

(19)日本国特許庁(JP)

(12)特許公報(B2)

(11)特許番号
特許第7658681号
(P7658681)

(45)発行日 令和7年4月8日(2025.4.8)

(24)登録日 令和7年3月31日(2025.3.31)

(51)国際特許分類

F I

G 0 6 F 16/903 (2019.01)

G 0 6 F 16/903

請求項の数 19 (全23頁)

(21)出願番号	特願2022-508750(P2022-508750)	(73)特許権者	390009531
(86)(22)出願日	令和2年8月21日(2020.8.21)		インターナショナル・ビジネス・マシー
(65)公表番号	特表2022-546228(P2022-546228		ンズ・コーポレーション
	A)		INTERNATIONAL BUSI
(43)公表日	令和4年11月4日(2022.11.4)		NESS MACHINES CORPO
(86)国際出願番号	PCT/IB2020/057852		RATION
(87)国際公開番号	WO2021/044246		アメリカ合衆国10504 ニューヨー
(87)国際公開日	令和3年3月11日(2021.3.11)		ク州 アーモンク ニュー オーチャード
審査請求日	令和5年1月20日(2023.1.20)		ロード
(31)優先権主張番号	16/559,351		New Orchard Road, A
(32)優先日	令和1年9月3日(2019.9.3)		rmonk, New York 105
(33)優先権主張国・地域又は機関	米国(US)	(74)代理人	04, United States of
			America
			100112690
			弁理士 太佐 種一
			最終頁に続く

(54)【発明の名称】 構造化及び非構造化されたデータを使用する照会の解決

(57)【特許請求の範囲】

【請求項1】

コンピュータ実装方法であって、
自然言語フォーマットで質問を受領すること、
前記質問から複数のサブクエリーを導出すること、
前記複数のサブクエリーの第1のサブクエリーがキュレーテッド・ナレッジ・ベースを使用して解決できないことを判断すること、
受領した前記質問から導出された前記第1のサブクエリーに基づいて、非構造化テキストにより生成された検索インデックスを使用して、不完全なタプルを有する前記第1のサブクエリーを解決すること、
完全なタプルを有する、解決された前記第1のサブクエリーを追加して、前記キュレーテッド・ナレッジ・ベースを適応させること
を含む、コンピュータ実装方法。

【請求項2】

前記質問が、質問・回答（QA）アプリケーションにより受領される、請求項1に記載のコンピュータ実装方法。

【請求項3】

前記複数のサブクエリー各々が、前記質問の自然言語フォーマットを論理形式にパースすることにより導出される、請求項1または2に記載のコンピュータ実装方法。

【請求項4】

前記不完全なタプルが、不完全な主語／述語／目的語のトリプルを含み、前記完全なタプルは、完全な主語／述語／目的語のトリプルを含む、請求項 1～3 のいずれか 1 項に記載のコンピュータ実装方法。

【請求項 5】

前記非構造化テキストは、テキスト・コーパスを含み、前記キュレーテッド・ナレッジ・ベースは、前記テキスト・コーパスから生成されたものである、請求項 1～4 のいずれか 1 項に記載のコンピュータ実装方法。

【請求項 6】

クエリー・リラクゼーションが、前記検索インデックスのクエリー内に含ませるための追加の語を生成するために使用される、請求項 1～5 のいずれか 1 項に記載のコンピュータ実装方法。

【請求項 7】

前記第 1 のサブクエリーを解決することが、前記検索インデックスに照会し、照会の結果を受領することを含む、請求項 1～6 のいずれか 1 項に記載のコンピュータ実装方法。

【請求項 8】

前記第 1 のサブクエリーを解決することが、前記検索インデックスの照会の結果を使用して新たなトリプルを生成することを含む、請求項 7 に記載のコンピュータ実装方法。

【請求項 9】

前記第 1 のサブクエリーを解決することが、前記新たなトリプルを使用して前記第 1 のサブクエリーを完全化させることを含む、請求項 8 に記載のコンピュータ実装方法。

【請求項 10】

さらに前記複数のサブクエリーの第 2 のサブクエリーを、適応させた前記キュレーテッド・ナレッジ・ベースを使用することにより解決することを含む、請求項 1～9 のいずれか 1 項に記載のコンピュータ実装方法。

【請求項 11】

解決された前記第 1 のサブクエリー及び解決された前記第 2 のサブクエリーが前記受領した質問に対する回答を生成するために使用される、請求項 10 に記載のコンピュータ実装方法。

【請求項 12】

前記不完全なタプルは、前記キュレーテッド・ナレッジ・ベースとの比較により完全化することができない不完全なタプルである、請求項 1～11 のいずれか 1 項に記載のコンピュータ実装方法。

【請求項 13】

請求項 1～12 のいずれか 1 項に記載のコンピュータ実装方法をコンピュータ実行可能なコンピュータ・プログラム。

【請求項 14】

システムであって、
プロセッサと、

前記プロセッサに一体化され、前記プロセッサにより実行可能であるか、又は前記プロセッサに一体化された論理であって、前記論理が、請求項 1～12 のいずれか 1 項に記載のコンピュータ実装方法を実行するように構成される論理と

を含むシステム。

【請求項 15】

コンピュータ実装方法であって、

自然言語フォーマットで質問を受領すること、

前記質問から第 1 のサブクエリー及び第 2 のサブクエリーを導出すること、

前記第 1 のサブクエリーが、非構造化テキストに基づき生成されたキュレーテッド・ナレッジ・ベースを使用して解決できないことを判断すること、

クエリー・リラクゼーションを使用して前記第 1 のサブクエリーのための追加の検索語を生成すること、

10

20

30

40

50

受領した前記質問から導出された前記第1のサブクエリー及び追加の前記検索語を使用して、前記非構造化テキストにより生成された検索インデックスを照会すること、
前記検索インデックスの照会に応答して検索結果を受領すること、
前記検索結果に基づいて新たな主語／述語／目的語のトリプルを生成すること、
前記新たな主語／述語／目的語のトリプルを使用して、不完全な主語／述語／目的語のトリプルを有する前記第1のサブクエリーを完全化すること、
完全な主語／述語／目的語のトリプルを有する、完全化した前記第1のサブクエリーを追加して、前記キュレーテッド・ナレッジ・ベースをアップデートすること、
前記第2のサブクエリーがアップデートされた前記キュレーテッド・ナレッジ・ベースを使用して解決できることを判断すること、及び
アップデートされた前記キュレーテッド・ナレッジ・ベースを使用して前記第2のサブクエリーを完全化させること
を含むコンピュータ実装方法。

10

【請求項16】

前記非構造化テキストは、テキスト・コーパスを含み、前記キュレーテッド・ナレッジ・ベースは、前記テキスト・コーパスから生成されたものである、請求項15に記載のコンピュータ実装方法。

【請求項17】

前記不完全なトリプルは、前記キュレーテッド・ナレッジ・ベースとの比較により完全化することができない不完全なトリプルである、請求項15または請求項16に記載のコンピュータ実装方法。

20

【請求項18】

コンピュータ実装方法であって、
自然言語入力を主語／述語／目的語（SPO）のトリプルに落とし込む自然言語のパースで、質問に対応する前記自然言語入力を受領すること、
キュレーテッド・ナレッジ・ベースに対するセマンティック的な関連性を考慮して、不完全なSPOトリプルを有する前記SPOトリプルの第1のサブセットをフリーテキストに対する照会モジュールに、かつ前記第1のサブセットとは異なるSPOトリプルの第2のサブセットを、キュレーテッド・ナレッジ・ベースに対する照会モジュールに向けること、

30

前記フリーテキストに対する照会モジュールにより、前記SPOトリプルの前記第1のサブセットにおいて興味のある断片を識別するためにインデックス付けされたテキストデータにアクセスすること、

興味のある前記断片から前記SPOトリプルの前記第1のサブセットに対応するSPOトリプルを抽出すること、

抽出されたSPOトリプルをそれらの前記SPOトリプルの前記第1のサブセットに対するセマンティック的な関連性にしたがってランク付けすること、

ランク付けされ、抽出されたSPOトリプルの上位から所定数のSPOトリプルを選択すること、及び

40

選択された前記所定数のSPOトリプルを、前記SPOトリプルの第2のサブセットに一体化して、前記質問に対する回答を形成すること、

を含み、前記コンピュータ実装方法は、さらに、

前記不完全なSPOトリプルを有する前記第1のサブセットを、興味のある前記断片に基づいて完全化し、完全なSPOトリプルを有する、完全化された前記第1のサブセットを前記キュレーテッド・ナレッジ・ベースに追加することをさらに含む、コンピュータ実装方法。

【請求項19】

前記一体化することが、選択された所定数の前記SPOトリプルと、前記SPOトリプルの前記第2のサブセットとの間の前記自然言語のパースにより生成された依存性の観点において実行される、請求項18に記載のコンピュータ実装方法。

50

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、概ねデータ分析に関し、より具体的にはインデックスされた非構造化データ及びキュレーテッド・ナレッジ・ベースの組み合わせを使用する照会の解決に関する。

