

(19)日本国特許庁(JP)

(12)公開特許公報(A)

(11)公開番号
特開2023-120862
(P2023-120862A)

(43)公開日 令和5年8月30日(2023.8.30)

(51)国際特許分類
G 0 6 F 16/90 (2019.01)

F I
G 0 6 F 16/90 1 0 0

テーマコード (参考)
5 B 1 7 5

審査請求		未請求	請求項の数	12	O L	(全34頁)
(21)出願番号	特願2022-23965(P2022-23965)					
(22)出願日	令和4年2月18日(2022.2.18)					
(71)出願人	314012076 パナソニックIPマネジメント株式会社 大阪府門真市元町2番6号					
(74)代理人	110002000 弁理士法人栄光事務所					
(72)発明者	池田 幹直 東京都港区東新橋2-12-7 住友東 新橋ビル2号館 パナソニックソリュー ションテクノロジー株式会社内					
(72)発明者	武田 鷹広 東京都港区東新橋2-12-7 住友東 新橋ビル2号館 パナソニックソリュー ションテクノロジー株式会社内					
(72)発明者	橋本 拓児 東京都港区東新橋2-12-7 住友東 新橋ビル2号館 パナソニックソリュー ションテクノロジー株式会社内					

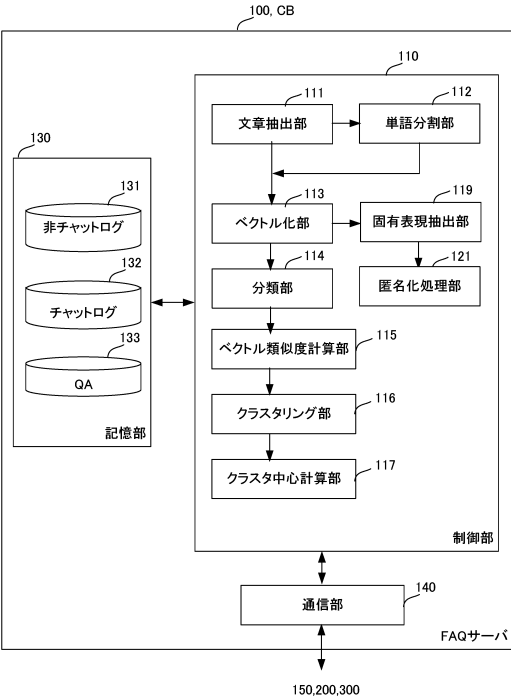
(54)【発明の名称】 データ管理装置、データ管理方法、及びプログラム

(57)【要約】

【課題】FAQシステムの利便性を向上できるデータ管理装置を提供する。

【解決手段】質問文及び質問文に対する回答文を管理するデータ管理装置であって、制御部110と通信部140とを備える。制御部は、複数の文章データを取得し、前記複数の文章データをベクトル化して複数の文章ベクトルを作成し、前記複数の文章ベクトルに基づいて、前記複数の文章データから複数の質問文を抽出し、前記複数の質問文に対応する複数の文章ベクトルである複数の質問文ベクトルに基づいて、抽出された前記複数の質問文を、相互に類似する質問文が含まれる質問クラスタにクラスタリングし、質問クラスタ毎に、質問クラスタの中心に対応する質問クラスタに属する質問文を代表する代表質問文を決定し、通信部は、決定された代表質問文を記憶装置に登録することを推奨する登録推奨情報を、データ管理装置を管理する管理者が使用する管理者端末へ送信する。

【選択図】図2



【特許請求の範囲】

【請求項 1】

質問文及び前記質問文に対する回答文を管理するデータ管理装置であって、
制御部と通信部とを備え、
前記制御部は、
複数の文章データを取得し、
前記複数の文章データをベクトル化して複数の文章ベクトルを作成し、
前記複数の文章ベクトルに基づいて、前記複数の文章データから複数の質問文を抽出し

、
前記複数の質問文に対応する複数の文章ベクトルである複数の質問文ベクトルに基づいて、抽出された前記複数の質問文を、相互に類似する質問文が含まれる質問クラスタにクラスタリングし、

前記質問クラスタ毎に、前記質問クラスタの中心に対応する前記質問クラスタに属する質問文を代表する代表質問文を決定し、

前記通信部は、決定された前記代表質問文を記憶装置に登録することを推奨する登録推奨情報を、前記データ管理装置を管理する管理者が使用する管理者端末へ送信する、
データ管理装置。

【請求項 2】

前記登録推奨情報は、前記代表質問文と前記代表質問文に対応する回答文である代表回答文とを関連付けて前記記憶装置に登録することを推奨する情報を含む、

請求項 1 に記載のデータ管理装置。

【請求項 3】

前記制御部は、

前記複数の文章ベクトルに基づいて、前記複数の文章データから複数の回答文を抽出し

、
前記複数の回答文に対応する複数の文章ベクトルである複数の回答文ベクトルに基づいて、抽出された前記複数の回答文を、相互に類似する回答文が含まれる回答クラスタにクラスタリングし、

前記回答クラスタ毎に、前記回答クラスタの中心に対応する前記回答クラスタに属する回答文を代表する前記代表回答文を決定する、

請求項 2 に記載のデータ管理装置。

【請求項 4】

前記制御部は、前記質問文又は前記回答文に所定の住所又は電話番号が含まれていることを検出した場合、前記所定の住所又は電話番号を削除し、

前記通信部は、前記所定の住所又は電話番号の削除後に、前記登録推奨情報を送信する

、
請求項 3 に記載のデータ管理装置。

【請求項 5】

前記通信部は、利用者が使用する利用者端末から任意の質問文を受信し、

前記制御部は、

前記任意の質問文に基づいて所定の検索処理を行い、検索結果として前記利用者が追加で質問する候補である複数の質問候補文を抽出し、

前記質問候補文をベクトル化して複数の質問候補文の文章ベクトルである複数の質問候補文ベクトルを作成し、

複数の質問候補文ベクトルに基づいて、抽出された複数の質問候補文を、相互に類似する質問候補文が含まれる質問候補クラスタにクラスタリングし、

前記質問候補クラスタ毎に、前記質問候補クラスタの特徴を示す特徴語を抽出し、

前記通信部は、各質問候補クラスタの特徴語を前記利用者端末へ送信する、

請求項 2 に記載のデータ管理装置。

【請求項 6】

前記制御部は、

前記質問候補クラスタが複数存在する場合、前記質問候補クラスタの特徴を示す特徴語が前記質問候補クラスタ毎に異なるように抽出する、

請求項 5 に記載のデータ管理装置。

【請求項 7】

前記制御部は、類似する単語を規定する類似語テーブルに基づいて、各質問候補クラスタの特徴語が異なる単語であるか否かを判定する、

請求項 5 に記載のデータ管理装置。

【請求項 8】

前記制御部は、抽出された前記質問候補文の数に応じて、クラスタリングにより得られる前記質問候補クラスタの数を決定する、

請求項 5 に記載のデータ管理装置。

【請求項 9】

前記制御部は、抽出された複数の特徴語の共起関係を示す共起関係情報を導出し、

前記通信部は、前記共起関係情報を前記利用者端末に送信する、

請求項 5 に記載のデータ管理装置。

【請求項 10】

質問文及び前記質問文に対する回答文を管理するデータ管理装置と、データ管理装置を管理する管理者端末と、を備えるデータ管理システムであって、

前記データ管理装置は、

複数の文章データを取得し、

前記複数の文章データをベクトル化して複数の文章ベクトルを作成し、

前記複数の文章ベクトルに基づいて、前記複数の文章データから複数の質問文を抽出し

、
前記複数の質問文に対応する複数の文章ベクトルである複数の質問文ベクトルに基づいて、抽出された前記複数の質問文を、相互に類似する質問文が含まれる質問クラスタにクラスタリングし、

前記質問クラスタ毎に、前記質問クラスタの中心に対応する前記質問クラスタに属する質問文を代表する代表質問文を決定し、

決定された前記代表質問文を記憶装置に登録することを推奨する登録推奨情報を、前記データ管理装置を管理する管理者が使用する管理者端末へ送信し、

前記管理者端末は、前記登録推奨情報を受信して表示する、

データ管理システム。

【請求項 11】

質問文及び前記質問文に対する回答文を管理するデータ管理方法であって、

プロセッサにより、

複数の文章データを取得するステップと、

前記複数の文章データをベクトル化して複数の文章ベクトルを作成するステップと、

前記複数の文章ベクトルに基づいて、前記複数の文章データから複数の質問文を抽出するステップと、

前記複数の質問文に対応する複数の文章ベクトルである複数の質問文ベクトルに基づいて、抽出された前記複数の質問文を、相互に類似する質問文が含まれる質問クラスタにクラスタリングするステップと、

前記質問クラスタ毎に、前記質問クラスタの中心に対応する前記質問クラスタに属する質問文を代表する代表質問文を決定するステップと、

決定された前記代表質問文を記憶装置に登録することを推奨する登録推奨情報を、前記質問文及び前記回答文を管理する管理者が使用する管理者端末へ送信するステップと、

を有するデータ管理方法。

【請求項 12】

請求項 11 に記載のデータ管理方法をコンピュータに実行させるプログラム。

10

20

30

40

50

【発明の詳細な説明】

【技術分野】

【0001】

本開示は、データ管理装置、データ管理方法、及びプログラムに関する。

【背景技術】

【0002】

従来、対話ログから、質問候補（Ｑ候補）及び回答候補（Ａ候補）をパターン認識してＦＡＱ（Frequently Asked Questions）を作成するＦＡＱシステムが知られている。ここでのパターン認識のためのパターン認識モデルとして、質問者と回答者との一連の通話をテキストとした対話ログを形態素解析したものについて、一連の通話を構成する個々の発話であるレコードごとに算出される、当該発話あたりの形態素数である発話長と、当該発話の前後で共通する単語の出現回数である共通キーワード出現回数と、当該発話あたりの周辺キーワードの重要度の総和であるキーワード重要度総和と、当該発話の後方１以上５以下の発話内に存在する間投詞の数である間投詞数と、の素性のうち少なくとも１つを含む素性値を、レコードごとの入力とし、かつ、質問者による質問に相当するＱ候補、又は回答者による回答に相当するＡ候補であるか否かを示す候補ラベルを出力として、分類器を教師あり学習によりトレーニングしたパターン認識モデルが知られている（特許文献１参照）。

10

【0003】

また、従来、文書クラスタリングを行って、全文検索結果により目的の文書への到達を支援する情報処理装置が知られている。このような情報処理装置として、文書の全文検索を行なって検索条件に該当する検索文書を検索する文書検索手段と、前記検索条件との適合度を表す適合度スコアによって前記検索文書を順序付ける文書スコアリング手段と、前記文書を識別するための文書識別子と、前記文書の全文データから抽出された特徴語と、当該特徴語の重みを表す重み値とが対応付けられて登録された特徴語ファイルデータベースと、前記検索文書の前記文書識別子をキーとして、前記特徴語ファイルデータベースから取得した前記特徴語と当該特徴語の重み値とに基づいて、前記検索文書に対してクラスタリング処理を行う文書クラスタリング手段と、を有し、検索結果一覧を作成する情報処理装置であって、前記適合度スコアに基づいて前記検索文書をグループ化する文書グループ化手段を有し、前記文書クラスタリング手段は、前記文書グループ化手段によりグループ化されたグループごとに、そのグループ内の前記検索文書に対してクラスタリング処理を行う情報処理装置が知られている（特許文献２参照）。

20

30

【先行技術文献】

【特許文献】

【0004】

【特許文献１】特開２０２０－７１６９０号公報

【特許文献２】特開２０１０－９５７７号公報

【発明の概要】

【発明が解決しようとする課題】

【0005】

40

特許文献１のパターン認識モデルを用いたＦＡＱシステムである場合、及び特許文献２の情報処理装置をＦＡＱシステムに適用した場合には、ＦＡＱシステムの利便性が不十分であり、改善の余地がある。

【0006】

本開示は、ＦＡＱシステムの利便性を向上できるデータ管理装置、データ管理システム、データ管理方法、プログラムを提供する。

【課題を解決するための手段】

【0007】

本開示の一態様は、質問文及び前記質問文に対する回答文を管理するデータ管理装置であって、制御部と通信部とを備え、前記制御部は、複数の文章データを取得し、前記複数

50

の文章データをベクトル化して複数の文章ベクトルを作成し、前記複数の文章ベクトルに基づいて、前記複数の文章データから複数の質問文を抽出し、前記複数の質問文に対応する複数の文章ベクトルである複数の質問文ベクトルに基づいて、抽出された前記複数の質問文を、相互に類似する質問文が含まれる質問クラスタにクラスタリングし、前記質問クラスタ毎に、前記質問クラスタの中心に対応する前記質問クラスタに属する質問文を代表する代表質問文を決定し、前記通信部は、決定された前記代表質問文を記憶装置に登録することを推奨する登録推奨情報を、前記データ管理装置を管理する管理者が使用する管理者端末へ送信する、データ管理装置である。

【0008】

本開示の一態様は、質問文及び前記質問文に対する回答文を管理するデータ管理装置と、データ管理装置を管理する管理者端末と、を備えるデータ管理システムであって、前記データ管理装置は、複数の文章データを取得し、前記複数の文章データをベクトル化して複数の文章ベクトルを作成し、前記複数の文章ベクトルに基づいて、前記複数の文章データから複数の質問文を抽出し、前記複数の質問文に対応する複数の文章ベクトルである複数の質問文ベクトルに基づいて、抽出された前記複数の質問文を、相互に類似する質問文が含まれる質問クラスタにクラスタリングし、前記質問クラスタ毎に、前記質問クラスタの中心に対応する前記質問クラスタに属する質問文を代表する代表質問文を決定し、決定された前記代表質問文を記憶装置に登録することを推奨する登録推奨情報を、前記データ管理装置を管理する管理者が使用する管理者端末へ送信し、前記管理者端末は、前記登録推奨情報を受信して表示する、データ管理システムである。

