ユーザから送信される語に関係のある、提案される検索語のリストを、前記強化されたクラスタ群を利用して生成する手順であって、前記提案される検索語は、前記ユーザから送信される語の受信に応答して生成される、手順と、

次の

【数２】

$$\begin{cases} a(X) = \beta L_X{}^T h(X) + (1-\beta) L_{XY} i(Y) \\ h(X) = \beta L_X a(X) + (1-\beta) L_{XY} i(Y) \\ i(X) = a(X) + h(X) \end{cases}$$

$$\begin{cases} a(Y) = \gamma L_Y{}^T h(Y) + (1-\gamma) L_{YX} i(X) \\ h(Y) = \gamma L_Y a(Y) + (1-\gamma) L_{YX} i(X) \\ i(Y) = a(Y) + h(Y) \end{cases}$$

を使用してノードの権威スコアおよびハブスコアを更新することによって、オブジェクト型内および異なるオブジェクト型の間での前記マルチ型データオブジェクトの個々のオブジェクト重要度の相互強化を行う手順であって、$X = \{x_1, x_2, \cdots x_m\}$ および $Y = \{y_1, y_2, \cdots y_n\}$ は R_X、R_Y、R_{XY}、および R_{YX} の関係を有する異質のオブジェクト型のそれぞれのオブジェクトの集合を表現し、方向性が考慮される場合、L_X および L_Y はそれぞれ集合Ｘおよびｙ内の関係を識別するリンクの隣接する行列を表現し、L_{XY} および L_{YX} はＸの中のオブジェクトからＹの中のオブジェクトへの関係を識別するリンクの隣接する行列を表現し、ａ（Ｘ）およびｈ（Ｘ）はそれぞれＸ内のノードの権威スコアおよびハブスコアであり、ａ（Ｙ）およびｈ（Ｙ）はそれぞれＹ内のノードの権威スコアおよびハブスコアを表し、ｉ（Ｘ）およびｉ（Ｙ）はそれぞれＸおよびＹの中のノードの重要度を表し、βおよびγは異なる関係に由来するリンクの影響を調整するための重みパラメータである、手順と

を実行させるためのプログラムを記録したコンピュータ読取可能な記録媒体。

【請求項１３】
前記レイヤ間関係は、コンテントに関係のある情報、関連トピックに対するユーザの関心、および関連Ｗｅｂページに対するユーザの関心のうち１つを含むことを特徴とする請求項１２に記載のコンピュータ読取可能な記録媒体。
【請求項１４】
前記レイヤ内関係は、クエリ改良、推奨されるＷｅｂページ、およびそれぞれのユーザの間の関係のうち少なくとも１つを含むことを特徴とする請求項１２に記載のコンピュータ読取可能な記録媒体。
【請求項１５】
前記マルチ型データオブジェクトのそれぞれは、検索クエリデータオブジェクト型、選択されたＷｅｂページ型、およびユーザ情報型のうち少なくとも１つに関係のあることを特徴とする請求項１２に記載のコンピュータ読取可能な記録媒体。
【請求項１６】
前記決定された関係のうち少なくとも１つに重みを付けることによって、前記マルチ型データオブジェクトの関連するオブジェクトに対する重要度を示すことを特徴とする請求項１２に記載のコンピュータ読取可能な記録媒体。
【請求項１７】
前記識別する手順および繰り返しクラスタリングする手順は、検索語提案のために実行されることを特徴とする請求項１２に記載のコンピュータ読取可能な記録媒体。
【請求項１８】
前記繰り返しクラスタリングする手順は、第１の反復のクラスタリング結果を前記マルチ型データオブジェクトのすべての関係のあるデータオブジェクトに伝播させる手順を含み、前記関係のあるデータオブジェクトの少なくとも２つは異質のデータの型であり、前記クラスタリング結果により形成される特徴空間は、強化されたクラスタリング操作の第２の反復において前記マルチ型データオブジェクトのそれぞれのオブジェクトのクラスタリングを高めるために使用されることを特徴とする請求項１２に記載のコンピュータ読取可能な記録媒体。
【請求項１９】
前記繰り返しクラスタリングする手順は、前記マルチ型データオブジェクトの個々のオブジェクトの間の類似度を決定する手順を含み、前記類似度は、オブジェクトコンテンツの類似度および前記識別された関係の少なくとも１つの間の類似度の少なくとも１つの関数であることを特徴とする請求項１２に記載のコンピュータ読取可能な記録媒体。
【請求項２０】
前記繰り返しクラスタリングする手順は、前記マルチ型データオブジェクトの関係のあるオブジェクトをマージすることによって前記関係のあるオブジェクトの特徴空間の次元を縮小する手順を含むことを特徴とする請求項１２に記載のコンピュータ読取可能な記録媒体。
【請求項２１】
前記コンピュータに、オブジェクト型内および異なるオブジェクト型の間で前記マルチ型データオブジェクトの個々のオブジェクトの重要度の相互強化を行う手順をさらに実行させることを特徴とする請求項１２に記載のコンピュータ読取可能な記録媒体。
【請求項２２】
前記コンピュータに、前記送信される語と前記強化されたクラスタの中のオブジェクトの特徴空間とを比較して、前記提案される検索語を識別する手順を実行させることを特徴とする請求項１２に記載のコンピュータ読取可能な媒体。
【発明の詳細な説明】
【技術分野】
【０００１】
本発明は、データマイニングに関し、より詳細には、異質のオブジェクトをクラスタリングすることにより、検索語提案のためのシステムおよび方法を高めることに関する。

【背景技術】
【０００２】
キーワードまたは句は、WWW（Ｗｏｒｌｄ　Ｗｉｄｅ　Ｗｅｂ）上で関連するＷｅｂページ／サイトを検索する場合、Ｗｅｂサーファ（ｓｕｒｆｅｒ）によって検索エンジンに送信される単語または語の集合である。検索エンジンは、ページ／サイト上に現れるキーワードおよびキーワード句に基づき、Ｗｅｂサイトの関連性を決定する。かなりの割合のＷｅｂサイトトラフィックは、検索エンジンの使用に起因するので、Ｗｅｂサイト主催者は、適切なキーワード／句選択が、所望のサイトの登場を獲得するためにサイトトラフィックの増加に不可欠であることを理解している。検索エンジン結果を最適化するためにＷｅｂサイトと関係のあるキーワードを識別する技術には、例えば、関係のあるキーワードを識別するためにＷｅｂサイトのコンテントおよび目的の人手による評価を含む。この評価には、キーワード人気度ツール（ｋｅｙｗｏｒｄ　ｐｏｐｕｌａｒｉｔｙ　ｔｏｏｌ）の使用を含むことができる。そのようなツールは、どれだけの数の人々が、特定のキーワードまたは特定のキーワードを含む句を検索エンジンに送信したかを決定する。Ｗｅｂサイトと関係があり、検索クエリを生成する際により頻繁に使用されると決定されたキーワードを、一般に、Ｗｅｂサイトに関して、検索エンジン結果を最適化するために選択する。
【０００３】
Ｗｅｂサイトの検索エンジン結果を最適化するためのキーワードの集合を識別した後、主催者は、（他のＷｅｂサイトの検索エンジン結果の表示位置と比較して）検索エンジン結果においてより上位にＷｅｂサイトを昇進させることを所望する可能性がある。この目的のために、主催者は、（複数の）キーワードを入札することにより、（複数の）キーワードに結びつく主催者のリストをＷｅｂサーファがクリックするたび毎に主催者がいくら支払うかを表す。すなわち、キーワード入札は、ペイパークリック（ｐａｙ－ｐｅｒ－ｃｌｉｃｋ）入札である。同一のキーワードに対する他の入札と比較してキーワード入札の総額がより大きいほど、検索エンジンにより、キーワードに基づく検索結果において結びつくＷｅｂサイトがより上位に（重要度に関してより目立つように）表示される。
【０００４】
Ｗｅｂサイトのコンテントと関係があり、入札する（複数の）語を識別するための従来システムおよび技術は、通常、クラスタリングアルゴリズムを使用することにより、同じクラスタからのオブジェクトは類似し、異なるクラスタからのオブジェクトは類似しないような仕方で、オブジェクトの集合をグループまたはクラスタに分割する。クラスタリングアプローチでは、クラスタリングされるデータオブジェクトが、独立であり、同一のクラスのオブジェクトであり、しばしば、固定長ベクトルの特徴／属性値でモデル化されていると仮定する。最近急増しているデータマイニング研究において、この古典的な問題が、大型データベースの文脈において再検討されている。しかし、クラスタリングされるデータオブジェクトの均質性は、たとえ、Ｗｅｂマイニングおよび協調フィルタリングなどいくつかのアプリケーションの出現により、そのような仮定に対する異議を唱えていても、依然、基本的な仮定であるように思われる。そのようなアプリケーションでは、データオブジェクトは、異なる型のデータオブジェクトであり、非常に相互に関係がある。残念ながら、たとえ異質のオブジェクト型にわたって分けられたオブジェクトが非常に相互に関係がある可能性があっても、従来のクラスタリング操作は、通常、それぞれのオブジェクト型を個々に、異なるオブジェクト型の相互に関係がある様相を考慮することなしにクラスタリングする。
【０００５】
このことの１つの理由は、異なる型のデータオブジェクト間の関係が、希薄であり識別するのが困難である場合が多いからである。別の理由は、それぞれのオブジェクトに結びついた静的な固定長の値のベクトルを有する、あらゆるそのような関係の表現は、ここで、ベクトルはオブジェクト属性および異なる型の関係のあるオブジェクトの属性の両方を表すが、非常に高次元（特徴空間）のオブジェクト属性／特徴ベクトルを作り出すからで

ある。このような高次元は望ましくない。なぜならば、特徴空間内でデータが互いに遠く離れ、効率的なモデルを小さい領域内の希薄な量のデータで十分に扱うことができないからである。

【発明の開示】

【発明が解決しようとする課題】

【０００６】

したがって、異質のデータオブジェクトにわたる関係の観点から、関係のあるオブジェクト（例えば、語）を識別しグループ化する、より良いクラスタリング技術は有用である。クラスタリング技術を使用することにより、例えば、検索エンジン最適化および語の入札のための（複数の）語を識別するシステムおよび方法を提供し、それによってシステムと方法の両方において、関係のある（複数の）語を識別する確率を大いに高めることができる。

【課題を解決するための手段】

【０００７】

関係のある語提案のためのシステムおよび方法を説明する。一態様では、２つ以上のマルチ型データオブジェクトのうちのそれぞれのオブジェクトの間にレイヤ内および／またはレイヤ間関係を識別する。マルチ型データオブジェクトのそれぞれのオブジェクトには、第１型の少なくとも１つのオブジェクト、および第１型とは異なる第２型の少なくとも１つのオブジェクトを含む。マルチ型データオブジェクトを、関係のうちのそれぞれの関係の観点から繰り返しクラスタリングすることによって、強化されたクラスタを生成する。

【０００８】

図では、構成要素参照符号の左端の数字が、その構成要素が最初に現れる特定の図を識別する。

【発明を実施するための最良の形態】

【０００９】

（概要）

図１は、相互に関係のある異質のオブジェクトデータ型の例示的フレームワーク１００を示す。フレームワーク１００は、異質のデータオブジェクト／ノードの複数のレイヤ１０２、ならびに結びつけられたレイヤ間およびレイヤ内データオブジェクトリンク／関係を含む。各レイヤ１０２－１ないし１０２－Ｎは、同一型（均質の）データオブジェクトまたはノードのそれぞれの集合を含む。すなわち、ノード集合Ｐは、同一データ型のおのおのである１つまたは複数のデータオブジェクトp_1ないしp_jを含み、ノード集合Ｕは、同一データ型のおのおのである１つまたは複数のデータオブジェクトu_1ないしu_kを含み、以下同様である。このため、異なるそれぞれのレイヤ１０２にあるデータオブジェクトの型は、お互いに関して異質である。

【００１０】

この実装では、例えば、

・レイヤ１０２－１は、（複数の）検索クエリデータオブジェクト／（複数の）ノードp_1ないしp_jを含むマイニングされた（ｍｉｎｅｄ）検索クエリである。検索クエリオブジェクトは、（複数の）クエリの語を含み、以下に説明するとおり、クエリログからマイニングされた複数の履歴クエリのそれぞれのクエリを表す。