【背景技術】

【0002】

複数のナレッジ・ベース（K B s）は、データを中心に据えるクエスチョン・アンサー（Q A）アプリケーションといったアプリケーション内で頻繁に使用される。このようなナレッジ・ベースは、キュレートされた非構造化テキストを使用して生成され、高精度の知識に基づいて正確な答えを提供する。しかしながら、ナレッジ・ベースは、正確な結果を情報の特定のサブセットについて提供するが、これらは特定のサブセット外の照会に対して答えを提供することはできない。

【発明の概要】

【0003】

1つの特徴によるコンピュータ実装方法は、自然言語フォーマットで質問を受領すること、質問から複数のサブクエリーを導出すること、複数のサブクエリーの第1のサブクエリーがキュレーテッド・ナレッジ・ベースを使用して解決できないことを判断すること、第1のサブクエリーを、非構造化テキストを使用して解決すること、及び解決された第1のサブクエリーを使用してキュレーテッド・ナレッジ・ベースを適応させることを含む。

【0004】

もう1つの特徴によれば、構造化及び非構造化されたデータを使用して照会を解決するためのコンピュータ・プログラム製品は、それに実体化されたプログラム命令を有するコンピュータ可読な記録媒体を含み、コンピュータ可読な記録媒体は、それ自体では過渡的な信号ではなく、プログラム命令は、プロセッサにより実行可能であって、プロセッサをして、プロセッサにより、自然言語フォーマットでの質問を受領すること、プロセッサにより、質問から複数のサブクエリーを導出すること、プロセッサにより、複数のサブクエリーの第1のサブクエリーがキュレーテッド・ナレッジ・ベースを使用して解決できないことを判断すること、プロセッサにより、第1のサブクエリーを、非構造化テキストを使用して解決すること、及びプロセッサにより、解決された第1のサブクエリーを使用してキュレーテッド・ナレッジ・ベースを適応させることを実行させる。

【0005】

もう1つの特徴によれば、システムは、プロセッサ及びプロセッサに論理的に一体化され、プロセッサにより実行可能又はプロセッサに一体化されるか、又は実行可能な論理を含み、この論理は、自然言語フォーマットで質問を受領し、質問から複数のサブクエリーを導出し、複数のサブクエリーの第1のサブクエリーがキュレーテッド・ナレッジ・ベースを使用して解決できないことを判断し、第1のサブクエリーを、非構造化テキストを使用して解決し、及び解決された第1のサブクエリーを使用してキュレーテッド・ナレッジ・ベースを適応させるように構成される。

【0006】

もう1つの特徴によれば、コンピュータ実装方法は、自然言語フォーマットで質問を受領すること、質問から第1のサブクエリー及び第2のサブクエリーを導出すること、複数のサブクエリーの第1のサブクエリーがキュレーテッド・ナレッジ・ベースを使用して解決できないことを判断すること、クエリー・リラクゼーションを使用して、第1のサブクエリーに対する追加の検索語を生成すること、第1のサブクエリー及び追加の検索語を使用して、非構造化テキストについて検索インデックスを照会すること、検索インデックスの照会に応答して検索結果を受領すること、検索結果に基づいて新たな主語／述語／目的語のトリプルを生成すること、新たな主語／述語／オブジェクトのトリプルを使用して第1のサブクエリーを完全化すること、完全化した第1のサブクエリーを含むように、キュレーテッド・ナレッジ・ベースをアップデートすること、第2のクエリーがアップデート

されたナレッジ・ベースを使用して解決できることを判断すること、及びアップデートしたナレッジ・ベースを使用して第2のサブクエリーを完全化することを含む。

【0007】

さらにもう1つの特徴によれば、オンデマンドな抽出及び一体化を実装するためのコンピュータ・プログラム製品は、それに実体化されたプログラム命令を有するコンピュータ可読な記録媒体を含み、このコンピュータ可読な記録媒体は、それ自体が過渡的ではなく、かつプログラム命令は、プロセッサにより実行可能であり、プロセッサをして、プロセッサにより自然言語フォーマットで質問を受領すること、プロセッサにより、質問から複数のサブクエリーを導出すること、プロセッサにより、複数のサブクエリーの第1のサブクエリーがキュレーテッド・ナレッジ・ベースを使用して解決できないことを判断すること、プロセッサにより、クエリー・リラクゼーションを使用して、第1のサブクエリーに対して追加の検索語を生成すること、プロセッサにより、第1のサブクエリー及び追加の検索語を使用して、非構造化テキストについて検索インデックスを照会すること、プロセッサにより、検索インデックスの照会に回答して検索結果を受領すること、プロセッサにより、検索結果に基づいて新たな主語／述語／目的語のトリプルを生成すること、プロセッサにより、新たな主語／述語／目的語のトリプルを使用して第1のサブクエリーを完全化すること、プロセッサにより、完全化した第1のサブクエリーを含むようにキュレーテッド・ナレッジ・ベースをアップデートすること、プロセッサにより、第2のクエリーがアップデートされたナレッジ・ベースを使用して解決できることを判断すること、及びプロセッサにより、アップデートしたキュレーテッド・ナレッジ・ベースを使用して第2のサブクエリーを完全化する方法を実行させる。

【0008】

本発明の他の特徴及び実施形態は、後述する詳細な説明から明らかとなるであろうが、ここで、図面と組み合わせられる場合、図面は、本発明の原理の実施例の目的において図示する。

【図面の簡単な説明】

【0009】

【図1】図1は、本発明の1実施形態によるネットワーク・アーキテクチャを示す。

【図2】図2は、本発明の1実施形態による図1のサーバ又はクライアント又はそれら両方に伴う代表的なハードウェア環境を示す。

【図3】図3は、本発明の1実施形態による構造化及び非構造化データを使用した照会解決のための方法を示す。

【図4】図4は、本発明の1実施形態によるオンデマンド抽出及び一体化を実装するための方法を示す。

【図5】図5は、本発明の1実施形態による非構造化及び構造化データを一体化する、例示的な質問・回答（QA）システムを示す。

【発明を実施するための形態】

【0010】

以下の説明は、本発明の一般的な原理を例示する目的でなされ、本明細書の請求項の発明的概念の限定を意味しない。さらに本明細書で説明する特定の特徵は、種々の可能な組み合わせ及び置換のそれぞれにおいて、他の説明した特徴と組み合わせにおいて使用することができる。

【0011】

本明細書で特に他に規定されない限り、すべての用語は、当業者により理解されるか又は辞書、論文において規定されるものなど又はこれらの組み合わせに加え、明細書から暗示される意味を含む、それらの最も広く、かつ可能な解釈を与えるものである。

【0012】

本明細書及び添付する請求の範囲で使用した用語は、特定の実施例のみを説明するための目的であり、本発明を限定することを意図しないことはまた明記されるべきである。本明細書で使用されるように、単数形式“a”、“an”、及び“the”は、文脈が明確にそれ

10

20

30

40

50

以外を示さない限り、同様に複数の形態を含むことを意図する。さらに、用語“含む”及び“含んでいる”は、本明細書において使用される場合、言及された特徴、整数、ステップ、操作、要素、及びコンポーネント又はそれらの組み合わせの存在を特定するが、1つ又はそれ以上の他の特徴、整数、ステップ、操作、要素、コンポーネント、及びそれらのグループ又はそれらの組み合わせを除外しない。

【0013】

後述する説明は、構造化及び非構造化データを使用した照会の解決のための、システム、方法及びコンピュータ・プログラム製品のいくつかの好ましい実施形態を開示する。

【0014】

1つの一般的な実施形態においては、コンピュータ実装方法は、自然言語フォーマットで質問を受領すること、質問から複数のサブクエリーを導出すること、複数のサブクエリーの第1のサブクエリーがキュレーテッド・ナレッジ・ベースを使用して解決できないことを判断すること、第1のサブクエリーを、非構造化テキストを使用して解決すること、及び解決された第1のサブクエリーを使用してキュレーテッド・ナレッジ・ベースを適応させることを含む。

【0015】

もう1つの一般的な実施形態においては、構造化及び非構造化されたデータを使用して照会を解決するためのコンピュータ・プログラム製品は、それに実体化されたプログラム命令を有するコンピュータ可読な記録媒体を含み、コンピュータ可読な記録媒体は、それ自体では過渡的な信号ではなく、プログラム命令は、プロセッサにより実行可能であり、プロセッサをして、プロセッサにより、自然言語フォーマットでの質問を受領すること、プロセッサにより、質問から複数のサブクエリーを導出すること、プロセッサにより、複数のサブクエリーの第1のサブクエリーがキュレーテッド・ナレッジ・ベースを使用して解決できないことを判断すること、プロセッサにより、第1のサブクエリーを、非構造化テキストを使用して解決すること、及びプロセッサにより、解決された第1のサブクエリーを使用してキュレーテッド・ナレッジ・ベースを適応させることを実行させる。