【0009】

本開示の一態様は、質問文及び前記質問文に対する回答文を管理するデータ管理方法であって、プロセッサにより、複数の文章データを取得するステップと、前記複数の文章データをベクトル化して複数の文章ベクトルを作成するステップと、前記複数の文章ベクトルに基づいて、前記複数の文章データから複数の質問文を抽出するステップと、前記複数の質問文に対応する複数の文章ベクトルである複数の質問文ベクトルに基づいて、抽出された前記複数の質問文を、相互に類似する質問文が含まれる質問クラスタにクラスタリングするステップと、前記質問クラスタ毎に、前記質問クラスタの中心に対応する前記質問クラスタに属する質問文を代表する代表質問文を決定するステップと、決定された前記代表質問文を記憶装置に登録することを推奨する登録推奨情報を、前記データ管理装置を管理する管理者が使用する管理者端末へ送信するステップと、を有するデータ管理方法である。

【0010】

本開示の一態様は、上記のデータ管理方法をコンピュータに実行させるプログラムである。

【発明の効果】

【0011】

本開示によれば、FAQシステムの利便性を向上できる。

【図面の簡単な説明】

【0012】

【図1】本開示の第1の実施形態に係るFAQ処理システムの構成例を示す模式図

【図2】FAQサーバの構成例を示すブロック図

【図3】FAQチャット画面の概要を説明する図

【図4】FAQ処理システムの動作概要を示すシーケンス図

【図5】問合せメールの一例を示す図

【図6】問合せメールの分類の一例を示す図

【図7】導出された質問文のクラスタリングと、クラスタの中心データに対応するQAデータの登録推奨情報と、の一例を示す図

【図8】FAQサーバによるFAQナレッジベースへのQA登録支援に係る動作例を示すフローチャート

10

20

30

40

50

【図 9】 Q A テキスト抽出処理の一例を示すフローチャート
【図 10】 代表質問文に対応する代表回答文の抽出の他例を示すフローチャート
【図 11 A】 本開示の第 2 の実施形態に係る F A Q サーバの構成例を示すブロック図
【図 11 B】 質問の絞り込みが行われる F A Q チャット画面のイメージを示す図
【図 12】 第 2 の実施形態におけるクラスタリングのイメージを示す図
【図 13】 Q 番号及び質問候補文を含む検索結果（候補）の一例を示す図
【図 14】 質問候補文をベクトル化した結果の一例を示す図
【図 15】 ベクトルをクラスタリングし、各クラスタと質問候補文との一例を示す図
【図 16】 各クラスタの質問候補文をベクトル化した結果の一例を示す図
【図 17】 クラスタ毎に抽出された特徴語の一例を示す図
【図 18】 質問の絞り込みを考慮した F A Q チャット画面の一例を示す図
【図 19】 各特徴語の関係性の一例を示すネットワーク図
【図 20】 第 3 の実施形態に係る F A Q サーバの構成例を示すブロック図
【図 21】 メール宛先の対応関係と、メール本文での確認必要部分と、を強調表示した情報の一例を示す図

10

【発明を実施するための形態】

【0013】

以下、適宜図面を参照しながら、実施形態を詳細に説明する。但し、必要以上に詳細な説明は省略する場合がある。例えば、既によく知られた事項の詳細説明や実質的に同一の構成に対する重複説明を省略する場合がある。これは、以下の説明が不必要に冗長になることを避け、当業者の理解を容易にするためである。尚、添付図面及び以下の説明は、当業者が本開示を十分に理解するために提供されるものであり、これらにより特許請求の範囲に記載の主題を限定することは意図されていない。

20

【0014】

（第 1 の実施形態を得るに至った経緯）

特許文献 1 のパターン認識モデルを用いた F A Q システムでは、テキストを形態素解析した後に、人が設計したような特徴量を質問文 Q や回答文 A の候補として抽出するので、抽出精度が不十分である。また、F A Q を作成する元となるメールやチャットログに含まれる不要なテキストが、Q 候補又は A 候補内に混じることがある。また、人手で Q や A の代表的な質問群を抽出する場合には、工数がかかる。第 1 の実施形態では、上記の点を改善する。

30

【0015】

（F A Q システムの構成）

図 1 は、本開示の第 1 の実施形態に係る F A Q システム 5 のブロック図である。F A Q システム 5 は、F A Q サーバ 100 と、F A Q ナレッジベース 150 と、利用者端末 200 と、管理者端末 300 と、を含む構成である。F A Q サーバ 100 と利用者端末 200 と管理者端末 300 とは、例えば、有線又は無線により通信可能に接続される。F A Q システム 5 は、データ管理システムの一例である。

【0016】

F A Q サーバ 100 は、F A Q に関する処理を行う装置であり、後述するように、F A Q ナレッジベース 150 への登録支援処理、質問の絞り込み処理、等を行う。F A Q サーバ 100 は、質問文や回答文の検索を行う検索機能を有する。F A Q サーバ 100 は、F A Q チャットボット C B としても動作可能である。F A Q チャットボット C B は、利用者からの質問を受け付け、F A Q ナレッジベース 150 を参照して、質問に対する回答を行う。

40

【0017】

F A Q ナレッジベース 150 は、F A Q に関するナレッジマネジメントのためのデータベースであり、例えばデータベースサーバとして構築される。F A Q ナレッジベース 150 は、一般的なサーバと同様に、例えば、制御機能、記憶機能、及び通信機能を有してよい。なお、F A Q ナレッジベース 150 は、F A Q サーバ 100 に含まれてもよく、F A

50

Qサーバ100と一体で構成されてもよい。

【0018】

利用者端末200は、利用者（ユーザ）に使用される端末である。この利用者は、FAQシステム5が提供するFAQサービスを利用する者（つまり質問者）である。管理者端末300は、管理者（オペレータ）に使用される端末である。この管理者は、FAQシステム5が提供するFAQサービスを管理する者である。管理者は、質問者による質問に対する回答を行う回答者を含んでもよい。また、管理者は回答者を含まず、回答者が使用する回答者端末が、管理者端末とは別に設けられてもよい。ここでは一例として、管理者に回答者が含まれることを想定する。利用者端末200及び管理者端末300は、直接に通信可能である。

10

【0019】

FAQサーバ100は、利用者端末200から質問文を受信し、質問文に対する回答を試みる。FAQサーバ100は、質問文に対して回答可能である場合、利用者端末200に対して回答文を送信する。一方、質問に対して回答不可能である場合、利用者端末200と管理者端末300とは、メール機能又はチャット機能を用いて質問文とこの質問に対する回答文とを通信する。管理者端末300と利用者端末200とは、FAQサーバ100を経由して通信されてもよい。

【0020】

図2は、FAQサーバ100の構成例を示すブロック図である。FAQサーバ100は、制御部110と、記憶部130と、通信部140と、を備える。

20

【0021】

制御部110は、例えばプロセッサにより構成され、記憶部130に保持されたプログラムを実行することで、各種機能を実現する。プロセッサは、MPU（Micro processing Unit）、CPU（Central Processing Unit）、DSP（Digital Signal Processor）、等を含んでよい。プロセッサは、各種集積回路（例えばLSI（Large Scale Integration）又はFPGA（Field Programmable Gate Array））等を含んでよい。制御部110は、FAQサーバ100の各部の動作を統括し、各種処理を行う。例えば、制御部110は、FAQに関する処理を行う。

【0022】

記憶部130は、一次記憶装置（例えばRAM（Random Access Memory）やROM（Read Only Memory））を含む。記憶部130は、二次記憶装置（例えばHDD（Hard Disk Drive）やSSD（Solid State Drive））や三次記憶装置（例えば光ディスク、SDカード）を含んでよい。記憶部130は、その他の記憶装置を含んでよい。記憶部130は、FAQサーバ100から取り外し可能であってもよい。記憶部130は、各種データ、情報、又はプログラム等を記憶する。

30

【0023】

記憶部130は、非チャットログデータベース（DB）131と、チャットログデータベース（DB）132と、QAデータベース（DB）133と、を含む。

【0024】

非チャットログデータベース131は、チャットログ以外の問合せログを保持する。チャットログ以外の問合せのログである非チャットログは、例えば、メール又は電話等による問合せに関するログ（履歴情報）である。非チャットログは、電子メール本文又は電話の内容を示すデータも含んでよい。

40

【0025】

チャットログデータベース132は、チャットによる問合せのログ（履歴情報）を保持する。チャットログは、チャット本文も含んでよい。

【0026】

このように、非チャットログデータベース131及びチャットログデータベース132には、過去に利用者と回答者の間で通信された各種データやメールが保持される。

【0027】

50

Q A データベース 1 3 3 は、チャットログ又は非チャットログを基に得られた Q A データを保持する。Q A データは、質問文と、その質問文に対する回答文と、を含む。また、Q A データベース 1 3 3 は、Q A データに付随するデータも保持してよく、後述するクラスタリングでのマッピング結果、文章データから得られた質問文の特徴ベクトル（質問文ベクトル）、回答文の特徴ベクトル（回答文ベクトル）、Q A データを参照可能なリンクの情報、等を保持してもよい。クラスタリングでのマッピング結果や質問文ベクトルや回答文ベクトルは、複数の Q A データの類似性を判断する判断材料となる。

【0028】

通信部 1 4 0 は、各種データ又は情報等を通信する。通信部 1 4 0 による通信方式は、有線通信又は無線通信の方式でよく、L A N（Local Area Network）、W A N（Wide Area Network）、電力線通信、又は携帯電話用のセルラー通信（例えば 5 G、L T E（Long Term Evolution））等の通信方式を含んでよい。通信部 1 4 0 は、例えば、F A Q ナレッジベース 1 5 0、利用者端末 2 0 0、及び管理者端末 3 0 0 と通信可能であり、その他の通信装置と通信可能であってもよい。

10

【0029】

制御部 1 1 0 は、機能構成として、文章抽出部 1 1 1 と、単語分割部 1 1 2 と、ベクトル化部 1 1 3 と、分類部 1 1 4 と、ベクトル類似度計算部 1 1 5 と、クラスタリング部 1 1 6 と、クラスタ中心計算部 1 1 7 と、固有表現抽出部 1 1 9 と、匿名化処理部 1 2 1 と、を含む構成である。

【0030】

文章抽出部 1 1 1 は、記憶部 1 3 0 に保持されたチャットログ又は非チャットログから、文章データ（テキストデータ）を抽出する。単語分割部 1 1 2 は、例えば自然言語処理モデルを用いて、抽出された文章データを形態素解析により単語単位に分割する。なお、制御部 1 1 0 は、単語単位に限らず文字やサブワード等の単位に分割してよい。

20

【0031】

ベクトル化部 1 1 3 は、自然言語処理モデル（例えば、B E R T（Bidirectional Encoder Representations from Transformers）又はその他のモデル）を用いて、抽出された文章データをベクトル化し、特徴ベクトル（文章ベクトル）を生成する。

【0032】

分類部 1 1 4 は、自然言語処理モデルを用いて、生成された文章ベクトルに基づいて、文章データを分類する。分類部 1 1 4 は、例えば、予め分類のためのアノテーションを実施しておき、抽出された文章データを分類し、文章データから質問文と回答文とを抽出する。分類部 1 1 4 は、抽出された質問文と回答文とに対して Q A 対応付け処理を行って、質問文とその回答文とを含む Q A データを生成する。Q A 対応付け処理では、分類部 1 1 4 は、記憶部 1 6 0 に保持されたメール本文中に質問文が 1 つである場合、管理者からの返答メールから回答文を抽出し、質問文とその回答文とを対応付けてよい。分類部 1 1 4 は、記憶部 1 6 0 に保持されたメール本文中に質問文が複数存在する場合、各質問文に、この各質問メールに対する各回答メールを対応付けてよい。

30

【0033】

なお、文章ベクトルのうち、質問文がベクトル化されたものを質問文ベクトルとも称し、回答文がベクトル化されたものを回答文ベクトルとも称する。また、後述する質問文の候補（質問候補文）がベクトル化されたものを質問候補文ベクトルとも称する。

40

【0034】

ベクトル類似度計算部 1 1 5 は、自然言語処理モデルを用いて、ベクトル化部 1 1 3 により生成された各文章ベクトルの類似度（相関度）を計算する。

【0035】

クラスタリング部 1 1 6 は、所定のクラスタリング手法（例えば k - M e a n s 法又はその他のクラスタリング手法）を用いて、導出された各文章ベクトル（各質問文ベクトル）の類似度に基づいて、質問文をクラスタリングし、複数のクラスタ（質問クラスタ）に分割する。クラスタリングにより、複数の質問文が複数の質問クラスタのいずれかに属す

50

るように分類される。同様に、所定のクラスタリング手法（例えば k-Means 法）を用いて、導出された各文章ベクトル（各回答文ベクトル）の類似度に基づいて、回答文をクラスタリングし、複数のクラスタ（回答クラスタ）に分割する。また、回答文は、質問文に対応付けてクラスタリングされてよい。クラスタリングにより、複数の回答文が複数の回答クラスタのいずれかに属するように分類される。

【0036】

クラスタリングでは、教師データが不要である。したがって、FAQサーバ100は、事前学習が不要で、高速に、例えば質問文及び回答文の分類が可能である。

【0037】

クラスタ中心計算部117は、所定のクラスタリング手法に基づいて、クラスタ毎に、クラスタ中心のデータを抽出する。つまり、クラスタ中心計算部117は、クラスタ中心にある質問文を代表質問文QRとして導出する。なお、クラスタ中心計算部117は、k-Means 法以外の方法で、クラスタ中心のデータを抽出してよい。

【0038】

固有表現抽出部119は、例えば自然言語処理モデルを用いて、文章データに含まれる単語を解析し、人名、会社名、住所、又は電話番号等の固有名詞（固有表現）を抽出する。

【0039】

匿名化処理部121は、抽出された固有表現に対して匿名化処理を行う。匿名化処理は、個人を特定可能な情報を削除又は変更するデータ処理である。匿名化処理は、例えば、固有表現を汎用的又は一般的な名称に変更すること、固有表現を特定の表現（例えばXXX）に変更すること、固有表現を削除すること、又はその他の匿名にする処理を含んでよい。

