

(12) 特許協力条約に基づいて公開された国際出願

(19) 世界知的所有権機関
国際事務局

(43) 国際公開日
2021年11月25日(25.11.2021)

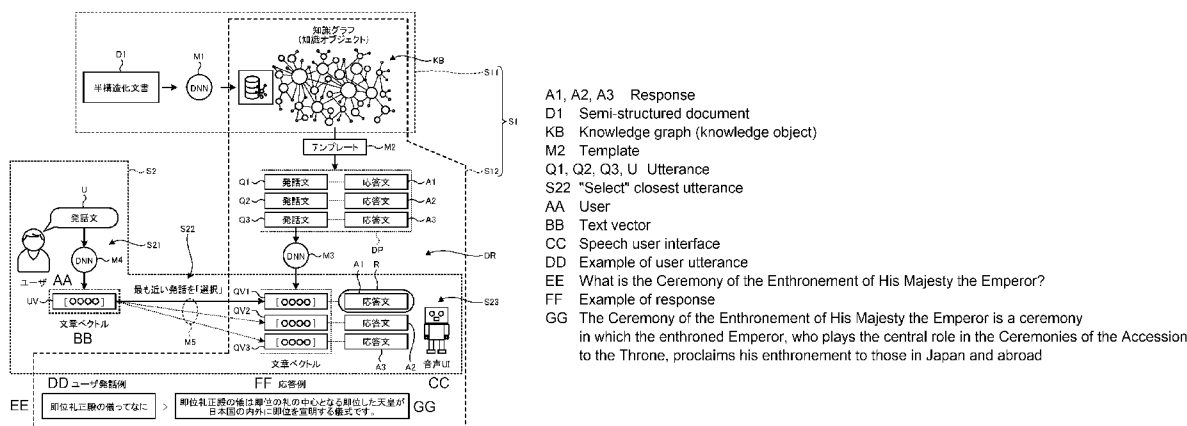


(10) 国際公開番号
WO 2021/235225 A1

- (51) 国際特許分類:
G06F 16/83 (2019.01) G06F 40/30 (2020.01)
G06F 16/90 (2019.01) G06F 40/56 (2020.01)
G06F 40/205 (2020.01)
- (71) 出願人: ソニーグループ株式会社(SONY GROUP CORPORATION) [JP/JP]; 〒1080075 東京都港区港南1丁目7番1号 Tokyo (JP).
- (72) 発明者: 本間 文規 (HOMMA, Fuminori); 〒1080075 東京都港区港南1丁目7番1号 ソニー株式会社内 Tokyo (JP).
- (74) 代理人: 特許業務法人酒井国際特許事務所 (SAKAI INTERNATIONAL PATENT OFFICE); 〒1000013 東京都千代田区霞が関3丁目8番1号 虎の門三井ビルディング Tokyo (JP).
- (81) 指定国(表示のない限り、全ての種類の国内保護が可能): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH,
- (21) 国際出願番号: PCT/JP2021/017336
- (22) 国際出願日: 2021年5月6日(06.05.2021)
- (25) 国際出願の言語: 日本語
- (26) 国際公開の言語: 日本語
- (30) 優先権データ:
特願 2020-089205 2020年5月21日(21.05.2020) JP

(54) Title: INFORMATION PROCESSING DEVICE AND INFORMATION PROCESSING METHOD

(54) 発明の名称: 情報処理装置及び情報処理方法



(57) Abstract: An information processing device according to the present invention is provided with a knowledge-object generation unit and a dialog-pair generation unit. The knowledge-object generation unit extracts a plurality of terms from a semi-structured document and generates a knowledge graph representing relationships among the plurality of extracted terms. The dialog-pair generation unit generates a dialog repository including a plurality of dialog pairs on the basis of the knowledge graph. Each of the plurality of dialog pairs is a set of a question and a response to the question. A first term included in the question and a second term included in the response are two terms having a connected relationship in the knowledge graph.

(57) 要約: 本開示に係る情報処理装置は、知識オブジェクト生成部と、対話ペア生成部とを備える。前記知識オブジェクト生成部は、半構造化文書から複数の用語を抽出し、抽出された前記複数の用語の間の関係を示す知識グラフを生成する。前記対話ペア生成部は、前記知識グラフに基づいて、複数の対話ペアを含む対話リポジトリを生成する。前記複数の対話ペアの各々は、質問文及び前記質問文に対する応答文の組である。前記質問文に含まれる第1の用語と、前記応答文に含まれる第2の用語とは、前記知識グラフにおいて接続された関係にある2つの用語である。

WO 2021/235225 A1

CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ,
EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN,
HR, HU, ID, IL, IN, IR, IS, IT, JO, JP, KE, KG, KH,
KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY,
MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ,
NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT,
QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL,
ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG,
US, UZ, VC, VN, WS, ZA, ZM, ZW.

- (84) 指定国(表示のない限り、全ての種類の広域保護が可能): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), ユーラシア (AM, AZ, BY, KG, KZ, RU, TJ, TM), ヨーロッパ (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

添付公開書類:

- 一 国際調査報告 (条約第21条(3))

明 細 書

発明の名称： 情報処理装置及び情報処理方法

技術分野

[0001] 本開示は、情報処理装置及び情報処理方法に関する。

背景技術

[0002] スマートフォンやスマートスピーカーの普及に伴い、ユーザから受け付けた発話やユーザの行動等に反応してユーザに適切な応答を行うための対話システムが広く利用されている。対話システムでは、ユーザ発話に対して適切な速度や内容で応答を行うことが求められる。

[0003] 例えば、ユーザからの質問や問合せに対して、ナレッジベース（knowledge base）化された各種の知識情報に応じて、ユーザへの応答の内容を生成するナレッジベースの対話生成に関する技術が知られている。ここで、ナレッジベースとは、コンピュータで利用できる形式で蓄積された知識のデータベースを言う。

先行技術文献

特許文献

[0004] 特許文献1：特開2019-207648号公報

発明の概要

発明が解決しようとする課題

[0005] しかしながら、上記の従来技術では、ユーザ発話を契機として、ユーザ発話に応じてナレッジベースを探索し、ナレッジベースから得られた知識を応答文に変換する。このような中、ナレッジベースの規模が拡大すると、ナレッジベースの探索速度の低下に伴う応答の遅延や探索ルール設計の煩雑化により、ユーザ発話に対する適切な応答が困難であった。

[0006] そこで、本開示では、ナレッジベースの規模を拡大した場合であっても、ユーザ発話に対して適切に応答することができる情報処理装置及び情報処理方法を提案する。

課題を解決するための手段

[0007] 本開示によれば、情報処理装置は、半構造化文書から複数の用語を抽出し、抽出された前記複数の用語の間の関係を示す知識グラフを生成する知識オブジェクト生成部と、前記知識グラフに基づいて、複数の対話ペアを含む対話リポジトリを生成する対話ペア生成部とを備え、前記複数の対話ペアの各々は、質問文及び前記質問文に対する応答文の組であり、前記質問文に含まれる第1の用語と、前記応答文に含まれる第2の用語とは、前記知識グラフにおいて接続された関係にある2つの用語である。

[0008] また、本開示によれば、情報処理装置は、ユーザからの質問文を取得する取得部と、半構造化文書から抽出された第1の用語を含む質問文と、前記半構造化文書から抽出された複数の用語の間の関係を示す知識グラフにおいて前記第1の用語に接続された第2の用語を含む応答文との組である対話ペアを、取得された前記ユーザからの質問文に応じて選択し、選択された前記対話ペアの応答文を前記ユーザからの質問文に対する応答文として出力する推論部とを備える。

図面の簡単な説明

[0009] [図1]実施形態に係る情報処理システムにより実行される情報処理の概要を示す図である。

[図2]実施形態に係る情報処理システムにより実行される情報処理の概要を示す図である。

[図3]実施形態に係る情報処理における入出力データの概要を示す図である。

[図4]実施形態に係る情報処理とは異なる、対話に特化したナレッジベースが生成されない場合の処理の概要を示す図である。

[図5]実施形態に係る情報処理システムの応答生成装置の機能構成の一例を示すブロック図である。

[図6]実施形態に係る情報処理システムの対話リポジトリ生成装置の機能構成の一例を示すブロック図である。

[図7]実施形態に係る情報処理において、半構造化文書から知識グラフを生成

する際のパターンマッチに用いる、半構造化文書の代表的な論理構造を説明するための図である。

[図8]実施形態に係る情報処理において、半構造化文書から知識グラフを生成する際のパターンマッチに用いる、半構造化文書の代表的な参照表記を説明するための図である。

[図9]実施形態に係る情報処理における半構造化知識の知識オブジェクト化の具体例を説明するための図である。

[図10]実施形態に係る情報処理における図9の知識分類ごとの知識オブジェクト化の具体例を説明するための図である。

[図11]実施形態に係る情報処理における図9の知識分類ごとの知識オブジェクト化の具体例を説明するための図である。

[図12]実施形態に係る情報処理における図9の知識分類ごとの知識オブジェクト化の具体例を説明するための図である。

[図13]実施形態に係る情報処理における図9の知識分類ごとの知識オブジェクト化の具体例を説明するための図である。

[図14]実施形態に係る情報処理における図9の知識分類ごとの知識オブジェクト化の具体例を説明するための図である。

[図15]実施形態に係る情報処理における図9の知識分類ごとの知識オブジェクト化の具体例を説明するための図である。

[図16]実施形態に係る情報処理における図9の知識分類ごとの知識オブジェクト化の具体例を説明するための図である。

[図17]実施形態に係る情報処理に係るナレッジベースからの対話ペア群の生成におけるイレギュラーなケースの対応の一例を示す図である。

[図18]実施形態に係る情報処理において、明示的に約款に記載されていない内容を、ペルソナを定義することにより自動補完する具体例を説明するための図である。

[図19]実施形態に係る情報処理において、約款記載の専門用語を、ペルソナのパラメタから算出した数値を用いて具体的な例文に置き換えることにより

自動補完する具体例を説明するための図である。

[図20]実施形態に係る情報処理において、明示的に約款に記載されていない内容を、半構造化文書の他の知識を用いて自動補完する具体例を説明するための図である。

[図21]実施形態に係る情報処理において、明示的に約款に記載されていない内容を、半構造化文書の他の知識を用いて自動補完する具体例を説明するための図である。

[図22]実施形態に係る情報処理において、明示的に約款に記載されていない内容を、約款のアップデート内容を用いて自動補完する具体例を説明するための図である。

[図23]実施形態に係る情報処理におけるコンテキスト情報を用いた対話ペアの選択について説明するための図である。

[図24]実施形態に係る情報処理におけるコンテキスト情報を用いた対話ペアの選択について説明するための図である。

[図25]実施形態に係る情報処理としての対話リポジトリ生成処理の一例を示すフローチャートである。

[図26]実施形態に係る文章ベクトルの生成について説明するための図である。

[図27]実施形態に係る対話リポジトリ生成処理により半構造化文書から生成された対話ペアのデータの一例を示す図である。

[図28]実施形態に係る情報処理としての応答生成処理の一例を示すフローチャートである。

[図29]実施形態に係る情報処理システムの各装置のハードウェア構成の一例を示すブロック図である。

発明を実施するための形態

[0010] 以下に、本開示の実施形態について図面に基づいて詳細に説明する。なお、以下の各実施形態において、同一の部位には同一の符号を付することにより重複する説明を省略する。

[0011] 以下に示す項目順序に従って本開示を説明する。

1. 実施形態

1-1. 実施形態に係る情報処理の概要

1-2. 実施形態に係る情報処理システムの機能構成

1-2-1. 半構造化文書からの対話ペアの生成

1-2-2. 対話ペアの補完

1-2-3. コンテキスト情報の利用

1-3. 実施形態に係る情報処理の手順

1-4. ハードウェア構成

2. 本開示に係る情報処理システムの効果

[0012] (1. 実施形態)

[1-1. 実施形態に係る情報処理の概要]

図1～図4を用いて、実施形態に係る情報処理の概要を説明する。図1及び図2は、実施形態に係る情報処理システム1により実行される情報処理の概要を示す図である。図3は、実施形態に係る情報処理における入出力データの概要を示す図である。図4は、実施形態に係る情報処理とは異なる、対話に特化したナレッジベースKBが生成されない場合の処理の概要を示す図である。

[0013] 図1に示すように、実施形態に係る情報処理は、半構造化文書D1から対話リポジトリDRを生成する処理S1と、ユーザからの質問（発話文U）に対する応答Rを生成する処理S2とを含む。対話リポジトリDRを生成する処理S1は、ユーザからの質問に対する応答Rを生成する処理S2に先立って実行される。ここで、対話リポジトリDRの生成に使用される半構造化文書D1（データセット）としては、約款などの契約文書や法律文書、家庭の医学などの医学書、教科書などのテキスト文書が適宜利用可能である。

[0014] 対話リポジトリDRを生成する処理S1は、半構造化文書D1から複数の対話ペアDPを含む対話リポジトリDRを自動生成する処理である。より詳細には、対話リポジトリDRを生成する処理は、半構造化文書D1から「対