【0016】

もう1つの一般的な特徴においては、システムは、プロセッサ及びプロセッサに論理的に一体化され、プロセッサにより実行可能又はプロセッサに一体化されてプロセッサにより実行可能な論理を含み、この論理は、自然言語フォーマットで質問を受領し、質問から複数のサブクエリーを導出し、複数のサブクエリーの第1のサブクエリーがキュレーテッド・ナレッジ・ベースを使用して解決できないことを判断し、第1のサブクエリーを、非構造化テキストを使用して解決し、及び解決された第1のサブクエリーを使用してキュレーテッド・ナレッジ・ベースを適応させるように構成される。

【0017】

もう1つの一般的な実施形態においては、コンピュータ実装方法は、自然言語フォーマットで質問を受領すること、質問から第1のサブクエリー及び第2のサブクエリーを導出すること、第1のサブクエリーがキュレーテッド・ナレッジ・ベースを使用して解決できないことを判断すること、クエリー・リラクゼーションを使用して、第1のサブクエリーに対して追加の検索語を生成すること、第1のサブクエリー及び追加の検索語を使用して、非構造化テキストについて検索インデックスを照会すること、検索インデックスの照会に応答して検索結果を受領すること、検索結果に基づいて新たな主語／述語／目的語のトリプルを生成すること、新たな主語／述語／目的語のトリプルを使用して第1のサブクエリーを完全化すること、完全化した第1のサブクエリーを含むようにキュレーテッド・ナレッジ・ベースをアップデートすること、第2のクエリーがアップデートされたキュレーテッド・ナレッジ・ベースを使用して解決できることを判断すること、及びアップデートしたキュレーテッド・ナレッジ・ベースを使用して第2のサブクエリーを完全化することを含む。

【0018】

もう1つの一般的な実施形態においては、オンデマンドな抽出及び一体化を実装するた

10

20

30

40

50

めのコンピュータ・プログラム製品は、それに実体化されたプログラム命令を有するコンピュータ可読な記録媒体を含み、このコンピュータ可読な記録媒体は、それ自体が過渡的ではなく、かつプログラム命令は、プロセッサにより実行可能であり、プロセッサをして、プロセッサにより自然言語フォーマットで質問を受領すること、プロセッサにより、質問から第1のサブクエリー及び第2のサブクエリーを導出すること、プロセッサにより、第1のサブクエリーがキュレーテッド・ナレッジ・ベースを使用して解決できないことを判断すること、プロセッサにより、クエリー・リラクゼーションを使用して、第1のサブクエリーに対して追加の検索語を生成すること、プロセッサにより、第1のサブクエリー及び追加の検索語を使用して、非構造化テキストについて検索インデックスを照会すること、プロセッサにより、検索インデックスの照会に応答して検索結果を受領すること、プロセッサにより、検索結果に基づいて新たな主語／述語／オブジェクトの3つを生成すること、プロセッサにより、新たな主語／述語／オブジェクトの3つを使用して第1のサブクエリーを完全化すること、プロセッサにより、完全化した第1のサブクエリーを含むようにキュレーテッド・ナレッジ・ベースをアップデートすること、プロセッサにより、第2のクエリーがアップデートされたキュレーテッド・ナレッジ・ベースを使用して解決できることを判断すること、及びプロセッサにより、アップデートしたキュレーテッド・ナレッジ・ベースを使用して第2のサブクエリーを完全化することを実行させる。

【0019】

図1は、1実施形態におけるアーキテクチャ100を示す。図1に示すように、第1のリモート・ネットワーク104及び第2のリモート・ネットワーク106を含む、複数のリモート・ネットワーク102が提供される。ゲートウェイ101は、リモート・ネットワーク102と、近接ネットワーク108との間を結合することができる。本アーキテクチャ100のコンテキストにおいて、ネットワーク104、106は、LAN、インターネットといったWAN、公衆切替電話ネットワーク（PSTN）、内部電話ネットワークなどを含む如何なる形態でも取ることができるが、これらに限定されることはない。

【0020】

使用において、ゲートウェイ101は、リモート・ネットワーク102から近接ネットワーク108への入口ポイントを提供する。そのようにして、ゲートウェイ101は、ルータとして機能することができ、ルータは、ゲートウェイ101に到着した所与のデータ・パケットを管理することができ、スイッチは、所与のパケットのためのゲートウェイ101の出入りにおける実際の経路を提供する。

【0021】

さらに含まれるものは、近接ネットワーク108に接続された、少なくとも1つのデータ・サーバ114であり、これはゲートウェイ101を介して、複数のリモート・ネットワーク102からアクセス可能である。データ・サーバ（複数でも良い）114は、如何なるタイプのコンピューティング・デバイス／グループウェアでも含むことができることについて、明記されるべきである。それぞれのデータ・サーバ114に結合されるものは、複数のユーザ・デバイス116である。ユーザ・デバイス116はまた、ネットワーク104、106、108の1つを通して直接的に結合されてもよい。そのようなユーザ・デバイス116は、デスクトップ・コンピュータ、ラップトップ・コンピュータ、ハンドヘルド・コンピュータ、プリンタ又は如何なる他のタイプの論理を含むことができる。ユーザ・デバイス111はまた、1つの実施形態では如何なるネットワークにでも直接結合することができることが留意されるべきである。

【0022】

周辺機器120又は複数の周辺機器120の連なり、例えばファクシミリ機、プリンタ、ネットワーク又はローカル又はそれら両方のストレージ・ユニット又はシステムは、1つ又はそれ以上のネットワーク104、106、108に結合することができる。データベース又は追加的なコンポーネント又はそれらの両方は、ネットワーク104、106、108に結合された如何なるタイプのネットワーク要素と共に用いられ、また一体化されることができることが留意されるべきである。本明細書のコンテキストにおいては、ネッ

トワーク要素は、ネットワークの否かるコンポーネントでも参照することができる。

【0023】

いくつかのアプローチにより、本明細書で説明する方法及びシステムは、仮想化システムと共に又は仮想化システム上に、又はそれら両方に、若しくはIBM zOS環境をエミュレートするUNIX（登録商標）システム、MICROSOFT WINDOWS（登録商標）環境を仮想的にホストするUNIX（登録商標）システム、IBM z/OS環境をエミュレートするMICROSOFT WINDOWS（登録商標）システムなどといった、1つ又はそれ以上のシステムをエミュレートするシステムに実装することができる。この仮想化又はエミュレーション又はそれらの両方は、いくつかの実施形態では、VMWARE（登録商標）を通して強化することができる。

10

【0024】

さらなるアプローチにおいては、1つ又はそれ以上のネットワーク104、106、108は、“クラウド”として共通的に参照されるシステムのクラスタを表すことができる。クラウド・コンピューティングにおいては、処理能力、周辺機器、ソフトウェア、データ、サーバなどといった共有リソースがクラウド内の如何なるシステムに対してもオンデマンドの関係において提供され、それによって、多数のコンピューティング・システムにわたってサービスへのアクセス及び配布を可能とする。クラウド・コンピューティングは、典型的にはクラウド内において動作しているシステムの間インターネット接続を含むが、システムを結合する他の技術もまた、使用することができる。

【0025】

図2は、図1のユーザ・デバイス116又はサーバ114又はそれら両方に伴われる代表的なハードウェア環境を示す。図2は、マイクロプロセッサといった中央処理ユニット210及びシステム・バス212を介して相互接続された多数の他のユニットを含むワークステーションの典型的なハードウェア構成を示す。

20

【0026】

図2に示すワークステーションは、ランダム・アクセス・メモリ（RAM）214、リード・オンリー・メモリ（ROM）216、ディスク・ストレージ・ユニット220といった周辺機器をバス212へと結合するI/Oアダプタ218、キーボード224、マウス226、スピーカ228、マイクロホン232又はタッチスクリーン及びデジタル・カメラ（不図示）といった他のユーザ・インタフェース・デバイスをバス212へと結合するためのユーザ・インタフェース・アダプタ222、ワークステーションを通信ネットワーク235（例えばデータ処理ネットワーク）に結合するための通信アダプタ234、及びバス212をディスプレイ・デバイス238結合するためのディスプレイ・アダプタ236を含む。

30

【0027】

ワークステーションは、それ上に、MICROSOFT WINDOWS（登録商標）オペレーティング・システム、MAC OS（登録商標）、UNIX（登録商標）OSなどといったオペレーティング・システムの格納部を有する。好ましい実施形態は、また、上述したものではないプラットフォーム及びオペレーティング・システム上に実装することができる。好ましい実施形態は、オブジェクト指向プログラミング方法論と共に、XML、C、又はC++言語又はそれらの組み合わせ、又は他のプログラミング言語を使用して記述することができる。

40

【0028】

ここで図3を参照すると、1実施形態による方法300のフローチャートが示される。方法300は、種々の実施形態の中でもとりわけ、図1～2及び5に図示される環境の如何なるものにおいても、本発明により実行することができる。当然ながら、図3に特定の記載されたものではなく、本明細書を読んだ当業者により理解されるであろうように、より多くの、又はより少ない操作が方法300に含まれてもよい。