【0040】

次に、利用者端末200及び管理者端末300の構成について説明する。

【0041】

利用者端末200及び管理者端末300は、例えば、PC（Personal Computer）、スマートフォン、又はタブレット端末を含む。利用者端末200及び管理者端末300は、例えば、制御部、記憶部、通信部、操作部、及び表示部を備える。

【0042】

次に、FAQシステム5の全体概要について説明する。

図3は、FAQチャット画面の概要を説明する図である。

【0043】

FAQナレッジベース150は、QAデータを保持する。利用者端末200が質問文QをFAQチャットボットCBに送信し、この質問文Qに対する回答文Aを得られなかった場合、利用者端末200は、管理者端末300へ質問文Qを送信し、Q&Aとして問合せする。この場合、FAQサーバ100の制御部110は、利用者端末200からの質問文Qと管理者端末300からの回答文Aを分析し、新たにQAデータを生成する。生成されたQAデータは、FAQナレッジベース150に登録可能である。これにより、FAQナレッジベース150に登録されたQAデータが増加するので、FAQチャットボットCBが回答可能な質問の数を増やすことができる。FAQサーバ100は、管理者の意思により管理者端末300を用いてによりFAQナレッジベース150へQAデータを登録する場合、FAQナレッジベース150への登録支援処理を行う（第1の実施形態）。

【0044】

また、FAQサーバ100は、任意の質問文Q（Q0）を受け、FAQナレッジベース150に保持されたQAデータから、この質問に関連するFAQを検索する。検索結果の候補（結果候補）は質問候補文QCとして、複数のQAデータが得られる。結果候補は、質問によって変動する。そのため、FAQサーバ100は、結果候補を絞り込む特徴語（キーワード）KWを結果候補に応じて都度作成することで、特徴語KWを選定する。このように、FAQサーバ100は、質問の絞り込み処理を行う（第2の実施形態）。

【 0 0 4 5 】

図 4 は、F A Q システム 5 の動作概要を示すシーケンス図である。

【 0 0 4 6 】

まず、F A Q サーバ 1 0 0 では、F A Q チャットボット C B を用いて利用者からの質問文 Q に対して回答文 A を生成するために、制御部 1 1 0 は、事前に（F A Q チャットボット C B の運用前に）F A Q の質問文 Q と回答文 A とを含む Q A データを F A Q ナレッジベース 1 5 0 に登録しておく（S 1 1）。なお、F A Q チャットボット C B の運用中に F A Q の Q A データを F A Q ナレッジベース 1 5 0 に登録していてもよい。なお、F A Q の質問文 Q と回答文 A とのいずれか一方のみが F A Q ナレッジベース 1 5 0 に登録されているものが存在してもよい。

10

【 0 0 4 7 】

利用者端末 2 0 0 は、例えば F A Q チャット画面 G 1 を用いて、F A Q チャットに任意の質問文 Q 0 を入力する（S 2 1）。利用者端末 2 0 0 は、この質問文 Q 0 を F A Q サーバ 1 0 0（F A Q チャットボット C B）に送信する。

【 0 0 4 8 】

F A Q サーバ 1 0 0 では、通信部 1 4 0 が質問文 Q 0 を受信し、制御部 1 1 0 が、質問文 Q 0 に関連する質問候補文 Q C を F A Q ナレッジベース 1 5 0 から抽出する（S 1 2）。制御部 1 1 0 は、抽出された質問候補文 Q C の中で類似する質問候補を含む質問候補群（質問候補クラスタ）を生成する（S 1 3）。制御部 1 1 0 は、質問候補群の分類を示す単語（例えば分類中の特徴を示す特徴語 K W）を設定する（S 1 4）。制御部 1 1 0 は、通信部 1 4 0 を介して、設定された特徴語 K W の表示を利用者端末 2 0 0 に指示する（S 1 5）。

20

【 0 0 4 9 】

利用者端末 2 0 0 は、F A Q サーバ 1 0 0 から 1 つ以上の特徴語 K W と特徴語 K W の表示の指示を受信し、この指示に従って特徴語 K W を表示部に表示する。利用者端末 2 0 0 は、例えば操作部を介して利用者による選択操作を受け付けて、受信された各特徴語 K W のうち、利用者が一番問い合わせたい内容に近い特徴語 K W を選択する（S 2 2）。利用者端末 2 0 0 は、選択された特徴語 K W を F A Q サーバ 1 0 0 に送信する。

【 0 0 5 0 】

F A Q サーバ 1 0 0 では、通信部 1 4 0 が利用者端末 2 0 0 から選択された特徴語 K W を受信する。制御部 1 1 0 は、選択された特徴語 K W に対応する質問候補群に含まれる質問候補文 Q C の表示を、利用者端末 2 0 0 に指示する（S 1 6）。利用者端末 2 0 0 は、F A Q サーバ 1 0 0 から、選択された特徴語 K W に対応する質問候補群に含まれる質問候補文 Q C とその表示の指示を受信し、受信された質問を表示する。

30

【 0 0 5 1 】

または、F A Q サーバ 1 0 0 は、選択された特徴語 K W に対応する質問候補群に含まれる質問が多数存在する場合、この質問候補群の中から、ステップ S 1 3 と同様に、類似する質問候補を含む質問候補群を再度作成する（S 1 6）。ステップ S 1 3 における質問候補群の作成と比較すると、ステップ S 1 6 における質問候補群の再作成では、質問候補の関連性の範囲が狭くなっており、質問が絞り込まれている。なお、質問候補群を再作成する場合には、ステップ S 1 3 に後続するステップ S 1 4 ～ S 1 6 の処理を再度実施することになる。

40

【 0 0 5 2 】

なお、ステップ S 1 1 は、F A Q ナレッジベース 1 5 0 への登録推奨処理に関連し、第の実施形態で詳細に説明する。ステップ S 1 2 ～ S 1 6 及びステップ S 2 1，S 2 2 は、質問の絞り込み処理に関連し、第 2 の実施形態で詳細に説明する。

【 0 0 5 3 】

次に、F A Q ナレッジベース 1 5 0 への登録支援処理について詳細に説明する。

【 0 0 5 4 】

図 5 は、問合せメール M 1 の一例を示す図である。図 6 は、問合せメール M 1 の分類の

50

一例を示す図である。問合せメールM1は、質問文Qを含む質問メールである。図7は、導出された質問文Qのクラスタリングと、クラスタの中心データに対応するQAデータの登録推奨情報と、の一例を示す図である。

【0055】

問合せメールM1は、例えば、FAQサーバ100が任意の利用者の利用者端末200から取得され、記憶部130の非チャットログデータベース131に保持されている。文章抽出部111は、この問合せメールM1から文章データを抽出する。文章抽出部111は、問合せメールM1に含まれる引用文及び署名等を削除したり、問合せメールM1に含まれる文章データを正規化したりする等、前処理を行ってもよい。ベクトル化部113は、抽出された文章データをベクトル化し、文章ベクトルを生成する。分類部114は、生成された文章ベクトルに基づいて、文章データを分類する。問合せメールM1は、例えば、図6に示されるような、利用者による挨拶を含む挨拶文、質問を含む質問文Q、質問の背景を含む背景文、メール末尾に添えられるよろしく文、等に分類される。つまり、問合せメールM1から質問文Qが抽出可能である。

10

【0056】

同様に、上記の問合せメールM1に対する回答メールが、例えば、FAQサーバ100が任意の管理者端末300から取得し、記憶部130の非チャットログデータベース131に保持されている。文章抽出部111は、この回答メールから文章データを抽出する。文章抽出部111は、回答メールに含まれる引用文及び署名等を削除したり、回答メールに含まれる文章を正規化したりする等、前処理を行ってもよい。ベクトル化部113は、抽出された文章を特徴ベクトル化し、文章ベクトルを生成する。分類部114は、自然言語処理モデルを用いて、生成された文章ベクトルに基づいて、文章データを分類する。回答メールは、図示はしていないが、回答者による挨拶を含む挨拶文、質問文Qに対する回答を含む回答文A、回答の理由を含む理由文、メール末尾に添えられるよろしく文、等に分類される。なお、回答文Aに理由文が含まれてもよい。つまり、回答メールから回答文Aが抽出可能である。

20

【0057】

制御部110は、問合せメールM1から導出された質問文Qと、この問合せメールM1に対する回答メールから導出された回答文Aと、をペアとして含むQAデータを生成し、QAデータベース133に保持させる。このように、制御部110は、非チャットログデータベース131に蓄積された多数の問合せメールM1及び多数の回答メールから、多数のQAデータを生成し、QAデータベース133に保持させてよい。また、ここでは、非チャットログであるメールを基にQAデータを生成することを例示したが、制御部110は、同様にして、チャットログデータベース132に蓄積された問合せのチャットログ及び回答のチャットログを基に、QAデータを生成し、QAデータベース133に保持させてもよい。これにより、過去の実績のあるQAデータが、QAデータベース133に多数保持されることになる。

30

【0058】

なお、ここでQAデータベース133に保持されたQAデータは、粗い状態のQAデータである。具体的には、QAデータベース133には、内容的に類似したQAデータも多く含まれ得る。

40

【0059】

クラスタリング部116は、ベクトル化部113により得られた質問文ベクトルに基づいて、質問文Qをクラスタリングする。図7では、質問クラスタC1、C2、C3のマッピング情報MPが例示されている。なお、図7では、二次元のクラスタリングの結果が示されているが、2次元以上（3次元等）のクラスタリングでもよい。クラスタ中心計算部117は、質問クラスタ毎に、クラスタ中心のデータを抽出する、つまり、クラスタ中心計算部117は、クラスタ中心の質問文Qを代表的な質問である代表質問文QRとして導出する。

【0060】

50

同様に、クラスタリング部 116 は、ベクトル化部 113 により得られた回答文ベクトルに基づいて、質問文 Q に対する回答文 A をクラスタリングする。クラスタ中心計算部 117 は、回答クラスタ毎に、クラスタ中心の回答文 A を代表的な回答である代表回答文 A R として導出する。代表質問文 Q R と代表回答文 A R とを含めて代表 Q A データとする。なお、質問文 Q に対する回答文 A が複数存在する場合には、回答文 A のクラスタリングを行うが、質問文 Q に対する回答文 A が 1 つである場合には、回答文 A のクラスタリングは不要であり、1 つの回答文 A が代表回答文 A R となる。なお、同様に、クラスタリング部 116 が質問文 Q についてもクラスタリングを行い、クラスタ中心計算部 117 が代表質問文 Q R を算出してよい。

【0061】

通信部 140 は、代表 Q A データを F A Q ナレッジベース 150 に登録することを推奨する登録推奨情報 I 1 を表示することを、管理者端末 300 に指示する。管理者端末 300 は、登録推奨情報 I 1 の表示の指示を受信し、この指示に従って登録推奨情報 I 1 を表示する。図 7 では、質問クラスタ C 3 の代表質問文 Q R とその代表回答文 A R とが表示されている。

【0062】

管理者端末 300 は、例えば、操作部を介して、代表 Q A データを確認した管理者により代表 Q A データを登録することが指定された場合、代表 Q A データを F A Q ナレッジベース 150 に登録する（保持させる）。これにより、F A Q ナレッジベース 150 に保持されていなかった Q A データが追加される。よって、F A Q チャットボット C B は、この Q A データに基づいて、利用者端末 200 からの質問に対して回答可能である。

【0063】

このようにして、F A Q サーバ 100 は、制御部 110 が、Q A データベース 133 に保持された類似した質問文 Q、類似した回答文 A のうち、各クラスタの中心のデータを抽出する。各クラスタの中心のデータは、類似する質問文 Q や回答文 A が含まれる各クラスタにおける代表的な質問文 Q や回答文 A であることが想定される。つまり、F A Q サーバ 100 は、クラスタ中心にある代表質問文 Q R 及び代表回答文 A R を抽出可能である。よって、F A Q ナレッジベース 150 では、類似した質問文 Q のうち代表質問文 Q R と、類似した回答文 A のうち代表回答文 A R と、が登録されることになる。つまり、F A Q サーバ 100 は、余分な質問文 Q 及び回答文 A が F A Q ナレッジベース 150 に登録されることを抑制できる。

【0064】

図 8 は、F A Q サーバ 100 による F A Q ナレッジベース 150 への Q A 登録支援に係る動作例を示すフローチャートである。

【0065】

まず、制御部 110 は、非チャットログデータベース 131 からメールを入力する（S 31）。制御部 110 は、チャットログデータベース 132 からチャットログを入力する（S 31）。制御部 110 は、Q A テキスト抽出処理を行う（S 32）。Q A テキスト抽出処理では、入力されたメール等から質問文 Q 及び回答文 A を抽出する。

【0066】

図 9 は、ステップ S 32 における Q A テキスト抽出処理の一例を示すフローチャートである。Q A テキスト抽出処理では、ベクトル化部 113 は、メール又はチャットログ毎に、メール中又はチャットログ中の文章データをベクトル化し、文章ベクトルを生成する（S 41）。分類部 114 は、生成された文章ベクトルに基づいて、文章データを分類する（S 42）。文章データの分類により、例えば、質問文 Q、回答文 A、等が得られる。制御部 110 は、得られた質問文 Q と、質問文 Q に対応する回答文 A（対応回答文）と、質問文 Q の特徴ベクトル（質問文ベクトル）及び回答文 A の特徴ベクトル（回答文ベクトル）と、を Q A データベース 133 に登録する（S 43）。

【0067】

図 8 に戻り、クラスタリング部 116 は、Q A データベース 133 に登録された質問文

10

20

30

40

50

ベクトルをクラスタリングする（S 3 3）。クラスタ中心計算部 1 1 7 は、クラスタリングにより得られた各質問クラスタの各質問文 Q から、各クラスタの中心となる質問文 Q（代表質問文 Q R）を抽出する（S 3 4）。クラスタ中心計算部 1 1 7 は、抽出された代表質問文 Q R に対応する回答文 A（代表回答文 A R）を、例えば Q A データベース 1 3 3 から抽出する（S 3 5）。通信部 1 4 0 は、管理者が使用する管理者端末 3 0 0 へ、代表質問文 Q R 及び代表回答文 A R を表示し、F A Q ナレッジベースへの登録を推奨する登録推奨情報を表示するよう、指示する（S 3 6）。

