(19)日本国特許庁(JP)

(12)特許公報(B2)

(11)特許番号
特許第7610573号
(P7610573)

(45)発行日　令和7年1月8日(2025.1.8)

(24)登録日　令和6年12月24日(2024.12.24)

(51)国際特許分類
G06F 16/35 (2025.01)

FI
G06F 16/35

請求項の数　12　（全41頁）

(21)出願番号　特願2022-505355(P2022-505355)
(86)(22)出願日　令和2年8月7日(2020.8.7)
(65)公表番号　特表2022-545335(P2022-545335A)
(43)公表日　令和4年10月27日(2022.10.27)
(86)国際出願番号　PCT/US2020/045390
(87)国際公開番号　WO2021/034517
(87)国際公開日　令和3年2月25日(2021.2.25)
審査請求日　令和5年6月13日(2023.6.13)
(31)優先権主張番号　62/888,998
(32)優先日　令和1年8月19日(2019.8.19)
(33)優先権主張国・地域又は機関
米国(US)
(31)優先権主張番号　16/985,131
(32)優先日　令和2年8月4日(2020.8.4)
最終頁に続く

(73)特許権者　502303739
オラクル・インターナショナル・コーポレイション
アメリカ合衆国 ９４０６５ カリフォルニア州 レッドウッド ショアーズ，メール ストップ ５オーピー７ オラクル パークウェイ ５００
(74)代理人　110001195
弁理士法人深見特許事務所
(72)発明者　マラック，マイケル
アメリカ合衆国、９４０６５ カリフォルニア州、レッドウッド・シティー、オラクル・パークウェイ、５００
(72)発明者　リーバス，ルイス・イー
アメリカ合衆国、９４０６５ カリフォ
最終頁に続く

(54)【発明の名称】　新語分類技術

(57)【特許請求の範囲】
【請求項１】
方法であって、
プロセッサとメモリとを備えるサーバコンピュータによって、第１の文字列を含む入力単語を受信するステップと、
前記サーバコンピュータによって、前記入力単語についての第１の複数のトリグラムを判定するステップと、
前記サーバコンピュータによって、判定された前記第１の複数のトリグラムのそれぞれについての複数の単語埋込みベクトル値を算出するステップと、
前記サーバコンピュータによって、前記複数の単語埋込みベクトル値を行列の複数の行にそれぞれ入力するステップと、
前記サーバコンピュータによって複数の最も近い名前を判定するステップとを含み、前記複数の最も近い名前は、前記入力単語の前記第１の文字列に類似する第２の文字列を有する名前であり、前記方法はさらに、
前記サーバコンピュータによって、前記複数の最も近い名前の各々についての第２の複数のトリグラムを判定するステップと、
前記サーバコンピュータによって、前記入力単語の前記第１の複数のトリグラムと前記複数の最も近い名前の各々の前記第２の複数のトリグラムとに共通するトリグラムを判定するステップと、
前記サーバコンピュータによって、前記複数の最も近い名前のうち前記共通するトリグ

ラムを含む名前の各々の分類に基づいて、前記第１の複数のトリグラムの各々についての複数の分類値を算出するステップと、
　前記サーバコンピュータによって、前記複数の分類値に対応するトリグラムの単語埋込みベクトル値を含む前記行列の行に前記複数の分類値が含まれるように前記行列を更新するステップと、
　前記サーバコンピュータによって、畳み込みニューラルネットワークに、更新された前記行列を入力して、前記畳み込みニューラルネットワークの出力に基づいて前記入力単語の分類を判定するステップとを含む、方法。
【請求項２】
　前記入力単語は、新語である、請求項１に記載の方法。
【請求項３】
前記第１の複数のトリグラムの各々についての前記複数の分類値を算出するステップは、
　前記入力単語から前記第１の複数のトリグラムを識別するステップと、
　前記複数の最も近い名前の各々から前記第２の複数のトリグラムを識別するステップと、
　前記入力単語からの前記第１の複数のトリグラムを、前記複数の最も近い名前の各々からの前記第２の複数のトリグラムと比較するステップと、
　前記入力単語における前記第１の複数のトリグラムと、前記複数の最も近い名前の各々からの前記第２の複数のトリグラムとの間の一致の数および前記複数の最も近い名前のうち一致するトリグラムを含む名前の各々の分類に基づいて、前記第１の複数のトリグラムの各々についての前記複数の分類値を算出するステップとを含む、請求項１または２に記載の方法。
【請求項４】
前記第１の複数のトリグラムの各々についての前記複数の分類値を算出するステップは、
　前記複数の分類値にそれぞれ対応する複数の分類の各々ごとにカウンタを開始するステップと、
　前記入力単語の前記第１の複数のトリグラムから第１のトリグラムを選択するステップと、
　前記複数の最も近い名前の各々の前記第２の複数のトリグラムから前記入力単語の前記第１のトリグラムに一致する１つ以上のトリグラムを判定するステップと、
　前記複数の最も近い名前のうち前記１つ以上のトリグラムの各々を含む名前の各々の分類から前記１つ以上のトリグラムの各々の分類を判定するステップと、
前記第１のトリグラムについて、前記１つ以上のトリグラムの各々の分類に対応する前記カウンタをインクリメントするステップと、
　前記カウンタの値に基づいて、前記第１のトリグラムについての前記複数の分類値を算出するステップとを含む、請求項１～３のいずれか１項に記載の方法。
【請求項５】
　前記複数の単語埋込みベクトル値は、単語埋込みモデルを用いて算出され、
　前記単語埋込みモデルは、前記入力単語の種類に従ってトレーニングされる、請求項１～４のいずれか１項に記載の方法。
【請求項６】
　前記単語埋込みモデルは、トリグラムについての単語埋込みベクトル値を算出するようにトレーニングされる、請求項５に記載の方法。
【請求項７】
　前記第１の複数のトリグラムのうちの１つのトリグラムは、前記入力単語からの連続した順序の３つの字、文字または記号を含む、請求項１～６のいずれか１項に記載の方法。
【請求項８】
　前記サーバコンピュータは、データエンリッチメントシステムの新語分類サーバである、請求項１～７のいずれか１項に記載の方法。
【請求項９】
　前記入力単語は、前記入力単語をエンリッチ化するように構成された前記データエンリ

ッチメントシステムの対話型ユーザインターフェイス上で受信される、請求項８に記載の方法。
【請求項１０】
前記対話型ユーザインターフェイスは、１つ以上の入力単語についての新語分類を実行するための推薦を表示するように構成される、請求項９に記載の方法。
【請求項１１】
サーバコンピュータの１つ以上のプロセッサに請求項１～１０のいずれか１項に記載の方法を実行させるためのコンピュータプログラム。
【請求項１２】
プロセッサと、
メモリと、
前記プロセッサに結合されたコンピュータ可読記憶媒体とを備え、前記コンピュータ可読記憶媒体は、請求項１１に記載のコンピュータプログラムを格納している、サーバコンピュータ。
【発明の詳細な説明】
【技術分野】
【０００１】
関連出願の相互参照
本願は、米国特許法第１１９条（ｅ）に準拠して、２０１９年８月１９日出願の「新語分類技術（NEOLOGISM CLASSIFICATION TECHNIQUES）」と題された米国仮特許出願第６２／８８８，９９８号に基づく優先権を主張するものであって、その内容全体があらゆる目的のために引用により本明細書に援用されている。
【０００２】
背景
本開示は、概して、データの分析および処理に関する。より特定的には、新語の分析および処理のための技術が開示される。
【背景技術】
【０００３】
ビッグデータのユーザは、彼らの顧客に関する大量のデータ（たとえば、データセット）を有し得る。ユーザらは、企業などのデータセットのユーザを含み得る。データセットはスプレッドシートおよびテーブルの形式であり得るとともに、顧客ＩＤ、ファーストネーム（名）、ラストネーム（姓）、住所などの顧客に関する情報を含み得る。ユーザらは、さまざまなソースからのデータセットを有し得る。当該データセットは名前、人口統計および地理的情報などを含み得る。
【０００４】
データセットは、追加の情報が当該データセットから判定され得る場合、ユーザにとってより有用になり得る。たとえば、性別、世代、および他の人口統計学的情報などの顧客に関する追加の情報が容易に入手できない場合、このような追加情報を知ることは有益であるだろう。
【発明の概要】
【発明が解決しようとする課題】
【０００５】
データセットは、これまで見たことのない単語（たとえば、新語）を含み得る。たとえば、顧客の名前は固有の名前または名前の固有のスペルであるかもしれない。さらに、顧客が勤務する企業の名前は、未知の名前であるかもしれないし造語であるかもしれない。これまで見たことのない単語についての追加情報を判定することは困難である。なぜなら、そのような単語について情報を得るための既存の情報（たとえば、属性、特徴など）が存在しないからである。
【０００６】
単語間の類似性を判定するための技術が存在する。たとえば、或る未知語を想定した場

合、いくつかの単語のコーパス内に含まれる最も類似する単語が見出され得る。しかしながら、このような技術では正確な結果が得られない。たとえば、単語のトレーニングコーパス内に含まれる最も類似した単語を用いても正確な結果は得られず、未知語について推定される属性は不正確である。具体的には、いくつかの単語全体を用いても高精度な結果は得られない。

【０００７】

したがって、未知語についての特徴を識別するより正確な方法が必要とされている。例示的な実施形態では、これらおよび他の問題に対処する。

【０００８】

他の実施形態は、本明細書に記載される方法に関連するシステム、デバイス、およびコンピュータ可読媒体を対象とする。以下の詳細な説明および添付の図面を参照することで、例示的な実施形態の性質および利点がより良く理解され得る。

【課題を解決するための手段】

【０００９】

概要

例示的な実施形態は、これまで見たことのない単語または名前（たとえば、新語）の現実世界の特徴に関する予測を行なうように構成される。

【００１０】

テキストの本体（たとえば、データセット、スプレッドシートなど）の場合、これまで見たことのない単語が存在する可能性がある。たとえば、新しい名前または異なるスペルの名前がデータセットに含まれている可能性がある。別の例として、新しい単語が作成されるかまたは新しいバリエーションの単語が作成される可能性がある。これまで見たことのない単語または名前は、その単語または名前に関連する意味および／または特徴が現在知られていないような単語または名前である。これまで見たことのない単語または名前は新語として知られ得る。未知の単語または名前は本明細書では未知語と称する。

【００１１】

例示的な一実施形態は、データセット情報がより有用になり得るようにデータセットをエンリッチ化する。たとえば、或るユーザが大規模なデータセット（たとえば、ビッグデータ）を有する可能性がある。当該データセットは、追加の情報および特徴が当該データセットに基づいて判定され得る場合、より有用になり得る。例示的な一実施形態は、未知語から意味を判定することができ、これにより、当該データセットをより有用なものにすることができる。当該データセットは、データの列および行を含むスプレッドシートの形式であり得る。当該データセットは、１つ以上の新語を含むデータの列を含み得る。たとえば、ファーストネームを対象とするデータセットの列は、これまで見たことのないいくつかの名前を含み得る。

【００１２】

例示的な一実施形態は、未知語に基づいて追加の情報を解読することができる。たとえば、例示的な一実施形態は、未知の名前に基づいて、業界、言語、性別、世代などの特徴を判定することができる。業界、言語、性別、および世代を例として説明しているが、データセット内のデータの種類（たとえば、ファーストネーム、企業名など）に応じてデータセットについての他の特徴および属性を判定することができる。

【００１３】

さらに、例示的な一実施形態は、高精度で未知語に関連する特徴および追加情報を判定することができる。したがって、ユーザは分類に関して推測する必要がない。

【００１４】

特定の名前または単語は既知の特徴を有し得る。たとえば、「Alice」などの名前は女性名として関連付けることができる。具体的には、履歴情報および事前の単語分析を考慮すると、「Alice」などの名前は女性に関連付けられている。しかしながら、これらの名前を「Allys」と綴る人がいるかもしれない。これはそのような名前の主要な具体例であるかもしれず、または、この名前とは何の関連性がない可能性もある。したがって、名前「Allys

」が特定の特徴に関連付けられない可能性がある。名前に加えて、これまで見たことのない単語が存在する可能性もある。新しい単語または単語の新しいスペルが作成される可能性があり、このため、これら新しい単語または単語の新しいスペルは歴史的背景または関連付けを持たない可能性がある。

【００１５】

例示的な一実施形態は、単語または名前がどの言語に由来しているかを予測し、名前（たとえば、企業名）に基づいて業界を予測し、人の名前に基づいて性別を予測することができる。性別を予測することは、市場細分化の分析時の集計に有用であり得る。

【００１６】

例示的な一実施形態は、未知語のスペルに基づいて意味を推定することができる。意味はトリグラムを用いて推定することができ、トリグラムは、未知語に関する情報を提供するのに用いることができる。具体的には、３文字のトリグラムを言語として扱うことによって未知語の意味を判定することができる。すなわち、意味は、或る文章内の語順が意味を持つのと同様に、単語内のトリグラムの順序に基づいて推定することができる。たとえば、トリグラムおよび履歴データに基づいて、母音で終わる名前が女性に関連付けられることが多いと判定することができる。

【００１７】

例示的な一実施形態は、ニューラルネットワーク（たとえば、Word2Vec）を介してトレーニングセットの単語から抽出されたトリグラムを実行する。さらに、未知語の属性および特徴についての予測を行なうために教師あり学習を実行することができる。上述したように、未知語の属性および特徴、すなわち、教師あり学習についてのラベルは、単語が由来する言語、業界、性別、世代などを含み得る。未知語の属性および特徴は、未知語について判定することができる追加の情報を含む。したがって、ユーザは、単に、たとえばファーストネームのデータセットを有するだけではなく、その顧客の性別、その顧客が勤めている業界などの追加情報（たとえば、これらのデータ項目についてのラベル）を有する。これにより、データセットはユーザにとってより有用になる。

【００１８】

図面の簡単な説明

本開示は、同様の参照番号が同様の要素を示している添付の図面に関連付けて以下の詳細な説明によって容易に理解されるだろう。

【図面の簡単な説明】

【００１９】

【図１】いくつかの例示的な実施形態に従った分析環境を示すブロック図である。

【図２】いくつかの例示的な実施形態に従った、データエンリッチメントシステムの新語分類サーバを示すブロック図である。

【図３】いくつかの例示的な実施形態に従った、新語を分類するための方法のフローチャートである。

【図４】いくつかの例示的な実施形態に従った、新語を分類するための変数の概要である。

【図５】いくつかの例示的な実施形態に従った、入力単語についてのトリグラムの判定を示す図である。

【図６】いくつかの例示的な実施形態に従った、トリグラム単語埋込みモデルベクトルを伴う行列を示す図である。

【図７】いくつかの例示的な実施形態に従った、単語についてのベクトルを算出するためのプロセスを示す図である。

【図８】いくつかの例示的な実施形態に従った、入力単語について最も近い名前の判定を示す図である。

【図９】いくつかの例示的な実施形態に従った、キューのリストを表わすテーブルである。

【図１０】いくつかの例示的な実施形態に従った、分類値を算出するための方法のフローチャートである。

【図１１】いくつかの例示的な実施形態に従った、分類値を算出するための方法の詳細な

フローチャートである。
【図１２】いくつかの例示的な実施形態に従った、分類値を含む行列を示す図である。
【図１３】いくつかの例示的な実施形態に従った畳み込みニューラルネットワークを示す図である。
【図１４】いくつかの例示的な実施形態に従った、新語分類を実行するためのユーザインターフェイスを示す図である。
【図１５】いくつかの例示的な実施形態に従った分散型システムを示す簡略図である。
【図１６】いくつかの例示的な実施形態に従った、サービスがクラウドサービスとして提供され得るシステム環境の１つ以上のコンポーネントを示す簡略ブロック図である。
【図１７】いくつかの例示的な実施形態に従った、例示的な一実施形態を実現するために用いられ得る、例示的なコンピュータシステムを示す図である。
【発明を実施するための形態】
【００２０】
詳細な説明
以下の説明では、説明の目的で、本発明のさまざまな実施形態を十分に理解できるようにするために多数の具体的な詳細が述べられる。しかしながら、当業者にとっては、これらの具体的な詳細のうちいくつかがなくても本発明の実施形態が実施され得ることが明らかであるだろう。他の例では、周知の構造およびデバイスをブロック図の形態で示す。
【００２１】
新語は新しい単語または語句であり得る。新語は、新しく考案された単語または通常使用されない単語であり得る。新語は文化および技術の変化によって推進され得る。さらに、新語はソーシャルメディアおよび大衆文化によって推進され得る。新しい単語および名前ならびに単語および名前のスペルは絶えず作成されている。例示的な一実施形態は、新語と関連付けることのできる属性または特徴を判定することができる。
【００２２】
本明細書中では、ユーザは企業などのデータのユーザである。しかしながら、ユーザは、新語分類が有益となるであろういずれの人またはエンティティをも含み得る。データが追加情報を含む場合、当該データはユーザにとってより有意義なものになり得る。データを充実させることにより、当該データはユーザにとってより有用になる。データは、たとえば、顧客データ（たとえば、ファーストネーム、ラストネーム、ＩＤ、社会保障番号など）および購入データ（たとえば、店舗名、購入したアイテムなど）を含み得る。
【００２３】
例示的な実施形態はさまざまな状況で用いることができる。たとえば、どの言語に属するのかが不明確である所与の単語が存在する可能性もある。別の例として、或る企業の名前などの単語を想定した場合、その企業に関連付けられる業界が識別される。別の例として、例示的な実施形態は、たとえばファーストネームを考慮することで性別を予測するのに用いることができる。例示的な実施形態は、顧客のファーストネームに基づいて顧客の年齢または世代を予測するために用いることができる。これらは、追加の特徴が識別され得ることによりデータがユーザにとってより有益なものになるという例示的な状況に過ぎない。
【００２４】
例示的な実施形態は、たとえば市場細分化に有用であり得る。市場細分化は、いくつかの種類の共有の特徴に基づいて顧客などの広範な消費者またはビジネスの市場を消費者の下位グループに分割することを含み得る。
【００２５】
例示的な一実施形態は単語を分類するための解決策を提供する。たとえば、参照リストは、ファーストネームと、典型的にはファーストネームによって表わされる性別とからなるリストを含み得る。たとえば、名前「David」は、性別「男性」に関連付けることができる。関連付けられた性別は、履歴データに基づいた関連付けの頻度に基づき得る。
【００２６】