・レイヤ１０２－２は、（複数の）Ｗｅｂページデータオブジェクト／（複数の）ノードu_1ないしu_kを含むマイニングされたＷｅｂページレイヤである。

・レイヤ１０２－３は、マイニングされたユーザレイヤであり、（複数の）ユーザ情報オブジェクト／複数のノードw_1ないしw_mを含む。

・レイヤ１０２－Ｎは、それぞれの異なるオブジェクト型x_1ないしx_oを備えるレイヤ１０２はあらゆる個数が可能であることの例示を示す。

【００１１】

一組のデータオブジェクトの間に張るライン／リンクは、それぞれのデータオブジェク

トの間に存在すると決定されたそれぞれのマイニングされた関係を表す。クラスタリングの所定の実施形態では、ライン／リンクを「エッジ」と呼ぶ。本明細書では、一般化された語のラインまたはリンクを使用することにより、リンク、エッジ、またはオブジェクト間の関係を説明するあるオブジェクトと別のオブジェクトとのあらゆる連結子を説明する。リンク方向を（それは、データオブジェクト間の関係を示す矢印により提供されるが）、関与する（ｐａｒｔｉｃｉｐａｔｉｎｇ）オブジェクト属性の相関的要素として、いずれかの方向に向けることができる。リンクは、例示的であり、範囲を限定するものではないと見なされる。フレームワーク１００で表されるようなＷｅｂ環境における所定のリンクを、１つの方向により適切に向けることができ、矢印の方向は、通常、以下に説明する強化されたクラスタリング操作に影響を与えない。

【００１２】

オブジェクトの組の間のリンクを、レイヤ内またはレイヤ間リンクとして分類することができる。レイヤ内リンクは、同一型の異なるオブジェクトの間の識別された関係を例示する。このように、レイヤ内リンク１０４は、同一レイヤ１０２内のオブジェクトを接続する。例えば、データオブジェクトのそれぞれの組の間における実線１０４は、レイヤ内リンクを表す。この例では、レイヤ内リンクは、Ｗｅｂページオブジェクトu_2から別のＷｅｂページオブジェクトu_3に張り、異なるＷｅｂページ間の（複数の）関係を表す。

【００１３】

レイヤ間リンクは、異なる型のデータオブジェクト間の関係を説明する。レイヤ間リンクは、異質のオブジェクトから成る組のそれぞれのオブジェクト間に張られるため、データオブジェクトから成る関与する組のおのおのを、異なるそれぞれのデータオブジェクト／ノード集合レイヤ１０２上に示す。図１に示すとおり、実線ではないオブジェクトからなる組を接続するあらゆるラインは、レイヤ間リンクである。例えば、リンク１０６は、オブジェクトから成る組の第１のオブジェクトから、オブジェクトから成る組の第２のオブジェクトへの参照（例えば、ハイパーリンク）を示し、リンク／ライン１０８は、オブジェクトから成る組の第１のオブジェクトから、オブジェクトから成る組の第２のオブジェクトに共有される／参照される問題（例えば、主題要素）を示し、リンク／ライン１１０は、オブジェクトから成る組の第１のオブジェクトから、オブジェクトから成る組の第２のオブジェクトへのブラウズリンクを示す。別の例では、リンクを、ユーザオブジェクトw_4から検索クエリオブジェクトp_5およびＷｅｂページオブジェクトu_5へ張り、ユーザにより関係があるとして選択されたＷｅｂページを戻すクエリをユーザが送信することを表すことが可能である。

【００１４】

図１の実施例では、レイヤ内およびレイヤ間リンクのそれぞれのリンクで示されるとおり、異なるオブジェクト型（ｐ、ｕ、ｗ、．．．）が関係している。例えば、ユーザ（オブジェクトｗで表される）が、クエリ（オブジェクトｐ）を発行する。ユーザは、発行されたクエリの受信に応答して検索エンジンによって戻されたＷｅｂページ（オブジェクトｕ）をブラウズし、各検索クエリ（オブジェクトｐ）は、１つまたは複数のそれぞれのＷｅｂページ（オブジェクトｕ）を参照する。以上の観点から、Ｗｅｂユーザ情報をクラスタリングする場合、ユーザがブラウズした（複数の）Ｗｅｂページ、およびそれぞれの（複数の）Ｗｅｂページを獲得するのに使用されたクエリは、より高い類似性を有し、クラスタリングプロセスにおいて一緒にクラスタリングされる傾向にあるはずである。同様に、Ｗｅｂページをクラスタリングする場合、Ｗｅｂページがどのようにユーザによって使用され、どのようにそれぞれの検索クエリによって参照されたかも考慮に入れなければならない。これに対処するため、以下に説明するとおり、強化されたクラスタリングアルゴリズムが、複数のデータオブジェクトのそれぞれのオブジェクトの間のマイニングされた関係の相関的要素として、そのような異質のデータオブジェクトをクラスタリングする。

【００１５】

本発明の一態様は、本来備わっている相互関係に基づき、クラスタリングされるオブジェクトに、他のオブジェクトへのリンクを提供する。各オブジェクトに接続する複数のリ

ンク（およびそれらのリンクが接続する先の複数のオブジェクト）の所定のリンクには、そのオブジェクトに対するリンクの関連性を反映するように異なる重要度で重みを付けることができる。例えば、クラスタリングされた複数のオブジェクトと同一型の複数のオブジェクトには、異なる型の複数のオブジェクトより大きい重要度を提供することができる。本発明は、様々なレベルの重要度を異なるオブジェクトまたは異なる型のオブジェクトに割り当てることができることによる仕組みを提供する。異なるオブジェクト（または異なる型のオブジェクト）に異なるレベルの重要度を割り当てることを、本明細書では、重要度付きクラスタリングと呼ぶ。異なるオブジェクトの様々なレベルの重要度により、クラスタリングの結果および効果を改善する場合が多い。次に、検索語提案のためのマルチ型データオブジェクトの強化されたクラスタリングの、以上および他の態様を説明する。

【００１６】

Ｗｅｂサイトと関係があり、検索クエリを生成する際にエンドユーザによってより頻繁に使用されたと判定された（複数の）語／（複数の）キーワードは、一般に、Ｗｅｂサイトに関して検索エンジン結果を最適化するために、Ｗｅｂサイト主催者／広告主によって選択される。これに留意して、以下の本発明のシステムおよび方法は、本実装において、検索語提案であり手近なタスクと相互に関係があると決定されたマルチ型データオブジェクトをマイニングする。そのようなマルチ型データオブジェクトには、履歴クエリを検索エンジンに送信することによって獲得された結果からマイニングされた意味論的文脈（例えば、テキスト、ＵＲＬ、結果タイトル、および各結果の短い説明など）で高められマイニングされた履歴検索クエリの（複数の）語、特定の履歴検索クエリに応答してユーザによって選択されたＷｅｂページの集合、ユーザに固有の情報（例えば、ユーザのＷｅｂサイトアクセス情報、検索クエリを生成するのに使用されたマシンのＩＰアドレスなど）、および／または関係のあるデータオブジェクトの型を含む。

【００１７】

マルチ型データオブジェクト間の類似度を、識別され重みが付けられたコンテキストの類似度と、計算されたオブジェクト間およびオブジェクト内関係の類似度との線形結合として決定する。データオブジェクトには、オブジェクト間および型内関係に由来するリンク構造を分析することにより、異なる重みを割り当てる。したがって、複数のデータオブジェクトのそれぞれのオブジェクト間の類似度には、オブジェクト自体の属性の類似度だけでなく、オブジェクトの関係の類似度も含む。

【００１８】

以上の計算されたマルチ型オブジェクト関係の観点から、強化されたクラスタリングアルゴリズムは、各オブジェクトの識別されたオブジェクト間およびオブジェクト内関係属性の相関的要素として、マルチ型データオブジェクトを繰り返しクラスタリングする。本実装において、クラスタ内のオブジェクトの重み付けされた合計を使用することによってクラスタ中心（ｃｅｎｔｒｏｉｄ）を算出するのに、変更された直接ｋ平均（ｄｉｒｅｃｔ－ｋ－ｍｅａｎｓ）アルゴリズムを使用する。これは、オブジェクトのそれぞれの関係属性を更新することによってすべての関係のあるデータオブジェクトにクラスタリング結果を伝播させる反復プロセスである。すなわち、１つの型のオブジェクトのクラスタリング結果が、新たな特徴空間を形成し、次に、この空間を、他の関係のある異なる型のオブジェクトに射影し、伝播させる。次に、関係のある型のオブジェクトに対するクラスタリングを、その更新された特徴空間を使用して実行する。この反復強化プロセスを、各オブジェクト型に対して実行することにより、特徴空間の次元を減らすよう大いに関係のあるクラスタノードをマージし、すべてのマルチ型オブジェクトにわたるクラスタリング結果が収束するまで続ける。これにより、大いに高い関係のあるマルチ型データオブジェクトの強化されたクラスタとなる。

【００１９】

エンドユーザから語を受け取ることに応答して、システムおよび方法は、語／クエリオブジェクト型に基づき、強化されたクラスタ内の（複数の）語のそれぞれの語とその（複数の）語を比較する。強化された語のクラスタは、互いに文脈上、別の語と関係のある語

を含むので、送信された入札をクラスタ内の語と比較する場合、語句を複数の関係のある文脈、すなわち「意味」の観点から評価する。さらに、各々の強化された語のクラスタは、高い関係のあるマルチ型オブジェクトの集合に由来するので、アルゴリズムは、純粋に文脈ベースの方法の欠点を克服することができる。すなわち、クエリの語の間の意味関係を効率的に強化し、語の文脈における雑音の影響を抑えることができる。受け取られた語と強化されたクラスタ内のオブジェクトの特徴空間とを比較することに応答して、１つまたは複数の検索語提案を識別する。これらの検索語提案をエンドユーザに伝達する。

【００２０】

（例示的システム）

必要ではないが、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する。プログラムモジュールには、一般に、特定のタスクを実行する、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、データ構造などが含まれる。本システムおよび方法を上述の文脈で説明するが、以下に説明する動作および操作は、ハードウェアに実装することもできる。

【００２１】

図２は、検索語提案のためのマルチ型データオブジェクトの強化されたクラスタリングを行うための例示的システム２００を示す。本実装において、システム２００は、ネットワーク２０４を介してクライアントコンピュータ処理装置２０６に結合された編集確認サーバ（ＥＶＳ）２０２を含む。例えば、クライアントコンピュータ処理装置２０６またはＥＶＳ２０２上で実行される別のアプリケーション（図示せず）から、（複数の）語２０８を受け取ったことに応答して、ＥＶＳ２０２は、提案される語リスト２１０を生成し、クライアントコンピュータ処理装置２０６に伝達することにより、エンドユーザが、実際に（複数の）語の入札を行う前に、（複数の）語２０８に意味的および／または文脈的に関係のある語の集合を評価することができる。ネットワーク２０４は、オフィス、企業全体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で一般的であるような、ローカルエリアネットワーク（ＬＡＮ）通信環境と一般的なワイドエリアネットワーク（ＷＡＮ）通信環境とのあらゆる組み合わせを含むことが可能である。システム２００がクライアントコンピュータ処理装置２０６を含む場合、クライアントコンピュータ処理装置は、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タ処理装置（例えば、セルラー電話機、パーソナルデジタルアシスタント、またはハンドヘルドコンピュータ）などのあらゆる型のコンピュータ処理装置である。

【００２２】

ＥＶＳ２０２は、提案される語リスト２１０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む。複数のコンピュータプログラムモジュールには、例えば、検索語提案（ＳＴＳ）モジュール２１２が含まれる。本実装において、説明および例示的図解の目的で、ＳＴＳモジュール２１２は、履歴クエリの語マイニング、Ｗｅｂページ検索、特徴抽出、特徴空間次元の縮小および正規化、マルチ型データオブジェクトの強化されたクラスタリング、ユーザ入札の（複数の）語を強化されたクラスタの内容とマッチングすることにより検索語提案を実行すること、および語分類などの、複数の機能を実行するとして説明する。以上の複数の操作のそれぞれの操作は、ＳＴＳモジュール２１２と通信する１つまたは複数の他のコンピュータプログラムモジュール（図示せず）によって実行されることも可能であることを認めることができよう。

【００２３】

（意味論的文脈でマイニングされた履歴検索クエリの強化）

ＳＴＳモジュール２１２は、本実装において、検索語提案である、手近なタスクと相互に関係があると決定された複数のマルチ型データオブジェクト（ＭＤＯ）２１４をマイニングする。そのような複数のマルチ型データオブジェクト２１４には、複数の履歴クエリ２１６のそれぞれのクエリを検索エンジンに送信することによって獲得された検索結果からマイニングされた意味論的文脈（例えば、テキスト、ＵＲＬ、結果タイトル、および各

結果の短い説明など）でＳＴＳモジュール２１２によって高められ、マイニングされた複数の履歴検索クエリ２１６の（複数の）語、および特定の履歴検索クエリに応答してユーザによって選択されたＷｅｂページの集合、ユーザに固有の情報（例えば、ユーザのＷｅｂサイトアクセス情報、検索クエリを生成するのに使用されたマシンのＩＰアドレスなど）、および／または関係のあるマルチ型データオブジェクトの型などの「他のＭＤＯ」２１８を含む。