話に特化した」ナレッジベースKBを生成する処理S11と、生成されたナレッジベースKBから対話ペアDPの群（対話ペア群）を生成する処理S12とを含む。

[0015] ナレッジベースKBを生成する処理S11は、半構造化文書D1から知識グラフなどの知識オブジェクト、すなわちナレッジベースKBを自動生成する処理である。より具体的には、ナレッジベースKBを生成する処理S11は、図1に示すように、半構造化文書D1からモデルM1を用いて複数の用語を抽出して知識オブジェクト化することにより、知識グラフ（知識オブジェクト）を生成する。つまり、知識グラフとは、半構造化文書D1から抽出された複数の用語の間の関係を示す半構造化文書D1の知識構造である。半構造化文書D1からの用語の抽出及び知識オブジェクト化は、例えば半構造化文書D1に対するパターンマッチ（モデルM1）により実現可能である。あるいは、半構造化文書D1からの用語の抽出及び知識オブジェクト化は、機械学習モデル（モデルM1）を用いて実現することもできる。機械学習モデルとしては、一例として、半構造化文書D1からのパターンマッチにより抽出されたデータを教師データとして学習された、例えばRNN-LSTM（Recurrent Neural Networks-Long short-term memory）を用いたSeq2seq（Sequence to sequence model）などのDNN（Deep Neural Network）が利用可能である。

[0016] 対話ペア群を生成する処理S12は、生成された知識グラフ、すなわちナレッジベースKBに基づいて、複数の対話ペアDPを含む対話リポジトリDRを自動生成する処理である。知識グラフとして構築されたナレッジベースKBからの対話ペアDPの生成は、テンプレート（モデルM2）に基づき自動生成する。ここで、複数の対話ペアDPの各々は、質問文としての発話文Q1、Q2、Q3（Q文）と、質問文（Q文）に対する応答文A1、A2、A3（A文）との組（QAペア、QA文）である。対話リポジトリDRを生成する処理S1は、DNN（モデルM3）を用いて、対話ペアDPのうち少なくとも質問文（発話文Q1、Q2、Q3）を、それぞれ文章ベクトル（

発話文ベクトル $QV1$ 、 $QV2$ 、 $QV3$)に変換(文章ベクトル化)する処理を含む。生成された対話リポジトリ DR は、 QA 推論モデル(モデル $M5$)の対話リポジトリとして登録する。

[0017] ユーザからの質問(発話文 U)に対する応答 R を生成する処理 $S2$ は、ユーザからの質問(発話テキスト、発話文 U)を取得する処理 $S21$ を含む。ユーザからの質問を取得する処理 $S21$ は、ユーザからの発話文 U の入力に応じた DNN (モデル $M4$)の出力を発話文ベクトル UV として取得する処理を含む。また、応答 R を生成する処理 $S2$ は、対話リポジトリ DR からユーザへの応答 R (応答テキスト、応答文)として出力する対話ペア DP を選択する処理 $S22$ を含む。対話ペア DP を選択する処理 $S22$ は、ユーザ発話の文章ベクトル(発話文ベクトル UV)と、予め生成された対話リポジトリの Q 文の文章ベクトル(発話文ベクトル $QV1$ 、 $QV2$ 、 $QV3$)との比較に基づいて行われる。より具体的には、発話文ベクトル UV に最も近い、かつ、しきい値以上(しきい値により規定される差未満)の対話リポジトリ DR の発話文ベクトル $QV1$ 、 $QV2$ 、 $QV3$ を検索し、検索された対話ペア DP に対応する応答文が選択される。図1に示す例では、発話文ベクトル $QV1$ に対応する応答文 $A1$ が選択される。また、応答 R を生成する処理 $S2$ は、選択された応答 R を出力する処理 $S23$ を含む。この応答 R を出力する処理 $S23$ では、上述のようにして選択された応答文 $A1$ が出力され、ユーザに提示される。このように、ユーザへの応答 R は、ユーザ発話(発話文 U)と対話ペア DP の質問文(発話文 $Q1$ 、 $Q2$ 、 $Q3$)との文章ベクトルの比較結果に基づいて一意に決めることができる。

[0018] また、対話リポジトリ DR を生成する処理 $S2$ においては、図2に示すように、半構造化文書 $D1$ (あるいはナレッジベース KB)に明示的に記載されていない部分を自動的に具体化した QA 対話文を生成する。対話リポジトリ DR を生成する処理 $S2$ は、約款などの半構造化文書 $D1$ を含むデータセット $DS1$ から約款内知識の対話ペア $DP1$ を生成するとともに、他の一般知識を含むデータセット $DS2$ から、約款外知識をさらに含む対話ペア DP

2を生成する。ここで、図2に例示する書き言葉・知識変換器C1は、図1に例示するモデルM1の一例である。また、図2に例示する知識・対話ペア変換器C2は、図1に例示するモデルM2（あるいはモデルM2及びモデルM3）の一例である。また、図2に例示する発話テキストUTは、図1に例示する発話文Uを示すテキストの一例である。また、図2に例示する応答テキストRTは、図1に例示する応答文A1を示すテキストの一例である。これにより、対話ペアDPを一般知識に基づいて補完することができる。対話ペアDPの補完は、例えばペルソナを用いて、個人情報に触れずに個人情報相当の具体的、かつ、不要な情報を除去した対話ペアDPを補完することにより行われる。あるいは、対話ペアDPの補完は、約款に記載されていない一般知識に関する対話を補完することにより行われる。

[0019] なお、実施形態に係るナレッジベースKBにおいて、対話ペアDPの質問文及び応答文の組は、半構造化文書D1から抽出された用語であって、知識グラフにおいて接続された関係にある2つの用語である。図3に例示するように、ナレッジベースKBにおいて「\$（用語）」と「\$（説明文）」とが接続された関係にあるとする。このとき、対話ペアDPとして、「\$（用語）とはなんですか？」という質問文（SRC）と、「\$（説明文）のことで。」という応答文（TGT）との組が生成される。より具体的に説明する。図1は、対話ペアDPにおけるユーザ発話例（Q文）と、応答例（A文）とを例示する。図1に例示するユーザ発話例（Q文）は、「即位礼正殿の儀ってなに」である。また、応答例（A文）は、「即位礼正殿の儀は即位の礼の中心となる即位した天皇が日本国の内外に即位を宣明する儀式です。」である。この例において、「即位礼正殿の儀」と、「即位の礼の中心となる即位した天皇が日本国の内外に即位を宣明する儀式」とは、知識グラフにおいて接続された関係にある2つの用語の一例である。図3に例示する他の用語（「\$（用語）」及び「\$（別称）」、「\$（用語）」及び「\$（具体化）」）についても、同様にして対話ペアDPが生成される。

[0020] 一方で、図4に示すように、本実施形態に係る情報処理とは異なり、対話

に特化したナレッジベースKBが生成されない場合について説明する。対話
に特化したナレッジベースKBが生成されない場合とは、約款などのデータ
セットDS1'を書き言葉・知識変換器C1'により変換したものをナレッ
ジベースKB'とする場合である。この場合、ユーザ発話（発話テキストU
T'）を発話・クエリ変換器C3'により検索クエリSQ'に変換し、検索
クエリSQ'によりナレッジベースKB'を探索する。そして、探索で得ら
れた知識（検索結果SR'）を結果・応答変換器C4'により応答文に変換
することにより、応答文（応答テキストRT'）が出力される。

[0021] したがって、本実施形態に係る情報処理とは異なり、対話に特化したナレ
ッジベースKBが生成されない場合には、ナレッジベースの規模（データサ
イズ）が大きくなると、検索速度が低下する。検索速度の低下は、即時性が
求められるエージェント応答の遅延を発生させる。さらに、ナレッジベース
の規模が大きい場合には、検索ルール設計コストが大きくなる。換言すれ
ば、その知識や対話に精通した設計者がナレッジベースの探索ルールを実装
又は更新する必要があるため、膨大な量の知識のナレッジベースへの変換に
要する人的リソースが大きくなる。

[0022] 以上説明したように、例えば約款のような半構造化文書を知識グラフ化し
、知識グラフに基づいて対話ペア（対話リポジトリ）を生成する実施形態に
係る情報処理によれば、End-to-Endでテキスト文書から直接対話
文を生成できる仕組みを実現することができる。換言すれば、実施形態に係
る情報処理によれば、応答速度遅延問題の改善と、人手による設計を必要と
しないスケーラブルな対話文生成を実現することができる。

[0023] [1-2. 実施形態に係る情報処理システムの機能構成]

図5及び図6を用いて、実施形態に係る情報処理システム1の機能構成の
一例を説明する。図5は、実施形態に係る情報処理システム1の応答生成装
置10の機能構成の一例を示すブロック図である。図6は、実施形態に係る
情報処理システム1の対話リポジトリ生成装置20の機能構成の一例を示す
ブロック図である。ここで、応答生成装置10及び対話リポジトリ生成装置

20は、それぞれ本開示に係る情報処理装置の一例である。

[0024] 図5及び図6に示すように、実施形態に係る情報処理システム1は、応答生成装置10、対話リポジトリ生成装置20及び外部サーバ30を含む。応答生成装置10、対話リポジトリ生成装置20及び外部サーバ30は、ネットワークNを介して通信可能に接続される。

[0025] 実施形態に係る情報処理のうち、ユーザからの質問に対する応答を生成する処理（応答生成処理）は、図5に示す応答生成装置10と、図6に示す対話リポジトリ生成装置20とによって実行される。一方で、実施形態に係る情報処理のうち、半構造化文書から対話リポジトリを生成する処理（対話リポジトリ生成処理）は、図6に示す対話リポジトリ生成装置20によって実行される。

[0026] 応答生成装置10は、本開示に係る応答生成処理を実行する情報処理端末（情報処理装置）である。応答生成装置10が実行する音声認識及び音声による応答処理等は、エージェント（Agent）機能と称される場合がある。また、応答生成装置10は、エージェント機器と称される場合がある。

[0027] 応答生成装置10は、情報処理機能を有する種々のスマート機器により実現されてもよいし、専用デバイスとして実現されてもよい。つまり、応答生成装置10は、汎用のコンピュータに限らず、スマートフォンやタブレット端末、スマートスピーカー、腕時計型や眼鏡型端末などのウェアラブルデバイス、テレビ、エアコン、冷蔵庫等のスマート家電、自動車などのスマートビークル、ドローン、家庭用ロボット等であってもよい。情報処理機能を有する種々の機器としての応答生成装置10は、応答生成処理を実現するためのプログラム（アプリケーション）を実行することにより、本開示に係る応答生成装置10として機能する。

[0028] 応答生成装置10は、図5に示すように、センサ11、入力部12、通信部13、記憶部14、取得部15、応答生成部16及び出力部17を含む。

[0029] センサ11は、各種情報を検知する。例えば、センサ11は、ユーザが発話した音声を集音するマイクロフォンや、ユーザの行動を映像として取得す

るカメラ、ユーザの体温や呼吸、心拍数などのバイタルサインを検知するバイタルセンサを含む。

[0030] 入力部12は、ユーザから各種操作を受け付けるためのデバイスである。例えば、入力部12は、キーボードやマウス、タッチパネル等によって実現される。

[0031] 通信部13は、ネットワークNと有線又は無線で接続され、ネットワークNを介して、対話リポジトリ生成装置20等との間で情報の送受信を行う。

[0032] 記憶部14は、例えば、RAM (Random Access Memory)、フラッシュメモリ (Flash Memory) 等の半導体メモリ素子や、ハードディスク、光ディスク等の記憶装置によって実現される。実施形態では、記憶部14は、例えば、ユーザ情報テーブル141、モデル情報142、制御プログラム143を有する。ユーザ情報テーブル141は、応答生成装置10を利用するユーザに関する情報を記憶する。モデル情報142は、応答生成処理に使用されるモデル (関数) と、パラメタとを記憶する。制御プログラム143は、応答生成処理を実現するためのプログラム (アプリケーション) を記憶する。

[0033] 取得部15は、各種情報を取得する処理部である。なお、取得部15は、入力情報とともに、入力情報が入力された際のコンテキスト情報を取得してもよい。取得部15は、検知部151、登録部152及び受信部153を含む。検知部151は、センサ11を介して、各種情報を検知する。例えば、検知部151は、センサ11の一例であるマイクロフォンを介して、ユーザが発話した音声を検知する。また、検知部151は、センサ11の一例であるカメラを介して、コンテキスト情報として使用される画像情報を検知する。また、検知部151は、センサ11の一例であるバイタルセンサを介して、ユーザの体温や呼吸、心拍数などのバイタルサインを検知する。登録部152は、入力部12を介して、ユーザからの登録を受け付ける。例えば、登録部152は、タッチパネルやキーボードを介して、応答生成装置10を利用するユーザのユーザプロファイル (属性情報) の登録を受け付ける。属性情報は、例えばペルソナの選択に利用される。受信部153は、各種情報を