【0 0 6 8】

なお、図 8 に示したステップ S 3 5 の処理ではなく、他の処理によって代表回答文 A R を抽出してもよい。図 1 0 は、代表質問文 Q R に対応する代表回答文 A R の抽出の他例を示すフローチャートである。例えば、質問文クラスタには、1 つ以上の質問文 Q（質問文ベクトル）が属するだけでなく、1 つ以上の回答文 A（回答文ベクトル）も含まれてもよい。クラスタリング部 1 1 6 は、各質問クラスタ内の回答文 A の特徴ベクトルをクラスタリングし、クラスタリングにより得られた回答クラスタの中心となる回答文 A を、代表回答文 A R として抽出してよい（S 3 5 A）。

10

【0 0 6 9】

なお、問合せメールにおいて質問文 Q が 1 つだけ含まれていた場合、制御部 1 1 0 は、この問合せメールに対する回答メールから回答文 A を抽出し、この質問文 Q と回答文 A とを対応付けて、Q A データを生成してよい。また、問合せメールにおいて質問文 Q が複数含まれていた場合、制御部 1 1 0 は、非チャットログデータベース 1 3 1 に保持された複数の回答文 A を参照し、管理者端末 3 0 0 の操作部を介して、人手（つまり管理者）によって各質問文 Q と各回答文 A とを対応付けて、複数の Q A データを生成してよい。

20

【0 0 7 0】

つまり、F A Q ナレッジベース 1 5 0 への Q A 登録支援では、以下の処理が行われるとも言える。制御部 1 1 0 は、メール又はチャットログ等から、質問文 Q と回答文 A のテキストのみを抽出し、粗い F A Q 情報を作成し、Q A データベース 1 3 3 に保持させる。制御部 1 1 0 は、粗い F A Q 情報の質問文 Q をクラスタリングすることで、類似する質問文 Q をグループ化（クラスタリング）する。制御部 1 1 0 は、グループ（クラスタ）の中心となる代表質問文 Q R とこの代表質問文 Q R に対応する代表回答文 A R を、管理者へ提示するよう指示する。また、制御部 1 1 0 は、同じ質問文グループの回答文 A をクラスタリングすることで、代表回答文 A R を管理者へ提示してもよい。

30

【0 0 7 1】

なお、登録推奨情報を管理者端末 3 0 0 へ送信する前に、F A Q サーバ 1 0 0 のベクトル類似度計算部 1 1 5 が、代表質問文 Q R が過去の質問文（Q A データベース 1 3 3 に保持された Q A データの質問文 Q）と一致しない又は類似しない（類似度が低い）か否かを判定してもよい。類似しないとは、例えば、両者の類似度が所定閾値以下であることである。代表質問文 Q R が過去の質問文（Q A データベース 1 3 3 に保持された Q A データの質問文 Q）と一致しない又は類似しない場合に、登録推奨情報を送信し、一致する又は類似する場合に、登録推奨情報を送信しなくてよい。これにより、F A Q サーバ 1 0 0 は、管理者によって過去の質問文 Q を含む Q A データと類似する代表 Q A データが F A Q ナレッジベース 1 5 0 に登録されることを抑制できる。

40

【0 0 7 2】

次に、本実施形態の F A Q システム 5 のバリエーションについて説明する。

【0 0 7 3】

F A Q サーバ 1 0 0 では、制御部 1 1 0 は、以下を実施してもよい。例えば、制御部 1 1 0 は、自然言語処理モデルを用いて、1 つ以上の質問文 Q を要約してよい。これにより、質問文 Q における冗長な情報を減らし、高速に読むことができる質問文 Q が得られる。

【0 0 7 4】

また、制御部 1 1 0 は、自然言語処理モデルを用いて、1 つ以上の質問文 Q を日本語の構造として正しい質問文 Q に作り直してよい。これにより、分かりやすい質問文 Q が得ら

50

れる。

【0075】

また、制御部110は、固有表現抽出部119が、例えば、自然言語処理モデルを用いて、質問文Qの単語の特徴ベクトルに基づいて、質問文Qに含まれる所定の表現（例えば誤字脱字チェック、コンプライアンス違反、不快な表現、又は金額の情報）を検出してよい。制御部110は、検出された表現の箇所を修正してもよい。

【0076】

また、制御部110は、固有表現抽出部119が、例えば、自然言語処理モデルを用いて、質問文Qの単語の特徴ベクトルに基づいて、質問文Qに含まれる個人情報（例えば名前、住所又は電話番号）を検出してよい。匿名化处理部121は、検出された個人情報に対して匿名化处理（例えば削除）してよい。これにより、個人情報を含まない質問文Qが得られる。この場合、FAQサーバ100は、通信部140が、個人情報の匿名化处理が行われた旨を示す匿名化情報を管理者端末300に送信してよい。管理者端末300は、匿名化情報を受信すると、個人情報が含まれる質問文Qと個人情報が含まれない質問文Qとのどちらかを選択するための選択画面を表示部に表示してよい。管理者端末300は、操作部を介して、選択画面からこの選択の操作を行い、選択の操作に応じて、FAQナレッジベース150に個人情報が含まれる質問文Qを含む代表QAデータか、個人情報が含まれない質問文Qを含む代表QAデータか、のいずれかを登録するよう推奨してよい。また、通信部140は、個人情報の匿名化处理後に、登録推奨情報を管理者端末300へ送信するようにしてよい。

【0077】

利用者がFAQチャットボットCBに質問して回答を得ようとする場合、事前処理として、上記のようなFAQナレッジベース150への登録支援処理を用いて、FAQナレッジベース150に代表QAデータを登録しておく。FAQチャットボットCBは、FAQナレッジベース150を参照し、利用者端末200から回答可能な質問文Qを受信した場合、質問文Qに対する回答文Aを利用者端末200に送信する。FAQチャットボットCBは、FAQナレッジベース150を参照し、利用者端末200から回答不可能な質問文Qを受信した場合、質問文Qに対して回答不可能である旨を利用者端末200へ回答する。

【0078】

そして、利用者が、利用者端末200の表示部を介して質問文Qに対して回答を得られないことを確認すると、利用者端末200が、メール又はチャット等によってこの質問文Qを管理者端末300へ送信する（以降、「直接質問」とも称する）。また、利用者がFAQチャットボットCBによる回答を望まない場合、利用者端末200が管理者端末300へ直接質問することもある。同様に、管理者端末300が、メール又はチャット等によって受けた質問文Qに対する回答文Aを管理者端末300へ送信することを、「直接回答」とも称する。直接質問又は直接回答が行われる場合、以下のような処理が実施されてよい。

【0079】

利用者端末200が管理者端末300へ直接質問する場合、管理者端末300は、利用者端末200から質問文Qを受信すると、FAQサーバ100に質問文Qを含めて問合せる。FAQサーバ100は、制御部110が、QAデータベース133又はFAQナレッジベース150に保持されたQAデータのうち、管理者端末300からの問い合わせに含まれる質問文Qに類似する質問文Qを含むQAデータを取得し、通信部140が、管理者端末300に送信する。質問文Q同士が類似するか否かは、質問文ベクトルに基づいて、ベクトル類似度計算部115により導出されてよい。管理者端末300は、FAQサーバ100からQAデータを受信し、利用者端末200からの質問文Qに対して、QAデータに含まれる回答文Aを利用者端末200へ送信する。これにより、管理者端末300を使用する回答者は、質問者からの質問に類似した過去のQAを参考にして回答でき、回答文Aの作成工数を削減できる。

【0080】

また、利用者端末200が管理者端末300へ直接質問する場合、管理者端末300は、利用者端末200から質問文Qを受信すると、FAQサーバ100に質問文Qを含めて問合せする。FAQサーバ100は、制御部110が、QAデータベース133を参照し、問合せに含まれる質問文Qを含むQAデータ又はこの質問文Qに類似する質問文Q（類似質問文）を含むQAデータ（類似QAデータ）等を参照可能なリンク先の情報を取得し、通信部140が、管理者端末300に送信する。管理者端末300は、FAQサーバ100から上記のリンク先の情報を受信し、利用者端末200に送信する。これにより、管理者端末300を使用する回答者は、リンク先にアクセスすることでQAデータ又は類似QAデータに含まれる回答文Aを確認でき、この回答文Aを参考にして回答でき、回答文Aの作成工数を削減できる。

10

【0081】

また、利用者端末200が管理者端末300へ直接質問する場合、管理者端末300は、利用者端末200から質問文Qを受信すると、FAQサーバ100に質問文Qを含めて問合せする。FAQサーバ100は、制御部110が、問合せに含まれる質問文Qを含むQAデータが属するクラスタを特定し、QAデータベース133を参照し、このクラスタに属する他のQAデータを抽出する。また、制御部110は、抽出されたクラスタの各QAデータを二次元平面又は三次元空間にマッピングしたマッピング情報を生成し、また、ツールヒントの情報を生成する。通信部140は、生成されたマッピング情報及びツールヒントの情報を利用者端末200へ送信する。管理者端末300は、FAQサーバ100からマッピング情報及びツールヒントの情報を受信して、表示部に表示させる。これにより、管理者端末300を使用する回答者は、マッピング情報を確認することで、受けた質問が属する質問クラスタ（質問群）における各QAデータの関係性を直感的に把握できる。よって、管理者端末300を使用する回答者は、質問者からの質問に類似した過去のQAを参考にして回答でき、回答文Aの作成工数を削減できる。

20

【0082】

また、利用者端末200が管理者端末300へ直接質問する場合、管理者端末300は、利用者端末200から質問文Qを受信すると、FAQサーバ100に質問文Qを含めて問合せする。FAQサーバ100は、制御部110が、例えば、自然言語処理モデルを用いて、QAデータベース133又はFAQナレッジベース150に保持されている各質問クラスタの各QAデータの各質問文Qと、管理者端末300からの問合せに含まれる質問文Qと、の類似度を算出する。この場合、制御部110は、問合せに含まれる質問文Qの特徴ベクトルに基づいて、QAデータベース133又はFAQナレッジベース150に保持されている各QAデータの各質問文Qとの類似度を算出してよい。通信部140は、この類似度が所定閾値未満である場合に、問合せに含まれる質問文Qを含むQAデータをQAデータベース133又はFAQナレッジベース150に早期に登録することを推奨する早期登録推奨情報を、管理者端末300に送信してよい。管理者端末300は、FAQサーバ100から早期登録推奨情報を受信して表示部に表示してよい。

30

【0083】

FAQシステム5を用いたサービスが開始されると、利用者と回答者との間でのやりとりに基づいて新たなQAデータが急激に増加する。これに対し、FAQシステム5によれば、新たに利用者により質問された質問文Qが、QAデータベース133又はFAQナレッジベース150に保持されているQAデータの質問文Qとは類似度が大きく異なる場合には、回答者に知らせて、質問者からの質問文Qを含むQAデータを早期にQAデータベース133又はFAQナレッジベース150に登録するよう促すことができる。よって、FAQシステム5は、不足度の高いQAデータを優先的に登録でき、利用者による質問及び回答者による回答に係る工数を削減できる。

40

【0084】

利用者端末200と管理者端末300との間で直接質問と直接回答とが完了した後、管理者端末300は、完了した質問文Qと回答文Aとを含むQAデータを含みこのQAデー

50

タの登録を申請する登録申請情報を、FAQサーバ100に送信する。FAQサーバ100は、通信部140がQAデータを含む登録申請情報を受信し、制御部110がQAデータをQAデータベース133又はFAQナレッジベース150に保持させる。また、制御部110が、質問文Qの特徴ベクトルに基づいて質問文Qに対応する質問クラスタにクラスタリングし、この質問クラスタに対応する質問文Qの種類をタグ付けする。このタグ付けの情報は、QAデータベース133又はFAQナレッジベース150に保持されてよい。

【0085】

また、FAQサーバ100は、制御部110が、質問文Qを単語に分割し、人名又は会社名等の固有名詞（固有表現）を抽出する。そして、通信部140は、QAデータベース133又はFAQナレッジベース150に登録申請情報に係るQAデータが登録されたことを示す登録完了情報と、タグ付けされたことを示すタグ付け完了情報と、固有表現を抽出したことを示す固有表現抽出情報とを、管理者端末300に送信する。管理者端末300は、登録完了情報とタグ付け完了情報と固有表現抽出情報とを受信して表示部に表示させる。これにより、利用者及び回答者によるQAデータの登録申請意欲が向上する。また、実績のある質問文Q及び回答文Aを含むQAデータが、過去の実績情報としてQAデータベース133やFAQナレッジベース150に登録されて活用可能となる。

10

【0086】

また、利用者端末200と管理者端末300との間で直接質問と直接回答とが完了した後、管理者端末300は、完了した質問文Qと回答文Aとを含むQAデータをFAQサーバ100に送信する。FAQサーバ100は、通信部140がこの完了したQAデータを受信し、制御部110が、QAデータベース133又はFAQナレッジベース150を参照し、この完了したQAデータに類似する類似QAデータを検索する。制御部110は、類似QAデータの有無に基づいて、完了したQAデータをQAデータベース133又はFAQナレッジベース150に保持させるか否かを判定する。つまり、類似QAデータは、FAQナレッジベース150への新規登録のための判断材料となる。FAQサーバ100は、類似QAデータが存在しない場合、完了したQAデータをQAデータベース133又はFAQナレッジベース150に保持させ、類似QAデータが存在する場合、完了したQAデータをQAデータベース133又はFAQナレッジベース150に保持させない。これにより、FAQサーバ100は、QAデータベース133又はFAQナレッジベース150に完了したQAデータが登録可能であり、利用者による質問及び回答者による回答にかかる工数を削減できる。