しかしながら、新しい名前および新しいスペルの名前が作成される可能性がある。例示的な一実施形態は、これまで見たことのない名前についての性別を提供することができる。すなわち、名前または単語の辞書において識別されなかった名前についての性別である。たとえば、「Chauna」という名前は名前の辞書にない可能性がある。したがって、当該名前が男性に関連付けられるのかまたは女性に関連付けられるのかが不明である。「Chauna」という名前の性別を判定するために、バイナリ分類を実行することができる。すなわち、２つの分類グループを識別することができる。個人に関する予測を行なうことが倫理的な問題を伴う可能性がある使用事例の場合、この技術は、市場細分化の識別などのために母集団が集約される使用事例において有用であり得る。しかしながら、これは例示的な実施形態であり、所望の分類に基づいて追加の変更を行なうことができる。例示的な実施形態は２つの分類に限定されない。ユーザのニーズに基づいて追加の分類を行なうことができる。

【００２７】

Ｉ．分析環境

図１は、いくつかの例示的な実施形態に従った分析環境１００のブロック図を示す。

【００２８】

データ分析環境１００は、分析システム１１０、データエンリッチメントシステム１２０、データソース１３０、クライアント１４０、クライアントデータストア１４１、分散型ストレージシステム１５０、データターゲット１６０、および知識ソース１７０を含み得る。データエンリッチメントシステム１２０は、ユーザインターフェイス１２１、インジェストサーバ１２２、準備サーバ１２３、変換サーバ１２４、推薦サーバ１２５、感情分析サーバ１２６、パブリッシュサーバ１２７、知識サーバ１２８、エンリッチサーバ１２９、および新語分類サーバ２００を含み得る。新語分類サーバ２００は図２に関連付けてより詳細に説明される。

【００２９】

分析環境１００はクラウドベースの環境であり得る。分析システム１１０は、自己サービス可視化、強力なインラインデータ準備、事業報告、高度な分析、および事前の洞察を提供する自己学習分析を含む単一の統合型プラットフォームを提供する。分析システム１１０は、たとえば、Oracle Analytics Cloudを含み得る。

【００３０】

クライアントまたはユーザ１４０は、分析システム１１０のデータエンリッチメントシステム１２０にデータエンリッチメント要求を提出し得る。クライアントは、クライアントに関連付けられたデータを格納するためのクライアントデータストア１４１を含み得る。データエンリッチメントシステム１２０は、データソース１３０の１つ以上（またはその部分、たとえば、特定のテーブル、データセットなど）を識別することができる。次いで、データエンリッチメントシステム１２０は、識別されたデータソース１３０に対してデータを処理するよう要求し得る。

【００３１】

いくつかの実施形態では、データソースがサンプリングされ得るとともに、サンプリングされたデータがエンリッチ化のために分析され得ることで、大規模なデータセットがより管理し易くなり得る。識別されたデータは、受信され得るとともに、データエンリッチメントサービスにアクセス可能な分散型ストレージシステム（Hadoop（登録商標）分散型ストレージ（Hadoop Distributed Storage（ＨＤＦＳ））システムなど）に追加され得る。データは、いくつかの処理段階（本明細書ではパイプラインまたはセマンティックパイプラインとして説明される）によって意味論的に処理され得る。これらの処理段階は、準備サーバ１２３を介する準備段階、パブリッシュサーバ１２７を介するパブリッシュ段階、エンリッチサーバ１２９を介するエンリッチ段階を含み得る。

【００３２】

いくつかの実施形態では、準備段階は、さまざまな処理下位段階を含み得る。これは、データソースフォーマットを自動的に検出することと、コンテンツ抽出および／または修

復を実行することとを含み得る。データソースフォーマットが識別されると、データソースは、データエンリッチメントサービスによって処理可能なフォーマットに自動的に正規化され得る。いくつかの実施形態では、データソースは、準備されると、エンリッチサーバ１２９によって処理され得る。いくつかの実施形態では、インバウンドデータソースは、（たとえば、データエンリッチメントサービスに通信可能に結合されたＨＤＦＳシステムなどの）データエンリッチメントシステム１２０にアクセス可能な分散型ストレージシステム１５０にロードされ得る。

【００３３】

分散型ストレージシステム１５０は、インジェストされたデータファイルのための一時的な格納空間を提供するとともに、中間処理ファイルを格納し得るとともに公開前の結果を一時的に格納し得る。いくつかの実施形態では、強化またはエンリッチ化された結果を分散型ストレージシステムに格納することもできる。いくつかの実施形態では、インジェストされたデータソースに関連付けられた、エンリッチ中に捕捉されたメタデータは、分散型ストレージシステム１５０に格納され得る。（たとえば、データソースの位置、結果、処理履歴、ユーザセッション、実行履歴、および構成などを示す）システムレベルメタデータは、分散型ストレージシステム内に、または、データエンリッチメントサービスにアクセス可能な別個のリポジトリ内に格納することができる。

【００３４】

いくつかの実施形態では、データエンリッチメントシステム１２０は、感情分析サーバ１２６を介して感情分析を行ない得る。感情分析サーバ１２６は、本明細書に開示される技術を用いて、さまざまなデータソースからのデータの感情を分析するための機能を含む。技術は、畳み込みニューラルネットワーク（convolutional neural network：ＣＮＮ）、語彙共起ネットワーク、および分析の精度を向上させるために感情分析を実行するためのバイグラム単語ベクトルの適用を含む。

【００３５】

いくつかの実施形態では、パブリッシュサーバ１２７を介するパブリッシュ段階は、エンリッチ化中およびいずれかのデータソースエンリッチ化または修復中に捕捉されたデータソースメタデータを分析のために１つ以上の視覚化システムに提供する（たとえば、推薦されたデータの変換、エンリッチ化および／または他の変更をユーザに表示する）ことができる。パブリッシュサブシステムは、処理されたデータを１つ以上のデータターゲットに配信することができる。データターゲットは、処理されたデータを送信することができる場所に対応し得る。この場所は、たとえば、メモリ、コンピューティングシステム、データベース、またはサービスを提供するシステム内の位置であり得る。たとえば、データターゲットは、Oracle Storage Cloud Service（ＯＳＣＳ）、ＵＲＬ、第三者ストレージサービス、ウェブサービス、および、他のクラウドサービス、たとえば、Oracle Business Intelligence（ＢＩ）、サービスとしてのデータベース（Database as a Service）、およびサービスとしてのデータベーススキーマ（Database Schema as a Service）などを含み得る。いくつかの実施形態では、シンジケートエンジンは、結果を閲覧し、選択し、サブスクライブするためのＡＰＩのセットを顧客に提供する。サブスクライブされて新しい結果が生成されると、結果データは、外部のウェブサービスエンドポイントへの直接フィードとして、または大容量ファイルダウンロードとして提供され得る。

【００３６】

以下でさらに説明するように、データは、未知語または新語についての追加の関連情報を含むようにエンリッチ化することができる。新語分類サーバ２００は、新語を分析して、新語についての属性または特徴を判定することができる。データは、データソース１３０から取得することができ、新語分類サーバ２００は、新語についての属性および／または特徴を判定することができる。属性は、特定の単語または名前を有する人々または物に特徴的であるかまたは固有である品質または特徴を含み得る。

【００３７】

例示的な一実施形態は、ユーザがすべてのデータを分析しなくてもすむように、機械学

習を用いるデータエンリッチメントシステムを提供する。さらに、データエンリッチメントシステムは、ユーザの介入なしに判定を行なうことができる。したがって、データエンリッチメントシステムはユーザのために分析を実行することができる。

【００３８】

本開示を通じて、いくつかの実施形態に従ったプロセスを示すさまざまなフローチャートおよび技術が開示される。個々の実施形態は、フローチャート、フロー図、データフロー図、構造図、またはブロック図として示されるプロセスとして説明され得る。フローチャートは動作を連続したプロセスとして説明する可能性もあるが、動作の多くは並列にまたは同時に実行されてもよい。加えて、動作の順序は並べ替えられてもよい。プロセスは、その動作が完了したときに終了するが、図に含まれない追加のステップを含む可能性もある。プロセスは方法、関数、手順、サブルーチン、サブプログラムなどに対応し得る。プロセスが関数に対応する場合、その終了は、呼出し関数またはメイン関数への関数の戻りに対応し得る。

【００３９】

図に示されるプロセスは、１つ以上の処理ユニット（たとえば、プロセッサコア）によって実行されるソフトウェア（たとえば、コード、命令、プログラム）、ハードウェア、またはそれらの組合せで実現され得る。たとえば、データエンリッチメントシステム１２０は、図のいずれかを参照して説明されるプロセスのためのコンピュータシステムによって実現することができる。プロセスはいずれもサービスとして実現され得る。いくつかの実施形態では、図中の要素はいずれも、図に示されるよりも多いかまたは少ないサブシステムおよび／もしくはモジュールで実現されてもよく、２つ以上のサブシステムおよび／もしくはモジュールを組合わせてもよく、または、異なる構成もしくは配置サブシステムおよび／もしくはモジュールを有していてもよい。サブシステムおよびモジュールは、ソフトウェア（たとえば、プログラムコード、プロセッサによって実行可能な命令）、ファームウェア、ハードウェア、またはそれらの組合せで実現され得る。いくつかの実施形態では、ソフトウェアは、メモリ（たとえば、非一時的なコンピュータ可読媒体）、メモリデバイス、または他の何らかの物理メモリに格納され得るとともに、１つ以上の処理ユニット（たとえば、１つ以上のプロセッサ、１つ以上のプロセッサコア、１つ以上のＧＰＵなど）によって実行され得る。

【００４０】

図中の特定の一連の処理ステップは限定を意図するものではない。他の一連のステップも、代替的な実施形態に従って実行され得る。たとえば、代替的な実施形態は、上記で概説されるステップを異なる順序で実行してもよい。さらに、図に示される個々のステップは、当該個々のステップに適したさまざまな順序で実行され得る複数のサブステップを含み得る。さらに、特定の用途に応じて付加的なステップが追加または削除されてもよい。当業者であれば、多くの変形例、変更例および代替例を認識するだろう。

【００４１】

いくつかの実施形態では、データは、１つ以上のデータ構造を用いて格納され得る。データ構造は、データがどのように、何に、および／またはどこに格納されるのかに応じて、さまざまな方法で編成され得る。特定のデータを含むデータ構造の各々が示されているが、データを格納するために実現されるデータ構造はより多くてもよく、またはより少なくてもよい。データ構造は他のデータ構造への参照を含み得る。データ構造は、連結リスト、アレイ、ハッシュテーブル、マップ、レコード、グラフ、または他の種類のデータ構造を含むがこれらに限定されない１つ以上の種類のデータ構造を用いて実現され得る。データ構造は階層的に実現されてもよい。データ構造の各々は、ユーザによる入力に基づいて宣言的に定義され得る。データ構造は、テンプレートに基づいて、たとえば、拡張マークアップ言語（Extended Markup Language：ＸＭＬ）などのマークアップ言語に基づいて定義されるテンプレートに基づいて、定義することができる。データ構造は、文書フォーマットとも称される１つ以上のフォーマットを有し得る。

【００４２】

ＩＩ．新語分類サーバ

図２は、いくつかの例示的な実施形態に従った、データエンリッチメントシステム１２０の新語分類サーバ２００のブロック図を示す。

【００４３】

新語分類サーバ２００は、新語分類システム２１０、トリグラム分析システム２１１、畳み込みニューラルネットワーク（ＣＮＮ）２１２、単語埋込みモデル２１３、データストア２１４、および、辞書２１５を含み得る。

【００４４】

新語分類システム２１０は、未知語または新語のエンリッチ化を実行することができる。新語分類システム２１０は、トリグラム分析システム２１１、畳み込みニューラルネットワーク（ＣＮＮ）２１２、単語埋込みモデル２１３、データストア２１４、および辞書２１５から受信した情報に基づいて、新語についての分類を提供することができる。

【００４５】

トリグラム分析システム２１１は、トリグラムを用いて新語についての分析を実行することができる。トリグラムは、文字、音節、または単語などの３つの連続した書き言葉単位のグループである。例示的な実施形態では、トリグラムは３つの連続する文字である。トリグラム分析システムは所与の単語についてのトリグラムを生成する。生成されたトリグラムは重複する文字を含む。たとえば、各トリグラムの２つの文字が重複し得る。所与の単語に関して生成される各トリグラムは、前のトリグラムの右側にある１文字で始まる。たとえば、名前「ＢＯＢ」の場合、トリグラムは「＾ＢＯ」、「ＢＯＢ」、および「ＯＢ＄」を含むだろう。トリグラム分析は図５に関連付けてより詳細に説明される。

【００４６】

畳み込みニューラルネットワーク（ＣＮＮ）２１２は、ディープニューラルネットワークのクラスである。ＣＮＮは値に重み付けを加える。ＣＮＮは、たとえば、数分程度で名前のセットを分析するようにトレーニングされ得る。畳み込みニューラルネットワークは図１３に関連付けてより詳細に説明される。

【００４７】

単語埋込みモデル２１３は、単語埋込みベクトルを生成するために使用可能なモデルである。単語埋込みは、語彙からの単語または語句が実数のベクトルにマッピングされる自然言語処理（natural language processing：ＮＬＰ）を含み得る。例示的な一実施形態において用いることができる例示的な単語埋込みモデルはWord2Vecである。Word2Vecは、あらゆる目的のために引用により本明細書に援用されている、Mikolovらによる「機械翻訳のための言語間の類似点の活用（Exploiting Similarities among Languages for Machine Translation）」（２０１３）（http://arxiv.org/pdf/1309.4168.pdf）に開示される技術を用いて実現され得る。

【００４８】

単語埋込みモデルは、語彙のコーパスが供給されるアルゴリズムを含み得る。（たとえば、記事、新聞からの）大量のテキストが取得されて、単語埋込みモデルのアルゴリズムに供給される。単語埋込みモデルベクトルは、重み変数（たとえば、トリグラムに対する３００のアスペクト）を含む。重み変数は分析されているトリグラムに次元を追加する。

【００４９】

名前の場合、コーパス単語は国勢調査局からの名前を含み得る。単語埋込みモデルは、国勢調査局からの名前のトリグラムを用いてトレーニングすることができる。単語埋込みモデルは図７に関連付けてより詳細に説明される。

【００５０】

データストア２１４は、新語分類が実行されるべきデータセットからのデータを格納するために用いられ得る。たとえば、クライアントは、データの列および行を含むデータのスプレッドシートを提供することができる。新語分類サーバ２００によって分析されるべきデータはデータストア２１４に格納され得る。データストア２１４はまた、カウンタ値と、単語埋込みモデルベクトル値および分類値を含む行列とを格納し得る。

【００５１】
辞書２１５は、ディープラーニング（深層学習）ライブラリ（たとえば、DeepLearning for JVM, Tiny Neural Network（ＴＩＮＮ）など）を含み得る。ディープラーニングライブラリは、新語分類サーバ２００のためのコードを生成するために使用可能な機能およびモジュールのライブラリを提供し得る。ディープラーニングライブラリは、新語分類サーバ２００のための機械学習を促進するのに役立ち得る。
【００５２】
ＩＩＩ．新語分類を実行するための方法の概要
例示的な一実施形態は新語分類を対象とする。分類は、ロジスティック回帰を用いて実行することができる。ロジスティック回帰は、ロジスティック関数を用いて変数をモデリングする統計モデルを用いることを含み得る。記載される例では、バイナリ分類（たとえば、２つの分類）が用いられるが、例示的な実施形態はバイナリ分類に限定されない。たとえば、世代識別は、３つ以上の分類（たとえば、サイレント、ブーマー、Ｘ、ミレニアル、Ｚなど）を含み得る。分類は、起こり得るいくつかの属性または特徴に基づき得る。いくつかの分類は、ユーザによって入力されるパラメータに基づき得る。
【００５３】
図３は、いくつかの例示的な実施形態に従った、新語を分類するための方法３００のフローチャートを示す。図３は、図４に示される変数に関連付けて説明される。
【００５４】
図４は、いくつかの例示的な実施形態に従った、新語を分類するための変数４００の概要である。図４に示される変数は、入力変数または入力値であり得る。変数「ｗ」は、新語または未知語または未知の名前である入力単語を表わし、変数「Ｍ」は、単語埋込みモデルベクトルおよび分類値がポピュレートされる行列を表わし、変数「ｋ」は、いくつかの既知の最近傍の入力値を表わし、変数「Ｓ」は、既知の最も近い名前を含むセットを表わし、セットＳ内の名前の数は入力値ｋ内の数に対応することとなり、変数ｗｑは入力単語「ｗ」についての識別されたトリグラムを表わし、変数「Ｑ」はＮ個のキューのリストを表わしている。この場合、各キューは、入力単語ｗと既知の最近傍のセットＳとの間の最長共通部分列（longest common subsequence：ＬＣＳ）に従って初期化される。Ｎ個のキューのリスト内のＮはキューの数を表わす。各キューは、式ＬＣＳ（ｗ，Ｓ［ｉ］）に従って初期化することができる。図４におけるＮの値は、３つの最近傍が存在するので３となる。変数およびそれらの値は以下により詳細に説明される。
【００５５】
ステップ３１０において、初期入力が受信される。初期入力は、新語として、または未知語もしくは未知の名前として知ることができる。新語は変数「ｗ」によって表わされる。初期入力は、新語の属性を分類するように新語分類サーバをトレーニングするために、ユーザによって選択することができる。例示を目的として、１つの入力単語が受信される。しかしながら、例示的な一実施形態では、データの列（たとえば、スプレッドシート内のデータの列）内の複数の単語などの複数の入力を受信することができる。
【００５６】
初期入力はデータエンリッチメントシステムのユーザインターフェイス上で受信され得る。未知語は、ユーザによって手動で入力可能である。たとえば、新語は、データエンリッチメントシステムのユーザインターフェイスを介して入力することができる。代替的には、未知語は、新語分類サーバによって自動的に入力され得る。新語分類サーバは、データエンリッチメントシステムによって生成される１つ以上の未知語を取得することができるか、または、データソースから１つ以上の未知語を取得することができる。
【００５７】
図４に示す例では、入力単語ｗは「ＪＯＡＮＮＡ」である。しかしながら、これは単に説明を目的としたものであり、分析のために如何なる名前または単語を用いることもできる。例示および説明の容易さを目的として、分類がどのように実行されるかを実証するために既知の名前が用いられる。例示的な実施形態はまた、既知の名前についての属性を識

