

【公報種別】特許法第17条の2の規定による補正の掲載

【部門区分】第6部門第3区分

【発行日】令和4年6月29日(2022.6.29)

【国際公開番号】WO2020/005986

【公表番号】特表2021-529385(P2021-529385A)

【公表日】令和3年10月28日(2021.10.28)

【出願番号】特願2020-572428(P2020-572428)

【国際特許分類】

G 0 6 F 1 6 / 9 0 (2 0 1 9 . 0 1)

G 0 6 F 1 6 / 9 0 3 2 (2 0 1 9 . 0 1)

G 0 6 F 1 3 / 0 0 (2 0 0 6 . 0 1)

【F I】

G 0 6 F 1 6 / 9 0 1 0 0

G 0 6 F 1 6 / 9 0 3 2

G 0 6 F 1 3 / 0 0 5 5 0 A

10

【手続補正書】

【提出日】令和4年6月21日(2022.6.21)

【手続補正1】

【補正対象書類名】特許請求の範囲

【補正対象項目名】全文

【補正方法】変更

【補正の内容】

【特許請求の範囲】

【請求項1】

複数のエンティティーの間の関係と対話するためのコンピュータプログラムであって、当該コンピュータプログラムは、非一時的なコンピュータ読み取り可能な媒体の中に具現化されているコンピュータ実行可能なコードを含み、1つ又は複数のコンピューティングデバイスにおいて実行するときに、

30

検索エンジンのためのユーザインターフェイスの検索入力フォームフィールドを介してユーザからキーワード検索を受信するステップと、

前記キーワード検索に同一指示アルゴリズムを適用して、前記ユーザが意図しているとともに、1つ又は複数のエンティティー関連のグループの言及によって特徴付けられるエンティティーを予測するステップと、

前記ユーザインターフェイスの中で前記ユーザに第1の検索結果を提示するステップであって、前記第1の検索結果は、第1の複数のエンティティータグ及びドキュメントの集合のうちの複数のドキュメントを含み、前記複数のドキュメントのうちの各々のドキュメントは、前記1つ又は複数のエンティティー関連のグループの言及のうちの少なくとも1つのうちで前記検索エンジンが配置する言及を含み、前記複数のエンティティータグのうちの各々のエンティティータグは、前記言及のうちの少なくとも1つに対応する、ステップと、

40

前記第1の複数のエンティティータグのうちで第1のエンティティー関連のグループの言及によって特徴付けられる第1のエンティティーのための第1のエンティティータグの前記ユーザからの第1の選択を受信するステップであって、前記第1の選択は、前記キーワード検索が前記第1のエンティティーを参照することを意図していたというユーザの確認を示す、ステップと、

前記検索入力フォームフィールドに前記第1のエンティティータグを追加するステップと、

、
関連性アルゴリズムを適用して、1つ又は複数の他のエンティティー関連のグループの言

50

及によって特徴付けられる1つ又は複数の他のエンティティと前記第1のエンティティとの間の関係の証拠を提供する第2の複数のドキュメントの中のテキストを識別するステップと、

各々が前記ユーザインターフェイスの中の前記1つ又は複数の他のエンティティのうちの1つに対応する第2の複数のエンティティータグを提示するステップと、

前記第2の複数のエンティティータグのうちの第2のエンティティータグの前記ユーザから第2の選択を受信するステップと、

前記第1のエンティティと前記第2のエンティティとの間の関係を具体化する前記第2の複数のドキュメントのうちの1つ又は複数を含む第2の検索結果を提示するステップと、を実行する、

10

コンピュータプログラム。

【請求項2】

前記ユーザインターフェイスの中に、前記第1のエンティティの第1のアイコン、前記第2のエンティティの第2のアイコン、及び前記第2のアイコンに前記第1のアイコンを視覚的に結合するコネクタとして、前記関係をグラフィカルに表示するステップを実行するコードをさらに含む、請求項1に記載のコンピュータプログラム。

【請求項3】

関係オペレータによって前記第2のエンティティータグに結合されている前記第1のエンティティータグとして、前記検索入力フォームフィールドの中の前記関係を表示するステップを実行するコードをさらに含む、請求項1に記載のコンピュータプログラム。

20

【請求項4】

前記関係オペレータは、シンボル"< >"を含む、請求項3に記載のコンピュータプログラム。

【請求項5】

前記ドキュメントの集合は、前記キーワード検索が受信されるユーザデバイスにおいてローカルにホストされるドキュメントを含む、請求項1に記載のコンピュータプログラム。

【請求項6】

前記ドキュメントの集合は、広域ネットワークにおいて配信されるドキュメントを含む、請求項1に記載のコンピュータプログラム。

【請求項7】

30

前記複数のドキュメントは、リモートクラウド記憶装置においてホストされているドキュメントを含む、請求項1に記載のコンピュータプログラム。

【請求項8】

前記第2の検索結果を提示するステップは、グラフの中の端部によって、各々が前記第1のエンティティに接続される複数のエンティティータグをグラフィカルに提示するステップを含む、請求項1に記載のコンピュータプログラム。

【請求項9】

方法であって、

検索エンジンのためのユーザインターフェイスの検索入力フォームフィールドにおいてユーザからキーワード検索を受信するステップと、

40

前記ユーザが意図している1つ又は複数のエンティティを予測するステップと、

前記ユーザインターフェイスの中で前記ユーザに第1の検索結果を提示するステップであって、前記第1の検索結果は、第1の複数のエンティティータグ及び複数のドキュメントを含み、前記複数のドキュメントのうちの各々のドキュメントは、前記1つ又は複数のエンティティのうちの少なくとも1つの前記検索エンジンが配置する言及を含み、前記複数のエンティティータグのうちの各々のエンティティータグは、前記言及のうちの少なくとも1つに対応する、ステップと、

前記第1の複数のエンティティータグのうちの第1のエンティティのための第1のエンティティータグの前記ユーザからの第1の選択を受信するステップであって、前記第1の選択は、前記キーワード検索が前記第1のエンティティを参照することを意図していた

50

というユーザの確認を示す、ステップと、
前記検索入力フォームフィールドに前記第1のエンティティータグを追加するステップと、
、
1つ又は複数の他のエンティティータグと前記第1のエンティティータグとの間の関係の証拠を提供する第2の複数のドキュメントの中のテキストを識別するステップと、
前記ユーザインターフェイスの中の前記1つ又は複数の他のエンティティータグのための第2の複数のエンティティータグを提示するステップと、
前記第2の複数のエンティティータグのうちの第2のエンティティータグのための第2のエンティティータグの前記ユーザから第2の選択を受信するステップと、
前記第1のエンティティータグと前記第2のエンティティータグとの間の関係を具体化する前記第2の複数のドキュメントのうちの1つ又は複数を含む第2の検索結果を提示するステップと、を含む、
方法。

10

【請求項10】

前記ユーザインターフェイスの中に、前記第1のエンティティータグの第1のアイコン、前記第2のエンティティータグの第2のアイコン、及び前記第2のアイコンに前記第1のアイコンを視覚的に結合するコネクタとして、前記関係をグラフィカルに表示するステップをさらに含む、請求項9に記載の方法。

【請求項11】

関係オペレータによって前記第2のエンティティータグに結合されている前記第1のエンティティータグとして、前記検索入力フォームフィールドの中の前記関係を表示するステップをさらに含む、請求項9に記載の方法。

20

【請求項12】

前記関係オペレータは、シンボル"< >"を含む、請求項11に記載の方法。

【請求項13】

前記キーワード検索において前記ユーザが意図している1つ又は複数のエンティティータグを予測するステップは、前記キーワード検索に同一指示アルゴリズムを適用して、前記ユーザが意図している1つ又は複数のエンティティータグを予測するステップを含む、請求項9に記載の方法。

【請求項14】

1つ又は複数の他のエンティティータグと前記第1のエンティティータグとの間の関係の証拠を提供する前記複数のドキュメントの中のテキストを識別するステップは、関連性アルゴリズムを適用して、1つ又は複数の他のエンティティータグと前記第1のエンティティータグとの間の関係の証拠を提供する前記複数のドキュメントの中のテキストを識別するステップを含む、請求項9に記載の方法。

30

【請求項15】

前記複数のドキュメントは、前記キーワード検索が受信されるユーザデバイスにおいてローカルにホストされるとともにドキュメントの集合から得られるドキュメントを含む、請求項9に記載の方法。

【請求項16】

前記複数のドキュメントは、広域ネットワークにおいて配信されるとともにドキュメントの集合から得られるドキュメントを含む、請求項9に記載の方法。

40

【請求項17】

前記複数のドキュメントは、リモートクラウド記憶装置においてホストされているとともにドキュメントの集合から得られるドキュメントを含む、請求項9に記載の方法。

【請求項18】

前記第1のエンティティータグと前記第2のエンティティータグとの間の関係を特徴付ける関係記述を格納するステップをさらに含む、請求項9に記載の方法。

【請求項19】

前記第1のエンティティータグの第1のアイコン、前記第2のエンティティータグの第2のアイコン

50

、及び前記関係を示すために前記第1のアイコンを前記第2のアイコンに視覚的に結合するコネクタをグラフィカルに表示するステップをさらに含み、前記コネクタは、前記第2の複数のドキュメントの中の前記関係の証拠へのユーザによるアクセスを提供するユーザインターフェイス要素を含む、請求項9に記載の方法。

【請求項20】

システムであって、

検索エンジンと、

データネットワークと、

前記データネットワークを通じて前記検索エンジンに結合されるコンピューティングデバイスであって、前記コンピューティングデバイスは、プロセッサ、メモリ、及びディスプレイを含み、前記メモリは、前記プロセッサによって実行可能であるコードを格納し、前記プロセッサは、ユーザインターフェイスの検索入力フォームフィールドにおいてユーザからキーワード検索を受信するステップと、前記ユーザが意図している1つ又は複数のエンティティを予測するステップと、前記ユーザインターフェイスの中で前記ユーザに第1の検索結果を提示するステップであって、前記第1の検索結果は、第1の複数のエンティティタグ及び複数のドキュメントを含み、前記複数のドキュメントのうちの各々のドキュメントは、前記1つ又は複数のエンティティのうちの少なくとも1つのエンティティの前記検索エンジンが配置する言及を含み、前記複数のエンティティタグのうちの各々のエンティティタグは、前記言及のうちの少なくとも1つに対応する、ステップと、前記第1の複数のエンティティタグのうちの第1のエンティティのための第1のエンティティタグの前記ユーザからの第1の選択を受信するステップであって、前記第1の選択は、前記キーワード検索が前記第1のエンティティを参照することを意図していたというユーザの確認を示す、ステップと、前記検索入力フォームフィールドに前記第1のエンティティタグを追加するステップと、1つ又は複数の他のエンティティと前記第1のエンティティとの間の関係の証拠を提供する第2の複数のドキュメントの中のテキストを識別するステップと、前記ユーザインターフェイスの中の前記1つ又は複数の他のエンティティのための第2の複数のエンティティタグを提示するステップと、前記第2の複数のエンティティタグのうちの第2のエンティティのための第2のエンティティタグの前記ユーザから第2の選択を受信するステップと、前記第1のエンティティと前記第2のエンティティとの間の関係を具体化する前記第2の複数のドキュメントのうちの1つ又は複数を含む第2の検索結果を提示するステップと、を実行する、システム。

【請求項21】

複数のエンティティの間の関係と対話するためのコンピュータプログラムであって、当該コンピュータプログラムは、非一時的なコンピュータ読み取り可能な媒体の中に具現化されているコンピュータ実行可能なコードを含み、1つ又は複数のコンピューティングデバイスにおいて実行するときに、

検索エンジンのためのユーザインターフェイスを表示するステップであって、前記ユーザインターフェイスは、検索入力フォームフィールドを含む、ステップと、

前記検索入力フォームフィールドへのユーザ入力を受信するステップと、

前記ユーザ入力を構文解析して、第1のエンティティを識別する第1のテキストストリング、第2のエンティティを識別する第2のテキストストリング、及び前記第1のテキストストリングと前記第2のテキストストリングとの間のオペレータを識別するステップであって、前記オペレータは、前記第1のエンティティと前記第2のエンティティとの間の関係の証拠を検索するための前記検索エンジンへの要求を指定する、ステップと、前記オペレータに応答して、ドキュメントの集合のうちの1つ又は複数のドキュメントを検索するステップであって、前記ドキュメントの集合は、同一指示アルゴリズムが、著者が前記第1のエンティティを参照することを意図していた言及であると予測する言及、及び、同一指示アルゴリズムが、前記著者が前記第2のエンティティを参照することを意図していた言及であると予測する他の言及を含むとともに、関連性アルゴリズムが、前

10

20

30

40

50

記第1のエンティティと前記第2のエンティティとの間の関係の証拠を提供すると予測するテキストを提供する、ステップと、
前記ユーザインターフェイスの中で前記1つ又は複数のドキュメントを提示するステップと、を含む、
方法。

【手続補正2】

【補正対象書類名】明細書

【補正対象項目名】全文

【補正方法】変更

【補正の内容】

10

【発明の詳細な説明】

【技術分野】

【0001】

関連出願

本出願は、2018年6月25日に出願された米国仮特許出願第62/689,737号、2019年1月18日に出願された米国仮特許出願第62/794,177号及び2019年4月10日出願された米国仮特許出願第62/832,085号の優先権を主張する。これらの出願の各々の全内容は、参照により本明細書に組み込まれる。

【0002】

本出願はまた、2018年6月6日に出願された米国特許出願第16/001,874号及び2015年5月12日に出願された米国特許出願第14/710,342号(現在の米国特許第9,275,132号)に関連する。これらの出願の各々の全内容は、参照により本明細書に組み込まれる。

20

【0003】

本開示は、概して、エンティティ(entities)間の関係を照会(querying)又は調査及び管理するための、システム、方法、及びユーザインターフェイスを含むエンティティベースの関係の調査に関する。

【背景技術】

【0004】

情報を収集し、ソートするのに役立つ様々なツールが開発されている。しかし、エンティティベースのデータ及び関係を発見し管理するための改善されたツールに対する必要性が残っている。

30

【発明の概要】

【0005】

2つの異なるエンティティ間の接続は、2つの異なるエンティティの各々の共起する言及(co-occurring mentions)を含む1つ又は複数のドキュメントを識別するために、ドキュメントの集合(collection of documents)を処理することによって識別される。この関係は、エンティティを相互接続するグラフィカルリンクと共に、2つのエンティティの各々についてアイコン等を有するユーザインターフェイス内にグラフィカルに表示され得る。グラフィカルリンクは、2つのエンティティ間の接続を実証する(substantiate)ドキュメントの集合内の証拠へのアクセスを提供することによって、ユーザインタラクションに応答するユーザインターフェイスのアクティブな要素であることができる。一態様では、ユーザインターフェイス内の検索入力フォームフィールドを使用して、2つのエンティティ間の関係を実証するドキュメントを明示的に要求し得る。別の態様では、ユーザは、関心のあるエンティティに言及するドキュメントを明示的に選択することによって、エンティティの言及を基礎として用い(ground)得る。

40

【図面の簡単な説明】

【0006】

本明細書に記載されたデバイス、システム、及び方法の前述の及び他の目的、特徴、及び利点は、添付の図面に例示されているように、その特定の実施形態の以下の説明から明ら

50

かになる。図面は必ずしも正確な縮尺ではなく、代わりに、本明細書に記載のデバイス、システム、及び方法の原理を図示することに重点が置かれる。

【 0 0 0 7 】

【図 1】エンティティー中心の情報検索及び集約のためのネットワーク環境を示す。

【 0 0 0 8 】

【図 2】コンテンツを推薦するシステムを示す。

【 0 0 0 9 】

【図 3】コンテンツを推薦する方法のフローチャートを示す。

【 0 0 1 0 】

【図 4】コンテンツを推薦するシステムを示す。

10

【 0 0 1 1 】

【図 5】コンテンツを推薦する方法のフローチャートを示す。

【 0 0 1 2 】

【図 6】コンピュータ支援知識発見プロセスにおけるアクティビティをジャーナリング(journaling activity)する方法のフローチャートを示す。

【 0 0 1 3 】

【図 7】コンピュータ支援研究及び知識発見のためのシステムを示す。

【 0 0 1 4 】

【図 8】コンピュータシステムを示す。

【 0 0 1 5 】

20

【図 9】エンティティー関係を調査する方法のフローチャートを示す。

【 0 0 1 6 】

【図 10】エンティティー間の関係を調査する方法のフローチャートを示す。

【 0 0 1 7 】

【図 11】エンティティー間の関係を調査するためのユーザインターフェイスを示す。

【 0 0 1 8 】

【図 12】エンティティー間の関係を調査するためのユーザインターフェイスを示す。

【 0 0 1 9 】

【図 13】エンティティー間の関係を調査するためのユーザインターフェイスを示す。

【 0 0 2 0 】

30

【図 14】エンティティー間の関係を調査するためのユーザインターフェイスを示す。

【 0 0 2 1 】

【図 15】エンティティー間の関係を調査するためのユーザインターフェイスを示す。

【 0 0 2 2 】

【図 16】エンティティー間の関係を調査するためのユーザインターフェイスを示す。

【発明を実施するための形態】

【 0 0 2 3 】

以下、好適な実施形態が示されている添付の図面を参照して実施形態を説明する。しかしながら、前述のものは、多くの異なる形態で具体化することができ、本明細書に記載された例示的な実施形態に限定されるものと解釈されるべきではない。

40

【 0 0 2 4 】

本明細書で言及された全てのドキュメントは、その全体が参照により本明細書に組み込まれる。単数形のアイテムへの言及は、文脈から別段の記述がない限り又は明確でない限り、複数形のアイテムを含むものと理解されるべきであり、その逆もまた同様である。文法上の接続詞は、文脈から別段の記述がない限り又は明確でない限り、結合された句、文章、単語などのすべての分離的及び結合的な組み合わせを表現することを意図している。したがって、用語「又は」は、一般に「及び/又は」などを意味すると理解されるべきである。

【 0 0 2 5 】

本明細書中で値の範囲の記述は、本明細書中で別段の指示がない限り、限定することを意

50

図するものではなく、代わりに、その範囲内にある任意の全ての値を個別に参照する。さらに、そのような範囲内の各別個の値は、あたかもそれが本明細書中に個別に列挙されているかのように、明細書に組み込まれる。「約」、「略」等の語句は、数値を伴う場合、当業者によって意図された目的のために満足に作動することが理解されるような偏差を示すものと解釈されることになる。値及び/又は数値の範囲は、本明細書では例としてのみ提供され、記載された実施形態の範囲に対する限定を構成しない。本明細書で提供される任意の及び全ての例、又は例示的な用語(「例えば」、「など」等)の使用は、単に実施形態をより良く明らかにすることを意図したものであり、実施形態又は請求項の範囲を限定するものではない。明細書のいかなる文言も、請求項に記載されていない要素が実施形態の実施に不可欠であることを示すものと解釈してはならない。

10

【0026】

以下の説明において、「第1」、「第2」、「第3」、「上」、「下」などの用語は、便宜上の用語であり、特に断らない限り、これらの用語を限定するものと解釈されないことが理解される。

【0027】

図1に、エンティティー中心の情報検索(information retrieval)及び集約のためのネットワーク環境を示す。一般に、環境100は、通信関係で複数の参加デバイスを相互接続するデータネットワーク102を含み得る。参加デバイスは、例えば、任意の数のクライアントデバイス104、サーバ106、コンテンツソース108、及び他のリソース110を含み得る。

20

【0028】

データネットワーク102は、環境100内の参加者間でデータ及び情報を通信するのに適した任意のネットワーク(複数可)又はインターネットワーク(複数可)であり得る。これは、インターネットのような公衆ネットワーク、プライベートネットワーク、公衆交換電話ネットワークのような電気通信ネットワーク、或いは第3世代(例えば、3G又はIMT-2000)、第4世代(例えば、LTE(E-UTRA)又はWiMAX-Advanced(IEEE 802.16m))及び/又はその他の技術を使用するセルラーネットワーク、並びに環境100の参加者間でデータを運ぶために使用され得る様々な企業エリア又はローカルエリアネットワーク及びその他のスイッチ、ルータ、ハブ、ゲートウェイ等を含み得る。

【0029】

データネットワーク102の参加者の各々は、例えば、ネットワークインターフェイスカードを含む適切なネットワークインターフェイスを含み得、この用語は、本明細書では、有線及び/又は無線通信の確立及び維持に適した任意のハードウェア(ソフトウェア、ファームウェアなどと共に、それらの動作を制御するために使用される)を含む。ネットワークインターフェイスカードは、有線イーサネットネットワークインターフェイスカード(“NIC”)、無線802.11ネットワークカード、無線802.11USBデバイス、又は有線若しくは無線ローカルエリアネットワーク用の他のハードウェアを含み得るが、これらに限定されない。ネットワークインターフェイスは、また或いは代わりに、セルラーネットワークハードウェア、ワイドエリアワイヤレスネットワークハードウェア、又は集中型、アドホック、ピアツーピア、若しくはネットワークに接続してデータを搬送するために使用され得る他の無線通信用の他のハードウェアを含み得る。別の態様では、ネットワークインターフェイスは、デスクトップコンピュータなどのローカルコンピューティングデバイスに直接接続するためのシリアルポート又はUSBポートを含み得、デスクトップコンピュータは、データネットワーク102へのより一般的なネットワーク接続性を提供する。

30

40

【0030】

クライアントデバイス104は、本明細書で意図されるような、エンティティー中心情報検索及び集約技術を実施するためにユーザによって操作される環境100内の任意のデバイスを含み得る。具体的には、クライアントデバイス104は、本明細書で意図されるシステム及び方法に含まれるツール、プラットフォーム、及びデバイスを管理、監視、又は

50

その他の方法で相互作用するだけでなく、検索を開始及び実行し、情報を収集し、エンティティプロフィールを作成し、他の研究タスクを実行する等のための任意のデバイスを含み得る。例として、クライアントデバイス104は、1つ又は複数のデスクトップコンピュータ、ラップトップコンピュータ、ネットワークコンピュータ、タブレット、モバイルデバイス、ポータブルデジタルアシスタント、メッセージングデバイス、セルラー電話、スマートフォン、ポータブルメディア又はエンターテインメントデバイス、又は本明細書で意図されるような環境100に参加することができる他の任意のコンピュータデバイスを含み得る。上述のように、クライアントデバイス104は、ネットワーク環境100とインタラクションするために使用され得る、任意の無線の、バッテリーで給電されるデバイスなどの任意の形態のモバイルデバイスを含み得る。また、クライアントデバイス104のうちの1つは、サーバ106、コンテンツソース108、又は他のリソース110のうちの1つのような他のエンティティによって実行されるような、関連する機能(例えば、エンティティプロフィールの検索、記憶など)を調整し得ることも理解されよう。

10

【0031】

各クライアントデバイス104は、概して、本明細書に記載される任意のユーザインターフェイスのいずれかなどのユーザインターフェイスを提供し得る。ユーザインターフェイスは、例えば、エンティティに関するサーバ106及びコンテンツソース108からデータを受信するクライアントデバイス104のうちの1つでローカルに実行するアプリケーションによって維持され得る。他の実施形態では、サーバ106又は他のリソース110のうちの1つが、クライアントデバイス104のうちの1つで実行しているウェブブラウザ又は類似のクライアント内に表示されることができ1つ又は複数のウェブページなどを介して情報を提供するウェブサーバを含む場合のように、ユーザインターフェイスは、クライアントデバイス104のうちの1つで遠隔的にサービスされ、提示され得る。ユーザインターフェイスは、一般に、クライアントデバイス104のうちの1つのディスプレイデバイス上にユーザインタラクションのための適切な視覚的提示を生成し得、例えば、キーボード、マウス、タッチパッド、タッチスクリーン、手のジェスチャ、又は他の使用される入力デバイス(複数可)からの入力を含む、任意の適切な形態のユーザ入力を受け取るために提供され得る。

20

【0032】

サーバ106は、データ記憶デバイス、ネットワークインターフェイス、並びにプロセッサ及び/又は他の処理回路を含み得る。以下の説明では、サーバ106の機能又は構成が記載されている場合、これはサーバ106のプロセッサの対応する機能又は構成(例えば、プログラミングによる)を含むことを意図している。一般に、サーバ106(又はそのプロセッサ)は、本明細書で説明するエンティティ中心の情報検索及び集約技術に関連する様々な処理タスクを実行し得る。例えば、サーバ106は、クライアントデバイス104のうちの1つ又は複数から受信した情報を管理し、データの検索及び管理などの関連する支援機能を提供し得る。サーバ106は、また或いは代わりに、クライアントデバイス104のうちの1つ又は複数でユーザによって実行されるアクションに反応するバックエンドアルゴリズムを含み得る。バックエンドアルゴリズムはまた或いは代わりに、環境100内の他の場所に配置され得る。

30

40

【0033】

サーバ106はまた、クライアントデバイス104によるサーバ106の能力へのウェブベースのアクセスを容易にするウェブサーバ又は同様のフロントエンドを含み得る。サーバ106はまた或いは代わりに、クライアントデバイス104上のユーザインターフェイスを介してユーザに提供するための情報を得るために、コンテンツソース108及び他のリソース110と通信し得る。ユーザが、検索タイプ、言語フィルタ、関連性基準(例えば、検索結果のエンティティへの関連性を判断するための)、信頼性基準(例えば、結果が特定のエンティティに関連するかどうかを判断するための)などの検索基準を指定する、又は検索基準が、その他の方法で、例えば、クライアントデバイス104に含まれるエンティティプロフィール上で実行されるアクションを通じて、指定される場合、この情報は、

50

サーバ106(及び任意の関連アルゴリズム)によって、関連する又は新規の情報を検索し、検索結果を再ランク付けするために、コンテンツソース108又は他のリソース110などの他のリソースにアクセスするために使用され得る。追加の処理は、ユーザに対して新しい検索戦略を推薦する、又はエンティティプロフィールに追加するためにユーザに対して潜在的に新しい情報を推薦するなど、このコンテキストにおいて有用に実行され得る。

【0034】

サーバ106はまた、クライアントデバイス104のユーザが、本明細書で提供される技術のいずれかをを用いて(例えば、エンティティプロフィール上で実行されるアクションを介して自動的に)データベースコンテンツの検索及び取得を実行するためのインターフェイスと共に、コンテンツのデータベース112を維持し得る。従って、一態様では、サーバ106(又はサーバ106を含む任意のシステム)は、エンティティ情報のデータベース112を含み得、サーバ106は、データベース112内の特定の属性の場所を見つけ(locationing)、場所を見つけられたドキュメントの検索などの支援サービスを提供するための検索エンジンを提供するサーバとして機能し得る。

10

【0035】

別の態様では、サーバ106は、データネットワーク102上の遠隔位置のコンテンツを定期的に検索し、得られたコンテンツをクライアント104によるその後の検索のためにインデックス付けすることによって、検索アクティビティをサポートし得る。これは、特定のドキュメントの位置又はアドレス情報を記憶すること、ならびに語、画像、媒体、メタデータなどを識別するために任意の適切な方法でドキュメントを解析すること、ならびに類似性タイプの比較、非類似性の比較、又は他の分析を支援するために特徴ベクトル又は他の微分データを作成することを含み得る。一態様では、データベース112は、任意の所望の基準に従って手動で収集され得る。サーバ106は、クライアント104でユーザに提供され得る、本明細書に記載されているインターフェイスのいずれかのようなインターフェイスを提供し得る、又は他の方法でそれをサポートし得る。

20

【0036】

サーバ106はまたあるいはその代わりに、コンテンツのソースへのサブスクリプション(subscriptions)を管理するためのインターフェイスを提供することによってなど、シンジケートされたコンテンツ等を追跡するように構成され得る。これは、既存のサブスクリプションを検索する、新しいソースを見つける若しくは指定する、コンテンツのソースにサブスクライブするなどためのツールを含み得る。一態様では、サーバ106は、サブスクリプションを管理し、ユーザからの入力に応じて、これらのサブスクリプションからクライアントデバイス104へ新しいコンテンツを自動的に誘導し得る。従って、クライアントデバイス104は、ネットワークインターフェイスを介してコンテンツのソースに自律的にサブスクライブし、そのようなソースから直接新しいコンテンツを受信することが考えられるが、この機能はサーバ106などのリモートリソースを介して維持され得ることも考えられる。一態様では、サーバ106は、検索エンジンアルゴリズム、サーチ技術、データストア、又は本明細書に記載される他のアルゴリズム、処理などのいずれかのための検索エンジン又は他のホストを含み得、これらには、エンティティベースのサーチツール、グラウンデッド(grounded)エンティティ言及ツール、推薦(recommendation)エンジン、曖昧さ除去プロセス、機械学習又は機械分析プラットフォームなどが含まれるが、これらに限定されない。

30

40

【0037】

コンテンツソース108は、例えば、ユーザによって作成されるエンティティプロフィールを更新又は改良するために、本明細書に記載される技術によって利用可能な任意の構造化、半構造化又は非構造化フォーマットの任意のデータ又は情報のソースを含み得る。例えば、コンテンツソース108は、ウェブページ(例えば、公開又は非公開ページ)、検索エンジン又は検索サービス、各種検索サービスへのインターフェイス、リモートデータソースへのアプリケーションプログラムインターフェイス(API)、ローカル又はリモート

50

データベース(例えば、非公開データベース、企業データベース、政府データベース、機関データベース、教育データベースなど)、ライブラリ、その他のオンラインリソース、ソーシャルネットワーク、コンピュータプログラム及びアプリケーション、その他のエンティティプロフィールなどを含むが、これらに限定されない。コンテンツソース108は、テキスト情報(例えば、書籍、雑誌、定期刊行物、雑誌、新聞、論文、報告書、法的ドキュメント、記者、辞書、百科事典、ブログ、ウィキなど)、グラフィックス情報(例えば、図表、グラフ、表など)、画像又は他の視覚データ(例えば、写真、図面、絵画、図解、レンファリング、モデル、スケッチ、図表、コンピュータ支援設計など)、音声データ、数値データ、地理データ、科学データ(例えば、化学組成、科学的な公式など)、数学データなどを含むが、これらに限定されない様々なタイプの情報及びデータを含み得る。

10

【0038】

他のリソース110は、本明細書に記載されるデバイス、システム、及び方法において有用に使用され得る任意のリソースを含み得る。例えば、他のリソース110は、他のデータネットワーク、人間の関係者(例えば、プログラマ、研究者、アノテータ(annotator)、エディタ等)、センサ(例えば、オーディオ又はビジュアルセンサ)、テキストマイニングツール、ウェブクローラ、知識ベース加速(KBA)ツール又は他のコンテンツモニタリングツール等を含み得るが、これらに限定されない。他のリソース110はまたあるいはその代わりに、本明細書に意図されているように、ネットワーク化されたアプリケーションにおいて有用に使用され得る任意の他のソフトウェア又はハードウェアリソースを含み得る。例えば、他のリソース110は、コンテンツサブスクリプション、コンテンツ購入、又はその他のための支払いを認可するために使用される支払い処理サーバ又はプラットフォームを含み得る。別の例として、他のリソース110は、例えば、ユーザによって実行されるエンティティプロフィール又は他の研究を共有するために、又はエンティティ情報の追加ソースとして使用され得るソーシャルネットワークキングプラットフォームを含み得る。別の態様では、他のリソース110は、身元の第三者検証、コンテンツの暗号化又は復号化などのための証明書サーバ又は他のセキュリティリソースを含み得る。別の態様では、他のリソース110は、クライアントデバイス104のうちの1つと共存する(例えば、シリアル又はUSBケーブルと共に、又はそれを介して直接結合された同一のローカルエリアネットワーク上の)デスクトップコンピュータなどを含み得る。この場合、他のリソース110は、クライアントデバイス104のための補足機能を提供し得る。他のリソース110は、スキャナ、カメラ、プリンタ等の補足的なリソースも含む。