【0029】

方法300の各ステップは、オペレーティング環境の如何なる好適なコンポーネントに

50

よっても実装することができる。例えば、種々の実施形態においては、方法300は、1つ又はそれ以上のサーバ、コンピュータにより、又は1つ又はそれ以上のプロセッサを内部に有するいくつかの他のデバイスによって、部分又は全体が実行されることができる。例えば、ハードウェア又はソフトウェア又はそれら両方に実装される処理回路（複数でも良い）、チップ（複数でも良い）、又はモジュール（複数でも良い）及び好ましくは少なくとも1つのハードウェア・コンポーネントを有するプロセッサは、本方法300の1つ又はそれ以上のステップを実行するために如何なるデバイスの内部でも使用することができる。例示的なプロセッサは、これらに限定されないが、中央処理ユニット（CPU）、特定用途集積回路（ASIC）、フィールド・プログラマブル・ゲートアレイ（FPGA）など、これらの組み合わせ又は本技術において知られている如何なる他の好適なコンピューティング・デバイスを含む。

10

【0030】

図3に示されるように、方法300は、操作302で開始することができ、ここで、質問は、自然言語フォーマットで受領される。1つの実施形態においては、質問は、質問・回答（QA）アプリケーションにより受領される。もう1つの実施形態では、質問は、OAアプリケーションにより回答されるべき照会を含むことができる。さらなるもう1つの実施形態では、自然言語フォーマットは、人間による日常会話において使用されるフォーマットを含むことができる。

【0031】

加えて、1つの実施形態においては、自然言語フォーマットは、プログラム・コードを使用する形式的な言語などから区別することができる。もう1つの実施形態においては、自然言語フォーマットにおける例示的な質問は、質問「2018年の会社Xの競合者の収益はどうか？」を含むことができる。

20

【0032】

さらに、方法300は、操作304に進むことができ、ここでは、複数のサブクエリーが、質問から導出される。1つの実施形態においては、サブクエリーは、質問の自然言語フォーマットを論理形式にパースすることにより、質問から導出することができる。例えば、質問がパースされて、質問の1つ又はそれ以上のコンポーネントを識別することができる。もう1つの実施例においては、1つ又はそれ以上のコンポーネントが分析されて、質問の1つ又はそれ以上のサブクエリーを生成することができる。

30

【0033】

さらに、1つの実施形態においては、それぞれのサブクエリーは、トリプル（例えば、主語／述語／目的語のトリプル、など）といったn重タプルを含むことができる。例えば、不完全なトリプルは、主語、述語、目的語のフィールドの1つ又はそれ以上についてのデータが無いことがある。もう1つの実施形態では、主語／述語／目的語のトリプルは、質問のパースの結果に基づいて記入される、主語、述語、及び目的語を含むデータ・オブジェクトを含むことができる。例えば、質問「2018年の会社Xの競合者の収益はどうか？」について、導出される主語／述語／目的語のトリプルは、以下の不完全なトリプルを含む可能性がある：第1のサブクエリー（? y “の収益” ? x）及び第2のサブクエリー（? x, “競合者の”, “会社X”）。

40

【0034】

また、方法300は、操作306に進むことができ、ここでは、複数のサブクエリーの第1のサブクエリーが、キュレーテッド・ナレッジ・ベースを使用して解決することができないと判断される。1つの実施形態では、キュレーテッド・ナレッジ・ベースは、非構造化テキストから収集された所定のドメインを確認するためのデータを含むことができる。例えば、非構造化テキスト（例えば、フリーテキストなど）は、所定のドメインにマッチするテキストデータ（例えば単語、センテンスなど）を判断するためにパース（手作業又は自動的に又はそれら両方で）されることができる。もう1つの実施形態においては、キュレーテッド・ナレッジ・ベースは、複数の完全なトリプレットを含む可能性がある。

【0035】

50

加えて、1つの実施例では、所定のドメインは、キュレーテッド・ナレッジ・ベースのための、1つ又はそれ以上の所定のトピックス（例えば、会社の財務情報など）を含むことができる。もう1つの実施例では、所定のドメインにマッチするテキストデータは、所定のドメインのためのキュレーテッド・ナレッジ・ベースに追加することができる。さらにもう1つの実施例では、キュレーテッド・ナレッジ・ベースは、複数の完全なトリプルを含むことができる。

【0036】

さらに、1つの実施形態では、複数のサブクエリーの第1のサブクエリーは、キュレーテッド・ナレッジ・ベースを使用するQAアプリケーションによっては解決することができない場合がある。複数のサブクエリーのうちの第1のサブクエリーは、不完全なトリプルをキュレーテッド・ナレッジ・ベースとの比較により完全化することができない、不完全なトリプルを含むことがある。もう1つの実施例では、複数のサブクエリーのうちの第2のサブクエリーは、キュレーテッド・ナレッジ・ベースを使用するQAアプリケーションにより解決することができる場合がある。例えば、複数のサブクエリーの第2のサブクエリーは、不完全なトリプルをキュレーテッド・ナレッジ・ベースと比較することにより完全化することができる不完全なトリプルを含む場合がある。もう1つの実施形態においては、第1のサブクエリーは、キュレーテッド・ナレッジ・ベースに対する第1のサブクエリーのセマンティックな関連性の分析に基づいて、キュレーテッド・ナレッジ・ベースを使用して解決できないと判断される場合がある。

【0037】

さらに、方法300は、操作308に進むことができ、ここでは、第1のサブクエリーが、非構造化テキストを使用して解決される。1つの実施形態では、非構造化テキストは、そこからキュレーテッド・ナレッジ・ベースが生成されたテキスト・コーパスを含むことができる。もう1つの実施形態においては、非構造化テキストは、所定の構造を全く有しない、履歴的なテキスト・ドキュメント（例えば、テキスト・ファイル、HTMLなど）を含むことができる。さらにもう1つの実施形態においては、第1のサブクエリーは、不完全なトリプルを含むことができる。1つの実施形態においては、第1のサブクエリーは、第1のサブクエリーが、キュレーテッド・ナレッジ・ベースを使用して解決できないとの判断に応答して、非構造化テキストを使用して解決される場合がある。

【0038】

また、1つの実施形態では、第1のサブクエリーの解決は、非構造化テキストを使用して以前に生成された検索インデックスを照会することを含むことができる。例えば、非構造化テキストは、事前処理（例えば、パースされ、分析されるなど）されて、検索インデックスを生成することができる。もう1つの実施例では、検索インデックスは、非構造化テキストのために生成されたキーワード・ベースの検索エンジンを含むことができる。さらにもう1つの実施例では、検索インデックスは、非構造化テキストのために生成されたキーワード・ベースの検索エンジンを含むことができる。例示的には、検索エンジンは、第1のサブクエリーの不完全なトリプル内に見出された語を使用して検索することができる。

【0039】

加えて、1つの実施形態では、クエリー・リラクゼーションは、検索インデックスのクエリーに含ませる追加的語を生成するために使用することができる。例えば、クエリー・リラクゼーションを実行することは、第1のサブクエリー内の不完全なトリプル内の語について、類似語（例えば、シノニムなど）を決定することを含むことができる。もう1つの実施例では、検索インデックスの照会は、不完全なトリプルの語に加えてこれらの類似語を含むように修正することができる。

【0040】

さらに、1つの実施形態では、第1のサブクエリーを解決することは、検索インデックスの照会の結果を受領することを含むことができる。例えば、結果は、興味のある1つ又はそれ以上の断片を含む可能性がある。もう1つの実施例では、1つ又はそれ以上のテキ

ストの断片のそれぞれは、検索インデックスの照会にマッチする、1つ又はそれ以上のテキストの部分（例えば、単語、フレーズ、センテンス、パラグラフなど）を含む場合がある。さらにもう1つの実施例では、1つ又はそれ以上のテキスト断片は、1つ又はそれ以上の基準（例えば、検索インデックスの照会に対する関連性など）に従ってランク付けすることができる。

【0041】

さらに、1つの実施形態においては、第1のサブクエリーを解決することは、検索インデックスの照会の結果を使用して新たなトリプルを生成することを含む。例えば、新たな主語／述語／目的語のトリプルは、検索インデックスの照会の結果として受領された、1つ又はそれ以上のテキスト断片を使用して生成することができる。もう1つの実施例では、これらの新たなトリプルが完全化する可能性はある（例えば、それらは、主語、述語、及び目的語のフィールドについての値を有することができる）。

10

【0042】

また、1つの実施形態では、第1のサブクエリーを解決することは、新たなトリプレットを使用して第1のサブクエリーを完全化することを含むことができる。例えば、第1のサブクエリーは、不完全トリプルを含む場合がある。もう1つの実施例では、第1のサブクエリーの不完全なトリプルは、新たなトリプルに基づいて完全化する可能性がある。さらにもう1つの実施例では、無いフィールドまたは不完全なトリプルについての値を決定するために、第1のサブクエリーの不完全なトリプルについて、1つ又はそれ以上の整列ベースの回答識別及び類似の操作が新たなトリプレットに基づいて実行される。

20

【0043】

また、1つの実施形態においては、非構造化データは、定期的にアップデートされて、以前に検索されたものの、非構造化データ内には見出されなかった情報を含ませることができる。例えば、検索インデックスの照会に応答して戻されるテキスト断片の量が、所定の閾値を下回る場合、検索は、検索インデックスの照会に、さらに関連すると判断される結果を含む追加の非構造化データについて行われる。この追加の非構造化データは、現在の非構造化データに組み込まれて非構造化データのコーパスを増強する。