【００２４】

詳細には、ＳＴＳモジュール２１２は、（複数の）クエリログ２２０から履歴クエリ２１６の集合を検索する。（複数の）履歴クエリ２１６は、１名または複数名のユーザによって検索エンジンに以前に送信された検索クエリの語を含む。ＳＴＳモジュール２１２は、出現頻度の相関的要素として履歴クエリ群２１６を評価することより、高いＦＯＯ（出現頻度）の検索語２２２と比較的低い出現頻度の検索語２２４を識別する。本実装において、設定可能な閾値を使用することにより、履歴クエリが比較的高い出現頻度を有するか、または比較的低い出現頻度を有するかを決定する。例えば、少なくとも閾値回数、出現する複数の履歴クエリ２１６の中の検索クエリの語は、高い出現頻度を有するとされる。類似して、閾値回数より少ない回数、出現する複数の履歴クエリ２１６の中の検索クエリの語は、低い出現頻度を有するとされる。例示の目的で、そのような閾値を「他のデータ」２２６のそれぞれの部分として示す。

【００２５】

ＳＴＳモジュール２１２は、各クエリを１つずつ（検索クエリ２２７）、検索エンジン２２８に送信することにより、高い出現頻度のクエリの語２２２の意味論上／文脈上の意味をマイニングする。検索クエリ２２７を受け取ったことに応答して、検索エンジン２２８は、（複数の）検索結果２３０の中でランク付けされたリスト（数が設定可能である）をＳＴＳモジュール２１２に戻す。ランク付けされたリストは、ＵＲＬ、結果タイトル、ならびに送信された検索クエリ２２７に関係のあるクエリの語の短い説明および／または文脈を含む。ランク付けされたリストは、検索結果２３０の中に格納される。そのような検索結果の検索が、各検索クエリ２２７に関して行われる。

【００２６】

ＳＴＳモジュール２１２は、ＷｅｂページＨＴＭＬ（ハイパーテキストマークアップ言語）を解析して、検索された各検索結果２３０から、各クエリの語２２２に関するクエリの語のＵＲＬ、結果タイトル、ならびに短い説明および／または文脈を抽出する。クエリの語のＵＲＬ、結果タイトル、短い説明および／または文脈、ならびに検索された検索結果２３０を獲得するのに使用された検索クエリ２２７は、抽出された特徴２３２のそれぞれのレコードの中にＳＴＳモジュール２１２によって格納される。

【００２７】

高い出現頻度のクエリの語２２２に関して検索結果２３０を解析した後、ＳＴＳモジュール２１２は、抽出された特徴２３２に対してテキスト前処理操作を実行して、抽出された特徴から個々のキーワードに入る言語トークンを生成する（トークン化する（ｔｏｋｅｎｉｚｅ））。トークンの次元を縮小するため、ＳＴＳモジュール２１２は、例えば、ポーターステミング（Ｐｏｒｔｅｒ　ｓｔｅｍｍｉｎｇ）アルゴリズムを使用して、あらゆるストップワード（ｓｔｏｐ－ｗｏｒｄ）（例えば、「ｔｈｅ」、「ａ」、「ｉｓ」など）を削除し、一般的な接尾辞を削除することにより、キーワードを正規化する。ＳＴＳモジュール２１２は、もたらされる抽出された特徴２３２を、１つまたは複数の語に基づくＭＤＯ（マルチ型データオブジェクト）ベクトル２３４に構成する。

【００２８】

各語に基づくマルチ型データオブジェクトベクトル２３４は、語の頻度に基づく次元、および逆ドキュメント頻度（ＴＦＩＤＦ）スコアを有する。第ｉ番のベクトルの第ｊ番のキーワードに対する重みは、以下のとおり計算される。すなわち、

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

ただし、TF_{ij}は、語の頻度（第ｉ番のレコード内のキーワードｊの出現回数）を表し

、Ｎは、クエリの語の総数であり、ＤＦ$_j$は、キーワードｊを含むレコードの数である。

【００２９】

各クエリの語のベクトル表現を所与として、コサイン関数を使用することにより、一組の語の間の類似度を測定する（ベクトルは正規化されていることを想い起こされたい）。すなわち、

【００３０】

【数１】

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

【００３１】

すなわち、２つの語の間の距離（類似測度）は、以下のとおり定義される。すなわち、

ｄｉｓｔ（ｑ$_j$，ｑ$_k$）＝１－ｓｉｍ（ｑ$_j$，ｑ$_k$）

このような類似測度を、「他のデータ」２２６のそれぞれの部分として示す。例示的なこのような類似度値を、以下に説明する表１の例示的な提案される語のリスト２１０の中で示す。

【００３２】

（ユーザが選択したＷｅｂページおよびユーザ情報のマイニング）

Ｗｅｂサイトの検索エンジン結果を最適化するために実質的に最も関係のある語の集合（検索語提案）を識別するため、ＳＴＳモジュール２１２は、複数の履歴クエリ２１６とは異なる／異質のマルチ型データオブジェクト２１４をマイニングする。説明の目的のために、これらのマイニングされる複数のオブジェクトを「他のＭＤＯ」２１８と表す。所定の実装において、「他のＭＤＯ」２１８には、例えば、エンドユーザが選択したＷｅｂページおよび／またはユーザ固有の情報が含まれ、ただし、ユーザは、履歴クエリ２１６を検索エンジン２２８に送信することに関連するユーザである。ＳＴＳモジュール２１２は、クエリログ２２０から、エンドユーザが選択したＷｅｂページを抽出する。エンドユーザが選択したＷｅｂページは、希薄であることも、そうでないことも可能であり、希薄は平均すると、例えば、履歴クエリ２１６当たり２から３のＷｅｂページになる。ＳＴＳモジュール２１２は、（複数の）クエリログ２２０または他のデータソースからユーザ固有の情報を抽出する。ユーザ固有の情報には、例えば、複数の履歴クエリ２１６のそれぞれのクエリを送信するのに使用されたマシンのＩＰ（インターネットプロトコル）アドレス、ＧＵＩＤ、および／またはＷｅｂサイトアクセス情報（例えば、マイクロソフト社のドットネットパスポートの情報）が含まれる。

【００３３】

（強化されたマルチ型データオブジェクトのクラスタリング）

ＳＴＳモジュール２１２は、クラスタリング分析のためにマルチ型の相互に関係のあるデータオブジェクト（ＭＤＯ２１４）間の関係を十分に詮索する。複数のマルチ型データオブジェクト２１４には、ｎ個の異なる型のオブジェクト、Ｘ$_1$、Ｘ$_2$、．．．、Ｘ$_n$（例えば、複数の履歴クエリ２１６および「他のＭＤＯ」２１８）が含まれる。各型のデータオブジェクトＸ$_i$は、特徴集合Ｆ$_i$で記述される。同一型内の複数のデータオブジェクトは、型内関係、Ｒ$_i$⊆Ｘ$_i$×Ｘ$_i$で相互に関係がある。２つの異なる型からのデータオブジェクトは、型間関係、Ｒ$_{ij}$⊆Ｘ$_i$×Ｘ$_j$を介して関係がある。これらの関係と区別するため、Ｆ$_i$を、データオブジェクトのコンテント特徴（ｃｏｎｔｅｎｔ　ｆｅａｔｕｒｅ）と呼ぶ。特定のオブジェクトｘ∈Ｘ$_i$に関して、ｘ．Ｆ$_i$を使用してそのオブジェクトのコンテント特徴を表し、ｘ．Ｒ$_i$⊆Ｘ$_i$およびｘ．Ｒ$_{ij}$⊆Ｘ$_j$を使用して、それぞれＸ$_i$およびＸ$_j$の中でそのオブジェクトに関係のあるオブジェクトを表す。マルチ型の相互に関係のあるデータオブジェクトをクラスタリングすることの問題は、各型の複数のオブジェクトＸ$_i$をＫ$_i$個のクラスタに分けて、各クラスタ内のデータオブジェクトが高い類似度を有し、異なるクラスタからのオブジェクトが似ていないようにすることであ

る。

【００３４】

マルチ型データオブジェクト群２１４のオブジェクトが、コンテント特徴と、複数のマルチ型データオブジェクト２１４の中の他の(複数の)オブジェクトとの関係をともに有することを考慮すると、２つのオブジェクト間の類似度を、以下の数式に従って決定する。すなわち、

$$S = \alpha \cdot s_f + \beta \cdot s_{intra} + \gamma \cdot s_{inter} \quad (1)$$

ただし、s_fは、コンテンツ類似度であり、s_{intra}およびs_{inter}はそれぞれ、型内類似度および型間類似度であり、α、β、およびγは、異なる類似度に対する重みであり、α＋β＋γ＝１である。

【００３５】

（１）から、２つのオブジェクト間の類似度は、コンテンツ類似度と関係類似度の線形結合である。α、β、およびγに異なる値を割り当てることにより、ＳＴＳモジュール２１２は、全体的な類似度における異なる類似度の重みを調整／構成することができる。例えば、α＝１、β＝γ＝０である場合、コンテンツ特徴間の類似度が考慮される。β＝０に設定することにより、ＳＴＳモジュール２１２は、型内類似度の効果を顧慮しない。

【００３６】

等式１における類似度は、通常、オブジェクトの型およびアプリケーションによって決定され、異なる関数を使用して定義することができる。例えば、２つのＷｅｂページ間のコンテンツ類似度は、Ｗｅｂページのコンテントに由来する２つのキーワードベクトルのコサイン関数$x \in X$、$y_i \in x$．R_Yとして定義することが可能である。

【００３７】

特定のオブジェクトの関係特徴は、エントリが特定のオブジェクトに関係のあるオブジェクトに対応するＭＤＯベクトル２３４によって表される。所定の実装において、各エントリは、関係の重みに対応する数値である。例えば、２つのオブジェクト型、$X = \{x_1, x_2, L\ x_m\}$および$Y = \{y_1, y_2, L\ y_n\}$を所与として、オブジェクトの型間関係は、$V_X = [v_1, v_2, L, v_n]^T$として定義され、ただし、$v_i \neq 0$の場合であり、それ以外の場合、$v_i = 0$である。すると、Ｘの中の２つのオブジェクト間の型間関係R_{XY}上の類似度$S_{inter-XY}$も、２つのベクトルのコサイン関数として定義することが可能である。

【００３８】

X_iの中のオブジェクトが、複数のデータオブジェクト型と型間関係を有する場合、最終的な型間類似度は、すべての型間類似度の線形結合であることが可能である。

【００３９】

定義された類似度関数を使用して、ＳＴＳモジュール２１２は、複数の履歴クエリ２１６および「他のＭＤＯ」２１８の間のレイヤ内関係／リンクおよびレイヤ間リンクを識別する。クラスタリングにおけるレイヤ間リンクの使用は、所定の型のオブジェクトのクラスタリングが別の型のオブジェクトによって影響を及ぼされる可能性があることを認識する。例えば、Ｗｅｂページオブジェクトのクラスタリングは、ユーザオブジェクトの構成、状態、および特性によって影響を及ぼされる可能性がある。したがって、それらのマイニングされたレイヤ内関係およびレイヤ間関係は、以下に説明するとおり、相互に関係のあるデータオブジェクトのクラスタ品質を向上させるのに使用される。マイニングされたレイヤ間およびレイヤ内データオブジェクト関係は、各オブジェクトのそれぞれのＭＤＯベクトル２３４の中に格納される。

【００４０】

所定の実装において、識別されたレイヤ間リンク／関係は、例えば、以下の１つまたは複数を表す。すなわち、

・コンテントに関係のある情報、例えば、履歴クエリ２１６内のリンク、およびユーザが選択した（クリックスルーされた（ｃｌｉｃｋｅｄ－ｔｈｒｕ））Ｗｅｂページに対応するリンク

・例えば、履歴クエリ２１６内のリンク、およびユーザ固有の情報により決定された、関連するトピックに対するユーザの関心
・例えば、ユーザ固有の情報と選択されたＷｅｂページの間のリンクを介して決定された、選択されたＷｅｂページに対するユーザの関心
所定の実装において、識別にされたレイヤ内リンク／関係（同一のデータ型のオブジェクト間の関係）は、例えば、以下の１つまたは複数を表す。すなわち、
・以下により詳細に説明する、クエリ内のリンク
・ユーザが選択したＷｅｂページ内の内向き／外向きの（ｄｉｒｅｃｔｅｄ　ｉｎ／ｏｕｔ）ハイパーリンクで示される、推奨される（複数の）Ｗｅｂページ
・例えば、それぞれのユーザの間で識別された関係／リンクで示される人間関係。所定の実装において、この型の関係情報は、ユーザプロファイルの計算された類似度からマイニングされる。ユーザプロファイルには、例えば、人口統計、地理的位置、（複数の）関心などが含まれる。所定の実装において、ユーザプロファイルは、それぞれのユーザが供給するＷｅｂサイトアクセス情報を介してアクセスされる。