20

30

【0087】

また、利用者端末200と管理者端末300との間で直接質問と直接回答とが完了した後、管理者端末300は、完了した質問文Qと回答文Aとを含むQAデータをFAQサーバ100に送信する。FAQサーバ100は、通信部140がこの完了したQAデータを受信し、制御部110が、QAデータに含まれる質問文Q及び回答文Aから上記の固有表現を抽出し、抽出された固有表現に対して匿名化处理し、匿名化处理されたQAデータをQAデータベース133又はFAQナレッジベース150に保持させてよい。通信部140は、匿名化处理された質問文Q及び回答文Aを含むQAデータを管理者端末300へ送信する。固有表現は、例えば、個人情報又は又はその他の秘匿化したい情報を含む。管理者端末300は、秘匿化处理されたQAデータを受信し表示部に表示させる。これにより、管理者端末300は、操作部を介して管理者による登録操作の指示を受けることで、QAデータベース133又はFAQナレッジベース150に匿名化处理されたQAデータを保持させることができる。よって、個人情報や秘匿化したい情報などを情報漏洩の可能性を低減して、QAデータベース133又はFAQナレッジベース150に登録でき、利用者による質問及び回答者による回答にかかる工数を削減できる。

40

【0088】

また、利用者端末200と管理者端末300との間で直接質問と直接回答とが完了した後、管理者端末300は、完了した質問文Qを含む問合せメールと回答文Aを含む回答メ

50

ールとをF A Qサーバ1 0 0に送信する。F A Qサーバ1 0 0は、通信部1 4 0が、完了した問合せメールと回答メールを受信し、制御部1 1 0が、これらのメールから署名情報を取得し、署名情報に含まれる所属の情報や氏名の情報に対して匿名化処理を行ってよい。そして、F A Qサーバ1 0 0は、匿名化処理が行われたメールを、非チャットログデータベース1 3 1に保持させてよい。これにより、F A Qサーバ1 0 0は、過去の問合せメールや回答メールを行った人物を特定不能な状態で、問合せメールや回答メールからF A QチャットボットC Bや回答者が必要な過去の実績のあるQ Aデータを抽出し、蓄積できる。また、個人情報等について確実に解読不能にできる。

【0089】

また、F A Qナレッジベース1 5 0にQ Aデータが登録された場合、F A Qサーバ1 0 0は、通信部1 4 0が、F A Qナレッジベース1 5 0にQ Aデータが登録されたことを示すナレッジベース登録完了情報を、管理者端末3 0 0に送信してよい。管理者端末3 0 0は、F A Qサーバ1 0 0からナレッジベース登録完了情報を受信して表示部に表示してよい。ナレッジベース登録完了情報は、例えば、「このやりとりはF A Q登録しました。〇〇〇のU R Lで確認できます」のようなメッセージを含んでよい。これにより、管理者は、正常に登録対象のQ AデータがF A Qナレッジベース1 5 0に登録されたことを確認でき、F A Qナレッジベース1 5 0を参照して回答を行うF A QチャットボットC Bが成長していることを確認できる。

【0090】

また、上記では、主に、F A QチャットボットC Bを用いた利用者からの質問文Qに対する自動回答を行う前（つまり運用前）に、事前にF A Qナレッジベース1 5 0に登録することを促すことを先に説明したが、これに限られない。例えば、利用者端末2 0 0と管理者端末3 0 0との間で直接質問と直接回答とが完了した後に、管理者端末3 0 0は、完了した質問文Qを含む問合せメールと回答文Aを含む回答メールとをF A Qサーバ1 0 0に送信する。F A Qサーバ1 0 0は、通信部1 4 0が、完了した問合せメールと回答メールを受信し、制御部1 1 0が、非チャットログデータベース1 3 1に保持させる。制御部1 1 0が、非チャットログデータベース1 3 1に保持された各メール又はチャットログデータベース1 3 2に保持された各チャットログから抽出された質問文Qをクラスタリングし、各クラスタの中心にある質問文Qを代表質問文Q Rとして抽出する。通信部1 4 0が、代表質問文Q Rと代表質問文Q Rに対する代表回答文A Rとを含む代表Q AデータをF A Qナレッジベース1 5 0に登録することを推奨する登録推奨情報を、利用者端末2 0 0に送信する。管理者端末3 0 0は、登録推奨情報を受信して表示部に表示する。これにより、F A Qシステム5は、直接質問と直接回答とが完了した後も、F A Qナレッジベース1 5 0に登録するよう推奨できる。さらに、F A Qシステム5は、F A QチャットボットC Bの運用前、運用中、及び運用後において、任意のタイミングでF A Qナレッジベース1 5 0への登録支援処理を繰り返して実施してもよい。これにより、F A Qナレッジベース1 5 0に、順次、Q Aデータ（代表Q Aデータ）が登録可能である。

【0091】

また、利用者端末2 0 0がF A Qサーバ1 0 0（F A QチャットボットC B）へ質問文Qの全体又は一部を送信する。F A Qサーバ1 0 0は、通信部1 4 0がこの質問文Qの全体又は一部を受信した場合、制御部1 1 0が、この質問文Qの全体又は一部に類似し、F A Qナレッジベース1 5 0に保持された質問文Qを含むQ Aデータを特定し、通信部1 4 0が、このQ Aデータを参照可能なリンク先の情報（例えばU R L）を利用者端末2 0 0へ送信する。利用者端末2 0 0は、リンク先の情報を受信して表示部に表示する。これにより、利用者は、リンク先にアクセスすることで、入力した質問文Qの全体又は一部に類似する質問文Qを容易に確認でき、質問に要する工数を削減できる。

【0092】

このように、本実施形態のF A Qシステム5は、管理者端末3 0 0に登録推奨情報を表示するので、管理者がQ AデータのF A Qナレッジベース1 5 0への登録の可否を容易に判断できる。また、F A Qシステム5は、質問文がベクトル化された質問文ベクトルを用

10

20

30

40

50

いて特徴量を抽出してクラスタリングするので、パターン認識モデルへの入力データとなる特徴量が不要であり、クラスタリングに係る管理者の処理負荷を低減できる。また、人が設計したような特徴量を質問文Qや回答文Aの登録候補として抽出されることが抑制可能であり、FAQナレッジベース150への登録精度が向上する、また、FAQを作成する元となるメールやチャットログに含まれる不要な文章データの一部が登録対象の質問文Qや回答文Aに含まれることを抑制できる。また、複数のQAデータから代表QAデータを抽出することに人手が不要である。したがって、FAQシステム5の利便性を向上できる。

【0093】

(第2の実施形態を得るに至った経緯)

FAQシステムは、任意の単語又は文章等を入力することで、検索結果を返す。しかし、引用文献2の技術をFAQシステムに適用した場合、このFAQシステムは、検索結果の候補が大量に存在する場合、ユーザ所望の検索結果を短時間で得ることは困難である。また、従来の全文検索システムであるFESSは、事前に候補を絞り込むための単語(ラベル)を付与することで検索結果の絞り込みが可能であるが、単語を手動で設定する必要がある。よって、このFESSをFAQシステムに適用して質問を絞り込もうとすると、人によって単語を手動で設定するため、同様な単語を含む候補が絞り込みの結果として選ばれる可能性が高く、質問を好適に絞り込むことが困難である。第2の実施形態では、上記の点を改善する。

【0094】

(第2の実施形態)

第2の実施形態では、FAQシステム5は、FAQチャットボットCBを用いて質問に対して自動回答する場合に、質問の絞り込み処理を行う。

【0095】

図11Aは、本開示の第2の実施形態に係るFAQサーバの構成例を示すブロック図である。図11Aでは、図2に示したFAQサーバ100と同様の構成については、同一の符号を付し、その説明を省略又は簡略化する。

【0096】

本実施形態のFAQサーバ100Aは、制御部110A、記憶部130A、及び通信部140を備える。記憶部130Aは、非チャットログデータベース(DB)131と、チャットログデータベース(DB)132と、QAデータベース(DB)133と、検索結果データベース(DB)134と、を含む。制御部110Aは、文章抽出部111、単語分割部112、ベクトル化部113、ベクトル類似度計算部115、クラスタリング部116、及び特徴語抽出部118を備える。

【0097】

検索結果データベース134は、利用者からの質問文に基づく所定の検索処理により得られた検索結果を保持する。検索結果データベース134は、過去の検索結果が履歴として蓄積されてもよいし、検索毎に書き換えられてもよい。ここでの検索結果は、質問文を用いて該当するFAQを検索した結果得られる、追加の質問候補文となる質問文である。検索結果として、質問候補文は複数導出され得る。

【0098】

文章抽出部111は、QAデータベース133又は検索結果データベース134から、文章データ(テキストデータ)を抽出する。この場合、文章抽出部111は、QAデータベース133に保持されたQAデータの質問文を質問候補文として抽出してよい。また、文章抽出部111は、検索結果データベース134に保持された質問候補文を抽出してよい。

【0099】

ベクトル化部113は、自然言語処理モデル(例えばWord2Vec又はTF-IDF(Term Frequency - Inverse Document Frequency)又はその他のクラスタリング手法)を用いて、質問候補文をベクトル化し、質問候補文の特徴ベクトル(質問補文

10

20

30

40

50

ベクトル)を得る。

【0100】

ベクトル類似度計算部115は、自然言語処理モデルを用いて、ベクトル化部113により生成された各特徴ベクトル(各質問補文ベクトル)の類似度(相関度)を計算する。

【0101】

クラスタリング部116は、所定のクラスタリング手法(例えばk-means法又はその他のクラスタリング手法)を用いて、導出された各質問候補文ベクトルの類似度に基づいて、質問候補文が複数のクラスタ(質問候補クラスタ)のうちのいずれかに属するように、分類する。

【0102】

単語分割部112は、クラスタリングにより得られた各質問候補クラスタに属する各質問候補文を単語単位に分割する。

【0103】

特徴語抽出部118は、自然言語処理モデル(例えばTF-IDF又はその他の特徴語抽出手法)を用いて、各質問候補クラスタに属する各質問候補文に含まれる各単語のうち、例えば出現頻度の高い単語を特徴語KWとして抽出する。例えば、特徴語抽出部118は、各質問候補クラスタの質問文から1つ以上の特徴語を抽出する。また、特徴語抽出部118は、質問候補クラスタが複数存在する場合、質問候補クラスタの特徴を示す特徴語KWが質問候補クラスタ毎に異なるように抽出してよい。

【0104】

通信部140は、抽出された各クラスタの各特徴語KWを利用者端末200へ送信する。利用者端末200は、FAQサーバ100Aから各クラスタの各特徴語KWを受信し、各特徴語KWを質問の聞き返しの候補として表示部に表示させる。これにより、利用者は、質問の聞き返し候補を確認可能であり、検索結果として所望の結果(質問文)を得ることができるように検索結果を絞ることができる。

【0105】

なお、第2の実施形態では、例えば第1の実施形態で説明したFAQナレッジベース150への登録支援処理によりFAQナレッジベース150にQAデータが登録されているが、これに限られない。例えば、第1の実施形態以外の方法で、QAデータがFAQナレッジベース150に登録されていてもよい。

【0106】

図11Bは、本実施形態の質問の絞り込みが行われるFAQチャット画面G2のイメージを示す図である。図11Bに示すように、利用者が利用者端末200を用いて任意の質問文Q0を送信すると、FAQサーバ100Aは、質問文Q0を用いて検索処理を行う。検索処理の結果、図11Bでは100件程の登録済みのFAQの質問候補が得られている。ここで、FAQサーバ100Aは、質問を絞り込むための特徴語KWを利用者端末200に送信して表示させ、利用者により特徴語KWを指定させるように促している。

【0107】

本実施形態では、FAQサーバ100Aは、制御部110Aが、クラスタリングを用い、候補である文書をクラスタリングして複数のクラスタ(グループ)に分割する。そして、制御部110Aが、各クラスタに含まれる文書から特徴語KWを抽出し、その特徴語KWを、聞き返しで用いる候補とする。そして、利用者端末200は、操作部を介して聞き返し候補としての特徴語KWを指定すると、この指定操作情報をFAQサーバ100Aに送信する。FAQサーバ100Aは、通信部140が指定操作情報を受信し、制御部110Aが、指定された特徴語KWに関連する質問候補クラスタに含まれる質問候補文QCを特定し、通信部140が特定された質問候補文QCを利用者端末200に送信する。利用者端末200は、FAQサーバ100Aから質問候補文QCを受信して表示部に表示する。このようにして、利用者は、任意の質問文Q0から関連する特徴語KWを得て、特徴語KWを指定することで、特徴語KWに関連する質問候補文QCを得ることができる。

【0108】

10

20

30

40

50

図12は、クラスタリングのイメージを示す図である。図12に示すように、FAQサーバ100Aは、制御部110Aが、例えば大量の検索結果から少数の質問候補クラスタCL1～CL3に分割する。ここで、通常のクラスタリングでは、利用者は、各質問候補クラスタの特徴を容易に把握することができない。一方、図12に示すように、FAQサーバ100Aは、各質問候補クラスタの特徴語KWを検索結果の候補とすることで、人間（利用者）から見て理解できる。FAQサーバ100Aは、動的な単語聞き返しにより、各単語の特徴差異が明確となり、高精度に質問を絞り込みできる。例えば、FAQサーバ100Aは、質問候補クラスタCL1に対応する特徴語KWが選択されると、特徴語KWにより特徴付けられた質問候補クラスタCL1に属する各質問候補文QCに絞り込むことができ、例えば5択にすることができる。

10

【0109】

図13は、Q番号及び質問候補文QCを含む検索結果（候補）の一例を示す図である。図13では、質問候補文QCを識別するQ番号と、質問候補文QCとが、関連付けて示されている。図13では、検索結果（候補）として質問候補文QCが6つ得られている。

【0110】

図14は、質問候補文QCをベクトル化した結果の一例を示す図である。図14では、Q番号と、質問候補文QCがベクトル化された質問候補文ベクトルとが、関連付けて示されている。

【0111】

図15は、質問候補文ベクトルをクラスタリングし、各クラスタと質問候補文QCとの一例を示す図である。図15では、クラスタを識別するクラスタ番号と、各クラスタに属する1つ以上の質問候補文QCと、が関連付けて示されている。図15では、4つのクラスタが生成されており、例えばクラスタ0には2つの質問候補文QCが属している。

20

【0112】

図16は、各クラスタの質問候補文QCをベクトル化した結果の一例を示す図である。図16では、クラスタ番号と、各クラスタに属する1つ以上の質問候補文QCの質問候補文ベクトルと、が関連付けて示されている。