20

30

【0039】

環境100は、環境100内の他の参加者のいずれかへの又はそれらからのウェブベースのアクセスを提供する1つ又は複数のウェブサーバ114を含み得る。別個のネットワークエンティティとして示されているが、ウェブサーバ114は、本明細書に記載されている他のデバイスのうちの1つと論理的又は物理的に関連付けられ得、例えば、データネットワーク102を介して、例えばクライアントデバイス104からのユーザインタラクションを可能にする方法で、サーバ106(又はそれらに結合されているデータベース112)のうちの1つ、コンテンツソース108のうちの1つ、又は他のリソース110のいずれかへのウェブアクセスのためのユーザインターフェイスを含み得る又は提供し得ることは容易に理解されよう。

40

【0040】

環境100の参加者は、本明細書に記載されている種々の機能を実行するための任意のハードウェア又はソフトウェアを含み得ることが理解されよう。例えば、クライアントデバイス104及びサーバ106のうちの1つ又は複数は、メモリ及びプロセッサを含み得る。

【0041】

上述のネットワーク環境100の種々の構成要素は、種々の方法で本明細書に記載の技術をサポートするように配置され、構成され得る。例えば、一態様では、クライアントデバイス104は、データネットワーク102を介して、エンティティ中心の情報検索及び集約に関連する様々な処理タスクを実行するサーバ106に接続する。例えば、サーバ106

50

は、エンティティに関連する情報を検索する、取得する(retrieving)、及びランキングするためのクエリとして使用されるエンティティプロフィールをユーザが構築するエンティティ中心情報検索及び集約プログラムを実行するウェブサイトホストし得る。このようにして、ユーザがクライアントデバイス104上に表示されるインターフェイス上にエンティティプロフィールを構築すると、サーバ106は、コンテンツソース108、他のリソース110、又はデータベース112を使用して、エンティティプロフィールに関連する新しい関連情報の検索を更新し得る。以下にさらに詳細に説明するように、サーバ106(又は環境100の別の参加者)は、検索を定義し、サーバ106がエンティティ中心情報検索及び集約プログラムで行われるアクション、例えば、エンティティプロフィールに対して行われる修正又は情報の選択に反応することを可能にする1つ又は複数のアルゴリズムを含み得る。より一般的には、本明細書に記載される情報の検索、処理、及び提示の態様は、任意の適切な方法で分散され得る。例えば、検索エンジンの検索機能は、ドキュメントを見つけて処理する1つ又は複数のサーバ間、及びクライアントデバイス間に分散され得、クライアントデバイスは、例えば、ユーザインターフェイス内でローカルにクエリを解析し、リモート検索エンジンへ情報要求を送信し、検索エンジンからの結果を提示し、又は本明細書に記載される情報及び他の機能の種々の検索、処理、及び表示に参加し得る。同様に、ローカルデバイス又はクラウドストレージ設備上のドキュメントが検索内に含まれる場合、検索エンジンは、分散配置されるか、又は関心のあるドキュメントリポジトリへのアクセスを提供され得る。

10

20

【0042】

図2はコンテンツを推薦するシステムを示す。特に、推薦エンジン202は、ユーザインタラクションに基づいて、ユーザインターフェイス206内の人間ユーザに対して、例えば、ドキュメントのコーパス(corpus)204からの関連ドキュメント及び概念に関する推薦を提供し得る。

【0043】

ユーザインターフェイス206は、例えば、知識グラフ(knowledge graph)208などのコンテンツを人間のユーザに表示し、グラフィカルインターフェイス入力、テキスト入力などのような入力を人間のユーザから受け取るのに適した任意のディスプレイ、インターフェイスデバイス、インターフェイス構成要素等を含み得る。これは、例えば、上述のクライアントデバイスのいずれかのためのディスプレイを含み得る。一般的に、知識グラフ208は、概念アイコンによってユーザインターフェイス206内で表現される1つ又は複数の概念とともに、ドキュメントアイコンによってユーザインターフェイス206内で表現される1つ又は複数のドキュメントを含み得る。知識グラフ208を形成するために、これらのドキュメント及び概念は、ユーザインターフェイス206内の記述の視覚インジケータとして表されるグラフエッジ(graph edges)などの関係を介して関連付けられ得る。図示のように、ディスプレイで描画されるユーザインターフェイス206は、ドキュメントアイコン210、第1の概念アイコン212、及び第1の視覚インジケータ216を含み得る。ドキュメントアイコン210は、例えば、単一のデータベース若しくはデータストアに記憶され得る、又はデータネットワーク又は他の分散環境にわたって分散され、検索エンジンなどによる識別のために適切にインデックス付けされ得るドキュメントのコーパス204内の第1のドキュメント218と関連付けられ得る。ユーザインターフェイス206とは別に描かれているが、第1のドキュメント218及びドキュメントのコーパス204内の1つ又は複数の他のドキュメントは、ユーザインターフェイス206をホストして制御するデバイス上に配置され得るが理解されよう。ドキュメントは、またあるいはその代わりに、デバイスから離れていてもよく、又はこれらの組み合わせであってもよい。

30

40

【0044】

知識グラフ208は、ここでは、視覚的に情報を提示する適切な方法として参照されるが、対応するデータの任意の他の視覚的表現を使用することもできるし、その代わりに使用することもできることを理解されたい。例えば、本明細書に記載されているように、知識グラフ208のような視覚的表現とのユーザ及びマシンのインタラクションは、関連する

50

アクティビティのためのデータリポジトリとして役立つ操作のジャーナル(journal of operations)に記録される。操作のジャーナル内の記録は、ユーザへのプレゼンテーションのための知識グラフ208を作成するために集約され得るが、操作のジャーナルによって表されるプロジェクトの任意の他の適切な視覚的表現を作成するために、記録はまた或いは代わりに集約され得る。従って、例えば、テキストが、便宜上、知識グラフ208を参照している場合、テキストは、以下に説明する要約カード及び言及ハイライトのような他の視覚表現、並びに、そのような情報に関するユーザへの情報の構成を提示する及び/又はユーザの入力を受け取るのに適した任意の他の視覚表現も参照することが意図されていることが理解されるべきである。

【0045】

一般に、第1の概念は、第1の概念アイコン212をユーザインターフェイス206内のドキュメントアイコン210に関連付ける第1の視覚インジケータ216によって表されるように、第1のドキュメント218において言及され得る。ユーザインターフェイス206は、図2において概して(a)として示されている第1の概念も言及するドキュメントのコーパス204内の他のドキュメントを、自動的に、又はユーザの要求に応答して識別するように構成され得る。推薦エンジン202は、自動的に、又は明示的なユーザの要求に応答して、第1の概念、(a)及び第2の概念、(b)を識別するドキュメントの集合内のドキュメント213を識別し、次いで第2の概念、(b)知識グラフ208に含めるために推薦し得る。一般に、推薦エンジン202は、ユーザインターフェイス206内の第1の概念アイコン212を第2の概念アイコン214に視覚的に関連付けるための第2の視覚インジケータ222と共に、第2の概念を識別する第2の概念アイコン214を自動的に作成し得る。推薦エンジン202はまた或いは代わりに、候補関係を識別し、ユーザインターフェイス206をホストするデバイス上でローカルに実行するエージェントにその関係を通信し得、エージェントは、視覚インジケータ及び/又はアイコンを構成及び表示する最良の方法を決定することができる。また、推薦エンジン202は、ユーザインターフェイス206内に知識グラフ208を自動的に投入し得る、又は、ユーザが提案された追加の説明(clarification)を受け入れる、拒否する、修正する、又は要求することができるよう、ユーザインターフェイス206を通してユーザに候補ドキュメント、概念、又は関係が提示され得ることも理解されよう。

【0046】

推薦エンジン202はまた、提案された追加のための様々な支援情報を提供し得る。例えば、推薦エンジン202は、複数の他のドキュメント213((a)及び(b)の記述を含む)のランク付けされたリスト224を作成し、ランク付けされたリスト224をディスプレイ、例えば、ユーザインターフェイス206における提示のためにデバイスに送信し得る。これには、例えば、(a)と(b)の間の関係を実証するスニペット(snippets)、コンテンツの抜粋等、又はドキュメント(複数可)内の対応する概念の両方のいずれかの言及を含み得る。例えば、後述するように、ランク付けされたリスト224内のドキュメントを選択し、順位付けするために、種々のランク付け技術が使用され得る。

【0047】

図3はコンテンツを推薦する方法のフローチャートを示す。一般に、ユーザは、ユーザインターフェイス内に表示される知識グラフとインタラクトし得、推薦エンジンは、特定のユーザ要求及び他のコンテキスト情報に基づいて、知識グラフへの追加のための推薦を応答的に生成し得る。このプロセスは、例えば、以下に記載されるように、ユーザが開始したコンピュータ支援による推薦及び選択のシーケンスを介して知識グラフの作成を支援するために、反復的に継続し得る。

【0048】

ステップ302に示すように、方法300は、ドキュメントアイコン、第1の概念アイコン、及びディスプレイにおける提示のための第1の視覚インジケータを提供することから開始し得る。ドキュメントアイコンは、例えば、ローカルドキュメントリポジトリ、リモートドキュメントリポジトリ、検索エンジンによってインデックス付けされたドキュメントの集合、又はこれらの何らかの組み合わせなどのドキュメントのコーパス内の第1のドキュ

10

20

30

40

50

メントと関連付けられ得る。第1の概念のアイコンは、第1のドキュメントで言及され、また、ドキュメントのコーパス内の複数の他のドキュメントで言及された第1の概念と関連付けられ得る。本明細書にさらに記載されるように、視覚インジケータは、ドキュメントアイコンをディスプレイ内の第1の概念アイコンと視覚的に関連付けられ得る。これらの表示要素は、リモートソースによって提供され得る、知識グラフのために遠隔的に記憶されたモデルに応答してローカルに生成され得る、(本明細書に記載されるように)操作のジャーナルの態様を表示するためにローカルに生成され得る、又は、ディスプレイ内のユーザインターフェイス内の知識グラフを視覚的に提示するために別の方法で提供され若しくは生成され得る。上述したように、知識グラフは、操作のジャーナルからの情報を提示するための1つの便利な手法であるが、情報の構成を伝達すること及び/又は操作のジャーナルに関連するユーザの入力を受け取ることに適した操作のジャーナルの任意の他の視覚的表現もまた或いは代わりに使用され得る。

10

【0049】

ステップ304に示されるように、方法300は、全てが本明細書で概して意図されるように、例えば、ドキュメント、概念、及びそれらの間の言及を視覚的に示し、関連付ける知識グラフを表示することを含み得る。一般に、知識グラフは、ユーザインターフェイス内のインタラクティブなオブジェクトであり得、例えば、追加の支援情報の検索(retrieval)及び表示、又は追加の推薦の要求等を容易にすることによって、個々の頂点(vertices)(例えば、ドキュメント又は概念)及びエッジ(例えば、ドキュメント/概念間の言及)とのユーザインタラクションをサポートし得る。知識グラフは、他の有用なコンテキスト情報と共に表示され得る。例えば、方法300は、知識グラフの第1の概念と知識グラフの第2の概念との関係を実証するディスプレイ内での提示のために、複数の他のドキュメントからの抜粋を提示することを含み得る。

20

【0050】

ステップ306に示されるように、方法300は、第1の概念に言及するドキュメントのコーパスの複数の他のドキュメントで言及されている第1の概念とは異なる第2の概念など、第2の概念のための推薦を作成することを含み得る。これは、推薦エンジンなどを用いて推薦を作成することを含み得る。一態様では、推薦は、ユーザインターフェイス内の知識グラフとのインタラクションなど、推薦についての明示的なユーザ要求に応答して作成され得る。別の態様では、推薦は、例えば、知識グラフ若しくは他のユーザコンテキスト、ユーザアクティビティ等を通じたナビゲーションに応答して自動的に生成され得る。

30

【0051】

ステップ308に示すように、方法300は、推薦のための視覚要素を提供することを含み得る。例えば、これは、第2の概念を識別する第2の概念アイコンと、第1の概念アイコンをディスプレイ内の第2の概念アイコンに視覚的に関連付けるための第2の視覚インジケータとを、ディスプレイ内の提示のための第2の概念アイコンと第2の視覚インジケータを提供することと共に、生成することを含み得る。上述のように、これは、ユーザインターフェイスをホストするクライアントにおいてローカルに、又は遠隔的に、例えば知識グラフの推薦を提供する推薦エンジンによって、又は知識グラフ提示プラットフォームをホストするサーバもしくは他のクラウドサービスなどによって実行され得る。一態様では、第1の視覚インジケータは、第2の視覚インジケータから視覚的に区別可能な外観を有し得、これにより、例えば、ユーザは、ユーザが関係の作成順序を区別することができるように、又は、知識グラフ内に含まれる関係の異なるソース及びタイプを区別するのに有用な他の情報を提供するために、ユーザ選択関係と機械選択関係とを区別することができる。

40

【0052】

ステップ310に示されるように、方法300は、ディスプレイにおける提示のための第1の概念に言及する複数の他のドキュメントのランク付けされたリストを作成することを含み得る。ドキュメントは、例えば、現在のグラフにつながるユーザ選択の履歴、機械生成の推薦などの知識グラフのコンテキストに基づいて、有利にランク付けされ得る。リストは

50

また或いは代わりに、ドキュメントが2つの概念アイコン間の関係をどの程度うまくサポートしているか、あるいは、各ドキュメントが知識グラフへの追加に役立つかもしれない新しい情報を含んでいるかどうかに基づいてなど、様々な潜在的な使用のためにランク付けされ得る。非限定的な例として、ランク付けされたリストは、第1の概念に対する類似性、第1のドキュメントとの相違、又は第2の概念の第1の概念に対する実証の強さのうちの少なくとも1つにしたがってランク付けすることを含み得る。ランク付けされたリストを作成することは、例えば、ローカルプログラム/リソースを用いて又はリモートリソースから、第2の視覚インジケータのユーザ選択にตอบสนองして、ディスプレイ内で提示するためのランク付けされたリストを提示することを含み得る。

【0053】

10

ステップ312に示されるように、方法300は、ユーザ入力を受信することを含み得る。これは、キーボード入力、マウス操作、又は知識グラフ及び関連情報のコンテキスト内でユーザが望むアクションを示す他のユーザ入力などを含み得る。

【0054】

ステップ314に示されるように、方法300は、追加の推薦を生成することを含み得る。例えば、ユーザ入力第2のドキュメントのユーザ選択を含む場合、ユーザ選択にตอบสนองして追加の推薦を生成することは、第2のドキュメントアイコン、第3の概念アイコン、及びディスプレイにおける提示のための第3の視覚インジケータを提供することを含み得、第2のドキュメントアイコンは第2のドキュメントに関連付けられ、第3の概念アイコンは第2のドキュメントにおける概念に関連付けられ、第3の視覚インジケータは第2のドキュメントアイコンをディスプレイにおける第3の概念アイコンに視覚的に関連付けられる。これは、ディスプレイにおける追加の推薦の提示を容易にし得る。追加の推薦は、明示的なユーザの入力にตอบสนองして生成され得るが、追加の推薦はまた或いは代わりに、例えば、知識グラフの状態に基づくバックグラウンドタスクとして、又はユーザが追加の推薦を要求することを示唆するユーザのアクションにตอบสนองして、ユーザの入力がない場合にも生成され得るか、又は、ユーザが既に利用可能な追加の推薦に関心を持つかもしれないことに留意されたい。

20

【0055】

一態様では、追加の推薦は、例えば、第1の概念のアイコン及び/又は第2の概念のアイコンのユーザインターフェイスにおけるユーザ選択にตอบสนองして、第3の概念の推薦を含み得る。さらに、プロセスは反復的であってもよく、潜在的な関心のある新しいドキュメント又は概念に加えて、知識グラフの要素間の関係を識別することを含んでもよい。したがって、例えば、追加の推薦は、第2のドキュメントのユーザ選択を受信することに対応して、第1の概念のアイコン及び第2の概念のアイコンのうちの少なくとも1つを、ディスプレイ内の第3の概念のアイコンに視覚的に関連付けるための第4の視覚インジケータを作成することを含み得る。

30

【0056】

このような推薦を生成する推薦エンジンは、ディスプレイをレンダリングするクライアント上でローカルに実行し得る。推薦エンジンはまた或いは代わりに、ディスプレイをレンダリングするクライアントから遠隔で実行し得る。一般に、推薦エンジンは、様々なデータベース及び他のデータソース、コンテンツのインデックス、処理リソース、サービスなどのいずれかによってサポートされ得る。従って、図3には示されていないが、本方法300は、推薦エンジンにアクセス可能な1つ又は複数の場所にドキュメントのコーパスを記憶することを含み得ることが理解されるであろう。別の態様では、方法300は、ドキュメントのコーパスのためのインデックスを記憶することを含み得、インデックスは、ドキュメントのコーパス内の各ドキュメントの少なくとも1つの概念を識別する。

40

【0057】

ステップ316に示されるように、方法300は、操作のジャーナルの知識グラフ又は他の視覚表現を更新することを含み得る。これは、例えば、知識グラフに関する情報(例えば、ドキュメント及び概念間の関係、ならびに、そのような関係に関する情報の実証)を含

50

むデータ構造を更新することを含み得る。これはまた或いは代わりに、ユーザインタラクション並びに知識グラフ及び/又は知識グラフの要素とのマシンインタラクションの履歴記録を含む、操作のジャーナルなどのデータ構造を更新することを含み得る。例えば、方法300は、ドキュメントアイコン、第1の概念アイコン、並びに第2の概念アイコン及び推薦エンジンからの少なくとも1つの推薦のうちの少なくとも1つ又は複数のユーザインタラクションを含む操作のジャーナルを記憶することを含み得る。操作のジャーナルを記憶することは、操作のジャーナルを、ディスプレイをホストするデバイスとは独立して、永続メモリに記憶することを含み得る。操作のジャーナルを保存することはまた或いは代わりに、推薦エンジンをホストするデバイスとは無関係に、操作のジャーナルを永続メモリに保存することを含み得る。

10

【0058】

知識グラフを更新することはまた或いは代わりに、デバイス上のディスプレイをローカルに更新すること、又はリモートリソースからディスプレイを遠隔的に更新することを含み得る。知識グラフのディスプレイを更新することを含み得る。一態様では、ディスプレイは、サーバ又は他のリモートリソースによって制御されるユーザインターフェイスを有するリモートクライアントと関連付けられ得、ドキュメントアイコン、第1の概念アイコン、及び第1の視覚インジケータを提供することは、ドキュメントアイコン、第1の概念アイコン、及び第1の視覚インジケータを、リモートクライアントのユーザインターフェイスにおける提示のために、リモートクライアントに送信することを含み得る。

20

【0059】

方法300のステップの一部又は全てが反復的に繰り返して、知識グラフを発展させ得ることが理解されよう。例えば、追加の入力はステップ312のように受信され得、追加の推薦は知識グラフをさらに更新するために生成され使用され得る。

【0060】

以上のことから、本明細書には、コンテンツを推薦するためのシステムも開示されている。システムは、ディスプレイと、デバイスのプロセッサと、デバイスと通信する関係で結合された推薦エンジンとを含むデバイスを含み得る。プロセッサは、ドキュメントアイコンと、第1の概念アイコンと、ディスプレイにおける提示のための第1の視覚インジケータとを提供するように構成され得、ドキュメントアイコンは、第1のドキュメントに関連付けられ、第1の概念アイコンは、第1のドキュメントにおいて言及されかつ複数の他のドキュメントにも言及される第1の概念に関連付けられ、第1の視覚インジケータは、ディスプレイにおける第1の概念アイコンにドキュメントアイコンを視覚的に関連付ける。デバイスからリモート又はローカル、あるいはこれらの組み合わせであり得る推薦エンジンは、第1の概念に関連する概念についてのリクエストをデバイスから受信し、複数の他のドキュメントにおいて言及された第2の概念についての推薦を作成するように構成され得、第2の概念は、第1の概念とは異なり且つ第1の概念に関連し、推薦エンジンはさらに、推薦をデバイスに通信するように構成され、第1の概念と第2の概念の間の関係を実証する複数の他の文献からの抜粋を含む。

30

【0061】

図4はコンテンツを推薦するシステムを示す。一般に、システム400は、本明細書に記載されているような自動化されたコンテンツ推薦の人間が読むことができる実証を提示するように構成される。

40

【0062】

システム400は、ファイルリスト、アイコンのグループ、又はドキュメントのコーパス408内のファイルの他の任意の適切な表現など、いくつかのドキュメント406の表現を表示するユーザインターフェイス404を有するディスプレイ402を含み得る。ユーザインターフェイス404内のドキュメント403の選択に応答して、システム400は、本明細書に概して記載されるような推薦エンジン412を用いて、ドキュメントのコーパス408からの1つ又は複数の他のドキュメント410の推薦による検索を開始し得る。特に、推薦エンジン412は、選択されたドキュメント403において第1の概念(a)を言及し、第1の概

50

念(a)に潜在的に関連する第2の概念(b)を言及するドキュメントを検索し得る。本明細書にさらに記載されるように、推薦エンジン412は、ドキュメント及び概念のグラフ表示を、それらの間の関係とともに、返し得る。推薦エンジン412はまた、ランク付けされたリストのようなリスト414などに提示され得る選択されたドキュメント403との関係を実証するドキュメント410からのスニペット又は抜粋などの情報を返し得る。

【0063】

リスト414は、さらに、サポートドキュメント410へのリンク、知識グラフ416内の概念の位置へのリンク等などのインタラクティブコンテンツを含み得る。

【0064】

推薦エンジン412、又はシステム400内の他のいくつかの適切なサービス若しくはエンティティはまた、本明細書に記載されるように、操作のジャーナル418を維持し得、これは、概して、共有され、編集され、さらなる推薦のためのコンテキストとして使用され得る、知識グラフ416との人間ユーザインタラクション及び機械ユーザインタラクションの記録を記憶する。また、システム400は、知識グラフ416の反復的生成をサポートするように構成され得ることが理解されよう。例えば、リスト414は、各新しい推薦のリストが作成されると、これは、追加の概念及び推薦を検索するためにユーザによって使用され得るように、ドキュメント406の表現として機能し得る。

【0065】

従って、知識グラフ416は、概して拡張可能であり得、さらに、他のユーザと、例えば、共同知識発見のためのプラットフォームを提供するための操作のジャーナル418として、共有され得る。

【0066】

図5はコンテンツを推薦する方法のフローチャートを示す。

【0067】

ステップ502に示されるように、方法500は、例えば、上述のように、デバイスのユーザインターフェイス内のドキュメントのコーパスからの第1の複数のドキュメントの表現を表示することによってなど、ドキュメントを表示することから始まる。これは、ファイルリスト、ドキュメントのアイコンを有するウィンドウ、又はユーザインターフェイス内の表示及び操作のためのドキュメントの他の任意の適切な表現を含み得る。

【0068】

ステップ504に示されるように、方法500は、第1の複数のドキュメントから第1のドキュメントの第1の概念のユーザインターフェイス内のユーザ選択を受信することによってなど、ユーザ選択を受信することを含み得る。これは、キーボード操作、マウスクリック、又は任意の他の適切なユーザインターフェイスインタラクションを含み得る。

【0069】

ステップ506に示されるように、方法500は、知識グラフのための視覚要素を作成するために、いくつかのステップを自動的に実行することを含み得る。例えば、これは、第1のドキュメントで言及された第1の概念に関連する第1の概念のアイコンを選択することと、ユーザインターフェイスで表示するための第1の概念のアイコンを提示することとを含み得る。これはまた、第1のドキュメントのドキュメントアイコンを第1の概念アイコンに視覚的に接続する第1の視覚インジケータをユーザインターフェイス内に提示することを含み得る。これはまた、(例えば、ユーザの選択に自動的に応答して)デバイスから離れた推薦エンジンによる第1の概念とは異なる第2の概念の推薦を作成することを含み得る。第2の概念は、例えば、第1の概念(ユーザによって選択された)と第2の概念の両方への言及を全て含む、ドキュメントのコーパスからの第2の複数のドキュメントにおいて言及された概念であり得る。これはまた、(この場合も同様に、オプションで、ユーザの選択に自動的に応答して)ユーザインターフェイス内の表示のための第2の概念に関連する第2の概念アイコンを提示することと、ユーザインターフェイス内の第2の視覚インジケータを提示することとを含み得、第2の視覚インジケータは、第1の概念アイコンを第2の概念アイコンに視覚的に接続する。

10

20

30

40

50

【 0 0 7 0 】

ステップ508に示されるように、方法500は、第1の概念及び第2の概念の言及を含む、第2の複数のドキュメントからの1つ又は複数のコンテンツ選択のランク付けされたリストを作成することを含み得る。リストは、例えば、ドキュメントのタイトル、ファイル名、作成日等、及び人間のレビューを容易にするための概念に関連する情報のスニペットを含み得る。ランク付けされたリストは、知識グラフを編集、更新、検証、又はその他の方法でのレビューもしくは修正するのに有用な任意の適切な方法でランク付けされ得る。例えば、ランク付けされたリストは、2つの関連する概念の間のグラフ距離に従ってランク付けされ得る。ランク付けされたリストは、第1の概念と第2の概念との間の関係の実証の推定に従ってランク付けされ得、これは、例えば、機械学習、人工知能、セマンティック処理、又はドキュメントコンテンツの自動評価及び比較のための任意の他のツールを用いて評価され得る。ランク付けされたリストはまた或いは代わりに、第1の概念の言及と第2の概念との間の第2の複数のドキュメントの各々内の距離に従ってランク付けされ得、これは、ドキュメント内の2つの概念の関係の代理(proxy)として機能し得る。ランク付けされたリストは、また或いは代わりに、第1の概念及び第2の概念のいくつかの言及、又は、概念間の関係又は知識グラフとの関連性を評価又は推定するための、任意の他の適切なメトリック(metric)若しくはメトリックの組み合わせに従って、ランク付けされ得る。

10

【 0 0 7 1 】

ステップ510に示されるように、方法500は、ユーザインターフェイス内に表示するために、ランク付けされたリストの少なくとも一部を提示することを含み得る。このようにして、二次概念を介して互いに関連付けられるドキュメントのグループ及びユーザによって選択された第1の概念が、さらなるユーザインタラクションのためにユーザインターフェイス内のリストとして自動的に生成され、提示され得る。1つ又は複数のコンテンツ選択のランク付けされたリストは、例えば、第1の概念のアイコンと第2の概念のアイコンとの間の関係をサポートするコンテンツを含み得る。

20

【 0 0 7 2 】

ステップ512に示されるように、方法500は、ユーザのための推薦を作成することを含み得る。例えば、ランク付けされたリストが提示された後、追加の関連概念推薦のためのユーザ要求が受信され得る。このようなユーザの要求に応答して、方法500は、ドキュメントのコーパスからユーザインターフェイスへのドキュメント内で言及された概念に関連する1つ又は複数の追加の概念アイコンを追加すること、及び各追加の概念アイコンに対して、ユーザインターフェイス内に表示された概念アイコンのうちの1つ又は複数の他のアイコンに追加の概念アイコンを視覚的に関連付ける少なくとも1つの視覚インジケータを追加することを含み得る。

30

【 0 0 7 3 】

ステップ514に示されるように、方法500は、知識グラフを反復的に探索、拡張、及び洗練するために、任意の数の追加ユーザ要求を受信及び処理することを含み得る。例えば、方法500は、2つの概念アイコンの間の関係を実証するコンテンツに対する第2のユーザ要求を受信することを含み得る。この第2の要求に応答して、方法は、概念アイコンの2つの間の接続を記述するドキュメントのコーパスからのドキュメント内の一つ以上のコンテンツ項目を提示することを含み得る。

40

【 0 0 7 4 】

これはまた或いは代わりに、ドキュメントの提示、知識グラフ、ドキュメントのランク付けされたリスト、及びユーザインターフェイスの任意の他の部分を更新することによってなど、ユーザインターフェイス内のディスプレイを更新することを含み得る。これはまた或いは代わりに、知識グラフとの人及びマシンインタラクションを記憶する操作のジャーナルを更新することを含み得る。

【 0 0 7 5 】

図6は、コンピュータ支援知識ディスカバリープロセスにおけるアクティビティをジャー

50

ナルする(journaling activity)方法を示す。

【0076】

ステップ602に示されるように、方法600は、情報を表示することを含み得る。一般に、これは、知識、情報、データソース、関係などの任意の構造化された表現を含み得る。これは、例えば、本明細書に記載されるような知識グラフ、又はデバイスのユーザインターフェイスにおける提示に適した任意の他の知識表現などを含み得る。従って、一態様では、これは、第1の表面上の視覚表示要素内にグラフを提示することを含み得る。これは、例えば、コンピュータ、タブレット、スマートフォン等のディスプレイ、並びに、このような物理的なディスプレイ媒体内のアプリケーション、プロセス等のウィンドウ又は他のアクティブなグラフィカル部分を含み得る。知識グラフ等の場合、グラフは、グラフの頂点として提示される1つ又は複数のドキュメント及び概念、並びにグラフのエッジとして提示される1つ又は複数のドキュメント及び概念の間の1つ又は複数の関係を含む複数のグラフ要素を含み得る。

10

【0077】

ステップ604に示されるように、この方法は、ユーザクションを受信することを含み得る。これは、例えば、視覚表示要素内の複数のグラフ要素の1つを用いて人間のユーザからのユーザ入力を受信すること、又は、知識グラフ若しくはその他の知識表現をレンダリングするユーザインターフェイス又はデバイスを用いて人間のユーザからの任意の他の入力を受信することを含み得る。一態様では、これは、推薦の要求、知識グラフにおける関係の実証のための要求、知識グラフにおけるドキュメントアイコンによって識別されるドキュメントのための要求、又は知識グラフのコンテンツに関連する若しくはそれに基づく任意の他の情報を含み得る。別の態様では、ユーザアクションは、知識グラフの操作であり得る。例えば、第1のアクションは、グラフに頂点として第1のドキュメントを追加すること、グラフから頂点を削除すること、グラフからエッジを削除すること、又はグラフにエッジを追加することを含み得る。

20

【0078】

一態様では、ユーザインターフェイスが、例えば、概念アイコンとして又はドキュメント内のテキストとして、概念の言及を表示する場合、ユーザによる第1のアクションは、第1の言及の選択、関連ドキュメントの要求、関連概念の要求等のような、第1の言及とのユーザインタラクションを含み得る。

30

【0079】

ステップ606に示されるように、方法600は、第1のアクションの記録(record)を記憶することを含み得る。記録は、知識グラフの状態、ユーザインタラクションの性質、又はアクション若しくはアクションが実行されたコンテキストを識別若しくは解釈するのに有用な他の情報に関する任意の情報を有用に含み得る。一態様では、記録は、1つ又は複数の他のドキュメント、例えば、ディスプレイ内の他のドキュメント又はユーザに提供される検索若しくは推薦内の他のドキュメントとの第1のドキュメントの関係を含み得る。別の態様では、記録は、1又は複数の他のドキュメントにおける1又は複数の他の言及との第1の言及の関係を含み得る。より一般的には、記録は、知識グラフ又は他の知識表現、ユーザがユーザインターフェイスとインタラクトする方法、ユーザによって提供される任意の要求又は命令の特定の又は一般的な性質などに関する任意の情報を含み得る。一態様では、記録は、第1のアクションに関連する第1のドキュメントの識別情報と、第1のアクションに関連する第1のドキュメント内の第1の概念の第1の言及とを含み得る。

40

【0080】

記録を保存することはまた或いは代わりに、操作のジャーナルに第1のアクションの記録を保存することを含み得る。本明細書に記載されるように、操作のジャーナルは、第1の表面によってアクセス可能であり、及び、推薦などを生成するための機械ベースのアルゴリズムなどの第1の表面をホストするデバイスから独立して実行する計算プラットフォーム上で動作する機械ベースのアルゴリズムによってもアクセス可能なデータストア(data store)に永続的に記憶され得る。