受信する。例えば、受信部 153 は、対話リポジトリ生成処理において選択された応答文を対話リポジトリ生成装置 20 から受信する。

[0034] 応答生成部 16 は、ユーザからの質問に対する応答を生成する処理（応答生成処理）を実行する。応答生成部 16 は、信号処理部 161、音声認識部 162、応答部 163 及び送信部 164 を含む。信号処理部 161 は、センサ 11 からのユーザの発話音声や画像、バイタルサイン等の入力情報に関する信号処理を行う。音声認識部 162 は、信号処理部 161 によって処理された信号を音声として認識する。応答部 163 は、ユーザからの質問に対する応答を生成する処理において、ユーザからの質問文（発話テキスト、発話文）を取得する処理を実行する。応答部 163 は、音声認識部 162 によって認識されたユーザの発話音声を文章ベクトルに変換する。また、応答部 163 は、選択された応答文を出力する処理を実行する。応答部 163 は、取得部 15 によって取得された対話リポジトリ生成装置 20 からの応答文に基づいて、ユーザへの応答を生成する。例えば、応答部 163 は、取得された応答文を音声データに変換する。送信部 164 は、応答部 163 によって生成されたユーザ発話音声の文章ベクトルを通信部 13 に送信する。送信部 164 は、応答部 163 によって変換された音声データを出力部 17 に送信する。

[0035] 出力部 17 は、種々の情報を出力するための機構である。出力部 17 は、例えばスピーカーである。出力部 17 は、応答生成部 16 から送信された音声データを音声出力する。なお、出力部 17 は、取得部 15 によって取得された応答文を表示するディスプレイであってもよい。

[0036] 対話リポジトリ生成装置 20 は、本開示に係る対話リポジトリ生成処理を実行する情報処理装置である。対話リポジトリ生成装置 20 は、例えば汎用のコンピュータやサーバ等であってもよい。対話リポジトリ生成装置 20 は、対話リポジトリ生成処理を実現するためのプログラム（アプリケーション）を実行することにより、本開示に係る対話リポジトリ生成装置 20 として機能する。

- [0037] 対話リポジトリ生成装置 20 は、図 6 に示すように、通信部 21、記憶部 22、取得部 23 及び対話リポジトリ生成部 24 を含む。
- [0038] 通信部 21 は、ネットワーク N と有線又は無線で接続され、ネットワーク N を介して、応答生成装置 10 や外部サーバ 30 との間で情報の送受信を行う。
- [0039] 記憶部 22 は、例えば、RAM (Random Access Memory)、フラッシュメモリ (Flash Memory) 等の半導体メモリ素子や、ハードディスク、光ディスク等の記憶装置によって実現される。実施形態では、記憶部 14 は、例えば、ユーザ情報テーブル 221、モデル情報 222、対話リポジトリ 223 及び制御プログラム 224 を有する。ユーザ情報テーブル 221 は、応答生成装置 10 の記憶部 14 から取得された情報であり、応答生成装置 10 を利用するユーザに関する情報を記憶する。モデル情報 222 は、対話リポジトリ生成処理に使用されるモデル (関数) と、パラメタとを記憶する。モデル情報 222 は、半構造化文書の知識構造を抽出してナレッジベースを生成する処理に使用されるパターンを記憶する。モデル情報 222 は、ナレッジベースから対話ペア群を生成する処理に使用されるテンプレートを記憶する。対話リポジトリ 223 は、生成された対話ペア群を記憶する。制御プログラム 224 は、対話リポジトリ処理を実現するためのプログラム (アプリケーション) を記憶する。記憶部 22 は、半構造化文書を記憶してもよい。
- [0040] 取得部 23 は、各種情報を取得する処理部である。取得部 23 は、受信部 231 を含む。受信部 231 は、例えば知識オブジェクト化の対象の半構造化文書や一般知識などのテキスト情報を外部サーバ 30 から取得する。また、受信部 231 は、応答生成装置 10 を利用するユーザに関する情報やユーザの発話テキスト (質問文, Q 文)、コンテキスト情報などを応答生成装置 10 から取得する。
- [0041] 対話リポジトリ生成部 24 は、半構造化文書から対話リポジトリを生成する処理 (対話リポジトリ生成処理) を実行する。対話リポジトリ生成部 24 は、知識オブジェクト生成部 241、対話ペア生成部 242、推論部 243

及び送信部 244 を含む。知識オブジェクト生成部 241 は、ナレッジベースを生成する処理を実行する。対話ペア生成部 242 は、生成されたナレッジベースから対話ペア群を含む対話リポジトリを生成する処理を実行する。このとき、推論部 243 は、対話ペアの少なくとも質問文に関して文章ベクトル化する。推論部 243 は、ユーザからの質問に対する応答を生成する処理において、対話リポジトリ 223 からユーザへの応答文として出力する対話ペアを選択する処理と、選択された応答文を出力する処理とを実行する。送信部 244 は、例えば、対話リポジトリ生成処理において選択された応答文を応答生成装置 10 へ通信部 21 を介して送信する。

[0042] 外部サーバ 30 は、知識オブジェクト化の対象の半構造化文書（データセット）を記憶する不揮発性メモリ等のストレージを含む装置である。なお、外部サーバ 30 あるいは当該不揮発性メモリ等のストレージは、ネットワーク N を介さず、対話リポジトリ生成装置 20 に直接接続されるように構成されていてもよい。なお、外部サーバ 30 は、対話リポジトリ生成装置 20 により生成された対話リポジトリ 223 を記憶してもよい。

[0043] なお、応答生成装置 10、対話リポジトリ生成装置 20 及び外部サーバ 30 のうちの少なくとも 2 つの装置は、1 つの装置として実現されても構わない。また、応答生成装置 10、対話リポジトリ生成装置 20 及び外部サーバ 30 は、それぞれ複数の装置により実現されてもよい。

[0044] [1-2-1. 半構造化文書からの対話ペアの生成]

ここで、図 7～図 17 を用いて、半構造化文書から対話リポジトリを生成する処理（対話リポジトリ生成処理）について、より詳細に説明する。

[0045] 図 7 は、半構造化文書から知識グラフを生成する際のパターンマッチに用いる、半構造化文書の代表的な論理構造を説明するための図である。図 7 は、約款の場合を例示するが、他の半構造化文書も同様の論理構造を有する。図 7 に示すように、半構造化文書は、例えば用語定義 S T 1、リンク定義 S T 2、条件文 S T 3、例外事由 S T 4 及び範囲 S T 5 などの論理構造を有する。用語定義 S T 1 は、例えば「A は B である。」のように用語を示す論理

構造である。リンク定義ST2は、例えば「AのとおりB。」のように相関（リンク）を示す論理構造である。条件文ST3は、例えば「Aの場合、B。」や「Aの場合に、Bのときは、C。」、「Aの場合、B。 この場合、C。」のように条件を示す論理構造である。例外事由ST4は、例えば「Aの場合、B。 ただし、Cの場合、D。」のように例外事由を示す論理構造である。範囲ST5は、例えば「A、ただしBを含む。」や「A、ただしBを除く。」のように範囲を示す論理構造である。

[0046] 図8は、半構造化文書から知識グラフを生成する際のパターンマッチに用いる、半構造化文書の代表的な参照表記を説明するための図である。図8は、条件文の場合を例示するが、他の論理構造についても同様に参照表記を有する。図8に示すように、条件文ST3の論理構造は、例えば丸括弧表記E1、複文表記E2、単文表記E3及び入れ子表記E4などの参照表記を有する。丸括弧表記E1は、例えば「B（Aの場合に限ります）。」のように丸括弧を用いて表現される参照表記である。複文表記E2は、例えば「A。この場合、B。」のように複文を用いて表現される参照表記である。単文表記E3は、例えば「Aの場合、B。」のように単文を用いて表現される参照表記である。入れ子表記E4は、例えば「1）次の場合、B。 a）A。 b）A'。」のように入れ子を用いて表現される参照表記である。

[0047] ここで、図9～図16を用いて、半構造化知識の知識オブジェクト（ナレッジベース）化について、より具体的に説明する。図9は、半構造化知識の知識オブジェクト化の具体例を説明するための図である。図9は、半構造化文書（半構造化知識）としての約款に関する。図10～図16は、それぞれ、図9の知識分類ごとの知識オブジェクト化の具体例を説明するための図である。なお、図9～図16に示す分類やパターン、テンプレートなどは一例であり、適宜変更、追加及び削除することができる。

[0048] 図9に例示する知識分類(大)及び知識分類(小)は、それぞれ知識グラフにおけるノードの一例である。図9に示す例では、知識分類(大)は、「用語」、「概要」、「指示」、「支払義務」、「権利」、「禁止」、「行為」、

「参照」、「範囲」、「期日」、「人間」、「計算式」、「条件」、「適用」及び「その他」の項目を含む。

- [0049] 「用語」の知識分類(小)は、「別称」、「定義」及び「読替え」の項目を含む。「概要」の知識分類(小)は、「約款概要」、「条概要」及び「項概要」の項目を含む。「指示」の知識分類(小)は、「払込指示」、「提出指示」、「変更指示」、「通知指示」、「指定指示」及び「請求指示」の項目を含む。「支払義務」の知識分類(小)は、「支払あり」、「支払なし」及び「支払可能性あり」の項目を含む。「権利」の知識分類(小)は、「請求権」、「選択権」、「変更権」、「解除権」、「解約権」、「取消権」、「前納権」、「返済権」、「復活権」、「貸付件」、「復旧権」、「照会権」、「承諾許可」及び「返済権」の項目を含む。「禁止」の知識分類(小)は、「復活禁止」、「解除禁止」、「変更禁止」、「対抗禁止」、「使用禁止」及び「公開禁止」の項目を含む。「行為」の知識分類(小)は、「確認」、「交付」、「割引」、「継続」、「通知」、「取扱い」、「処理」、「払戻」、「登録」、「設定」、「返済」、「判断」、「付加」及び「積増し」の項目を含む。「参照」の知識分類(小)は、「次のとおり」の項目を含む。「範囲」の知識分類(小)は、「含む」及び「限る」の項目を含む。「期日」の知識分類(小)は、「効力失効日」、「責任発生時」、「猶予期間」、「登録期間」、「更改日」、「基準日」及び「効力発生日」の項目を含む。「人間」の知識分類(小)は、「受取人」、「契約者」及び「代理人」の項目を含む。「計算式」の知識分類(小)は、「金属定義」、「計算方法」、「最低金額」、「差引」、「繰入」、「充当」、「割合」、「切捨て」及び「減額」の項目を含む。「条件」の知識分類(小)は、「金額」、「払込」及び「告知」の項目を含む。「適用」の知識分類(小)は、「適用」、「準用」及び「適用外」の項目を含む。「その他」の知識分類(小)は、「責任の所在」、「管轄裁判所」、「効力無効」、「消滅」、「配当なし」及び「払戻なし」の項目を含む。

- [0050] 図10は、「用語」ノードに関する知識オブジェクト化及び対話ペア化を例示する。図10に、知識分類「用語」を半構造化文書から抽出するための

パターンの一例を示す。「別称」のパターンは、例えば「\$（用語）（いわゆる「\$（別称）」）」である。「定義」のパターンは、例えば「\$（定義）（以下「\$（用語）」といいます。）」、「\$（用語）（\$（定義）のことをいいます。以下同じ。）」及び「\$（用語）は、\$（定義）とします。」である。「読替え」のパターンは、例えば「「\$（用語）」は「\$（読替え）」と読替えます。」及び「\$（用語1）、\$（用語2）、\$（用語3）、\$（用語4）、\$（用語5）、\$（用語6）とあるのは、…それぞれ、\$（読替え1）、\$（読替え2）、\$（読替え3）、\$（読替え4）、\$（読替え5）、\$（読替え6）と読替えます。」である。「範囲」のパターンは、例えば「□□については、\$（範囲）を\$（用語）とします。」である。

[0051] 上述したように、半構造化文書は、ある論理構造及び参照表記を用いて記述されている文書である。ここで、図9及び図10に示すように、各知識分類は、知識グラフの各ノード（node）に対応する。また、論理構造及び参照表記に応じたパターン（モデル情報222）は、知識分類（ノード）ごとに規定される。

[0052] 例えば、図10に示すように、例えば、半構造化文書から、「用語」、「別称」のパターンである「\$（用語）（いわゆる「\$（別称）」）」を用いて、「「頸部症候群」（いわゆる「むちうち症」）」との記載が抽出されたとする。ここで、パターンマッチで使用するパターン（モデル情報222）のうちの下線を付している部分は、知識グラフ化の際に削除される。このとき、「頸部症候群」と、「（いわゆる「むちうち症」）」とは、図10に示すように、知識グラフにおいて用語の別称として接続されることになる。このように、知識オブジェクト生成部241は、論理構造及び参照表記に応じたパターン（モデル情報222）を用いてパターンマッチを行うことにより、半構造化文書の知識構造を抽出することができる。

[0053] 次に、図10を用いて、生成されたナレッジベースからの対話ペア群の生成について、より具体的に説明する。図9は、各知識分類の値（用語）から

生成される対話ペアの種類である「QAの種類」の一例と、各知識分類について生成される対話ペアの一例とを例示する。ここで、図9に例示する各対話ペアにおいて、下線部は、対話ペアの生成に用いるテンプレート（モデル情報222）の部分を示す。

[0054] 対話ペア生成部242は、図10に示すように、例えば知識グラフにおいて接続されている「用語」及び「別称」をテンプレート（モデル情報222）に挿入することにより、対話ペアを生成する。具体的には、対話ペア生成部242は、「頸部症候群」をテンプレートに挿入して、「頸部症候群とはなんですか?」というQ文を生成する。同様に、対話ペア生成部242は、「いわゆる「むちうち症」」をテンプレートに挿入して、「いわゆる「むちうち症」のことです。」というA文を生成する。