【0044】

加えて、方法300は、操作310に進んで、ここでは、キュレーテッド・ナレッジ・ベースが解決された第1のサブクエリーを使用して適応される。1つの実施形態においては、キュレーテッド・ナレッジ・ベースを適応させることは、キュレーテッド・ナレッジ・ベースに解決された第1のサブクエリーを追加することを含む。もう1つの実施形態においては、解決された第1のサブクエリーは、解決された第1のサブクエリーと、キュレーテッド・ナレッジ・ベース内の既存のトリプルとの間の依存性の判断に基づいて、キュレーテッド・ナレッジ・ベース内の既存の完全化したトリプルに追加されることになる完全化したトリプルを含むことができる。

30

【0045】

さらに、1つの実施形態においては、複数のサブクエリーの第2のサブクエリーは、適応されたキュレーテッド・ナレッジ・ベースを使用して解決することができる。例えば、第2のサブクエリーを解決することは、適応されたキュレーテッド・ナレッジ・ベース内のトリプルを使用して第2のサブクエリーを完全化することを含むことができる。もう1つの実施例では、複数のサブクエリーの第2のサブクエリーは、適応されたキュレーテッド・ナレッジ・ベース内（例えば、解決された第1の完全化したトリプルを含む）の完全化したトリプルと比較される不完全なトリプルを含む可能性がある。

40

【0046】

さらに、1つの実施例においては、1つ又はそれ以上の整列ベースの回答識別及び類似操作は、無いフィールドまたは不完全なトリプルのフィールドについての値を決定するために、適応されたキュレーテッド・ナレッジ・ベース内の完全化したトリプルに基づいて第2のサブクエリーの不完全なトリプルについて実行することができる。もう1つの実施例では、第2のサブクエリーの、無いフィールドまたは不完全なトリプルのフィールドに

50

ついで、その値を、不完全なトリプルに追加して、完全なトリプルを生成することができる（それにより、第2のサブクエリーを解決する）。

【0047】

また、1つの実施形態では、解決された第1のサブクエリー及び解決された第2のサブクエリーは、受領した質問の回答を生成するために使用することができる。例えば、解決された第1のサブクエリー及び解決された第2のサブクエリーのトリプルのフィールド内の情報は、自然言語フォーマットを有する応答へと変換することができる。もう1つの実施例では、応答は、自然言語フォーマットで戻すことができる。

【0048】

この方法において、非構造化データ（非構造化テキスト）及び構造化データ（キュレーテッド・ナレッジ・ベース）の両方を、QAシステムにより質問に回答するために互いに使用することができる。加えて、キュレーテッド・ナレッジ・ベースは、非構造化テキスト検索及び分析の結果を使用して、動的に増強することができる。これがキュレーテッド・ナレッジ・ベースの妥当性を進歩させ、このことがキュレーテッド・ナレッジ・ベースを使用して実行される将来のサブクエリー検索の正確性及び性能を改善することができる。キュレーテッド・ナレッジ・ベースの妥当性の改善により、キュレーテッド・ナレッジ・ベースを使用するコンピューティング・デバイスはまた、改善された性能からの利益を得ることができる。

【0049】

ここで、図4を参照すると、1実施形態による、オンデマンド抽出及び一体化を実装するための方法400のフローチャートが示される。方法400は、種々の実施形態の中でもとりわけ、図1～2、及び5に図示された如何なる環境においても本発明により実行することができる。当然ながら、図4に特定的に記載されたもの以外により多く、又はより少ない操作は、本明細書を読んだ当業者において理解されるであろうように、方法400において含まれることができる。

【0050】

方法400のステップのそれぞれは、オペレーティング環境の如何なる好適なコンポーネントによっても実行することができる。例えば、種々の実施形態においては、方法400は、1つ又は複数のサーバ、コンピュータ、又はその内部に1つ又はそれ以上のプロセッサを有するいくつかの他のデバイスにより、部分又は全体を実行することができる。例えば、ハードウェア又はソフトウェア又はそれら両方に実装される処理回路（複数でも良い）、チップ（複数でも良い）、又はモジュール（複数でも良い）及び好ましくは少なくとも1つのハードウェア・コンポーネントを有するプロセッサは、本方法400の1つ又はそれ以上のステップを実行するために如何なるデバイスの内部にあっても使用することができる。例示的なプロセッサは、これらに限定されないが、中央処理ユニット（CPU）、特定用途集積回路（ASIC）、フィールド・プログラマブル・ゲートアレイ（FPGA）など、これらの組み合わせ又は本技術において知られている如何なる他の好適なコンピューティング・デバイスをも含む。

【0051】

図4に示されるように、方法400は、操作402で開始することができ、ここで、質問は、自然言語フォーマットで受領される。追加的に方法400は、操作404に進み、ここで、第1のサブクエリー及び第2のサブクエリーが、質問から導出される。1つの実施形態においては、第1のサブクエリー及び第2のサブクエリーは、両方とも不完全な主語／述語／目的語のトリプルを含む。もう1つの実施形態では、第1のサブクエリーは、第2のサブクエリーとは異なる。例えば、第1のサブクエリーは、第2のサブクエリーに比較して異なるトリプルのフィールド値を有することができる。

【0052】

追加的に方法400は、操作406に進むことができ、ここで、第1のサブクエリーが非構造化テキストに基づいてキュレーテッド・ナレッジ・ベースを使用して解決できないことを判断する。さらに方法400は、操作408に進むことができ、ここで、クエリー

・リラクゼーションを使用して、追加の検索語が第1のサブクエリーについて生成される。さらに、方法400は、操作410に進むことができ、ここで第1のサブクエリー及び追加の検索語を使用して、非構造化テキストについて検索インデックスが照会される。1つの実施形態においては、検索インデックスは、非構造化テキストを使用して以前に生成されたものとして行うことができる。もう1つの実施形態では、非構造化テキストのための検索インデックスは、第1のサブクエリーがキュレーテッド・ナレッジ・ベースを使用して解決できないと判断されたことに応答して照会を行うことができる。

【0053】

また、方法400は、操作412に進み、ここで、検索結果が検索インデックスの照会に応答して受領される。加えて、方法400は、操作414に進み、ここで、検索結果に基づいて、新たな主語／述語／目的語のトリプルを生成する。例えば、検索結果は、複数の関連する断片を含むことができ、オープン・インフォメーション抽出を、断片から新たなトリプルを抽出するために使用することができる。もう1つの実施形態においては、新たなトリプルは、第1のサブクエリー及び第2のサブクエリーに対するそれらのセマンティック的な関連性に従ってランク付けを行うことができ、上位にランク付けされたトリプルの所定数が、新たなトリプルとして選択される。さらに、方法400は、操作416に進み、ここで、第1のサブクエリーが、新たな主語／述語／目的語のトリプルを使用して完全化される。さらに、方法400は、操作418に進み、ここで、キュレーテッド・ナレッジ・ベースが、完全化された第1のサブクエリーを含むようにアップデートされる。

【0054】

また、方法400は、操作420に進み、ここで、第2のサブクエリーがアップデートされたキュレーテッド・ナレッジ・ベースを使用して解決できることを判断する。1つの実施形態においては、第2のサブクエリーは、第1のサブクエリーが完全化されるまで保留される。例えば、第2のサブクエリーは、第1のサブクエリー及び第2のサブクエリーの不完全なトリプルの分析に基づいて第1のサブクエリーに応じて判断を行うことができる。例示として、第2のサブクエリーで無いフィールドは、第1のサブクエリー内に配置されていることがあり、第2のサブクエリーの無いフィールドが第1のサブクエリーの完全化時に完全化できると判断を行うことができる。もう1つの実施形態では、第2のサブクエリーは、この判断に応答して第1のサブクエリーが完全化するまで、保留することができる。

【0055】

追加的に方法400は、操作422に進むことができ、ここで、第2のサブクエリーがアップデートされたキュレーテッド・ナレッジ・ベースを使用して完全化される。1つの実施形態では、第2のサブクエリーは、第1のサブクエリーが完全化し、かつキュレーテッド・ナレッジ・ベースに追加されたことの判断に応答して完全化を行うことができる。もう1つの実施形態では、第2のサブクエリーは、完全化した第1のサブクエリーのトリプルを含むアップデートされたキュレーテッド・ナレッジ・ベース内の主語／述語／目的語のトリプレットを使用して解決される。

【0056】

この方法において、キュレーテッド・ナレッジ・ベース及び非構造化テキスト両方の検索は、キュレーテッド・ナレッジ・ベースの増強及び解決した質問に正確に応答するように遂行することができる。

【0057】

図5は、例示的な実施形態による、非構造化データ及び構造化データを一体化させた、質問・回答(QA)システム500を示す。QAシステム500は、非構造化データ(フリーテキスト)及び構造化データ(キュレーテッド・ナレッジ・ベース502)の両方からリアルタイムで情報を一体化することにより質問に回答する。

【0058】

図示されるように、システムへの入力は、自然言語の質問Q504である。質問Q504は、自然言語理解ユニット506によりパースされ、これは、質問を、機械により理解

10

20

30

40

50

することができるサブクエリー508に変換する。例えば、サブクエリー508は、不完全な主語／述語／目的語のトリプルを含むことができる。

【0059】

追加的に、サブクエリー508のそれぞれのサブクエリーq_xについて、照会計画生成器510は、類似のクエリーのセットを使用してq_xを拡張することにより、クエリー拡張を実行する。例えばq_xが、

(会社, 取得, ?)