50

【0081】

ステップ608に示されるように、方法600は、ユーザからの第1のアクションに応答する機械ベースのアルゴリズムからの第2のアクションなどのコンピュータアクションを受信することを含み得る。第2のアクションは、ユーザアクションに対する任意の適切な応答であり得る。例えば、第2のアクションは、第2のドキュメントにおいてユーザによって選択される概念の第2の言及を識別し得る。別の態様では、これは、機械学習アルゴリズム、ニューラルネットワーク、又は他のパターンマッチングアルゴリズムなどによって識別される、関連する又は潜在的に関連するドキュメント又は概念を含む、本明細書で意図される推薦エンジンなどによって提供され得る任意の推薦を含み得る。別の態様では、これは、データ処理又は操作を含み得る。例えば、コンピュータアクションは、知識グラフの中で、知識グラフに追加するための可能な推薦のリスト内で、又はこれらの何らかの組み合わせで、第1の概念の第1の言及と第2の概念の第2の言及との間の関係を実証する、第2のドキュメントからの1つ又は複数の抜粋の自動化された選択を含み得る。これはまた或いは代わりに、頂点、エッジ等の追加など、知識グラフへの推薦される変更を含み得る。例えば、機械ベースのアルゴリズムは、既に知識グラフ内にある2つの概念間の関係、又は知識グラフに関連する新しい概念を識別し得、第2のアクションは、その関係を示すグラフにエッジを追加することを含み得る。

10

【0082】

ステップ610に示されるように、方法600は、コンピュータアクションの記録を記憶することを含み得る。これは、例えば、第2のアクション(又は第2のアクションの説明)を、操作のジャーナル又は他の適切なリポジトリに記憶することを含み得る。これはまた、記録のその後の使用に適した任意のコンテキスト情報等を記録することを含み得る。

20

【0083】

ステップ612に示されるように、方法600は、例えば、ユーザインターフェイスにおける知識グラフ若しくは他の知識表現の表示(display)、又はドキュメントのランク付けされたリスト、推薦などの関連項目を更新することを含み得る。一態様では、これは、第1の表面上の視覚表示要素においてステップ608で推薦される第2のドキュメントを提示することを含み得る。

【0084】

ステップ614に示されるように、方法600は、操作のジャーナルにおける蓄積された記録に基づいて、又は知識グラフ等における人的及び機械的操作の記録の他の同様の蓄積に基づいて、実行されることが出来る他のアクティビティを含み得る。上述したように、操作のジャーナルは、例えば、複数のユーザ間での共有及びコラボレーションを可能にする、ネットワークアクセス可能なロケーションに記憶され得る。知識グラフの最新バージョンは、操作のジャーナルから得られ、任意の数のデバイス上に所望通りに表示され得る。より一般的には、操作のジャーナルは、操作のジャーナルの知識グラフ又は他の視覚的表現を表示、共有、編集、又は他の方法で操作するために、任意の適切な方法で適用され得る。

30

【0085】

例えば、これは、第2の表面上にグラフを表示するために、操作のジャーナルを適用することを含み得る。一態様では、これは、ユーザが知識グラフ内に含まれる特定の研究プロジェクトを、ユーザが使用する別のデバイスなどの第2の表面に移植する(port)ことを可能にする。別の態様では、これは、別のユーザのために第2の表面上にグラフを表示するために操作のジャーナルを適用することによって、ユーザが他のユーザと知識グラフを共有することを可能にする。

40

【0086】

別の態様では、これは、これらの他のユーザがグラフを見る、修正する、コピーする、又は他の方法でインタラクトすることができるように、1又は複数の他のユーザとの操作のジャーナルの共有を容易にすることを含み得る。これは、例えば、データネットワークを介して、操作のジャーナルのデータ構造を1又は複数の他のユーザと共有すること、又は

50

、他者が使用するために、操作のジャーナルに基づくインタラクティブ知識グラフを発行することを含み得る。

【0087】

別の態様では、単一のユーザが、複数の研究プロジェクトを有し得、各研究プロジェクトは、別々の操作のジャーナルによって表される。従って、方法600は、複数の操作のジャーナルを記憶することを含み得、そのいくつかは、人間のユーザが書き込み許可を有する異なるプロジェクトに関連付けられ得る。ユーザクションがそのようなジャーナル間で正確に配布されることを確実にするために、ユーザインターフェイスは、特定のユーザクションが記録される場所を管理するために、種々のメカニズムのいずれかを採用し得る。例えば、一態様では、ユーザインターフェイスは、特定のプロジェクト又は操作のジャーナルの明示的な選択を要求し得、このプロジェクトは、ユーザが異なるプロジェクト又は操作のジャーナルを指定するまで、例えば、異なるデバイス、ドキュメント、アプリケーションなどにわたるすべてのユーザインタラク션을捕捉するために使用され得る。別の態様では、システムは、アプリケーションの変更、デバイスの変更などに基づいて変更を推測し得る。このように、方法600は、1つ又は複数の表面上の人間のユーザのアクティビティをモニタし、現在のアクティビティを記録するために複数の操作のジャーナルの1つを自動的に選択すること、又は、様々な条件下で若しくは様々な事象に応答して、複数の操作のジャーナルの1つのユーザ選択を要求することを含み得る。別の態様では、システムは、複数のジャーナルに操作を記録し得る。例えば、ユーザは、自身のすべての操作が記録されることを可能にする長期にわたる履歴ジャーナルと、ユーザの操作のサブセットも受け取る特定の探索又はタスクのための複数の他のジャーナルとを有し得る。

10

20

【0088】

別の態様では、操作のジャーナルは、他の種類の処理にも使用され得る。例えば、方法600は、操作のジャーナルにおけるデータの集約(aggregation)に基づいて、概念及び第2の概念に言及するドキュメントのリストをランク付けすることを含み得る。これは、例えば、知識グラフ又はドキュメントコンテンツの構造を明示的に指定するデータを含む。これはまた或いは代わりに、関係を意味する(implying)データを含み得る。例えば、ユーザアクションのコンテキストは、推薦のための要求が行われるコンテキストのような、関連性を意味し得る。別の態様では、ユーザが知識グラフに概念を追加した順序、又はグラフのエッジの実証を要求した順序など、ユーザクションの順序は、関連性を意味し得る。別の態様では、ユーザがグラフの特定のエッジの実証を要求する頻度又はユーザがグラフ内の頂点によって表される特定の概念に関連する概念を要求する頻度など、ユーザアクションの頻度は、関連性を意味し得る。

30

【0089】

上記によれば、コンピュータ支援知識発見プロセスにおけるアクティビティをジャーナルするためのシステムも本明細書に開示されている。一般に、システムは、データストア、データストアに記憶された操作のジャーナル、及びディスプレイとプロセッサを有するデバイスを含み得る。

【0090】

データストアは、例えば、本明細書において説明されているデータベース、データストア、データリポジトリ、又は他のメモリ等のいずれかのみならず、上記のものの組み合わせであってもよい。一般的に、操作のジャーナル(journal of operations)は、プロジェクト(project)とのヒューマンインタラクション及びマシンインタラクションの累積されている記録(accumulated record)を含んでもよく、それらの累積されている記録は、次に、操作のジャーナルの中に格納されている情報の知識グラフ(knowledge graph)又は他の視覚的な表現としてレンダリング(rendered)されてもよい。例えば、操作のジャーナルは、1つ又は複数のドキュメント及び1つ又は複数の概念等の表示の際における視覚的な表現との1つ又は複数のヒューマンインタラクション(human interactions)及び1つ又は複数のマシンインタラクション(machine interactions)の記録を含んでもよい。データストアは、デバイスがデータネットワークを通じてアクセスすること

40

50

が可能であるリモートデータストアであってもよく、共有されていてもよく、(例えば、他のユーザによるアクセスから保護されていてもよいといったように)プライベートであってもよく、又は、それらの何らかの組み合わせであってもよい。他の態様において、データストアは、表示と関連しているデバイスにおけるローカルデータストアであってもよい。

【0091】

デバイスのプロセッサは、操作のジャーナルの視覚的な表現の生成又は表示等の複数の関連する機能とともに、ジャーナルの作成及び使用を支援するように構成されてもよい。例えば、プロセッサは、操作のジャーナルに基づいて、表示の際にプロジェクトの視覚的な表現をレンダリングするように構成されてもよく、プロセッサは、ディスプレイ表示の際にその視覚的な表現とのユーザインタラクションを受信するように構成されてもよい。操作のジャーナルを維持するために、プロセッサは、操作のジャーナルにユーザインタラクションを追加するように構成されてもよい。例えば、データストアがローカルである場合に、プロセッサは、ジャーナルに記録を追加するプロセス等を直接的に実行してもよい。これに対して、データストアがリモートデータストア等である場合に、プロセッサは、デバイスが、データストアに適切な命令及び他の情報を送信して、遠隔位置にある操作のジャーナルの中にユーザインタラクションの記録を作成する、ようにさせてもよい。プロセッサは、また、推薦エンジン(recommendation engine)にユーザインタラクションと関連する推薦(recommendation)を要求する機能及び推薦エンジンによる操作のジャーナルへの応答を追加する機能等の他の関連する機能を実行してもよく、その推薦エンジンは、デバイスにおいて実行されるローカル推薦エンジンであってもよく又はデータネットワークを通じてアクセスすることが可能であるリモート推薦エンジンであってもよい。

【0092】

表示の際にレンダリングされるプロジェクトの視覚的な表現は、任意の適する形態をとってもよい。このことは、例えば、知識グラフを形成するアイコン又は頂点を含んでもよい。このことは、また、又は、代わりに、以下で説明されているドキュメントの抜粋を提供する略式のカードを含んでもよく、複数のドキュメントの中の潜在的に関連性のあるコンテンツのハイライト又はその他の視覚的なインジケータ、及び、言及及び概念等に関する略式の情報に言及してもよい。より一般的には、役立つようにユーザに情報を提示し及び/又はプロジェクトに関連するユーザの入力を容易にする操作のジャーナルにおける記録のいずれかの視覚的な表現は、本明細書において意図されているように、プロジェクトの視覚的な表現の中で使いやすいうように利用されてもよい。

【0093】

本明細書において説明されているシステム及び方法は、コンピュータによって支援される探求及び知識発見(computer-assisted research and knowledge discovery)を支援するために使用されてもよい。一般的に、コンピュータによって支援される探求及び知識発見は、本明細書において説明されているさまざまなアーキテクチャの特徴を含むシステム700によって支援される。例えば、知識グラフ又は同様の構成は、ユーザデバイス704のユーザインターフェイス702の中で、相互に関連する概念(interrelated concepts)の明示的な且つ視覚的な表現を使用して、支援用のドキュメント及びテキストへの便利なアクセスを容易にするとともに、ドキュメント710の全集(コーパス)(a corpus of documents 710)とのヒューマンインタラクション及び人間の理解を容易にする。同時に、例えば、推薦エンジン(recommendation engine)及び複数の検索エンジン(search engines)等を含むマシン分析システム(machine analysis system)706は、ドキュメント710の利用可能なコーパスの中の複数のエンティティー(entities)及び複数の概念(concepts)の関係に基づいて、新たな関係(relationships)及び洞察(insights)の発見を容易にする。操作のジャーナル708又は同様のデータ構造によって、この知識グラフを使用するヒューマンインタラクション及びマシンインタラクションを支援してもよく、その操作のジャーナル708又は同様のデータ構造は、知識グラフのためのコンテキスト及び履歴を保存し、それらのコンテキスト及び履歴は、一方では、複

数の人間ユーザの間の共有及び共同作業を支援する方法で集約されてもよく、他方では、マシナルゴリズムによる改善されたコンテキストの分析を支援する方法で集約されてもよい。

【0094】

このアーキテクチャを使用するマシンによって支援される知識発見プラットフォーム(machine-assisted knowledge discovery platform)のさらなる特徴及び態様は、本明細書において使用される複数の用語の複数の代表的な説明から開始して、以下において説明される。

【0095】

本明細書において使用されているように、推薦エンジンは、いずれかのシステム又は方法であってもよく、そのいずれかのシステム又は方法は、1つ又は複数の人間ユーザからのコンテキスト情報を解釈し、そして、ユーザが、例えば、推薦ドキュメントを開いて読むことによって、その推薦に基づいて動作することを選択するであろうアルゴリズム的予測(algorithmic prediction)に基づいて、ドキュメントのコーパスからコンテンツを推薦することによって応答する。

【0096】

"ユーザ"は、プログラム、Webサービス、又はディスプレイ等のツールと相互作用する人間又はコンピュータプロセスである。

【0097】

"インテリジェントアシスタント(intelligent assistant)"又は"インテリジェント仮想アシスタント(intelligent virtual assistant)"としても知られている"コラボレーションエージェント(collaborative agent)"は、いずれかのソフトウェアシステムであり、そのソフトウェアシステムは、電話での音声ダイアログ、データフィード(data feed)、電子メール、チャットアプリケーションにおけるテキストメッセージ等の通信の1つ又は複数のチャンネルを通じてユーザと対話する。典型的なコラボレーションエージェントは、単一チャンネルの限定された範囲に集中する。本明細書においては、人間及びコラボレーションエージェントの双方は、本明細書において考慮されている知識オペレーティングシステム(knowledge operating system(KOS))のユーザであるため、"コラボレーションエージェント"と相互に交換可能な方法で"マシンベースのユーザ"を使用する。

【0098】

本明細書において使用されている表面は、ユーザとの通信のためのいずれかの媒体を含んでもよい。表面は、いずれかのデバイス、スクリーン、アプリケーションウィンドウ、音響環境、振動アクチュエータ、及びユーザと対話するためのいずれかの他のセンサアクチュエータメカニズムを含んでもよい。広く使用されているタイプの表面は、コンピュータにおけるウェブブラウザ、マイクロソフトワードアプリケーション、PDFビューア、又は本明細書を読むことが可能である他のプログラム等のアプリケーションウィンドウである。"視覚的表面"は、視覚的構成要素を有するいずれかのそのような表面を含んでもよい。

【0099】

典型的には、人間ユーザは、1日を通じていくつかの表面を使用する。例えば、ラップトップコンピュータにおいて同時に電子メールプログラム及びウェブブラウザを開くこと、及び、また、携帯電話の通話機能を動作させているのと同時にチャットダイアログを同時進行で開くことはよく行われることである。これらの多くの表面は、ユーザが何をしているかを理解するための機械学習アルゴリズムにとって非常に大きな価値を有している。この文脈から、コラボレーションエージェントは、ユーザが次に必要とするものを推測することを試みてもよい。

【0100】

例えば、推薦エンジンによって"コンテンツを推薦する"ことは、ユーザが以前に実現したことがない場合があるか又は認識したことがない場合があるコンテンツにユーザがアクセス

10

20

30

40

50

スするのを支援するのを試みることを意味し、ユーザにとって役立つであろう。コンテンツの推薦のある1つの典型的な目的は、推薦用のディスプレイの中に十分な情報を提供して、推薦されているある1つのコンテンツの潜在的な有用性を特徴づけることである。推薦を取り巻くこのような状況は、人間ユーザが、推薦を受け入れるか又は推薦を使用することの利点を認識するのに役立つ場合がある。

【0101】

本明細書において考慮されている"知識"は、少なくとも概念及び複数の概念の間の関係を含む。知識は、複数のエンティティの間の関係を含んでもよく、複数のエンティティの間の関係は、ある1つの特別なタイプの概念である。"エンティティ"は、故郷、電話番号、記事の数、及びDNA等の強く定型化されている属性によって区別されてもよい。名付けられたエンティティは、一般的に、一意ではないが、ある与えられた名前を有する。エンティティの代名詞に言及する場合は、(例えば、彼、あなた等の)記述の事項を指す。名目上エンティティに言及する場合は、名称を提供することなく、例えば、"一般"又は"3つのジャーナリスト"等のある特定のエンティティを説明する。エンティティークラスの例は、人、会社、又は組織、施設又は建物、車両、(例えば、電話番号、スカイプハンドル、電子メールアドレス、IPアドレス等の)デバイス識別子又はサイバー空間の識別子、化学化合物、惑星オブジェクト、及びタンパク質を含む。

10

【0102】

"ドキュメント"は、概念の言及を含むデジタルメディアオブジェクトである。ドキュメントの印刷された形態等の非デジタル記録は、ある1つのアーティファクトであるにすぎず、本明細書において一般的に説明されているような語の意味でのドキュメントではない。ドキュメントは、多くのコンテキストで表示される。このことは、例えば、ユーザが参考資料として使用してもよいソースドキュメント、及び、例えば、電子メールスレッドの中の返事等のユーザが作成する作業ノートを含む。

20

【0103】

ソフトウェアシステムは、しばしば、さまざまな形態で言及と相互作用し、したがって、本明細書においては、言及を識別するための複数の概念及び用語を説明する。"表面形態"は、少なくとも一部のユーザがある概念を指していると認識するであろう画像又は他の未処理のデータ表現における文字のストリング又は音のシーケンス又はピクセルのセットである。上記で述べられているように、概念はコンテキストの中でその表面形態によってのみ定義されるという立場をとる。したがって、ある概念を識別するためには、ドキュメント及びそのドキュメントのうちの言及が存在する部分を識別する必要がある。多くのドキュメントは、階層的な見出し及び表等の豊富な構造を含み、これらは、ドキュメントの複数の部分を識別するのに使用されてもよい。多くの場合、表面形態の言及は、あるドキュメントの中の文字又はピクセルの特定の部分的範囲であり、これらの範囲は、"スパン"と呼ばれることがよくある。ドキュメントから言及のスパンを自動的に選択するための多くのアルゴリズムが開発されている。

30

【0104】

"ドキュメントのコーパス(corpus of documents)"は、ポータブルネットワークグラフィック(Portable Network Graphic(PNG))フォーマットの画像、.docxフォーマットのマイクロソフトワードファイル又はポータブルドキュメントフォーマット(Portable Document Format(PDF))、ハーバーテキストマークアップ言語(Hyper Text Markup Language(HTML))で書かれているウェブページ、又は多くの他のデータフォーマットのいずれか等のデータファイルの集合である。コーパスの中のドキュメントは、画像、ビデオ、オーディオ、テキスト、時系列、数値マトリクス、又は、複数のコンピュータシステムの間で送信することが可能であるようにバイナリ形態で表現可能ないずれかの他の情報を含む多種多様のデジタル媒体を含んでもよい。紙ドキュメントでさえ、紙を走査して画像を作ることによってコーパスの中で表現されてもよく、その作られた画像は、その次に、光学的文字認識等によってテキストに変換されてもよい。ドキュメントのコーパスは、新たなドキュメントが極めて頻繁に作成されるパブリックウェブのよう

40

50

に無制限のサイズであってもよい。

【0105】

テキストのドキュメントは、オーディオドキュメントによって話すことが可能である複数の単語を含んでいることが多い。そのような話すことが可能である情報は、その構造が、コンピュータ時代以前に人間の会話によって出現した自然現象であるため、自然言語と呼ばれることが多い。ドキュメント中の他のデータは、ハイパーテキストマークアップ言語(HTML)のような、機械による解釈のための構造化された情報を記述し、HTMLは、どのように情報を表示するかをウェブブラウザに伝達するタグを含む。典型的なウェブページは、構造化されたHTML及び自然言語の双方を含む。

【0106】

テキストの中での言及のスパンを識別するための広く使用されている技術は、文字の単一且つ連続した配列又はユニコードコードポイント(Unicode code points)として、ドキュメントのテキスト全体を記述することである。コードポイントのこの配列の中の開始インデックス及び終了インデックスによって、言及のスパンを一意に識別することが可能である。豊富な構造を有するドキュメントでさえ、単一のストリングへとシリアル化することが可能である。例えば、HTMLとしてウェブページを表現することが可能であり、ドキュメントの単一の且つ連続するストリングとして、HTMLドキュメントのストリングを使用することが可能である。同じドキュメントにおいて複数の言及のスパンの選択アルゴリズムを実行するために、それらの識別子が、コードポイントの同じアレイを使用するということを保証することは、役立つことが多い。このことは、複数の異なるアルゴリズムが同じ又は重複する言及スパン(the same or overlapping mention spans)を選択するとき、検出を容易にするとともに、複数の異なるアルゴリズムが複数のスパンを選択するときであっても、複数の言及スパンの順序決定を保存することを可能とする。例えば、電子メールアドレスを検出するための正規表現(regular expression)及びスキップチェーン条件付きランダムフィールドモデル(skip-chain conditional random field model)等の統計的シーケンスラベル付けモデル(statistical sequence labeling model)の双方を実行することが可能であり、そして、コードポイントの同じ根拠アレイ(same underlying array)においてこれらの双方を実行することによって、それらの2つのアルゴリズムが選択する言及スパンを組み合わせ、単一のセットとすることが可能である。ある特定のアルゴリズムが、HTMLマークアップタグを適切に処理することが不可能である場合に、後続の自然言語テキストの複数のアレイ位置が変更されないように、これらのタグを空白文字コードポイントで置き換えることが可能である。本開示は、"キャラクターオフセット(character offsets)"として、そのようなアレイの中のインデックス位置に言及してもよい。例えば、本開示が、位置ゼロとして前の文の中の文字"W"を扱う場合に、"refer"の語は、キャラクターオフセット[9,13](最終インデックス位置は、言及スパンの中に含まれる)又は[9,14](最終インデックス位置は、言及スパンの中に含まれない)を有する。

【0107】

ユニバーサルリソースロケータ(URL)は、ワールドワイドウェブ(WWW)及び他のコンテンツ管理システム(CMS)等のネットワーク接続システムにおけるドキュメントを識別するのに広く使用されている。IETF RFCs(<https://tools.ietf.org/html/rfc3986>及び<https://tools.ietf.org/html/rfc1738>)等のさまざまな標準化ドキュメントの中で定義されているように、URLは、ドキュメントにアクセスするための通信"スキーム"、"ホスト名"、及びファイル"パス"を提供する。URLは、問い合わせパラメータやハッシュフラグメント等の追加のデータ要素を含んでもよい。実際には、URLストリングは、可変の長さを有していることが多く、したがって、ソフトウェアシステムの動作上の問題を引き起こす場合がある。これらの問題に対処するための一般的な技術は、MD5、SHA1、又は、SHA256等の低い衝突率の一方向性ハッシュ関数を適用し、それらの一方向性ハッシュ関数が、16進数で簡単に表現することが可能である固定の長さのストリングを生成することである。ハッシュは、きわめてまれにしか衝突しないので、それらのハッ

10

20

30

40

50

シュは、ドキュメントのための一意の識別子であるかのように使われることが多い。本開示は、そのようなハッシュベースの識別子HashDocIDに言及するとともに、一般的に、"docID"としてのドキュメントのいずれかの種類のストリングベースの識別子に言及する。

【0108】

したがって、テキストによる言及は、docID及びキャラクターオフセットの組み合わせによって識別されてもよく、それらのキャラクターオフセットは、そのドキュメントの中の言及スパンの開始位置及び終了位置を識別する。そのような組み合わせは、"言及識別子(mention identifier)"又は"言及ID(mentionID)"と呼ばれる。例えば、"...を指すことが多い"によって終了する例示的な文章を再び考慮する。その1つの文章のみを含むドキュメントを想定し、そのドキュメントが、MD5ハッシュがda98292ac64ea61dc4706dac2a48881aであるURL"file://server/doc.txt"を有しているということを仮定する。このようにして、そのドキュメントの中の"refer"という語の言及IDは、"da98292ac64ea61dc4706dac2a48881a:9,14"であり、コロン(:)は、キャラクターオフセットの開始位置及び終了位置からdocIDを区切るのに使用される。

【0109】

同様の考察は、写真及びビデオの中の部分画像(sub-images)及びオーディオトラックにおける発話に有用な且つ具体的な識別子(usefully concrete identifiers)を提供する。そのような言及の識別子は、言及IDと称される。複数のエンティティーは、概念のサブセットとなっており、したがって、ある1つのエンティティーは、言及IDによって識別される。本明細書において複数の実施形態を説明するために、言及は、説明し又は取り扱うことが必要となるであろうエンティティーのただ1つの具体化(materialization)又は明示化(manifestation)である。明確化のために、歌手のBlack Francis等のある1つの特定のエンティティーの言及IDを指すために、システムは、"MentionID(Black Francis)"と記載してもよい。このことにより、言及IDを使用する複数の例をより簡単に明確化することが可能である。

【0110】

関係は、エンティティーの一種の属性であってもよい。関係は、複数の概念である2つのエンティティーによって識別されてもよく、したがって、それらの2つのエンティティーの各々について少なくとも1つの言及を行うことによって具体化される。すなわち、言及IDの対は、関係を識別するために必要な情報であってもよい。

【0111】

関係は、"タイプ"を有するように説明されてもよく、その"タイプ"は、オントロジーと呼ばれることがよくあるある特別な種類のドキュメントによって定義される。オントロジーにおける"のメンバー"等のさまざまな概念の言及は、"Black Francisは、ピクシーズのメンバーであった"等の複数のエンティティーの間の関係を記述するのに使用されてもよい。2タプルの(MentionID(Black Francis), MentionID(The Pixies))は、より基本的なオブジェクトであってもよい。タイプとの関係を参照するために、その関係のタイプを定義するオントロジーの一部及びオントロジードキュメントのための識別子のみならず、エンティティーの対を識別する2つのタプルの双方を識別することが可能である。

【0112】

2つの概念の間の関係を具体化する(substantiates)ドキュメントは、その関係のタイプの証拠を提供することが可能である。たとえば、人体が感冒(病気)を引き起こすウイルスと戦うのをビタミンCがどのようにして支援するのかを説明するジャーナルの論文を考えてみる。そのようなドキュメントは、概念"ビタミンC"及び概念"感冒"の関係を示す証拠を提供する。そのドキュメントの中の他の概念の言及は、その関係の特徴づけるコンテキストを提供する。

【0113】

同一指示決定(Coreference resolution)又は"coref"は、言及に意味を割り当てるブ

10

20

30

40

50

ロセスである。一般的に、同じ表面形状名(same surface form name)を有する2つの言及は、同じ概念を指す可能性がある候補(candidates)である。それらの2つの言及を取り巻くコンテキストが同様である場合に、人間は、その言及が同じ意味を有する、すなわち、同一指示的なものであると認識する可能性がより高い。アルゴリズムは、2つの言及を取り巻くコンテキストにおけるデータを比較し、そして、その言及が同一指示的なものであると人間が認識するであろう可能性を推定することが可能である。

【0114】

このプロセスは、言及を解釈する人間ベースのエージェント又はマシンベースのエージェントにおいて生起してもよい。そのプロセスは、複数の言及が同じ概念を指す、すなわち、同一指示であるということを断言することによって、曖昧さを減少させることが可能である。この曖昧さの減少は、各々の言及に意味を割り当てるアクター(actors)によって経験される。意味や真理を直接的に知ることができるものはだれもいないので、このことは微妙である。むしろ、証拠から真実を推論してもよい。このようにして、意味を割り当てるという文字通りのアイデアは、得ることが可能ではない場合はある。むしろ、曖昧さの除去の操作化可能な概念は、同一指示決定である。あるアクターが、ある1つの特定の言及が他の言及と同じ概念を指していると信じている場合に、そのアクターは、それらの2つの言及が同一指示である又は同一指示的である、すなわち、同じ意味を有するということを信じる。一方のものが、同一指示決定及びその関係のうちの一部の認識によって、他の概念とXを関連させるときに、その一方のものは、"Xを理解している"又は"Xの意味を知っている"と言う。同じ概念の他の言及にある1つの言及を結びつけることは、"coref ed"となった後に、その意味を解決すること、又は、"coref"することである。

【0115】

エージェントの"Coref"は、アクターを識別するので、エージェントの"Coref"は、余剰の重みを搬送する。アイデンティティは、現実の世界のアクターにとって多くの実際的な意味を有してもよい。多くのオンラインサイバーアクターは、自身のアイデンティティを隠しているので、ペルソナ(persona)という特別な語は、ある1つのオンラインアクター指すのに使用されてもよく、他のアクターは、いまだに、正確には、他のペルソナとそのオンラインアクターが"coref ed(同一指示である)"とすることはできない。

【0116】

概念及びエンティティーに関する情報を発見する際に、いくつかのタイプのドキュメントに遭遇する場合がある。このことは、例えば、"第一次情報源(primary source)"ドキュメントを含んでもよく、その第一次情報源ドキュメントは、通常は、例えば、あるイベントの写真又はあるインタビューの台本等のさまざまなイベントと同時に、複数のアクターの間の通信のために作成されるアーティファクトである。このことは、また、"第二次情報源(secondary sources)"を含んでもよく、その第二次情報源は、通常、人とのインタビュー又は紙のアーティファクトのスキャン等の単一の第一次情報源ドキュメントをドキュメント化する(documenting)単一の情報源報告書等の典型的なインテリジェンス情報誌等の以前のイベントに依存するか又は焦点を当てている以降のイベントの際に、第一次情報源ドキュメントを分析する。新聞記事は、第二次情報源材料の一般的な例である。"第三次情報源"は、一般的に、第二次情報源ドキュメントからの情報を編集して、例えば、多くのWikipedia記事等の適切に書かれている百科事典等のこれらの分析の要約又は圧縮版を提供する。

【0117】

本明細書中において使用されているように、コンテキストは、ドキュメントの一部であってもよく、そのドキュメントの一部は、そのコンテキストの中での言及の意図された伝達を人が理解することを可能とするか又は理解することを支援する。コンテキストは可変になっている。一般的に、コンテキストとしてドキュメントのより大きな部分を含めることは、読者がより多くのことを理解することを可能とするであろう。一方で、より大きなコンテキストは、人間の読者によって要約し、そして、機械によって処理するのにより多くの時間を必要とする。例えば、検索結果リストのテキストの断片は、ドキュメント全体に

アクセスするためにリンクを開くか否かを効率的に決定するのに十分なコンテキストを読者に与えることを意図している。本明細書において使用されているように、コンテキストは、また、アクションが取られるユーザコンテキスト又はコンピュータコンテキストを含んでもよいということが理解されるであろう。このことは、例えば、要求を作成するデバイス又はアプリケーション、及び、(日時等の)いずれかの他の観察可能な物理的コンテキスト、(ネットワークアドレス、実行プロセス等の)計算的コンテキスト、及び、(知識グラフの位置、情報要求のタイプ等の)探索コンテキスト等を含んでもよい。

【0118】

引用は、文献への参照である。URLは、引用の一種である。しかしながら、引用文献は、また、或いは、むしろ、読者が特定のバージョンのドキュメントの曖昧さを取り除くように何らかの情報を提供してもよい。あるドキュメントが他のドキュメントへの引用を含むときに、その引用のコンテキストは、双方において言及されている複数の概念を示してもよい。

10

【0119】

本明細書において考慮されているこのシステムは、有用性関数を使用してもよく、その有用性関数は、一般的に、潜在的に有用な情報についての情報源を評価するユーザの黙示的な有用性関数(implicit utility function)を模倣する(mimic)ように適合されていてもよい。ある1つの態様において、このことは、ある特定のユーザの探索活動を動的に模倣してもよく、それによって、マシンは、ユーザの有用性関数を追跡する有用性関数を動的に最適化することによって、ユーザと協働することが可能である。

20

【0120】

このことは、言及の周囲の(ドキュメント内の)コンテキスト((in-document) context)を捕捉し、そして、さらに、ある特定の言及がユーザの興味を高めるか否かについてのユーザからのフィードバックを捕捉することを含んでもよい。このことは、ソフトウェアエージェントの有用性関数が、ユーザの現在の作業コンテキストによって定義されてもよいということを意味する。具体的には、ソフトウェアエージェントの目的関数は、ユーザの作業コンテキストの中で言及されているエンティティとユーザの作業コンテキストの中にまだ存在していない他のエンティティとの間の関係の証拠を見つけることが可能であってもよい。そのような関連するエンティティ及び関連するコンテンツを動的にランク付けし直すことにより、システムは、ユーザの現在のプロジェクトにそれらのドキュメントを追加する(引用する)ようにユーザを動機付けることが可能である。本明細書において考慮されている推薦エンジン又は他のコンピュータによって支援される発見ツールは、このようにして、リモートシステムの照会、ドキュメントの検索、コンテンツの要約、エンティティの同一指示化、及び人間の共同作業者の作業コンテキストの間のずれの埋めようとする試みの人間のプロセスを模倣することが可能である。