[0055] 図11は、「概要」ノードに関する知識オブジェクト化及び対話ペア化を例示する。図11に、知識分類「概要」を半構造化文書から抽出する対象の一例を示す。図12は、「指示」ノード及び「条件付き指示」ノードに関する知識オブジェクト化及び対話ペア化を例示する。図13は、「支払義務」ノード及び「条件付き支払義務」ノードに関する知識オブジェクト化及び対話ペア化を例示する。図14は、「権利」ノード及び「条件付き権利」ノードに関する知識オブジェクト化及び対話ペア化を例示する。図15は、「禁止」ノード及び「条件付き禁止」ノードに関する知識オブジェクト化及び対話ペア化を例示する。図16は、「行為」ノード及び「条件付き行為」ノードに関する知識オブジェクト化及び対話ペア化を例示する。知識オブジェクト生成部241は、これらの場合も図10を用いて説明した場合と同様にして、抽出する対象のパターンに従いパターンマッチを行うことにより知識グラフを生成することができる。例えば図12～図15に示すように条件節が抽出されたとき、「条件付き〇〇」の知識分類（ノード）が生成される。例えば図12に例示する「条件付き権利」ノードには、条件節の記載（用語）と、主節の記載に関する「権利」ノードとが接続されることになる。また、対話ペア生成部242は、図10を用いて説明した場合と同様にして、知識

グラフにおいて接続されている用語をテンプレート（モデル情報 2 2 2）に挿入することにより、対話ペアを生成することができる。

[0056] このように、対話ペア生成部 2 4 2 は、半構造化文書から「対話に特化した」ナレッジベースが生成された後、ナレッジベースから対話ペア群を生成する。しかしながら、半構造化文書を対話 Q A 文に変換する際には、テンプレートだけでは生成できないイレギュラーなケースが発生する場合もある。図 1 7 は、ナレッジベースからの対話ペア群の生成におけるイレギュラーなケースの対応の一例を示す図である。図 1 7 に示すように、例えば約款と約款に付帯する F A Q との間で表現が異なる場合があり得る。図 1 7 は、約款に載っているのは「入院給付金日額」だが、F A Q では「保障額」と言っている場合と、約款に載っているのは「本条の請求をするとき」だが、F A Q では「解約をするとき」と言っている場合とを例示する。ここで、約款から抽出された用語は、テンプレートを用いて対話ペア化が可能である。したがって、対話ペア生成部 2 4 2 は、F A Q の用語（保障額、解約する）を用いた質問文を、言い回しバリエーションの質問文として追加で生成する。このとき、F A Q の用語を用いた質問文は、例えば約款から抽出された用語を用いた質問文と同一の応答文に対する対話ペアとして生成することができる。また、図 1 7 に示すように、約款には支払われるケースについて記載がある一方、支払われないケースについては記載されていないなどの場合があり得る。このようなとき、対話ペア生成部 2 4 2 は、約款に記載されているケース以外の場合に関する質問文を、ネガティブなケースの質問文として追加で生成する。ネガティブなケースの応答文は、テンプレートを使用する場合と同様に、知識グラフにおける接続関係から生成することができる。

[0057] [1-2-2. 対話ペアの補完]

ここで、図 1 8 ~ 図 2 2 を用いて、半構造化知識の知識オブジェクト（ナレッジベース）化における対話ペアの補完について、より具体的に説明する。対話ペア生成部 2 4 2 は、図 2 を用いて上述したように、半構造化文書（あるいはナレッジベース）に明示的に記載されていない部分を自動的に具体

化したQ A対話文を生成することができる。

[0058] 図18は、明示的に約款に記載されていない内容を、ペルソナを定義することにより自動補完する具体例を説明するための図である。図18は、ユーザ発話（質問文）と、その質問文に対応する応答文として約款知識から自動生成した応答文と、契約例に応じたペルソナを定義して自動生成した応答文との対応の一例を示す。図19は、約款記載の専門用語を、ペルソナのパラメタから算出した数値を用いて具体的な例文に置き換えることにより自動補完する具体例を説明するための図である。図19に示す例において、ペルソナは、項目事例として「被保険者」、「性別」、「1入院の支払限度」、「入院給付金日額」、「保険期間」、「保険料払込期間」及び「個別扱月払保険料」を含む。これらの各項目に対して、パラメタが予め定義されているとする。図18及び図19に示す例において、下線部分は、約款知識から上述のようにして自動生成した対話Q A文である。また、二重下線部分は、ペルソナを用いて具体化した文章であって、補完される対話Q A文である。ペルソナのパラメタは、例えばユーザ情報テーブル221が記憶している。例えば、対話ペア生成部242は、図18に示すように、「だれに支払われますか？」というユーザ発話（質問文）に対応する応答文として、約款知識からは「保険契約者に支払われます。」という応答文を自動生成する。一方で、対話ペア生成部242は、保険契約者のペルソナに応じて、「たとえば、お客様ご本人が保険契約者の場合、お客様に支払われます。」という表現を追加した応答文を自動生成する。例えば、対話ペア生成部242は、図18及び図19に示すように、「既払込保険料」という用語に対して、定義されたペルソナのパラメタと、約款に記載された既払込保険料の算出方式（演算式）とに基づいて、具体的な応答文を生成し、具体化したQ A対話文を自動的に追加する。つまり、対話ペア生成部242は、定義されたペルソナのパラメタに応じて払込保険料の金額を算出する。そして、対話ペア生成部242は、「たとえば、お客様が、月々4,610円の保険料を5年間支払っていた場合、支払った保険料の総額は・・・28万円になります。この28万円

が、既払込保険料です。」というように具体化した応答文を自動的に追加する。ここで、既払込保険料という用語と、算出方式とは、知識グラフにおいて接続された関係にある。

[0059] なお、図19に示すペルソナのパラメタは一例であり、これに限らない。また、定義されるペルソナは、2つ以上の複数であってもよい。定義されるペルソナは、例えば約款に対応するパンフレットなどで提示されている契約例の項目に対して、代表的なパラメタ違いのバリエーションを用意することができる。なお、パラメタ違いのバリエーションは、例えば全契約者のプロフィールをプロットし、プロットに対するクラスタリングにより代表的なパラメタを抽出して決めることもできる。

[0060] このように、ペルソナを用いることにより、対話ペア生成部242は、テンプレートを用いて約款知識をそのまま対話QA文に変換した場合と比較して、人間が説明する形に近く、分かりやすい応答文を生成し、対話リポジトリに追加することができる。

[0061] 図20及び図21は、明示的に約款に記載されていない内容を、半構造化文書の他の知識を用いて自動補完する具体例を説明するための図である。図20は、ユーザ発話（質問文）と、その質問文に対応する応答文として約款知識から自動生成した応答文と、一般知識を用いて自動生成した応答文との対応の一例を示す。図20及び図21に示す例において、下線部分は、約款知識から上述のようにして自動生成した対話QA文である。また、二重下線部分は、一般知識を用いて具体化した文章であって、補完される対話QA文である。図20に示すように、「骨髄移植給付金」という用語については、知識グラフに基づいて対話ペアを生成できているとする。このような中、「骨髄移植給付金」に対応する用語のうちの「公的医療保険制度」や「骨髄移植術」などの用語については、約款に定義などの記載がないとする。このように約款に記載がない用語の対話ペアについては、約款からのナレッジベースを用いて生成することができない。そこで、知識オブジェクト生成部241は、図21に例示するように、約款に記載がない用語を、オントロジーや

W i k i P e d i a などの半構造化文書の他の一般知識を用いて知識グラフ化する。図 20 及び図 21 の例では、例えば、一般知識から「骨髄移植術」とは、「白血病や再生不良性貧血などの血液難病の患者に、提供者（ドナー）の正常な骨髄細胞を静脈内に注入して移植する治療」であることが得られている。このように、一般知識を用いて具体化した Q A 対話文（対話ペア）を自動的に追加することができる。ここで、一般知識としては、ネットワーク N を介してインターネット上のウェブサイトに記載された知識も利用可能である。

[0062] 図 22 は、明示的に約款に記載されていない内容を、約款のアップデート内容を用いて自動補完する具体例を説明するための図である。図 22 に示す例において、下線部分は、約款のアップデートに伴い変更があった部分を示す。また、二重下線部分は、約款のアップデート内容を用いて具体化した文章であって、補完される対話 Q A 文である。対話ペア生成部 242 は、例えば、約款の更新と、契約者の引き継ぎとが可能な約款については、約款のアップデート前後の差分自体をナレッジベース化する。このとき、対話ペア生成部 242 は、アップデート前後の約款の各々について対話ペアを生成するとともに、アップデート前の旧版に関して、アップデート後の新版がある旨の対話ペアを自動生成し、対話リポジトリに追加する。図 22 は、約款のアップデートに伴い、契約時の約款における「月払保険料相当額×12×契約」という「既払込保険料（相当額）」の用語の定義が、最新の約款では「月払保険料相当額×11×契約」という定義に更新されている場合を例示する。このようなとき、対話ペア生成部 242 は、「たとえば・・・万円が、既払込保険料です。なお、2019年4月に更新された最新の約款では、12ヶ月から11ヶ月に変更・・・。」というように、約款のアップデート内容を補完した応答文を自動生成する。なお、約款の一部についてのみ内容がアップデートされる場合もあり得る。このような場合、対話ペア生成部 242 は、アップデート前後の両版の内容について、それぞれ対話ペアを生成する。このとき、対話ペア生成部 242 は、図 22 に例示するように、「なお、

新しい条項では」といった言い回しを挿入する。このように、約款のアップデート内容を用いて対話ペアを補完することもできる。

[0063] [1-2-3. コンテキスト情報の利用]

ここで、図23及び図24を用いて、対話リポジトリ223からユーザへの応答文として出力する対話ペアを選択する処理の一例について説明する。図23及び図24は、それぞれコンテキスト情報を用いた対話ペアの選択について説明するための図である。

[0064] 図23は、半構造化文書D1として、保険約款などの約款や、法庫などの法律文書、家庭の医学などの医療文書、管理規約などの規約、文法書などの教科書が適宜利用される場合を例示する。このようなとき、応答生成装置10の応答部163は、ユーザの発話文Uに加えて、ユーザプロフィール（属性情報）などのコンテキスト情報UC1やセンサ11により得られたバイタルサインや画像などのコンテキスト情報UC2をさらに文章ベクトル化する。つまり、応答生成装置10から対話リポジトリ生成装置20に出力される文章ベクトルは、図23に示すように、ユーザの発話文U及びコンテキスト情報UC1、UC2が結合された文書ベクトルである。なお、画像としては、医療情報であれば、診断書やノートなどの関連する書類の画像や、調べたい症状を撮影して得た症状画像などが適宜利用可能である。

[0065] なお、ユーザの発話文Uに加えてコンテキスト情報UC1、UC2が用いられる場合に、上述したように、複数のペルソナが定義されたり、一般文書が知識として利用されたりして対話ペアDPを補完することもできる。ただし、この場合には、用いられるコンテキスト情報UC1、UC2によっては、冗長な応答文が大量に生成される可能性がある。しかしながら、実施形態に係る推論部243は、応答生成装置10の応答部163からの文章ベクトルの類似度に応じて、ユーザへの応答文として出力する対話ペアDPを絞り込むことができる。具体的には、推論部243は、保険の契約者を示すユーザプロフィールに応じて、約款に関する対話ペアDPを優先して選択する。あるいは、推論部243は、複数のペルソナA、B、C、・・・、Xのうち

のユーザプロファイルに応じた定義のペルソナBを用いて生成された対話ペアDPBを優先して選択する。この構成によれば、ユーザの発話文Uがコンテキスト情報UC1, UC2により補われた文章ベクトルにより、対話ペアDPBを選択することができる。このため、ユーザの発話負担を低減することができる。

[0066] 図24は、半構造化文書のデータセットDS1として、家庭の医学などの医療文書が利用される場合を例示する。図24に示す例では、応答部163の画像・病名変換器C5は、症状画像I10を示すコンテキスト情報UC21を医療情報のデータセットDS3に基づいて病名に変換し、画像テキストUCT1の「水いぼ」を生成する。また、応答部163のバイタルセンサ・症状変換器C6は、バイタルセンサの出力を示すコンテキスト情報UC22の体温「36.8度」をバイタルサインのデータセットDS4に基づいて症状に変換し、センサテキストUCT2の「平熱」を生成する。これらのコンテキスト情報UC21, UC22を結合した発話テキストUTの文章ベクトルに基づいて、推論部243がユーザへの応答文（応答テキストRT）を選択することにより、疑似診断のような仕組みを実現することができる。換言すれば、実施形態に係る技術は、データセットが医療情報になった場合でも適用可能であり、ユーザ発話、画像、バイタルセンサの情報を入力として、医療知識から該当する症状を自動簡易診断するシステムを実現することができる。

[0067] [1-3. 実施形態に係る情報処理の手順]

次に、図25～図28を用いて、実施形態に係る情報処理の手順について説明する。

[0068] 図25は、実施形態に係る情報処理としての対話リポジトリ生成処理の一例を示すフローチャートである。

[0069] 知識オブジェクト生成部241は、任意の半構造化文書を開く（S101）。知識オブジェクト生成部241は、文書の各文を書き言葉・知識変換器に入力し、ナレッジベースを生成する（S102）。ここで、書き言葉・知