【００４１】

クエリ内のリンクに関して、クエリ内のリンクによって示されるレイヤ内関係は、初期の履歴クエリ２１６と後続のクエリ改良（ｒｅｆｉｎｅｍｅｎｔ）（複数の履歴クエリ２１６のそれぞれのクエリによっても表現される）との間の、または後続のクエリ改良間のリンクを表現する。所定の実装において、そのような情報は、（複数の）クエリログ２２０から検索されたクリックスルーＷｅｂページ情報から抽出される。より詳細には、初期検索クエリ結果が満足のいくものでないと決定されると、ユーザは、初期クエリが送信された時点から設定可能な合計時間内に、１つまたは複数の改良されたクエリを検索エンジン２２８へ送信することが想定される。設定可能な合計時間は、クエリセッションを表現する。１回または複数回のそのような検索クエリの語の改良の後、ユーザは、満足のいく検索結果を獲得することができる。例えば、ユーザが、製品サポートのＷｅｂサイトを訪れ、「クッキー（ｃｏｏｋｉｅ）」という初期クエリを送信することを考慮されたい。検索結果が満足のいくものでない（例えば、広すぎる）と決定されると、ユーザは、「クッキーを可能にする（ｅｎａｂｌｅ　ｃｏｏｋｉｅ）」クエリの語に変更／改良して、より満足のいく検索結果を獲得することができる。

【００４２】

所定の実装において、ＳＴＳモジュール２１２は、（複数の）クエリログ２２０の１つまたは複数の部分をそれぞれのクエリセッションにセグメント化することにより、クエリ内のリンクを識別する。各クエリセッションは、初期クエリ、１つまたは複数のクエリ改良、およびできる限り１つまたは複数のＷｅｂページクリックスルーの指示を含むことが可能である。初期クエリ、および１つまたは複数の関連するクエリ改良を分類するため、ＳＴＳモジュール２１２は、各クエリセッションのクエリ間の語の類似度を計算する。類似度の１つまたは複数の閾値基準を満たす検索クエリが、クエリ内、および対応するクエリ改良内のリンクを生成するために選択される。所定の実装において、クエリの類似度を、例えば、段落（００３８）で上述した例示的操作を使用して決定する。

【００４３】

複数のマルチ型データオブジェクト２１４の間の関係を複数のＭＤＯベクトル２３４の対応するベクトルにおいてモデル化された関係特徴としてマッピングした後、従来のクラスタリング技術を使用して（すなわち、本明細書で開示する強化されたクラスタリング操作を使用せずに）、各型のデータオブジェクトを個々にクラスタリングすることも可能である。しかし、データオブジェクトを個々にクラスタリングすることさえ、最初は、実行可能であるように思われる可能性があるものの、この技術は、実質的に限られており、問題を抱えている。そのことの１つの理由は、関係のための特徴ベクトルのサイズが非常に大きくなるにつれ、オブジェクトの数が非常に多くなるためである。また、関係のあるオブジェクトの正確なマッチングに基づく、関係特徴に関して定義された類似度が、０でないエントリの希薄さにより悪化する。別の理由は、データオブジェクト間の（複数の）関

係が、データオブジェクトに割り当てられた特徴に十分に反映されていない可能性があるが、クラスタリングプロセス自体の過程にだけ発見される可能性があることを、従来のクラスタリング技術が考慮しないためである。すなわち、既存のクラスタリング技術は、順にクラスタリング操作を行うことにより、後続の分析／クラスタリング操作においてデータを強化する際に有用な構造化された情報を提供できることを考慮しない。

【００４４】

ＳＴＳモジュール２１２は、１つのデータオブジェクト型のクラスタリング結果をそのオブジェクトと関係のあるすべてのデータオブジェクト型に、それらのデータオブジェクト型のそれぞれの関係特徴を更新することによって少なくとも伝播させることにより、従来のクラスタリング技術の以上の問題／限界に対処する。すなわち、ＳＴＳモジュール２１２は、強化された複数のクラスタ２３６のコンテントに基づく個々のマルチ型データオブジェクト２１４に対する指示されたデータオブジェクトの（複数の）関係を集約することにより、強化された複数のクラスタ２３６を生成する。例えば、クラスタリング後に２つの候補ノードが存在する場合、最も近接した２つの候補ノードを、例えば、その２つの候補ノードのベクトル値を平均することにより、マージすることができる。このマージが個々のノードを結合することにより、考慮すべきノードの数を減らすことが可能になる。真の意味において、（複数の）ＭＤＯベクトル２３４の次元を縮小する。次に、ＳＴＳモジュール２１２は、（複数の）ＭＤＯベクトル２３４をクラスタリングする。このプロセスは、すべてのオブジェクト型のクラスタリング結果が収束するまで、繰り返し実行される。

【００４５】

反復クラスタリング射影技術は、均質の型のオブジェクトを各レイヤが含む、別々のレイヤに構成された別々の型のオブジェクトからクラスタリング情報を獲得することに依拠する。ノード情報をリンク情報と組み合わせて使用することにより、クラスタリングが収束するまで（クラスタリングアルゴリズムは、レイヤ間で提供される）、クラスタリングされた結果を繰り返し射影し、伝播させる。すなわち、各型の異なる種類のノードおよびリンクが調べられることにより、クラスタリングのために使用することができる構造情報が獲得される。構造情報は、例えば、異なるデータオブジェクトを接続するリンクの型（例えば、リンクが、レイヤ間リンクであるか、またはレイヤ内リンクであるか）を考慮することにより獲得することができる。所定の型のオブジェクトの結果を別の型のオブジェクトのクラスタリング結果に繰り返しクラスタリングすることにより、データの希薄さに関連するクラスタリングの困難を削減することができる。この反復射影により、所定のレイヤクラスタリングにおける類似測度を、別の型のクラスタの個々のグループの代わりのクラスタで計算する。

【００４６】

例えば、２つのオブジェクト型、$X=\{x_1, x_2, \ldots x_m\}$および$Y=\{y_1, y_2, \ldots y_n\}$の観点からプロセスを例示すると、ＳＴＳモジュール２１２はまず、あらゆる従来のクラスタリング方法を使用して、Ｙの中のオブジェクトを$\{C_1, C_2, \ldots, C_k\}$で表されるｋ個のクラスタにクラスタリングする。$x \in X$の関係特徴ベクトルを含むＭＤＯベクトル２３４は、最初、$V_X=[v_1, v_2, L, v_n]^T$として定義され、各成分がＹの中の所定のオブジェクトに対応することを想い起こされたい。Ｙの中のクラスタにより、各成分がＹの所定のクラスタに対応し、$x.R_Y \cap C_i \neq \Phi$である場合、v_i'が０ではない、$V_x'=[v_1', v_2', L, v_k']^T$で$V_x$を置き換える。$v_i'$の数値は、オブジェクトｘからクラスタ$C_i$の中の複数のオブジェクトに対する関係の数を表現する、$|x.R_Y \cap C_i|$に、または関連するオブジェクトの重要度（オブジェクト重要度は、以下に説明する）に設定することも可能である。したがって、Ｘの中のオブジェクトのクラスタリングは、この新たな型間関係特徴に基づく。プロセスは、所定の型のクラスタリング結果を別の型にそれらの型のレイヤ間関係により繰り返し射影することによって、収束まで続けられる。

【００４７】

以上の強化されたクラスタリングアルゴリズムの利点は、クラスタリング結果が、コンテントからのデータ分布を反映するだけでなく、他のデータ型との関係も反映することである。このアルゴリズムは、データ希薄性の問題もある程度、解決することができる。固定の特徴空間に関して類似度を定義する既存のクラスタリングアプローチと比較すると、マルチ型データオブジェクトの強化されたクラスタリングを行うための説明するシステムおよび方法は、クラスタリングプロセス中に２つのオブジェクト間で類似度を更新することにより、新たに発見された関係特徴空間に適応する。さらに、所定の実装において、あらゆる従来のクラスタリングアルゴリズムをこの提案するフレームワークに埋め込むことにより、クラスタリングパフォーマンスを向上させることができる。

【００４８】

（リンク分析およびオブジェクトの重要度）

一部のデータオブジェクトおよびアプリケーションに関して、同一型内の複数のマルチ型データオブジェクト２１４は、クラスタリングプロセスにおいて異なる重要度を有する可能性がある。通常の実施例には、若干のＷｅｂページが権威のあるページであるため、より重要である場合のＷｅｂページ／ユーザクラスタリング、および一部のユーザが、アイテムのふさわしさ（ｂｅｌｏｎｇｉｎｇｎｅｓｓ）を決定する際により権威を有するべき場合の、共同フィルタリングなどのためのアイテム／ユーザクラスタリングが含まれる。オブジェクトをノード見なし、オブジェクト間の（複数の）関係をリンクと見なした場合、ＨＩＴＳアルゴリズムなどの従来のリンク分析方法を使用することにより、各データオブジェクトの固有値を計算する。しかし、複数の型のデータオブジェクトが関係する場合、この方法は、異なる型のオブジェクトの重要度が比較できないため、うまく作用しない。

【００４９】

この問題に対処するため、マルチ型データオブジェクトの強化されたクラスタリングのための説明するシステムおよび方法は、次のとおりＨＩＴＳアルゴリズムを拡張する。型内でオブジェクト重要度の相互強化を考慮するだけでなく、型間の相互強化も考慮する。各ノードには、ハブスコア（ｈｕｂ　ｓｃｏｒｅ）および権威スコア（ａｕｔｈｏｒｉｔｙ　ｓｃｏｒｅ）が割り当てられる。

【００５０】

簡単にするため、提案するアルゴリズムを例示する実施例として、２つの型の相互に関係のあるオブジェクトを含むケースを使用することを続ける。２つの型のオブジェクト、$X=\{x_1, x_2, \ldots x_m\}$、$Y=\{y_1, y_2, \ldots y_n\}$、ならびに$R_X$、$R_Y$、$R_{XY}$、および方向性が考慮される場合$R_{YX}$という関係を所与として、隣接する行列を使用してリンク情報を表現する。L_XおよびL_Yがそれぞれ、集合ＸおよびＹ内のリンク構造の隣接する行列を表す。L_{XY}およびL_{YX}が、Ｘの中のオブジェクトからＹの中のオブジェクトへのリンクの隣接する行列を表す。例えば、ノードx_iからノードy_jへの所定のリンクが存在する場合、$L_{XY}(i, j)=1$である。

【００５１】

次の２つのレベルの計算が存在する。すなわち、１つのレベルは、同一型からのオブジェクトのハブ値（ｈｕｂ　ｖａｌｕｅ）および権威値（ａｕｔｈｏｒｉｔｙ　ｖａｌｕｅ）が、型内関係によって互いを強化することであり、他方のレベルは、異なる型のノードの重要度が、型間関係によって互いを強化することである。このアプローチにおける計算は、以下のとおり書き表される。

【００５２】

【数２】

$$\begin{cases} a(X) = \beta L_X{}^T h(X) + (1-\beta) L_{XY} i(Y) \\ h(X) = \beta L_X a(X) + (1-\beta) L_{XY} i(Y) \\ i(X) = a(X) + h(X) \\ \\ a(Y) = \gamma L_Y{}^T h(Y) + (1-\gamma) L_{YX} i(X) \\ h(Y) = \gamma L_Y a(Y) + (1-\gamma) L_{YX} i(X) \\ i(Y) = a(Y) + h(Y) \end{cases} \quad (2)$$

【００５３】

ただし、ａ（Ｘ）およびｈ（Ｘ）はそれぞれ、Ｘ内のノードの権威スコアおよびハブスコアである。同様に、ａ（Ｙ）およびｈ（Ｙ）は、Ｙ内のノードの権威スコアおよびハブスコアであり、ｉ（Ｘ）およびｉ（Ｙ）はそれぞれ、ＸおよびＹの中のノードの重要度を表す。βおよびγは、異なる関係に由来するリンクの影響を調整する重みパラメータである。

【００５４】

計算の始めに、すべてのベクトル、ａ（Ｘ）、ｈ（Ｘ）、ａ（Ｙ）、およびｈ（Ｙ）は、１に初期設定される。ハブスコアおよび権威スコアは、各反復時に等式（２）を使用して更新される。各反復の終了時に、ベクトルは、次の反復計算のために正規化される。このアルゴリズムは、各オブジェクト型内で正規化された一様な重要度を提供し、型間関係を介して他の型の関連するオブジェクトの重要度を考慮することにより、より妥当な結果を獲得する。