別するために用いられ得る。
【００５８】
例示的な実施形態は、新語または未知語もしくは未知の名前について高精度の分類結果を提供することができる。例示的な実施形態は、これまで見たことのない名前または存在しない名前についての属性をユーザが名前のライブラリであると判定するのを支援するのに有益である。名前および単語のライブラリはデータストア２１４に格納することができる。名前について説明がなされているが、例示的な実施形態は如何なる単語にも適用されるものであって、名前に限定されない。さらに、単一の単語について説明がなされているが、例示的な実施形態は、単語のグループ、１つの表現または１つの句に適用され得る。
【００５９】
ステップ３２０において、入力単語についてトリグラムが判定される。識別されたトリグラムは変数ｗｑによって表わすことができる。トリグラムは、文字の順序が意味を有するので、文字が順序通りに識別される。トリグラムの順序は追加の意味を保持する。したがって、トリグラムは、入力単語に基づいた順序に並んだ３文字分に関するものである。トリグラムはトリグラム分析システム２１１によって判定することができる。
【００６０】
図５は、いくつかの例示的な実施形態に従った、入力単語についてのトリグラム５００を示す。図５に示されるように、６つのトリグラムが、ステップ３１０において入力された単語について識別される。トリグラムは、文字、音節、記号または単語などの３つの連続した書き言葉単位のグループである。図４に示す例では、トリグラムは文字および記号のグループである。入力単語「ＪＯＡＮＮＡ」について識別される６つのトリグラムは、「＾ＪＯ、ＪＯＡ、ＯＡＮ、ＡＮＮ、ＮＮＡ、ＮＡ＄」を含む。識別されるトリグラムは、単語「ＪＯＡＮＮＡ」において連続した順序で見出される。記号「＾」は文字列の先頭を表わし、「＄」は文字列の末尾を表わす。すなわち、「文字列の先頭」および「文字列の末尾」では、ゴースト文字（たとえば、＾および＄）が、単語のトリグラムへの分解に関与している。トリグラムの生成時、２つの文字が前のトリグラムから重複している。
【００６１】
ステップ３３０において、当該トリグラムについて単語埋込みベクトル値が算出される。各トリグラムｗｑごとに、単語埋込みベクトル値がトレーニングされる。単語埋込みベクトル値は、Word2Vecなどの単語埋込みモデルを用いて算出され得る。単語埋込みモデルは、特に分析されている単語の種類に関してトレーニングされる。したがって、説明される例では、単語埋込みモデルは、具体的にはファーストネームに関して、より具体的には、ファーストネームのトリグラムに関してトレーニングされる。
【００６２】
単語埋込みモデルは、いくつかの単語全体ではなく単語のトリグラムを分析するようにトレーニングされる。したがって、例示的な実施形態に従った単語埋込みモデルはトリグラムについてトレーニングされる。トリグラムが言語を形成しているので、トレーニングはトリグラムに対して実行される。単語内のトリグラムは、或る文内の単語が言語を形成するのと同様の態様で言語を形成する。文字列内の文字およびトリグラムの位置および順序は、入力単語の分類に影響を及ぼす可能性がある。たとえば、トリグラム分析に基づいて、女性名の方が男性名よりも母音で終わることが多いと判定することができる。
【００６３】
説明される例では、単語の種類はファーストネームである。単語の他の種類またはカテゴリはラストネームまたはビジネスネームなどを含み得る。しかしながら、これらは単なる例に過ぎず、さまざまな種類の名前または単語を用いることができる。単語埋込みモデルは、特に、分析されている単語の種類に対してトレーニングされるものであるが、一般的な単語に対してトレーニングされる一般的な単語埋込みモデルを使用しない。具体的には、一般的な単語埋込みモデルを用いても、特に、分析されている単語の種類に対してトレーニングされた単語埋込みモデルを用いるほど正確な結果は得られないだろう。
【００６４】

ステップ３４０において、算出されたベクトル値が行列Ｍに入力される。すなわち、行列Ｍは、トリグラムｗｑに関連付けられたベクトル値を含むようにポピュレートされる。行列Ｍへのベクトル値のポピュレートはステップ３４０で実行されるが、行列Ｍにおけるベクトル値のポピュレートは、畳み込みニューラルネットワークを介して行列を実行する前よりも後の時点で実行され得る。

【００６５】

図６は、いくつかの例示的な実施形態に従った、トリグラム単語埋込みモデルベクトルを伴う行列Ｍ６００を示す。

【００６６】

図６に示すように、行列Ｍ６００は、ステップ３２０で生成されたトリグラム６１０を含む。したがって、行列Ｍ６００は、入力単語「ＪＯＡＮＮＡ」に関して識別された６つのトリグラムについての６つの行を含む。行列Ｍ６００についての行の数は、新語について判定されるトリグラムの数に基づいて変化するだろう。行列Ｍ６００は単語埋込みモデルベクトル列６２０を含む。図示の例では、３００列の単語埋込みモデルベクトルが用いられる。行列Ｍにはトリグラム単語埋込みモデルベクトル６２０がポピュレートされる。したがって、単語埋込みモデルは３００の次元を有する。３００は、過小適合でも過剰適合でもない単語埋込みモデルについての数である。３００は、より多くのベクトルまたはより少ないベクトルがユーザのニーズに基づいて使用可能となる例において用いられる。

【００６７】

行列Ｍ６００はまた、分類列６３０を含む。この例では、２つの分類（たとえば、男性および女性）が判定されることとなる。分類列の数は、起こり得る分類の数またはユーザが所望する分類の数に依存するだろう。たとえば、世代識別では、起こり得る５つの異なる世代分類（たとえば、サイレント、ブーマー、Ｘ、ミレニアル、Ｚなど）に対応する５つの分類列が存在し得る。分類の種類は、分析されているデータに基づいて、またはユーザが所望する分類情報に基づいて、異なり得る。分類はまた、ラベルとしても知られ得る。

【００６８】

図６に示されるように、第１の分類列（たとえば、男性）についての第１の列６３１と、第２の分類（たとえば、女性）についての第２の列６３２とが存在する。図６には２つの分類が示されているが、例示的な実施形態は２つの分類に限定されない。したがって、分類列の数は、ユーザが判定を所望する分類に基づいて異なり得る。例示的な一実施形態は、行列Ｍの分類列を埋めるための技術を提供する。

【００６９】

行列Ｍが作成される箇所では、行が単語埋込みモデルベクトルであり、それらが現われる順序で次々に配置されて、行列の縦軸を時間次元の形式にする。単語埋込みモデルでモデリングされる「単語」は、感情分析技術で行われるように、単語ではなく（３文字の）トリグラムである。感情分析技術は、教師なしのニューラル言語モデルから得られた単語ベクトルに加えて１つの畳み込みの層でＣＮＮをトレーニングすることを含み得る。畳み込みニューラルネットワークは、単語埋込みモデル（たとえば、Word2Vec）の上に構築することができる。単語埋込みモデルは、入力トレーニングデータ（たとえば、ファーストネームの辞書）を用いてトレーニングされる。感情分析は、あらゆる目的のために引用により本明細書に援用されている、Ｋｉｍによる「文章分類のための畳み込みニューラルネットワーク（Convolutional Neural Networks for Sentence Classification）」（２０１４年）（https://arxiv.org/pdf/1408.5882.pdf）に開示された技術を用いて実現されてもよい。

【００７０】

単語埋込みモデルベクトル６２０を算出するためのプロセスが図７に関連付けて以下でより詳細に説明される。

【００７１】

ステップ３５０において、入力単語に関してｋ個の最も近い名前またはｋ個の最近傍が判定される。ｋ個の最近傍または最も近い名前は類似性メトリックを用いて識別可能であ

る。類似性メトリックは、JaccardまたはDiceなどの意味論的な類似性メトリックを含み得る。ｋ個の最も近い名前は、入力単語と類似するトリグラムを含む名前である。すなわち、ｋ個の最も近い名前は、入力名「ＪＯＡＮＮＡ」と類似するトリグラムを有するファーストネームである。

【００７２】

ｋ個の最近傍を識別するために、ファジー文字列一致を実行することができる。これまで見たことのない文字列を想定する場合、類似性メトリックにより、辞書から最も近い一致を見出すことができ、次いで、ｋ個の最近傍の形式でｋ個の最近傍の分類に基づいて、これまで見たことのない文字列の分類についての予測を行なうことができる。分類が予測されている単語（たとえば、これまで見たことのないファーストネーム）が、たとえば、StatSimのｋ個の最近傍に通されて、上位Ｋ（たとえば、３個）の一致が用いられる。上位Ｋ個の一致はセットＳ内にある。

【００７３】

ｋ個の最も近い名前を識別することができる。入力パラメータｋは最も近い名前の数を表わす。判定されたｋ個の最も近い名前はセットＳにおいて識別される。説明を簡潔にするために、３つの最も近い名前が識別される。したがって、ｋの値は３である。しかしながら、ｋの値は、ユーザが所望する結果に基づいて変更することができる。たとえば、代替的な実施形態では、ｋの値は１０であり得る。

【００７４】

図６の要素６４０は、変数ｋによって識別されるｋ個の最も近い名前の数が３であることを示す。したがって、３個の最も近い名前が入力単語について識別されている。

【００７５】

図８は、いくつかの例示的な実施形態に従った、入力単語についての最も近い名前のセットＳ８００を示す。入力単語「ＪＯＡＮＮＡ」についての最近傍のセットＳは「ＪＯＡＮＮＥ」、「ＪＯＨＮ」、および「ＡＮＮＡ」の３つの最も近い名前を含む。

【００７６】

ステップ３６０において、最長共通部分列が判定される。具体的には、入力単語からのトリグラムおよびｋ個の最近傍からのトリグラム（「ＪＯＡＮＮＥ」、「ＪＯＨＮ」および「ＡＮＮＡ」）に関して最長共通部分列が識別される。

【００７７】

図９は、いくつかの例示的な実施形態に従った、Ｎ個のキューのリストを表わすテーブルＱ９００を示す。各キューは、入力単語ｗとｋ個の最近傍のセットとの間の最長共通部分列（ＬＣＳ）に従って初期化される。各キューは、式ＬＣＳ（ｗ，Ｓ［ｉ］）に従って初期化することができる。ステップ３５０で識別されたｋ個の最も近い名前から、最長共通部分列が識別される。

【００７８】

テーブルＱ９００は、入力単語のトリグラムｗｑに共通するｋ個の最近傍のトリグラムを表わす列９４０と、ｋ個の最近傍を含む列９５０と、列９５０内のそれぞれのｋ個の最近傍についての分類（たとえば、性別）を識別する列９６０とを含む。図示の例では、３つの最近傍が存在する（ｋ＝３）ので、ｋ個の最近傍の各々に対して３つの行が存在する。しかしながら、これは単に説明を目的としたものに過ぎず、ｋ個の最近傍は３に限定されない。たとえば、１０個の最近傍が用いられてもよい。テーブルＱ９００は、ｋ個の最近傍の数に応じてより大きくなるだろう。

【００７９】

行９１０は名前「ＪＯＡＮＮＥ」に対応し、行９２０は名前「ＪＯＨＮ」に対応し、行９３０はｋ個の最近傍の名前「ＡＮＮＡ」に対応する。行９１０は入力単語「ＪＯＡＮＮＥ」に対応する。「ＪＯＡＮＮＥ」に対応する行９１０はトリグラム９１１（たとえば、＾ＪＯ、ＪＯＡ、ＯＡＮ、ＡＮＮ）を含む。名前「ＪＯＨＮ」に対応する行９２０はトリグラム９１２（たとえば、＾ＪＯ）を含む。名前「ＡＮＮＡ」に対応する行９３０はトリグラム９１３（たとえば、ＡＮＮ、ＮＮＡ、ＮＡ＄）を含む。

【００８０】
トリグラム９１１、９１２および９１３は、入力単語ｗ「ＪＯＡＮＮＡ」についてのトリグラムｗｑにおけるトリグラムと一致する。入力単語ｗ「ＪＯＡＮＮＡ」についてのトリグラムｗｑと一致しないｋ個の最近傍についてのトリグラム（ＪＯＡＮＮＥ、ＡＮＮＡ、ＪＯＨＮ）はテーブルＱ９００に含まれない。たとえば、名前「ＪＯＨＮ」はトリグラム「ＯＨＮ」を含む。トリグラム「ＯＨＮ」は、入力単語ｗ「ＪＯＡＮＮＡ」についてのトリグラムｗｑ内のトリグラムと一致しないので、トリグラム「ＯＨＮ」はｋ個の最近傍のうちの１つ「ＪＯＨＮ」についてのトリグラム９１２に含まれない。したがって、テーブルＱ９００には、ｋ個の最近傍と入力名とに共通するトリグラムが配置される。

【００８１】
ｋ個の最近傍と入力名とに共通するトリグラムは最長共通部分列（ＬＣＳ）を用いて判定される。２つの入力文字列（たとえば、入力名と最近傍のうちの１つと）があると想定する場合、ＬＣＳアルゴリズムを用いて、トリグラムの２つのリスト間で最長共通部分列を見出すことができる。ＬＣＳアルゴリズムは、あらゆる目的のために引用により本明細書に援用されている、Wagnerによる「文字列間の補正の問題（The String-to-String Correction Problem）」（１９７４）（http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.367.5281&rep=rep1&type=pdf）に開示される技術を用いて実現され得る。

【００８２】
ステップ３７０において、分類値が算出される。分類値の算出は、図１０および図１１に関連付けてより詳細に説明される。

【００８３】
図１２は、いくつかの例示的な実施形態に従った、第１の分類１２３１および第２の分類１２３２についての分類値を含む行列Ｍ１２００を示す。図１２は、分類値が判定されて行列にポピュレートされた後の行列Ｍを示す。

【００８４】
図１２に示すように、行列Ｍは、入力単語ｗについて識別された各トリグラムｗｑごとに６つの行１２１０を含む。行列内の各行はトリグラムを表わしており、各トリグラムは、辞書からのＮ個のStatSim一致単語のうちの１つ以上における対応するトリグラムと一致している。行列Ｍは、３００列のベクトルについての単語埋込みベクトル値１２２０を含む。

【００８５】
行列Ｍは分類列１２３０を含む。分類列は第１の分類列１２３１と第２の分類列１２３２とを含む。２つの分類列が説明されているが、分類またはラベルの数は、ユーザによって識別される分類の種類に基づいて異なり得る。分類の数は、たとえば、ユーザが実行したい販売営業活動の種類、ユーザがターゲットにしたい客層、またはユーザのビジネス戦略に応じて異なり得る。

【００８６】
分類値は、その特定のトリグラムを含む他の類似する名前が何個あるかを表わしている。他の類似する名前は既知の名前の辞書から得られる。説明される例では、名前が分析されているので、その辞書は既知の名前のものである。しかしながら、さまざまな種類の入力単語（たとえば、企業名）が分析されている場合、そのような名前の辞書がトリグラム比較のために用いられることとなる。

【００８７】
図１２に示すように、第１の分類列１１３１は、トリグラム「＾ＪＯ、ＪＯＡ、ＯＡＮ、ＡＮＮ、ＮＮＡ、ＮＡ＄」のそれぞれについての値「０．３、０．１、０．０、０．０、０．０、０．０」を含む。第２の分類列１２３２は、トリグラム「＾ＪＯ、ＪＯＡ、ＯＡＮ、ＡＮＮ、ＮＮＡ、ＮＡ＄」のそれぞれについての値「０．１、０．４、０．３、０．５、０．３、０．３」を含む。

【００８８】

分類結果に基づくと、第２の分類列１１３２は結果として第１の分類列１１３１よりも高い値をもたらす。したがって、入力単語「ＪＯＡＮＮＡ」は、第２の分類（たとえば、女性）に分類される可能性が高い。しかしながら、行列が畳み込みニューラルネットワーク（ＣＮＮ）を介して実行されるため、精度の高い結果が得られる。

【００８９】

ステップ３８０において、行列Ｍは畳み込みニューラルネットワークを介して実行される。行列Ｍに単語埋込みモデルベクトル値および分類値がポピュレートされた後、行列は畳み込みニューラルネットワークを通じて実行されることとなる。畳み込みニューラルネットワークは、入力単語が第１の分類に属する可能性が高いかまたは第２の分類に属する可能性が高いかを識別することができる。

【００９０】

例示的な実施形態におけるＣＮＮは、単語埋込みモデルからのベクトルと、トリグラム分析を用いて得られた分類値とを含む行列を分析する。ＣＮＮ内の緻密（Dense）層とグローバルマックスプーリング（Global Max-Pooling）層との間のアクティベーションが正規化線形ユニット（Rectified Linear Unit：ＲｅＬＵ）であるので、当該ＲｅＬＵの重みの初期化をＣＮＮのために用いることができる。畳み込みニューラルネットワークは、図１３に関連付けて以下でより詳細に説明される。

【００９１】

ＩＶ．単語埋込みモデル

図７は、いくつかの例示的な実施形態に従った、単語についてのベクトルを算出するための一般的なプロセス７００を示す。しかしながら、例示的な実施形態はトリグラムについてのベクトルを算出するだろう。

【００９２】

ステップ７１０において、単語が入力される。図７に示す例では、単語の入力セットは「Bridgestone」、「Firestone」、および「Michelin」を含み得る。しかしながら、例示的な一実施形態では、単語の入力セットは新語について識別されたトリグラムを含み得る。

【００９３】

ステップ７２０において、入力データセットが分析される。データは、入力データセットを分析するために、Word2Vecなどの機械学習技術を用いることによって分析することができる。Word2Vecは、あらゆる目的のために引用により本明細書に援用されている、Mikolovらによる「機械翻訳のための言語の類似性の活用（Exploiting Similarities among Languages for Machine Translation）」（２０１３）（http://arxiv.org/pdf/1309.4168.pdf）に開示される技術を用いて実現され得る。Word2Vecは、テキスト入力（たとえば、大規模なデータソースからのテキストコーパス）を受信し得るとともに、各入力単語のデータ構造（たとえば、ベクトル表現）を単語のセットとして生成し得る。データ構造は、本明細書では「モデル」または「Word2Vecモデル」と称されることもある。Word2Vecについて説明されているが、他の単語埋込みモデルを用いてデータ分析を実行することもできる。

【００９４】

ステップ７３０において、単語のセット内の各単語が複数の属性に関連付けられる。属性は、特徴、ベクトル、成分、および特徴ベクトルとも称され得る。たとえば、データ構造は、単語のセット内の各単語に関連付けられた３００個の特徴を含み得る。特徴は、たとえば、単語を記述する性別、国籍などを含み得る。特徴の各々は、感情との関連性に基づいてトレーニングされた機械学習（たとえば、教師あり機械学習）のための技術に基づいて判定され得る。

【００９５】

大規模なテキストコーパス（たとえば、ニュースアグリゲータ、またはGoogle（登録商標）ニュースコーパスなどの他のデータソース）を用いて構築されるWord2Vecモデルを用いて、対応する数値ベクトル値（たとえば、浮動小数点）が各入力単語ごとに識別され

得る。これらのベクトルが分析されると、ベクトルがベクトル空間内で（ユークリッド的に）「近接して」いると判定され得る。ステップ７４０に示すように、３つの入力単語は、ベクトル空間内で互いに密接にクラスタリングされる。