知識変換器は、図1～図3、図6などを用いて上述したように、モデル情報222が記憶するモデルの一例である。一例として、書き言葉・知識変換器は、半構造化文書の文の入力に応じて知識オブジェクトを出力するように、パターンマッチで半構造化文書から抽出されたデータを用いてパラメタが学習された、Seq2seqなどのDNNである。別の一例として、書き言葉・知識変換器は、パターンマッチにより半構造化文書の文の入力に応じて知識オブジェクトを出力するように設計された関数である。文書のすべてをナレッジベース化していないとき（S103：No）、図25の流れは、S101及びS102の処理を繰り返す。

[0070] 一方で、すべての文書をナレッジベース化したとき（S103：Yes）、対話ペア生成部242は、ナレッジベースから該当する知識を検索し、知識・対話変換器に入力し、Q文とA文との組、すなわち対話ペアを生成する（S104）。ここで、知識・対話変換器は、図1～図3、図6などを用いて上述したように、モデル情報222が記憶するモデルの一例である。一例として、知識・対話変換器は、ナレッジベースから検索された知識をテンプレートに挿入することにより、各用語の対話ペアを順次生成することができる。

[0071] また、対話ペア生成部242は、生成されたQA文、すなわち対話ペアの中から約款知識外の例文が追加でき得る用語を抽出する（S105）。S105の処理において約款知識外の例文が追加でき得る用語が抽出された場合には、対話ペア生成部242は、用語に関連する知識を拡張したQA文を追加（対話ペアを補完）する（S106）。その後、図25の流れは、ナレッジベースをすべてQA文に変換するまで（S107：No）、S104～S106の処理を繰り返し、ナレッジベースをすべてQA文に変換したとき（S107：Yes）、終了する。

[0072] なお、対話ペア生成部242は、S104及びS106の処理においてQ文が生成されたことを契機として、生成されたQ文を文章・ベクトル変換器に入力し、検索用の文章ベクトルを生成してもよい。図26は、文章ベクトル

ルの生成について説明するための図である。一例として、対話ペア生成部 242 は、図 26 に示すように、文章・ベクトル変換器（モデル M3, M4）を用いて Q 文を文章ベクトル化する。文章・ベクトル変換器としては、例えば学習済みの大規模モデル（機械学習モデル）である BERT (Bidirectional Encoder Representations from Transformers) が利用可能である。文章・ベクトル変換器は、モデル情報 222 が記憶するモデルの一例である。なお、文章ベクトル化の手段はこれに限らず、Word2vec や ELMo などの他の機械学習モデルが利用されても構わない。

[0073] 図 27 は、図 25 を用いて説明した対話リポジトリ生成処理により半構造化文書（知識）から生成された QA 文（対話ペア）のデータの一例を示す。このように、実施形態に係る対話リポジトリ生成処理によれば、半構造化文書（知識）から QA 文（対話ペア）を自動生成することができる。

[0074] 図 28 は、実施形態に係る情報処理としての応答生成処理の一例を示すフローチャートである。図 28 の流れは、図 27 に示すような対話リポジトリ生成処理により半構造化文書（知識）から生成された QA 文（対話ペア）のデータを用いて、ユーザ発話に対する応答を生成する。

[0075] 応答部 163 は、センサ 11 により得られた画像データを、画像・文章変換器に入力し、テキストデータに変換する (S201)。ここで、画像・文章変換器は、図 23 及び図 24 などを用いて上述したように、モデル情報 142 が記憶するモデルの一例である。一例として、画像・文章変換器は、画像データの入力に応じてテキストデータを出力するようにパラメタが決定された機械学習モデルである。応答部 163 は、センサ 11 により得られたバイタルセンサなどのセンシングデータを、センシングログ・文章変換器に入力し、テキストデータに変換する (S202)。ここで、センシングログ・文章変換器は、図 23 及び図 24 などを用いて上述したように、モデル情報 142 が記憶するモデルの一例である。一例として、センシングログ・文章変換器は、センシングデータの入力に応じてテキストデータを出力するようにパラメタが決定された機械学習モデルである。また、応答部 163 は、ユ

ーザプロファイルなどの他のコンテキスト情報をテキストデータに変換する（S203）。その後、応答部163は、図23及び図24などを用いて上述したように、任意のユーザ発話テキストデータと、変換後のテキストデータとを結合し、文章・ベクトル変換器を用いて文章ベクトルを生成する。この文章・ベクトル変換器は、モデル情報142が記憶するモデルの一例である。なお、文章・ベクトル変換器としては、BERTやWord2vec、ELMoなどの種々の機械学習モデルが適宜利用可能である。

[0076] 対話ペア生成部242は、図25を用いて上述した対話リポジトリ生成処理のS104及びS106の処理で変換されていない場合には、ナレッジベースから生成されたQ文を文章・ベクトル変換器に入力し、検索用の文章ベクトルを生成してもよい（S205）。このとき、A文をさらに文章ベクトルに変換しても構わない。

[0077] 対話ペア生成部242は、対話リポジトリに対応するQ文のうち、応答部163からのユーザ発話に対応する文章ベクトルに対してベクトル距離が最も近いQ文を検索して選択する（S206）。その後、対話ペア生成部242は、選択されたQ文に対応するA文を応答として応答生成装置10に送信する。応答生成装置10は、対話ペア生成部242からのA文をユーザに提示する。その後、図28の流れは終了する。

[0078] このように、実施形態に係る応答生成処理によれば、半構造化文書（知識）から自動生成されたQA文（対話ペア）から適切な応答を選択することができる。

[0079] なお、実施形態に係る各機械学習モデルは、複数の関数が合成されたパラメタ付き合成関数であり、複数の調整可能な関数及びパラメタの組合せにより定義されたとする。機械学習モデルは、複数の調整可能な関数及びパラメタの組合せにより定義される如何なるパラメタ付き合成関数であってもよい。なお、機械学習モデルは、畳み込みニューラルネットワーク（CNN）であってもよいし、全結合のネットワークであってもよい。なお、学習済みの機械学習モデルのパラメタは、例えば記憶部14及び記憶部22にそれぞれ

記憶されているとする。

[0080] [1-4. ハードウェア構成]

図29は、実施形態に係る情報処理システム1の各装置のハードウェア構成の一例を示すブロック図である。上述してきた実施形態に係る情報処理システム1の各装置（応答生成装置10、対話リポジトリ生成装置20及び外部サーバ30）等の情報機器は、例えば図29に示すような構成のコンピュータ1000によって実現される。

[0081] コンピュータ1000は、CPU1100、RAM1200、ROM (Read Only Memory) 1300、HDD (Hard Disk Drive) 1400、通信インターフェイス1500、及び入出力インターフェイス1600を有する。コンピュータ1000の各部は、バス1050によって通信可能に接続される。

[0082] CPU1100は、ROM1300又はHDD1400に格納されたプログラムに基づいて動作し、各部の制御を行う。例えば、CPU1100は、ROM1300又はHDD1400に格納されたプログラムをRAM1200にロードし、各種プログラムに対応した処理を実行する。

[0083] ROM1300は、コンピュータ1000の起動時にCPU1100によって実行されるBIOS (Basic Input Output System) 等のブートプログラムや、コンピュータ1000のハードウェアに依存するプログラム等を格納する。

[0084] HDD1400は、CPU1100によって実行されるプログラム、及び、かかるプログラムによって使用されるデータ等を非一時的に記録する、コンピュータが読み取り可能な記録媒体である。具体的には、HDD1400は、プログラムデータ1450の一例である本開示に係る制御プログラム143, 224を記録する記録媒体である。

[0085] 通信インターフェイス1500は、コンピュータ1000がネットワークNや外部ネットワーク1550（例えばインターネット）と接続するためのインターフェイスである。例えば、CPU1100は、通信インターフェイ

ス1500を介して、他の機器からデータを受信したり、CPU1100が生成したデータを他の機器へ送信したりする。

[0086] 入出インターフェイス1600は、入出力デバイス1650とコンピュータ1000とを接続するためのインターフェイスである。例えば、CPU1100は、入出インターフェイス1600を介して、キーボードやマウス等の入力デバイスからデータを受信する。また、CPU1100は、入出インターフェイス1600を介して、ディスプレイやスピーカーやプリンタ等の出力デバイスにデータを送信する。また、入出インターフェイス1600は、所定の記録媒体（メディア）に記録されたプログラム等を読み取るメディアインターフェイスとして機能してもよい。メディアとは、例えばDVD (Digital Versatile Disc)、PD (Phase change rewritable Disk) 等の光学記録媒体、MO (Magneto-Optical disk) 等の光磁気記録媒体、テープ媒体、磁気記録媒体、または半導体メモリ等である。

[0087] 例えば、コンピュータ1000のCPU1100は、RAM1200上にロードされた制御プログラム143を実行することにより、応答生成装置10の有する各機能を実現する。例えば、コンピュータ1000のCPU1100は、RAM1200上にロードされた制御プログラム224を実行することにより、対話リポジトリ生成装置20の有する各機能を実現する。また、HDD1400には、本開示に係る制御プログラム143、224や、記憶部14、22内のデータが格納される。なお、CPU1100は、プログラムデータ1450をHDD1400から読み取って実行するが、他の例として、外部ネットワーク1550を介して、他の装置からこれらのプログラムを取得してもよい。

[0088] (2. 本開示に係る情報処理システムの効果)

対話リポジトリ生成装置20（情報処理装置）は、知識オブジェクト生成部241と、対話ペア生成部242とを備える。知識オブジェクト生成部241は、半構造化文書から複数の用語を抽出し、抽出された複数の用語の間の関係を示す知識グラフを生成する。対話ペア生成部242は、知識グラフ

に基づいて、複数の対話ペアを含む対話リポジトリを生成する。複数の対話ペアの各々は、質問文及び質問文に対する応答文の組である。質問文に含まれる第1の用語と、応答文に含まれる第2の用語とは、知識グラフにおいて接続された関係にある2つの用語である。

[0089] これにより、対話リポジトリ生成装置20（情報処理装置）は、半構造化文書（知識）からQA文（対話ペア）を自動生成することができる。したがって、ナレッジベースの規模を拡大した場合であっても、探索ルール設計の煩雑化を抑制できるため、ユーザ発話に対して適切に応答することができる。

[0090] 対話リポジトリ生成装置20（情報処理装置）において、知識オブジェクト生成部241は、半構造化文書の他の知識情報から抽出された用語をさらに含む知識グラフを生成し、対話ペア生成部242は、半構造化文書の他の知識情報から抽出された用語を用いて対話ペアを補完する。

[0091] これにより、対話リポジトリ生成装置20（情報処理装置）は、約款などの半構造化文書に記載されていない一般知識を用いて、対話ペアを補完することができる。したがって、ユーザからの質問文に対する応答として適切な応答文を生成することができる。

[0092] 対話リポジトリ生成装置20（情報処理装置）において、対話ペア生成部242は、少なくとも1つのペルソナを用いて対話ペアを補完する。

[0093] これにより、テンプレートを用いて約款知識をそのまま対話QA文に変換した場合と比較して、人間が説明する形に近く、ユーザにとって分かりやすい応答文を生成することができる。

[0094] 対話リポジトリ生成装置20（情報処理装置）は、取得部23と、推論部243とを備える。取得部23は、ユーザからの質問文を取得する。推論部243は、半構造化文書から抽出された第1の用語を含む質問文と、半構造化文書から抽出された複数の用語の間の関係を示す知識グラフにおいて第1の用語に接続された第2の用語を含む応答文との組である対話ペアを、取得されたユーザからの質問文に応じて選択し、選択された対話ペアの応答文を

ユーザからの質問文に対する応答文として出力する。

- [0095] これにより、ユーザ発話を契機として応答文を生成するとき、ナレッジベースを検索することなく、対話リポジトリからユーザからの質問文に対する応答文を選択することができる。したがって、ナレッジベースの規模を拡大した場合であっても、ナレッジベースの探索速度の低下に伴う応答の遅延を抑制できるため、ユーザ発話に対して適切に応答することができる。
- [0096] 対話リポジトリ生成装置 20（情報処理装置）において、推論部 243 は、対話ペアの質問文と、ユーザからの質問文との間における文書ベクトルのベクトル距離に基づいて対話ペアを選択する。
- [0097] これにより、ナレッジベースの規模を拡大した場合であっても、ナレッジベースの探索速度の低下に伴う応答の遅延を抑制できるため、ユーザ発話に対して適切に応答することができる。
- [0098] 対話リポジトリ生成装置 20（情報処理装置）において、推論部 243 は、ユーザからの質問文及びユーザからの質問文に対応するコンテキスト情報を結合して生成された文書ベクトルに基づいて対話ペアを選択する。
- [0099] これにより、ユーザ発話の質問文をコンテキスト情報により補うことができる。したがって、ユーザ発話の負担を低減するとともに、ユーザからの質問文に対してより適切な応答文を生成することができる。
- [0100] 対話リポジトリ生成装置 20（情報処理装置）において、取得部 23 は、ユーザからの質問文に対応するコンテキスト情報をさらに取得し、推論部 243 は、取得されたコンテキスト情報に対応する対話ペアを優先的に選択する。
- [0101] これにより、ペルソナやコンテキスト情報の利用に伴い大量の対話ペアが補完された場合であっても、ユーザプロファイルなどのコンテキストに近い対話ペアを一意に選択することができる。したがって、ユーザ発話の負担を低減するとともに、ユーザからの質問文に対してより適切な応答文を生成することができる。
- [0102] 対話リポジトリ生成装置 20（情報処理装置）において、コンテキスト情