【００５５】

オブジェクトの重要度スコアを所与として、説明する強化されたクラスタリングプロセスは、オブジェクトの重要度を反映するように変更される。本実装において、ｋ平均クラスタリングアルゴリズムが、重み付けされたｋ平均アルゴリズムに変更される。すなわち、クラスタ中心を計算する際、クラスタメンバの重み付けされた合計を新たな中心として使用して、クラスタがそれらの重要なオブジェクトの方にバイアスされるようにする。

【００５６】

以上の観点から、ＳＴＳモジュール２１２は、複数のマルチ型データオブジェクト２１４の間における型間関係と型内関係の両方に基づき、マルチ型データオブジェクトの重要度を区別する。この重要度が、クラスタリングプロセスに組み込まれる。

【００５７】

（入札の語の例示的処理）

エンドユーザ（例えば、広告主、Ｗｅｂサイト主催者など）から（複数の）語２０８を受け取ったことに応答して、ＳＴＳモジュール２１２は、（複数の）語２０８を複数の強化された語のクラスタ２３６の中の語／句のそれぞれの語／句と比較する。複数の強化された語のクラスタ２３６は、Ｗｅｂページおよびユーザとの語の相互関係に由来する、文脈上で互いに関係するだけでなく、意味上でも互いに関係する複数の語を含むので、（複数の）語２０８は、複数の関係のある履歴上の文脈、すなわち「意味」の観点から評価される。

【００５８】

所定の実装において、ＳＴＳモジュール２１２により、（複数の）語２０８が、強化されたクラスタ２３６からの（複数の）語と合致すると決定する場合、検索語提案モジュール２１２は、強化されたクラスタ２３６から提案される語のリスト２１０を生成する。本実装において、合致は、正確な合致であっても、単数形／複数形、つづり間違い、句読点

などの少数の違いを伴う合致であってもよい。戻されるリストは、ＦＯＯと信頼度値の結合により順に並べられる。

【００５９】

所定の実装において、(複数の)語がクラスタからの語に合致する場合、そのクラスタが、提案される語のリストの中でエンドユーザに戻される。提案される語のリスト２１０は、（複数の）語に意味上、および／または文脈上、関係があると決定された語／句、それぞれの（複数の）語と（複数の）語との類似測度（信頼度値）、およびそれぞれの（複数の）語出現頻度（ＦＯＯ）を含む。戻されるリスト２１０は、ＦＯＯと信頼度値の結合により順に並べられる。

【００６０】

ＳＴＳモジュール２１２により、語２０８が複数の強化された語のクラスタ２３６の中の語と合致すると決定する場合、検索語提案モジュール２１２により、複数の強化された語のクラスタ２３６の複数のクラスタの中の語から、複数の提案された語リスト２１０を生成する。複数のリストは、クラスタサイズにより順に並べられ、各リスト内の語は、ＦＯＯと信頼度値の結合により順に並べられる。

【００６１】

合致するクラスタがないと識別された場合、クエリの語は、低いＦＯＯを有するクエリの語から生成された、拡張されたクラスタ群に対してさらにマッチングされる。所定の実装において、低いＦＯＯを有するクエリ語は、高い出現頻度の履歴クエリログの語から生成された、複数の強化された語のクラスタ２３６に関する分類子（ｃｌａｓｓｉｆｉｅｒ）（例えば、Ｋ－最近傍（ｎｅａｒｅｓｔ　ｎｅｉｇｈｂｏｒ）の分類子）を訓練することによってクラスタリングされる。低い出現頻度を有すると決定された履歴クエリの語が１つずつ、検索エンジンに送信される。次に、戻された検索結果の選定された検索結果（例えば、第１の最上位ランクのＷｅｂページ、および／または等）から、特徴が抽出される。抽出された特徴を、正規化し、低いＦＯＯを有するクエリの語を表現するのに使用する。次に、クエリの語は、既存の複数のクラスタに分類することにより、訓練された分類子に基づいて複数の拡張されたクラスタが生成される。次に、エンドユーザが送信した（複数の）語を、これらの拡張されたクラスタの観点から評価することにより、提案される語のリストを識別し、エンドユーザに戻す。

【００６２】

（低いＦＯＯ語の分類）

高いＦＯＯ（出現頻度）のクエリの語２２２から生成され、強化された語の複数のクラスタ２３６が、エンドユーザが入力した（複数の）語２０８と同一の語を含まない場合、ＳＴＳモジュール２１２により、高いＦＯＯ（出現頻度）のクエリログの語２２２から生成され、強化された語の複数のクラスタ２３６から、訓練された分類子２３８を生成する。強化された語の複数のクラスタ２３６の中の語は、分類操作に適切なベクトル空間モデルの中で対応するキーワードベクトルを既に有する。さらに、ストップワードの削除および単語のステミング（ｓｔｅｍｍｉｎｇ）（接尾辞削除）により、（複数のクラスタ２３６が基づく）語のベクトル２３４の次元が縮小される。所定の実装において、さらなる次元縮小技術、例えば、特徴選択またはパラメータ変更（ｒｅ－ｐａｒａｍｅｔｅｒｉｚａｔｉｏｎ）を使用することができる。

【００６３】

本実装において、クラスが未知のクエリの語２２２を分類するため、ＳＴＳモジュール２１２は、ｋ－最近傍の分類子（ｋ－Ｎｅａｒｅｓｔ　Ｎｅｉｇｈｂｏｒ　ｃｌａｓｓｉｆｉｅｒ）のアルゴリズムを使用することにより、クラスが既知のすべてのクエリの語２２２に対応する特徴ベクトルに依拠する、クラスが既知のすべてのクエリの語２２２の中でｋ個の最も類似した近傍を求め、近傍のクラスラベルの重み付けされたマジョリティ（ｗｅｉｇｈｔｅｄ　ｍａｊｏｒｉｔｙ）を使用することにより、新たなクエリの語のクラスを予測する。この場合、既に強化された語の複数のクラスタ２３６の中にあるおのおののクエリの語には、対応するクラスタのラベルと同一のラベルが割り当てられる一方で、

おのおのの強化されたクラスタ２３６には、単なるシーケンス番号でラベルが付けられる。それらの近傍には、Ｘに対する各近傍の類似度を使用して重みが付けられ、ただし、類似度は、２つのベクトルの間のユークリッド距離、またはコサイン値で測定される。コサイン類似度は、以下のとおりである。すなわち、

【００６４】

【数３】

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

【００６５】

ただし、Ｘは、試験語（ｔｅｓｔ　ｔｅｒｍ）、すなわち、ベクトルとして表現される、分類されるべきクエリの語であり、D_jは、第ｊ番の訓練中の語であり、t_iは、ＸとＤ$_j$が共有する語であり、x_iは、Ｘの中のキーワードt_iの重みであり、d_{ij}は、D_jの中のキーワードt_iの重みであり、

【００６６】

【数４】

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

【００６７】

は、Ｘのノルムであり、$\|D_j\|_2$は、D_jのノルムである。したがって、試験語Ｘのクラスラベルは、以下のとおり、すべての近傍のクラスラベルの重み付けされたマジョリティである。すなわち、

【００６８】

【数５】

$$\mathrm{label}(X) = \arg\max_{l_i} \left(\sum_{\text{ラベル}(D_j)=l_i\text{である場合のすべての } D_j} \mathrm{sim}(X, D_i) \right)$$

【００６９】

別の実装において、最近傍の分類技術以外の異なる統計分類－機械学習技術（例えば、回帰モデル、ベイズ分類子、判定ツリー、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む）を使用して、訓練された分類子２３８が生成される。

【００７０】

ＳＴＳモジュール２１２により、低いＦＯＯ（出現頻度）のクエリの語２２４を１つずつ（それぞれ検索クエリ２２７を介して）、検索エンジン２２８に送信する。特定の検索クエリ２２７に関連する（複数の）検索結果２３０を受け取ったことに応答し、既に説明した技術を使用して、ＳＴＳモジュール２１２は、（複数の）検索結果２３０によって識別された１つまたは検索された検索結果２３０から、特徴を抽出する（抽出された特徴２３２）。本実装において、特徴は、第１の最上位ランクの（複数の）検索結果２３０から抽出される。おのおの検索され、解析された（複数の）検索結果２３０に関して、ＳＴＳモジュール２１２は、抽出された特徴２３２のそれぞれのレコードの中に以下の情報を格納する。すなわち、ＵＲＬ、結果タイトル、クエリの語の短い説明および／または文脈、ならびに検索された検索結果２３０を獲得するのに使用された検索クエリ２２７である。次に、ＳＴＳモジュール２１２は、低いＦＯＯクエリの語２２４に由来する抽出済みの特徴２３２をトークン化し、次元を縮小し、正規化することにより、語の複数のベクトル２３４を生成する。次に、ＳＴＳモジュール２１２は、クエリの語を、語のクラスタ２３６のそれぞれの集合にクラスタリングする。このクラスタリング操作は、（高いＦＯＯクエリの語２２２から生成された）訓練された分類子２３８を使用して実行される。

【００７１】

ＳＴＳモジュール２１２は、（低いＦＯＯクエリの語２２４に基づいて生成された、）これらの拡張された語の複数のクラスタの観点からエンドユーザが送信した（複数の）語２０８を評価することにより、１つまたは複数の提案される語のリスト２１０を識別し、エンドユーザに戻す。例示的なそのような手続きを段落（００６３）および段落（００６６）で上述しており、以下のセクションで説明する。

【００７２】

（例示的な検索語提案のリスト）

提案される語のリスト２１０は、例えば、（複数の）語２０８に関係があると決定された語、それぞれの（複数の）語と（複数の）語２０８との類似測度（信頼度値）、およびそれぞれの（複数の）語のＦＯＯ（出現頻度）、すなわち、履歴クエリログ内の頻度を含む。関係のある（複数の）語を識別する、類似測度を生成する、およびＦＯＯ値を生成するための技術は、上述した。

【００７３】

表１は、「ｍａｉｌ」という語２０８に関係があると決定された語の例示的な提案される語のリスト２１０を示す。本実施例では、（複数の）語２０８に関係のある語が、「提案される語」という題名が付けられた第１列の中に示されている。

【００７４】

【表１】

表１

入札の語「ＭＡＩＬ」に関する例示的な提案される語のリスト

提案される語	類似度	頻度	〈文脈〉
hotmail	0.246142	93161	関係のあるオンライン電子メール
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	0.0962268	4481	
free email	0.230611	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	0.0200538	108	
webmail.yahoo.com	0.08789	71	
free email account	0.0234481	65	
提案される語	類似度	頻度	
mail	1	2191	関係のある従来のメール
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	

【００７５】

表１を参照して、提案される語のリストの中の語が、語類似度値（「類似度」という題名が付いた第２列を参照）および出現頻度スコア（「頻度」という題名が付いた第３列を参照）にマッピングされていることに留意されたい。「語のクラスタリング」という題名が付いたセクションにおいて後述するとおり計算される各語類似度値は、対応する提案される語（第１列）と、本実施例では「ｍａｉｌ」である（複数の）語２０８との間の類似測度を提供する。各頻度値、または頻度スコアは、履歴クエリログ内で提案される語が出

現する回数を示す。提案される語のリストを、事業目標の相関的要素としての語の類似度、および／または出現頻度スコアの相関的要素としてソートする。

【００７６】

あらゆる所与の（複数の）語２０８（例えば、ｍａｉｌなど）も、入札の語が使用されることが可能な１つに留まらない文脈を有することが可能である。そのことを考慮に入れるため、ＳＴＳモデル２１２は、いずれの提案される語が（複数の）語２０８の複数の文脈のいずれに対応するかという指示を、提案される語リスト２１０の中で提供する。例えば、表１を参照すると、「ｍａｉｌ」という（複数の）語２０８は、次の２つの文脈を有する。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である。関係のある語のそれぞれのリストを、それら２つの入札の語の文脈のおのおのに関して示すことに留意されたい。

【００７７】

さらに、あらゆる（複数の）語２０８に関して提案される語は、入札の語の複数の同義語であることも可能である。例えば、表１を参照すると、提案される語「ｕｓｐｓ」は、メールを扱う組織の頭字語であり、入札の語「ｍａｉｌ」の同義語ではない。しかし、「ｕｓｐｓ」も、「ｍａｉｌ」の入札の語に非常に強い関係の語であり、したがって、提案される語のリスト２１０の中で示されている。所定の実装において、ＳＴＳモデル２１２は、次の関連付け規則、すなわち、ｉｔｒ（Ｔ）→ｉｔｒ（Ｒ）の相関的要素として、関係のある語Ｒ（例えば、「ｕｓｐｓ」）と関係のある語Ｔ（例えば、「ｍａｉｌ」）との間の関係を決定する。ただし、「ｉｔｒ」は、「関心がある」を表現する。ユーザ（広告主、Ｗｅｂサイト主催者、および／または等）がＲに関心がある場合、そのユーザは、Ｔにも関心がある。