【0113】

図17は、クラスタ毎に抽出された特徴語KWの一例を示す図である。図17では、クラスタ番号と、各クラスタの特徴語KWと、が関連付けて示されている。クラスタ毎に、複数の特徴語KWが導出されていてもよい。

30

【0114】

このようにして、制御部110Aは、任意の質問文Q0に対して得られた検索結果から、検索結果を絞り込むための聞き返し候補を導出できる。また、聞き返し候補として得られた各特徴語KWは、利用者端末200に送信されて表示されることが可能である。

【0115】

図18は、質問の絞り込みを考慮したFAQチャット画面G3の一例を示す図である。図18では、任意の質問の後に、検索結果として20件が得られたことを表示しており、さらに、特徴語KWでの絞り込みを促している。ここでの特徴語KWは、図17に示したクラスタ毎の特徴語KWである。図18では、クラスタ毎に分けて特徴語KWが表示されている。さらに、利用者は、利用者端末200を用いて、クラスタ毎の特徴語KWのうち、クラスタ番号1の特徴語KWである契約期間、有効期限、自動更新を絞り込みの特徴語KWとして指定している。この結果、指定された特徴語KWに対応するクラスタ番号0のクラスタに属する質問候補文QCが表示されている。

40

【0116】

次に、本実施形態のバリエーションについて説明する。

【0117】

管理者端末300は、操作部を介して、FAQサーバ100Aの特徴語抽出部118により特徴語KWとして抽出される対象としないストップワードを入力し、ストップワードをFAQサーバ100Aに送信してよい。FAQサーバ100Aは、通信部140がスト

50

ストップワードの情報を受信し、制御部 110A がストップワードを設定してよい。設定されたストップワードは、記憶部 130 に保持されてよい。

【0118】

管理者端末 300 は、操作部を介して、類似する複数の単語が登録される類似語テーブルを生成する操作を受け付け、この操作情報を F A Q サーバ 100A に送信してよい。F A Q サーバ 100A は、通信部 140 がこの操作情報を受信し、制御部 110A がこの操作情報に基づいて類似語テーブルを生成してよい。類似語テーブルは、例えば、類似する複数の単語と、類似する複数の単語を代表する代表単語と、を関連付けて保存する。類似する単語は、例えば、表記ゆれの単語である。設定された類似語テーブルは、記憶部 130 に保持されてよい。特徴語抽出部 118 は、特徴語 K W が得られると、類似語テーブルにこの特徴語 K W に類似する単語が含まれる場合、この特徴語 K W に対応する代表特徴語を抽出する。

10

【0119】

これにより、F A Q サーバ 100A は、抽出される特徴語 K W の質を向上でき、つまり聞き返し候補の質を向上できる。よって、F A Q サーバ 100A は、検索結果の候補の絞り込みの速度又は絞り込みの精度等を向上できる。

【0120】

また、F A Q サーバ 100A は、制御部 110A が、クラスタリングにより得られるクラスタ数が自動又は手動で指定されてもよい。例えば、制御部 110A が、検索結果（候補）としての質問候補文 Q C の数に応じて、クラスタ数を決定してよい。また、管理者端末 300 は、操作部を介して、クラスタリングにより得られるクラスタ数を指定する操作を受け付け、この操作情報を F A Q サーバ 100A に送信してよい。F A Q サーバ 100A は、制御部 110A が、通信部 140 を介して管理者端末 300 から上記の操作情報を受信し、操作情報に基づいてクラスタ数を決定してよい。例えば、クラスタ数が大きな数に設定されると、質問候補文 Q C の分類が細くなる。

20

【0121】

F A Q サーバ 100A は、クラスタ数を適切な数に設定することで、効果の高い検索結果の候補の絞り込みが可能である。例えば、クラスタ数を 4 つ以下のいずれかの数に決定してよい。この場合、利用者は短時間で絞り込みの傾向を把握可能である。

【0122】

また、F A Q サーバ 100A は、制御部 110A が、共起ネットワークを用いて、特徴語 K W をノード N D として、各特徴語 K W をエッジ E D で結んで、ネットワーク図 N E を生成してよい。したがって、このネットワーク図 N E は、各特徴語 K W の関連性を示すものとなる。図 19 は、各特徴語 K W の関係性の一例を示すネットワーク図 N E である。

30

【0123】

このネットワーク図 N E では、特徴語 K W 同士の関係性が近いつまり類似度が高い程、この特徴語 K W を示すノード N D 同士の距離が短くなるようにノード N D が配置される。各特徴語 K W の類似度は、各特徴語 K W で特徴付けられる各クラスタの類似度に相当する。各クラスタの類似度は、ベクトル類似度計算部 115 により計算可能である。通信部 140 は、生成されたネットワーク図 N E を利用者端末 200 に送信してよい。利用者端末 200 は、F A Q サーバ 100A からネットワーク図 N E を受信して表示してよい。これにより、利用者は、ネットワーク図 N E の表示を確認することで、各特徴語 K W の関係性を視覚的に把握できる。

40

【0124】

また、ネットワーク図 N E に含まれる各ノード N D は、利用者端末 200 の操作部を介して、利用者により選択可能であってよい。例えば、利用者端末 200 は、操作部を介して、任意のノード N D を選択してよい。利用者端末 200 は、選択されたノード N D の情報を F A Q サーバ 100A へ送信する。F A Q サーバ 100A は、通信部 140 が、利用者端末 200 から選択されたノード N D の情報を受信し、制御部 110A が、選択されたノード N D に対応する特徴語 K W が指定されたことを認識する。つまり、F A Q サーバ 1

50

00Aは、FAQチャット画面上で絞り込み用の特徴語KWを指定する代わりに、ネットワーク図NEを用いてボタンを選択するようにノード選択することで、1つの特徴語KWを指定できる。よって、FAQサーバ100Aは、一層直感的に特徴語KWを用いた検索結果の候補の絞り込みを実現できる。

【0125】

また、ネットワーク図NEに含まれる各エッジEDは、利用者端末200の操作部を介して、利用者により選択可能であってよい。例えば、利用者端末200は、操作部を介して、任意のエッジEDを選択してよい。利用者端末200は、選択されたエッジEDの情報をFAQサーバ100Aへ送信してよい。FAQサーバ100Aは、通信部140が、利用者端末200から選択されたエッジEDの情報を受信し、選択されたエッジEDの両端に接続されたノードNDに対応する2つの特徴語KWが指定されたことを認識する。つまり、FAQサーバ100Aは、FAQチャット画面上で絞り込み用の特徴語KWを指定する代わりに、ネットワーク図NEを用いてボタンを選択するようにエッジ選択することで、2つの特徴語KWを同時に指定できる。よって、FAQサーバ100Aは、一層直感的に特徴語KWを用いた検索結果の候補の絞り込みを実現できる。

10

【0126】

このように、FAQサーバ100Aは、ネットワーク図NEを用いることで、各特徴語KWの関係性を容易に把握でき、視覚的な検索結果の候補の絞り込みが可能である。エッジEDを用いた絞り込みでは、検索結果の候補の新たな見せ方が可能である。例えば、「英語」及び「ひな形」の2つのノードを接続するエッジEDが選択されると、FAQサーバ100Aは、「英語」及び「ひな形」を絞り込みのための特徴語KWとすることができ、「英語のひな形」で検索結果（質問候補文QC）を絞り込みできる。

20

【0127】

また、制御部110Aは、各特徴語KWの類似度を基に、各特徴語KWの表示の優先順位を決定してもよい。例えば、利用者端末200の操作部を介して、利用者が1つの特徴語KWを指定した場合、指定された特徴語KWとの関係性が深い（類似度が所定閾値以上である）他の特徴語KWが、優先的に利用者端末200の表示部に表示され、選択可能とされてよい。この場合、共起ネットワークを用いたネットワーク図NEにおいて、ノードNDとして表示される特徴語KWが優先順位の高い特徴語KWに限定されることで、利用者は、指定済みの特徴語KWに似た特徴語KWを指定し易くなる。

30

【0128】

このように、共起ネットワークを用いたネットワーク図NEを用いることで、特徴語KW同士の関連性を視覚的に理解し易くなる。また、視覚的に検索結果の候補の絞り込みが可能である。

【0129】

また、検索結果の候補の出力は、文字情報だけでなく、画像情報が用いられてもよい。例えば、制御部110Aは、質問の絞り込み処理の前に、画像とその画像を説明する文章データとを学習させ、学習済みモデルを生成しておく。制御部110Aは、この学習済みモデルを用いて、検索結果の候補として文字情報（質問候補文QC）の代わりに、又は文字情報とともに、画像情報を出力してよい。また、制御部110Aは、1つ以上の特徴語KWに対応する画像情報を検索し、通信部140を介してこの画像情報を送信してよい。例えば、2つの特徴語KWとして「犬」と「白」とが抽出された場合、白い犬の画像が画像情報として決定され、利用者端末200に送信されてよい。利用者端末200は、FAQサーバ100Aから画像情報を受信して白い犬の画像を表示部に表示してよい。

40

【0130】

このように、本実施形態のFAQシステム5によれば、FAQサーバ100Aは、質問文Q0に基づく検索結果の候補としての質問候補文QCの類似性により、各質問候補文QCをクラスタリングして質問候補クラスタを生成するだけでなく、質問候補クラスタの特徴を示す単語（特徴語）を表示させることができる。これにより、FAQサーバ100Aは、利用者による質問の絞り込みを的確に支援できる。

50

【0131】

(第3の実施形態)

第3の実施形態は、第1の実施形態の変形例に相当する。第3の実施形態では、問合せメールや回答メールなどのメールに関する処理を行う。

【0132】

図20は、第3の実施形態に係るFAQサーバの構成例を示すブロック図である。図20では、図2に示したFAQサーバ100と同様の構成については、同一の符号を付し、その説明を省略又は簡略化する。

【0133】

本実施形態のFAQサーバ100Bは、制御部110B、記憶部130、及び通信部140を備える。制御部110Bは、第1の実施形態の制御部110と比較すると、逸脱処理部122と強調処理部123を含む構成である。

【0134】

逸脱処理部122は、メール（送受信メール）の宛先とメール本文の宛先との対応関係を判定し、両宛先が一致しているか不一致であるか（逸脱しているか）を判定する。両宛先が不一致である場合には、逸脱処理部122は、例えば、両宛先が不一致であることを示す警告情報の出力、メールの送受信を遅延させる送受信遅延、又は両宛先が不一致であるメールの強制分離、等の処理を行う。上記の警告情報の出力は、例えば管理者端末300への警告情報の送信でよい。この場合、管理者端末300は、警告情報を受信して提示（例えば表示部による表示）してよい。よって、FAQサーバ100Bは、上記の宛先の不一致がある場合には、メールの通常の送信が抑制されるので、メールの誤送信を抑制できる。

【0135】

強調処理部123は、メール本文において特に確認すべき部分（確認必要部分ともいう）を判別し、確認必要部分を強調表示するよう、利用者端末200又は管理者端末300に指示する。この確認必要部分は、例えば、当該メールにおける代表文章、又はセキュリティリスクの高い文章である。利用者端末200又は管理者端末300は、FAQサーバ100Bからの指示を受信すると、判別されたメール本文中の確認必要部分を強調表示する。これにより、利用者又は回答者は、メール中の重要な部分を容易に把握できる。

【0136】

なお、メール本文における確認必要部分の判別は、前述と同様に、自然言語処理モデルを用いて実施されてよい。例えば、制御部110Bは、メール本文の文章データをベクトル化し、文章ベクトルを用いて文章データを分類してよい。制御部110Bは、文章データを、代表文章又はセキュリティの高い文章等の確認必要部分を含む複数の種類（クラス）に分類可能であり、確認必要部分を確認可能となる。

【0137】

図21は、メールの宛先の対応関係と、メール本文での確認必要部分BS1、BS2と、を強調表示した情報の一例を示す図である。確認必要部分BS1は、例えば代表文章であり、確認必要部分BS2は、例えばセキュリティリスクの高い文章である。

【0138】

図21では、宛先（Cc:）に対応する送信先は、メールの宛先を示すメール送信先情報と、メール本文中と、のいずれにおいても「伊藤」である。一方、図21では、宛先（To:）に対応する送信先は、メールの宛先を示すメール送信先情報では「田中」であり、メール本文中では「鈴木」である。したがって、宛先（To:）は両方で不一致である。強調処理部123は、不一致である双方の宛先（To:）の表示形態を変更し、双方の宛先（To:）の視認性を向上させてよい。これにより、利用者又は回答者は、宛先（To:）が異なっていることを容易に確認でき、正しい宛先に容易に修正できる。また、逸脱処理部122は、宛先（To:）は両方で不一致であるので、両宛先が不一致であることを示す警告情報の出力、メールの送受信を遅延させる送受信遅延、又は両宛先が不一致であるメールの強制分離、等の処理を行ってよい。これにより、FAQサーバ100Bは

10

20

30

40

50

、利用者又は回答者によるメールの送信の意図によらずに、メール送信を抑制できる。

【0139】

また、強調処理部123は、分類により得られた確認必要部分BS1、BS2の表示態様を変更して、例えば太字にしたり目立つ文字色にしたりすることで、確認必要部分BS1、BS2の視認性を向上させてよい。

【0140】

このように、本実施形態のFAQシステム5によれば、例えば、回答者が利用者からQ&Aの問合せを受けた場合に、メールにおいて注意すべき点を強調表示できる。そのため、回答者は、質問文Qの内容確認を容易に行うことができ、回答文Aにおけるリスクを検出できる。