報は、画像、バイタルサイン、ユーザプロフィールのうちの少なくとも1つを含む。

[0103] これにより、ユーザ発話の負担を低減するとともに、ユーザからの質問文に対してより適切な応答文を生成することができる。

[0104] 情報処理方法は、半構造化文書から複数の用語を抽出し、抽出された複数の用語の間関係を示す知識グラフを生成することと、知識グラフに基づいて、複数の対話ペアを含む対話リポジトリを生成することとを含み、複数の対話ペアの各々は、質問文及び質問文に対する応答文の組であり、質問文に含まれる第1の用語と、応答文に含まれる第2の用語とは、知識グラフにおいて接続された関係にある2つの用語である。

[0105] これにより、対話リポジトリ生成装置20（情報処理装置）は、半構造化文書（知識）からQA文（対話ペア）を自動生成することができる。したがって、ナレッジベースの規模を拡大した場合であっても、探索ルール設計の煩雑化を抑制できるため、ユーザ発話に対して適切に応答することができる。

[0106] 情報処理方法は、ユーザからの質問文を取得することと、半構造化文書から抽出された第1の用語を含む質問文と、半構造化文書から抽出された複数の用語の間関係を示す知識グラフにおいて第1の用語に接続された第2の用語を含む応答文との組である対話ペアを、取得されたユーザからの質問文に応じて選択し、選択された対話ペアの応答文をユーザからの質問文に対する応答文として出力することを含む。

[0107] これにより、ユーザ発話を契機として応答文を生成するとき、ナレッジベースを検索することなく、対話リポジトリからユーザからの質問文に対する応答文を選択することができる。したがって、ナレッジベースの規模を拡大した場合であっても、ナレッジベースの探索速度の低下に伴う応答の遅延を抑制できるため、ユーザ発話に対して適切に応答することができる。

[0108] なお、本明細書に記載された効果はあくまで例示であって限定されるものではなく、また他の効果があってもよい。

[0109] なお、本技術は以下のような構成も取ることができる。

(1)

半構造化文書から複数の用語を抽出し、抽出された前記複数の用語の間の関係を示す知識グラフを生成する知識オブジェクト生成部と、

前記知識グラフに基づいて、複数の対話ペアを含む対話リポジトリを生成する対話ペア生成部と

を備え、

前記複数の対話ペアの各々は、質問文及び前記質問文に対する応答文の組であり、

前記質問文に含まれる第1の用語と、前記応答文に含まれる第2の用語とは、前記知識グラフにおいて接続された関係にある2つの用語である、

情報処理装置。

(2)

前記知識オブジェクト生成部は、前記半構造化文書の他の知識情報から抽出された用語をさらに含む前記知識グラフを生成し、

前記対話ペア生成部は、前記半構造化文書の他の知識情報から抽出された用語を用いて前記対話ペアを補完する、

前記(1)に記載の情報処理装置。

(3)

前記対話ペア生成部は、少なくとも1つのペルソナを用いて前記対話ペアを補完する、前記(1)又は(2)に記載の情報処理装置。

(4)

ユーザからの質問文を取得する取得部と、

半構造化文書から抽出された第1の用語を含む質問文と、前記半構造化文書から抽出された複数の用語の間の関係を示す知識グラフにおいて前記第1の用語に接続された第2の用語を含む応答文との組である対話ペアを、取得された前記ユーザからの質問文に応じて選択し、選択された前記対話ペアの応答文を前記ユーザからの質問文に対する応答文として出力する推論部と

を備える情報処理装置。

(5)

前記推論部は、前記対話ペアの質問文と、前記ユーザからの質問文との間における文書ベクトルのベクトル距離に基づいて前記対話ペアを選択する、前記(4)に記載の情報処理装置。

(6)

前記推論部は、前記ユーザからの質問文及び前記ユーザからの質問文に対応するコンテキスト情報を結合して生成された文書ベクトルに基づいて前記対話ペアを選択する、前記(4)又は(5)に記載の情報処理装置。

(7)

前記取得部は、前記ユーザからの質問文に対応するコンテキスト情報をさらに取得し、

前記推論部は、取得された前記コンテキスト情報に対応する対話ペアを優先的に選択する、

前記(4)から(6)のうちのいずれかに記載の情報処理装置。

(8)

前記コンテキスト情報は、画像、バイタルサイン、ユーザプロフィールのうちの少なくとも1つを含む、前記(6)又は(7)に記載の情報処理装置。

(9)

半構造化文書から複数の用語を抽出し、抽出された前記複数の用語の間の関係を示す知識グラフを生成することと、

前記知識グラフに基づいて、複数の対話ペアを含む対話リポジトリを生成することと

を含み、

前記複数の対話ペアの各々は、質問文及び前記質問文に対する応答文の組であり、

前記質問文に含まれる第1の用語と、前記応答文に含まれる第2の用語と

は、前記知識グラフにおいて接続された関係にある2つの用語である、
情報処理方法。

(10)

ユーザからの質問文を取得することと、

半構造化文書から抽出された第1の用語を含む質問文と、前記半構造化文書から抽出された複数の用語の間の関係を示す知識グラフにおいて前記第1の用語に接続された第2の用語を含む応答文との組である対話ペアを、取得された前記ユーザからの質問文に応じて選択し、選択された前記対話ペアの応答文を前記ユーザからの質問文に対する応答文として出力することとを含む情報処理方法。

符号の説明

[0110]	1	情報処理システム
	10	応答生成装置
	11	センサ
	12	入力部
	13	通信部
	14	記憶部
	141	ユーザ情報テーブル
	142	モデル情報
	143	制御プログラム
	15	取得部
	151	検知部
	152	登録部
	153	受信部
	16	応答生成部
	161	信号処理部
	162	音声認識部
	163	応答部

- 1 6 4 送信部
- 1 7 出力部
- 2 0 対話リポジトリ生成装置
- 2 1 通信部
- 2 2 記憶部
- 2 2 1 ユーザ情報テーブル
- 2 2 2 モデル情報
- 2 2 3 対話リポジトリ
- 2 2 4 制御プログラム
- 2 3 取得部
- 2 3 1 受信部
- 2 4 対話リポジトリ生成部
- 2 4 1 知識オブジェクト生成部
- 2 4 2 対話ペア生成部
- 2 4 3 推論部
- 2 4 4 送信部
- 3 0 外部サーバ

請求の範囲

- [請求項1] 半構造化文書から複数の用語を抽出し、抽出された前記複数の用語の間の関係を示す知識グラフを生成する知識オブジェクト生成部と、
前記知識グラフに基づいて、複数の対話ペアを含む対話リポジトリを生成する対話ペア生成部と
を具備し、
前記複数の対話ペアの各々は、質問文及び前記質問文に対する応答文の組であり、
前記質問文に含まれる第1の用語と、前記応答文に含まれる第2の用語とは、前記知識グラフにおいて接続された関係にある2つの用語である、
情報処理装置。
- [請求項2] 前記知識オブジェクト生成部は、前記半構造化文書の他の知識情報から抽出された用語をさらに含む前記知識グラフを生成し、
前記対話ペア生成部は、前記半構造化文書の他の知識情報から抽出された用語を用いて前記対話ペアを補完する、
請求項1に記載の情報処理装置。
- [請求項3] 前記対話ペア生成部は、少なくとも1つのペルソナを用いて前記対話ペアを補完する、請求項1に記載の情報処理装置。
- [請求項4] ユーザからの質問文を取得する取得部と、
半構造化文書から抽出された第1の用語を含む質問文と、前記半構造化文書から抽出された複数の用語の間の関係を示す知識グラフにおいて前記第1の用語に接続された第2の用語を含む応答文との組である対話ペアを、取得された前記ユーザからの質問文に応じて選択し、選択された前記対話ペアの応答文を前記ユーザからの質問文に対する応答文として出力する推論部と
を具備する情報処理装置。
- [請求項5] 前記推論部は、前記対話ペアの質問文と、前記ユーザからの質問文

との間における文書ベクトルのベクトル距離に基づいて前記対話ペアを選択する、請求項4に記載の情報処理装置。

[請求項6] 前記推論部は、前記ユーザからの質問文及び前記ユーザからの質問文に対応するコンテキスト情報を結合して生成された文書ベクトルに基づいて前記対話ペアを選択する、請求項4に記載の情報処理装置。