【００７８】

（例示的処理手順）

図３は、検索語提案のためにマルチ型データオブジェクトの強化されたクラスタリングを行うための例示的処理手順３００を示す。説明の目的のために、処理手順の操作は、図２の特徴に関連して説明する。（すべての参照符号は、構成要素が最初に導入される図面の番号で始める）。ブロック３０２では、ＳＴＳ（検索語提案）モジュール２１２（図２）により、クエリログ２２０から履歴クエリの語２１６を収集する。ＳＴＳモジュール２１２により、出現頻度の相関的要素として履歴クエリ２１６を構成する。ブロック３０４では、ＳＴＳモジュール２１２により、高い出現頻度のクエリの語２２２を検索エンジン２２８に送信し、対応する検索結果２３０を受け取る。ブロック３０６では、ＳＴＳモジュール２１２により、おのおのの（複数の）検索結果２３０からスニペット（ｓｎｉｐｐｅｔ）記述を抽出し、そのスニペット記述（抽出された特徴２３２）を一緒にマージすることにより、複数のＭＤＯベクトル２３４に基づく語を形成する。それぞれの高い出現頻度のクエリの語２２２のおのおのに関して、それぞれの語のベクトルが生成される。

【００７９】

ブロック３０８では、ＳＴＳモジュール２１２により、例えば、（複数の）クエリログ２２０、Ｗｅｂサイトのユーザアクセス情報などから、「他のＭＤＯ」２１８をマイニングする。ＳＴＳモジュール２１２により、マイニングされた「他のＭＤＯ」２１８の特徴空間を表現するそれぞれの複数のＭＤＯベクトル２３４が生成される。ブロック３１０では、ＳＴＳモジュール２１２により、ＭＤＯ２１４のそれぞれのＭＤＯの間におけるオブジェクト内およびオブジェクト間関係／リンクを識別する。ブロック３１２では、ＳＴＳモジュール２１２により、ＭＤＯ２１４のそれぞれのＭＤＯベクトル２３４に基づいてＭＤＯ２１４の強化されたクラスタリングを実行することにより、強化されたクラスタ２３６を生成する。ブロック３１２の異質なデータオブジェクトの強化されたクラスタリングの詳細を、図５を参照して以下に説明する。処理手順３００は、ページ上の参照符号「Ａ」で示すとおり、図４のブロック４０２に継続する。

【００８０】

図４は、検索語提案のためにマルチ型データオブジェクトの強化されたクラスタリング

を行うための図３における例示的処理手順３００の続きである。説明の目的のために、処理手順の操作は、図２の特徴に関連して説明する。ブロック４０２では、エンドユーザから（複数の）語２０８（図２）を受け取ったことに応答して、ＳＴＳモジュール２１２により、（複数の）語２０８に大いに類似し、関係があると決定された、強化された語の複数のクラスタ２３６からのあらゆる（複数の）語から、提案される語のリスト２１０を生成する。異なるオブジェクト型の間の相互関係を使用することにより、クラスタリングを向上させる。ブロック４０４では、ＳＴＳモジュール２１２により、キーワードクラスタ２３６からのあらゆる（複数の）語が、（複数の）語２０８に大いに類似する／関係のあると決定されたかどうかを決定する。類似する／関係のあると決定された場合、処理手順は、ブロック４０６に続き、ＳＴＳモジュール２１２により、対応する提案される語のリスト２１０をエンドユーザに送信する。類似しない／関係がない場合、処理手順は、ページ上の参照符号「Ｂ」で示すとおり、図５のブロック５０２に継続する。

【００８１】

図５は、検索語提案のためにマルチ型データオブジェクトの強化されたクラスタリングを行うための、図３および図４における例示的処理手順３００の続きである。説明の目的のために、処理手順の操作は、図２の特徴に関連して説明する。（すべての参照符号は、構成要素が最初に導入される図面の番号で始める）。ブロック５０２では、ＳＴＳモジュール２１２により、強化された語の複数のクラスタ２３６から分類子２３８（訓練された分類子）を生成し、複数のクラスタ２３６は、この時点で、高い出現頻度のクエリの語２２２に基づいている。ブロック５０４では、ＳＴＳモジュール２１２により、低い出現頻度のクエリの語２２４を１つずつ検索エンジン２２８に送信し、対応する検索結果２３０を受け取る。ブロック５０６では、ＳＴＳモジュール２１２により、検索結果２３０からスニペット記述（抽出された特徴２３２）を抽出し、その記述から語の複数のベクトル２３４を生成する。

【００８２】

ブロック５０８では、ＳＴＳモジュール２１２により、訓練された分類子２３８の観点から、低い出現頻度のクエリの語２２４から生成された語の複数のベクトル２３４を分類することにより、低い出現頻度のクエリの語２２４に基づくそれぞれの強化された語の複数のクラスタ２３６を生成する。ブロック５１０では、ＳＴＳモジュール２１２により、（複数の）語２０８と大いに類似すると決定された低い出現頻度のクエリの語２２４に基づく強化された語の複数のクラスタ２３６からのキーワード／重要句から、提案される語のリスト２１０を生成する。ブロック５１２では、ＳＴＳモジュール２１２により、提案される語のリスト２１０をエンドユーザに送信する。

【００８３】

図６は、図３のブロック３１２の強化されたクラスタリング操作の例示的な詳細を示す。説明の目的のために、ブロック３１０の操作は、図１および図２の特徴に関連して説明する。ＳＴＳモジュール２１２により実装される強化されたクラスタリングアルゴリズムへの入力には、グラフ１００のようなマルチレイヤのフレームワークグラフが含まれ、グラフ１００は、識別され、重みが付けられたオブジェクト間およびオブジェクト内関係を含む、対応する複数のノードのコンテント特徴ｆｉおよびｇｊを含む。このクラスタリングアルゴリズムの出力には、マルチ型データオブジェクトの強化されたクラスタリングを反映する新たなフレームワークグラフ１００が含まれる。新たなフレームワークグラフの若干の実装において、新たなノード位置に変更され、および／またはグラフ１００の次元を縮小するために別のノードとマージされた各々の古いノードの変化を例示することが可能である。

【００８４】

ブロック６０２では、元のフレームワークグラフが（各クラスタリング反復の前に）入力される。ブロック６０４では、考慮されている各ノードの重要度が、等式（２）を使用して決定または計算される。ブロック６０６では、任意のレイヤが、クラスタリングのために選択される。ブロック６０８では、選択されたレイヤの中のノードを適切となるよう

にクラスタリングすることにより（例えば、コンテント特徴に従って）、強化されたクラスタ２３６が生成される。若干の実装において、ノードは、所望のフィルタリングアルゴリズム（図示せず）を使用してフィルタ処理することにより、クラスタリングを向上させることができる。ブロック６１０では、各クラスタの複数のノードが１つのノードにマージされる。例えば、フィルタ処理の後に２つの候補ノードが存在する場合、最も近接した２つの候補ノードを、例えば、その２つの候補ノードのベクトル値を平均することにより、マージすることができる。このマージにより、個々のノードを結合することができることによって、考慮されなければならないノードの数を減らすことが可能になる。真の意味において、マージする操作を使用することにより、重複または重複に近いノード（ｎｅａｒ－ｄｕｐｌｉｃａｔｅ）の出現を減らすことができる。ブロック６１２では、６１０におけるマージに基づいて、対応するリンクが更新される。ブロック６１４では、クラスタリングアルゴリズムは、クラスタリングのために第２のレイヤに（任意に選択されたレイヤから）切り替える。ブロック３１２の操作は、ページ上の参照符号「Ｃ」で示されるとおり図７のブロック７０２に継続する。

【００８５】

図６の操作を参照すると、最初のクラスタリングパス（ｐａｓｓ）において、コンテント特徴だけが利用されることに留意されたい。ほとんどのケースでは、リンク特徴は、当初は、希薄すぎてクラスタリングに役立たない。図７を参照して以下に説明する、後続のクラスタリングパスでは、コンテント特徴とリンク特徴とを結合することにより、クラスタリングの有効性が高められる。コンテント特徴とリンク特徴とを結合することにより、重みを異なる値で指定し、結果を比較することが可能であり、向上した精度を有するクラスタリングを提供することが可能である。

【００８６】

図７は、図３および図６のブロック３１２の強化されたクラスタリング操作の例示的な続きを示す。ブロック７０２では、第２のレイヤの複数のノードが、複数のノードのコンテント特徴および更新されたリンク特徴に従ってクラスタリングされる。ブロック７０４では、各クラスタの複数のノードが１つのノードにマージされる。ブロック７０６では、他方のレイヤの元のリンク構造および元の複数のノードが復元される。ブロック７０８では、第２のレイヤの各クラスタの複数のノードがマージされ、対応するリンクが更新される。ブロック７１０では、この反復クラスタリングプロセスが、コンピュータ環境内で継続される。ブロック７１２では、フレームワークグラフ１００の改訂されたバージョンを出力する。

【００８７】

（例示的動作環境）

図８は、検索語提案のためにマルチ型データオブジェクトの強化されたクラスタリングを行うための、図２におけるシステム２００、および図３から図６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タ処理環境８００の例を例示する。例示的コンピュータ処理環境８００は、適切なコンピュータ処理環境の一例に過ぎず、本明細書で説明するシステムおよび方法の用法または機能の範囲について何ら限定を示唆することを意図するものではない。また、コンピュータ処理環境８００は、コンピュータ処理環境８００に例示したコンポーネントのいずれの１つ、または組み合わせに関連する依存関係または要件も有するものと解釈してはならない。

【００８８】

本明細書で説明する方法およびシステムは、他の多数の汎用または専用の、コンピュータ処理システムの環境または構成で機能する。使用に適する可能性がある周知のコンピュータ処理のシステム、環境、および／または構成の例には、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、以上のシステムまたは装置のいずれかを含む分散コンピュータ処理環境などが含まれるが、以上には限定されない。また、フレームワークのコンパクトまたはサブセットのバージョンを、ハンドヘ

ルドコンピュータ、または他のコンピュータ処理装置などの、限られたリソースのクライアントにおいて実装することもできる。本発明は、通信ネットワークを介してリンクされたリモート処理ユニット群によってタスクが実行される分散コンピュータ処理環境において実施される。分散コンピュータ処理環境では、プログラムモジュール群は、ローカルおよびリモートのメモリ記憶装置の両方の中に配置することができる。

【００８９】

図８を参照すると、検索語提案のためにマルチ型データオブジェクトの強化されたクラスタリングを行うための例示的システムは、コンピュータ８１０の形態で汎用コンピュータ処理装置を含む。コンピュータ８１０の以下に説明する態様は、クライアントコンピュータ処理装置ＰＳＳサーバ２０２（図２）および／またはクライアントコンピュータ処理装置２０６の例示的実装である。コンピュータ８１０のコンポーネントには、処理ユニット（群）８２０、システムメモリ８３０、ならびにシステムメモリから処理ユニット８２０までを含む様々なシステムコンポーネントを結合するシステムバス８２１を含むことができるが、以上には限定されない。システムバス８２１は、メモリバスまたはメモリコントローラ、周辺バス、およびあらゆる様々なバスアーキテクチャを使用するローカルバスを含め、いくつかの型のバス構造のいずれであることも可能である。例として、限定としてではなく、そのようなアーキテクチャには、インダストリスタンダードアーキテクチャ（ＩＳＡ）バス、マイクロチャネルアーキテクチャ（ＭＣＡ）バス、エンハンストＩＳＡ（ＥＩＳＡ）バス、ビデオエレクトロニクススタンダーズアソシエーション（ＶＥＳＡ）ローカルバス、およびメザニン（Ｍｅｚｚａｎｉｎｅ）バスとしても知られるペリフェラルコンポーネントインターコネクト（ＰＣＩ）バスを含むことが可能である。

【００９０】

コンピュータ８１０は、通常、様々なコンピュータ読取可能な媒体を含む。コンピュータ読取可能な媒体は、コンピュータ８１０によりアクセスすることができるあらゆる利用可能な媒体であることが可能であり、揮発性および不揮発性の媒体、取り外し可能および固定の媒体がともに含まれる。例として、限定としてではなく、コンピュータ読取可能な媒体は、コンピュータ記憶媒体および通信媒体を備えることが可能である。コンピュータ記憶媒体には、コンピュータ読取可能な命令、データ構造、プログラムモジュール、または他のデータなどの情報を格納するためにあらゆる方法または技術で実装された、揮発性および不揮発性の、取り外し可能および固定の媒体が含まれる。コンピュータ記憶媒体には、ＲＡＭ、ＲＯＭ、ＥＥＰＲＯＭ、フラッシュメモリもしくは他のメモリ技術、ＣＤ－ＲＯＭ、デジタル多用途ディスク（ＤＶＤ）もしく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もしくは他の磁気記憶装置、または所望の情報を格納するのに使用することができおよびコンピュータ８１０によりアクセスすることができるあらゆる他の媒体が含まれるが、以上には限定されない。