【００９６】

いくつかの実施形態では、Word2Vecモデルは、第三者プロバイダによって生成され得る。Word2Vecモデルは、プロバイダのアプリケーションプログラミングインターフェイス（application programming interface：ＡＰＩ)を介して取得され得る。ＡＰＩは、モデル内の各単語ごとの成分の数などの、単語埋込みモデルに関する情報を含むWord2Vecモデルを取得するための機能を提供し得る。

【００９７】

ステップ７３０は、トレーニングデータに基づいてデータ構造（たとえば、ベクトルデータ構造）を２次元行列として生成するステップを含み得る。行列における各軸（ｘ軸およびｙ軸）は座標または次元を有する。トレーニングデータの場合、１つ以上のアプリケーション（たとえば、Ｌａｍｂｄａアプリケーション）を利用して、最長テキスト文字列の長さに基づいてベクトルの高さを計算し得る。たとえば、データ構造は各メッセージごとに生成される。その高さは単一のレビューにおける単語の最大数である。２次元行列を構築する際に、各行が単語ベクトルとして規定されるとともに、各列が特徴ベクトルとして規定され得る。データ構造は、畳み込みニューラルネットワーク（ＣＮＮ）を実現するためのＡＰＩへの入力として作成される。２次元行列は、ｙ軸が単一のメッセージ中の各単語ごとにエントリを有するとともに、ｘ軸がベースライン感情分析アプローチについてのものとなるように作成される。ｘ軸上の各々のエントリまたは次元は、Word2Vecモデルにおける特徴のうちの或る特徴に対応する。ｘ軸上の単語に関して複数の特徴をリスト化することができる。各単語ごとの特徴の各々は、トレーニングデータに基づいて生成されるWord2Vecモデルから取得され得る。

【００９８】

図６の単語埋込みモデルベクトル列６２０は生成されたデータ構造（たとえば、ベクトルデータ構造）を例示している。

【００９９】

Ｖ．分類値の算出

Ａ．分類値の算出方法の概要

図１０は、いくつかの例示的な実施形態に従った、分類値を算出するための方法１０００のフローチャートを示す。図１０において実行されるステップは図３のステップ３７０に対応し得る。

【０１００】

ステップ１０１０において、入力単語からのトリグラムが識別される。識別されるトリグラムｗｑは図３のステップ３２０において識別されるトリグラムであり得る。

【０１０１】

ステップ１０２０において、入力単語についてのトリグラムに共通するｋ個の最近傍についてのトリグラムが識別される。図９に示すように、列９４０は、入力単語のトリグラムｗｑに共通するｋ個の最近傍のトリグラムを表わす。

【０１０２】

ステップ１０３０では、入力単語からのトリグラムが、ｋ個の最近傍の各々についてのトリグラムと比較される。入力単語からのトリグラム間で一致する頻度はｋ個の最近傍の各々についてのトリグラムと比較される。一致する頻度はカウンタを用いて判定することができる。

【０１０３】

ステップ１０４０において、分類値は、入力単語内のトリグラムとｋ個の最近傍からのトリグラムとの間で一致する頻度に基づいて算出される。

【０１０４】

ステップ１０５０において、未知語のトリグラムについての分類値は、算出された分類

値でポピュレートされる。分類値は行列にポピュレートすることができる。
【０１０５】
トリグラムの分類値が判定されて行列にポピュレートされた後、当該行列は畳み込みニューラルネットワークを通じて実行され得る。
【０１０６】
Ｂ．分類値を算出する詳細な方法
図１１は、いくつかの例示的な実施形態に従った、分類値を算出するための方法１１００の詳細なフローチャートを示す。図１１において実行されるステップは図３のステップ３７０に対応し得る。図１１は図１０において実行されるステップをより詳細に説明する。
【０１０７】
図１１に示す方法は以下の式で表わすことができる。
行ごとに行列Ｍを通るループ
Ｑを通るループ
ｗｑ［０］＝＝Ｑ［ｉ］［０］であれば、
ｐｏｐ　Ｑ［ｉ］
Ｓ［ｉ］に関連付けられたＣ［ｊ］をインクリメント
行列Ｍの追加の列にλ・Ｃ［ｊ］／ｋをポピュレート
変数Ｃはカウンタを表わす。カウンタＣは、起こり得る各分類ごとに開始することができる。カウンタは、一時的に格納される一時的なカウンタであり得る。カウンタは、たとえば、データストア２１４に格納することができる。説明される例では、判定されるべき２つの分類（たとえば、男性および女性）があるため、２つのカウンタを開始することができる。変数λは追加の特徴列をスケーリングするために用いられる。例示を目的として、変数λは値０．００１によって表わされる。変数λの値は、より容易に単語埋込みモデルベクトル値に類似するようにスケーリングまたは作成される必要があるデータの値に基づいている。変数Ｑは、図９に示すようにＮ個のキューのリストを表わす。
【０１０８】
変数Ｑ［ｉ］はｋ個の近傍のうちの１つを表わす。ｋ個の近傍は、ｋ個の最近傍（Nearest Neighbor：ＫＮＮ）アルゴリズムを用いて取得され得る。したがって、Ｑ［ｉ］は、トリグラムｗｑに共通するｋ個の近傍のうちの１つについてのトリグラムを表わす。Ｃ［ｊ］における変数ｊは分類を表わしている。したがって、変数Ｃ［ｊ］は分類のためのカウンタを表わしている。
【０１０９】
説明される例では、入力単語ｗについてのトリグラムｗｑは、「＾ＪＯ、ＪＯＡ、ＯＡＮ、ＡＮＮ、ＮＮＡ、ＮＡ＄」である。ｋ個の近傍についてのＱにおけるトリグラムは、「＾ＪＯ、ＪＯＡ、ＯＡＮ、ＡＮＮ」、「＾ＪＯ」および「ＡＮＮ、ＮＮＡ、ＮＡ＄」である。最も近い名前Ｓのセットは「ＪＯＡＮＮＥ」、「ＪＯＨＮ」および「ＡＮＮＡ」を含む。この例におけるＱは、３個の最近傍についてのＮ＝３のキューのリストを含む。
【０１１０】
ステップ１１１０において、カウンタＣは起こり得る各分類ごとに開始される。カウンタはＣ［ｊ］によって表わすことができる。このｊは分類（たとえば、第１、第２、第３など）を表わす。説明される例では、２つの分類（たとえば、男性および女性）があるため、２つのカウンタが開始される。第１のカウンタＣ［０］は第１の分類（たとえば、男性）に対応し得るとともに、第２のカウンタＣ［１］は第２の分類（たとえば、女性）に対応し得る。カウンタは、新語分類サーバのデータストアに格納することができる。カウンタは、各反復の後、ｗｑのトリグラムが分析された後に除去することができ、新しいカウンタを開始することができる。
【０１１１】
ステップ１１２０において、入力単語ｗについての行列Ｍ内のトリグラム（たとえば、＾ＪＯ）が分析のために識別または選択される。図１１の方法は、入力単語についての各トリグラムｗｑごとに繰り返され続けるだろう。分析のための行列Ｍ内のトリグラムは、

一度に１つずつ順々にループされる。ここで、ｗｑ「＾ＪＯ、ＪＯＡ、ＯＡＮ、ＡＮＮ、ＮＮＡ、ＮＡ＄」に６つのトリグラムがあるので、分析すべき最初のトリグラムは「＾ＪＯ」であり、分析すべき最後のトリグラムは「ＮＡ＄」である。したがって、６つのトリグラムｗｑを含む例の場合、ステップ１１２０～１１７０が各トリグラムごとに６回繰り返され得る。

【０１１２】

ステップ１１３０では、ステップ１１２０で識別されたトリグラムに一致するｋ個の最近傍（たとえば、ＪＯＡＮＮＥ、ＪＯＨＮ、ＡＮＮＡ）についてのＮ個のキューＱのリスト内のトリグラム（たとえば、＾ＪＯ）が判定される。この例では、トリグラム＾ＪＯは２回発生する。入力単語の第１のトリグラムに一致するｋ個の最も近い名前のトリグラムが判定される。ｋ個の最近傍「ＪＯＡＮＮＥ」に対して１回、ｋ個の最近傍「ＪＯＨＮ」に対して１回とされる。

【０１１３】

ステップ１１４０において、入力単語からのトリグラムに一致する特定の最近傍についての一致するトリグラムの分類が判定される。一致するトリグラム（たとえば、＾ＪＯ）の分類は、たとえば、それぞれの最近傍についての分類を示す列９６０から識別される。したがって、名前「ＪＯＡＮＮＥ」の分類（たとえば女性）が識別されるとともに、名前「ＪＯＨＮ」の分類（たとえば男性）が識別される。なぜなら、「ＪＯＡＮＮＥ」および「ＪＯＨＮ」がともにトリグラム「＾ＪＯ」を含んでいるからである。

【０１１４】

ステップ１１５０において、分類に関連付けられたカウンタがインクリメントされる。ｋ個の最も近い名前についての１つ以上のトリグラムの判定済み分類に対応する複数の分類の各々のためのカウンタがインクリメントされる。特定の最近傍「ＪＯＡＮＮＥ」に関する分類が「女性」であるので、第２の分類に関する第２のカウンタＣ［２］がインクリメントされる。特定の最近傍「ＪＯＨＮ」に関する分類が「男性」であるので、第１の分類に関する第１のカウンタＣ［１］もインクリメントされる。

【０１１５】

ステップ１１６０において、Ｎ個のキューＱのリストを更新することができる。入力単語トリグラム（たとえば、＾ＪＯ）に一致するトリグラム（たとえば、＾ＪＯ）は、Ｎ個のキューＱ（たとえば、名前「ＪＯＡＮＮＥ」および名前「ＪＯＨＮ」について）のリストから除去することができる。

【０１１６】

ステップ１１７０において、識別されたトリグラムｗｑについての行列Ｍに分類値をポピュレートすることができる。したがって、図１２の分類列１２３０内の行１２３３に示される分類値は、ステップ１１７０の後に行列に埋められるだろう。

【０１１７】

例示的な実施形態では、分類値は、各トリグラム（たとえば、＾ＪＯ、ＪＯＡ、ＯＡＮ、ＡＮＮ、ＮＮＡ、ＮＡ＄）が分類された後にポピュレートされる。しかしながら、行列Ｍにおける分類値のポピュレートは、入力単語についてのトリグラムｗｑがすべて分析された後に実行され得る。

【０１１８】

分類列は、式λ・Ｃ［ｊ］／ｋに基づいてポピュレートすることができる。ｋは最近傍の数または量である。変数λはスケーリングのために用いられる。スケーリングは、分類値が列１２２０内の単語埋込みモデル値と同じ桁になるように実行される。上述の例における単語埋込み値は小数点の前に３つの先行ゼロを有する。分類値がスケーリングされなかった場合、分類値内の余分な列がＣＮＮを押さえ込んでしまう可能性がある。

【０１１９】

１１８０において、分析されるべき入力単語についての追加のトリグラムｗｑが存在するかどうかが判定される。ステップ１０２０、１０３０、１０４０、１０５０、１０６０および１０７０は、入力単語についてのすべてのトリグラムｗｑがｋ個の最近傍について

のトリグラムと比較されるまで繰り返される。すなわち、入力単語ｗ「ＪＯＡＮＮＡ」についてのすべてのトリグラムｗｑがｋ個の最近傍についてのＮ個のキューＱのリスト内のすべてのトリグラムと比較されてしまうまで、ステップ１０２０、１０３０、１０４０、１０５０、１０６０および１０７０が繰り返される。

【０１２０】

分析されるべき入力単語ｗについての追加のトリグラムｗｑが存在する場合、ｗｑ内の次のトリグラムに対してステップ１０２０が繰り返される。たとえば、ｗｑにおいて分析されるべき次のトリグラムは「ＪＯＡ」である。

【０１２１】

ステップ１１８０において、分析されるべき入力単語ｗについての追加のトリグラムｗｑが存在しないと判定された場合、プロセスが終了する。

【０１２２】

Ｃ．分類値がポピュレートされた行列

図１２は、いくつかの例示的な実施形態に従った、分類値を含むように完成された行列Ｍを示す。図１２に示す行列Ｍは、図１０および図１１に示す方法を実行した後のポピュレートされた行列である。

【０１２３】

図１２に示すように、第１の分類列１２３１は、トリグラム「＾ＪＯ、ＪＯＡ、ＯＡＮ、ＡＮＮ、ＮＮＡ、ＮＡ＄」のそれぞれについての値「０．３、０．１、０．０、０．０、０．０、０．０」を含む。第２の分類列１２３２は、トリグラム「＾ＪＯ、ＪＯＡ、ＯＡＮ、ＡＮＮ、ＮＮＡ、ＮＡ＄」のそれぞれについての値「０．１、０．４、０．３、０．５、０．３、０．３」を含む。

【０１２４】

分類結果に基づいて、第２の分類列１２３２は結果として、第１の分類列１２３１よりも高い値をもたらす。したがって、入力単語「ＪＯＡＮＮＡ」は第２の分類（たとえば、女性）に分類される可能性が高い。

【０１２５】

しかしながら、分類結果の正確さを保証するために、行列ＭはＣＮＮを介して実行されるだろう。ＣＮＮからの出力は長さｎのｓｏｆｔｍａｘベクトルであり、ここでｎは起こり得る分類の数である。ｓｏｆｔｍａｘ関数は、Ｋ個の実数のベクトルｚを入力として取ることができるとともに当該入力を入力された数の指数に比例するＫ個の確率からなる確率分布に正規化することができる関数である。ｓｏｆｔｍａｘベクトルにおいて、各成分は、そのクラスにあるサンプルの確率（０．０～１．０）を表わす。しばしば、最も高い確率を有するクラスが予測として選択され、他の確率は無視される。

【０１２６】

ＶＩ．畳み込みニューラルネットワーク

図１３は、いくつかの例示的な実施形態に従った畳み込みニューラルネットワーク１３００を示す。

【０１２７】

図１３は、標準的な畳み込みニューラルネットワークの一例を示す。文章をモデリングするための畳み込みニューラルネットワーク（ＣＮＮ）は、あらゆる目的のために引用により援用されている、Kalchbrennerによる「文章をモデリングするための畳み込みニューラルネットワーク（A Convolutional Neural Network for Modelling Sentences）」（２０１４年）（http://www.aclweb.org/anthology/P14-1062）における技術を用いることができる。ＣＮＮは、１次元畳み込み層を用いて実現されてもよい。ＣＮＮは、第三者（たとえば、GitHub）によって提供されるパッケージまたはライブラリを用いて実現されてもよい。パッケージは、Java（登録商標）およびScala（登録商標）のためにPython（登録商標）および／またはDeepLearing4Jを用いて実現されるKerasであってもよい。さらに、ＣＮＮは、たとえば、３層または７層のＣＮＮであってもよい。これらは単なる例に過ぎず、さまざまなＣＮＮを用いることができる。さらに、ＣＮＮはユー

ザが独自に作成してもよい。
【０１２８】
少なくとも１つの実施形態では、各メッセージごとのベクトルを参照するアレイは、感情分析を判定するためにＣＮＮ技術とともに用いられ得る。技術の例は、あらゆる目的のために引用により援用されている「http://www.wildml.com/2015/11/understanding-convolutional-neural-networks-for-nlp/」および「https://github.com/fchollet/keras/blob/master/examples/imdb_cnn.py/」において見出される技術に基づいて実現され得る。ＣＮＮ技術を実現するために、変数フィルタ長＝３、CNNDenseDropout＝０．２、dense_dims＝２０、batch_size＝１０、nb_epoch＝５、validation_split＝０．１についての値が設定され得る。ＣＮＮは、上記で生成されたデータ構造を用いて以下の関数呼出しを実行することによって実現され得る。ＣＮＮは１次元構造としての単語についてのデータ構造に基づいて実現され得る。
【０１２９】
関数呼出しに基づいてＣＮＮを実行することにより、初期感情分析を実行するためにＣＮＮを用いてテストするためのトレーニングデータが実行される。
【０１３０】
ＶＩＩ．ユーザインターフェイス
図１４は、いくつかの例示的な実施形態に従った、新語分類を実行するためのユーザインターフェイス１４００を示す。
【０１３１】
ユーザインターフェイス１４００は、図１に示される分析システム１１０のデータエンリッチメントシステム１２０のディスプレイ上に表示され得る。ユーザインターフェイス１４００は、データセットからのデータの複数の列１４２０を含む。データセットは、たとえば、顧客情報、またはスプレッドシートにおいて利用可能な他の種類のデータを含み得る。図１４に示す例では、顧客情報は都市名、州、および郵便番号を含む。しかしながら、これは単なる一例に過ぎず、データセットはさまざまな種類の情報を含み得る。
【０１３２】
ユーザは、新語分類が実行されるべきデータセットを入力することができる。たとえば、ユーザは、データエンリッチメントシステムの対話型ユーザインターフェイス上でデータセットを選択することができる。ユーザインターフェイスは、複数のデータソースオプションを表示することができる。これら複数のデータソースオプションからはデータセットを得ることができる。ユーザは、エンリッチ化、分析、または視覚化したいデータセットをアップロードすることができる。データセットはスプレッドシートなどの列の形式であり得る。
【０１３３】
ユーザインターフェイス１４００は推薦パネル１４１０を含む。推薦パネルは、データセットに適用するべきアクションの１つ以上の推薦をユーザに提供する。たとえば、推薦は、データの列についての新語を判定することによってデータをエンリッチ化することを含み得る。推薦は、データセット内のデータの１つ以上の列について新語分類を実行することによってデータをエンリッチ化することを含み得る。たとえば、ユーザは、新語分類を実行することによって、データセットの都市列をエンリッチ化するために推薦１４３０を選択することができる。
【０１３４】
データエンリッチメントシステムは、データのエンリッチ化などのデータセットに対する処理を実行するための推薦を提供するように構成される。すなわち、データエンリッチメントシステムは、データセットを分析して、データセットに関する推薦アクションをユーザに提供することでデータをエンリッチ化することができる。データエンリッチメントシステムは、提供されたデータセットに基づいて、どのエンリッチメントがデータに有益であり得るかを自動的に判定することができる。
【０１３５】

その一部がパネル１４１０に示されているデータセットを想定する場合、新語分類の実行を選択することによって、データの列をエンリッチ化するようにとの推薦をユーザに提することができる（要素１４３０）。図１４は、ユーザが新語分類の実行を選択することができるユーザインターフェイスの簡略図を示す。追加のディスプレイが提供されてもよく、または、ユーザは、新語分類が実行される前に追加の情報を提供するように促されてもよい。