[請求項7] 前記取得部は、前記ユーザからの質問文に対応するコンテキスト情報をさらに取得し、

前記推論部は、取得された前記コンテキスト情報に対応する対話ペアを優先的に選択する、

請求項4に記載の情報処理装置。

[請求項8] 前記コンテキスト情報は、画像、バイタルサイン、ユーザプロフィールのうちの少なくとも1つを含む、請求項7に記載の情報処理装置。

[請求項9] 半構造化文書から複数の用語を抽出し、抽出された前記複数の用語の間関係を示す知識グラフを生成することと、

前記知識グラフに基づいて、複数の対話ペアを含む対話リポジトリを生成することと

を含み、

前記複数の対話ペアの各々は、質問文及び前記質問文に対する応答文の組であり、

前記質問文に含まれる第1の用語と、前記応答文に含まれる第2の用語とは、前記知識グラフにおいて接続された関係にある2つの用語である、

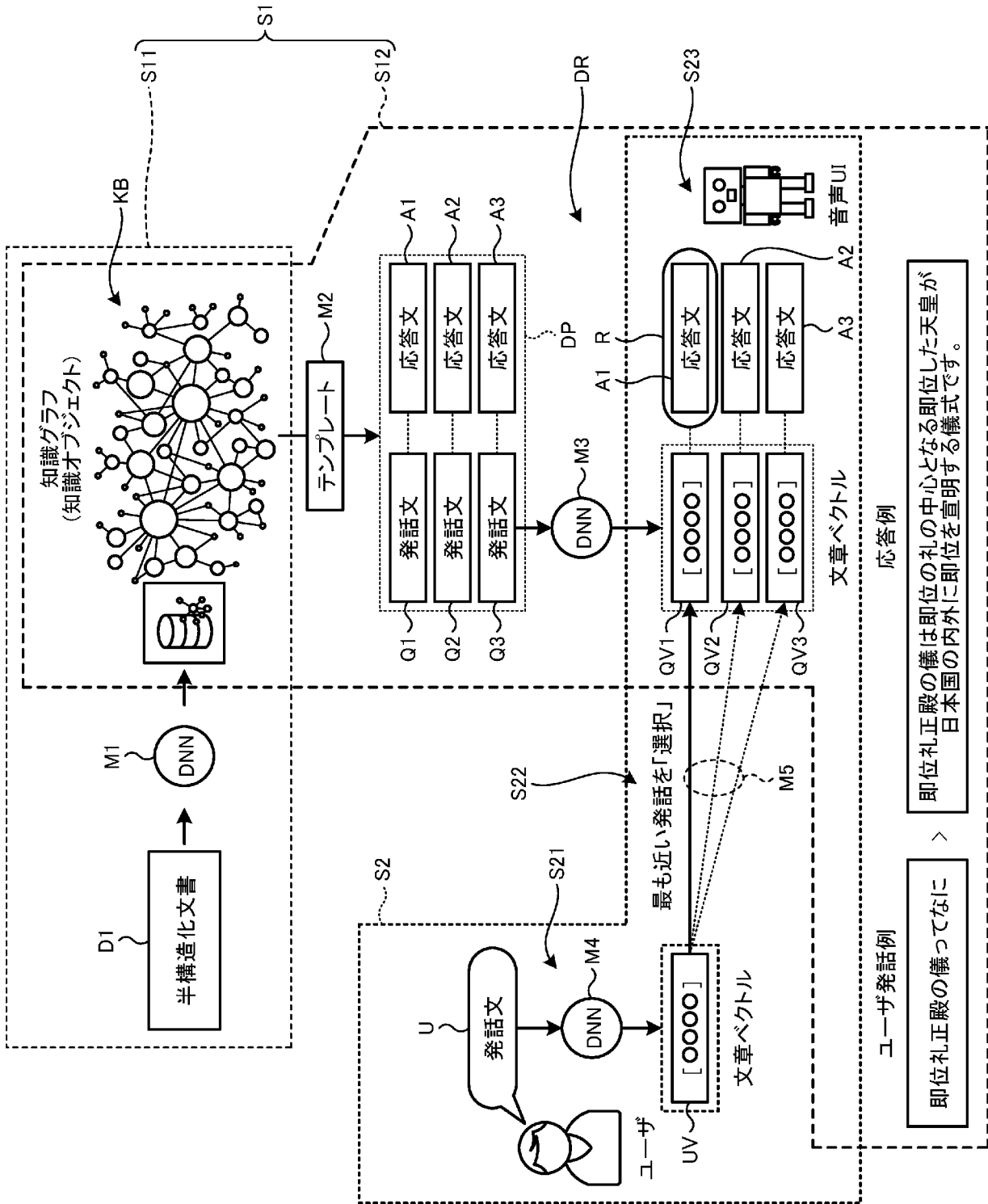
情報処理方法。

[請求項10] ユーザからの質問文を取得することと、

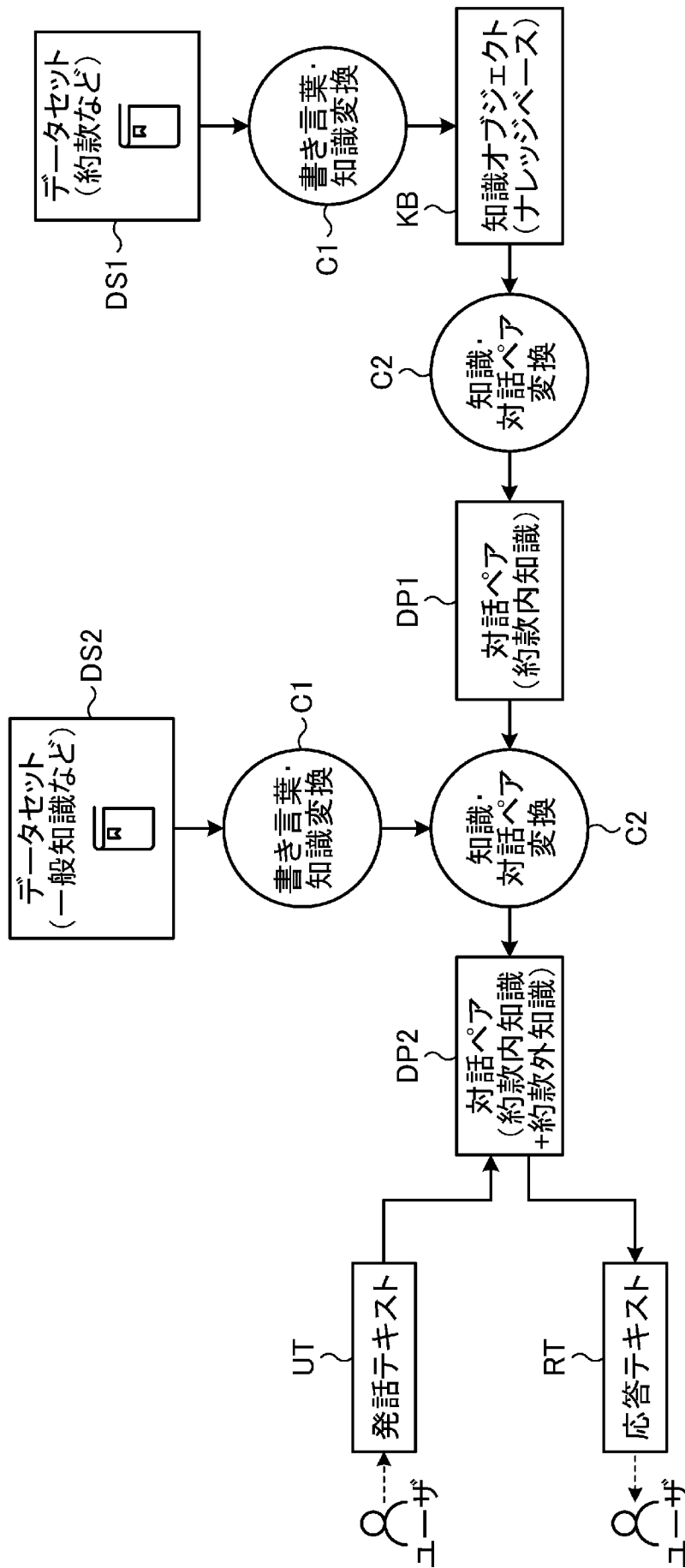
半構造化文書から抽出された第1の用語を含む質問文と、前記半構造化文書から抽出された複数の用語の間関係を示す知識グラフにおいて前記第1の用語に接続された第2の用語を含む応答文との組であ

る対話ペアを、取得された前記ユーザからの質問文に応じて選択し、
選択された前記対話ペアの応答文を前記ユーザからの質問文に対する
応答文として出力することと
を含む情報処理方法。

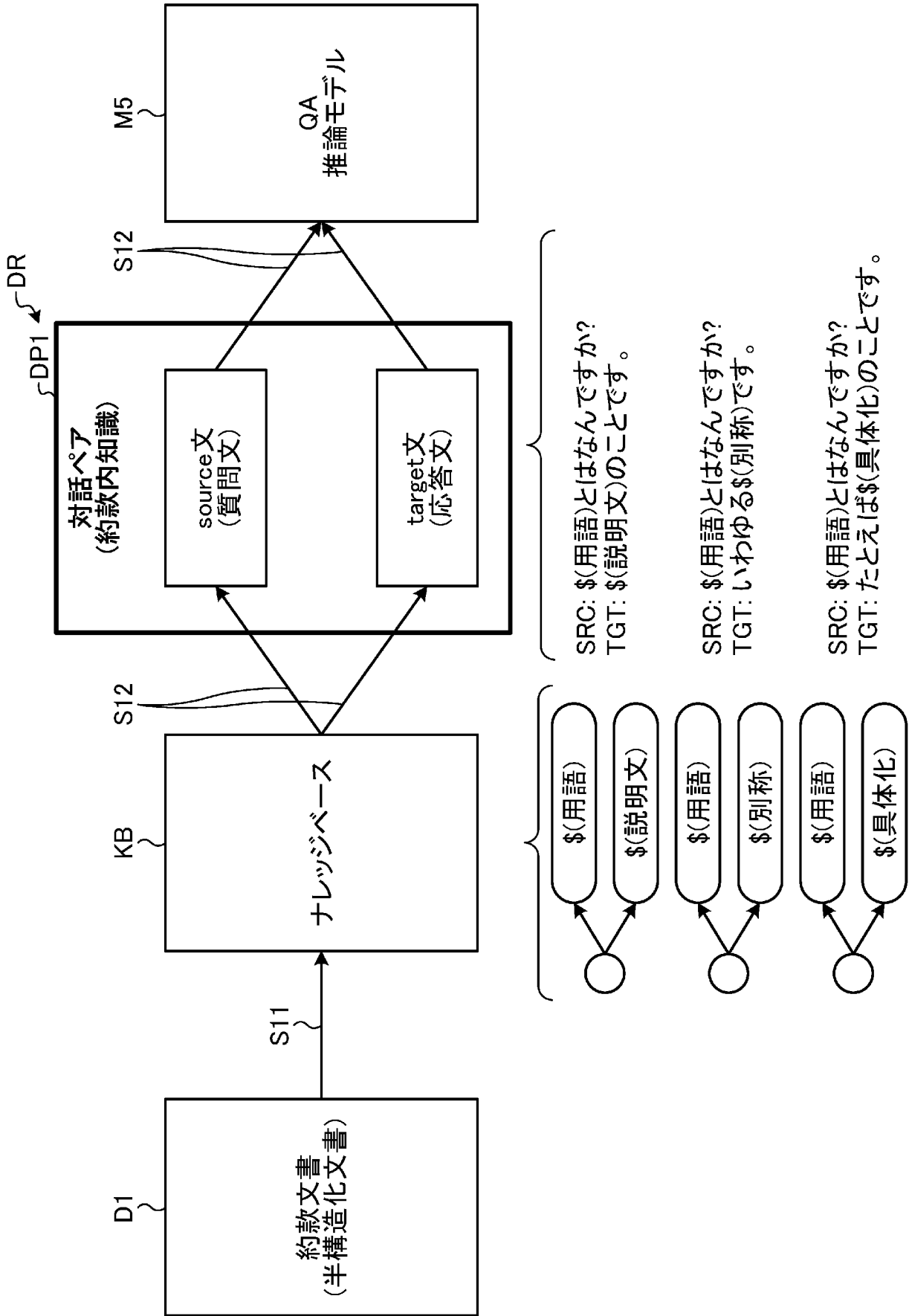
[図1]



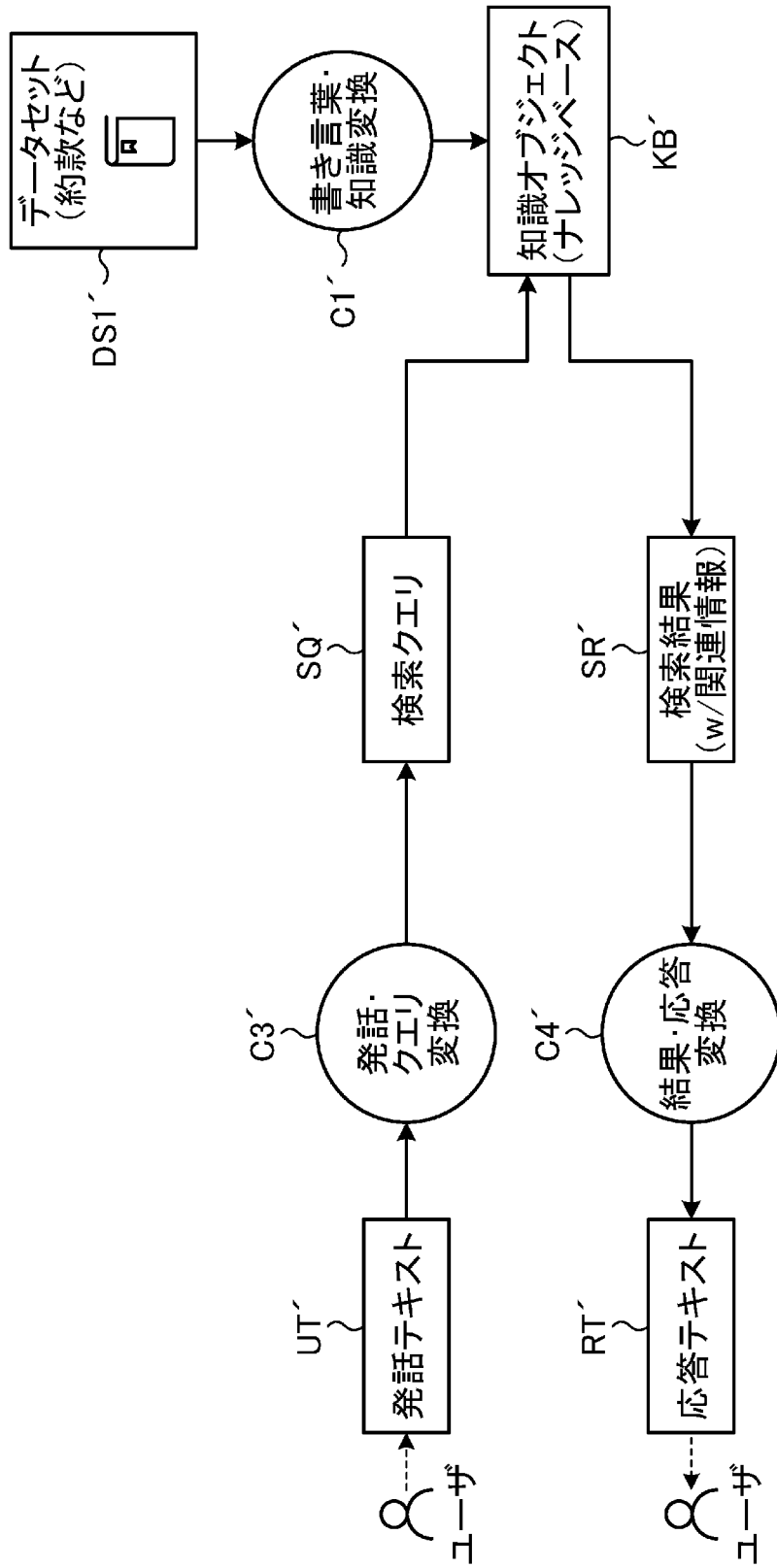
[図2]