【００９１】

通信媒体は、通常、コンピュータ読取可能な命令、データ構造、プログラムモジュールまたは搬送波もしくは他の移送機構などの変調されたデータ信号の中の他のデータを含み、およびあらゆる情報配信媒体が含まれる。「変調されたデータ信号」という用語は、信号内に情報を符号化するような形で、特性の１つまたは複数が設定または変更された信号を意味する。例として、限定としてではなく、通信媒体には、有線ネットワークまたは直接有線接続などの有線媒体、ならびに音響、ＲＦ、赤外線、および他の無線媒体などの無線媒体が含まれる。以上の媒体のいずれか媒体の組み合わせも、コンピュータ読取可能な媒体の範囲内に含められなければならない。

【００９２】

システムメモリ８３０は、読み出し専用メモリ（ＲＯＭ）８３１およびランダムアクセスメモリ（ＲＡＭ）８３２などの、揮発性および／または不揮発性メモリの形態でコンピュータ記憶媒体を含む。始動中などにコンピュータ８１０内部の要素間で情報を転送するのを助ける基本ルーチンを含む基本入出力システム（ＢＩＯＳ）８３３が、通常、ＲＯＭ８３１の中に格納される。ＲＡＭ８３２は、通常、処理ユニット８２０により即時にアク

セス可能および／または現在処理中のデータおよび／またはプログラムモジュール群を含む。例として、限定としてではなく、図８は、オペレーティングシステム８３４、アプリケーションプログラム群８３５、他のプログラムモジュール群８３６、およびプログラムデータ８３８を例示する。所定の実装において、コンピュータ８１０は、ＰＳＳサーバ２０２である。本シナリオにおいて、アプリケーションプログラム群８３５は、検索語提案モデル２１２を含む。この同一のシナリオでは、プログラムデータ８３８は、マルチ型データオブジェクト２１４、検索結果２３０、抽出された特徴２３２、ＭＤＯベクトル群２３４、強化されたクラスタ群２３６、訓練された分類子２３８、および他のデータ２２６を含む。

【００９３】

コンピュータ８１０は、他の取り外し可能／固定の、揮発性／不揮発性のコンピュータ記憶媒体も含むことが可能である。単に例として、図８は、固定の不揮発性の磁気媒体に対して読み出しまたは書き込みを行うハードディスクドライブ８４１、取り外し可能な不揮発性の磁気ディスク８５２に対して読み出しまたは書き込みを行う磁気ディスクドライブ８５１、およびＣＤ－ＲＯＭまたは他の光媒体などの取り外し可能な不揮発性の光ディスク８５６に対して読み取りまたは書き込みを行う光ディスクドライブ８５５を例示する。例示的動作環境において使用することができる他の取り外し可能な／固定の、揮発性／不揮発性のコンピュータ記憶媒体には、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、固体素子ＲＡＭ、固体素子ＲＯＭなどが含まれるが、以上には限定されない。ハードディスクドライブ８４１は、通常、インタフェース８４０などの固定のメモリインタフェースを介してシステムバス８２１に接続され、ならびに磁気ディスクドライブ８５１および光ディスクドライブ８５５は、通常、インタフェース８５０などの取り外し可能なメモリインタフェースによりシステムバス８２１に接続される。

【００９４】

以上に説明し、図８に例示した駆動装置、および関連するコンピュータ記憶媒体により、コンピュータ読取可能な命令、データ構造、プログラムモジュール、および他のデータのストレージがコンピュータ８１０に提供される。図８では、例えば、ハードディスクドライブ８４１が、オペレーティングシステム８４４、アプリケーションプログラム群８４５、他のプログラムモジュール群８４６、およびプログラムデータ８４８を格納しているのを例示する。これらのコンポーネントは、オペレーティングシステム８３４、アプリケーションプログラム群８３５、他のプログラムモジュール群８３６、およびプログラムデータ８３８と同一であることも、異なることも可能であることに留意されたい。オペレーティングシステム８４４、アプリケーションプログラム群８４５、他のプログラムモジュール群８４６、およびプログラムデータ８４８に、本明細書では、それらが少なくとも異なるコピーであることを例示するために異なる参照符号を付ける。

【００９５】

ユーザは、キーボード８６２および、マウス、トラックボール、またはタッチパッドと一般に呼ばれるポインティング装置８６１などの入力装置群を介して、コマンドおよび情報をコンピュータ８１０に入力することができる。他の入力装置群（図示せず）には、マイクロフォン、ジョイスティック、ゲームパッド、衛星受信アンテナ、スキャナなどを含むことが可能である。以上および他の入力装置群は、システムバス８２１に結合されたユーザ入力インタフェース８６０を介して処理ユニット８２０に接続される場合が多いが、パラレルポート、ゲームポート、またはユニバーサルシリアルバス（ＵＳＢ）などの、他のインタフェースおよびバス構造により接続してもよい。

【００９６】

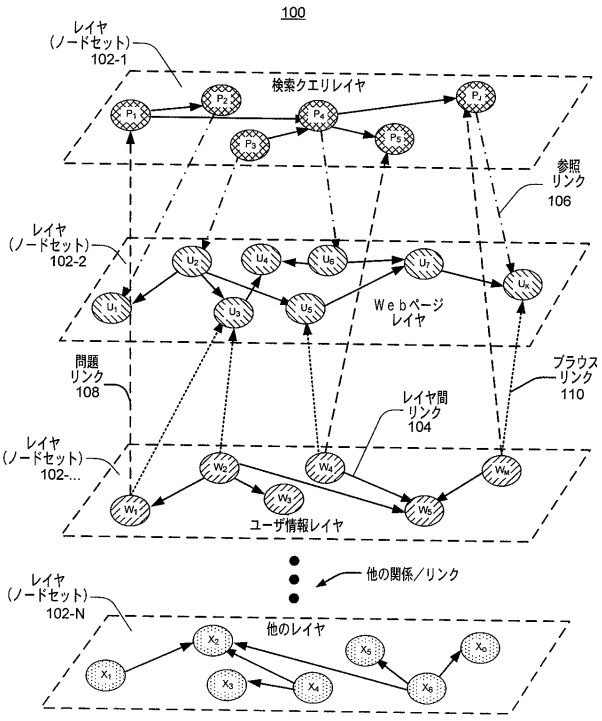
また、モニタ８９１または他の型のディスプレイ装置も、ビデオインタフェース８９０のようなインタフェースを介して、システムバス８２１に接続される。モニタに加えて、コンピュータは、出力周辺インタフェース８９５を介して接続することができるスピーカ８９８やプリンタ８９６などの、他の周辺出力装置群も含むことが可能である。

【００９７】
コンピュータ８１０は、リモートコンピュータ８８０など、１つまたは複数のリモートコンピュータへの論理接続を使用するネットワーク化された環境において動作する。リモートコンピュータ８８０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、ピア装置、または他の共通のネットワークノードであることが可能であり、コンピュータ８８０の特定の実装との相関的要素として、コンピュータ８１０に関係のある上述した要素の多くまたはすべてを含むことが可能であるが、メモリ記憶装置８８１だけを図８に例示している。図８に示す論理接続には、ローカルエリアネットワーク（ＬＡＮ）８８１およびワイドエリアネットワーク（ＷＡＮ）８８３を含むが、他のネットワークも含むことが可能である。そのようなネットワーキング環境は、オフィス、企業全体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で一般的に見られる。
【００９８】
ＬＡＮネットワーキング環境において使用される場合、コンピュータ８１０は、ネットワークインタフェースまたはネットワークアダプタ８８０を介してＬＡＮ８８１に接続される。ＷＡＮネットワーキング環境において使用される場合、コンピュータ８１０は、通常、モデム８８２、またはインターネットなどのＷＡＮ８８３を介して通信を確立するための他の手段を含む。モデム８８２は、内蔵型でも外付け型でもよく、ユーザ入力インタフェース８６０、または他の適切な機構を介してシステムバス８２１に接続することができる。ネットワーク化された環境では、コンピュータ８１０との関係を示したプログラムモジュール群、またはプログラムモジュール群の一部は、リモートメモリ記憶装置の中に格納することができる。例として、限定としてではなく、図８は、リモートアプリケーションプログラム群８８５がメモリ装置８８１上に存在するのを例示する。示したネットワーク接続は例示的であり、コンピュータ間で通信リンクを確立する他の手段も使用することができる。
【００９９】
（結論）
検索語提案のためにマルチ型データオブジェクトの強化されたクラスタリングを行うためのシステムおよび方法を、構造上の特徴、および／または方法上の操作もしくは動作に特有の言い回しで説明してきたが、添付の特許請求の範囲において定義する実装は、説明した特定の特徴または動作に必ずしも限定されないことを理解されたい。例えば、マルチ型データオブジェクトの強化されたクラスタリングを、検索語提案のアプリケーションに関して説明したが、マルチ型データオブジェクトの強化されたクラスタリングは、クラスタリングを利用する他の多くの型のアプリケーションにも適用することができる。したがって、特定の特徴および動作を、請求の対象を実装するための例示的形態として開示する。
【図面の簡単な説明】
【０１００】
【図１】異質なデータオブジェクト／ノードの複数のレイヤ１０２、ならびに関連するレイヤ間およびレイヤ内データオブジェクトリンク／関係を含むマルチレイヤフレームワークグラフ１００を示す図である。
【図２】検索語提案のためにマルチ型データオブジェクトの強化されたクラスタリングを行うための例示的システムを示す図である。
【図３】検索語提案のためにマルチ型データオブジェクトの強化されたクラスタリングを行うための例示的処理手順を示す図である。
【図４】検索語提案のためにマルチ型データオブジェクトの強化されたクラスタリングを行うための図３の例示的処理手順３００の続きを示す図である。
【図５】検索語提案のためにマルチ型データオブジェクトの強化されたクラスタリングを行うための図３および図４の例示的処理手順３００の続きを示す図である。
【図６】図３のブロック３１２の強化されたクラスタリング操作の例示的詳細を示す図である。

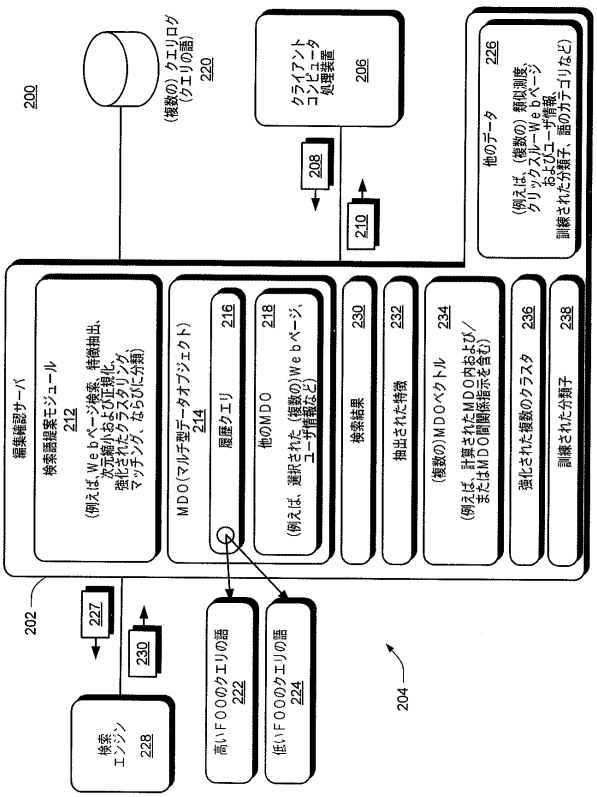
【図７】図３のブロック３１２および図６の強化されたクラスタリング操作の例示的続きを示す図である。

【図８】検索語提案のためにマルチ型データオブジェクトの強化されたクラスタリングを行うための上述のシステム、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例示的コンピュータ処理環境を示す図である。