【０１３６】

したがって、例示的な一実施形態では、データセットが新語分類を実行することができるデータを含む場合、データエンリッチメントシステムはユーザに新語分類を実行することを提案し得る。ユーザが新語分類の実行を選択する（たとえば、要素１４３０を選択する）と、選択されたデータの列に対して新語分類が実現され得る。

【０１３７】

例示的な実施形態は、精度が改善されたシステム、方法、およびコンピュータ可読媒体を提供する。したがって、例示的な実施形態は新語の分類の判定を改善させる。

【０１３８】

例示的な実施形態では非常に正確な分類が提供されるので、データエンリッチメントシステムのユーザに提供されるデータは、より完全なものになり得る。これにより、ユーザにとってより有用なデータとなる。

【０１３９】

例示的な一実施形態は、トリグラム、トリグラムについての単語埋込みモデル（たとえば、Word2Vec）、ｋ個の最も近い名前（たとえば、StatSim）、カウンタでの追加の分類（たとえば、男性／女性）、および畳み込みニューラルネットワーク（ＣＮＮ）の組合わせを含み得る。

【０１４０】

トリグラムにより、新語に関連付けられた特徴を正確に判定する際に得られる結果が改善されたものとなる。意味は、トリグラム自体の値だけではなく、トリグラムの順序から推定される。

【０１４１】

さらに、例示的な実施形態は、トリグラムに関する単語埋込みモデルＣＮＮの組み合わせを用いる。追加の列が追加の特徴として行列の右側に追加される。

【０１４２】

ＶＩＩＩ．ハードウェア概要

図１５は、ある実施形態を実現するための分散型システム１５００の簡略図を示す。図示される実施形態において、分散型システム１５００は、１つ以上の通信ネットワーク１５１０を介してサーバ１５１２に結合された１つ以上のクライアントコンピューティングデバイス１５０２、１５０４、１５０６、および１５０８を含む。クライアントコンピューティングデバイス１５０２、１５０４、１５０６、および１５０８は１つ以上のアプリケーションを実行するように構成され得る。

【０１４３】

さまざまな実施形態において、サーバ１５１２は、本開示に記載される正規表現の自動化された生成を可能にする１つ以上のサービスまたはソフトウェアアプリケーションを実行するように適合され得る。たとえば、特定の実施形態では、サーバ１５１２は、クライアントデバイスから送信されたユーザ入力データを受信し得る。この場合、ユーザ入力データは、クライアントデバイスで表示されたユーザインターフェイスを介して、クライアントデバイスによって受信される。次いで、サーバ１５１２は、ユーザ入力データを、ユーザインターフェイスを介して表示するためにクライアントデバイスに送信される正規表現に変換し得る。

【０１４４】

特定の実施形態では、サーバ１５１２はまた、非仮想環境および仮想環境を含み得る他のサービスまたはソフトウェアアプリケーションを提供し得る。いくつかの実施形態では

、これらのサービスは、クライアントコンピューティングデバイス１５０２、１５０４、１５０６、および／または１５０８のユーザに対して、サービスとしてのソフトウェア（Software as a Service：ＳａａＳ）モデルのようなウェブベースのサービスまたはクラウドサービスとして提供され得る。クライアントコンピューティングデバイス１５０２、１５０４、１５０６、および／または１５０８を操作するユーザは、１つ以上のクライアントアプリケーションを利用してサーバ１５１２とやり取りすることで、これらのコンポーネントによって提供されるサービスを利用し得る。

【０１４５】

図１５に示される構成では、サーバ１５１２は、サーバ１５１２によって実行される機能を実現する１つ以上のコンポーネント１５１８、１５２０および１５２２を含み得る。これらのコンポーネントは、１つ以上のプロセッサ、ハードウェアコンポーネント、またはそれらの組合わせによって実行され得るソフトウェアコンポーネントを含み得る。分散型システム１５００とは異なり得る多種多様なシステム構成が可能であることが認識されるはずである。したがって、図１５に示される実施形態は、実施形態のシステムを実現するための分散型システムの一例であり、限定するよう意図されたものではない。

【０１４６】

ユーザは、クライアントコンピューティングデバイス１５０２、１５０４、１５０６、および／または１５０８を用いて、本開示の教示に従って正規表現を生成し得る１つ以上のアプリケーションを実行してもよい。クライアントデバイスは、当該クライアントデバイスのユーザが当該クライアントデバイスと対話することを可能にするインターフェイスを提供し得る。クライアントデバイスはまた、このインターフェイスを介してユーザに情報を出力し得る。図１５は４つのクライアントコンピューティングデバイスだけを示しているが、任意の数のクライアントコンピューティングデバイスがサポートされ得る。

【０１４７】

クライアントデバイスは、ポータブルハンドヘルドデバイス、パーソナルコンピュータおよびラップトップのような汎用コンピュータ、ワークステーションコンピュータ、ウェアラブルデバイス、ゲームシステム、シンクライアント、各種メッセージングデバイス、センサまたはその他のセンシングデバイスなどの、さまざまな種類のコンピューティングシステムを含み得る。これらのコンピューティングデバイスは、各種モバイルオペレーティングシステム（たとえばＭｉｃｒｏｓｏｆｔ Ｗｉｎｄｏｗｓ（登録商標）Ｍｏｂｉｌｅ（登録商標）、ｉＯＳ（登録商標）、Ｗｉｎｄｏｗｓ Ｐｈｏｎｅ（登録商標）、Ａｎｄｒｏｉｄ（登録商標）、ＢｌａｃｋＢｅｒｒｙ（登録商標）、Ｐａｌｍ ＯＳ（登録商標））を含むさまざまな種類およびバージョンのソフトウェアアプリケーションおよびオペレーティングシステム（たとえばＭｉｃｒｏｓｏｆｔ Ｗｉｎｄｏｗｓ（登録商標）、Ａｐｐｌｅ Ｍａｃｉｎｔｏｓｈ（登録商標）、ＵＮＩＸ（登録商標）またはＵＮＩＸ系オペレーティングシステム、Ｌｉｎｕｘ（登録商標）またはＬｉｎｕｘ系オペレーティングシステム、たとえば、Ｇｏｏｇｌｅ Ｃｈｒｏｍｅ（登録商標）ＯＳ）を含み得る。ポータブルハンドヘルドデバイスは、セルラーフォン、スマートフォン（たとえばｉＰｈｏｎｅ（登録商標））、タブレット（たとえばｉＰａｄ（登録商標））、携帯情報端末（ＰＤＡ）などを含み得る。ウェアラブルデバイスは、Ｇｏｏｇｌｅ Ｇｌａｓｓ（登録商標）ヘッドマウントディスプレイおよびその他のデバイスを含み得る。ゲームシステムは、各種ハンドヘルドゲームデバイス、インターネット接続可能なゲームデバイス（たとえばＫｉｎｅｃｔ（登録商標）ジェスチャ入力デバイス付き／無しのＭｉｃｒｏｓｏｆｔ Ｘｂｏｘ（登録商標）ゲーム機、Ｓｏｎｙ ＰｌａｙＳｔａｔｉｏｎ（登録商標）システム、Ｎｉｎｔｅｎｄｏ（登録商標）が提供する各種ゲームシステムなど）を含み得る。クライアントデバイスは、各種インターネット関連アプリケーション、通信アプリケーション（たとえばＥメールアプリケーション、ショートメッセージサービス（ＳＭＳ）アプリケーション）のような多種多様なアプリケーションを実行可能であり得るとともに、各種通信プロトコルを使用し得る。

【０１４８】

ネットワーク１５１０は、利用可能な多様なプロトコルのうちのいずれかを用いてデータ通信をサポートできる、当該技術の当業者には周知のいずれかの種類のネットワークであればよく、上記プロトコルは、ＴＣＰ／ＩＰ（伝送制御プロトコル／インターネットプロトコル）、ＳＮＡ（システムネットワークアーキテクチャ）、ＩＰＸ（インターネットパケット交換）、ＡｐｐｌｅＴａｌｋ（登録商標）などを含むがこれらに限定されない。単に一例として、ネットワーク１５１０は、ローカルエリアネットワーク（ＬＡＮ）、Ｅｔｈｅｒｎｅｔ（登録商標）に基づくネットワーク、トークンリング、ワイドエリアネットワーク（ＷＡＮ）、インターネット、仮想ネットワーク、仮想プライベートネットワーク（ＶＰＮ）、イントラネット、エクストラネット、公衆交換電話網（ＰＳＴＮ）、赤外線ネットワーク、無線ネットワーク（たとえば電気電子学会（ＩＥＥＥ）８０２．１１プロトコルスイートのいずれかの下で動作する無線ネットワーク、Ｂｌｕｅｔｏｏｔｈ（登録商標）および／もしくは任意の他の無線プロトコル）、ならびに／または、これらおよび／もしくは他のネットワークの任意の組み合わせを含み得る。

【０１４９】

サーバ１５１２は、１つ以上の汎用コンピュータ、専用サーバコンピュータ（一例としてＰＣ（パーソナルコンピュータ）サーバ、ＵＮＩＸ（登録商標）サーバ、ミッドレンジサーバ、メインフレームコンピュータ、ラックマウント型サーバなどを含む）、サーバファーム、サーバクラスタ、またはその他の適切な構成および／または組み合わせで構成されてもよい。サーバ１５１２は、仮想オペレーティングシステムを実行する１つ以上の仮想マシン、または仮想化を伴う他のコンピューティングアーキテクチャを含み得る。これはたとえば、サーバに対して仮想記憶装置を維持するように仮想化できる論理記憶装置の１つ以上のフレキシブルプールなどである。各種実施形態において、サーバ１５１２を、上記開示に記載される機能を提供する１つ以上のサービスまたはソフトウェアアプリケーションを実行するように適合させてもよい。

【０１５０】

サーバ１５１２内のコンピューティングシステムは、上記したもののうちのいずれかを含む１つ以上のオペレーティングシステム、および、市販されているいずれかのサーバオペレーティングシステムを実行し得る。また、サーバ１５１２は、ＨＴＴＰ（ハイパーテキスト転送プロトコル）サーバ、ＦＴＰ（ファイル転送プロトコル）サーバ、ＣＧＩ（コモンゲートウェイインターフェイス）サーバ、ＪＡＶＡ（登録商標）サーバ、データベースサーバなどを含むさまざまな追加のサーバアプリケーションおよび／または中間層アプリケーションのうちのいずれかを実行し得る。例示的なデータベースサーバは、Ｏｒａｃｌｅ（登録商標）、Ｍｉｃｒｏｓｏｆｔ（登録商標）、Ｓｙｂａｓｅ（登録商標）、ＩＢＭ（登録商標）（International Business Machines）などから市販されているものを含むが、それらに限定されない。

【０１５１】

いくつかの実現例において、サーバ１５１２は、クライアントコンピューティングデバイス１５０２、１５０４、１５０６、および１５０８のユーザから受信したデータフィードおよび／またはイベントアップデートを解析および整理統合するための１つ以上のアプリケーションを含み得る。一例として、データフィードおよび／またはイベントアップデートは、センサデータアプリケーション、金融株式相場表示板、ネットワーク性能測定ツール（たとえば、ネットワーク監視およびトラフィック管理アプリケーション）、クリックストリーム解析ツール、自動車交通監視などに関連するリアルタイムのイベントを含み得る、１つ以上の第三者情報源および連続データストリームから受信される、Ｔｗｉｔｔｅｒ（登録商標）フィード、Ｆａｃｅｂｏｏｋ（登録商標）アップデートまたはリアルタイムのアップデートを含み得るがそれらに限定されない。サーバ１５１２は、データフィードおよび／またはリアルタイムのイベントをクライアントコンピューティングデバイス１５０２、１５０４、１５０６、および１５０８の１つ以上の表示デバイスを介して表示するための１つ以上のアプリケーションも含み得る。

【０１５２】

分散型システム１５００はまた、１つ以上のデータリポジトリ１５１４、１５１６を含み得る。特定の実施形態において、これらのデータリポジトリを用いてデータおよびその他の情報を格納し得る。たとえば、データリポジトリ１５１４、１５１６のうちの１つ以上を用いて、システムにより生成された正規表現と一致する新たなデータの列のような情報を格納してもよい。データリポジトリ１５１４、１５１６は、さまざまな場所に存在し得る。たとえば、サーバ１５１２が使用するデータリポジトリは、サーバ１５１２のローカル位置にあってもよく、またはサーバ１５１２から遠隔の位置にあってもよく、ネットワークベースの接続または専用接続を介してサーバ１５１２と通信する。データリポジトリ１５１４、１５１６は異なる種類であってもよい。特定の実施形態において、サーバ１５１２が使用するデータリポジトリは、データベース、たとえば、Oracle Corporation（登録商標）および他の製造業者が提供するデータベースのようなリレーショナルデータベースであってもよい。これらのデータベースのうちの１つ以上を、ＳＱＬフォーマットのコマンドに応じて、データの格納、アップデート、およびデータベースとの間での取り出しを可能にするように適合させてもよい。

【０１５３】

特定の実施形態では、データリポジトリ１５１４、１５１６のうちの１つ以上は、アプリケーションデータを格納するためにアプリケーションによって用いられてもよい。アプリケーションが使用するデータリポジトリは、たとえば、キー値ストアリポジトリ、オブジェクトストアリポジトリ、またはファイルシステムがサポートする汎用ストレージリポジトリのようなさまざまな種類のものであってもよい。

【０１５４】

特定の実施形態において、本開示に記載される機能は、クラウド環境を介してサービスとして提供され得る。図１６は、特定の例に従った、各種サービスをクラウドサービスとして提供し得るクラウドベースのシステム環境１６００の簡略ブロック図である。図１６に示される例において、クラウドインフラストラクチャシステム１６０２は、ユーザが１つ以上のクライアントコンピューティングデバイス１６０４、１６０６、および１６０８を用いて要求し得る１つ以上のクラウドサービスを提供し得る。クラウドインフラストラクチャシステム１６０２は、サーバ１５１２に関して先に述べたものを含み得る１つ以上のコンピュータおよび／またはサーバを含み得る。クラウドインフラストラクチャシステム１６０２内のコンピュータは、汎用コンピュータ、専用サーバコンピュータ、サーバファーム、サーバクラスタ、またはその他任意の適切な配置および／または組み合わせとして編成され得る。

【０１５５】

ネットワーク１６１０は、クライアント１６０４、１６０６、および１６０８と、クラウドインフラストラクチャシステム１６０２との間におけるデータの通信および交換を容易にし得る。ネットワーク１６１０は１つ以上のネットワークを含み得る。ネットワークは同じ種類であっても異なる種類であってもよい。ネットワーク１６１０は、通信を容易にするために、有線および／または無線プロトコルを含む１つ以上の通信プロトコルをサポートし得る。

【０１５６】

図１６に示される例は、クラウドインフラストラクチャシステムの一例にすぎず、限定を意図したものではない。なお、その他いくつかの例において、クラウドインフラストラクチャシステム１６０２が、図１６に示されるものよりも多くのコンポーネントもしくは少ないコンポーネントを有していてもよく、２つ以上のコンポーネントを組み合わせてもよく、または、異なる構成もしくは配置のコンポーネントを有していてもよいことが理解されるはずである。たとえば、図１６は３つのクライアントコンピューティングデバイスを示しているが、代替例においては、任意の数のクライアントコンピューティングデバイスがサポートされ得る。

【０１５７】

クラウドサービスという用語は一般に、サービスプロバイダのシステム（たとえばクラ

ウドインフラストラクチャシステム１６０２）により、インターネット等の通信ネットワークを介してオンデマンドでユーザにとって利用可能にされるサービスを指すのに使用される。典型的に、パブリッククラウド環境では、クラウドサービスプロバイダのシステムを構成するサーバおよびシステムは、顧客自身のオンプレミスサーバおよびシステムとは異なる。クラウドサービスプロバイダのシステムは、クラウドサービスプロバイダによって管理される。よって、顧客は、別途ライセンス、サポート、またはハードウェアおよびソフトウェアリソースをサービスのために購入しなくても、クラウドサービスプロバイダが提供するクラウドサービスを利用できる。たとえば、クラウドサービスプロバイダのシステムはアプリケーションをホストし得るとともに、ユーザは、アプリケーションを実行するためにインフラストラクチャリソースを購入しなくても、インターネットを介してオンデマンドでアプリケーションをオーダーして使用し得る。クラウドサービスは、アプリケーション、リソースおよびサービスに対する容易でスケーラブルなアクセスを提供するように設計される。いくつかのプロバイダがクラウドサービスを提供する。たとえば、ミドルウェアサービス、データベースサービス、Ｊａｖａ（登録商標）クラウドサービスなどのいくつかのクラウドサービスが、カリフォルニア州レッドウッド・ショアーズのOracle Corporation（登録商標）から提供される。

【０１５８】

特定の実施形態において、クラウドインフラストラクチャシステム１６０２は、ハイブリッドサービスモデルを含む、サービスとしてのソフトウェア（ＳａａＳ）モデル、サービスとしてのプラットフォーム（ＰａａＳ）モデル、サービスとしてのインフラストラクチャ（ＩａａＳ）モデルなどのさまざまなモデルを使用して、１つ以上のクラウドサービスを提供し得る。クラウドインフラストラクチャシステム１６０２は、各種クラウドサービスのプロビジョンを可能にする、アプリケーション、ミドルウェア、データベース、およびその他のリソースの一式を含み得る。

【０１５９】

ＳａａＳモデルは、アプリケーションまたはソフトウェアを、インターネットのような通信ネットワークを通して、顧客が基本となるアプリケーションのためのハードウェアまたはソフトウェアを購入しなくても、サービスとして顧客に配信することを可能にする。たとえば、ＳａａＳモデルを用いることにより、クラウドインフラストラクチャシステム１６０２がホストするオンデマンドアプリケーションに顧客がアクセスできるようにし得る。Oracle Corporation（登録商標）が提供するＳａａＳサービスの例は、人的資源／資本管理のための各種サービス、カスタマー・リレーションシップ・マネジメント（ＣＲＭ）、エンタープライズ・リソース・プランニング（ＥＲＰ）、サプライチェーン・マネジメント（ＳＣＭ）、エンタープライズ・パフォーマンス・マネジメント（ＥＰＭ）、解析サービス、ソーシャルアプリケーションなどを含むがこれらに限定されない。

【０１６０】

ＩａａＳモデルは一般に、インフラストラクチャリソース（たとえばサーバ、ストレージ、ハードウェアおよびネットワーキングリソース）を、クラウドサービスとして顧客に提供することにより、柔軟な計算およびストレージ機能を提供するために使用される。各種ＩａａＳサービスがOracle Corporation（登録商標）から提供される。