30

【0121】

知識グラフ又は他の視覚的な表現は、さまざまな視覚的表示要素を有効に利用することが可能である。ある1つの態様において、このことは、さまざまなドキュメントのコンテンツが具体化する(substantiated)複数の言及等のドキュメント及び概念の間の複数の言及を表す複数の端部(edges)とともに、(例えば、アイコンとして)ドキュメント又は概念を表す複数の頂点(vertices)を含んでもよい。さまざまな追加的な有用な視覚的表示要素又は視覚的知識要素を以下で説明する。

40

【0122】

1. コンテキストの中の言及ハイライト: テキストに最も近いものは、視覚的なハイライトであり、それらの視覚的なハイライトは、テキスト表示システムのフィールドの中で広く使用される。視覚的なハイライトは、何らかの目的のために、関心のある言及スパンを構成する文字のあるスパンの中で、文字の色又は色合いのコントラスト或いはその他の視覚的な側面を変化させてもよい。ハイライトは、また、或いは、むしろ、画像において使用されてもよく、形状(shapes)又はポリゴン(polygons)の輪郭は、ある画像の頂部に(on top of an image)レンダリングされて、関心のある部分画像(sub-images)

50

又は他の画像コンテンツを識別することが可能である。オーディオトラックは、(例えば、注釈によって)いずれかの適切な視覚的手段、音声手段、又はテキスト手段が識別する複数のサブセグメント(subsegments)によってセグメント化されてもよい。ビデオの中の複数の概念の言及をハイライトすることは、例えば、視覚的インジケータ及びオーディオインジケータによって簡単にすることが可能である。コンテキスト内の言及ハイライトについては、ビューア又はエディタに完全な情報源ドキュメントをロードしてもよい。このことは、一般的に、ドキュメントの生来の形態(native form)である。例えば、ウェブブラウザは、何らかのURLからロードされるHTMLウェブページ又はPDFドキュメントを表示することが可能であり、そのドキュメントのドキュメントオブジェクトモデル(document object model(DOM))の中に複数のハイライトを挿入する(injected)ことが可能である。たとえば、"dossier/html-highlighter"において、DARPA Memexの一部として作成されているとともに、GitHubでオープンソース化されているHTML Highlighterプロジェクトにおいて、ある1つの適切なハイライト挿入ツールが説明されている。これは、ドキュメントのコンテキストにハイライトを挿入するツールのある1つの例である。他のハイライト挿入ツール保、公に利用可能となっており、ドキュメントの中の概念の言及をハイライトする際の使用に適しているさまざまなリソース及び技術は、当業者によって容易に理解されるであろう。

10

【0123】

2. 要約カード: 検索結果の断片のリスト(lists of search result snippets)は、"キューの中の要約カード(summary cards in a queue)"と呼ばれるもののよく知られた形態である。要約カードの他のよく知られた形態は、Kanban-style work board in Trello or Jira or Zenhubの中のドラッグ可能なカード(draggable cards)であり、GitHubにアドオンする。そのようなカードベースの表示は、あるドキュメントから抽出されるか、又は、要約を生成するアルゴリズムによって要約として作成されるかのいずれかであるコンテンツの一部を示し、ドキュメントの中で言及されている概念のサブセットを有するドキュメントの表現がコンパクトな形態でユーザにハイライトされているということをユーザが理解することができるように表示することが可能である。典型的には、このコンパクトな形態により、ユーザは、視覚的な表示における1つのビューにおいて、複数のドキュメントの表現を同時に見ることが可能である。例えば、要約カードは、ある1つのドキュメント又は複数のドキュメントの中で言及されている1つ又は複数の概念の表面形態(surface forms)を示すことが可能である。要約カードは、複数のドキュメントを表してもよく、例えば、カードは、複数のドキュメントのすべてが、ともに、同じ2つの概念に言及しているということを示すことが可能であり、このようにして、複数の概念の間のある1つの関係の証拠(evidence)を提供する。要約カードは、また、画像、部分画像(sub-images)、或いは、ビデオ又はサウンドトラックの複数の部分を表示することが可能である。"断片(snippet)"という語は、あるドキュメントから抽出され又は要約され、要約カードの中に表示されるコンテンツの部分を指すのに使用されてもよい。要約カードの視覚的な表示は、リスト又はタブの形態、或いは、複数の列の表示で表示されてもよい。カードに対するユーザのアクション(user actions)は、キューの先頭の近傍にある場所から(from positions near the top of the queue)カードを除去することによってそのリストを選別することを可能とするため、要約カードのリストは、"キュー"と称されてもよい。

20

30

40

【0124】

3. グラフの頂点: 線で結ばれたアイコン又はシンボルの集合で、ドキュメントからの情報の視覚的な表現をレンダリングすることが可能である。そのようなグラフ表示における複数のアイコンは、複数の特定の概念又は概念の集合を表すことが可能である。アイコンは、ドキュメント又はドキュメントの特定の版又はバージョン、特定の要求者のための特定の時間におけるURLからのコンテンツのスナップショット、或いは、いずれかの他のドキュメント又はドキュメントの部分を表してもよい。アイコンは、また、或いは、むしろ、1つ又は複数のドキュメントからの言及の集合に基づいて、概念を表すことが可能

50

である。例えば、本開示は、概念"グラフ"に言及し、この関係の視覚的表示は、表面形態"グラフ"を使用してラベリングされている概念頂点にドキュメントアイコンを接続する線を使用してドキュメントアイコンをレンダリングすることによって、グラフの中の複数の頂点として描写されてもよい。グラフに対するラベリングは、コンテキスト内のハイライト及び要約カードが、ユーザにコンテキストを提供するのとほぼ同じ方法で、ユーザに視覚的コンテキストを提供してもよい。そのようなグラフ表示における複数の線は、さまざまな視覚的品質、特性、ラベル、及び他の記号論を有してもよい。例えば、ドキュメントの中で言及されているある概念に、あるドキュメントアイコンを結びつける線は、その言及されている概念を指す矢印を有する破線であってもよい。複数のドキュメントの中でも言及されている複数の概念に接続する線は、複数の異なる色に割り当てられているさまざまな意味を有する実線又は色付きの線として描写されてもよい。

10

【0125】

複数のコンテキスト内のハイライトは、コンテキストの中での言及スパンを識別することが可能である。要約カードは、複数のドキュメントを描写するビューにおいて関心のある(又は、潜在的な関心のある)言及を識別することが可能である。ある1つのグラフは、これらの複数のアイテムの間の関係を描写する複数の端部によって、概念/言及及びドキュメントを表すのに複数の頂点を使用してもよい。視覚的抽象化(visual abstraction)の3つのレベルのすべては、本明細書においては視覚的知識要素(visual knowledge elements)と称されてもよく、知識/情報の表示のみならず、関心のある言及を識別する入力、ユーザが関心がないと決定する言及を退ける入力、及び推薦を要求する入力等のユーザの入力をサポートしてもよい。これらの同じ視覚的知識要素は、また、知識構造との自動化されている対話又はコンピュータベースの対話を可能としてもよく、そして、ユーザにとって関心の可能性のある言及のマシンベースの選択及びマシンベースの提示、及びユーザインターフェイス又は他の視覚コンテキストの中の言及の提示への動的な更新等の複数の操作をサポートしてもよい。一般的に、システムがユーザに表示することが可能であるそれらのさまざまな視覚的知識要素は、ある1つのジャーナルからの複数の知識操作レコード(knowledge operation records(KORs))を処理するクライアントアプリケーションが生成する集合体である。例えば、要約カードのキューは、そのジャーナルからの複数の操作を処理するユーザインターフェイス表示アプリケーションによって構築されてもよい。このようにして、より一般的には、本明細書において考慮されている知識オペレーティングシステム(knowledge operating system)は、ジャーナルの中のデータの集合体(aggregations of data)を使用して結果として生ずる知識構造(resulting knowledge structure)のさまざまな表現とともに、操作のジャーナル(journal of operations)への関連するアクション及びドキュメントの組織化をサポートしてもよい。

20

30

【0126】

上記の視覚要素は、知識構造における情報の提示をサポートしてもよい。このシステムは、また、このコンテキストの中でのさまざまなアクションをサポートするための制御を含んでもよい。例えば、ボタン、又は、ドロップダウンメニュー、ドラッグアンドドロップ操作、ダイアル、及びスライダー等の他の制御を使用して、ユーザがアクションを要求することを可能としてもよい。例えば、電子メールプログラムは、"compose"とラベリングされているボタンを提示し、新たなドラフト電子メールドキュメントを開くか、又は、電子メールを送信するための"send"を提示する。そのような視覚的なアクションボタンは、前者がユーザに開始することを申し出るアクションによってラベリングされ、後者がユーザが読み取るか又はアクセスすることが可能である情報源データによってラベリングされているという点で視覚的知識要素とは異なってもよい。

40

【0127】

ある1つの例示的な実装において、ユーザが、コンテキスト内のハイライトを有する言及スパンの近傍にマウスポインタを運ぶときに、アイコン等の視覚的なアクションボタンが表示されてもよく、ユーザがそのアイコンをクリックする場合に、システムは、その言及

50

されている概念のための追加的な処理を有効化してもよい。スパンが視覚的知識要素によって特に識別されていなかった場合であっても、ユーザは、また、文字、イメージ、オーディオ、又は他のデジタルメディアのスパンを選択することが可能である。ユーザがそのようなカスタム言及スパンを選択する場合に、システムは、また、視覚的なアクションボタンを表示してもよく、それによって、ユーザは、ユーザの選択したスパンに基づいて、複数のアクションを開始することが可能である。

【0128】

ある1つの態様において、コンテキスト追跡は、知識発見を容易にするのに使用されてもよい。マシンベースのコラボレーションエージェント(machine-based collaborative agent)は、複数のドキュメントから複数のユーザアクションを受信することによって、ユーザの作業コンテキストを追跡することが可能である。この追跡は、本明細書において説明されている操作のジャーナルの中に記録されてもよく、その操作のジャーナルは、例えば、知識操作の追加のみのログとして実装されてもよく、知識操作は、人間ベースのエージェント又はマシンベースのエージェントがコーパスの中でとることが可能であるいずれかのアクションである。以下の説明において、この操作のジャーナルは、また、"知識操作ジャーナル(knowledge operations journal)"又は"KOJ"と称されてもよい。そのようなジャーナルは、既知の操作のリストを定義し、そして、その次に、タイムスタンプによってデータベースの中に複数の観察されている操作を記録することによって実装されてもよい。上記で説明されているように、各々のレコードは、また、或いは、むしろ、ユーザコンテキスト及び記録されている操作に関連する他の情報を含んでもよい。タイムスタンプを付されている操作のデータベースは、このようにして、プロジェクトを定義し、同等に、"知識プロジェクト(knowledge project)"と称される。

【0129】

重要な利点として、このような方式によって実装される知識プロジェクト(knowledge project)は、複数のドキュメント、複数のアプリケーション、及び複数のデバイス等にわたって、知識プロジェクトに関連する複数のユーザ活動とともに追跡することを可能とする。このことは、また、本明細書において説明されている複数の異なる種類の視覚的知識要素を同期させ及び協調させることを可能としてもよい。たとえば、Firefox等のウェブブラウザの中で1つ又は複数のドキュメントを開き、そして、また、Microsoft Wordで1つ又は複数のドキュメントを開き、そして、また、Outlookで1つ又は複数の電子メールを開き、そして、また、Gmail clientで他の電子メールを開き、そして、また、他のツールによって他のファイルを開く場合があるユーザを考慮する。ユーザが、これらの複数のドキュメントのうちのいくつかに推薦エンジン又はその他のマシンによって支援されるツール等のコラボレーションエージェントを導入する場合に、知識プロジェクトは、これらの複数のドキュメントの中で言及されている複数の概念を追跡することが可能であるとともに、ユーザのアクションに対応することが可能である。このことは、また、タブレット及びスマートフォン等の他のデバイスに対するユーザの複数のアクションを組み込むことを可能とする。ユーザは、また、複数の知識プロジェクトを開始し、そして、さまざまなプロジェクトにさまざまなドキュメントを追加することが可能である。ある1つのドキュメントは、1つ又は複数のプロジェクトの中に存在してもよく、複数のユーザベースのコラボレーションエージェント及び/又はマシンベースのコラボレーションエージェントは、プロジェクトを共有してもよい。

【0130】

プロジェクトへのドキュメントの追加は、対応する知識操作ジャーナルに対する操作であってもよい。ある1つの例示的な実装において、OpDocumentAdd操作は、ドキュメントを自動的に分析して、言及スパンを選択し、そして、OpConceptAdd操作によってプロジェクトに複数の概念を追加するようにコラボレーションエージェントのアルゴリズムをトリガする。

【0131】

ユーザアクション

10

20

30

40

50

【0132】

視覚的知識要素に対するユーザアクションの複数の例は、以下のユーザアクションを含んでもよい。

【0133】

1. 肯定的なアクション: ユーザは、追跡のために言及を選択してもよい。そのようなアクションは、複数の異なる視覚的知識要素を通じて受信されてもよい。ユーザは、ハイライトの近傍に表示されているコンテキスト内のハイライト又は視覚的なアクションボタンをクリックしてもよい。同様に、ユーザは、要約されているドキュメントの中で言及されている概念に言及する要約カードをクリックしてもよい。同様に、ユーザは、視覚的なグラフの中の頂点をクリックしてもよい。

10

【0134】

2. 否定的なアクション: ユーザは、要約カードにある視覚的なアクションボタンをクリックすることによって、コラボレーションエージェントからの提案を拒否してもよく、その視覚的なアクションボタンは、"誤ったエンティティー"又は"興味がない"等のアクションの記述によってラベリングされていてもよい。同様に、ユーザは、赤く着色されているコンテキスト内のハイライトを拒否して、ある特定の提案が役に立たなかったということアルゴリズムに伝えてもよい。同様に、ユーザは、視覚的なグラフ表示の中のある1つの頂点を破棄してもよい。ある1つの特定の实装は、複数の破棄アクションを提案してもよく、したがって、ユーザが、ある与えられた頂点又は端部又は証拠に対するフィードバックについてのより多くの情報を通信することを可能とする。

20

【0135】

3. 中間的な廃棄アクション: ユーザは、複数の自動的に生成された視覚的知識要素が間違っているということをシステムに伝える必要なく、それらの複数の自動的に生成された視覚的知識要素を取り除いてもよい。これらの非肯定的な且つ非否定的なアクションは、ユーザが、ハイライトを無効化し、要約カードを取り除き、又は、グラフ表示から複数の頂点又は複数の端部を削除することによって、視野をきれいにすることを可能としてもよい。

【0136】

4. プロセス開始アクション: 特定のコマンド及び制御プロセスは、複数の視覚的知識要素に基づいてユーザに公開されてもよい。例えば、ある1つの例示的な実装は、"自動構築者"ボタンを提供してもよく、その"自動構築者"ボタンは、ユーザが複数の概念のある特定のセットに関心があるということをユーザがシステムに伝え、そして、したがって、コラボレーションエージェントがそれらの概念に対して利用可能な探索アクションのすべてを実行することをユーザが要求することを可能とする。

30

【0137】

5. 探索アクション: マシンベースのユーザは、言及の表面形態、表面形態の変化させられている変形、又はその言及から導き出される代替的な語を使用する複数の問い合わせ(照会)を自動的に定式化し、そして、リモート検索エンジンにそれらの複数の問い合わせを送信して、言及が以下で説明されているインデックス作成のさまざまな層を通過することが可能であるドキュメントを検索し、そして、最終的にそれらの複数のドキュメントから言及のうちの一部をプロジェクトに追加するということを推薦することによって、その言及に基づいてアクションを実行してもよい。

40

【0138】

6. 言及の推薦(Recommending Mentions): マシンベースのユーザは、複数の他の言及が同一指示的である(co-referent)可能性が高いということをそのマシンベースのユーザが推定するそれらの複数の他の言及を推薦することによって、ある1つの言及に基づいてアクションを実行してもよい。

【0139】

7. 関連するエンティティーの推薦(Recommending related entities): マシンベ

50

ースのユーザは、ある1つの言及との間で同一指示的である可能性が高い複数の言及と同時に生起する(co-occur with)複数の他のエンティティの複数の言及を推薦することによって、その言及に基づいてアクションを実行してもよい。

【0140】

8. アクションの推薦: 人間ベースのコラボレーションエージェント又はマシンベースのコラボレーションエージェントは、極めて広範なアクションを推薦し、そして、開始してもよい。ある1つの側面において、マシンベースのユーザが提案し又は開始するアクションは、複数の言及に応答して生起する。

【0141】

一般的に、知識操作ジャーナルに複数のユーザアクションを追加してもよい。このことは、マシンベースのコラボレーションエージェントが、複数のアルゴリズムを適用して、ユーザが探している知識又はアクションについてのある人間ユーザとの協議(dialog)の履歴を維持することを可能としてもよい。そのジャーナルは、マシンベースのツールが、1つ又は複数のユーザの探索目的を進展させる方法で、ユーザの意図をより良く推測することを可能としてもよい。

【0142】

肯定的なアクション及び否定的なアクションのセットは、2つの言及スパンが同一のエンティティであるということ又は同一のエンティティではないということを示すことを含んでもよい。例えば、要約カードが、ユーザのドキュメントから、John Smithという名前の人物の言及等の言及スパンを識別するときに、コラボレーションエージェントアルゴリズムは、何らかの他のドキュメントの中に"John Smith"の言及を発見し、そして、そのドキュメントについての要約カードを提示してもよい。その要約カードによって又は他のドキュメントの全ビュー形態によって提供されるコンテキスト情報の読み取りから、ユーザは、他のドキュメントの中の言及が、ユーザのドキュメントの中で言及されているJohn Smithと同じ人物を指してはいないということを決定してもよい。この"誤ったエンティティ"の断定("wrong entity" assertion)を示すために、ユーザは、マイナス記号を示すボタンをクリックしてもよい。マシンは、このフィードバックを使用して、将来的な処理における同一指示決定処理(coreference resolution processing)を改善してもよい。反対に、ユーザが、その人物が同じ人物であるということを決定する場合に、そのユーザは、プラス記号を示す視覚的アクションボタンをクリックすることによって、マシンにフィードバックとしてこのことを提供してもよい。そのような複数のアクションは、知識ジャーナルの中の複数の操作の中に格納されている肯定的な同一指示の断定(positive coreference assertion)及び否定的な同一指示の断定(negative coreference assertion)を作成してもよい。

【0143】

そのシステムは、他のユーザアクションから同一指示の断定(coref assertion)を推測してもよい。例えば、上記のJohn Smithの例を継続すると、そのユーザが、ドラフト電子メール等の進行中のドキュメントを編集している最中であり、他のドキュメントに関する要約カードを使用して、引用を作成するということを考える。そのような引用アクションは、OpCitationAddと名付けられてもよい操作によって知識ジャーナルの中に記録されてもよい。以降の処理において、システムは、ユーザが、また、他のドキュメントの中で言及されているJohn Smithが、ユーザのドキュメントが同様に言及しているJohn Smithと同じであるということを信じている、すなわち、2つの言及が同一指示的である(coreferent)ということを信じているということを意味するということに、その操作を解釈してもよい。

【0144】

インデックス層

【0145】

いくつかのデータ構造及びインデックスを組み合わせて使用して、本明細書において考慮されている知識オペレーティングシステムの中でドキュメント検索、概念検索及び関連付

10

20

30

40

50

け、及びライブ知識構造の追跡等を提供してもよい。それらの複数のインデックスから、特に、以下で説明されている層3インデックス(tier3 index)から、複数の表面にわたる人間の操作及びコンピュータの操作のジャーナルを格納する記録を集約することによって、知識グラフ、推薦リスト、及び端部情報又は頂点情報等の知識オペレーティングシステムの他の側面を作成してもよい。

【0146】

"層0インデックス(Tier0 index)"は、コーパスの中の各々のドキュメントの全コンテンツの標準的なインデックスであってもよい。ある1つの態様において、このインデックス(又は、これらの複数のインデックス)は、複数のドキュメントを調べ(looking up)及び検索する(retrieving)ための高速な方法を提供してもよい。例えば、複数のリバースインデックス(reverse indexes)は、いわゆる"全テキスト"検索("full-text" searching)を可能にしてもよく、その全テキスト検索は、ユーザが入力する単語又はフレーズに言及するドキュメントを発見するためのキーワード及びフレーズ照会をサポートする。キーベースのインデックス(key-based index)は、メタデータ等のあるドキュメントと関連する特定の値の正確な一致の探索(exact-match lookups)に基づく探索を可能にしてもよい。ドキュメント及び知識操作ジャーナルとのそれらの複数の技術の使用は、以下で説明されているように、知識操作ジャーナルの有利な能力を可能にしてもよい。本明細書において使用されているように、"標準的なインデックス(standard index)"は、全テキスト検索技術及びキーベースの検索技術のいずれか又は双方を指す。

【0147】

"層1インデックス(Tier1 index)"は、各々の言及連鎖サブドキュメント(mention chain subdocument)の標準的なインデックスである。追加的な特徴データは、NER及び他の言及選択アルゴリズムによって計算されてもよく、それらの複数の言及スパンは、各々の言及連鎖サブドキュメントによって索引付けられているコンテキストの一部として使用され、したがって、互いに近接する2つの名前に言及する複数のドキュメントを効率的に検索するといったように、より洗練された検索照会を可能とする。名付けられたエンティティ認識(Named entity recognition(NER))及び他の自然言語処理(natural language processing(NLP))ツールは、すべての種類のドキュメントに関するメタデータを生成するのに本発明の技術分野において広く使用されている。テキスト、イメージ、オーディオ、及びビデオにおけるデータストリームのセグメンテーション及び概念とエンティティの複数の言及のラベリングのためのアルゴリズムが利用可能である。システムは、それらの複数の強化アルゴリズムを使用して、ドキュメントにおいて、本明細書において"言及連鎖(mention chains)"と称される関心のある言及のセットを識別してもよい。これらは、複数の特定の言及の周囲の"言及連鎖サブドキュメント"又はMCに組み込まれてもよい。これらのサブドキュメントは、複数の言及を取り巻くコンテキストを含み、それによって、(人間ベースの又はマシンベースの)ユーザは、それらの言及の意味を理解することが可能である。

【0148】

"層2インデックス(Tier2 index)"は、階層的凝集クラスタリング(hierarchical agglomerative clustering(HAC))と呼ばれるアルゴリズムによって生成され、そのアルゴリズムは、言及連鎖サブドキュメントを言及のネスト化されているセットとともにグループ化する。階層の中のリーフレベルに近いセットは、そのアルゴリズムが同一指示的なものである可能性がより高いと推定する複数の要素を含んでいてもよい。例えば、同じ表面形態名を有する2つの言及は、同じ概念を指す可能性がある候補である。それら2つの言及を取り巻くコンテキストが同様である場合に、人間は、それらの言及が同じ意味を有する、すなわち、同一指示的であると認識する可能性がより高い。あるアルゴリズムは、2つの言及を取り巻くコンテキストにおけるデータを比較し、そして、人間が、それらの言及が同一指示的であると認識するであろう可能性を推定してもよい。HACが生成するツリー構造は、同じ概念の複数の言及を発見するための高速ルックアップデータ構造を提供することが可能である。そのツリー構造は、言及連鎖サブドキュメントにキーベース

のメタデータとして格納されてもよい。例えば、MCが、リーフセット34に割り当てられ、そのリーフセット34が、サブセット97の中に含まれ、サブセット104の中に含まれ、サブセット14の中に含まれる場合に、そのMCのツリーアドレスは、キーの順序付けられたセットである[14,104,97,34]となる。それらの複数のキーは、高速検索のための検索キーとして検索インデックスのMCレコードに格納される。

【0149】

モデルは、各々のネスト化されているセットについて、同一指示である可能性のしきい値を定義してもよい。例えば、リーフレベルに最も近い定義されているセットは、90%の可能性、次の80%の可能性、及び次の70%の可能性等であってもよい。リーフレベルからより遠いセットは、より不確実となる場合がある。上記の例を参照すると、[14,104,3,402]のツリーアドレスを有する他のリーフは、セット104において最も少ない共通祖先(least common ancestor(LCA))を有し、したがって、70%が同一指示的なものである可能性が高い。

10

【0150】

このことは、以下のような効率的な検索を可能としてもよい。すなわち、HACが、ユーザのドキュメントからの少なくとも1つを含む言及連鎖サブドキュメントの集合に対する操作を実行した後に、ユーザのドキュメントの中で言及されている同じエンティティを指す可能性の高い他の言及が、先祖鍵を共有するドキュメントを要求することによって検索可能となる。

【0151】

20

層3インデックスは、ユーザの多くのウィンドウ、複数のアプリケーション、複数のデバイス、及び複数の知識作業者の間の協調同期を可能としてもよい。一般的に、層3インデックスは、例えば、操作のジャーナル又は同様のデータ構造の中に格納されているとともに、複数の表面にわたる1つ又は複数の人間ユーザ及び1つ又は複数のマシンユーザを含むユーザアクションの記録のシーケンスを含んでもよい。層3インデックスの集約を使用して、カード又はリスト等の関連する視覚的表示要素或いは知識グラフの表示等の知識オペレーティングシステムのさまざまな特徴を導出してもよい。

【0152】

アクチュエータフレームワーク

【0153】

30

アクチュエータのフレームワークは、システムが、データを発見し、そして、プロジェクトの中にそのデータを配列することを可能とする。

【0154】

メタサーチ(Metasearch)は、アクチュエータフレームワークの1つの特定の実装である。メタサーチは、データベースでバックアップされている(database-backed)ウェブサイトの中の検索フォームフィールドに照会をプッシュし(pushes)、そして、データベースでバックアップされているインデックスが返す検索結果の中のリンクからドキュメントを検索することによって、いわゆるディープウェブをクロールする(検索用インデックスに登録するためにファイル又はウェブサイトアクセスする)(crawls)。メタサーチは、操作ジャーナルの履歴を処理して、言及されているストリングを発見し、そして、それらのストリングからの照会を定式化する。メタサーチアクチュエータは、コラボレーションエージェントのアルゴリズムが、(例えば、証券取引委員会又は米国特許庁等の)連邦政府機関の検索エンジン及び商業データベース等の検索エンジンに、それらの照会ストリングをプッシュして、ドキュメントを発見することを可能とする。このプロセスは、人のアクション及び監督の必要性を回避するように自動化されていてもよい。検索エンジンの結果ページ(SERP)に戻されているリンクに自動的にしたがうことによって、メタサーチアクチュエータフレームワークは、プロジェクトに追加することが可能である追加的なドキュメントを取得してもよい。

40

【0155】

アクチュエータフレームワークの他の例は、本発明のアクティブウィキファイアシステム

50

(active wikifier system)である。ウィキファイアは、ウィキ(wiki)等の参照知識ベースにエンティティの言及をリンクするアルゴリズムである。本明細書において考慮されているシステムは、ユーザが、よく知られている参照情報源(reference sources)及びソーシャルメディアの中のプロファイルを検索することによって、プロジェクトに追加する言及に動的に回答し、それによって、そのシステムは、それらの外部の情報源にユーザの言及をリンクすることが可能である。システムは、共通のウィキフィケーションターゲット(common wikification targets)のキャッシュを使用してもよく、それによって、そのシステムは、最も一般的な概念を検索する必要がない。しかしながら、動的なウィキファイアアクチュエータフレームワーク(dynamic wikifier actuator framework)は、コラボレーションエージェントが、ウェブを通じてウィキリソース(wiki resources)の進化状態に遅れないようにすることを可能とする。

10

【0156】

知識オペレーティングシステム

【0157】

知識オペレーティングシステム(knowledge operating system(KOS))は、コンテンツ管理システム(content management system(CMS))であり、そのコンテンツ管理システムは、本明細書において"プロジェクト"と称されるKOJに対するさまざまなアクションを容易にする。ユーザは、新たなプロジェクトを作成し、その新たなプロジェクトに名前をつけ、そして、そのプロジェクトにドキュメントからの言及を追加してもよい。各々のプロジェクトは、ディスクに永続的に格納することが可能であるシリアル化可能なデータ構造の中に個別の明示化(manifestation)を有する。このように、プロジェクトは、新たな種類のドキュメントのようなものである。伝統的なドキュメントが、URL等の引用又は構造化された参照によって他のドキュメントを参照することが可能であるのと同じ方法で、KOJは、また、多くの他のドキュメントへの参照も含んでもよい。ほとんどの他の種類のドキュメントとは異なり、KOJは、KOJが参照しているドキュメントに関連するユーザのアクションを追跡してもよい。いずれかの他の種類のドキュメントとは異なり、KOJは、アルゴリズムがそれらの追跡されているアクションを使用することを可能として、KOJ自身を自動的に拡張してもよい。

20

【0158】

KOSは、JavaScript Object Notationを送信する代表状態転送Webサービス(Representational State Transfer web services)等のプログラムインターフェイスを提供してもよく、そのプログラムインターフェイスは、クライアントのプログラムが複数の表面と対話することを可能とし、それによって、他のドキュメントに対するユーザアクションを追跡することが可能である。KOSは、また、複数のアクチュエータフレームワークを提供してもよく、それらの複数のアクチュエータフレームワークは、複数の自動的なアルゴリズムが、複数の異なるデータ情報源からドキュメント及びデータを検索することを可能とし、それによって、KOJにそれらのドキュメント及びデータを追加することが可能である。KOSウェブサービスは、複数の表面が、同じプロジェクトから情報をレンダリングする視覚的知識要素を表示することを可能とする。さまざまな表面は、KOSインターフェイスを反復してポーリングしてもよく、又は、KOSは、KOSがジャーナルに変更をプッシュすることを可能にするプロトコルを使用してもよい。

30

40

【0159】

KOSは、有利には、ユーザが、複数のデバイス及び複数のアプリケーションにわたって複数のプロジェクトを制御し及び管理し、人的リソース及びマシンリソースとの間で知識発見を共同で実行することを可能としてもよい。伝統的な検索エンジンの検索履歴とは異なり、KOSは、ユーザが、複数のプロジェクトを同時に進行させることを可能としてもよい。ユーザは、いずれの場所に探索アクションを記録するのかが選択してもよく、それによって、各々のプロジェクトは、ある特定の活動領域又は作業領域についてユーザBDIを正確に追跡することが可能である。ユーザのクリック又は同様のアクションの観察のみから、そのユーザがいずれのプロジェクトを追跡しているのかを推測することを試みても

50

よく、このことは、人間がどのようにして他の人間とともに作業をするかということではない。2人の人間がさまざまなプロジェクトに対してともに作業を実行している場合に、一方の人間が他方の人間にいずれのプロジェクトについて自身が考えているのかを伝えることは普通のことである。チームメイト又は同僚が、"今すぐに他のプロジェクトに話題を変えよう..."又は"君も覚えていると思うけど、今度の会議について、重要な要素は..."といい、そして、その他のコンテキスト設定ステートメントをいうのをよく耳にする。このようにして、ユーザは、一般的に、プロジェクトを指定してもよい。他の態様において、KOSは、(例えば、ユーザへの通知なしにといったように)透過的に又は(例えば、明示的なユーザへの通知及び要求によってといったように)対話的に、又は、これらの何らかの組み合わせのうちのいずれかにより、プロジェクトを推測し、及び/又は、複数の話題の間で切り替えを実行してもよい。

10

【0160】

今日のコンピュータのほとんどの人間ユーザは、まるでそれが考えられない無生物のツールであるかのようにコンピュータソフトウェアを操作することに慣れている。インテリジェントな仮想アシスタントを可能にするアルゴリズムの出現により、ソフトウェアは、人間インテリジェントな仮想アシスタントを可能にするアルゴリズムの出現により、ソフトウェアは、人間の共同作業からそのようなプロジェクトコンテキスト設定信号(project context setting signals)を要求し(solicit)及び受信する能力を必要とする場合がある。KOSは、そのようなコンテキスト設定を有効化してもよい。