10

【0141】

次に、本実施形態のバリエーションについて説明する。

【0142】

本実施形態の機能は、メールの送受信ソフト又はメールの送受信ソフトのアドインソフトによって実現されてもよい。

【0143】

FAQサーバ100Bの制御部110Bは、送信メールのメール宛先（例えばTo:）と、メール本文中から抽出した宛先と、の対応関係を導出してよい。制御部110Bは、受信メールのメール宛先（例えばFrom:）と、本文中から抽出された送信者と、の対応関係を導出してよい。制御部110Bは、メール送信時において、対応関係から逸脱（例えば宛先の不一致）を判定し、逸脱している場合には、メール誤送信の可能性ありと判定し、警告、送信遅延、又は強制分離等の誤送信緩和措置を行ってよい。また、制御部110Bは、メール受信時において、対応関係（例えば宛先の一致）を学習させた所定の学習済みモデルの対応関係から逸脱している場合、成りすましの可能性ありと判定し、警告、受信遅延、又は強制分離等の誤受信緩和措置を行ってよい。

20

【0144】

また、制御部110Bは、送信メールにおけるメール宛先（例えばTo:）とメール本文中から抽出した宛先との対応関係と、受信メールにおけるメール宛先（例えばFrom:）とメール本文中から抽出した送信者との対応関係と、を学習させたAIモデル（学習済みモデル）によって、対応関係の逸脱を判定する機能を有してよい。

30

【0145】

また、制御部110Bは、上記のAIモデルの学習を、メールアドイン経由で都度行ってもよいし、非チャットログデータベース131に蓄積されたメールデータを基に行ってもよい。また、制御部110Bは、メール誤送信と成りすましとの判定のための推論は、メールアドイン経由で都度行ってもよい。

【0146】

また、利用者及び回答者は、メール作業中には利用者端末200及び管理者端末300による表示を確認しているので、AIモデルを用いたメール誤送信及び成りすましの判定の判定結果が正当であるか否かを判別可能である。AIモデルを用いたメール誤送信及び成りすましの判定の判定結果が正当でない（つまり誤判定である）場合、利用者端末200又は管理者端末300は、その端末の操作部を介して正当な判定の情報を入力し、フィードバック情報としてFAQサーバ100Bに送信する。制御部110Bは、通信部140を介してフィードバック情報を入力し、AIモデルの学習データとして入力してAIモデルを学習してよい。これにより、FAQシステム5は、AIモデルをより正しく育成可能であり、AIモデルを用いたメール誤送信及び成りすましの判定の判定精度を向上できる。

40

【0147】

以上のように、上記実施形態のFAQサーバ100（データ管理装置の一例）は、質問文Q及び質問文Qに対する回答文Aを管理する。FAQサーバ100は、制御部110と通信部140とを備える。制御部110は、複数の文章データを取得し、複数の文章デー

50

タをベクトル化して複数の文章ベクトルを作成し、複数の文章ベクトルに基づいて、複数の文章データから複数の質問文Qを抽出する。制御部110は、複数の質問文Qに対応する複数の文章ベクトルである複数の質問文ベクトルに基づいて、抽出された複数の質問文Qを、相互に類似する質問文Qが含まれる質問クラスタにクラスタリングする。制御部110は、質問クラスタ毎に、質問クラスタの中心に対応する質問クラスタに属する質問文Qを代表する代表質問文QRを決定する。通信部140は、決定された代表質問文QRをFAQナレッジベース150（記憶装置の一例）に登録することを推奨する登録推奨情報I1を、FAQサーバ100を管理する管理者が使用する管理者端末300へ送信する。

【0148】

これにより、FAQサーバ100は、文章データから質問文を抽出でき、QAデータの生成を省力化できる。また、パターン認識モデルへの入力データとなる特徴量が不要であり、質問文Qがベクトル化された質問文ベクトルを用いて特徴量を抽出してクラスタリングするので、クラスタリングに係る処理負荷を低減できる。また、FAQサーバ100は、クラスタ内の代表的な質問文Qを登録に推奨することで、例えば、管理者が管理者端末300を用いて質問文Q毎に微調整してFAQナレッジベース150に登録する場合でも、クラスタ内の質問文Qの管理者による修正量を低減できる。このように、FAQシステム5の利便性を向上できる。

【0149】

また、登録推奨情報I1は、代表質問文QRと代表質問文QRに対応する回答文Aである代表回答文ARとを関連付けてFAQナレッジベース150に登録することを推奨する情報を含んでよい。

【0150】

これにより、FAQサーバ100は、類似する複数の質問文Qの中の代表的な質問と回答を関連付けて管理できる。よって、利用者は、利用者端末200を用いてFAQチャットボットCBに質問することで、この質問文Qに関する代表的な回答文Aを得ることができる。

【0151】

また、制御部110は、複数の文章ベクトルに基づいて、複数の文章データから複数の回答文Aを抽出し、複数の回答文Aに対応する複数の文章ベクトルである複数の回答文ベクトルに基づいて、抽出された複数の回答文Aを、相互に類似する回答文Aが含まれる回答クラスタにクラスタリングしてよい。制御部110は、回答クラスタ毎に、回答クラスタの中心に対応する回答クラスタに属する回答文Aを代表する代表回答文ARを決定してよい。

【0152】

これにより、FAQサーバ100は、FAQサーバ100は、文章データから回答文Aを抽出でき、QAデータの生成を省力化できる。また、パターン認識モデルへの入力データとなる特徴量が不要であり、回答文Aがベクトル化された回答文ベクトルを用いて特徴量を抽出してクラスタリングするので、クラスタリングに係る処理負荷を低減できる。また、FAQサーバ100は、回答クラスタ内の代表的な回答文Aを登録に推奨することで、例えば、管理者が管理者端末300を用いて回答文毎に微調整してFAQナレッジベース150に登録する場合でも、回答クラスタ内の回答文Aの管理者による修正量を低減できる。このように、FAQシステム5の利便性を向上できる。

【0153】

また、制御部110は、質問文又は回答文に所定の住所又は電話番号が含まれていることを検出した場合、所定の住所又は電話番号を削除してよい。通信部140は、所定の住所又は電話番号の削除後に、登録推奨情報I1を送信してよい。所定の住所又は電話番号は、例えば、事前に記憶部130に登録されていない住所又は電話番号であってよい。

【0154】

これにより、管理者端末300が登録推奨情報を受信して表示することで、管理者は、登録推奨情報を確認して、所定の住所又は電話番号が削除されたQAデータをFAQナレ

10

20

30

40

50

ッジベース 150 に登録できる。

【0155】

また、通信部 140 は、利用者が使用する利用者端末 200 から任意の質問文 Q0 を取得してよい。制御部 110A は、任意の質問文 Q0 に基づいて所定の検索処理を行い、検索結果として利用者が追加で質問する候補である複数の質問候補文 QC を抽出し、質問候補文 QC をベクトル化して複数の質問候補文 QC の文章ベクトルである複数の質問候補文ベクトルを作成してよい。制御部 110A は、複数の質問候補文ベクトルに基づいて、抽出された複数の質問候補文 QC を、相互に類似する質問候補文 QC が含まれる質問候補クラスタにクラスタリングし、質問候補クラスタ毎に、質問候補クラスタの特徴を示す特徴語 KW を抽出してよい。通信部 140 は、各質問候補クラスタの特徴語 KW を利用者端末 200 へ送信してよい。

10

【0156】

これにより、FAQサーバ 100A は、質問候補文 QC を質問候補クラスタに単に分類するだけでなく、質問候補クラスタの特徴を示す特徴語 KW を導出し、つまり同一クラスタ内の質問候補文 QC で頻出する単語を特徴語 KW として導出し、この特徴語 KW を任意の質問文 Q0 に後続する追加質問の特徴語 KW として使用できる。よって、利用者は、おおざっぱな質問をすることで、質問を絞り込むためのヒントを得ることができ、質問効率が向上する。

【0157】

また、制御部 110 は、質問候補クラスタが複数存在する場合、質問候補クラスタの特徴を示す特徴語 KW が質問候補クラスタ毎に異なるように抽出してよい。

20

【0158】

これにより、FAQサーバ 100A は、特徴語 KW から質問候補クラスタが一意に決定可能である。よって、FAQサーバ 100 は、利用者が特徴語 KW を選択することで、利用者が意図する質問候補文 QC に絞り込む精度を向上できる。

【0159】

また、制御部 110 は、類似する単語を規定する類似語テーブルに基づいて、各質問候補クラスタの特徴語 KW が異なる単語であるか否かを判定してよい。

【0160】

これにより、FAQサーバ 100 は、表記ゆれ等により表記が異なり類似する単語を特定できる。

30

【0161】

また、制御部 110 は、抽出された質問候補文 QC の数に応じて、クラスタリングにより得られる質問候補クラスタの数を決定してよい。

【0162】

これにより、FAQサーバ 100 は、質問候補クラスタ内の質問候補文 QC の数を適切に維持可能である。例えば、FAQサーバ 100 は、質問候補クラスタの数に相当する質問の種別の数が過多となり、利用者が質問の種別を選択し難くなることを抑制できる。また、質問の種別に属する質問候補文 QC の数が過多となり、細かな質問が利用者に多数提示されることを抑制できる。

40

【0163】

また、制御部 110 は、抽出された複数の特徴語 KW の共起関係を示す共起関係情報（例えばネットワーク図 NE）を導出してよい。通信部 140 は、共起関係情報を利用者端末 200 に送信してよい。

【0164】

これにより、利用者端末 200 は、共起関係情報を取得して共起関係情報に基づいて共起関係を示すネットワーク図 NE を表示することで、各特徴語 KW の関係性を一見して把握できる。また、例えば、ネットワーク図 NE のノード ND を選択可能としておくことで、利用者端末 200 は、ネットワーク図 NE を用いて特徴語 KE を直感的に選択でき、容易に質問を絞り込むための質問候補を選択できる。

50

【0165】

以上、図面を参照しながら各種の実施形態について説明したが、本開示はかかる例に限定されないことは言うまでもない。当業者であれば、特許請求の範囲に記載された範囲内において、各種の変更例又は修正例に想到し得ることは明らかであり、それらについても当然に本開示の技術的範囲に属するものと了解される。また、開示の趣旨を逸脱しない範囲において、上記実施形態における各構成要素を任意に組み合わせてもよい。

【0166】

また、上記実施形態では、プロセッサは、物理的にどのように構成してもよい。また、プログラム可能なプロセッサを用いれば、プログラムの変更により処理内容を変更できるので、プロセッサの設計の自由度を高めることができる。プロセッサは、1つの半導体チップで構成してもよいし、物理的に複数の半導体チップで構成してもよい。複数の半導体チップで構成する場合、上記実施形態の各制御をそれぞれ別の半導体チップで実現してもよい。この場合、それらの複数の半導体チップで1つのプロセッサを構成すると考えることができる。また、プロセッサは、半導体チップと別の機能を有する部材（コンデンサ等）で構成してもよい。また、プロセッサが有する機能とそれ以外の機能とを実現するように、1つの半導体チップを構成してもよい。また、複数のプロセッサが1つのプロセッサで構成されてもよい。

【産業上の利用可能性】

【0167】

本開示は、FAQシステムの利便性を向上できるデータ管理装置、FAQシステム、データ管理方法、プログラム等に有用である。

【符号の説明】

【0168】

5 FAQシステム
100, 100A, 100B FAQサーバ
110, 110A, 110B 制御部
111 文章抽出部
112 単語分割部
113 ベクトル化部
114 分類部
115 ベクトル類似度計算部
116 クラスタリング部
117 クラスタ中心計算部
118 特徴語抽出部
119 固有表現抽出部
121 匿名化处理部
122 逸脱処理部
123 強調処理部
130, 130A 記憶部
131 非チャットログデータベース
132 チャットログデータベース
133 QAデータベース
134 検索結果データベース
140 通信部
150 FAQナレッジベース
200 利用者端末
300 管理者端末