【０１６１】

ＰａａＳモデルは一般に、顧客が、環境リソースを調達、構築、または管理しなくても、アプリケーションおよびサービスを開発、実行、および管理することを可能にするプラットフォームおよび環境リソースをサービスとして提供するために使用される。Oracle Corporation（登録商標）が提供するＰａａＳサービスの例は、Oracle Java Cloud Service（ＪＣＳ）、Oracle Database Cloud Service（ＤＢＣＳ）、データ管理クラウドサービス、各種アプリケーション開発ソリューションサービスなどを含むがこれらに限定されない。

【０１６２】

クラウドサービスは一般に、オンデマンドのセルフサービスベースで、サブスクリプシ

ョンベースで、柔軟にスケーラブルで、信頼性が高く、可用性が高い、安全なやり方で提供される。たとえば、顧客は、サブスクリプションオーダーを介し、クラウドインフラストラクチャシステム１６０２が提供する１つ以上のサービスをオーダーしてもよい。次いで、クラウドインフラストラクチャシステム１６０２は、処理を実行することにより、顧客のサブスクリプションオーダーで要求されたサービスを提供する。クラウドインフラストラクチャシステム１６０２を、１つ以上のクラウドサービスを提供するように構成してもよい。

【０１６３】

クラウドインフラストラクチャシステム１６０２は、さまざまなデプロイメントモデルを介してクラウドサービスを提供し得る。パブリッククラウドモデルにおいて、クラウドインフラストラクチャシステム１６０２は、第三者クラウドサービスプロバイダによって所有されていてもよく、クラウドサービスは一般のパブリックカスタマーに提供される。このカスタマーは個人または企業であってもよい。プライベートクラウドモデルでは、クラウドインフラストラクチャシステム１６０２がある組織内で（たとえば企業組織内で）機能してもよく、サービスはこの組織内の顧客に提供される。たとえば、この顧客は、人事部、給与部などの企業のさまざまな部署であってもよく、企業内の個人であってもよい。コミュニティクラウドモデルでは、クラウドインフラストラクチャシステム１６０２および提供されるサービスは、関連コミュニティ内のさまざまな組織で共有されてもよい。上記モデルの混成モデルなどのその他各種モデルが用いられてもよい。

【０１６４】

クライアントコンピューティングデバイス１６０４、１６０６、および１６０８は、異なる種類であってもよく（たとえば図１５に示されるデバイス１５０２、１５０４、１５０６、および１５０８）、１つ以上のクライアントアプリケーションを操作可能であってもよい。ユーザは、クライアントデバイスを用いることにより、クラウドインフラストラクチャシステム１６０２が提供するサービスを要求するなど、クラウドインフラストラクチャシステム１６０２とのやり取りを行い得る。

【０１６５】

いくつかの実施形態において、クラウドインフラストラクチャシステム１６０２が、管理関連サービスを提供するために実行する処理はビッグデータ解析を含み得る。この解析は、大きなデータセットを使用し、解析し、処理することにより、このデータ内のさまざまな傾向、挙動、関係などを検出し可視化することを含み得る。この解析は、１つ以上のプロセッサが、場合によっては、データを並列に処理し、データを用いてシミュレーションを実行するなどして、実行してもよい。たとえば、自動的に正規表現を決定するために、ビッグデータ解析がクラウドインフラストラクチャシステム１６０２によって実行されてもよい。この解析に使用されるデータは、構造化データ（たとえばデータベースに格納されたデータもしくは構造化モデルに従って構造化されたデータ）および／または非構造化データ（たとえばデータブロブ（blob）（binary large object：バイナリ・ラージ・オブジェクト））を含み得る。

【０１６６】

図１６の例に示されるように、クラウドインフラストラクチャシステム１６０２は、クラウドインフラストラクチャシステム１６０２が提供する各種クラウドサービスのプロビジョンを容易にするために利用されるインフラストラクチャリソース１６３０を含み得る。インフラストラクチャリソース１６３０は、たとえば、処理リソース、ストレージまたはメモリリソース、ネットワーキングリソースなどを含み得る。

【０１６７】

特定の実施形態において、異なる顧客に対しクラウドインフラストラクチャシステム１６０２が提供する各種クラウドサービスをサポートするためのこれらのリソースを効率的にプロビジョニングし易くするために、リソースを、リソースのセットまたはリソースモジュール（「ポッド」とも処される）にまとめてもよい。各リソースモジュールまたはポッドは、１種類以上のリソースを予め一体化し最適化した組み合わせを含み得る。特定の

実施形態において、異なるポッドを異なる種類のクラウドサービスに対して予めプロビジョニングしてもよい。たとえば、第１のポッドセットをデータベースサービスのためにプロビジョニングしてもよく、第１のポッドセット内のポッドと異なるリソースの組み合わせを含み得る第２のポッドセットをＪａｖａサービスなどのためにプロビジョニングしてもよい。いくつかのサービスについて、これらのサービスをプロビジョニングするために割り当てられたリソースをサービス間で共有してもよい。

【０１６８】

クラウドインフラストラクチャシステム１６０２自体が、クラウドインフラストラクチャシステム１６０２の異なるコンポーネントによって共有されるとともにクラウドインフラストラクチャシステム１６０２によるサービスのプロビジョニングを容易にするサービス１６３２を、内部で使用してもよい。これらの内部共有サービスは、セキュリティ・アイデンティティサービス、統合サービス、エンタープライズリポジトリサービス、エンタープライズマネージャサービス、ウィルススキャン・ホワイトリストサービス、高可用性、バックアップリカバリサービス、クラウドサポートを可能にするサービス、Ｅメールサービス、通知サービス、ファイル転送サービスなどを含み得るが、これらに限定されない。

【０１６９】

クラウドインフラストラクチャシステム１６０２は複数のサブシステムを含み得る。これらのサブシステムは、ソフトウェア、またはハードウェア、またはそれらの組み合わせで実現され得る。図１６に示されるように、サブシステムは、クラウドインフラストラクチャシステム１６０２のユーザまたは顧客がクラウドインフラストラクチャシステム１６０２とやり取りすることを可能にするユーザインターフェイスサブシステム１６１２を含み得る。ユーザインターフェイスサブシステム１６１２は、ウェブインターフェイス１６１４、クラウドインフラストラクチャシステム１６０２が提供するクラウドサービスが宣伝広告され消費者による購入が可能なオンラインストアインターフェイス１６１６、およびその他のインターフェイス１６１８などの、各種異なるインターフェイスを含み得る。たとえば、顧客は、クライアントデバイスを用いて、クラウドインフラストラクチャシステム１６０２がインターフェイス１６１４、１６１６、および１６１８のうちの１つ以上を用いて提供する１つ以上のサービスを要求（サービス要求１６３４）してもよい。たとえば、顧客は、オンラインストアにアクセスし、クラウドインフラストラクチャシステム１６０２が提供するクラウドサービスをブラウズし、クラウドインフラストラクチャシステム１６０２が提供するとともに顧客が申し込むことを所望する１つ以上のサービスについてサブスクリプションオーダーを行い得る。このサービス要求は、顧客と、顧客が申しむことを所望する１つ以上のサービスを識別する情報を含んでいてもよい。たとえば、顧客は、クラウドインフラストラクチャシステム１６０２によって提供される正規表現の自動生成関連サービスの申し込み注文を出すことができる。

【０１７０】

図１６に示される例のような特定の実施形態において、クラウドインフラストラクチャシステム１６０２は、新しいオーダーを処理するように構成されたオーダー管理サブシステム（order management subsystem：ＯＭＳ）１６２０を含み得る。この処理の一部として、ＯＭＳ２２２０は、既に作成されていなければ顧客のアカウントを作成し、要求されたサービスを顧客に提供するために顧客に対して課金するのに使用する課金および／またはアカウント情報を顧客から受け、顧客情報を検証し、検証後、顧客のためにこのオーダーを予約し、各種ワークフローを調整することにより、プロビジョニングのためにオーダーを準備するように、構成されてもよい。

【０１７１】

適切に妥当性確認がなされると、ＯＭＳ１６２０は、処理、メモリ、およびネットワーキングリソースを含む、このオーダーのためのリソースをプロビジョニングするように構成されたオーダープロビジョニングサブシステム（ＯＰＳ）１６２４を呼び出し得る。プロビジョニングは、オーダーのためのリソースを割り当てることと、顧客オーダーが要求するサービスを容易にするようにリソースを構成することとを含み得る。オーダーのため

にリソースをプロビジョニングするやり方およびプロビジョニングされるリソースの種類は、顧客がオーダーしたクラウドサービスの種類に依存し得る。たとえば、あるワークフローに従うと、ＯＰＳ１６２４を、要求されている特定のクラウドサービスを判断し、この特定のクラウドサービスのために予め構成されたであろうポッドの数を特定するように構成されてもよい。あるオーダーのために割り当てられるポッドの数は、要求されたサービスのサイズ／量／レベル／範囲に依存し得る。たとえば、割り当てるポッドの数は、サービスがサポートすべきユーザの数、サービスが要求されている期間などに基づいて決定してもよい。次に、割り当てられたポッドを、要求されたサービスを提供するために、要求している特定の顧客に合わせてカスタマイズしてもよい。

【０１７２】

クラウドインフラストラクチャシステム１６０２は、要求されたサービスがいつ使用できるようになるかを示すために、レスポンスまたは通知１６４４を、要求している顧客に送ってもよい。いくつかの例において、顧客が、要求したサービスの利益の使用および利用を開始できるようにする情報（たとえばリンク）を顧客に送信してもよい。特定の実施形態では、正規表現の自動生成関連サービスを要求する顧客に対して、応答は、実行されるとユーザインターフェイスを表示させる命令を含み得る。

【０１７３】

クラウドインフラストラクチャシステム１６０２はサービスを複数の顧客に提供し得る。各顧客ごとに、クラウドインフラストラクチャシステム１６０２は、顧客から受けた１つ以上のサブスクリプションオーダーに関連する情報を管理し、オーダーに関連する顧客データを維持し、要求されたサービスを顧客に提供する役割を果たす。また、クラウドインフラストラクチャシステム１６０２は、申し込まれたサービスの顧客による使用に関する使用統計を収集してもよい。たとえば、統計は、使用されたストレージの量、転送されたデータの量、ユーザの数、ならびにシステムアップタイムおよびシステムダウンタイムの量などについて、収集されてもよい。この使用情報を用いて顧客に課金してもよい。課金はたとえば月ごとに行ってもよい。

【０１７４】

クラウドインフラストラクチャシステム１６０２は、サービスを複数の顧客に並列に提供してもよい。クラウドインフラストラクチャシステム１６０２は、場合によっては著作権情報を含む、これらの顧客についての情報を格納してもよい。特定の実施形態において、クラウドインフラストラクチャシステム１６０２は、顧客の情報を管理するとともに管理される情報を分離することで、ある顧客に関する情報が別の顧客に関する情報からアクセスされないようにするように構成された、アイデンティティ管理サブシステム（ＩＭＳ）１６２８を含む。ＩＭＳ１６２８は、アイデンティティサービス、情報アクセス管理、認証および許可サービス、顧客のアイデンティティおよび役割ならびに関連する能力などを管理するためのサービスなどの、各種セキュリティ関連サービスを提供するように構成されてもよい。

【０１７５】

図１７は、いくつかの例示的な実施形態に従ったコンピュータシステム１７００の例を示す。いくつかの実施形態では、コンピュータシステム１７００は、上述のシステムのいずれかを実現するために用いられ得る。図１７に示されるように、コンピュータシステム１７００は、バスサブシステム１７０２を介して他のいくつかのサブシステムと通信する処理サブシステム１７０４を含むさまざまなサブシステムを含む。これらの他のサブシステムは、処理加速ユニット１７０６、Ｉ／Ｏサブシステム１７０８、ストレージサブシステム１７１８、および通信サブシステム１７２４を含み得る。ストレージサブシステム１７１８は、記憶媒体１７２２およびシステムメモリ１７１０を含む非一時的なコンピュータ可読記憶媒体を含み得る。

【０１７６】

バスサブシステム１７０２は、コンピュータシステム１７００のさまざまなコンポーネントおよびサブシステムに意図されるように互いに通信させるための機構を提供する。バ

スサブシステム１７０２は単一のバスとして概略的に示されているが、バスサブシステムの代替例は複数のバスを利用してもよい。バスサブシステム１７０２は、さまざまなバスアーキテクチャのうちのいずれかを用いる、メモリバスまたはメモリコントローラ、周辺バス、ローカルバスなどを含むいくつかの種類のバス構造のうちのいずれかであってもよい。たとえば、このようなアーキテクチャは、業界標準アーキテクチャ（Industry Standard Architecture：ＩＳＡ）バス、マイクロチャネルアーキテクチャ（Micro Channel Architecture：ＭＣＡ）バス、エンハンストＩＳＡ（Enhanced 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（Video Electronics Standards Association：ＶＥＳＡ）ローカルバス、およびＩＥＥＥ　Ｐ１３８６．１規格に従って製造されるメザニンバスとして実現され得る周辺コンポーネントインターコネクト（Peripheral Component Interconnect：ＰＣＩ）バスなどを含み得る。

【０１７７】

処理サブシステム１７０４は、コンピュータシステム１７００の動作を制御し、１つ以上のプロセッサ、特定用途向け集積回路（ＡＳＩＣ）、またはフィールドプログラマブルゲートアレイ（ＦＰＧＡ）を含み得る。プロセッサは、シングルコアまたはマルチコアプロセッサを含み得る。コンピュータシステム１７００の処理リソースを、１つ以上の処理ユニット１７３２、１７３４などに組織することができる。処理ユニットは、１つ以上のプロセッサ、同一のまたは異なるプロセッサからの１つ以上のコア、コアとプロセッサとの組み合わせ、またはコアとプロセッサとのその他の組み合わせを含み得る。いくつかの実施形態において、処理サブシステム１７０４は、グラフィックスプロセッサ、デジタル信号プロセッサ（ＤＳＰ）などのような１つ以上の専用コプロセッサを含み得る。いくつかの実施形態では、処理サブシステム１７０４の処理ユニットの一部または全部は、特定用途向け集積回路（ＡＳＩＣ）またはフィールドプログラマブルゲートアレイ（ＦＰＧＡ）などのカスタマイズされた回路を用いて実現され得る。

【０１７８】

いくつかの実施形態において、処理サブシステム１７０４内の処理ユニットは、システムメモリ１７１０またはコンピュータ可読記憶媒体１７２２に格納された命令を実行し得る。さまざまな例において、処理ユニットはさまざまなプログラムまたはコード命令を実行し得るとともに、同時に実行する複数のプログラムまたはプロセスを維持し得る。任意の所定の時点で、実行されるべきプログラムコードの一部または全部は、システムメモリ１７１０および／または潜在的に１つ以上の記憶装置を含むコンピュータ可読記憶媒体１７２２に常駐していてもよい。適切なプログラミングを介して、処理サブシステム１７０４は、上述のさまざまな機能を提供し得る。コンピュータシステム１７００が１つ以上の仮想マシンを実行している例において、１つ以上の処理ユニットが各仮想マシンに割り当てられてもよい。

【０１７９】

特定の実施形態において、コンピュータシステム１７００によって実行される全体的な処理を加速するように、カスタマイズされた処理を実行するために、または処理サブシステム１７０４によって実行される処理の一部をオフロードするために、処理加速ユニット１７０６が任意に設けられ得る。

【０１８０】

Ｉ／Ｏサブシステム１７０８は、コンピュータシステム１７００に情報を入力するための、および／またはコンピュータシステム１７００から、もしくはコンピュータシステム１７００を介して、情報を出力するための、デバイスおよび機構を含み得る。一般に、入力デバイスという語の使用は、コンピュータシステム１７００に情報を入力するためのすべての考えられ得る種類のデバイスおよび機構を含むよう意図される。ユーザインターフェイス入力デバイスは、たとえば、キーボード、マウスまたはトラックボールなどのポインティングデバイス、ディスプレイに組み込まれたタッチパッドまたはタッチスクリーン、スクロールホイール、クリックホイール、ダイアル、ボタン、スイッチ、キーパッド、音声コマンド認識システムを伴う音声入力デバイス、マイクロフォン、および他の種類の

入力デバイスを含んでもよい。ユーザインターフェイス入力デバイスは、ユーザが入力デバイスを制御しそれと対話することを可能にするＭｉｃｒｏｓｏｆｔ Ｋｉｎｅｃｔ（登録商標）モーションセンサ、Ｍｉｃｒｏｓｏｆｔ Ｘｂｏｘ（登録商標）３６０ゲームコントローラ、ジェスチャおよび音声コマンドを用いる入力を受信するためのインターフェイスを備えたデバイスなど、モーション検知および／またはジェスチャ認識デバイスも含んでもよい。ユーザインターフェイス入力デバイスは、ユーザから目の動き（たとえば、写真を撮っている間および／またはメニュー選択を行っている間の「まばたき」）を検出し、アイジェスチャを入力デバイス（たとえばＧｏｏｇｌｅ Ｇｌａｓｓ（登録商標））への入力として変換するＧｏｏｇｌｅ Ｇｌａｓｓ（登録商標）瞬き検出器などのアイジェスチャ認識デバイスも含んでもよい。また、ユーザインターフェイス入力デバイスは、ユーザが音声コマンドを介して音声認識システム（たとえばＳｉｒｉ（登録商標）ナビゲータ）と対話することを可能にする音声認識感知デバイスを含んでもよい。

【０１８１】

ユーザインターフェイス入力デバイスの他の例は、三次元（３Ｄ）マウス、ジョイスティックまたはポインティングスティック、ゲームパッドおよびグラフィックタブレット、ならびにスピーカ、デジタルカメラ、デジタルカムコーダ、ポータブルメディアプレーヤ、ウェブカム、画像スキャナ、指紋スキャナ、バーコードリーダ３Ｄスキャナ、３Ｄプリンタ、レーザレンジファインダ、および視線追跡デバイスなどの聴覚／視覚デバイスを含むが、それらに限定されない。加えて、ユーザインターフェイス入力デバイスは、たとえば、コンピュータ断層撮影、磁気共鳴撮像、陽電子放出断層撮影、および医療用超音波検査デバイスなどの医療用画像化入力デバイスを含んでもよい。ユーザインターフェイス入力デバイスは、たとえば、ＭＩＤＩキーボード、デジタル楽器などの音声入力デバイスも含んでもよい。