【0161】

20

ある1つの実装において、各々のユーザアクションは、KOJの中に記録される。ユーザのアクションは、MS Word又はウェブブラウザの中のタブ等の表面で生起する。ユーザアクションは、これらのアプリケーションの中で動作するソフトウェアプラグインによって、又は、親オペレーティングシステムの中で動作するスクリーン画像取り込みソフトウェアによって取り込まれてもよい。以下で、コンテキストを設定するためのアプローチの複数の例示的な実装のうちのいくつかを説明する。

【0162】

ユーザがいずれのKOJがアクションを受信するかを指定することを可能とするために、KOSは、KOJを選択するための視覚的アクションボタンを表示してもよい。ある1つの実装において、例えば、クローム(Chrome)又はファイアフォックス(Firefox)において各々のウェブブラウザタブの中のプロジェクトを選択するための別個のそのような視覚的アクションボタン、或いは、表面又はデバイスの単一の視覚的アクションボタンを表示する。KOSは、特定のユーザデバイスにデスクトップアプリケーション等として展開されてもよく、ユーザが、各々のドキュメント、アプリケーション、又はウィンドウ等にマシベースのコラボレーションエージェントを導入することを可能としてもよい。ある1つの態様において、ユーザは、複数の異なるプロジェクトに各々の新たなドキュメントを表すアイコンをドラッグアンドドロップ(drag-and-drop)してもよく、或いは、ユーザがプロジェクトと関連させることを望む各々の新たなドキュメント、ウィンドウ、又はアプリケーションにプロジェクトを表すアイコンをドラッグアンドドロップしてもよい。

30

【0163】

40

ある1つの態様において、ユーザは、例えば、KOSデスクトップアプリケーションのためのソフトウェア嗜好パネル(software preferences panel)において、自動プロジェクト切り替え(automatic project switching)を起動してもよい。このスイッチが"ON"に設定され、その次に、ユーザがある1つのウィンドウ又はアプリケーションの中でプロジェクトを選択するときに、他の場所で開いているKOSプロジェクトと関連する他の複数のドキュメント又はウィンドウのすべては、また、自動的にそのプロジェクトに切り替わってもよい。これにより、ユーザが、各々のウィンドウを新たなプロジェクトに手動で切り替える手間を省くことを可能とする。この自動プロジェクト切り替えは、ユーザが同じ時点において複数のプロジェクトをアクティブにすることをより困難にする不都合な点を有する場合があり、したがって、その自動切り替え特徴は、容易に無効化される場

50

合があるか、又は、条件付き切り替え特徴が提供され、ユーザは、新たなプロジェクト又はプロジェクトの切り替えを検出するときに、他のウィンドウ又はドキュメントを新たなプロジェクトに切り替えるか否かについて問い合わせられる場合がある。

【0164】

他の例示的な実装において、ユーザは、検索結果のリストからドキュメントを選択し、そして、"プロジェクトに追加"のためのボタンをクリックしてもよい。システムは、既存のプロジェクトの名前を示すセクタを表示してもよく、したがって、ユーザは、いずれのプロジェクトがその選択されたドキュメントから言及を受信するかを選択してもよい。システムは、また、ユーザが新たなプロジェクトを開始することを可能としてもよい。このワークフローにおいて、ユーザがいずれのプロジェクトを確認するためにクリックするか又は新たなプロジェクトを指名するときに、そのシステムは、自動的に、(1) そのシステムが、ドラフトメールを開いて、ユーザが同僚との間でプロジェクトを共有するのを支援する、(2) システムが、ドキュメント及びその他の頂点を含むプロジェクトのグラフビューの中の頂点を開く、という2つのアクションを実行する。グラフビューが開くときに、ユーザが検索プロセスの一部と言及又はストリングを特定した場合は、それらの複数の概念の頂点が自動的に拡張して、関連する概念を表示してもよい。

10

【0165】

システムは、各々のプロジェクトについての要約ビューを示すプロジェクトダッシュボードを提供してもよい。要約ビューは、プロジェクトの中の複数のドキュメント及び複数の概念のうちの一部をリスト化し、プロジェクトのグラフビューの中の頂点を開くためのリンクを提供する。ある1つの態様において、プロジェクトを共有することが可能であり、それによって、複数のユーザは、単一のプロジェクトに参加する。例えば、2人の人間ユーザ及び2つの異なるソフトウェアシステムは、単一のジャーナルに複数の操作を投稿してもよい。

20

【0166】

本明細書において考慮されているジャーナルは、そのジャーナルが新たな操作を継続的に蓄積することが可能であるという意味で、生きているドキュメントであってもよい。複数の他の生きているドキュメントと同様に、ジャーナルは、混ぜ込まれて(forked)、以降に原本(original)から分岐する(diverges)コピーを作成する。また、ジャーナルをロックすることが可能であり、それによって、新たな操作を追加することが不可能であるようにしてもよい。ジャーナルは、複数の形態のアクセス制御許可を有してもよい。読み取り専用の許可を有するジャーナルにアクセスすることが可能であるユーザは、1つのセットのアクセス制御許可によって定義されてもよく、書き込み許可を有するジャーナルにアクセスすることが可能であるユーザは、異なるセットのアクセス制御許可によって定義されてもよい。"所有者"ユーザのセットは、ユーザの読み取り専用リスト及び書き込み許可リストを変更する権限によって定義されてもよい。

30

【0167】

コラボレーションエージェントのアルゴリズムにユーザに代わって動作する権限を持たせるために、ジャーナルにエージェントが使用するアクセス制御の許可を与えてもよい。これらのアクセス許可は、"ジャーナルのアクセス権限"と呼ばれてもよい。ジャーナルは、悪意のあるユーザ自身がアクセスすることが不可能であるデータへのアクセスをそれらの悪意のあるユーザが得るための攻撃ベクトルとなるべきではないため、そのような許可は、それらの許可がジャーナルを制御するユーザからのものであるということを要求される場合がある。ジャーナルは、ユーザが、ジャーナルにおいて動作している複数のコラボレーションエージェントに特定のアクセス権限を割り当てることを可能としてもよい。

40

【0168】

例えば、ユーザが、検索結果のリストの中で発見されるドキュメントからプロジェクトを開始するときに、ユーザは、アクセス許可のセットを利用して、その検索を実行してもよい。デフォルトで、そのアクセス権限として、ジャーナルにそれらの同じアクセス許可を付与してもよく、それによって、ジャーナルの中のデータに反応するコラボレーションエ

50

ージェントは、関連するコンテンツを検索し続けて、ユーザに代わってそのプロジェクトを拡張してもよい。ユーザは、他のユーザとの間でそのプロジェクトを共有することを望む場合があるので、KOSは、ユーザが、ジャーナルのアクセス権限として、アクセス許可のより制限されたセットを選択することを可能としてもよい。このようにして、プロジェクトに招待される他の人々は、プロジェクトの元のクリエイターがアクセスすることが可能である同じデータのうちのすべてにはアクセスすることが不可能である場合がある。

【0169】

例えば、ジャーナルの元のクリエイターは、ユーザの電子メールのすべてを読み及びインデックスを付すことをコラボレーションエージェントに要請している場合がある。ユーザが、ある特定の問題について他のジャーナルを作成することを決定するときに、そのユーザは、そのジャーナルが使用することが可能である許可を制限してもよく、それによって、そのプロジェクトに招待されている他の人は、そのジャーナルによってユーザの電子メールにアクセスすることが不可能である。

10

【0170】

他の態様において、ジャーナルは、複数のユーザによる非同期修正をサポートするように構成されてもよい。以下の例は、エンティティ及び関係のグラフを使用して、協調的なマシンによって支援される知識発見(collaborative, machine-assisted knowledge discovery)のためのユーザインターフェイスのある1つの例示的实施形態に関する詳細を説明する。ユーザがある概念の頂点をダブルクリックして、その頂点に関連するより多くの概念を要求するときに、以下の一連のイベントが生起する場合がある。第1に、ユーザインターフェイスは、例えば、知識グラフの表示をサポートするデスクトップアプリケーション又は他のローカルソフトウェアエージェントによって、マシン分析システム706から、ある頂点のための関連する概念を要求してもよい。マシン分析システム706(又は、KOSをサポートする他の適切なバックエンド)は、作業ユニットを生成してもよく(may spawn)、その作業ユニットは、関連する概念の発見及びグラフへのそれらの発見した概念の追加の双方のための要求を完了することを意図している。この作業ユニットは非同期に実行される。マシン分析システム706は、ユーザが複数の関連する概念の要求を開始した時点を示すタイムスタンプを含める方法で、その操作ログに要求をコミットしてもよい(収容してもよい)(may commit)。そのコミットは、また、作業ユニットの識別子を含んでもよい。マシン分析システム706は、上記のコミットを含む複数の操作のすべてによって要求に応答してもよく、それらの複数の操作は、ユーザインターフェイス702による最後のレンダリングのときからログにコミットされる(収容される)。クライアントは、新たな操作を使用して、ある特定の頂点にロードアイコン(loading icon)をレンダリングしてもよい。"ロード(loading)"状態は、グラフのログの一部であるため、ユーザがグラフを閉じ、そして、そのグラフを再度開いた場合であっても、このロードアイコンは、有効に存続してもよいということに留意すべきである。コミットされた操作(committed operation)は、タイムスタンプを含むので、ユーザインターフェイス702は、あらかじめ定められているタイムアウトの後にロードアイコンを省略してもよい。

20

30

【0171】

非同期的に、作業ユニットは、実行を開始してもよく、関連する概念の検索は、マシン分析システム706によって実行されてもよい。その次に、操作ログ(operation log)に対する書き込みロック(write lock)を取得してもよい。このロックが保持されている間は、いかなる他のプロセスも、直接的にログに書き込むことは不可能である。このことは、実際のユーザ又は潜在的なユーザのためのユーザインターフェイス又はグラフの利用可能性には影響を与えないということに留意すべきである。この書き込みロックは、データレースを防止するためにグラフ全体の重複演算が実行されるときはいつでも必要となってもよい。ロックが設定されている場合には、いずれかの関連する概念は、現在のグラフに対して重複する場合がある。ある1つの態様において、このプロセスは、名称類似性発見の問題解決(name similarity heuristics)等を使用することによって実装されてもよい

40

50

。操作ログに書き込むことによって、グラフの中に新たな概念の頂点及び証拠の端部を作成してもよい。デフォルトで、新たな端部のすべてを非表示にしてもよい。その次に、書き込みロックを解放してもよい。

【0172】

非同期的に、次に、ユーザインターフェイスが、(ポーリングによって又は1-4と同様の他の操作によって)バックエンドから更新を受信したとき、ユーザインターフェイスは、操作ログに追加されている複数の追加的な頂点及び端部を含む場合がある。それらの複数の追加的な頂点及び端部は、デフォルトで隠されているので、ユーザインターフェイスは、いくつかの隠された端部が存在しているかを示すインジケータを頂点に表示してもよい。ユーザがそれらを見せるように要求するときに、ユーザインターフェイス(又は、より正確には、ユーザインターフェイスを制御するプロセス)は、それらを示す操作を実行してもよい。

10

【0173】

書き込みロックの単純な実装は、操作ジャーナルにおいてテーブルレベルのROW EXCLUSIVE PostgreSQLロックを取得することである。しかしながら、このことは、ある1つの特定のグラフではなく、すべてのグラフへの書き込み要求を遮断する場合がある。他の態様において、この問題は、分離モードをSERIALIZABLEに設定することによって取り扱われてもよいが、このことは、例えば、エンドユーザが、迅速且つ連続的に2つの別個の頂点をダブルクリックする場合のように、結局のところ最後には重複処理を引き起こす場合がある。他の態様において、システムは、PostgreSQLトランザクションレベル助言ロックを使用して、ジャーナルに対する書き込みロックを実装し、そして、複数のユーザによる同時の使用をサポートしてもよい。

20

【0174】

知識操作ジャーナル

【0175】

知識操作ジャーナル(Knowledge Operations Journal(KOJ))は、知識操作レコード(knowledge operation records(KORs))の集合である。それらの複数の集合は、順序付けられて、追加のみのログとされてもよい。操作のこのログは、そのジャーナルのユーザが開始する検索イベント又は知識指向のアクションの記録を形成してもよい。

【0176】

30

操作変換ジャーナルは、本発明の技術分野においては標準的である。Etherpad、Google Wave、Google Docs、Office 365、及びDropbox Paper等の共同編集ツール(Collaborative editing tools)は、そのようなジャーナル構造を使用して、複数人が編集可能なドキュメントの中の編集イベントを追跡する。そのようなシステムが機能するのを可能とするために、コンピュータサイエンスのサブフィールドは、衝突のない複製データタイプ(conflict-free replicated data types(CRDTs))を研究するために統合されている。CRDT及び関連する操作ジャーナルの技術分野において標準的であるように、チェックポイント操作(checkpoint operation)を形成することによって、あるジャーナルの中のデータを簡略化することが可能であり、チェックポイント操作は、その時点までの複数の以前の操作の集約効果(aggregate effect)を要約する(summarizes)。このことは、ジャーナルのクライアントが、操作の履歴全体を再処理することを必要としないということをも可能としてもよく、むしろ、クライアントは、直近のチェックポイントから開始し、以降の操作を行うことによって、そのジャーナルの状態の自身のビューを更新することが可能である。

40

【0177】

KORは、役立つように、そのアクションをとったユーザの識別子、コーパスのうちのあるドキュメントの中での言及の識別子、及び、適切である又は有用である場合には、識別されている言及を指す視覚的知識要素に対するアクション等のアクションの各々の記録のためのフィールドを含んでもよい。ユーザ識別子等のKORの複数の態様を使用して、複数の特定のユーザによる複数の特定の投稿を示す方式によって、あるプロジェクトの知識

50

グラフ又は他の視覚的な表現をレンダリングしてもよい。このようにして、例えば、複数の人間ユーザがあるプロジェクトに追加する場合に、各々のユーザが追加する頂点又は端部は、色によって符号化されてもよく、ユーザのアイコン又はアバターによってフラグ付けされてもよく、或いは、ある特定の追加的な情報源を識別するように視覚的に符号化されてもよい。マシンユーザによる複数の活動は、また、或いは、むしろ、この方式によって視覚的に符号化されてもよい。同様に、修正許可は、例えば、ユーザが頂点への変更を要求する頂点を追加したユーザに依存してもよい。このようにして、例えば、ユーザは、そのユーザのエントリーに対する変更を禁止し、許可し、又は条件付きで許可してもよい。

【0178】

10

一般的に、本明細書において考慮されているユーザは、人間ユーザ及びマシンユーザの双方を含んでもよい。例えば、マシンユーザは、"John Smith"に言及する要約カードを提示するアクションを実行するマシンベースのアルゴリズムを含んでもよく、人間ユーザは、その要約カードをクリックするアクションを実行して、例えば、その要約カードについてより多くの情報を要求してもよい。

【0179】

KORの複数の集合を集めることによって、人間ベースの及びマシンベースの双方のユーザは、多くの表面にわたってすべてのユーザの過去のアクションをレビューしてもよい。さらに、アクション及び複数のKORにわたるアクションは、ユーザの意図のアルゴリズム的推測を容易にすることが可能であり、それによって、マシンユーザは、人間の探索の関心を高め、及び/又は、そのジャーナルにおいてより多くのKORを生成する複数のさらなるアクションを利用してもよい。

20

【0180】

ある1つの例示的な実装において、"OpEdgeSubstantiateStart"及び"OpEdgeSubstantiateEnd"と呼ばれる操作は、関係の証拠を束ねること(relation evidence bundles)を可能とする。ある関係に関心を表明するユーザアクションは、OpEdgeSubstantiateStart操作として記録される。この操作は、ユーザが証拠を見ることを望む関係の中の2つの概念のための識別子を搬送してもよい。コラボレーションエージェントは、証拠を発見するためのアルゴリズムを実行することによって、この操作に反応する。コラボレーションエージェントが証拠を発見するときに、そのコラボレーションエージェントは、記憶システムに証拠を束ねたもの(evidence bundle)を格納する。その次に、そのコラボレーションエージェントは、格納位置へのポインタを使用して、ジャーナルにOpEdgeSubstantiateEnd操作を追加する。ジャーナルを処理するクライアントプログラムは、その次に、ポインタを使用することによって証拠を表示して、格納されている証拠にアクセスしてもよい。これらの2つの操作は、ジャーナルの中にグループを形成してもよい。取り消し操作は、操作又は操作のグループへのポインタを搬送してもよく、それによって、操作のグループ化は、取り消しを容易にする。引用操作は、ユーザが、ドキュメントからジャーナルへの文章を添付してもよい。このことは、ユーザが、多くの言及を含むデータのより長い部分に対してアクションを実行することを可能とする。

30

【0181】

40

KORのジャーナルを処理することによって、プロセスは、マシンユーザ及び人間ユーザによるアクションの履歴を表す集合体(an aggregate representative of a history of actions)を作成してもよい。集合体を構築するために、システムは、ジャーナルの中の操作のシーケンスを処理してもよい。チェックポイント操作は、その次に、過去の操作の何らかの態様の要約として、ジャーナルにその集約の結果を格納してもよい。単一のジャーナルから形成することが可能である多くの可能な集合体が存在する。例えば、ある1つのユーザの複数のアクション(actions)のすべてを収集して、その1つのユーザの活動(activity)及び/又は意図(activity)の記述(description)としてもよい。代替的に、複数のユーザのための複数の操作を使用して、例えば、ユーザのグループの思考プロセス又は検索目的を示すグループの意図をつかんでもよい。

50

【0182】

1つの例示的な実装において、1つ又は複数のユーザの意図をモデル化するために操作を集約するためのメカニズムは、以下ようになる。システムは、各々の言及を取り巻くコンテキストを記述する特徴ベクトルを生成してもよい。ベクトルは、高次元ベクトル空間の一部であってもよく、コーパスの中で言及されている各々の単語、句、又は概念は、ベクトル空間の次元として扱われる。ユーザのある与えられたセットの集合体を形成するために、システムは、要約ベクトルを作成してもよい。最初に、要約ベクトルは、ゼロベクトルである。システムは、その次に、ある与えられたセットの中のユーザによる操作のために、ジャーナルフィルタリングの操作を繰り返す。フィルタを通過するKORが、コーパスの中の言及のための識別子を搬送している場合に、システムは、その言及の特徴ベクトルを使用して、要約ベクトルを洗練してもよい。その操作が、肯定的なユーザアクションを搬送している場合に、言及の特徴ベクトルは、要約ベクトルに追加されるであろう。操作が否定的なユーザアクションを搬送している場合に、言及の特徴ベクトルは、要約ベクトルから減算されるであろう。言及が、中間的なユーザアクションを搬送しているか、あいまいなユーザアクションを搬送しているか、又はいかなるユーザアクションも搬送していない場合に、システムは、要約ベクトルに対する変更の大きさ及び符号を変化させる重み付け係数を使用して、要約ベクトルへの言及の特徴ベクトルの追加を行ってもよい。

10

【0183】

要約ベクトルは、ユーザの現在の知識状態をモデル化するデータ構造を提供することが可能である。要約ベクトルの中にゼロ以外の成分を有するベクトル空間の次元は、ユーザの知識の現在の状態を記述することが可能である。正の値を有する次元は、ユーザの望ましい知識を記述することが可能である。システムは、ジャーナルから集約されているこの要約ベクトルを使用して、言及の推薦及び関連する概念の推薦に順位をつけてもよい。システムは、要約特徴ベクトルの中に正の値を有するさまざまなストリングを使用して、上記で説明されている層0及び層1及び層2インデックスを照会することによって、候補の言及及び候補の関連する概念を生成してもよい。候補の言及の特徴ベクトルが、要約ベクトルと強い重複を有している場合に、システムは、その言及のコンテキストが、"本筋を外れていない(on topic)"、すなわち、そのユーザが研究していることに関して高い可能性が高いと断言してもよい。同様に、関係の推薦の場合に、同時に生起する概念の言及のための特徴ベクトルの組み合わせが、要約ベクトルと強い重複を有している場合に、システムは、本筋を外れていない可能性が高いと断言してもよい。

20

30

【0184】

同時生起(Co-occurrence)は、候補の関係(candidate relations)を生成する簡単な方法である。それらの複数の概念の言及が、同じドキュメントの中に出現する場合に、2つの概念は、"同時に生起"する(co-occur)と言われる。そのような同時生起が、2つの概念の間の有意義な関係を意味する可能性は、それらの複数の言及の近接度が増加するにしたがって高くなる。すなわち、テキスト又は画像、又は、他のメディアの中でともに近接している言及は、実際に関連している可能性がより高い。

【0185】

そのシステムは、単純な内積、マハラノビス距離(Mahalanobis distance)、又は他の対カーネル関数(pairwise kernel function)、又はベクトルの複数の部分に適用されるカーネルから構成されるポテンシャル関数の重みづけ総和を使用して、それらの複数のベクトルの重複を測定することが可能である。ある1つの態様において、特徴ベクトルは、いくつかのより小さなベクトル空間からの複数のベクトルの積であってもよい。たとえば、ある1つの部分空間は、言及のコンテキストの中に出現する人の名前であり、他の部分空間は、言及のコンテキストの中に出現する組織の名前である。

40

【0186】

ユーザのプロジェクトについて、言及又は関連する概念が本筋を外れていない(on topic)という可能性に加えて、要約ベクトルを使用して、その言及がユーザにとって有用であるという可能性を推定してもよい。ユーザが、そのユーザがまだ知らない新たな材料を

50

探している場合に、システムは、言及の特徴ベクトル(又は、複数の言及のベクトルの関連する概念の組み合わせ)の次元の数を計数し、そして、その次に、要約ベクトルの中に存在しない次元の数を計数してもよい。新しさのスコアは、候補のベクトルの中に存在する概念の数に対する集約要約ベクトル(aggregate summary vector)の中に存在しない概念の数の比として計算されてもよい。多くの候補の推薦のリストを考慮するとき、システムは、また、それらの候補の多様化を追求する場合がある。このことは、例えば、KOJからの集約要約ベクトルがユーザの意図を表すことができるこの特殊化されたコンテキストに対して、例えば、最大限の限界関連性(maximum marginal relevance)等のさまざまな多様化モデル(diversification models)のうちのいずれかを使用して達成されてもよい。

10

【0187】

KOJにとっての利点の1つは、複数の異なる視覚的表示要素がユーザへの知識の提示を同期させることが可能であるということである。例えば、ユーザは、KOSを使用して作業してもよく、そのKOSは、そのユーザが開いている各々のドキュメントの中に要約カードのキューを表示する。カードの1つのキューからカードを破棄するユーザアクションは、その同じプロジェクトを示す他のキューからそのカードを削除するようにさせてもよい。同様に、ユーザは、提案されているドキュメントにおける関心を確認するアクションをとってもよく、自動的なシステムは、その提案されているドキュメントが、関心のあるエンティティに言及しているということを信じている。そのアクションは、例えば、ドキュメントの中のコンテキスト内のハイライト等の何らかの視覚的知識要素を通じて受信されてもよい。KOJによって、これに応じて他の視覚ディスプレイを更新してもよい。例えば、ユーザがコンテキスト内の言及のハイライトに対するアクションボタンをクリックすると、そのプロジェクトのためのグラフ内の頂点の集約は、その概念についてのある頂点を他の頂点に接続する適切な線を使用してその概念についてのその頂点を示すことが可能である。

20

【0188】

KOJ構造は、また、有利には、マシン学習アルゴリズムの使用を促進することが可能である。肯定的なユーザアクション及び否定的なユーザアクションを記録することによって、KOJは、分類器及び他の学習可能なアルゴリズムの学習のための真の肯定及び偽の肯定の情報源を提供することが可能である。KOJは、ユーザ及びユーザの焦点を動的に追跡するので、KOJは、広範囲の機械学習モデル等のための学習ラベルの高品質の情報源を提供することが可能である。

30

【0189】

引用符(Quote notes)は、特別なタイプの操作であってもよく、その特別なタイプの操作においては、ユーザは、テキストの断片を取り込み、観察されたデータのスナップショットとしてジャーナルの中にそのテキストの断片を格納する。引用符は、その引用符から概念への言及の端部として現れる場合があるグラフの中の他の概念の頂点に言及してもよい。また、ユーザが言及連鎖(mention chain)を引用することを望むときに、引用符を作成することが可能である。ユーザが言及連鎖を引用するときに、その引用とともに新たな引用符を作成することが可能であり、新たな引用符は、対応する言及連鎖を含む概念の頂点への引用の端部を有してもよい。

40

【0190】

KOSによる解決に役立つ場合がある問題のうちの1つは、ドキュメントの重複又は同じ概念への参照である。例えば、ユーザが、ほぼ同時に、2つの異なる概念のために複数の関連する概念を要求することが可能であるということを検討する。この要求は、バックエンドによって非同期的に処理されてもよい。複数の新たな関連する概念が発見されるときに、それらは、個別の操作として操作ジャーナルに追加されてもよい。これらの概念のいくつかが同じである場合、すなわち、これらの概念のいくつかは、同一指示アルゴリズムが、同じ概念を指している可能性が高いということを検出することが可能である言及である場合には、以下で説明するように、複数の異なる解決方法を使用することが可能である。

50

【0191】

上記の問題に対して、複数の異なるハイレベルのアプローチが存在する。1つの方法は、識別情報に基づいていてもよく、その識別情報は、頂点のための識別子を採取するプロセスが、識別情報を頂点に提供するということの意味し、したがって、重複排除は、ユーザのグラフのコンテキストにおいて自動的に生起する。このアプローチに伴う問題は、すべての重複排除論理が概念のための識別情報を定義するように強いられるということであり、このことは、概念を定義する正規辞書を導入するということの意味することが多く、すべてのドキュメントが同じ辞書に固執することを要求するが、実際のデータを実装し又は実際のデータに強制するのは実用的ではない。

【0192】

他のアプローチは、重複排除問題(dedupe problem)に正面から立ち向かうことであってもよい。すなわち、頂点が知識グラフに自動的に追加されるときはいつでも、まずジャーナルの他の概念と比較して、その概念がそのプロジェクトの中のいずれかの場所すでに言及されている概念を指しているか否かを決定してもよい。その場合に、システムは、その概念を完全に省略する(drop it completely)か、又は、その概念とその頂点と統合する(merge it with that vertex)かを決定してもよい。省略し又は統合するアクションは、"重複排除(deduped)"という語で要約されてもよい。

【0193】

図8は、ある1つのコンピュータシステムを図示している。一般的に、コンピュータシステム800は、例えば外部デバイス804を通じてネットワーク802に接続されるコンピューティングデバイス810を含んでもよい。コンピューティングデバイス810は、上記の図1を参照して上記で説明されているいずれかのコンピューティングデバイス等の本明細書において説明されているいずれかのタイプのコンピューティングデバイスであってもよく、又は、それを含んでもよい。例えば、コンピューティングデバイス810は、デスクトップコンピュータワークステーションを含んでもよい。コンピューティングデバイス810は、また、或いは、むしろ、これらには限定されないが、ラップトップコンピュータ、デスクトップコンピュータ、パーソナルデジタルアシスタント、タブレット、及び携帯電話等を含むいずれかの他のデバイスであってもよく、そのいずれかの他のデバイスは、プロセッサを有するとともに、ネットワーク802を介して通信してもよい。コンピューティングデバイス810は、また、或いは、むしろ、サーバ、検索エンジン、及びクラウドコンピューティングリソース等を含んでもよく、又は、コンピューティングデバイス810は、サーバに、或いは、仮想的なサーバファーム又は物理サーバファームの中、或いは、他の物理的な又は仮想化された環境に又は形態で配置されてもよい。

【0194】

コンピューティングデバイス810は、図1を参照して上記で説明されているコンピューティングデバイスのいずれかであってもよい。例えば、コンピューティングデバイス810は、サーバ、クライアント、データベース、検索エンジン、或いは、本明細書において説明されている他の設備又はコンピューティングデバイスのいずれかであってもよい。複数の態様において、コンピューティングデバイス810は、(例えば、デスクトップコンピュータ等の)ハードウェア、(例えば、仮想マシン等の)ソフトウェア、又は、(例えば、デスクトップコンピュータにおいて実行されるプログラム等の)ソフトウェア及びハードウェアの組み合わせを使用して実装されてもよい。コンピューティングデバイス810は、スタンドアロンデバイス、他のエンティティ又はデバイスに統合されているデバイス、複数のエンティティにわたって分散されているプラットフォーム、仮想化環境の中で実行される仮想化デバイス、又はこれらの組み合わせであってもよい。

【0195】

ネットワーク802は、1つ又は複数のデータネットワーク又は相互接続ネットワーク等のいずれかのネットワーク又はネットワークの組み合わせを含んでもよく、それらのいずれかのネットワーク又はネットワークの組み合わせは、通信している関係にある複数の他のエンティティを結合し、そして、コンピュータシステム800の中の複数の参加者の間

10

20

30

40

50

でデータ及び制御情報を通信するのに適している。ネットワーク802は、インターネット等の公衆ネットワーク、プライベートネットワーク、及び公衆交換電話ネットワーク等の通信ネットワーク、又は、(例えば、3G又はIMT-2000等の)第3世代セルラー技術、(例えば、4G、LTE、MT-Advanced、E-UTRA等の)第4世代セルラー技術又はWiMAX-Advanced(IEEE 802.16m)及び/又はその他の技術を使用するセルラーネットワークのみならず、コンピュータシステム800の中の複数の参加者の間でデータを搬送するのに使用されてもよいいずれかのスイッチ、ルータ、ハブ、及びゲートウェイ等とともに、さまざまな企業エリア、大都市エリア、キャンパス、又はその他のローカルエリアネットワーク又は企業ネットワークのいずれかを含んでもよい。ネットワーク802は、また、データネットワークの組合せを含んでもよく、厳密には、パブリックネットワーク又はプライベートネットワークに限定される必要はない。 10

【0196】

外部デバイス804は、ネットワーク802を通じてコンピューティングデバイス810に接続するいずれかのコンピュータ又は他のリモートリソースであってもよい。これは、サーバ、検索エンジン、クライアント、データベース、ネットワーク記憶デバイス、コンテンツをホストするデバイス、又はネットワーク802を通じてコンピューティングデバイス810に接続されてもよいいずれかの他のリソース又はデバイスを含んでもよい。

【0197】

演算デバイス810は、プロセッサ812、メモリ814、ネットワークインターフェイス816、データストア818、及び1つ又は複数の入力/出力デバイス820を含んでもよい。コンピューティングデバイス810は、さらに、1つ又は複数の周辺機器822及び他の外部入力/出力デバイス224を含んでもよく、又は、それらと通信してもよい。 20

【0198】

プロセッサ812は、本明細書において説明されているいずれかののものであってもよく、一般的に、コンピューティングデバイス810又はコンピュータシステム800の中で実行するための命令を処理することが可能であってもよい。プロセッサ812は、本明細書において考慮されているように、データ及び命令を処理するのに適している単スレッドプロセッサ、マルチスレッドプロセッサ、マルチコアプロセッサ、又はいずれかの他のプロセッサ、処理回路、又はそれらの組み合わせを含んでもよい。プロセッサ812は、メモリ814の中に格納され又はデータストア818に格納されている命令を処理することが可能であってもよい。プロセッサ812は、また、或いは、むしろ、物理プロセッサをエミュレートする仮想化マシンを含んでもよい。 30

【0199】

メモリ814は、コンピューティングデバイス810又はコンピュータシステム800の中の情報を格納してもよい。メモリ814は、いずれかの揮発性メモリ又は不揮発性メモリ、或いは、これらには限定されないが、ランダムアクセスメモリ(RAM)、フラッシュメモリ、読み出し専用メモリ(ROM)、プログラム可能な読み出し専用メモリ(PROM)、消去可能なPROM(EPROM)、及びレジスタ等を含む他のコンピュータ読み取り可能な媒体を含んでもよい。メモリ814は、プログラム命令、プログラムデータ、実行可能プログラム、及びコンピューティングデバイス810の動作を制御するとともに、ユーザのための機能を実行するようにコンピューティングデバイス810を構成するのに有用な他のソフトウェア及びデータを格納してもよい。メモリ814は、コンピューティングデバイス810の動作の複数の異なる態様のための複数の異なるステージ及びタイプを含んでもよい。例えば、プロセッサは、特定のデータ又は命令へのより高速なアクセスのためのオンボードメモリ及び/又はキャッシュを含んでもよく、また、要求に応じて、メモリ容量を拡張するために、個別のメインメモリ等を含んでもよい。 40