10

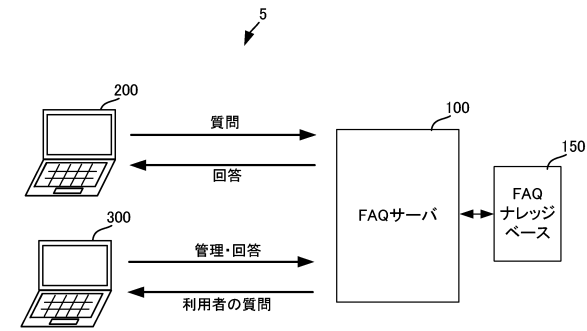
20

30

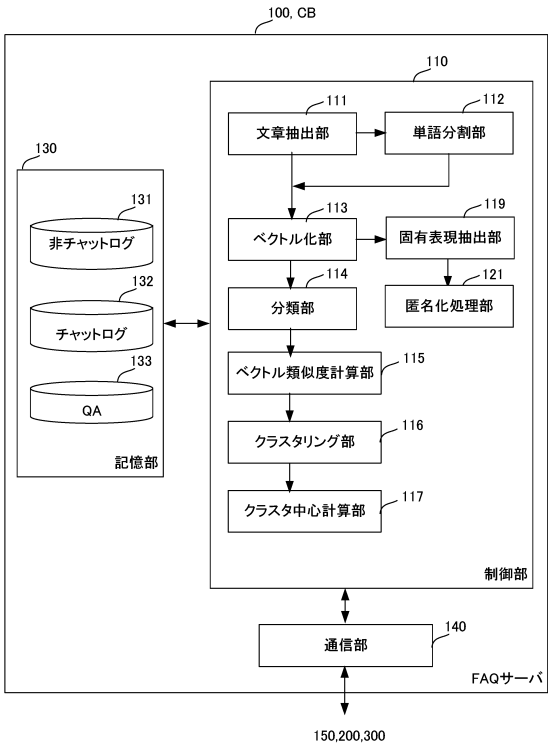
40

50

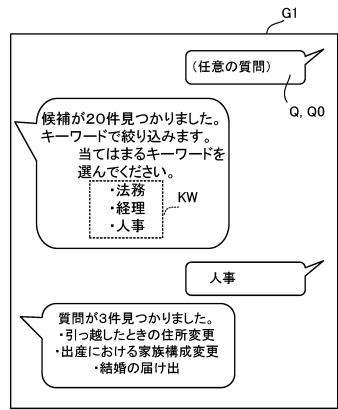
【図面】
【図 1】



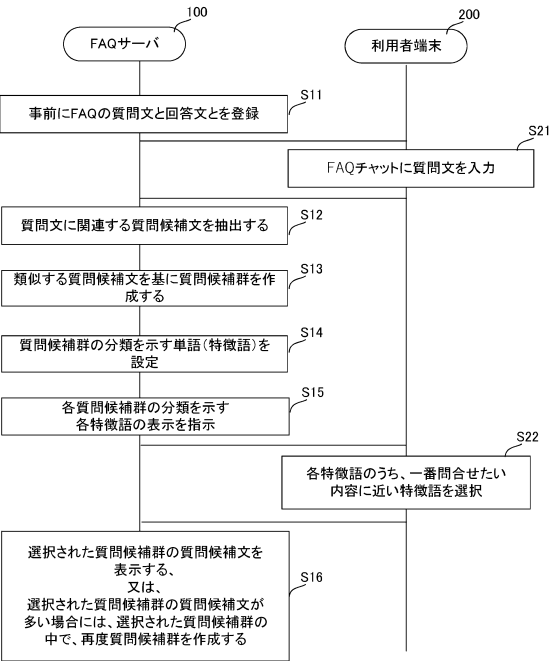
【図 2】



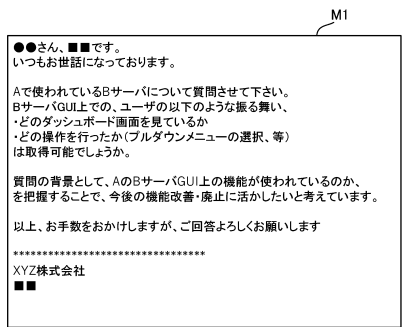
【 図 3 】



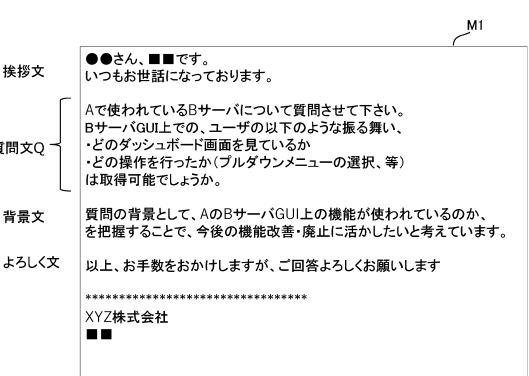
【 図 4 】



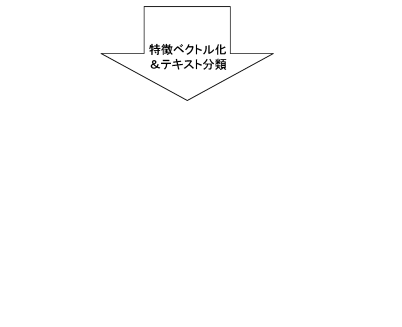
【 図 5 】



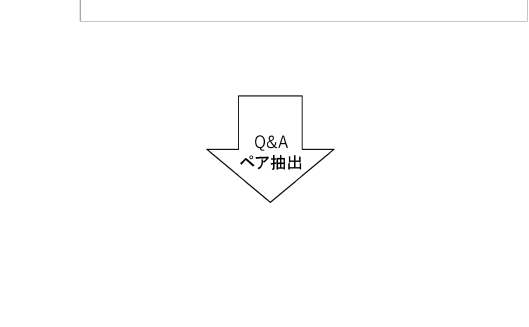
【 図 6 】



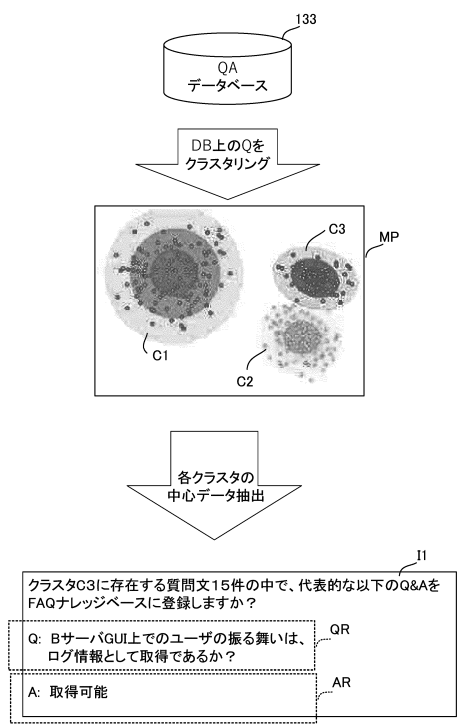
【 図 7 】



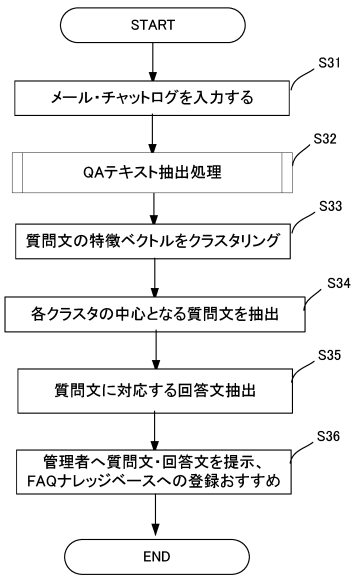
【 図 8 】



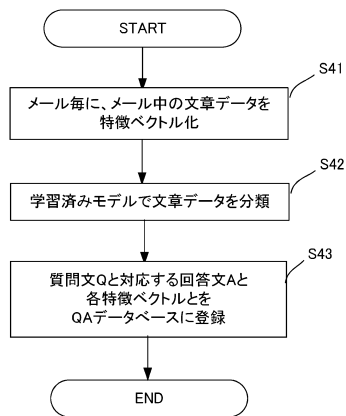
【 図 7 】



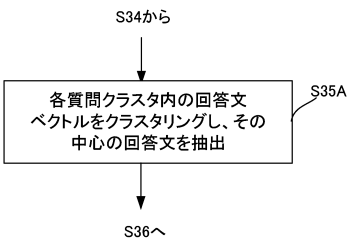
【 図 8 】



【 図 9 】



【 図 1 0 】



10

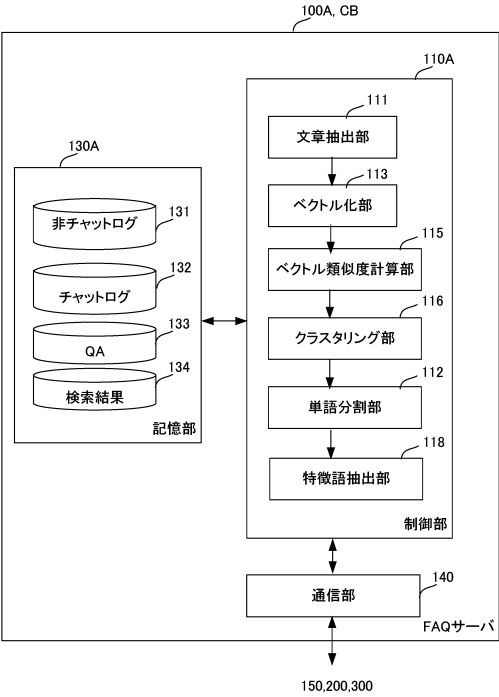
20

30

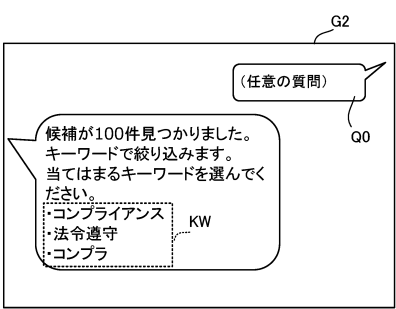
40

50

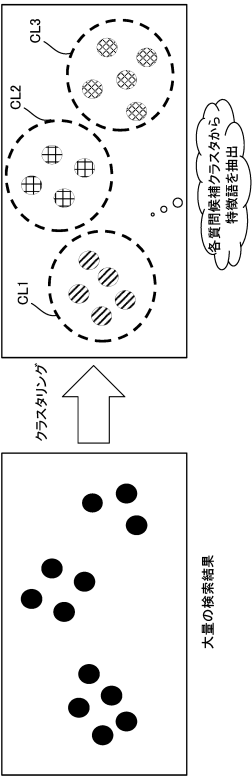
【図 1 1 A】



【図 1 1 B】



【図 1 2】



【図 1 3】

<検索結果(候補)>

Q番号	質問候補文
1	NDAの契約期間の設定について教えてください
2	NDAの署名者はルールがあるのでしょうか？
3	有効期間を永久または自動更新とすることは可能ですか？
4	三社間でNDAを締結することはできますか？
5	外国とのNDA契約の雛形は日本と違いますか
6	英語の雛形が欲しい



10

20

30

40

50

【図 1 4】

Q番号	質問候補文ベクトル
1	0. 0. 0.03688225 0. 0. 0.07117276, 0.02372425 0. ...
2	[0. 0. 0. 0. 0. 0.04304484, 0. 0. 0. 0. 0. 0., 0.05185589 0.06691847
3	[0. 0. 0. 0. 0. 0.08784263, 0. 0.06828106 0. 0. 0. 0.06828106, 0...
4	0. 0. 0.03688225 0. 0. 0.07117276, 0.02372425 0. ...
5	[0. 0. 0. 0. 0. 0.04304484, 0. 0. 0. 0. 0. 0., 0.05185589 0.06691847
6	[0. 0. 0. 0. 0. 0.08784263, 0. 0.06828106 0. 0. 0. 0.06828106, 0...

クラスタリング

➡

【図 1 5】

クラスタ	質問候補文
0	NDAの契約期間の設定について教えてください有効期間を永久または自動更新することは可能ですか？
1	NDAの署名者はルールがあるのでしょうか？
2	三社間でNDAを締結することはできますか？
3	外国とのNDA契約の雛形は日本と違いますか語の雛形が欲しい

ベクトル化

➡

10

【図 1 6】

クラスタ	質問候補文ベクトル
0	0. 0. 0.03688225 0. 0. 0.07117276, 0.02372425 0. ...
1	[0. 0. 0. 0. 0. 0.04304484, 0. 0. 0. 0. 0. 0., 0.05185589 0.06691847
2	[0. 0. 0. 0. 0. 0.08784263, 0. 0.06828106 0. 0. 0. 0.06828106, 0...
3	0. 0. 0.03688225 0. 0. 0.07117276, 0.02372425 0. ...

特徴語抽出

➡

【図 1 7】

クラスタ	特徴語
0	「契約期間」「有効期限」「自動更新」
1	「署名」「ルール」
2	「締結」「終結」
3	「海外」「ひな形」「英語」

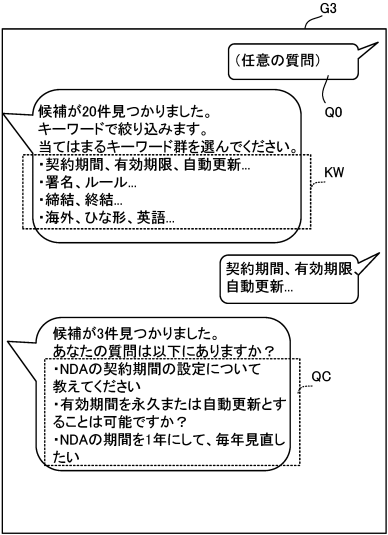
20

30

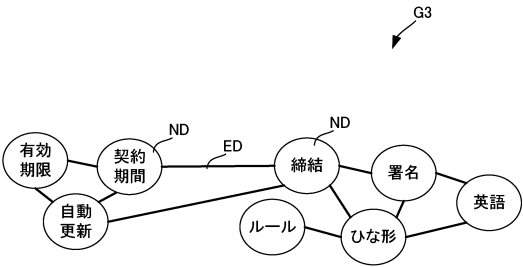
40

50

【図 1 8】



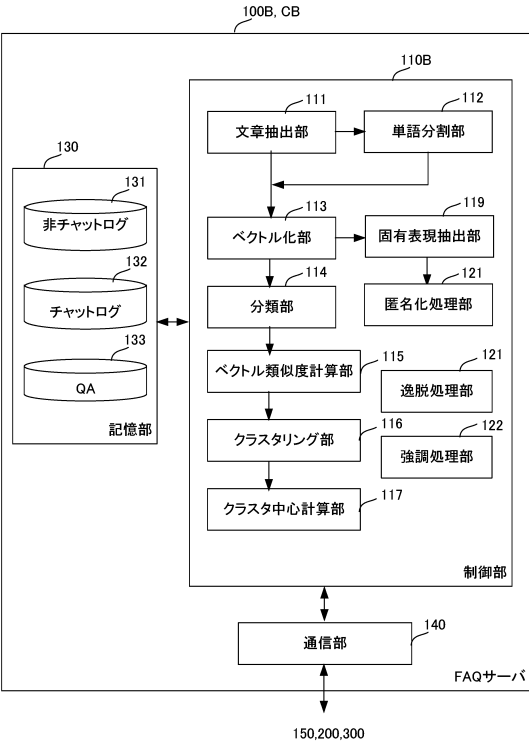
【図 1 9】



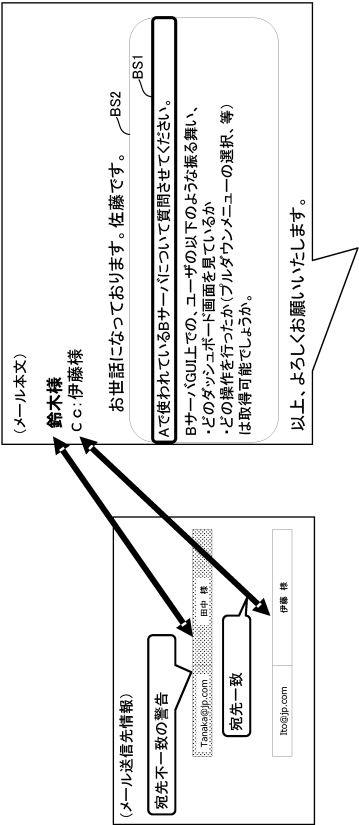
10

20

【図 2 0】



【図 2 1】



30

40

50

フロントページの続き

新橋ビル2号館 パナソニックソリューションテクノロジー株式会社内
Fターム (参考) 5B175 DA01 EA01