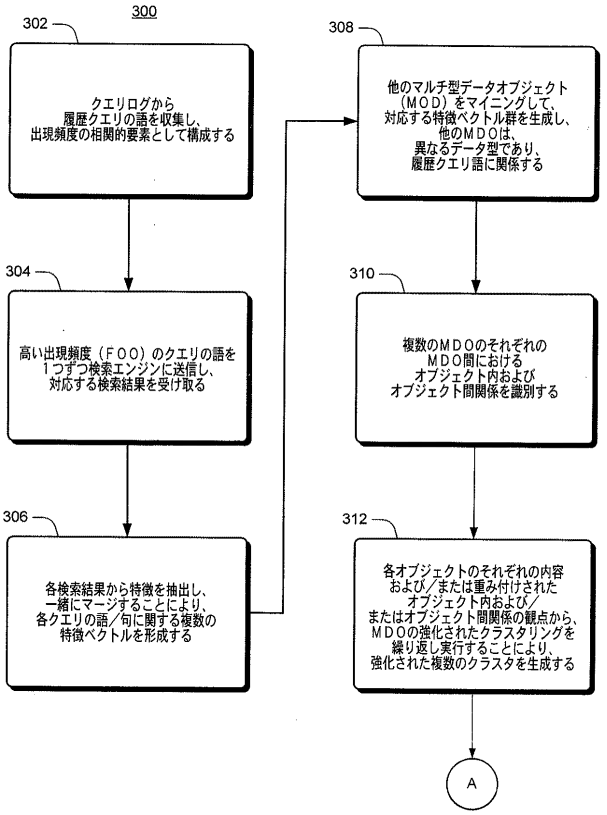
【図１】



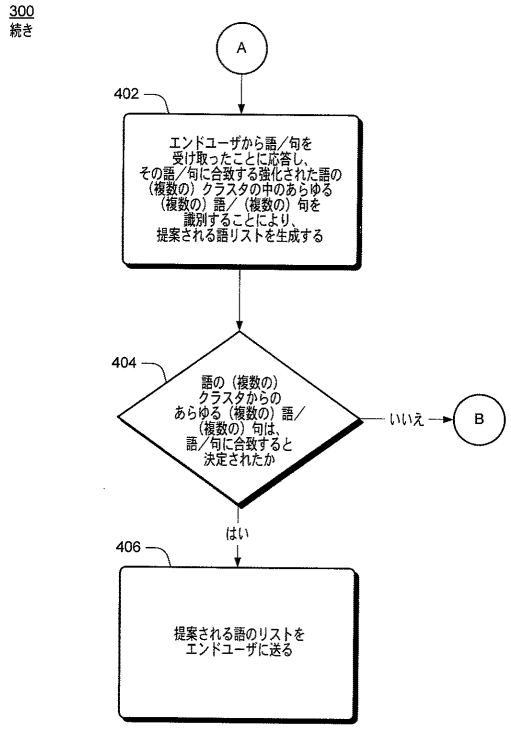
【図２】



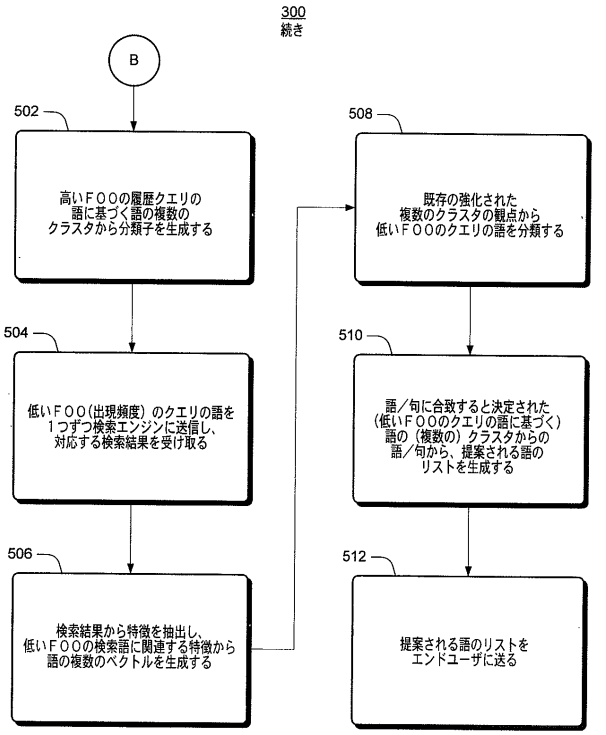
【図３】


300
302
クエリログから
履歴クエリの語を収集し、
出現頻度の相関的要素として構成する
304
高い出現頻度（ＦＯＯ）のクエリの語を
１つずつ検索エンジンに送信し、
対応する検索結果を受け取る
306
各検索結果から特徴を抽出し、
一緒にマージすることにより、
各クエリの語／句に関する複数の
特徴ベクトルを形成する
308
他のマルチ型データオブジェクト
（ＭＯＤ）をマイニングして、
対応する特徴ベクトル群を生成し、
他のＭＤＯは、
異なるデータ型であり、
履歴クエリ語に関係する
310
複数のＭＤＯのそれぞれの
ＭＤＯ間における
オブジェクト内および
オブジェクト間関係を識別する
312
各オブジェクトのそれぞれの内容
および／または重み付けされた
オブジェクト内および／
またはオブジェクト間関係の観点から、
ＭＤＯの強化されたクラスタリングを
繰り返し実行することにより、
強化された複数のクラスタを生成する
A

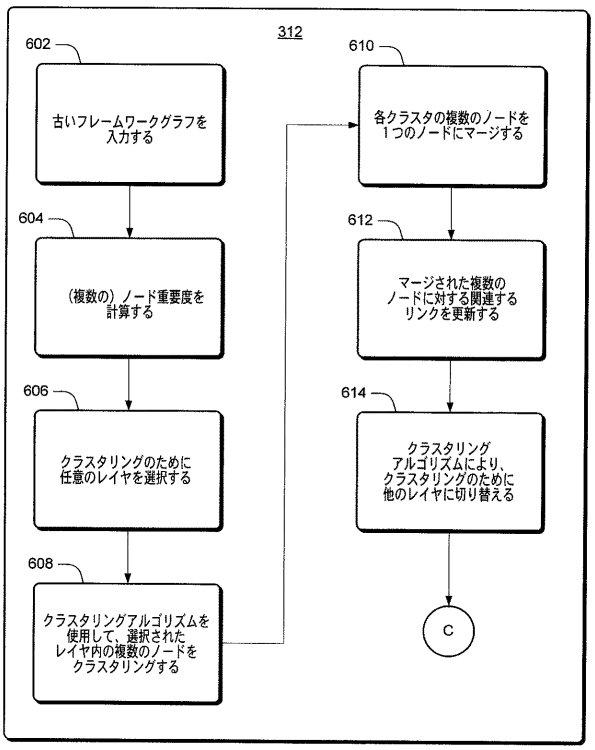
【図４】


300
続き
A
402
エンドユーザから語／句を
受け取ったことに応答し、
その語／句に合致する強化された語の
（複数の）クラスタの中のあらゆる
（複数の）語／（複数の）句を
識別することにより、
提案される語リストを生成する
404
語の（複数の）
クラスタからの
あらゆる（複数の）語／
（複数の）句は、
語／句に合致すると
決定されたか
いいえ
B
はい
406
提案される語のリストを
エンドユーザに送る

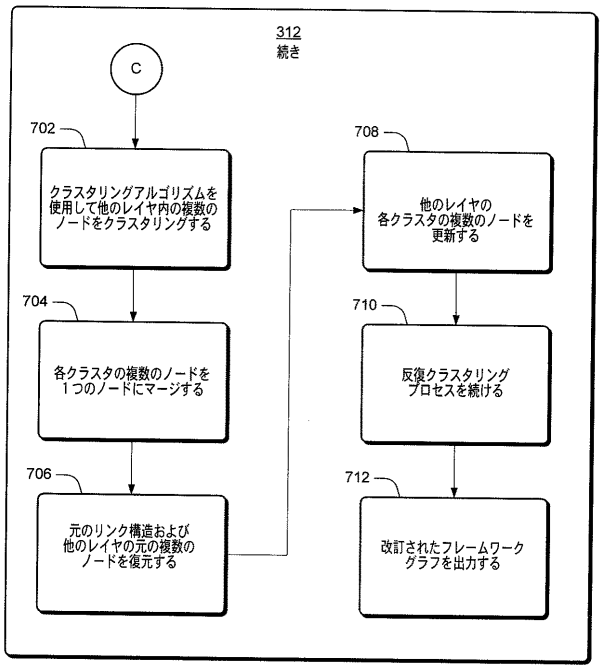
【図５】


300
続き
B
502
高いＦＯＯの履歴クエリの
語に基づく語の複数の
クラスタから分類子を生成する
504
低いＦＯＯ（出現頻度）のクエリの語を
１つずつ検索エンジンに送信し、
対応する検索結果を受け取る
506
検索結果から特徴を抽出し、
低いＦＯＯの検索語に関連する特徴から
語の複数のベクトルを生成する
508
既存の強化された
複数のクラスタの観点から
低いＦＯＯのクエリの語を分類する
510
語／句に合致すると決定された
（低いＦＯＯのクエリの語に基づく）
語の（複数の）クラスタからの
語／句から、提案される語の
リストを生成する
512
提案される語のリストを
エンドユーザに送る

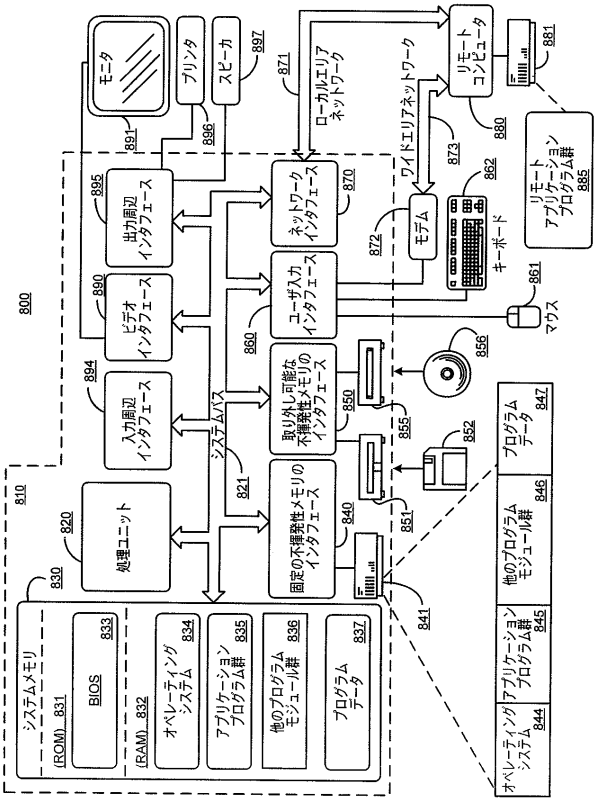
【図６】


312
602
古いフレームワークグラフを
入力する
604
（複数の）ノード重要度を
計算する
606
クラスタリングのために
任意のレイヤを選択する
608
クラスタリングアルゴリズムを
使用して、選択された
レイヤ内の複数のノードを
クラスタリングする
610
各クラスタの複数のノードを
１つのノードにマージする
612
マージされた複数の
ノードに対する関連する
リンクを更新する
614
クラスタリング
アルゴリズムにより、
クラスタリングのために
他のレイヤに切り替える
C

【図７】


312
続き
C
702
クラスタリングアルゴリズムを使用して他のレイヤ内の複数のノードをクラスタリングする
704
各クラスタの複数のノードを1つのノードにマージする
706
元のリンク構造および他のレイヤの元の複数のノードを復元する
708
他のレイヤの各クラスタの複数のノードを更新する
710
反復クラスタリングプロセスを続ける
712
改訂されたフレームワークグラフを出力する

【図８】


800
モニタ
プリンタ
スピーカ
ローカルエリアネットワーク
ワイドエリアネットワーク
リモートコンピュータ
リモートアプリケーションプログラム群
モデム
キーボード
マウス
ネットワークインタフェース
ユーザ入力インタフェース
出力周辺インタフェース
ビデオインタフェース
入力周辺インタフェース
取り外し可能な不揮発性メモリのインタフェース
固定の不揮発性メモリのインタフェース
システムバス
処理ユニット
システムメモリ
(ROM)
(RAM)
BIOS
オペレーティングシステム
アプリケーションプログラム群
他のプログラムモジュール群
プログラムデータ

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 岩間 直純

(56)参考文献 特開２００２－２１５６７４（ＪＰ，Ａ）
BEEFERMAN D, BERGER A, Agglomerative clustering of a search engine query log, Proceedings of the sixth ACM SIGKDD international conference on Knowledge discovery and data mining, 米国, ACM, ２０００年, pp.407-416
WEN J.R, et al., Query clustering using user logs, ACM Transactions on Information Systems (TOIS), 米国, ACM, ２００２年 １月, Vol.20, No.1, pp.59-81

(58)調査した分野(Int.Cl., ＤＢ名)
Ｇ０６Ｆ １７／３０