を含むことができる。

【0060】

クエリー拡張は、その後、

(会社, 買収, ?) 及び

(?, により買収, 会社)

といった、また推定されたクエリーについて適用される。

【0061】

1つの実施形態においては、上述のクエリー拡張は、1つ又はそれ以上のパラフレーズ・モデルを使用して完全化することができる。その後、ストリンガー類似性マッチング、埋め込み（例えば、事前学習させたword2vec埋め込み）マッチング、及びパラフレーズに基づくマッチングが、どのデータ・ソース（例えば、非構造化テキスト512又はキュレーテッド・ナレッジ・ベース502）へと、それぞれのサブクエリー508を送付するべきかを判断するために使用することができる。サブクエリー508は、判断されたデータ・ソースに基づいて、2つのサブセット—フリーテキスト・サブクエリー514及びキュレーテッド・ナレッジ・ベース・サブクエリー516に分割される。

【0062】

さらに、フリーテキスト・サブクエリー514は、フリーテキスト照会モジュール518へと送付され、これはインデックス付けされたテキスト・コーパス520を検索し、フリーテキスト・サブクエリー514に対する回答を含む可能性のある、関連するテキスト断片522の収集を識別する。非構造化テキスト512は、インデックス付け524されて、テキスト・コーパス520を生成する。これは、関連するテキスト断片522を迅速に識別し、検索するために、インデックス付け技術（例えば、TF-IDFに基づく方法など）を使用して完全化される。インデックス付けされたコーパス520を検索する場合、フリーテキスト・サブクエリー514は、包括性を増加させるために、パラフレーズ・モデルを使用して拡張することができる。

【0063】

さらに、関連するテキスト断片522は、情報抽出器526（リラクゼーション抽出器といった）に渡されて、それぞれのフリーテキスト・サブクエリー514に対する回答の候補を抽出し（これにより、それぞれのフリーテキスト・サブクエリー514が完全化する）、それらの関連性スコアに基づいてそれらをランク付けする。

【0064】

また、完全化されたフリーテキスト・サブクエリー514の全部またはいくつかはその後、キュレーテッド・ナレッジ・ベース502に追加され、キュレーテッド・ナレッジ・ベース502をアップデートする。1つの実施形態では、完全化されたフリーテキスト・サブクエリー514の全部、又はいくつかは、完全化されたキュレーテッド・ナレッジ・ベース・サブクエリー516と一体化され（完全化された、とは、例えば、キュレーテッド・ナレッジ・ベース502を使用するキュレーテッド・ナレッジ・ベースに対する照会モジュール528による）、自然言語の質問Q504に回答して提供される最終的な回答を判断することができる。もう1つの実施形態では、1つ又はそれ以上のキュレーテッド・ナレッジ・ベース・サブクエリー516は、アップデートされたキュレーテッド・ナレッジ・ベース502を使用するキュレーテッド・ナレッジ・ベースに対する照会モジュール528により完全化することができ、完全化されたナレッジ・ベース・サブクエリー516は、その後、自然言語の質問Q504に回答して提供されるべき最終的な回答を判断

10

20

30

40

50

するために使用することができる。

【0065】

もう1つの実施形態においては、自然言語の質問Q504から識別されたアグリゲート・サブクエリーが存在する場合、1つ又はそれ以上のアグリゲーションを実行することができる。

【0066】

オンデマンド抽出及び一体化

ナレッジ・ベース（KB）は、多くのデータを中心とした、質問・回答（QA）アプリケーションといったアプリケーションの核となった。大規模なナレッジ・ベースは、異なったエンティティに関する膨大な量の知識を格納し、システムがエンティティを中心とする質問に対する回答を提供することを可能とする。これらのナレッジ・ベースの中でも、キュレートされたKBsは、システムが高い精度の知識にわたって理由付けをし、かつ正確な回答を返すことを可能とする。

【0067】

ナレッジ・ベースは、ドメインに特化したスキーマ及びキュレーションのワークフローの提案が要求されるドメイン・エキスパートにより記述され、かつキュレートされて、スキーマに肯定される事実を抽出することが可能である。高い精度であったとしても、これらのKBsは、世界の知識が、選択された言語的な述語を使用してエンコードすることができることを仮定するため、不完全である可能性がある。しばしば、KBsからの情報が無いことは、KBsが立脚する非構造化テキストにおいて存在する。

【0068】

例示的に、KBは、公的な会社に関する情報及びそれらのいくつかの財務的基準を含む可能性が有るが、取得者、競合者、及び会社の間に存在する子会社といった関係を含まない場合がある。これらの情報は、使用可能な場合、情報技術を使用してテキストからキュレートして、完全化させ、かつKBsに無い情報を補足することができる。これは、キュレートされたKBsの不完全性に遭遇するQAシステムの汎用化を支援することができる。ドメイン・エキスパートは、自動的にキュレートされたテキストを取得し、かつシステムにより提供された回答からの洞察を使用して、キュレーション・ワークフローをさらに改善する、という質問を依頼する可能性がある。

【0069】

例示として、ドメイン・エキスパートが公的な会社について、これらの会社の財務報告内のテーブル（例えば10-kのレポート）から、これらの会社の正確な財務基準を含むKBを構築しようとする。しかしながら、ユーザが興味のあるクエリーは、同じように見えるであろう。

「私に2016年における会社Xの競合者の収益を示せ」

【0070】

このクエリーは、キュレートされたKBからだけではサポートすることができず、かつ2つのサブクエリーに分解することができる：

(1) (?y, の収益, ?x)、及び

(2) (?x, “の競合者”, “会社X”)、

ここで、クエリー(1)は、我々がクエリー(2)に対する回答を知るや否やキュレーテッド・ナレッジ・ベースにより回答することができ、この回答は、会社Xの財務報告のフリーテキストにおいて見出すことができる（報告が実際に興味のある情報を含んでいるものと仮定する。）。

【0071】

この結果、1実施形態においては、不均質なデータ・ソース、構造化KB、及び非構造化されたテキストを結合し、構造化KBの正確さ及び非構造化テキストからの事実の包括性に資する、動的なハイブリッド・ナレッジ・ベースを構築するために使用することができる。所与の自然言語の質問又は自然言語の質問の解釈について、システムは、全体として、非構造化テキスト又は構造化KBを選択的に照会し、必要に応じて情報を抽出し、一

10

20

30

40

50

体化させ、結果を一体化してクエリーに回答することができる。

【0072】

例示的な特徴は、以下のものを含む：

1. 所与の自然言語クエリー（NLQ）のローカルな代表を、論理クエリーとして参照すると、対応する照会計画が生成される。この計画は、既存のナレッジ・ベース及び非構造化データに対して照会を行って中間的な結果を生成し、中間的な回答を結合し、そして最終的な結果をアグリゲートする照会操作器から構成される。

2. 非構造化データにアクセスする照会操作器については、操作器に関連する関係のある断片が、非構造化データから検索される。

a. 関係のある断片からの事実が、抽出され、ランク付けされる。

b. 入力した照会操作器についての結果が、抽出された事実から判断される。

3. キュレーテッド・ナレッジ・ベースにアクセスした照会操作器のために、ナレッジ・ベースに対する照会が行なわれ、関連する事実を検索する。

4. 元の照会計画についての最終的な結果が生成される。

【0073】

1つの例示的なアプローチは、オンデマンド様式でフリーテキストを紹介して、無い情報を見出すことで、照会に回答することができる。このアプローチは、静的なナレッジ・ベースへのアプローチに比較した場合、より多くの照会をサポートする。加えて、無い情報は、オンデマンド様式で、フリーテキストから抽出される。さらに、本アプローチは、関連する情報がキュレートされたKBに無い場合、複雑なクエリーを、キュレートされたKB、又は抽出されたKBにより、より信頼性をもって回答することができるようにする、多数のサブクエリーへと分解する。

【0074】

如何なる質問についてでも回答する前に、システムは、ワнтаイムのプリプロセッシング・ステップを必要とする。オフライン・プロセスにおいては、システムは、複数のテキスト・コーパスの上で、照会時におけるキーワード検索を効率化することを可能とする、転置インデックスを生成する。

【0075】

システムは、オフライン処理で、複数のテキスト・コーパスの上で、キーワード検索を効率化することを可能とする、転置インデックスを生成する。これに加えて、年、名前空間、及びテキストに伴われるエンティティといったいくつかのメタデータがインデックス付けされる。