【０１８２】

一般に、出力デバイスという語の使用は、コンピュータシステム１７００からユーザまたは他のコンピュータに情報を出力するための考えられるすべての種類のデバイスおよび機構を含むことを意図している。ユーザインターフェイス出力デバイスは、ディスプレイサブシステム、インジケータライト、または音声出力デバイスなどのような非視覚化ディスプレイなどを含み得る。ディスプレイサブシステムは、陰極線管（ＣＲＴ）、液晶ディスプレイ（ＬＣＤ）またはプラズマディスプレイを使うものなどのフラットパネルデバイス、投影デバイス、タッチスクリーンなどであってもよい。たとえば、ユーザインターフェイス出力デバイスは、モニタ、プリンタ、スピーカ、ヘッドフォン、自動車ナビゲーションシステム、プロッタ、音声出力デバイスおよびモデムなどの、テキスト、グラフィックスおよび音声／映像情報を視覚的に伝えるさまざまな表示デバイスを含み得るが、それらに限定されない。

【０１８３】

ストレージサブシステム１７１８は、コンピュータシステム１７００によって使用される情報およびデータを格納するためのリポジトリまたはデータストアを提供する。ストレージサブシステム１７１８は、いくつかの例の機能を提供する基本的なプログラミングおよびデータ構成を格納するための有形の非一時的なコンピュータ可読記憶媒体を備える。処理サブシステム１７０４によって実行されると上述の機能を提供するソフトウェア（たとえばプログラム、コードモジュール、命令）が、ストレージサブシステム１７１８に格納されてもよい。ソフトウェアは、処理サブシステム１７０４の１つ以上の処理ユニットによって実行されてもよい。ストレージサブシステム１７１８はまた、本開示の教示に従って使用されるデータを格納するためのリポジトリを備えてもよい。

【０１８４】

ストレージサブシステム１７１８は、揮発性および不揮発性のメモリデバイスを含む１つ以上の非一時的メモリデバイスを含み得る。図１７に示すように、ストレージサブシステム１７１８は、システムメモリ１７１０およびコンピュータ可読記憶媒体１７２２を含む。システムメモリ１７１０は、プログラム実行中に命令およびデータを格納するための

揮発性主ランダムアクセスメモリ（ＲＡＭ）と、固定命令が格納される不揮発性読取り専用メモリ（ＲＯＭ）またはフラッシュメモリとを含む、いくつかのメモリを含み得る。いくつかの実現例において、起動中などにコンピュータシステム１７００内の要素間における情報の転送を助ける基本的なルーチンを含むベーシックインプット／アウトプットシステム（basic input/output system：ＢＩＯＳ）は、典型的には、ＲＯＭに格納されてもよい。典型的に、ＲＡＭは、処理サブシステム１７０４によって現在操作および実行されているデータおよび／またはプログラムモジュールを含む。いくつかの実現例において、システムメモリ１７１０は、スタティックランダムアクセスメモリ（ＳＲＡＭ）、ダイナミックランダムアクセスメモリ（ＤＲＡＭ）などのような複数の異なる種類のメモリを含み得る。

【０１８５】

限定ではなく一例として、図１７に示されるように、システムメモリ１７１０は、ウェブブラウザ、中間層アプリケーション、リレーショナルデータベース管理システム（ＲＤＢＭＳ）などのような各種アプリケーションを含み得る、実行中のアプリケーションプログラム１７１２、プログラムデータ１７１４、およびオペレーティングシステム１７１６を、ロードしてもよい。一例として、オペレーティングシステム１７１６は、さまざまなバージョンのＭｉｃｒｏｓｏｆｔ Ｗｉｎｄｏｗｓ（登録商標）、Ａｐｐｌｅ Ｍａｃｉｎｔｏｓｈ（登録商標）および／もしくはＬｉｎｕｘオペレーティングシステム、市販されているさまざまなＵＮＩＸ（登録商標）もしくはＵＮＩＸ系オペレーティングシステム（さまざまなＧＮＵ／Ｌｉｎｕｘオペレーティングシステム、Ｇｏｏｇｌｅ Ｃｈｒｏｍｅ（登録商標）ＯＳなどを含むがそれらに限定されない）、ならびに／または、ｉＯＳ（登録商標）、Ｗｉｎｄｏｗｓ（登録商標） Ｐｈｏｎｅ、Ａｎｄｒｏｉｄ（登録商標） ＯＳ、ＢｌａｃｋＢｅｒｒｙ（登録商標） ＯＳ、Ｐａｌｍ（登録商標） ＯＳオペレーティングシステムのようなモバイルオペレーティングシステムなどを、含み得る。

【０１８６】

コンピュータ可読記憶媒体１７２２は、いくつかの例の機能を提供するプログラミングおよびデータ構成を格納し得る。コンピュータ可読記憶媒体１７２２は、コンピュータシステム１７００のための、コンピュータ可読命令、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を提供し得る。処理サブシステム１７０４によって実行されると上記機能を提供するソフトウェア（プログラム、コードモジュール、命令）はストレージサブシステム１７１８に格納されてもよい。一例として、コンピュータ可読記憶媒体１７２２は、ハードディスクドライブ、磁気ディスクドライブ、ＣＤ ＲＯＭ、ＤＶＤ、Ｂｌｕ－Ｒａｙ（登録商標）ディスクなどの光ディスクドライブ、またはその他の光学媒体のような不揮発性メモリを含み得る。コンピュータ可読記憶媒体１７２２は、Ｚｉｐ（登録商標）ドライブ、フラッシュメモリカード、ユニバーサルシリアルバス（ＵＳＢ）フラッシュドライブ、セキュアデジタル（ＳＤ）カード、ＤＶＤディスク、デジタルビデオテープなどを含み得るが、それらに限定されない。コンピュータ可読記憶媒体１７２２はまた、フラッシュメモリベースのＳＳＤ、エンタープライズフラッシュドライブ、ソリッドステートＲＯＭなどのような不揮発性メモリに基づくソリッドステートドライブ（ＳＳＤ）、ソリッドステートＲＡＭ、ダイナミックＲＡＭ、スタティックＲＡＭのような揮発性メモリに基づくＳＳＤ、ＤＲＡＭベースのＳＳＤ、磁気抵抗ＲＡＭ（ＭＲＡＭ）ＳＳＤ、およびＤＲＡＭとフラッシュメモリベースのＳＳＤとの組み合わせを使用するハイブリッドＳＳＤを含み得る。

【０１８７】

特定の実施形態において、ストレージサブシステム１７１８は、コンピュータ可読記憶媒体１７２２にさらに接続され得るコンピュータ可読記憶媒体リーダ１７２０も含み得る。リーダ１７２０は、ディスク、フラッシュドライブなどのメモリデバイスからデータを受取り得るとともに、当該データを読取るように構成され得る。

【０１８８】

特定の実施形態において、コンピュータシステム１７００は、処理およびメモリリソー

スの仮想化を含むがこれに限定されない仮想化技術をサポートし得る。たとえば、コンピュータシステム１７００は、１つ以上の仮想マシンを実行するためのサポートを提供し得る。特定の実施形態において、コンピュータシステム１７００は、仮想マシンの構成および管理を容易にするハイパーバイザなどのプログラムを実行し得る。各仮想マシンには、メモリ、演算（たとえばプロセッサ、コア）、Ｉ／Ｏ、およびネットワーキングリソースが割当てられてもよい。各仮想マシンは通常、他の仮想マシンから独立して実行される。仮想マシンは、典型的には、コンピュータシステム１７００によって実行される他の仮想マシンによって実行されるオペレーティングシステムと同じであり得るかまたは異なり得るそれ自体のオペレーティングシステムを実行する。したがって、潜在的に複数のオペレーティングシステムがコンピュータシステム１７００によって同時に実行され得る。

【０１８９】

通信サブシステム１７２４は、他のコンピュータシステムおよびネットワークに対するインターフェイスを提供する。通信サブシステム１７２４は、他のシステムとコンピュータシステム１７００との間のデータの送受のためのインターフェイスとして機能する。たとえば、通信サブシステム１７２４は、コンピュータシステム１７００が、１つ以上のクライアントデバイスとの間で情報を送受信するために、インターネットを介して１つ以上のクライアントデバイスへの通信チャネルを確立することを可能にし得る。

【０１９０】

通信サブシステム１７２４は、有線および／または無線の通信プロトコルの両方をサポートし得る。特定の実施形態において、通信サブシステム１７２４は、（たとえば、セルラー電話技術、３Ｇ、４ＧもしくはＥＤＧＥ（グローバル進化のための高速データレート）などの先進データネットワーク技術、ＷｉＦｉ（ＩＥＥＥ８０２．ＸＸファミリー規格、もしくは他のモバイル通信技術、またはそれらのいずれかの組み合わせを用いて）無線音声および／またはデータネットワークにアクセスするための無線周波数（ＲＦ）送受信機コンポーネント、グローバルポジショニングシステム（ＧＰＳ）受信機コンポーネント、および／または、他のコンポーネントを含み得る。いくつかの実施形態において、通信サブシステム１７２４は、無線インターフェイスに加えてまたはその代わりに、有線ネットワーク接続（たとえばEthernet（登録商標））を提供し得る。

【０１９１】

通信サブシステム１７２４は、さまざまな形式でデータを受信および送信し得る。いくつかの実施形態において、通信サブシステム１７２４は、他の形式に加えて、構造化および／または非構造化されたデータフィード１７２６、イベントストリーム１７２８、イベントアップデート１７３０などの形式で入力通信を受信してもよい。たとえば、通信サブシステム１７２４は、ソーシャルメディアネットワークおよび／またはＴｗｉｔｔｅｒ（登録商標）フィード、Ｆａｃｅｂｏｏｋ（登録商標）アップデート、Ｒｉｃｈ　Ｓｉｔｅ　Ｓｕｍｍａｒｙ（ＲＳＳ）フィードなどのウェブフィード、および／または１つ以上の第三者情報源からのリアルタイムアップデートなどのような他の通信サービスのユーザから、リアルタイムでデータフィード１７２６を受信（または送信）するように構成されてもよい。

【０１９２】

特定の実施形態において、通信サブシステム１７２４は、連続データストリームの形式でデータを受信するように構成されてもよく、当該連続データストリームは、明確な終端を持たない、本来は連続的または無限であり得るリアルタイムイベントのイベントストリーム１７２８および／またはイベントアップデート１７３０を含んでもよい。連続データを生成するアプリケーションの例としては、たとえば、センサデータアプリケーション、金融株式相場表示板、ネットワーク性能測定ツール（たとえばネットワーク監視およびトラフィック管理アプリケーション）、クリックストリーム解析ツール、自動車交通監視などが含まれ得る。

【０１９３】

通信サブシステム１７２４は、コンピュータシステム１７００からのデータを他のコン

ピュータシステムまたはネットワークに伝えるように構成されてもよい。このデータは、構造化および／または非構造化データフィード１７２６、イベントストリーム１７２８、イベントアップデート１７３０などのようなさまざまな異なる形式で、コンピュータシステム１７００に結合された１つ以上のストリーミングデータソースコンピュータと通信し得る１つ以上のデータベースに、伝えられてもよい。

【０１９４】

コンピュータシステム１７００は、ハンドヘルドポータブルデバイス（たとえばｉＰｈｏｎｅ（登録商標）セルラーフォン、ｉＰａｄ（登録商標）コンピューティングタブレット、ＰＤＡ）、ウェアラブルデバイス（たとえばＧｏｏｇｌｅ　Ｇｌａｓｓ（登録商標）ヘッドマウントディスプレイ）、パーソナルコンピュータ、ワークステーション、メインフレーム、キオスク、サーバラック、またはその他のデータ処理システムを含む、さまざまな種類のうちの１つであり得る。コンピュータおよびネットワークの性質が常に変化しているため、図１７に示されるコンピュータシステム１７００の記載は具体例として意図されているに過ぎない。図１７に示されるシステムよりも多くのコンポーネントまたは少ないコンポーネントを有するその他多くの構成が実現可能である。当業者であれば、本明細書における開示および教示に基づいて、さまざまな例を実現するための他の態様および／または方法を認識するだろう。

【０１９５】

特定の例について説明してきたが、さまざまな変形、変更、代替構成、および均等物が可能である。例は、特定のデータ処理環境内の動作に限定されず、複数のデータ処理環境内で自由に運用することができる。加えて、いくつかの例を特定の一連のトランザクションおよびステップを用いて説明してきたが、これが限定を意図しているのではないことは当業者には明らかであるはずである。いくつかのフローチャートは動作を逐次的プロセスとして説明しているが、これらの動作のうちの多くは並列または同時に実行されてもよい。加えて、動作の順序を並べ替えてもよい。プロセスは図に含まれない追加のステップを有し得る。上記の例の各種特徴および局面は、個別に使用されてもよく、またはともに使用されてもよい。

【０１９６】

さらに、いくつかの例をハードウェアとソフトウェアとの特定の組み合わせを用いて説明してきたが、ハードウェアとソフトウェアとの他の組み合わせも可能であることが理解されるはずである。いくつかの例は、ハードウェアでのみ、またはソフトウェアでのみ、またはそれらの組み合わせを用いて実現されてもよい。本明細書に記載されたさまざまなプロセスは、同じプロセッサまたは任意の組み合わせの異なるプロセッサ上で実現されてもよい。

【０１９７】

デバイス、システム、コンポーネントまたはモジュールが特定の動作または機能を実行するように構成されると記載されている場合、そのような構成は、たとえば、動作を実行するように電子回路を設計することにより、動作を実行するようにプログラミング可能な電子回路（マイクロプロセッサなど）をプログラミングすることにより、たとえば、非一時的なメモリ媒体に格納されたコードもしくは命令またはそれらの任意の組み合わせを実行するようにプログラミングされたコンピュータ命令もしくはコード、またはプロセッサもしくはコアを実行するなどにより、達成され得る。プロセスは、プロセス間通信のための従来の技術を含むがこれに限定されないさまざまな技術を使用して通信してもよく、異なる対のプロセスは異なる技術を使用してもよく、同じ対のプロセスは異なる時間に異なる技術を使用してもよい。

【０１９８】

上述の例が十分に理解されるように、本開示では具体的な詳細が示されている。しかしながら、例はこれらの具体的な詳細がなくとも実施され得るものである。たとえば、周知の回路、プロセス、アルゴリズム、構造、および技術は、例が曖昧になるのを避けるために不必要な詳細事項なしで示している。本明細書は例示的な例のみを提供し、他の例の範

囲、適用可能性、または構成を限定するよう意図されたものではない。むしろ、例についての上記説明は、各種例を実現することを可能にする説明を当業者に提供するだろう。要素の機能および構成の範囲内でさまざまな変更が可能である。

【０１９９】

したがって、明細書および図面は、限定的な意味ではなく例示的なものとみなされるべきである。しかしながら、請求項に記載されているより広範な精神および範囲から逸脱することなく、追加、削減、削除、ならびに他の修正および変更がこれらになされ得ることは明らかであろう。このように、具体的な例を説明してきたが、これらは限定を意図するものではない。さまざまな変形例および同等例は添付の特許請求の範囲内にある。

【０２００】

上記の明細書では、本開示の局面についてその具体的な例を参照して説明しているが、本開示がそれに限定されるものではないことを当業者は認識するだろう。上記の開示のさまざまな特徴および局面は、個々にまたは一緒に用いられてもよい。さらに、例は、明細書のさらに広い精神および範囲から逸脱することなく、本明細書に記載されているものを超えて、さまざまな環境および用途で利用することができる。したがって、明細書および図面は、限定的ではなく例示的であると見なされるべきである。

【０２０１】

上記の説明では、例示の目的で、方法を特定の順序で記載した。代替例では、方法は記載された順序とは異なる順序で実行されてもよいことを理解されたい。また、上記の方法は、ハードウェアコンポーネントによって実行されてもよいし、マシン実行可能命令のシーケンスで具体化されてもよいことも理解されたい。当該マシン実行可能命令は、当該命令でプログラムされた汎用もしくは専用のプロセッサまたは論理回路などのマシンに上記方法を実行させるために用いられ得るものである。これらのマシン実行可能命令は、ＣＤ－ＲＯＭもしくは他の種類の光ディスク、フロッピー（登録商標）ディスク、ＲＯＭ、ＲＡＭ、ＥＰＲＯＭ、ＥＥＰＲＯＭ、磁気もしくは光学カード、フラッシュメモリのような、１つ以上の機械可読媒体、または電子命令を格納するのに適した他の種類の機械可読媒体に保存され得る。代替的には、これらの方法は、ハードウェアとソフトウェアとの組み合わせによって実行されてもよい。

【０２０２】

構成要素が特定の動作を実行するように構成されるとして記載されている場合、そのような構成は、たとえば、特定の動作を実行するよう電子回路もしくは他のハードウェアを設計することによって、特定の動作を実行するようプログラミング可能な電子回路（たとえばマイクロプロセッサもしくは他の好適な電子回路）をプログラミングすることによって、またはそれらの任意の組み合わせによって達成されてもよい。

【０２０３】

本願の説明のための例をここに詳細に記載したが、本発明の概念が他の態様で様々に具現化および採用され得ること、ならびに特許請求の範囲が、先行技術によって限定される場合を除き、そのような変形例を含むように解釈されるよう意図されていることを理解されたい。

【０２０４】

構成要素が特定の動作を実行する「ように構成される」ものとして記載されている場合、そのような構成は、たとえば、特定の動作を実行するよう電子回路もしくは他のハードウェアを設計することによって、特定の動作を実行するようプログラミング可能な電子回路（たとえばマイクロプロセッサもしくは他の好適な電子回路）をプログラミングすることによって、またはそれらの任意の組み合わせによって達成されてもよい。