【0200】

メモリ814は、一般的に、コンピュータコードを含む不揮発性コンピュータ読み取り可能媒体を含んでもよく、そのコンピュータコードは、コンピューティングデバイス810によって実行されると、例えば、プロセッサファームウェア、プロトコルスタック、デー 50

データベース管理システム、オペレーティングシステム、又はそれらの組み合わせを構成するコード、及び/又は本明細書に記載されているさまざまなフローチャート及び他のアルゴリズムの説明に記載されている複数のステップの一部又はすべてを実行するコード等の対象となるコンピュータプログラムの実行環境を生成する。単一のメモリ814が示されているが、コンピューティングデバイス810の中に、役立つように、いずれかの数のメモリを組み込んでもよいということが理解されるであろう。例えば、たとえ、コンピューティングデバイス810の電源が切られているときであっても、第1のメモリは、ファイル及びコードの永続的な又は長期的な格納のためにディスクドライブ等の不揮発性記憶デバイスを提供してもよい。ランダムアクセスメモリ等の第2のメモリは、処理を実行するための命令及びデータを格納するための揮発性の(一方で、より高速の)メモリを提供してもよい。他の態様において、メモリ814は、物理メモリリソースをエミュレートする仮想メモリを含んでもよい。

【0201】

ネットワークインターフェイス816は、ネットワーク802を通じて他のリソースと通信する関係にあるコンピューティングデバイス810を接続するためのいずれかのハードウェア及び/又はソフトウェアを含んでもよい。このことは、例えば、(例えば、イーサネット等の)物理的接続、(例えば、WiFi等の)無線周波数通信、(例えば、光ファイバ又は赤外線等の)光通信、超音波通信、又はコンピューティングデバイス810と他のデバイスとの間でデータを搬送するのに使用されてもよいこれらの媒体又は他の媒体のいずれかの組み合わせを使用する短距離通信プロトコルを使用することによって利用可能となるローカルリソースのみならず、インターネットを通じてアクセス可能なリモートリソース等のリソースへの接続を含んでもよい。ネットワークインターフェイス816は、例えば、ルータ、モデム、ネットワークカード、赤外線トランシーバ、無線周波数(RF)トランシーバ、近距離通信インターフェイス、無線周波数識別(RFID)タグリーダ、又はいずれかの他のデータ読み取りリソース又は書き込みリソース等を含んでもよい。

【0202】

データストア818は、ディスクドライブ、光学ドライブ、磁気ドライブ、フラッシュドライブ、又はコンピューティングデバイス810のための大容量記憶装置を提供することが可能である他のデバイス等のコンピュータ読み取り可能な媒体を提供するいずれかの内部メモリストアであってもよい。データストア818は、以降の検索及び使用のために不揮発性の形態で、コンピューティングデバイス810又はコンピュータシステム800のためのコンピュータ読み取り可能な命令、データ構造、プログラムモジュール、及び他のデータを格納してもよい。データストア818は、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、及び他のプログラムモジュール、ソフトウェアオブジェクト、ライブラリ、実行可能なプログラム等のためのコンピュータ実行可能なコードを格納してもよい。データストア818は、また、プログラムデータ、データベース、ファイル、メディア等を格納してもよい。

【0203】

入力/出力インターフェイス820は、コンピューティングデバイス810に結合されてもよい他のデバイスへの入力及び他のデバイスからの出力をサポートしてもよい。周辺機器822は、いずれかのデバイス又はデバイスの組み合わせを含んでもよく、そのいずれかのデバイス又はデバイスの組み合わせは、コンピューティングデバイス810に情報を提供し又はコンピューティングデバイス810から情報を受信するのに使用される。これは、キーボード、マウス、マウスパッド、トラックボール、ジョイスティック、マイクロホン、フットペダル、カメラ、タッチスクリーン、スキャナ、又はユーザ830がコンピューティングデバイス810に入力を提供するのに使用されてもよい他のデバイス等の人間の入力/出力(I/O)デバイスを含んでもよい。これは、また、或いは、むしろ、ディスプレイ、スピーカ、プリンタ、プロジェクタ、ヘッドセット、又は、ユーザに情報を提示するための、又は、そうでない場合には、コンピューティングデバイス810からマシン使用可能な出力又は人間が使用可能な出力を提供するための他のいずれかの視聴覚デバイスを

含んでもよい。他のハードウェア826は、コプロセッサ、ディジタル信号処理システム、数学コプロセッサ、グラフィックスエンジン、及びビデオドライバ等のコンピューティングデバイス810の中に組み込まれていてもよい。他のハードウェア826は、また、或いは、むしろ、拡張入力/出力ポート、追加メモリ、(例えば、DVDドライブ又は他のアクセサリ等の)追加ドライブ等を含んでもよい。バス832又はバスの組み合わせは、プロセッサ812、メモリ814、ネットワークインターフェイス816、他のハードウェア826、データストア818、及び入力/出力インターフェイス等のコンピューティングデバイス810の構成要素を相互接続するための電気機械プラットフォームとして機能してもよい。図に示されているように、コンピューティングデバイス810の複数の構成要素の各々は、情報を通信するためのシステムバス832又は他の通信機構を使用して相互接続されてもよい。

10

【0204】

コンピュータシステム800のプロセッサ812を使用して、本明細書において説明されている方法及びシステムを実現して、メモリ814に含まれている命令の1つ又は複数のシーケンスを実行し、そして、あらかじめ定められているタスクを実行してもよい。複数の実施形態において、コンピューティングデバイス810は、性能を向上させるために、ともにコードを実行するように同期している複数の並列プロセッサとして配置されてもよく、或いは、コンピューティングデバイス810は、仮想化環境において実装されてもよく、その仮想化環境においては、ハイパーバイザ又は他の仮想化管理機能におけるソフトウェアは、コンピューティングデバイス810のハードウェアのインスタンス化の機能の一部又はすべてを再現するのに適するように、コンピューティングデバイス810の構成要素をエミュレートする。

20

【0205】

図9は、複数のエンティティ間の関係(relationships)を調査する方法のフローチャートである。一般的に、検索エンジンは、複数のエンティティの言及、より具体的には、エンティティに関連するグループの言及、及びドキュメントのコーパス(corpus of documents)の中のそのようなエンティティに関連するグループの言及の間の関係を識別する情報を検索してもよい。関連するユーザインターフェイスは、ユーザが複数のエンティティ(又は、エンティティに関連するグループの言及)及び関心のある関係の選択及び提示を案内するために、ユーザに対話的に情報を提示してもよい。

30

【0206】

ステップ902に示されているように、方法900は、ドキュメントの集合へのアクセスを提供することから開始してもよい。このことは、検索エンジン及びユーザのディスプレイを関心のあるドキュメントに接続するためのさまざまな技術のいずれかを含んでもよい。例えば、このことは、ローカルディレクトリ、クラウド記憶設備、又はワールドワイドウェブ等の公衆ネットワーク、又はこれらの組み合わせ等の検索エンジンへの接続を提供するステップを含んでもよく、その検索エンジンは、利用可能なドキュメントを検索し、分析し、及び索引付けしている。このことは、また、例えば、ユーザインターフェイスをホストするデバイスとドキュメント情報源との間の接続をサポートすることによって、直接的なアクセスを提供するステップを含んでもよく、その直接的なアクセスは、ユーザの照会を受信するときに、ドキュメントを動的に検索し及び処理するのに使用されてもよい。より一般的には、ユーザの検索、プログラマチックな検索及び検索、(例えば、複数のエンティティ間の関係を具体化する、又は、その他の点では、エンティティに関連するグループの言及をサポートする、エンティティの言及を含む証拠(本明細書においては、また、"証拠の節"と称される)等の)分析、ユーザのディスプレイ及びレビュー、及び、基礎となるコンテンツとのユーザの対話をサポートするためのいずれかの技術を、ステップ902に示されているように、ドキュメントへのアクセスを提供するのに使用してもよい。広範な索引付け技術及びネットワーク検索エンジン等が利用可能であるということが理解されるであろう。同様に、データホスト及びデータリポジトリシステムは、Kerberos, Active Directory, OAuth, 公開鍵インフラストラクチャ(public key infra

40

50

structure(PKI))、及びその他のデータ転送、サインイン技術、及び/又はクラウドベースのストレージ技術を含む広範なアクセス制御技術及びアクセス統合技術を含む。上記のいずれかは、単独で、又は推薦エンジン又はマシン分析システム等と組み合わせて使用して、ドキュメントへのアクセスを提供するとともに、本明細書に説明されているようなドキュメントの中のエンティティに関連するグループの言及のユーザの調査をサポートしてもよい。

【0207】

ステップ904に示されているように、方法900は、ユーザ入力を受信するステップを含んでもよい。このユーザ入力は、本明細書において説明されているエンティティ調査プラットフォームによって、さらなる処理を開始するいずれかのユーザ入力であってもよい。ステップ906に示されるように、方法900は、ドキュメントを処理するステップを含んでもよい。これに関連して、ドキュメントを処理するステップは、エンティティデータベース又はインデックスコンテンツ等を作成するための検索及び分析等の一般的な背景処理を含んでもよく、又は、ドキュメントを処理するステップは、ある特定のユーザ要求又は他のユーザ入力に応答して特定の処理を含んでもよい。一般的に、ユーザ入力を受信するステップは、特定のユーザインタラクションに応じて、ドキュメントを処理するステップの前、処理の最中、又は処理の後、又は、これらのいずれかの組み合わせの時点で生起してもよい。

【0208】

例えば、ある1つの態様において、ユーザ入力を受信するステップは、検索入力フォームフィールドにテキスト入力を提供して、新たなサーチ又はプロジェクトを開始するステップを含んでもよい。ドキュメントの処理は、その次に、ユーザに提示するドキュメントの利用可能なコーパスの中で複数の候補のエンティティの言及を検索するステップを含んでもよい。ある1つの態様において、ユーザは、検索入力フォームフィールドの中の2つ又はそれ以上のエンティティを直接的に識別するしてもよく、方法900は、ユーザインターフェイスの中に、例えば、タグ等として、複数の候補の言及を表示するステップを含んでもよい。他の態様において、ユーザインターフェイスの中に、さまざまなエンティティの言及を含むドキュメント又はそのようなドキュメントからの抜粋を表示してもよく、そのドキュメントの中の言及を識別する1つ又は複数のタグとともに、各々のドキュメントを表示してもよい。次に、ユーザは、その次に、特に、関心のある言及を選択してもよい。

【0209】

本明細書中において使用されている"エンティティ"という語は、物理的な世界におけるある特定の既知のエンティティを意味するものではないということが理解されるであろう。本明細書において考慮されているデータ処理が実世界におけるエンティティについての結論に達する点は、通常は、存在しない。むしろ、本明細書において説明されているエンティティは、複数のドキュメントの中に位置する複数の言及の間の関係の推測であり、同一指示の言及のグループ分けは、エンティティと称される単一のものを指すと推定され又は予測されてもよい。このようにして、本明細書中で考慮されている"エンティティ"は、実世界のエンティティの実際の調査から得た真実の識別情報(ground truth identification)ではなく、言及のそのようなエンティティベースのグループ分けを特徴付けるか又は言及のそのようなエンティティベースのグループ分けを存続させるデータ構造、或いは、言及が出現するドキュメントの周囲のコンテキストに起因して、同一指示されるように見える言及のグループである。このようにして、文脈から特に又は別途記述されない限り、添付の説明及び複数の請求項の中で使用されている"エンティティ"という語は、単一のエンティティを指しているか或いはエンティティ関連のグループ分けを特徴付けるか又は反映するデータ構造等を指しているように見えるドキュメントの中の言及のエンティティ関連のグループ分けの簡略化として役立つものと理解されるべきである。一般的に、このエンティティ関連のグループ分けにおける対応する言及は、同一指示又はそのような同一指示によって示唆される単一エンティティの現

10

20

30

40

50

実世界の意味についての結論を意味することなく、同一指示であるように見える言及である。

【0210】

他の態様において、方法900のステップ904及び906は、第1のエンティティ及び第2のエンティティを識別するステップを含んでもよく、各々は、第1のエンティティ及び第2のエンティティを参照するドキュメントの集合の中のテキストのスパンを選択すること等によって、ドキュメントの集合の中で言及される。このことは、また、或いは、むしろ、複数のエンティティのうちから指定し又は選択するために、複数の異なるユーザインタラクションを含んでもよい。例えば、ユーザは、プラットフォームが提供するエンティティのリストから2つのエンティティを明示的に選択してもよく、又は、ユーザは、2つのエンティティの記述を入力してもよく、プラットフォームは、ユーザが記述する複数のエンティティに対応する可能性が高い複数のドキュメントの中の言及を示唆してもよい。他の態様において、ユーザは、1つのエンティティを識別してもよく、プラットフォームは、例えば、ドキュメントの中の同一指示に基づいて関心のある可能性のある他の関連エンティティを検索してもよい。ユーザは、その次に、プラットフォームが識別するとともに、ユーザインターフェイスの中でユーザに提示されている複数のエンティティタグ又は複数の言及の中から第2のエンティティを選択してもよい。

10

【0211】

エンティティが識別され、ユーザ入力がドキュメントの中の関係の証拠を裏付けるための要求を含むある1つの態様において、ドキュメントを処理するステップは、ドキュメントの集合を処理して、第1のエンティティと第2のエンティティとの間の接続を具体化する(例えば、複数の証拠の節等の)証拠を識別するステップを含み、その証拠は、ドキュメントの集合のうちの少なくとも1つのドキュメントの中での第1のエンティティ及び第2のエンティティの同時生起する言及を含む。このことは、また、或いは、むしろ、ドキュメントの集合を処理して、第1のエンティティの第1の言及及び第2のエンティティの第2の言及を含む証拠の節等の第1のエンティティと第2のエンティティとの間の接続を具体化するドキュメントの集合のうちの1つ又は複数のドキュメントの中の証拠の節を識別するステップを含んでもよい。

20

【0212】

その他のインタラクションは、また、或いは、むしろ、ドキュメントの集合の中で利用可能な証拠に基づいて、複数のエンティティの間の関係を特定し及び選択するのに使用されてもよい。例えば、その方法は、複数のエンティティの間の同一指示に基づいて、ユーザに複数の候補接続を提示するステップと、ユーザによる複数の接続のうちの1つの選択に応答して、(例えば、以下に説明されるように)ユーザインターフェイスの中に表示するための接続を作成するステップと、を含んでもよい。他の態様において、方法900は、ユーザによる第1のエンティティの選択を受信するステップと、第2のエンティティへの接続を自動的に示唆するステップと、を含んでもよく、したがって、ドキュメントの集合の中の情報が示唆する複数の可能性のある関係の中から識別し及び選択する際に、ユーザに支援を提供する。他の態様において、その方法900は、ユーザによる第1のエンティティ及び第2のエンティティの選択を受信するステップと、ドキュメントの集合の中の関連を具体化するための証拠の節を検索するステップと、を含んでもよい。

30

40

【0213】

ステップ908に示されているように、方法900は、第1のエンティティと第2のエンティティとの間の接続を具体化するドキュメントの集合の中からの証拠を特徴付ける関係記述を格納するステップを含んでもよく、又は、そうでない場合に、ドキュメントの集合の中の1つ又は複数のドキュメントの中の情報に基づいて、2つのエンティティの間の可能な関係のためのサポートを提供してもよい。例えば、このことは、ドキュメントの全文又は双方のエンティティに言及する1つ又は複数の抜粋とともに、2つのエンティティの同一指示を含むドキュメントのタイトル及び/又は位置を含んでもよい。このことは、また、或いは、むしろ、言及が関心のあるエンティティを指す可能性、2つのエン

50

ティティの言及の間の(例えば、単語数等の)距離、(例えば、ドキュメントの意味分析(semantic analysis)等に基づく)エンティティの間の関係の明示的な陳述の識別、同一指示を含む他のドキュメントの数、或いは、集合の中の1つ又は複数のドキュメントに基づく複数のエンティティの間の関係を具体化し又は特徴づけるのに有用な他の情報等の関心のある関連情報を含んでもよい。

【0214】

この情報は、ユーザインターフェイスをサポートするプラットフォームがアクセス可能であるジャーナル、データベース、又は他のデータリポジトリの中に関係記述909として格納されてもよい。(例えば、エンティティ関連のグループの言及等の)基礎となるエンティティ及び1つ又は複数の関係を識別してもよく、そして、関係記述909は、ユーザの要求に回答して、自動的に、手動で、又はこれらの何らかの組み合わせによって作成されてもよい。例えば、上記のように、ユーザ入力を受信するステップは、複数のエンティティの言及の同時生起に基づいて、ユーザに候補として複数の接続を提示するステップと、ユーザによる複数の接続のうちの1つの選択に回答して、関係記述909を作成するステップと、を含んでもよい。他の態様において、ユーザ入力を受信するステップは、ユーザによる第1のエンティティの選択を受信し、そして、第2のエンティティへの接続を自動的に示唆するステップを含んでもよい。他の態様において、ユーザ入力を受信するステップは、ユーザによる第1のエンティティ及び第2のエンティティの選択を受信するステップと、ドキュメントの集合の中の接続を具体化するための(例えば、複数の証拠の節等の)証拠を検索するステップと、を含んでもよい。ユーザ入力を受信するステップは、また、或いは、むしろ、ユーザによる2つ又はそれ以上のエンティティの選択を受信するステップと、ドキュメントの集合の中の接続の具体化を検索するステップと、を含んでもよい。より一般的には、関係記述909は、本明細書において説明されているいくつかの技術を使用して自動的に及び/又は対話的に作成されてもよい。

【0215】

ステップ910に示されるように、方法900は第1のエンティティと第2のエンティティとの間の関係を視覚的に表示することを含み得る。一般に、一度2つのエンティティ間の関係が特定又は選択されると、多数の技術を用いてグラフィカルユーザインターフェイス内でさらなるユーザインタラクション(user interactions)がサポートされ得る。例えば、これは、ユーザインターフェイス内で第1のエンティティのための第1のアイコンと、第2のエンティティのための第2のアイコンを表示することと、ユーザインターフェイス内で第1のアイコンと第2のアイコンとを結合するコネクタシンボル等を表示することとを含み得る。コネクタシンボルは、第1のアイコンを第2のアイコンに視覚的に連結し、第1のエンティティと第2のエンティティとの間のつながりを表す線、矢印又は他の視覚的特徴等であり得る。ユーザインタラクションをサポートするために、コネクタシンボル(本明細書では単に「コネクタ」ともいう)は、つながりをサポートする証拠の関係記述909にアクセスするためにユーザがアクティブにできるユーザインターフェイス要素も含み得る。

【0216】

ステップ912に示されるように、方法900は、ディスプレイとのユーザインタラクションを受信することを含み得る。例えば、これは、ユーザがユーザインターフェイス内でコネクタシンボルの上をクリックすること又はホバリングすること、第1のアイコン及び第2のアイコンを選択すること又は追加の情報の要求、追加の入力の提供又は関係記述909又は裏付け証拠とのやりとりのために所定のやり方でグラフィカルディスプレイとやりとりすることを含み得る。

【0217】

ステップ914に示されるように、方法900はユーザインタラクションに回答することを含み得る。上述したように、方法900は、多数のドキュメントを処理し、その中でのエンティティの言及を特定し、さらに、関係記述として記憶され得る2つのエンティティ間の関係を証明する係るドキュメント内の重要な情報をさらに特定するステップを含み得る

。この情報を用いて、ユーザインターフェイスは、特定の関係の根拠の調査を容易にすることに加えてドキュメント群により示唆される追加の関連エンティティ及び関係を検索するために様々な有用なやりとりをサポートし得る。

【0218】

一態様では、ユーザインタラクションに応答することは、ユーザインタラクション内で証拠へのアクセスを提供することによりコネクタとのユーザインタラクションに応答することを含み得る。例えば、2つのアイコンを視覚的につなげるコネクタとのユーザインタラクションに応答して、これは、例えば、以前に記憶された関係記述の一部を表示することにより、示された関係をサポートする証拠へのアクセスをユーザインターフェイス内で提供することを含み得る。この文脈において、アクセスを提供することは、証拠の少なくとも一部をテキスト等として表示することを含み得る。例えば、証拠は、第1のエンティティ及び第2のエンティティの共起言及(co-occurring mentions)を含むドキュメントのコレクション内のドキュメントの1つ以上からのテキストを含んでもよく、証拠へのアクセスを提供することはテキスト中の共起言及を強調することを含んでもよい。

10

【0219】

別の態様では、マウスオーバー又はマウス選択等のユーザインターフェイス内に表示されたコネクタシンボルとのユーザインタラクションに応答することは、第1のエンティティと第2のエンティティとの間のつながりを実証するドキュメントについての検索要求を含めるためにユーザインターフェイスの検索入力フォームフィールドを更新することを含み得る。

20

【0220】

上記によれば、本明細書では、検索エンジン及びコンピュータを含むシステムも説明される。検索エンジンは上述した装置のいずれかでホストされてもよく、第1のエンティティ及び第2のエンティティの識別を受信し、ドキュメントのコレクションを処理して第1のエンティティと第2のエンティティと間のつながりを実証する証拠を特定するように構成されてもよく、証拠はドキュメントのコレクションうちの少なくとも1つのドキュメント内での第1のエンティティ及び第2のエンティティの共起言及を含む。コンピュータはプロセッサ、メモリ、ディスプレイ及びユーザ入力装置を含んでもよく、コンピュータは、本明細書に記載されるようにエンティティの関係を調査するために関係記述909及びドキュメントのコレクションとのユーザインタラクションをサポートするように構成され得る。例えば、プロセッサは、第1のエンティティ及び第2のエンティティを特定するステップと、第1のエンティティ及び第2のエンティティを検索エンジンに提出するステップと、検索エンジンからの結果に応答して、第1のエンティティのための第1のアイコン、第2のエンティティのための第2のアイコン及びコネクタシンボルをユーザインターフェイス内に表示するステップであって、コネクタシンボルは第1のアイコンを第2のアイコンに視覚的に結合し、それらの間のつながりを表し、コネクタシンボルはユーザインターフェイス内で証拠へのアクセスを提供するためにユーザインターフェイスを介して操作可能である、ステップと、を行うためにコンピュータ実行可能コードにより構成され得る。一態様では、検索エンジンはコンピュータ上でローカルに実行する。別の態様では、検索エンジンはデータネットワークを介してコンピュータによりアクセス可能な遠隔検索エンジンである。検索エンジンは様々な形でこれらの装置間で分散されてもよい。例えば、コンピュータは、例えば明示的な検索識別子を特定するために又は検索エンジンにクエリを送る前にテキストを前処理するためにクエリのテキストをローカルでパースし得る。別の態様では、コンピュータ上のローカル検索エンジンは、ローカルでホストされるドキュメントの調査をサポートしてもよく、広域ネットワーク上で分散されているか又は当該コンピュータ以外のコンピュータリソース上でホストされているか又は利用可能なドキュメントのコレクションを処理する遠隔検索エンジンと協働してもよい。

30

40

【0221】

図10はエンティティ間の関係を調査するための方法のフローチャートを示す。

【0222】

50

ステップ1002に示されるように、方法1000は、本明細書に記載の検索エンジンのいずれか又は他のエンティティベースのツール等の検索エンジンのためのユーザインターフェイスの検索入力フォームフィールドを介して、ユーザからキーワード検索を受け取ることから始まり得る。キーワード探索は、ユーザが意図する検索の明示的なステートメントを形成する任意のテキスト入力と共に任意の演算子又は他のテキスト等を含み得る。

【0223】

ステップ1004に示されるように、方法1000は、キーワード検索に基づいてユーザの意図を予測することを含み得る。例えば、これは、キーワード探索に共参照アルゴリズムを適用することによりユーザがキーワード検索で参照することを意図する1つ以上のエンティティを予測すること又はキーワード探索のテキストを分析してそこで特定される可能性が高いエンティティを特定することを含み得る。上述したように、この文脈での「エンティティ」という用語は、共参照として特定されたエンティティに関連する言及群(entity-related group of mentions)の略称であると理解すべきである。そのため、ステップ1004はより具体的には、ユーザが意図するとともに1つ以上のエンティティに関連する言及群により特徴付けられるエンティティを予測するために又は別の言い方をするとユーザが参照することを意図するエンティティに関連する言及群を予測するためにキーワード検索に共参照アルゴリズムを適用することを含み得ることが理解され得る。

10

【0224】

ステップ1006に示されるように、方法1000は第1の検索結果を提示することを含み得る。これは、例えば、第1の複数のエンティティタグ(本明細書では単に「タグ」ともいう)と、例えば、ドキュメントのコレクションから得られた複数のドキュメントとを含む検索結果をユーザインターフェイス内に表示することを含み得る。一般に、各ドキュメントはその全体が表示されるわけではない。むしろ、ユーザインターフェイス内でのドキュメントの表示は、タイトル、日付、ソース及び/又は他の情報等のメタデータに加えて、証拠パッセージ(evidence passage)又は他の抜粋等の代表的なコンテンツと、任意で証拠パッセージ内でのエンティティの言及のハイライトを表示することを含む。複数のドキュメントのそれぞれは、1つ以上のエンティティのうちの少なくとも1つの検索エンジンにより特定された言及(又は1つ以上のエンティティに関連する言及群における検索エンジンにより検索された言及)を含んでもよく、複数のエンティティタグのそれぞれは言及のうちの少なくとも1つに対応し得る。そのため、各エンティティタグは共参照又は関連するものと特定されているエンティティに関連する一言及群からの特定の言及を表すことも理解されよう。一般に、ドキュメントのコレクションはキーワード探索が受信されるユーザ装置上でローカルにホストされるか、広域ネットワーク上で分散されるか、遠隔クラウドストレージ施設上でホストされるか又はこれらの何らかの組み合わせであり得る。

20

30

【0225】

ステップ1008に示されるように、方法1000は、例えば、ユーザによる第1の複数のエンティティタグからの第1のエンティティのための第1のエンティティタグの選択を受信することにより、エンティティタグのうちの1つのユーザ選択を受信することを含み得る。上述したように、第1のエンティティは、第1のエンティティに関連する言及群により特徴付けられ得るか又は対応するエンティティに関連する言及群であり得る。選択は、ユーザインターフェイス内でエンティティタグをクリックするか又は他の操作から機械的に構成され得る。一般に、この第1の選択は、キーワード検索が第1のエンティティを参照することを意図するものであったことをユーザが確認したことを示し得るため、ユーザにより選択された特定のドキュメントにおける対応するエンティティ言及の基礎とされ得る。

40

【0226】

ステップ1010に示されるように、方法1000は第1のエンティティタグを検索入力フォームフィールドに追加することを含み得る。そのため、テキストの一部は、対応するエンティティ及び/又はユーザによって選択されたドキュメント内のそのエンティティのために提供される特定の文脈で置き換えられ得る。一態様では、エンティティタグを検索入力

50

フォームフィールドに配置するというこのアクションは、特定のエンティティが1つ以上の特定のドキュメントを参照して選択されたという視覚的なフィードバックをユーザに提供するユーザインターフェイス機能である。別の態様では、これは、ユーザにより最初に提供されたキーワードの代わりに用いることができる情報を提供してエンティティについて改善されたエンティティベースの検索をサポートする。

【0227】

ステップ1012に示されるように、方法1000は他のエンティティとの関係の証拠を見つけることを含み得る。この文脈では、証拠は、関連性アルゴリズム又は他のツール若しくはメトリックに従ってエンティティ間の関係を実証する傾向がある任意の証拠パッセージ等であり得る。また、本明細書で述べたように、各エンティティは、エンティティに関連する言及群により特徴付けられ得るか又は含み得る。そのため、方法1000は、第1のエンティティと1つ以上の他のエンティティとの間の関係の証拠(例えば証拠パッセージ)を提供するテキストを第2の複数のドキュメント内で特定するために関連性アルゴリズム又は他のツール若しくはメトリックを適用すること又は他のエンティティとの可能性のある関係の存在を実証する情報を特定することを含み得る。

10

【0228】

ステップ1014に示されるように、方法1000は、他のエンティティとの関係の証拠の検索の間に特定された1つ以上の他のエンティティ(又は1つ以上の他のエンティティに関連する言及群)のためにユーザインターフェイス内で第2の複数のエンティティタグを提示することを含み得る。これらのタグは、ユーザによる便利な視覚的な分析及び選択のために、例えば検索入力フォームフィールドの真下に又は視覚的に隣接する他の位置で表示され得る。

20

【0229】

ステップ1016に示されるように、方法1000は、例えば第2の複数のエンティティタグ内のタグのうちの1つに向けてユーザインターフェイス内でポイント及びクリックすることを介してユーザが第2の複数のエンティティタグから第2のエンティティ(又はエンティティに関連する言及群)のための第2のエンティティタグを選択したものである第2の選択を受信することを含み得る。

【0230】

ステップ1018に示されるように、第2のタグをユーザが選択したことに応答して、方法1000は第2のタグに基づく検索を行うことを含み得る。とりわけ、第1のエンティティと第2のエンティティとの間の関係を実証するドキュメントについて、例えば証拠パッセージ等を通じて新たな検索が行われ得る。これは、任意の好適なドキュメントのコレクションの完全に新しい検索を含んでもよいし又はこれは、例えば、上記のステップ1012で関係の証拠を検索したときに先に得られた結果の読み出しを含み得る。

30

【0231】

ステップ1020に示されるように、方法1000は、例えば、第1のエンティティと第2のエンティティとの間の関係を実証する第2の複数のドキュメントのうちの1つ以上を含む第2の検索結果を提示することにより、新たな検索の結果を表示することを含み得る。これらの結果は様々な方法のいずれかで表示され得る。例えば、これは、第1のエンティティと第2のエンティティとの間の関係を、第1のエンティティのための第1のアイコン、第2のエンティティのための第2のアイコン及び第1のアイコンを第2のアイコンに視覚的に結合して、例えば関係を視覚的に表すコネクタシンボルとしてユーザインターフェイス内に視覚的に表示することを含み得る。コネクタシンボルは、ユーザによる第2の複数のドキュメント内の関係の証拠へのアクセスを提供するユーザインターフェイス要素を含み得る。結果を表示することは、上記に加えて又は代えて「<>」シンボル等の関係演算子(relationship operator)により第2のエンティティタグに結合された第1のエンティティタグとして、検索入力フォームフィールド内で関係を表示することを含み得る。これに加えて又は代えて、追加の関連する処理が行われ得ることが理解されるであろう。例えば、第1のエンティティと第2のエンティティとの間の関係を特徴付ける関係記述(relatio

40

50

nship description)が、例えばユーザによるユーザインターフェイス内でのさらなる使用のために任意の好適なやり方及び場所に記憶され得る。

【0232】

別の態様では、本明細書の技術を用いて特定された多数のエンティティタグ及び関係が視覚的に表示され得る。そのため、例えば、第2の検索結果を提示することは、グラフ内のエッジにより第1のエンティティにそれぞれが接続された複数のエンティティタグを視覚的に提示することを含んでもよく、各エッジは、ユーザが裏付けとなる証拠パッセージを検索するためにユーザインターフェイス内で操作可能であり、各エンティティタグは、グラフ表示により反映された知識グラフから対応するエンティティに関連する言及群を追加又は削除するためにユーザにより操作可能である。