【0076】

システムへの入力、論理クエリー（例えば、自然言語での質問の1つの解釈）である。論理クエリーは、多数のクローズの組み合わせであり、ここで、それぞれのクローズは、1つ又はそれ以上の可変的な要旨を伴うn重のタプルである。実体的な計画生成器は、論理クエリーを翻訳して、照会のための実体的な計画を生成する。実体的な計画は、クローズの実行順、及びクローズをキュレートされたKB又はフリーテキストに対して実行することが必要か否かを決定する。加えて、実体的な計画は、どのようにして中間的な結果を組み合わせ、任意的にアグリゲートするかを決定する。

【0077】

クエリー・クローズがフリーテキストに対して実行される必要がある場合、クエリー・クローズは、より柔軟化された照会計画を生成するために、先ずメタデータを使用して書き換えられる。これは、興味のある情報を表現しない可能性のあるテキスト断片を検索することを支援する。柔軟化された照会計画のそれぞれは、同一の事実が、フリーテキスト内に表現されるであろう数多くのやり方を処理するため、その後、パラフレーズ化される。照会計画は、元のクエリーに対するそれらのセマンティック的な類似性に基づいてランク付けされる。上位k個の計画が最適化され、照会計画にわたった共通のキーワードのクエリーが識別され、かつ、一度に実行される。

【0078】

キーワード・クエリーがインデックス付けされたテキスト・コーパスに対して発行されて、関連するテキスト断片を検索する。オープン・ドメインのインフォメーション抽出器が、テキスト断片から事実のトリプル（主語－述語－目的語）を抽出する。これらのトリプルは、ノイズが多く、かつ規格化されていないので、システムは、さらにこれら进行处理してアノテートし、回答を正確に識別することを支援する。トリプルは、異なる照会計画を横断してアグリゲートされ、かつそれらの推定された信頼性に基づいてランク付けされる。システムは、その後、整列ベースの回答識別アルゴリズムを使用してトリプルから回答を識別する。これらの回答は、さらに非曖昧化され、さらなる利用のために計画生成器に戻される。

10

【0079】

クエリー・クローズをキュレートされたKBに対して実行する必要がある場合、それは、先ず高レベルの宣言的言語におけるクエリーに翻訳されなければならない。オントロジ・クエリー言語（OQL）は、キュレートされたKBに格納された情報にアクセスするために使用することができる。クエリー分解は、クエリーを多数の単純なクローズに分解するので、手作業で作成されたテンプレートを、クローズをOQLクエリーに翻訳するために使用することができる。論理クエリーと、OQLとの間のスキーマ要素の情報のオーバーラップは、キュレートされたKBに照会するためのテンプレートを選択するために使用することができる。選択されたテンプレートは、インスタンス化され、対応するOQLクエリーがKBに対して発行され、タプルのセットが戻される。

20

【0080】

1つの実施形態においては、不均質なデータ・ソースは、構造化KBを含むことができ、これは、高精度の実世界の事実及びフリーテキストを有し、フリーテキストは、KB内の事実を捕捉する膨大なセンテンスのレポジトリを含む。ハイブリッド・システムは、キュレートされたKBの精度及びテキスト形式の事実の包括性による利益を得、かつドメイン・エキスパートがキュレートされたKBsに無い述語を識別することを支援する。

【0081】

加えて、1実施形態では、キーワード・クエリーは、クエリー・リラクゼーション及び書き換えについてのメタデータを使用して関連するドキュメントのスーパーセットにルートさせることができる。加えて、パラフレージング・モデルは、クエリーの拡張のために使用することができる。

30

【0082】

本発明は、いかなる可能な技術的に詳細な一体化レベルであっても、システム、方法、又はコンピュータ・プログラム製品又はこれらの組み合わせとすることができる。コンピュータ・プログラム製品は、プロセッサに対して本発明の特徴を遂行させるためのコンピュータ可読なプログラム命令をそれ上に有する記録媒体（又は複数の媒体）を含むことができる。

【0083】

コンピュータ可読な記録媒体は、命令実行デバイスが使用するための複数の命令を保持し格納することができる有形のデバイスとすることができる、コンピュータ可読な媒体は、例えば、これらに限定されないが、電氣的記録デバイス、磁氣的記録デバイス、光学的記録デバイス、電気磁氣的記録デバイス、半導体記録デバイス又はこれらのいかなる好ましい組み合わせとすることができる。コンピュータ可読な記録媒体のより具体的な実施例は、次のポータブル・コンピュータ・ディスク、ハードディスク、ランダム・アクセス・メモリ（RAM）、リード・オンリー・メモリ（ROM）、消去可能なプログラマブル・リード・オンリー・メモリ（EPROM又はフラッシュ・メモリ（登録商標））、スタティック・ランダム・アクセス・メモリ（SRAM）、ポータブル・コンパクト・ディスク・リード・イオンリー・メモリ（CD-ROM）、デジタル多目的ディスク（DVD）、メモリ・スティック、フロッピー・ディスク（登録商標）、パンチ・カード又は命令を記録した溝内に突出する構造を有する機械的にエンコードされたデバイス、及びこれらの好

40

50

ましい如何なる組合せを含む。本明細書で使用するように、コンピュータ可読な記録媒体は、ラジオ波又は他の自由に伝搬する電磁波、導波路又は他の通信媒体（例えば、光ファイバ・ケーブルを通過する光パルス）といった電磁波、又はワイヤを通して通信される電気信号といったそれ自体が一時的な信号として解釈されることはない。

【0084】

本明細書において説明されるコンピュータ・プログラム命令は、コンピュータ可読な記録媒体からそれぞれのコンピューティング／プロセッシング・デバイスにダウンロードでき、又は例えばインターネット、ローカル・エリア・ネットワーク、ワイド・エリア・ネットワーク又はワイヤレス・ネットワーク及びそれからの組み合わせといったネットワークを介して外部コンピュータ又は外部記録デバイスにダウンロードすることができる。ネットワークは、銅通信ケーブル、光通信ファイバ、ワイヤレス通信ルータ、ファイアウォール、スイッチ、ゲートウェイ・コンピュータ及びエッジ・サーバ又はこれらの組み合わせを含むことができる。それぞれのコンピューティング／プロセッシング・デバイスにおけるネットワーク・アダプタ・カード又はネットワーク・インタフェースは、ネットワークからコンピュータ可読なプログラム命令を受領し、このコンピュータ可読なプログラム命令を格納するためにそれぞれのコンピューティング／プロセッシング・デバイス内のコンピュータ可読な記録媒体内に転送する。

【0085】

本発明の操作を遂行するためのコンピュータ可読なプログラム命令は、アセンブラ命令、命令セット・アーキテクチャ（ISA）命令、機械語命令、マシン依存命令、マイクロ・コード、ファームウェア命令、状態設定データ、集積回路のための構成データ、又は1つ又はそれ以上の、Smalltalk（登録商標）、C++などのオブジェクト指向プログラミング言語、“C”プログラミング言語又は類似のプログラム言語といった手続き型プログラミング言語を含むプログラミング言語のいかなる組合せにおいて記述されたソース・コード又はオブジェクト・コードのいずれかとすることができる。コンピュータ可読なプログラム命令は、全体がユーザ・コンピュータ上で、部分的にユーザ・コンピュータ上でスタンドアローン・ソフトウェア・パッケージとして、部分的にユーザ・コンピュータ上で、かつ部分的にリモート・コンピュータ上で、又は全体がリモート・コンピュータ又はサーバ上で実行することができる。後者のシナリオにおいて、リモート・コンピュータは、ローカル・エリア・ネットワーク（LAN）、ワイド・エリア・ネットワーク（WAN）を含むいかなるタイプのネットワークを通してユーザ・コンピュータに接続することができ、又は接続は、外部コンピュータ（例えばインターネット・サービス・プロバイダを通じて）へと行うことができる。いくつかの実施形態では、例えばプログラマブル論理回路、フィールド・プログラマブル・ゲートアレイ（FPGA）、又はプログラマブル論理アレイ（PLA）を含む電気回路がコンピュータ可読なプログラム命令を、コンピュータ可読なプログラム命令の状態情報を使用して、本発明の特徴を実行するために電気回路をパーソナライズして実行することができる。

【0086】

本明細書で説明した本発明の実施形態を、本発明の実施形態にしたがい、フローチャート命令及び方法のブロック図、又はそれらの両方、装置（システム）、及びコンピュータ可読な記録媒体及びコンピュータ・プログラムを参照して説明した。フローチャートの図示及びブロック図又はそれら両方及びフローチャートの図示におけるブロック及びブロック図、又はそれらの両方のいかなる組合せでもコンピュータ可読なプログラム命令により実装することができることを理解されたい。

【0087】

これらのコンピュータ可読なプログラム命令は、汎用目的のコンピュータ、特定目的のコンピュータ、または他のプロセッサ又は機械を生成するための他のプログラマブル・データ・プロセッシング装置に提供することができ、コンピュータのプロセッサ又は他のプログラマブル・データ・プロセッシング装置による実行がフローチャート及びブロック図のブロック又は複数のブロック又はこれらの組み合わせで特定される機能／動作を実装す