【図面】

【図１】

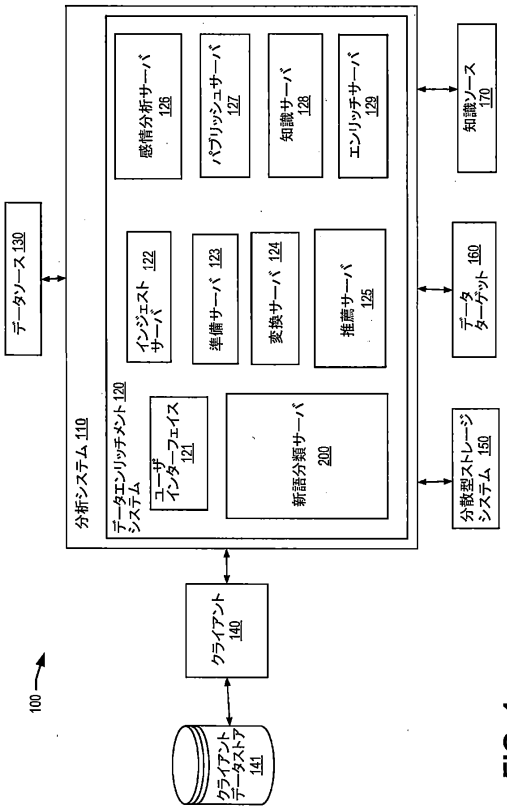


FIG. 1

【図２】

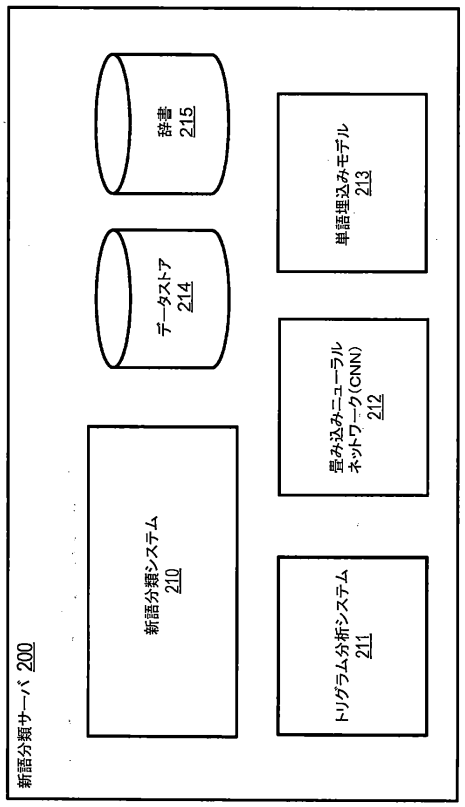


FIG. 2

【図３】

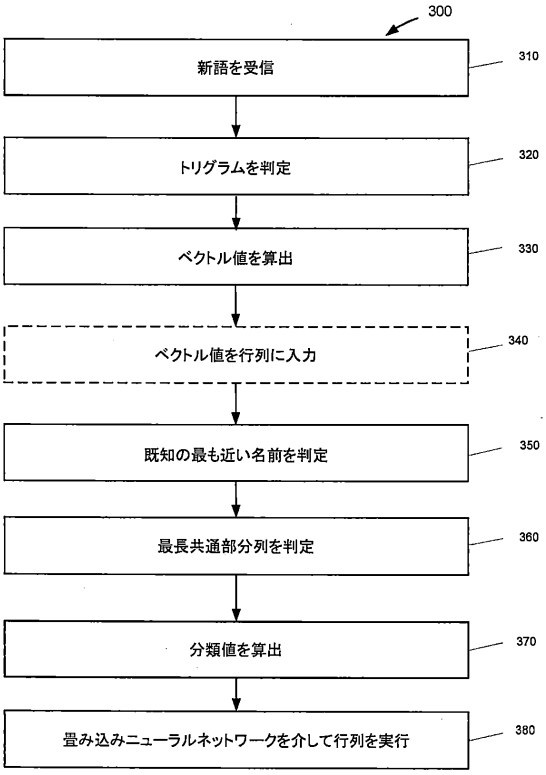


FIG. 3

【図４】

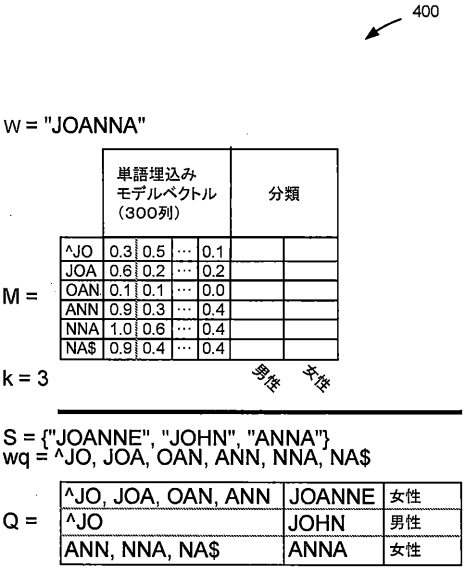


FIG. 4

【図５】

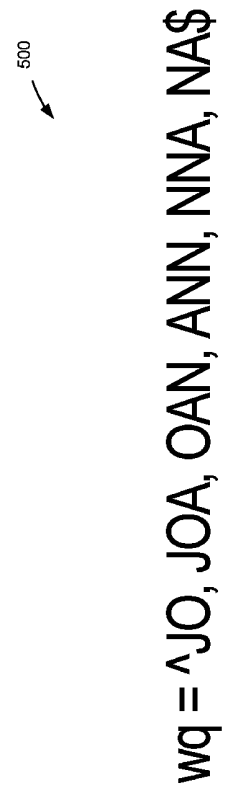


FIG. 5

【図６】

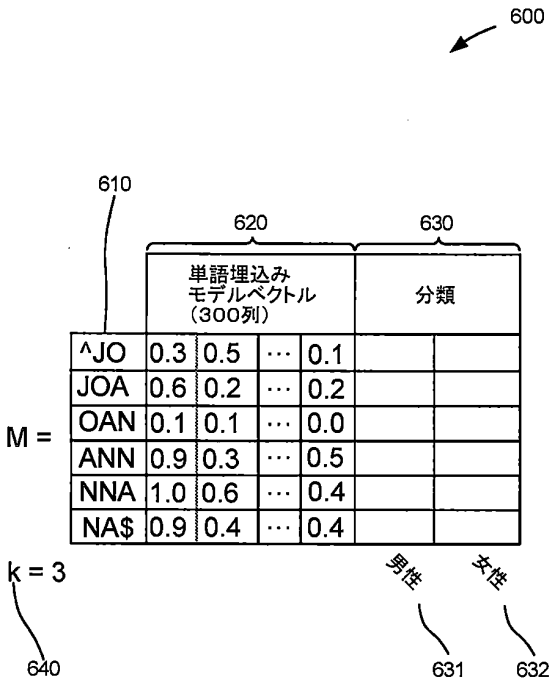


FIG. 6

【図７】

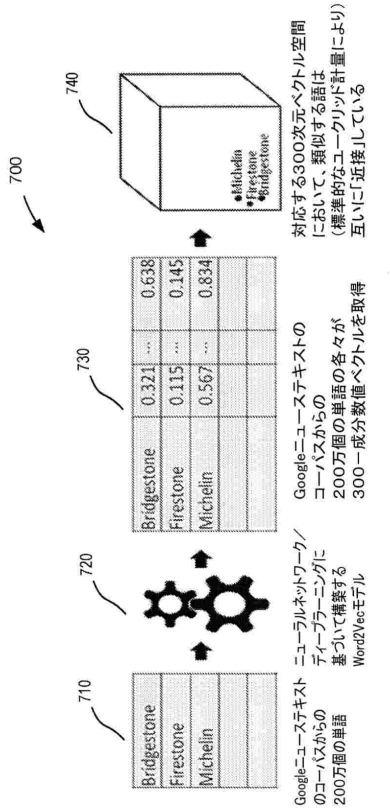


FIG. 7

【図８】



FIG. 8

【図９】

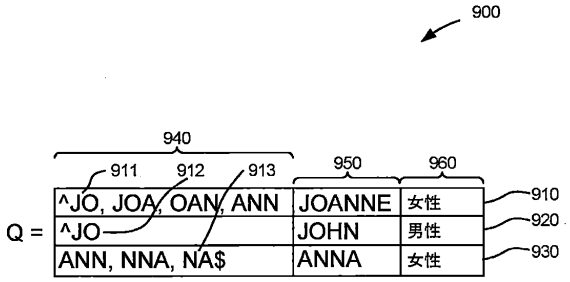


FIG. 9

【図１０】

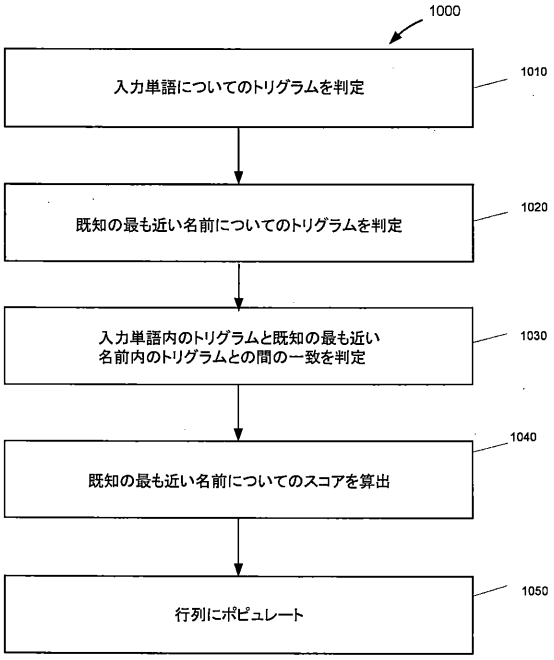


FIG. 10

【図１１】

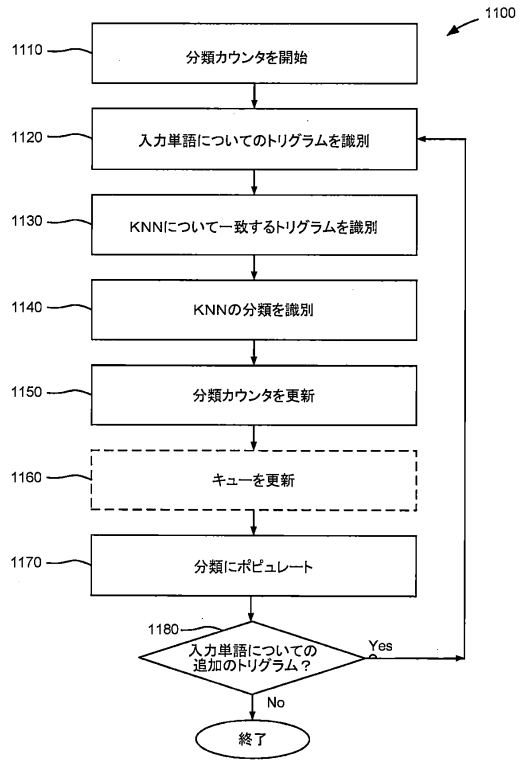


FIG. 11

【図１２】

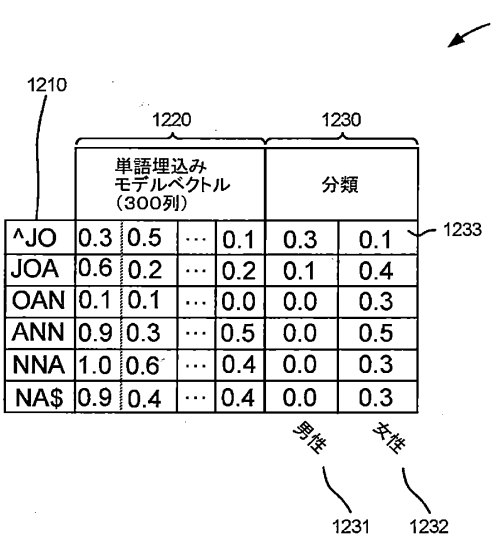


FIG. 12

【図１３】

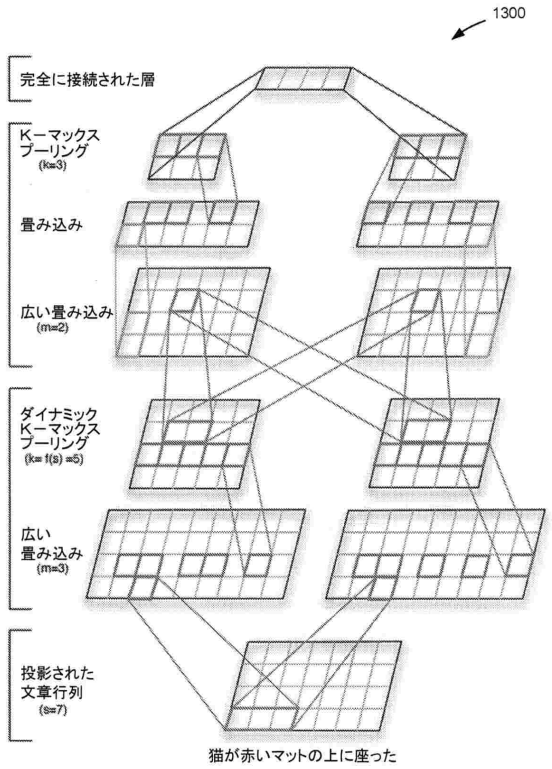


FIG. 13

【図１４】

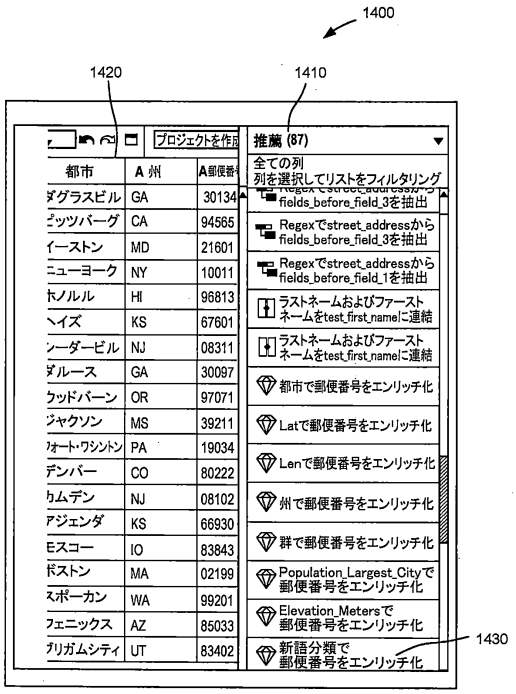


FIG. 14

【図１５】

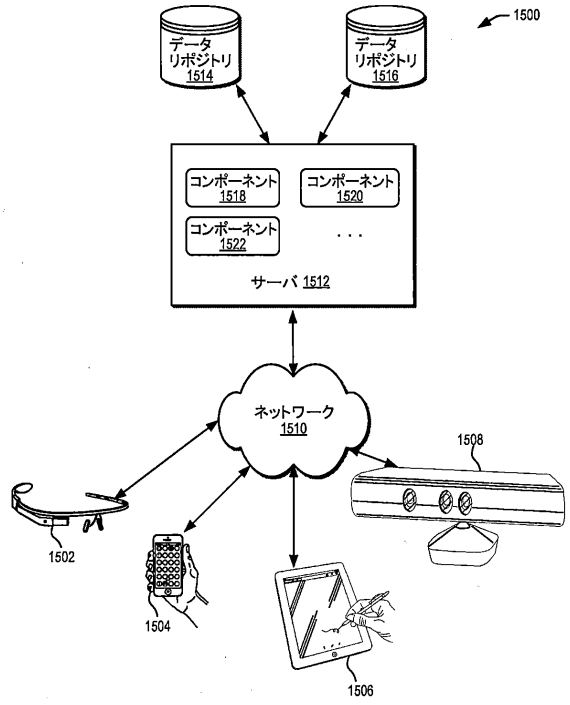


FIG. 15

【図１６】

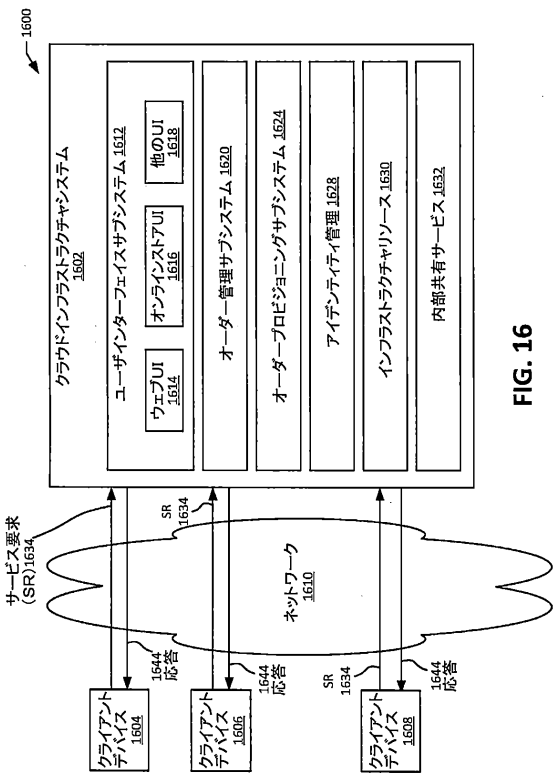


FIG. 16

【図１７】

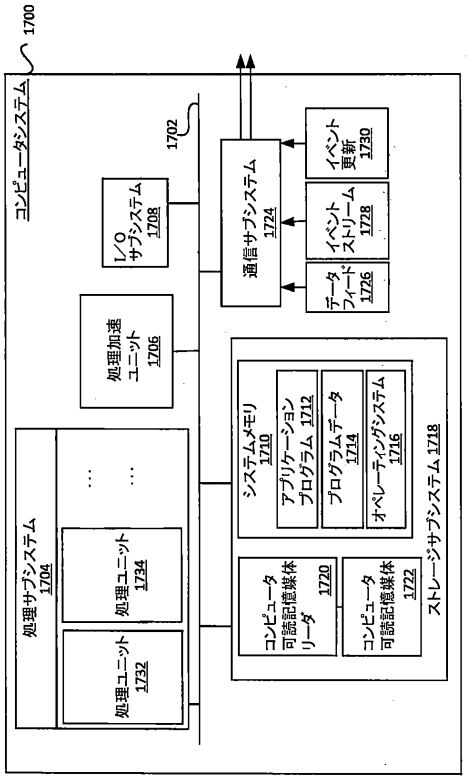

1700
コンピュータシステム
1702
処理サブシステム
1704
処理ユニット
1732
処理ユニット
1734
処理加速
ユニット
1706
I/O
サブシステム
1708
システムメモリ
1710
アプリケーション
プログラム 1712
プログラムデータ
1714
オペレーティングシステム
1716
コンピュータ
可読記憶媒体
リーダ
1720
コンピュータ
可読記憶媒体
1722
ストレージサブシステム 1718
通信サブシステム
1724
データ
フィード
1726
イベント
ストリーム
1728
イベント
更新
1730

FIG. 17

フロントページの続き

(33)優先権主張国・地域又は機関
米国(US)
ルニア州、レッドウッド・シティー、オラクル・パークウェイ、５００
(72)発明者　クライダー，マーク・リー
アメリカ合衆国、９４０６５ カリフォルニア州、レッドウッド・シティー、オラクル・パークウェイ、５００
審査官　酒井 恭信
(56)参考文献　特開２００４－３４８２３９（ＪＰ，Ａ）
国際公開第２０１９／０９２８６８（WO，Ａ１）
特開２０１１－２４８４８３（ＪＰ，Ａ）
特開２０１３－２１８６２８（ＪＰ，Ａ）
米国特許出願公開第２００９／００４３７２０（ＵＳ，Ａ１）
(58)調査した分野　(Int.Cl.，ＤＢ名)
Ｇ０６Ｆ　１６／００ － １６／９５８
Ｇ０６Ｆ　４０／００ － ４０／５８