10

【0233】

別の態様では、上述した方法1000をサポートするシステムが本明細書で開示される。例えば、本明細書で開示のシステムは、検索エンジン、データネットワーク及びデータネットワークを介して検索エンジンに連結されるコンピュータ装置を含み得る。コンピュータ装置はプロセッサ、メモリ及びディスプレイを含んでもよく、このメモリは、ユーザインターフェイスの検索入力フォームフィールド内でユーザからのキーワード検索を受信するステップと、ユーザが意図する1つ以上のエンティティを予測するステップと、ユーザインターフェイス内でユーザに第1の検索結果を提示するステップであって、第1の検索結果は第1の複数のエンティティタグ及び複数のドキュメントを含み、複数のドキュメントのそれぞれは、1つ以上のエンティティのうちの少なくとも1つの検索エンジンにより特定された言及を含み、複数のエンティティタグのそれぞれは言及のうちの少なくとも1つに対応する、ステップと、第1の複数のエンティティタグから第1のエンティティのための第1のエンティティタグのユーザによる第1の選択を受信するステップであって、第1の選択はキーワード検索が第1のエンティティを参照することを意図したものであったとのユーザの確認を示す、ステップと、第1のエンティティタグを検索入力フォームフィールドに追加するステップと、第1のエンティティと1つ以上の他のエンティティとの間の関係の証拠を提供するテキストを第2の複数のドキュメント内で特定するステップと、1つ以上の他のエンティティのための第2の複数のエンティティタグをユーザインターフェイス内で提示するステップと、第2の複数のエンティティタグから第2のエンティティのための第2のエンティティタグのユーザによる第2の選択を受信するステップと、第1のエンティティと第2のエンティティとの間の関係を実証する第2の複数のドキュメントのうちの1つ以上を含む第2の検索結果を提示するステップと、を行うためにプロセッサが実行可能なコードを記憶し得る。

20

30

【0234】

エンティティベースの関係を発見及びナビゲートすることをサポートする追加のユーザインターフェイス技術も本明細書で開示される。例えば、一態様では、本明細書で検討される方法は、検索エンジンのためのユーザインターフェイスを表示するステップであって、ユーザインターフェイスは検索入力フォームフィールドを含む、ステップと、検索入力フォームフィールドへのユーザ入力を受けるステップと、ユーザ入力をパースして第1のエンティティを特定する第1のテキスト文字列、第2のエンティティを特定する第2のテキスト文字列及び第1のテキスト文字列と第2のテキスト文字列との間の演算子を特定し、第1のエンティティと第2のエンティティとの間の関係の証拠について検索するための要求を検索エンジンに指定するステップと、著者が第1のエンティティを参照することを意図している共参照アルゴリズムが予測する言及と、著者が第2のエンティティを参照することを意図している共参照アルゴリズムが予測する言及とを含むとともに、第1のエンティティと第2のエンティティとの関係の証拠を提供すると関連性アルゴリズムが予測するテキストを提供するドキュメントのコレクション内の1つ以上のドキュメントを検索することと、1つ以上のドキュメントのうちの複数のユーザインターフェイス内で提示することを含む。

40

【0235】

50

図11は、エンティティの間の関係を調査するためのユーザインターフェイスを示す。一般に、ユーザインターフェイス1100はコンピュータ装置のディスプレイ上にレンダリングされてもよく、ユーザからテキスト入力を受けるように構成された検索入力フォームフィールド1102を含み得る。図11に示すように、例えば、検索結果からの関心のあるドキュメントの要約1104がドキュメントにより記述されている可能性が高いエンティティを参照する1つ以上のエンティティタグ1106と共に表示され得る。エンティティタグ1106はユーザにより選択可能であってもよく、上述したように(a)選択されたタグを検索入力フォームフィールド1102内に視覚的に配置すること及び/又は(b)検索結果に示される特定のドキュメントにおけるエンティティの言及を基礎として用いることによりユーザの選択に応答し得る。本明細書でさらに説明するように、特定のドキュメントにおいて基礎として用いられるエンティティの言及を用いて、ユーザインターフェイス1100は次に検索入力フォームフィールド1102で特定されるエンティティに関連し得る多くの追加のエンティティのためのタグを提示し得る。ユーザは、例えばエンティティタグ1106とのやりとりを通じて対応するエンティティを選択してもよく、これらのユーザ関係表示は、本明細書に記載されているように、実証された関係の視覚的な可視化を作成するために関係を実証する好適な証拠パッセージを検索するのに用いられ得る。

10

【0236】

図12は、エンティティ間の関係を調査するためのユーザインターフェイスを示す。図12のユーザインターフェイス1200では、2つのエンティティが、2つのエンティティ間の関係を実証するドキュメントの検索を示す演算子シンボル1204(本例では「<>」)と共に検索入力フォームフィールド1202内に提示され、結果1206は例えば、迅速な視覚的識別のためにハイライトされた対応するエンティティの言及と共にユーザインターフェイス1200内で概して提示される。特定のユーザ検索又は対応する関係記述及び他の情報を含む既存のプロジェクト等のいずれかに基づいて、エンティティの関係の視覚的なレンダリング1208も提供され得る。なお、関係を示すために線又は矢印等のコネクタシンボル1210を用いてもよいし、コネクタシンボル1210は関係の存在又は性質をサポートする証拠パッセージを含むドキュメントの検索等の機能を行うためにユーザにより選択可能であってもよい。

20

【0237】

上記の方法及びシステムの非限定的な例示の実施を通じて追加の詳細及び説明を提供する。根拠として用いられるエンティティの言及及び関連技術は2019年4月10日に出願されるとともにその全体が参照により本願に組み込まれる米国仮出願第62/832085号における非限定例においても見出すことができる。

30

【0238】

本明細書で用いられる概念の言及とは人間の読者が共有概念を参照していると認識し得る文字列であり得る。言及は、読者が言及群の意味を解釈して曖昧さを減らすことができる他の概念についての十分な追加の言及を含む場合に基礎として用いられ得る。このような曖昧さの排除プロセスは基礎として用いること(grounding)と関連して起こる。より徹底的に根差したものであればあるほど、概念はその意味についての豊富な議論にあり、曖昧さがより徹底的に取り除かれるほど、読者はより正確に理解できるようになる。概念は人々の間で伝えることができる一般的な考えであり、話し合う人達は自分達が共通の考えを有すると考える。イベントとは空間及び時間に根ざした概念であり、そのような共通の考えについて話し合う人達は何処でそして何時そのイベントが起こったのかについての共通の理解を有する。イベントは概してそのイベントが起こるか又は起こっている空間及び期間の範囲を有する。「エンティティ」は、区別可能な属性によっても特定される概念の特殊なサブセットであり、あるエンティティについて話し合う人達は自分達がどのエンティティについて話をしているかについて共通の理解を有する。人及び企業は身近な種類のエンティティである。エンティティは一般に名前を持ち、人々が所与のエンティティについて話し合う場合、実世界のエンティティを参照する共有概念としてその名前が通常用いられる。前述の説明はエンティティを強調しているが、本明細書のシステム及び方法は

40

50

、エンティティの代わりに又はエンティティに加えて、任意の他の概念と共に用いられ得ることが理解されるであろう。

【0239】

あるドキュメントが2つの概念と一緒に言及する場合、概念は「共起」する。2つの異なるエンティティを参照する言及の共起は、2つのエンティティが直接的な関係を有していることを必ずしも含意しない。例えば、ウェブページにファンプロワーを販売する企業のリストが含まれている場合があり、2つの異なる企業の名前がそのリストにあるからといって、その共起はこの2つの企業が提携関係等の直接的な関係を有することを必ずしも意味しない。そのような共起は、双方が同様の又は関連する製品の製造業者であるといった間接的な関係を実証するにすぎないかもしれない。しかしながら逆もしばしば起こる。すなわち、2つのエンティティが直接的な関係を有する場合、その直接的な関係は双方のエンティティを参照するテキストのパッセージ又は他のメディアオブジェクトでしばしば述べられているか又は言及されている。すなわち、2つのエンティティへの共起参照を含む。したがって、共起参照を含むドキュメントを見つけることがしばしば有用となる。何故なら、それらは2つのエンティティ又は概念間の関係の証拠を実証するか又は提供する候補ドキュメントだからである。

10

【0240】

本明細書で検討する検索エンジンインターフェイスは、関係を実証し得るそのような候補ドキュメントの要求をユーザが伝えることができるようにし得る。ビジュアルインターフェイスは、(i)第1のエンティティの曖昧さをなくす第1の基礎として用いられる言及の第1のビジュアルアイコン、(ii)第2のエンティティの曖昧さをなくす第2の基礎として用いられる言及の第2のビジュアルアイコン及び(iii)第1のビジュアルアイコンと第2のビジュアルアイコンとを相互につなげる視覚的リンクという少なくとも3つの要素を含み得る。このような入力は、検索エンジンのクエリパーサにより、第1のエンティティと第2のエンティティとの間の可能性のある関係を実証するドキュメントを検索するための要求として解釈できる。アイコンは、基礎として用いられる言及、例えば、ドキュメントの著者が何らかの現実世界のエンティティを参照することを意図している参照として人間が認識する可能性が高い特定のドキュメントの文字スパンを表す。

20

【0241】

好ましい実施形態では、ユーザは異なるエンティティのために1つ、2つ、3つ又はそれ以上のビジュアルアイコンを入力できる。基礎として用いられる言及を示すそのような1つのビジュアルアイコンを入力することにより、ユーザは示された基礎として用いられる言及と同様の又は文字列が同じ名前の別のエンティティに言及し得る他のドキュメントではなく、その特定のエンティティに言及するドキュメントを要求し得る。様々なエンティティに対する基礎として用いられる言及を示す2つ以上のビジュアルアイコンを入力することにより、ユーザはこれらの複数のエンティティに言及するドキュメントを要求する。このようなクエリは「マルチエンティティクエリ」と呼ばれることがある。

30

【0242】

一実施形態では、検索エンジンアルゴリズムはそのようなマルチエンティティクエリに対して、特定されたエンティティにより多く言及するドキュメントをリストの上位にランク付けすることにより応答する。そのようなクエリはSTRICT ANDよりはむしろSHOULD BOOSTINGクエリと呼ばれることがある。検索エンジンアルゴリズムがクエリに対してSTRICT ANDセマンティクスを適用する場合、それはユーザのマルチエンティティクエリ入力により特定されるエンティティの全てに言及するドキュメントにしか反応しないことになる。それとは対照的に、SHOULD BOOSTINGセマンティクスはランキングアルゴリズムが、マルチエンティティクエリ入力により特定されるエンティティの全てよりも少ない数のエンティティに言及するドキュメントに対応させることができる。エンティティの全て又は大半に言及するドキュメントはより高くランク付けされるが、結果のリストは、エンティティのうちの多分1つだけを含むエンティティの全てよりも少ない数のエンティティに言及するドキュメントを含み得る。結果のリストは、ランキングアルゴリ

40

50

ズムが、それらの結果の内容が、ユーザのマルチエンティティクエリ入力における様々なエンティティのビジュアルアイコンに示される基礎として用いられる言及を取り巻く文脈に関連すると判断した場合に、入力クエリにおけるエンティティのいずれにも言及していない結果を含むことになる。

【0243】

基礎として用いられる言及を取り巻く文脈は人間の読者が意味を認識し、ソース/著者により意図される共参照を類推できるようにする。この文脈は、検索結果を改善しユーザによりフラグ付けされたエンティティの特定の言及に検索行動を固定するために本明細書に記載のシステム及び方法と共に用いられ得る。

【0244】

1つの好ましい実施形態では、ユーザインターフェイスは、クエリ内の各ビジュアルアイコンのための複数の基礎として用いられる言及をユーザが特定できるようにする。例えば、ジョン・スミスのビジュアルアイコンは、その名前の特定の人物のいくつかの言及の群を表すことができる。これらのドキュメントは、各ドキュメント内の言及を取り巻く様々な文脈的なパッセージを有する複数のドキュメントに現れ得る。所定のドキュメントはエンティティに複数回言及し、そのような言及を取り巻く言葉は文脈を提供する。好ましい実施形態では、検索エンジンアルゴリズムは、これらの1つ以上の言及を取り巻く文脈の次元数縮退ベクトル表現(dimensionally reduced vector representation)を構築するため、各ビジュアルアイコンは、これらの言及を取り巻く集約された又は凝集された文脈の機械表現に対応する。そのようなベクトル表現は当該技術分野では一般的であり、多くの形態をとることができる。例えば、1つの表現は、3つの文の窓(three-sentence window)といった各言及の周囲の単語の窓からの単語及びフレーズを数える単語の袋を用いり得る。別の例では、埋め込みベクトルは、ニューラルネットワークを自然発生の言葉のシーケンスに基づいてフレーズを予測するように訓練した後でニューラルネットワークの内部アクティベーションから構築される。別の例示的アプローチでは、埋め込みベクトルは、共起統計量のマトリクスを因数分解することにより学習される。他のアプローチは、これらのアプローチと他のベクトル化アプローチとを組み合わせ得る。これらの全てのアプローチの1つの顕著な特徴は、検索クエリ入力がユーザの意図する検索対象の曖昧さをなくすユーザ自身の基礎として用いられる言及を集め、ひいては検索エンジンアルゴリズムがこれらの改善された入力に基づいて結果のランク付けを改善することができる点である。

【0245】

「エンティティ参照セット(entity reference set)」又はERSという用語は、本明細書ではエンティティへの言及群を意味し、より具体的には(本明細書に記載のインターフェイスを用いて)ユーザにより集められてクエリの一部を基礎とする一式にされたエンティティベースの言及群を指す。この言及群は本明細書ではときおり単に「エンティティ」とも呼ばれるが、エンティティ参照セットはより正確には、ドキュメント及び各ドキュメント内の特定のサブレンジに対する識別子群であり、ドキュメントの内容は対象のエンティティを参照する。ERSは当該エンティティを参照していないとユーザがマークしたエンティティへの言及である「否定的言及(negative reference)」も含むことができる。ユーザが、当該エンティティを正しく参照しているマークした言及は「肯定的言及(positive reference)」と呼ばれる。例えば、ゼネラルモーターズ社の会長であるジョン・スミス氏に関するコンテンツを探索する場合に、ユーザがその正しい人物に言及するドキュメントを見つけると、ユーザはそのドキュメントをジョン・スミス氏のERSにおいて肯定的言及としてマークすることができる。ユーザがフットボールコーチのジョン・スミス氏の言及を見つけた場合、ユーザは後者の言及は意図した人物を参照していないため、それゆえ否定的言及としてマークすることができる。このように、ユーザは、同様の又は同一の表層形式の名前(surface form name)の文字列を有する2つの異なるERSを示すビジュアルアイコンを有するクエリを有することができる。一方のERSは会長を参照し、他方のERSはコーチを参照することができ、それぞれのERSにおける基礎として用いら

10

20

30

40

50

れる言及は、たとえ表層形式の名前の文字列が同じであっても意図した意味の曖昧さをなくす。

【0246】

エンティティの言及を集めるプロセスは、「言及レベルブックマーキング(mention-level bookmarking)」とも呼ばれ得る。何故なら、例えば、関係記述又は他のデータ構造においてシステムにより記憶される識別子はドキュメントと特定の言及のドキュメントの双方を特定するからである。1つのエンティティのためのこのようなブックマーク群はERSである。いくつかのERSを一緒に用いて、エンティティ間の関係の証拠のためのマルチエンティティクエリを指定することができる。エンティティ参照セット群は「プロジェクト」として追跡することもできる。上述のように、知識操作ジャーナル(knowledge operations journal)又は「KOJ」は、1つ以上のグラウンデッドエンティティ、すなわち、曖昧さが取り除かれた対象エンティティへの言及を有するプロジェクトを実施できる。このようなプロジェクトは、プレゼンテーションファイルやメモ等の他の種類のオフィスファイルアーチファクトのようにユーザがプロジェクト名をつけて保存し、将来戻ることができる知識アーチファクトである。検索エンジンアルゴリズムは、ユーザがクエリの文脈として特定のプロジェクトを特定できるようにし、次いで、検索エンジンはプロジェクトに基づいてドキュメントを違った形でランク付けできる。検索コンテキストのためにプロジェクトを用いるための好ましい実施形態は、プロジェクト内の全てのエンティティ参照セット内の全ての基礎として用いられる言及の全ての文脈から集約ベクトル(aggregate vector)を構築し、そのベクトルを検索結果に対して多様化のための影響(diversifying influence)として用いる。多様化のための影響は、プロジェクトと重複しているように見えるドキュメントを軽く抑え得るため、ユーザは検索結果のヒットから新たな情報を学ぶ可能性が高くなる。

10

20

【0247】

各ERSのために構築されたベクトルは、肯定的及び否定的言及を異なった形で用いることができる。肯定的言及は、検索エンジンにより検索されるドキュメントを拡張する促進的フレーズ(boosting phrase)を提供でき、否定的言及は誤ったドキュメントを拒絶する抑制的フレーズを提供できる。例えば、ユーザが会長についてのドキュメントのみを探している場合、検索ランキングアルゴリズムが正しいジョン・スミス氏についてのドキュメントのランキング位置を高めるために用いることができるコンテキストワードとして「自動車」を提供し得る。反対に、否定的言及は、検索ランキングアルゴリズムが間違ったジョン・スミス氏についてのドキュメントのランクを下げるために用いることができるコンテキストワードとして「スタジアム」を提供し得る。

30

【0248】

検索エンジンランキングアルゴリズムはいくつかの方法でランキングを改善するためにそのような基礎として用いられる言及のコンテキスト情報を用いることができる。ランキングの改善の別の例として、検索エンジンは肯定的言及の文脈と同様のドキュメントを抑えることができる。何故なら、これはユーザが対象のエンティティについて学ぶのを助ける、新しく、より新規である可能性の高い情報を提供するからである。あるいは、検索エンジンは、肯定的言及の文脈と同様のドキュメントを増やすことができる。何故なら、これは、ユーザがすでに集められたドキュメントの言及から得た理解を裏付ける可能性が高いドキュメントを提供するからである。

40

【0249】

そのようなランキング技術は、Luceneインデックスエンジンを用いるエラスティックサーチ又はSOLR等の標準的なキーワード検索エンジンの上で実施できる。ERS群はユーザインターフェイス内で行われてもよく、クエリ処理モジュールは、エンティティ参照セットのマルチエンティティクエリを解釈して、肯定的及び否定的なブースト文脈を含む複雑なクエリを受信するように構成されたより低レベルの検索インデックスのためのクエリを構築できる。否定的なブーストは、SHOULD NOT節で肯定的なブーストとして実施できる。

50

【0250】

クエリがそのような検索エンジンからドキュメントのコレクションを検索すると、検索エンジンに他のエンティティの言及等の他の共起的用語の集合を生成するか又は共参照として曖昧性除去エンジンが評価した言及群のための識別子を設定できる。ユーザインターフェイスは、推奨される他の対象エンティティとしてそのような集約から用語を提示することができる。クエリにより検索されたドキュメントのサブセット内で頻繁に言及される他のエンティティは「関連エンティティ」である。システムは関連エンティティを集め、それらをマルチエンティティクエリを用いた更なる探索のための奨励としてユーザに提示する。

【0251】

例えば、マルチエンティティクエリは、3つのERSにおける基礎として用いられる言及により特定される3つのエンティティを有し得る。第1がGMの会長のジョン・スミス氏であり、第2がデトロイトであり、第3がアリソン・トランスミッションである。このクエリに対応して、検索エンジンは、それらのエンティティに言及するドキュメントを探し、次に集約クエリはこれらのドキュメントで頻繁に言及されている他のエンティティ(例えばスミス氏が役員であるデルタ航空等)を探す。システムはこのアプローチを用いて、ユーザがまだ探索することを考えていなかったであろう他のエンティティへの多数の候補的つながりを提示し得る。そのような関連エンティティを提案することにより、システムはユーザにより多くのマルチエンティティクエリを探索することを促す。例えば、ユーザはデトロイト及びアリソン・トランスミッションを削除し、デルタを加えてスミス氏とデルタとの関係の証拠を調査するために新たな2エンティティクエリを作成し得る。

【0252】

関連エンティティの示唆を生成するそのような集約クエリは、表層形式の名前文字列を生成し得るが、それはとりわけ複数のエンティティが同様の又は同一の表層形式の名前文字列を共有する場合に基礎として用いられる言及のERSを形成するのに十分ではない可能性がある。そのため、システムは、例えばユーザが選択した基礎に基づいて、提案された関連エンティティに言及する特定のドキュメントに関連エンティティをさらに合わせることができるため、ユーザがそのような示唆を選択した場合、エンティティのビジュアルアイコンにより示されるERSとして検索問合せフォームフィールドに入る。

【0253】

ドキュメントが1つの特定のエンティティについて1回以上言及する場合、ユーザインターフェイスは、ドキュメントアイコンと概念アイコンとを結び、矢印が概念アイコンを指す視覚的な線によりこれを示すことができる。特定のドキュメントからそのような矢印群はERSを視覚化する別の方法である。

【0254】

関係の証拠はドキュメントのコレクションの中から集めることができる。2つの概念間の関係を実証するドキュメントは、関係の種類の証拠をさらに提供することができる。例えば、人体が風邪(病気)を生じさせるウイルスを撃退するのをビタミンCがどのように助けるかを説明するジャーナル記事の場合を考えてみる。そのようなドキュメントは、「ビタミンC」という概念と「風邪」という概念との関係の証拠を提供する。そのドキュメント中の他の概念の言及は、その関係の特徴付ける文脈を提供する。

【0255】

上述したように、共参照解析又は「共参照」とは言及に意味を割り当てるプロセスである。共参照解析を支援するために、ユーザアクションのための文脈は、推奨のリクエストが行われる文脈等の関連性を含意し得る。別の態様では、ユーザアクションの順序は、ユーザが知識グラフに概念を加える順序又はグラフのエッジの実証を要求した順序等の関連性を含意し得る。別の態様では、ユーザアクションの頻度は、ユーザがグラフの特定のエッジの実証を要求する頻度又はユーザがグラフ内の頂点により表される特定の概念に関連する概念を要求する頻度等の関連性を含意し得る。

【0256】

話し言葉を読むか又は理解するプロセスは、著者又は話し手が意図する意味を解釈するプロセスである。人間の心は、著者が何を意図しているか又は意味しているかを絶えず推測している。コミュニケーターが言及しようとしているものについて読者又は聞き手が一度も観察するか又は考えたことがない場合、読者又は聞き手にとって理解するのが困難になる。それとは対照的に、読者又は聞き手が依然に考えに遭遇したことがある場合、この同等的なもの(equivalence)を検出することにより理解する行為が大幅に簡素化される。同様に、検索エンジンにキーワード検索を提示する代わりに、本明細書に記載の技術は、ある人間により書かれた検索ユーザの関心のエンティティを参照することを意図する特定のメッセージを特定することにより、特定の時間に明示的な人間による入力を求めて意図する検索クエリをより正確に表現する。例えば、「ジョン・スミス氏」のためのキーワード検索をタイプして、その名前を持つ多くの異なる人物を参照する検索結果のヒットを除外することに代えて、本明細書で検討される検索エンジンのユーザは、特定のジョン・スミス氏に言及するメッセージを特定することにより自身のクエリを洗練できる。そうすることで、検索ユーザはこれらの言及を取り巻く豊かな文脈を活用し、その文脈は、著者がその特定のジョン・スミス氏を参照するという著者の意図により条件付けられた考え方の中で書いた文脈である。

10

【0257】

本明細書で説明する技術と、使用する用語とのさらなる例として次のテキストを考えてみる。「1992年以来ゼール社のCEOであるジョン・スミス氏は彼自身の報酬が21%増加して120万ドルになった。これは、3年間にわたってその金融サービス会社の取締役会長を務めた人物であるサンドラ・ジョーンズ氏に代わって47歳のスミス氏がその職に就任したからである。」

20

【0258】

上記のメッセージは特定のエンティティであるジョン・スミス氏、ゼール社及びサンドラ・ジョーンズ氏を参照する言及を含む。これらのエンティティは、この特定の言及ストリングにおいて共起する。しかしながら、メッセージ内の他の言及ストリングと同じ現実世界のエンティティを参照するものとして人間の読者が自然に解決する「彼」や「人物」等の名目上の言及も存在する。この場合、「彼」は「ジョン・スミス」が参照するものと同じのエンティティを参照し、「人物」は「サンドラ・ジョーンズ」が参照するものと同じのエンティティを参照する。すなわち、「彼」は「ジョン・スミス」と共参照関係にあり、「人物」は「サンドラ・ジョーンズ」と共参照関係にある。これらの関係の混同は、例えば「人物」がジョン・スミスを参照するか又は「彼」がサンドラ・ジョーンズを参照する等の参照エラーにつながる。他のドキュメント内の他のメッセージもこれらのエンティティに言及していることがあり、別のドキュメント内の「ジョン・スミス」に対する別の言及ストリングが上記のメッセージの「ジョン・スミス」と同じエンティティを参照しているかどうかについて人間の読者は共参照プロセスを行う。

30

【0259】

例えば、同じ表層形式の名前を有する2つの言及はおそらく同じ概念を参照する候補である。これら2つの言及を取り巻く文脈が同様である場合、人間はその言及が同じ意味を持つ、すなわち共参照であると認識する可能性がより高い。アルゴリズムは2つの言及を取り巻く文脈におけるデータを比較し、その言及が共参照であると人間が認識し得る可能性を推定できる。

40

【0260】

このメッセージは共起エンティティ間の関係も含む。例えば、ジョン・スミス、ゼール社及びサンドラ・ジョーンズはこのメッセージの内容を通じて関連している。ドキュメントにわたってこれらの言及を処理することは、関連する共起エンティティを追いながら、各エンティティに対応する言及文字列を抽出することを含み得る。エンティティ間の関係は、エンティティを相互につなげるグラフィカルリンクを用いて可視化できる。ドキュメントが1つの特定のエンティティに何度も言及する場合、ドキュメントアイコンを指す矢印付きのエッジにより示すことができる。そして、このメッセージは、2つ以上の共起エン

50

ティティ間の関係を実証する証拠、例えば証拠パッセージとなり得る。

【0261】

関係とその証拠は、個々のエンティティについてさらに学ぶだけでなく、エンティティ間のつながりを理解するために検索する上でユーザにとって重要なターゲットである。上述したように、例えば、グラフィカルディスプレイにおける又は2つのエンティティタグの選択を介したユーザによる特定のつながりの選択は検索入力フォームフィールドにおけるクエリを変更し得る。この検索入力フォームフィールドでは、「<>」シンボル(又は別のシンボル)は、シンボルにより結合された2つのエンティティ間の関係を実証するドキュメントのクエリを明示的に示す検索演算子として機能し得る。キーワード検索入力は、先ず第1の検索文字列を表示し、第1の検索文字列に回答して一覧となった検索結果を表示し得る。知識オブジェクト要約カード(又は知識操作ジャーナルの他の表現)は関連エンティティの推奨を表示し、ユーザが関連エンティティの推奨をクリックすることに応答してユーザインターフェイスはキーワード入力を変更して、関連エンティティを表すクエリ文字列と、関連エンティティ間の関係の証拠を見つけるための検索演算子とを表示し得る。ユーザインターフェイスは、関連エンティティに言及するドキュメントを表示するために検索結果リストを追加で変更し、一覧となった結果のうち少なくとも1つは関連エンティティの強調表示された言及を示し得る。例えば、ユーザは「ジョン・スミス<>ゼール社」を検索し、これは、これらの2つのエンティティ間の関係の検索である。この検索の間に、ユーザが新規性及び関連性に加えて肯定的共参照及び否定的共参照等を制御して関係を実証するより多くの証拠を見ることが、ユーザがまだ見たことがない新たな関係の証拠を見ることがとを区別的にトレードオフできるようにするために追加のユーザインターフェイスコントロールが提供され得る。

【0262】

図13～図16は本教示の例及びさらなる説明として、エンティティ間の関係を調査するための様々なユーザインターフェイスを示す。図13のユーザインターフェイス1300では、キーワード検索1302はクエリ入力フォームフィールド内にある。入力キーワードサーチ1302に応答して、検索エンジンは検索結果ヒットのリスト1304を表示している。通常、結果はテキストの抜粋1306を示す。本教示のシステムでは、結果は1つ以上のタグ1308も表示し得る。各タグは、ドキュメント内のエンティティの1つ以上の言及を示す。タグ1308はドキュメントから例示の表層形式の名前文字列を示し得る。一実施形態では、共参照アルゴリズムの予測がキーワード検索1302によりユーザが意図したエンティティを参照する可能性が高いというタグ1308にカラーリング、太字又は他の視覚的強調が加えられ得る。キーワード検索1302はまだ基礎として用いられてないため、ユーザインターフェイス1300はユーザがタグ1308を選択することにより自身のクエリを基礎として用いることができるようにする。タグ1308が選択されると、システムは、図14のユーザインターフェイスに示すように、検索ボックス内にタグの視覚的表現を挿入する。具体的には、図14のユーザインターフェイス1400に示されるように、フォームフィールド1402内にあるタグ1416は、図13の結果リスト内の特定のドキュメントから特定のタグを選択するユーザアクションに応答してフォームフィールド1402内に表示される。このアクションは、ユーザが図13のキーワード検索1302により意図したものと同一エンティティを参照しているとして特定の言及を特定した。クエリを(フォームフィールド1402に表示される)基礎として用いるエンティティ言及検索クエリにアップグレードすることによって、インタラクションは、システムが曖昧さを除去し、ユーザが意図する対象エンティティを参照している可能性が低いものとして、図13の低位の検索結果ヒット1304を適切に抑えることができるようにするため、図13の低位の検索結果ヒット1304は抑えられており、図14には表示されない。図14は、図13の低位の検索結果ヒット1304が表示されたであろう場所に空の結果1414を有していることは注目に値する。図13の低位の検索結果ヒット1304は同じ名前の異なるエンティティに言及し、タグ1416を有する改善されたクエリは、検索エンジンのランキングアルゴリズムが空の結果1414をもたらす誤った結果を排除することを可能にする。

10

20

30

40

50

【 0 2 6 3 】

図13のユーザインターフェイス1300に戻って、システムはユーザに複数の関連エンティティ1310を、例えばこの特定の例では、コーパス1314内のドキュメントからエンティティであるジャック・ワン(Jack Wang)1312の言及を提示する。検索エンジンアルゴリズムは、ユーザのキーワード検索1302クエリに関連する様々なドキュメントでジャック・ワンに言及されていることを特定し、統計的ランキングアルゴリズムを用いてジャック・ワンの言及を他のエンティティの言及よりも上位にランク付けするため、それを図13のユーザインターフェイス1300に表示することを決定する。

【 0 2 6 4 】

図13のユーザインターフェイス1300では、ユーザはジョン・スミスのタグ1416についてのERSクエリの結果を閲覧している。このERSは、ユーザがそのタグを以前クリックしたため、1408により示される言及を含む。ユーザは1つ以上の追加のタグを選択してそれらをERSに追加することができる。例えば、ユーザはスミス会長1422をクリックして、その追加の言及をクエリフォームフィールド内のERSに追加することができる。一実施形態では、ユーザインターフェイスはユーザがERS内の言及を一覧表示する第2のウィンドウを開くことができるようにし、ユーザがERSから言及を削除するか又はそれらを別個のERSに分割できるようにする。一実施形態では、ユーザインターフェイスはユーザが一定期間にわたって集めた様々なエンティティ参照セット(ERSs)を追跡する。

【 0 2 6 5 】

図14のユーザインターフェイス1400では、システムは、ジョン・スミスのタグ1416に対するERSのための基礎として用いる言及の検索クエリに対応する結果を表示している。ユーザは、例えば、キンチュン(QingChun)社1418等のタグをクリックすることができる。このタグは結果により表されるドキュメント内の特定の連続の言及を表すため、ユーザがそのタグを選択すると、ユーザはユーザが意図するキンチュン社についての曖昧さのないコンテキストを提供する1つ以上のドキュメント1420内の言及群を選択する。キンチュン社1418のタグをクリックすることにより、ユーザは、タグ1416により示される言及により特定される2つのエンティティと、キンチュン社1418のタグとの間の関係を実証し得る証拠パッケージを見たいと自身が望むことをシステムに示す。そのようなマルチエンティティクエリは図15のユーザインターフェイス1500に示され、そこでは、関係演算子1524(すなわち、< >)は、1528により示されるERS及び1526により示されるERSにより示される言及により特定されるエンティティ間の関係を実証し得るドキュメントを見つけるための命令として、検索エンジンのクエリパーザによりパースされる。検索エンジンユーザインターフェイス1500は、2つの対象エンティティについて言及し得るとアルゴリズムが推測するテキストの部分列に視覚的な強調を提供する。例えば、1530により示される結果は、スミス会長1516及びキンチュン社1518の言及を有する。