るための手段を生成する。コンピュータ、プログラマブル・データ・プロセッシング装置及び他の装置又はこれらの組み合わせが特定の仕方で機能するように指令するこれらのコンピュータ可読なプログラム命令は、またコンピュータ可読な記録媒体に格納することができ、その内に命令を格納したコンピュータ可読な記録媒体は、フローチャート及びブロック図のブロック又は複数のブロック又はこれらの組み合わせで特定される機能／動作の特徴を実装する命令を含む製造品を構成する。

【0088】

コンピュータ可読なプログラム命令は、またコンピュータ、他のプログラマブル・データ・プロセッシング装置、又は他のデバイス上にロードされ、コンピュータ、他のプログラマブル装置、又は他のデバイス上で操作ステップのシリーズに対してコンピュータ実装プロセスを生じさせることで、コンピュータ、他のプログラマブル装置又は他のデバイス上でフローチャート及びブロック図のブロック又は複数のブロック又はこれらの組み合わせで特定される機能／動作を実装させる。

【0089】

図のフローチャート及びブロック図は、本発明の種々の実施形態にしたがったシステム、方法及びコンピュータ・プログラムのアーキテクチャ、機能、及び可能な実装操作を示す。この観点において、フローチャート又はブロック図は、モジュール、セグメント又は命令の部分を表すことができ、これらは、特定の論理的機能（又は複数の機能）を実装するための1つ又はそれ以上の実行可能な命令を含む。いくつかの代替的な実装においては、ブロックにおいて記述された機能は、図示した以外で実行することができる。例えば、連続して示された2つのブロックは、含まれる機能に応じて、實際上1つのステップとして遂行され、同時的、実質的に同時的に、部分的又は完全に一時的に重ね合わされた仕方で実行することができ、又は複数のブロックは、時として逆の順番で実行することができる。またブロック図及びフローチャートの図示、又はこれらの両方及びブロック図中のブロック及びフローチャートの図示又はこれらの組み合わせは、特定の機能又は動作を実行するか又は特定の目的のハードウェア及びコンピュータ命令を遂行する特定目的のハードウェアに基づいたシステムにより実装することができることを指摘する。

【0090】

さらに、種々の実施形態によるシステムは、プロセッサ及びプロセッサに一体化又は実行可能な又はそれら両方のロジックを含むことができ、この論理は、本明細書に掲載した1つ又はそれ以上のプロセス・ステップを実行するように構成される。プロセッサは、本明細書において説明したとおり、ディスクリットなプロセッサ又はプロセッシング・ハードウェア、メモリ、I/Oインタフェースなどの多くのコンポーネントを含む、如何なる構成のものにでもすることができる。一体化されることが意味するものは、プロセッサが特定用途集積回路（ASIC）、FPGAなどのハードウェア論理としてそれに埋め込まれた論理を有することを意味する。プロセッサにより実行可能ということが意味するものは、論理がファームウェアといったハードウェア論理、ソフトウェア論理、オペレーティング・システムの部分、アプリケーション・プログラムの部分など、又はハードウェア及びプロセッサによりアクセス可能で、かつプロセッサをしてプロセッサによる実行に際していくつかの機能を実行させるように構成されたソフトウェア論理を意味する。ソフトウェア論理は、ローカル又はリモート又はそれら両方の如何なるメモリ形式のメモリに格納することができる。本技術において知られた、ソフトウェア・プロセッサ・モジュール又はASIC、FPGA、中央処理装置（CPU）、集積回路（IC）、グラフィックス・プロセッシング・ユニット（GPU）などのハードウェア・プロセッサといった、如何なるプロセッサが使用可能である。

【0091】

上述したシステム又は方法論又はそれら両方は、如何なる仕方においても組み合わせることができ、上記に提示した説明からの複数の組み合わせが生成できることが明らかであろう。

【0092】

さらに、本発明の実施形態は、カスタマーの代わりに配置され、オンデマンドにサービスを提供する形態において提供することができる。

【0093】

本開示の種々の実施形態の説明は、例示の目的のために提示されたが、開示された実施形態への排他又は限定を意図するものではない。多くの変更例又は変形例は、本開示の範囲及び精神から逸脱することなく、当業者において自明である。本明細書で使用する用語は、本実施形態の原理、実用的用途、又は市場において見出される技術を超える技術的改善を最良に説明するため、又は本明細書において開示された実施形態を当業者の他の者が理解できるようにするために選択したものである。

10

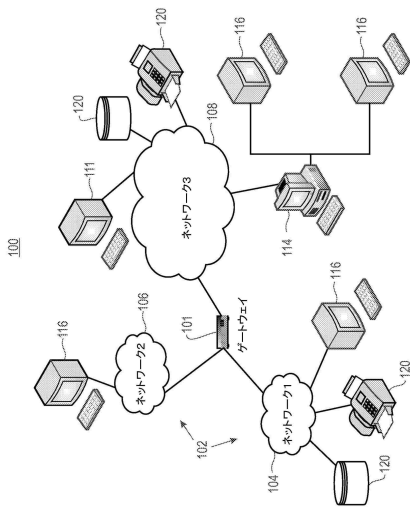
20

30

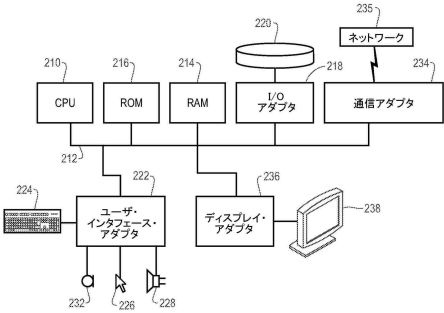
40

50

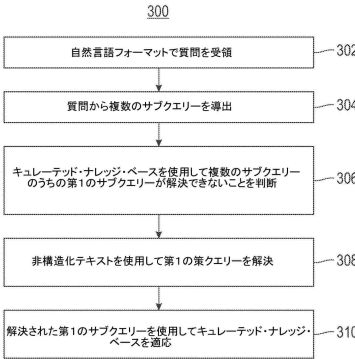
【図面】
【図 1】



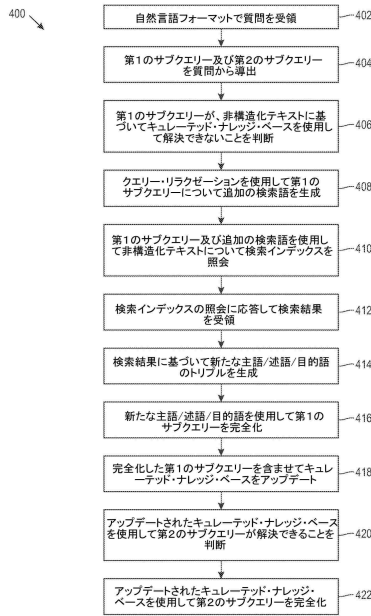
【図 2】



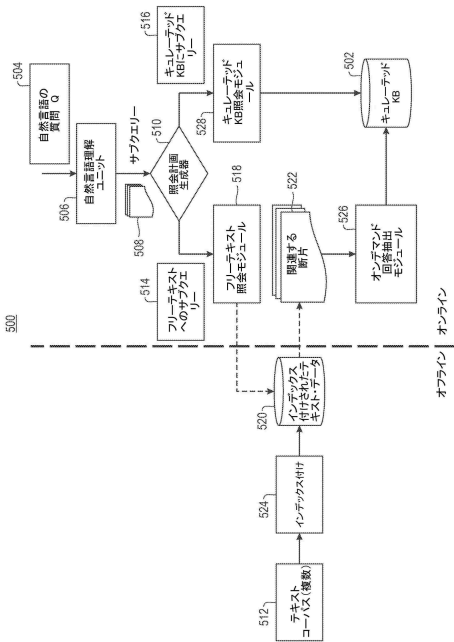
【図 3】



【図 4】



【図 5】



10

20

30

40

50

フロントページの続き

- (72)発明者 クイアン、クン
アメリカ合衆国、９５１２０－６０９９、カリフォルニア州、サンジョーゼ、ハーリー・ロード６
５０、アルマーデン・リサーチ・センター
- (72)発明者 リ、ユンヤオ
アメリカ合衆国、９５１２０－６０９９、カリフォルニア州、サンジョーゼ、ハーリー・ロード６
５０、アルマーデン・リサーチ・センター
- (72)発明者 ブータニ、ニキータ
アメリカ合衆国、９５１２０－６０９９、カリフォルニア州、サンジョーゼ、ハーリー・ロード６
５０、アルマーデン・リサーチ・センター
- 審査官 松尾 真人
- (56)参考文献 米国特許出願公開第２０１９／００４２９８８（ＵＳ，Ａ１）
米国特許出願公開第２０１４／０３５８８９０（ＵＳ，Ａ１）
特開２００７－０８７４０１（ＪＰ，Ａ）
特開２０１６－１４３３１１（ＪＰ，Ａ）
中国特許出願公開第１０６６４９３９４（ＣＮ，Ａ）
- (58)調査した分野 (Int.Cl.，ＤＢ名)
Ｇ０６Ｆ １６／００－１６／９５８