【 0 2 6 6 】

図14のユーザインターフェイス1400も特定の結果についてのタグを示す。結果1404はジョン・スミスという名の人物に言及し、その言及は1408により示される。ユーザインターフェイス1400は、検索クエリ入力フォームフィールド1402内のERSにより特定されるエンティティとの関係を有し得る連続の提案されたエンティティ1410も示す。例えば、ビジュアルアイコン1412はジャック・ワンという名前の人物の言及群を示し、ユーザはこのビジュアルアイコン1412をクリックして、それをビジュアルアイコン1412の言及により特定されるエンティティとジョン・スミス(タグ1416)の言及により特定されるエンティティとの間の関係証拠のための新たなクエリとして検索ボックス内に置くことができる。

【 0 2 6 7 】

図16のユーザインターフェイス1600は、ドキュメントの集成からのエンティティの言及を示す2つのタグアイコン間の関係コネクタシンボル1608を表示する方法を示す。図16では、ユーザインターフェイス1600は、検索入力ツール1602内のクエリの結果を示している。結果は、様々なエンティティの言及を含むドキュメントを特定し、結果の表示

10

20

30

40

50

は、それらの言及のうちの一部(例えば、ジャック・ワンの言及を示すタグ1610)を示すタグを示し得る。タグにより特定されるそのようなエンティティは検索入力ツール1602のクエリ内にあってもよいし、なくてもよい。この例では、第1の結果1609に示されるドキュメントは、関心のある可能性のある関係としてユーザに表示されるビジュアルアイコン1604により示されるドキュメントの集成からのドキュメントのうちの1つであり得る。ジャック・ワンのためのERSを示すビジュアルアイコン1604は、コネクタシンボル1608によりサリー・スミスのためのERSを示すビジュアルアイコン1606につなげられている。ビジュアルアイコン1604及び1606の双方は、共参照アルゴリズムが同一のエンティティの言及である可能性が高いとしてグループ化したドキュメントの集成からの1つ以上の言及群を特定する。システムはこれらの2つのエンティティ間の関係を提案し得る。何故なら、関係推奨アルゴリズムがそのような関係を実証するドキュメントを読むことにユーザは関心があるかもしれないと予測するからであるユーザが、例えば、コネクタシンボル1608をクリックすることによりコネクタシンボル1608とやりとりした場合、ユーザインターフェイス1600は検索入力ツール1602内のクエリを変更することにより反応し得る。この更新されたクエリを第2の検索入力ツール1611に示す。ここでは、アイコン1614はジャック・ワンのためのERSを示し、ビジュアルアイコン1616はサリー・スミスのためのERSを示し、関係検索演算子1612は、これらの2つのエンティティ間の関係を実証する可能性が高いドキュメントを検索すべきであることを検索ランキングアルゴリズムに示す。

10

【0268】

20

図16のユーザインターフェイス1600は、この更新されたクエリに対応する結果を示す。結果1618はクエリからの2つのエンティティの言及1620の視覚的強調を示す。結果1622は、エンティティの言及の太字化等の視覚的強調も示す。肯定的及び否定的共参照入力インジケータ1624は、所与の言及が、第2の検索入力ツール1611のクエリにおけるERSアイコン1614の他の言及と共参照であるという共参照アルゴリズムの予測にユーザが同意又は不同意できるようにする。一実施形態では、カラーコーディング又は他のインジケータは、システムがどのエンティティ参照セットにどの言及をユーザが追加することを提案しているかをユーザが効率的に認識するのを助ける。例えば、サリーの言及は紫色であってもよく、ジャックの言及は緑色であってもよい。

【0269】

30

図16のユーザインターフェイス1600は、第2のコネクタシンボル1626を用いた追加の関係提案も示し得る。これにより、関連するエンティティのネットワークを反復的に探索することをユーザに奨励する。

【0270】

好ましい実施形態では、ユーザインターフェイスは、ユーザが1クリックだけ等の少数回のクリックで肯定的又は否定的共参照ラベルのバルク割り当てを入力できるようにする。例えば、ユーザが特定のジャック・ワンを調査している場合を考えてみる。サリー・スミス等のジャック・ワンに関連するエンティティについての関係クエリをユーザが開始すると、結果の全てはユーザの探索の対象とは異なるジャック・ワンを参照していることがある。例えば、ユーザがジャック・ワンとサリー・スミスとの間の提案された関係を見て、コネクタシンボル1608をクリックすると、ユーザは探求中であり、この関係が、対象人物と同じ名前(「ジャック・ワン」)の別の人物に関連することを発見し得る。その場合、バルクラベル機能によりユーザは全ての結果を1回の素早いアクションで破棄することができる。このような破棄は、クエリ内のERSと、ユーザに表示された結果中で言及されている様々な(例えば、トップ10の)ジャック・ワンとの間に否定的共参照ラベルを割り当てる。このような破棄アクションにより推奨された関係からサリー・スミスの言及が削除され得るため、ユーザはその関係のさらなる提案によって困惑することがない。

40

【0271】

上記のシステム、装置、方法、プロセス等は、特定の用途に好適なハードウェア、ソフトウェア又はこれらの任意の組み合わせで実現され得る。ハードウェアは汎用コンピュータ

50

及び/又は専用のコンピュータ装置を含み得る。これは、1つ以上のマイクロプロセッサ、マイクロコントローラ、埋め込み型マイクロコントローラ、プログラマブルデジタル信号プロセッサ又は他のプログラマブルデバイス若しくはプログラマブル回路と合わせて、内部及び/又は外部メモリにおいて実現することを含む。これに加えて又は代えて、1つ以上の特定用途向け集積回路、プログラマブルゲートアレイ、プログラマブルアレイロジックコンポーネント又は電子信号を処理するように構成され得る任意の他のデバイスを含み得る。上述のプロセス又は装置の実現は、C等の構造化プログラミング言語、C++等のオブジェクト指向プログラミング言語又は上述の装置のうちの1つで実行されるように記憶、コンパイル又は解釈され得る任意の他のハイレベル又はローレベルのプログラミング言語(アセンブリ言語、ハードウェア記述言語、データベースプログラミング言語及び技術を含む)に加えて、プロセッサ、プロセッサアーキテクチャの異種の組み合わせ又は異なるハードウェア及びソフトウェアの組み合わせを用いて作成されるコンピュータ実行可能コードを含み得ることがさらに理解されよう。別の態様では、上記の方法は、そのステップを行うシステムで実施されてもよく、様々なやり方でデバイスにわたって分散され得る。それと同時に、処理は、上述した様々なシステム等のデバイスにわたって分散されてもよいし、機能の全てを専用のスタンドアロンデバイス又は他のハードウェアに統合してもよい。別の態様では、上述のプロセスに関連するステップを行うための手段は、上述のハードウェア及び/又はソフトウェアのいずれかを含み得る。全てのそのような置換及び組み合わせは本開示の範囲内に含まれることを意図する。

10

【0272】

20

本明細書で開示の実施形態は、1つ以上のコンピュータ装置上で実行された場合に、ステップのいずれか及び/又は全てを行うコンピュータ実行可能コード又はコンピュータ使用可能コードを含むコンピュータプログラム製品を含み得る。コードは、コンピュータメモリに非一時的に記憶されてもよく、該コンピュータメモリはそこからプログラムが実行されるメモリ(例えば、プロセッサに関連するランダムアクセスメモリ)であってもよいし、ディスクドライブ、フラッシュメモリ又は任意の他の光学、電磁、磁気、赤外線若しくは他の装置又は装置の組み合わせ等の記憶装置であってもよい。別の態様では、上述のシステム及び方法のいずれも、コンピュータ実行可能コード及び/又はそれらの入力又は出力を搬送する任意の適切な伝送又は伝搬媒体で実施され得る。

【0273】

30

図全体にわたるフローチャート及びブロック図を含む、本明細書に記載及び図示される要素は、要素間の論理的境界を含意する。しかしながら、ソフトウェア又はハードウェアのエンジニアリング手法によれば、図示した要素及びその機能は、モノリシックなソフトウェア構造として、スタンドアロンソフトウェアモジュールとして又は外部ルーチン、コード、サービス等を用いるモジュールとして又はこれらの任意の組み合わせとして記憶されたプログラム命令を実行可能なプロセッサを有するコンピュータ実行可能媒体を介してマシン上で実施してもよく、そのような実施の全ては本開示の範囲内にある。このような機械の例としては、限定されないがパーソナルデジタルアシスタント、ラップトップ、パーソナルコンピュータ、携帯電話、他のハンドヘルドコンピュータ装置、医療装置、有線又は無線通信装置、トランスデューサ、チップ、計算機、衛星、タブレットPC、電子書籍、ガジェット、電子機器、人工知能を有する装置、コンピュータ装置、ネットワーク機器、サーバ、ルータ等を含み得る。さらに、フローチャート及びブロック図に示す要素又は他の論理コンポーネントは、プログラム命令を実行可能なマシン上で実施され得る。そのため、前述の図面及び説明は開示のシステムの機能的な側面を記載しているが、これらの機能的な側面を実施するためのソフトウェアの特定の配置は、明示的に記載されているか又は文脈から明確でない限りこれらの記載から推論すべきではない。同様に、先に特定及び説明した様々なステップは変化されてもよく、ステップの順序を本明細書で開示の技術の特定の用途に適合され得ることが理解されよう。そのような変更及び修正の全ては本開示の範囲内にあることを意図している。そのため、様々なステップの順序の図示及び/又は説明は、特定のアプリケーションによって要求されるか、明示的に記載されているか又

40

50

は文脈から明確である場合を除いてそれらのステップに対する実行の特定の順序が必要であると理解すべきでない。反対の明示的な指示がない限り、開示のステップは、本開示の範囲から逸脱することなく変更、補足、省略及び/又は再順序付けされ得る。

【0274】

本明細書に記載の実施の方法ステップは、異なる意味が明示的に提供されているか又は文脈から明確でない限り、下記の特許請求の範囲の特許性と一致して、係る方法ステップを実施させる任意の好適な方法を含むことを意図している。そのため、例えば、Xのステップを行うことは、遠隔ユーザ、遠隔処理リソース(例えば、サーバ又はクラウドコンピュータ)又はマシン等の別の当事者にXのステップを行わせるための任意の適切な方法を含む。同様に、ステップX、Y及びZを行うことは、そのようなステップの利益を得るためにそのような他の個人又はリソースの任意の組み合わせにステップX、Y及びZを行わせるか又は制御する任意の方法を含み得る。そのため、本明細書に記載の実施の方法ステップは、別の意味が明示的に提供されているか又は文脈から明確でない限り、下記の特許請求の範囲の特許性と一致して、1つ以上の他の当事者又はエンティティにステップを行わせる任意の好適な方法を含むことを意図している。そのような当事者又はエンティティは、任意の他の当事者又はエンティティの指揮又は管理下にある必要はなく、特定の法域内にいる必要はない。

10

【0275】

上述の方法及びシステムは一例として記載され、限定されるものではないことが理解されるであろう。多数の変更、追加、省略及び他の修正は当業者には明らかであろう。加えて、明細書及び図面における上記の方法ステップの順序又は提示は、特定の順序が明確に要求されるか又は文脈から明確でない限り、記載されたステップを行うためにこの順序が必要であることを意図していない。そのため、特定の実施形態を図示説明してきたが、本開示の精神及び範囲から逸脱することなく形態及び詳細に様々な変更及び修正が加えられることがあり、法律によって許容される最も広い意味で解釈されるべき下記の特許請求の範囲により定義される本発明の一部を形成することを意図していることは当業者には明らかであろう。

20

【0276】

本発明の一例にしたがって、本願は、さらに、以下の実施形態を提供する。

実施形態1: 複数のエンティティの間の関係と対話するためのコンピュータプログラムであって、当該コンピュータプログラムは、非一時的なコンピュータ読み取り可能な媒体の中に具現化されているコンピュータ実行可能なコードを含み、1つ又は複数のコンピューティングデバイスにおいて実行するときに、

30

ドキュメントの集合へのアクセスを提供するステップと、

第1のエンティティ及び第2のエンティティの各々を参照する複数のドキュメントの集合にテキストのスパンを表示することによって、前記第1のエンティティ及び前記第2のエンティティを識別するステップと、

ドキュメントの前記集合を処理して、前記第1のエンティティと前記第2のエンティティとの間の接続を具体化するドキュメントの前記集合のうちの1つ又は複数のドキュメントの中の証拠の節を識別するステップであって、前記証拠の節は、前記第1のエンティティの第1の言及及び前記第2のエンティティの第2の言及を含む、ステップと、

40

ユーザインターフェイスの中に前記第1のエンティティのための第1のアイコン、前記第2のエンティティのための第2のアイコン、及びコネクタシンボルを表示するステップであって、前記コネクタシンボルは、前記第1のアイコンを前記第2のアイコンに視覚的に結合するとともに、前記第1のエンティティと前記第2のエンティティとの間の前記接続を表す、ステップと、

前記コネクタシンボルとのユーザインタラクションにตอบสนองして、前記ユーザインターフェイスの中に前記証拠の節へのアクセスを提供するステップと、を実行する、コンピュータプログラム。

実施形態2: 前記証拠の節へのアクセスを提供するステップは、前記証拠の節の少なくとも

50

も一部を表示するステップを含む、実施形態1に記載のコンピュータプログラム。

実施形態3: 前記証拠の節は、前記第1のエンティティ及び前記第2のエンティティの同時生起の言及を含むドキュメントの前記集合のうちの前記ドキュメントの1つ又は複数からのテキストを含む、実施形態1に記載のコンピュータプログラム。

実施形態4: 前記テキストの中の前記同時生起の言及をハイライトするステップを実行するコードをさらに含む、実施形態3に記載のコンピュータプログラム。

実施形態5: 前記コネクタシンボルは、前記第1のアイコンを前記第2のアイコンに結合する線を含む、実施形態1に記載のコンピュータプログラム。

実施形態6: ドキュメントの前記集合を処理するステップは、ユーザの要求に応答してドキュメントの前記集合を処理するステップを含む、実施形態1に記載のコンピュータプログラム。

実施形態7: 複数のエンティティの間の同一指示に基づいて、ユーザに複数の候補接続を提示し、そして、前記ユーザによる複数の接続のうちの1つの選択に応答して、前記ユーザインターフェイスの中に表示のための前記接続を作成するステップを実行するコードをさらに含む、実施形態1に記載のコンピュータプログラム。

実施形態8: ユーザによる前記第1エンティティの選択を受信し、そして、前記第2エンティティへの前記接続を自動的に示唆するステップを実行するコードをさらに含む、実施形態1に記載のコンピュータプログラム。

実施形態9: ユーザによる前記第1のエンティティ及び前記第2のエンティティの選択を受信し、そして、ドキュメントの前記集合の中の前記接続を具体化するための応答証拠の節を検索するステップを実行するコードをさらに含む、実施形態1に記載のコンピュータプログラム。

実施形態10: 前記第1のエンティティと前記第2のエンティティとの間の前記接続を具体化するドキュメントの検索要求を含むように、前記ユーザインターフェイスの検索入力フォームフィールドを更新することによって、前記コネクタとの第2のユーザインタラクションに応答するステップを実行するコードをさらに含む、実施形態1に記載のコンピュータプログラム。

実施形態11: 方法であって、

第1のエンティティと第2のエンティティとの間の接続を具体化するドキュメントの集合の中からの証拠を特徴付ける関係記述を格納するステップと、

ユーザインターフェイスの中に前記第1のエンティティのための第1のアイコン及び前記第2のエンティティのための第2のアイコンを表示するステップと、

前記ユーザインターフェイスの中に前記第1のアイコンと前記第2のアイコンとを結合するコネクタシンボルを表示するステップであって、前記コネクタシンボルは、前記第1のアイコンと前記第2のアイコンとの間の前記接続を表し、前記コネクタシンボルは、ユーザインターフェイス要素を含み、ユーザは、前記ユーザインターフェイス要素を有効化して、証拠の前記関係記述にアクセスする、ステップと、

前記ユーザインターフェイスの中に前記証拠へのアクセスを提供することによって、前記コネクタシンボルとのユーザインタラクションに応答するステップと、を含む、方法。

実施形態12: 前記証拠の節へのアクセスを提供するステップは、前記関係記述の少なくとも一部を表示するステップを含む、実施形態11に記載の方法。

実施形態13: 前記関係記述は、前記第1のエンティティ及び前記第2のエンティティの同時生起する言及を含むドキュメントの前記集合のうちの前記ドキュメントの1つ又は複数からのテキストを含む、実施形態12に記載の方法。

実施形態14: 前記テキストの中の前記第1のエンティティ及び前記第2のエンティティの前記同時生起する言及をハイライトするステップをさらに含む、実施形態13に記載の方法。

実施形態15: 前記関係記述を格納するステップは、ユーザの要求に応答して、前記関係記述を作成するステップをさらに含む、実施形態11に記載の方法。

10

20

30

40

50

実施形態16: 複数のエンティティの言及の同時生起に基づいて、ユーザに複数の接続を提示し、そして、前記ユーザによる前記複数の接続のうちの1つの選択に応答して、前記関係記述を作成するステップをさらに含む、実施形態11に記載の方法。

実施形態17: ユーザによる前記第1エンティティの選択を受信し、そして、前記第2エンティティへの前記接続を自動的に示唆するステップをさらに含む、実施形態11に記載の方法。

実施形態18: ユーザによる前記第1のエンティティ及び前記第2のエンティティの選択を受信し、そして、ドキュメントの前記集合の中の前記接続を具体化するための1つ又は複数の応答証拠の節を検索するステップをさらに含む、実施形態11に記載の方法。

実施形態19: ユーザによる2つ又はそれ以上のエンティティの選択を受信し、そして、ドキュメントの前記集合の中の前記接続の具体化を検索するステップをさらに含む実施形態11に記載の方法。

10

実施形態20: 前記第1のエンティティと前記第2のエンティティとの間の前記接続を具体化するドキュメントの検索要求を含むように、前記ユーザインターフェイスの検索入力フォームフィールドを更新することによって、前記コネクタとの第2のユーザインタラクションに응答するステップをさらに含む、実施形態11に記載の方法。

実施形態21: システムであって、

第1のエンティティ及び第2のエンティティの識別情報を受信し、そして、ドキュメントの集合を処理して、前記第1のエンティティと前記第2のエンティティとの間の接続を具体化する証拠の節を識別する、ように構成される検索エンジンであって、前記証拠の節は、ドキュメントの前記集合の少なくとも1つのドキュメントの中の前記第1のエンティティ及び前記第2のエンティティの同時生起の言及を含む、検索エンジンと、プロセッサ、メモリ、ディスプレイ、及びユーザ入力デバイスを有するコンピュータであって、前記プロセッサは、コンピュータ実行可能なコードによって構成され、前記第1のエンティティ及び前記第2のエンティティを識別するステップと、前記検索エンジンに前記第1のエンティティ及び前記第2のエンティティを提示するステップと、前記検索エンジンからの結果に응答して、ユーザインターフェイスの中に、前記第1のエンティティの第1のアイコン、前記第2のエンティティの第2のアイコン、及びコネクタシンボルを表示するステップと、を実行し、前記コネクタシンボルは、前記第1のアイコンを前記第2のアイコンに視覚的に結合するとともに、前記第1のエンティティと前記第2のエンティティとの間の前記接続を表し、前記コネクタシンボルは、前記ユーザインターフェイスを介して動作可能であり、前記ユーザインターフェイスの中の前記証拠の節へのアクセスを提供する、コンピュータと、を含む、システム。

20

30

実施形態22: 前記検索エンジンは、前記コンピュータにおいてローカルに実行される、実施形態21に記載のシステム。

実施形態23: 前記検索エンジンは、データネットワークを通じてコンピュータによってアクセス可能なりモート検索エンジンである、実施形態21に記載のシステム。

実施形態24: 複数のエンティティの間の関係と対話するためのコンピュータプログラムであって、当該コンピュータプログラムは、非一時的なコンピュータ読み取り可能な媒体の中に具現化されているコンピュータ実行可能なコードを含み、1つ又は複数のコンピューティングデバイスにおいて実行するときに、

40

検索エンジンのためのユーザインターフェイスの検索入力フォームフィールドを介してユーザからキーワード検索を受信するステップと、

前記キーワード検索に同一指示アルゴリズムを適用して、前記ユーザが意図しているとともに、1つ又は複数のエンティティ関連のグループの言及によって特徴付けられるエンティティを予測するステップと、

前記ユーザインターフェイスの中で前記ユーザに第1の検索結果を提示するステップであって、前記第1の検索結果は、第1の複数のエンティティタグ及びドキュメントの集合のうちの複数のドキュメントを含み、前記複数のドキュメントのうちの各々のドキュメン

50

トは、前記1つ又は複数のエンティティー関連のグループの言及のうちの少なくとも1つのうちで前記検索エンジンが配置する言及を含み、前記複数のエンティティータグのうちの各々のエンティティータグは、前記言及のうちの少なくとも1つに対応する、ステップと、

前記第1の複数のエンティティータグのうちで第1のエンティティー関連のグループの言及によって特徴付けられる第1のエンティティーのための第1のエンティティータグの前記ユーザからの第1の選択を受信するステップであって、前記第1の選択は、前記キーワード検索が前記第1のエンティティーを参照することを意図していたというユーザの確認を示す、ステップと、

前記検索入力フォームフィールドに前記第1のエンティティータグを追加するステップと、

—
関連性アルゴリズムを適用して、1つ又は複数の他のエンティティー関連のグループの言及によって特徴付けられる1つ又は複数の他のエンティティーと前記第1のエンティティーとの間の関係の証拠を提供する第2の複数のドキュメントの中のテキストを識別するステップと、

各々が前記ユーザインターフェイスの中の前記1つ又は複数の他のエンティティーのうちの1つに対応する第2の複数のエンティティータグを提示するステップと、

前記第2の複数のエンティティータグのうちの第2のエンティティータグの前記ユーザから第2の選択を受信するステップと、

前記第1のエンティティーと前記第2のエンティティーとの間の関係を具体化する前記第2の複数のドキュメントのうちの1つ又は複数を含む第2の検索結果を提示するステップと、を実行する、

コンピュータプログラム。

実施形態25: 前記ユーザインターフェイスの中に、前記第1のエンティティーの第1のアイコン、前記第2のエンティティーの第2のアイコン、及び前記第2のアイコンに前記第1のアイコンを視覚的に結合するコネクタとして、前記関係をグラフィカルに表示するステップを実行するコードをさらに含む、実施形態24に記載のコンピュータプログラム。

実施形態26: 関係オペレータによって前記第2のエンティティータグに結合されている前記第1のエンティティータグとして、前記検索入力フォームフィールドの中の前記関係を表示するステップを実行するコードをさらに含む、実施形態24に記載のコンピュータプログラム。

実施形態27: 前記関係オペレータは、シンボル"< >"を含む、実施形態26に記載のコンピュータプログラム。

実施形態28: 前記ドキュメントの集合は、前記キーワード検索が受信されるユーザデバイスにおいてローカルにホストされるドキュメントを含む、実施形態24に記載のコンピュータプログラム。

実施形態29: 前記ドキュメントの集合は、広域ネットワークにおいて配信されるドキュメントを含む、実施形態24に記載のコンピュータプログラム。

実施形態30: 前記複数のドキュメントは、リモートクラウド記憶装置においてホストされているドキュメントを含む、実施形態24に記載のコンピュータプログラム。

実施形態31: 前記第2の検索結果を提示するステップは、グラフの中の端部によって、各々が前記第1のエンティティーに接続される複数のエンティティータグをグラフィカルに提示するステップを含む、実施形態24に記載のコンピュータプログラム。

実施形態32: 方法であって、

検索エンジンのためのユーザインターフェイスの検索入力フォームフィールドにおいてユーザからキーワード検索を受信するステップと、

前記ユーザが意図している1つ又は複数のエンティティーを予測するステップと、

前記ユーザインターフェイスの中で前記ユーザに第1の検索結果を提示するステップであって、前記第1の検索結果は、第1の複数のエンティティータグ及び複数のドキュメントを含み、前記複数のドキュメントのうちの各々のドキュメントは、前記1つ又は複数のエ

10

20

30

40

50

ンティティーのうちの少なくとも1つの前記検索エンジンが配置する言及を含み、前記複数のエンティティータグのうちの各々のエンティティータグは、前記言及のうちの少なくとも1つに対応する、ステップと、

前記第1の複数のエンティティータグのうちの第1のエンティティーのための第1のエンティティータグの前記ユーザからの第1の選択を受信するステップであって、前記第1の選択は、前記キーワード検索が前記第1のエンティティーを参照することを意図していたというユーザの確認を示す、ステップと、

前記検索入力フォームフィールドに前記第1のエンティティータグを追加するステップと、

1つ又は複数の他のエンティティーと前記第1のエンティティーとの間の関係の証拠を提供する第2の複数のドキュメントの中のテキストを識別するステップと、

前記ユーザインターフェイスの中の前記1つ又は複数の他のエンティティーのための第2の複数のエンティティータグを提示するステップと、

前記第2の複数のエンティティータグのうちの第2のエンティティーのための第2のエンティティータグの前記ユーザから第2の選択を受信するステップと、

前記第1のエンティティーと前記第2のエンティティーとの間の関係を具体化する前記第2の複数のドキュメントのうちの1つ又は複数を含む第2の検索結果を提示するステップと、を含む、

方法。

実施形態33: 前記ユーザインターフェイスの中に、前記第1のエンティティーの第1のアイコン、前記第2のエンティティーの第2のアイコン、及び前記第2のアイコンに前記第1のアイコンを視覚的に結合するコネクタとして、前記関係をグラフィカルに表示するステップをさらに含む、実施形態32に記載の方法。

実施形態34: 関係オペレータによって前記第2のエンティティータグに結合されている前記第1のエンティティータグとして、前記検索入力フォームフィールドの中の前記関係を表示するステップをさらに含む、実施形態32に記載の方法。

実施形態35: 前記関係オペレータは、シンボル"< >"を含む、実施形態34に記載の方法。

実施形態36: 前記キーワード検索において前記ユーザが意図している1つ又は複数のエンティティーを予測するステップは、前記キーワード検索に同一指示アルゴリズムを適用して、前記ユーザが意図している1つ又は複数のエンティティーを予測するステップを含む、実施形態32に記載の方法。

実施形態37: 1つ又は複数の他のエンティティーと前記第1のエンティティーとの間の関係の証拠を提供する前記複数のドキュメントの中のテキストを識別するステップは、関連性アルゴリズムを適用して、1つ又は複数の他のエンティティーと前記第1のエンティティーとの間の関係の証拠を提供する前記複数のドキュメントの中のテキストを識別するステップを含む、実施形態32に記載の方法。

実施形態38: 前記複数のドキュメントは、前記キーワード検索が受信されるユーザデバイスにおいてローカルにホストされるとともにドキュメントの集合から得られるドキュメントを含む、実施形態32に記載の方法。

実施形態39: 前記複数のドキュメントは、広域ネットワークにおいて配信されるとともにドキュメントの集合から得られるドキュメントを含む、実施形態32に記載の方法。

実施形態40: 前記複数のドキュメントは、リモートクラウド記憶装置においてホストされているとともにドキュメントの集合から得られるドキュメントを含む、実施形態32に記載の方法。

実施形態41: 前記第1のエンティティーと前記第2のエンティティーとの間の関係を特徴付ける関係記述を格納するステップをさらに含む、実施形態32に記載の方法。

実施形態42: 前記第1のエンティティーの第1のアイコン、前記第2のエンティティーの第2のアイコン、及び前記関係を示すために前記第1のアイコンを前記第2のアイコンに視覚的に結合するコネクタをグラフィカルに表示するステップをさらに含む、前記コネク

10

20

30

40

50

タは、前記第2の複数のドキュメントの中の前記関係の証拠へのユーザによるアクセスを提供するユーザインターフェイス要素を含む、実施形態32に記載の方法。

実施形態43: システムであって、
検索エンジンと、

データネットワークと、

前記データネットワークを通じて前記検索エンジンに結合されるコンピューティングデバイスであって、前記コンピューティングデバイスは、プロセッサ、メモリ、及びディスプレイを含み、前記メモリは、前記プロセッサによって実行可能であるコードを格納し、前記プロセッサは、ユーザインターフェイスの検索入力フォームフィールドにおいてユーザからキーワード検索を受信するステップと、前記ユーザが意図している1つ又は複数のエンティティを予測するステップと、前記ユーザインターフェイスの中で前記ユーザに第1の検索結果を提示するステップであって、前記第1の検索結果は、第1の複数のエンティティタグ及び複数のドキュメントを含み、前記複数のドキュメントのうちの各々のドキュメントは、前記1つ又は複数のエンティティのうちの少なくとも1つのエンティティの前記検索エンジンが配置する言及を含み、前記複数のエンティティタグのうちの各々のエンティティタグは、前記言及のうちの少なくとも1つに対応する、ステップと、前記第1の複数のエンティティタグのうちの第1のエンティティのための第1のエンティティタグの前記ユーザからの第1の選択を受信するステップであって、前記第1の選択は、前記キーワード検索が前記第1のエンティティを参照することを意図していたというユーザの確認を示す、ステップと、前記検索入力フォームフィールドに前記第1のエンティティタグを追加するステップと、1つ又は複数の他のエンティティと前記第1のエンティティとの間の関係の証拠を提供する第2の複数のドキュメントの中のテキストを識別するステップと、前記ユーザインターフェイスの中の前記1つ又は複数の他のエンティティのための第2の複数のエンティティタグを提示するステップと、前記第2の複数のエンティティタグのうちの第2のエンティティのための第2のエンティティタグの前記ユーザから第2の選択を受信するステップと、前記第1のエンティティと前記第2のエンティティとの間の関係を具体化する前記第2の複数のドキュメントのうちの1つ又は複数を含む第2の検索結果を提示するステップと、を実行する、システム。

実施形態44: 複数のエンティティの間の関係と対話するためのコンピュータプログラムであって、当該コンピュータプログラムは、非一時的なコンピュータ読み取り可能な媒体の中に具現化されているコンピュータ実行可能なコードを含み、1つ又は複数のコンピューティングデバイスにおいて実行するときに、

検索エンジンのためのユーザインターフェイスを表示するステップであって、前記ユーザインターフェイスは、検索入力フォームフィールドを含む、ステップと、

前記検索入力フォームフィールドへのユーザ入力を受信するステップと、

前記ユーザ入力を構文解析して、第1のエンティティを識別する第1のテキストストリング、第2のエンティティを識別する第2のテキストストリング、及び前記第1のテキストストリングと前記第2のテキストストリングとの間のオペレータを識別するステップであって、前記オペレータは、前記第1のエンティティと前記第2のエンティティとの間の関係の証拠を検索するための前記検索エンジンへの要求を指定する、ステップと、前記オペレータに応答して、ドキュメントの集合のうちの1つ又は複数のドキュメントを検索するステップであって、前記ドキュメントの集合は、同一指示アルゴリズムが、著者が前記第1のエンティティを参照することを意図していた言及であると予測する言及、及び、同一指示アルゴリズムが、前記著者が前記第2のエンティティを参照することを意図していた言及であると予測する他の言及を含むとともに、関連性アルゴリズムが、前記第1のエンティティと前記第2のエンティティとの間の関係の証拠を提供すると予測するテキストを提供する、ステップと、

前記ユーザインターフェイスの中で前記1つ又は複数のドキュメントを提示するステップと、を含む、

10

20

30

40

50

方法。

10

20

30

40

50