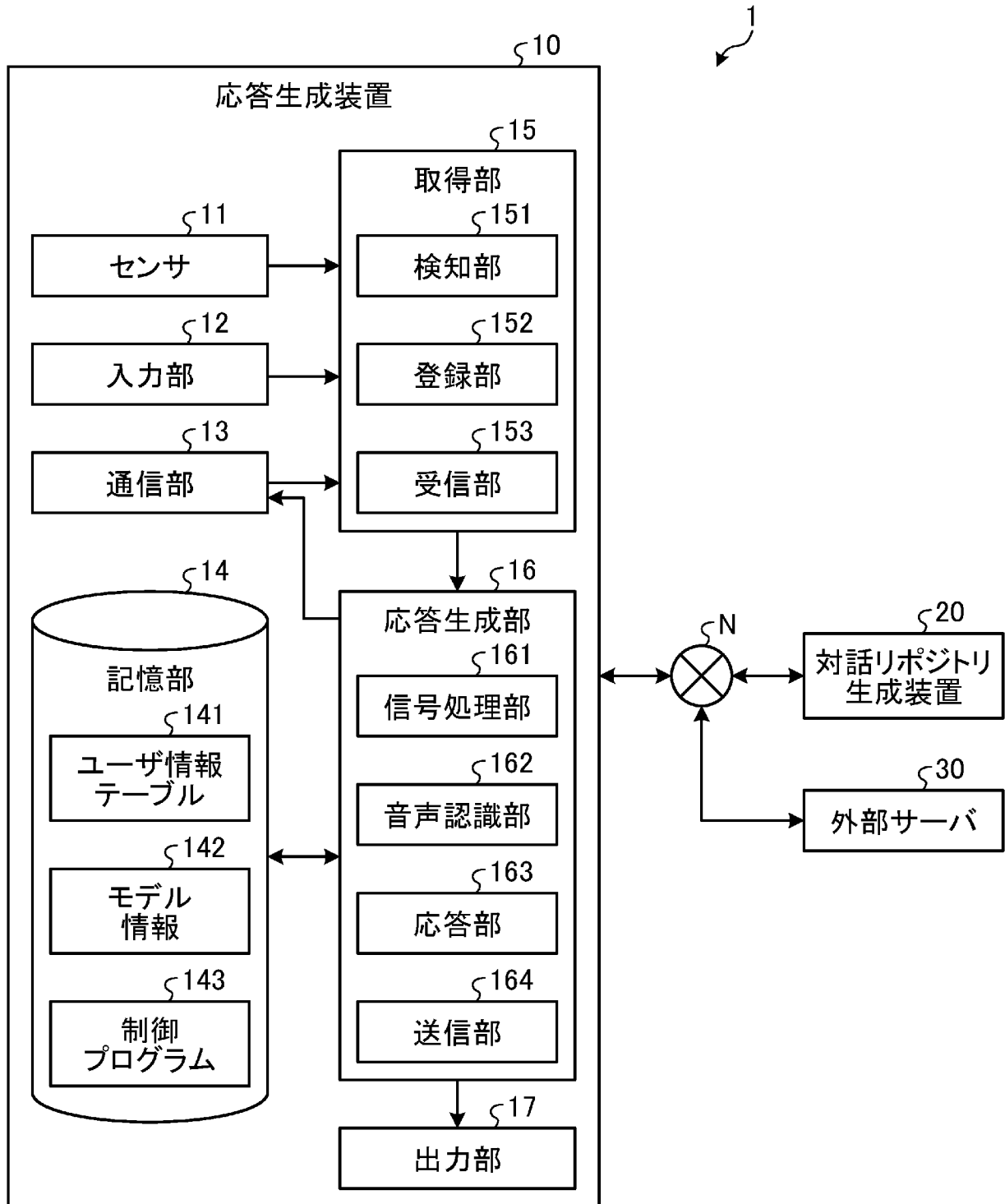
[図3]



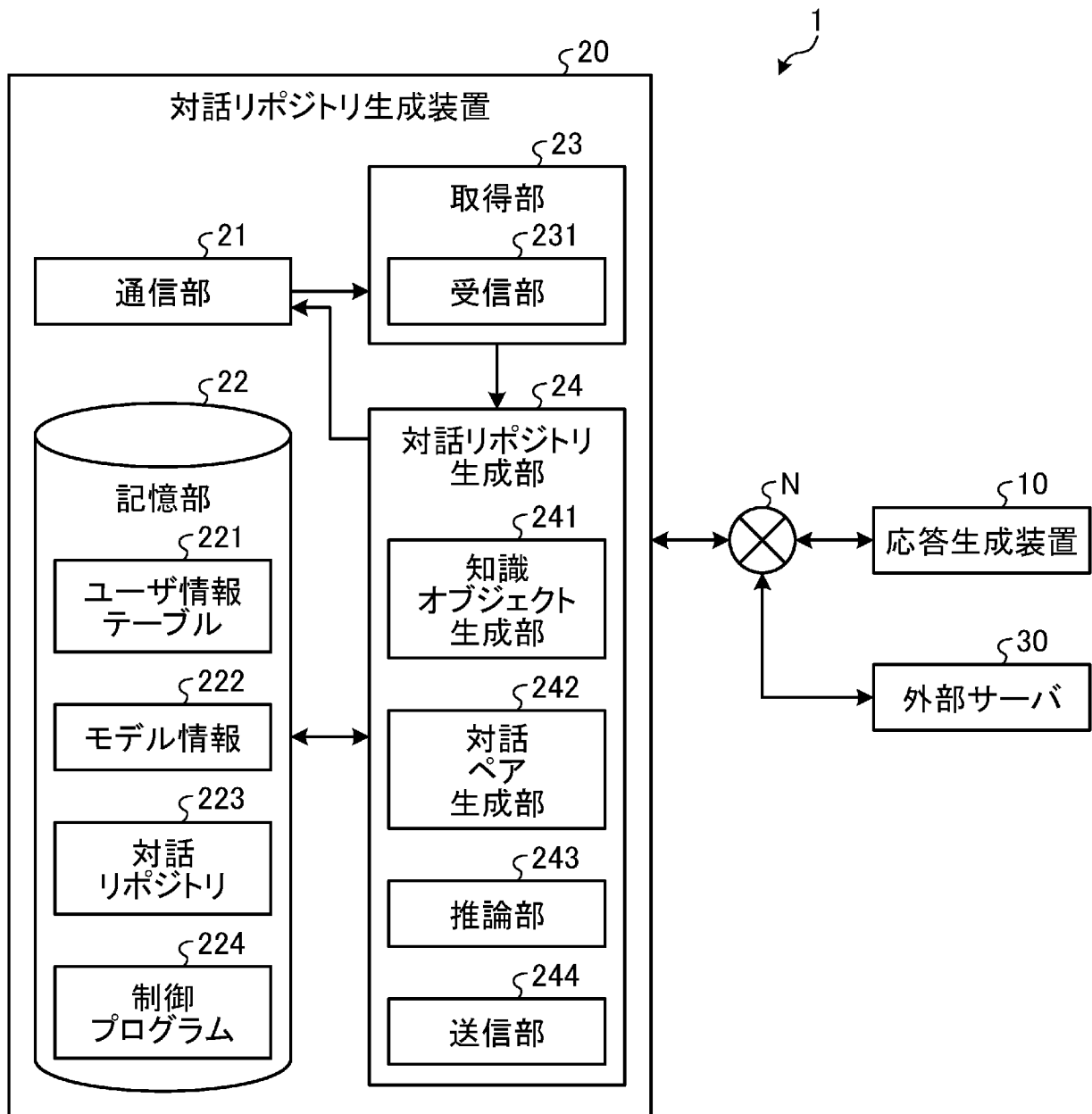
[図4]



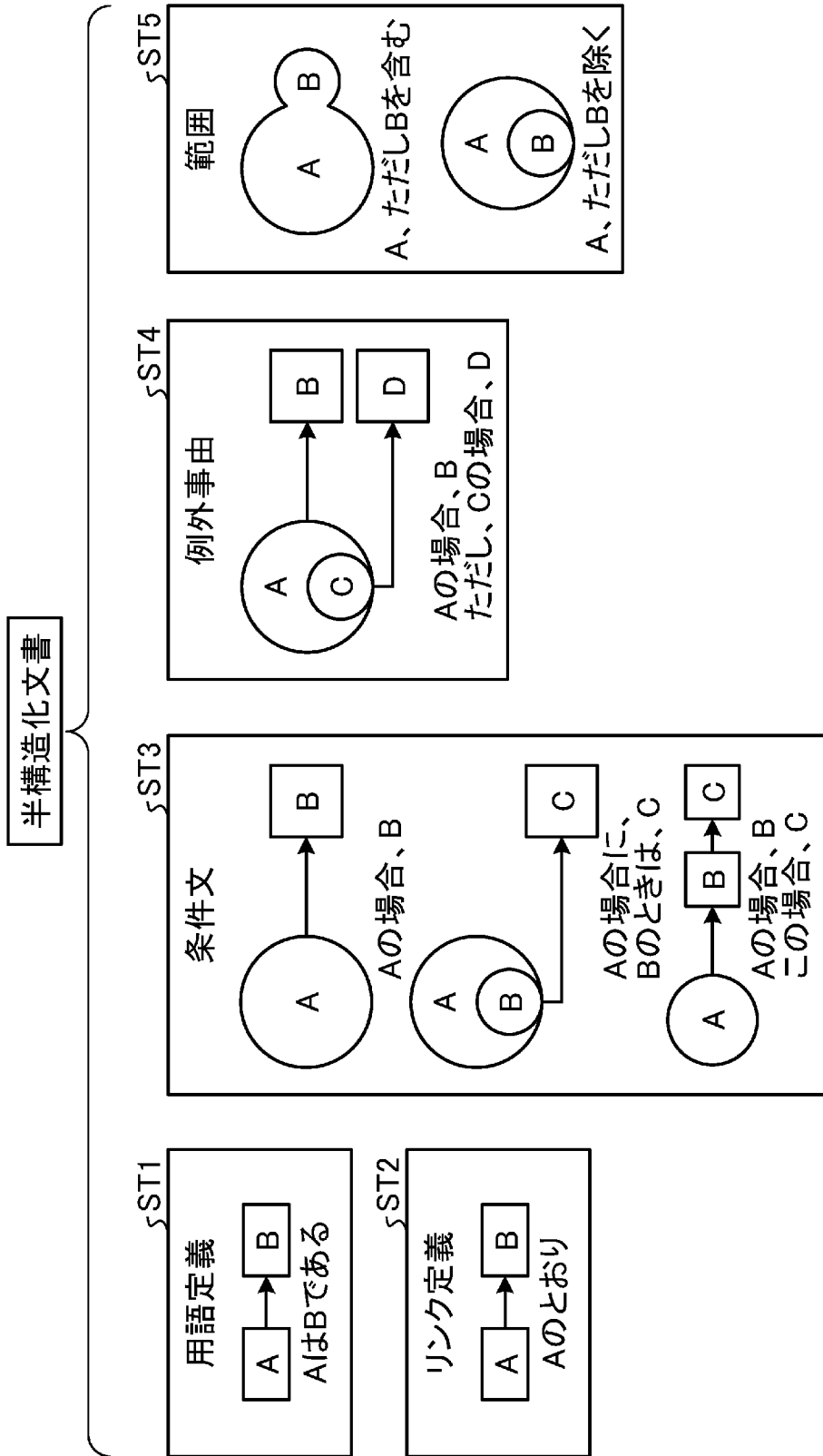
[図5]



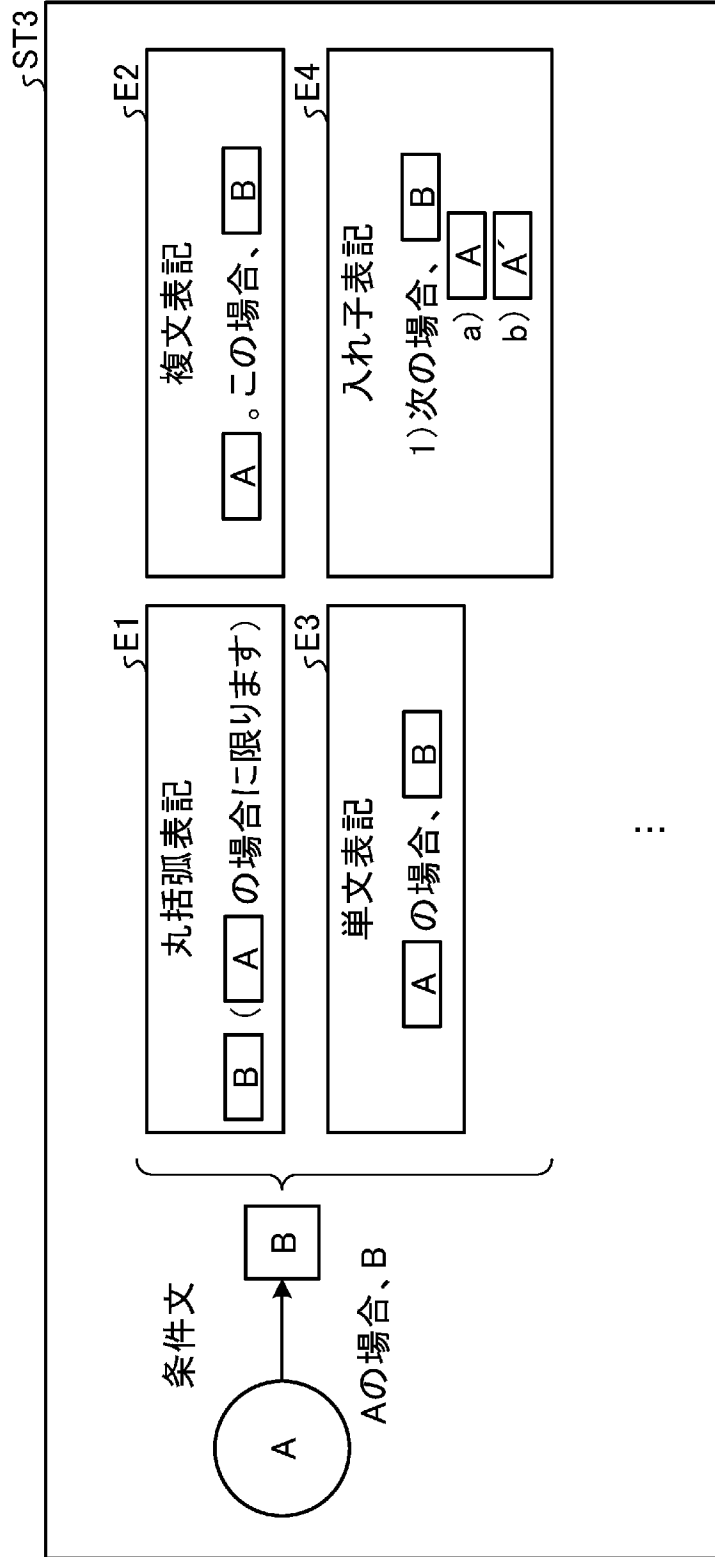
[図6]



[図7]



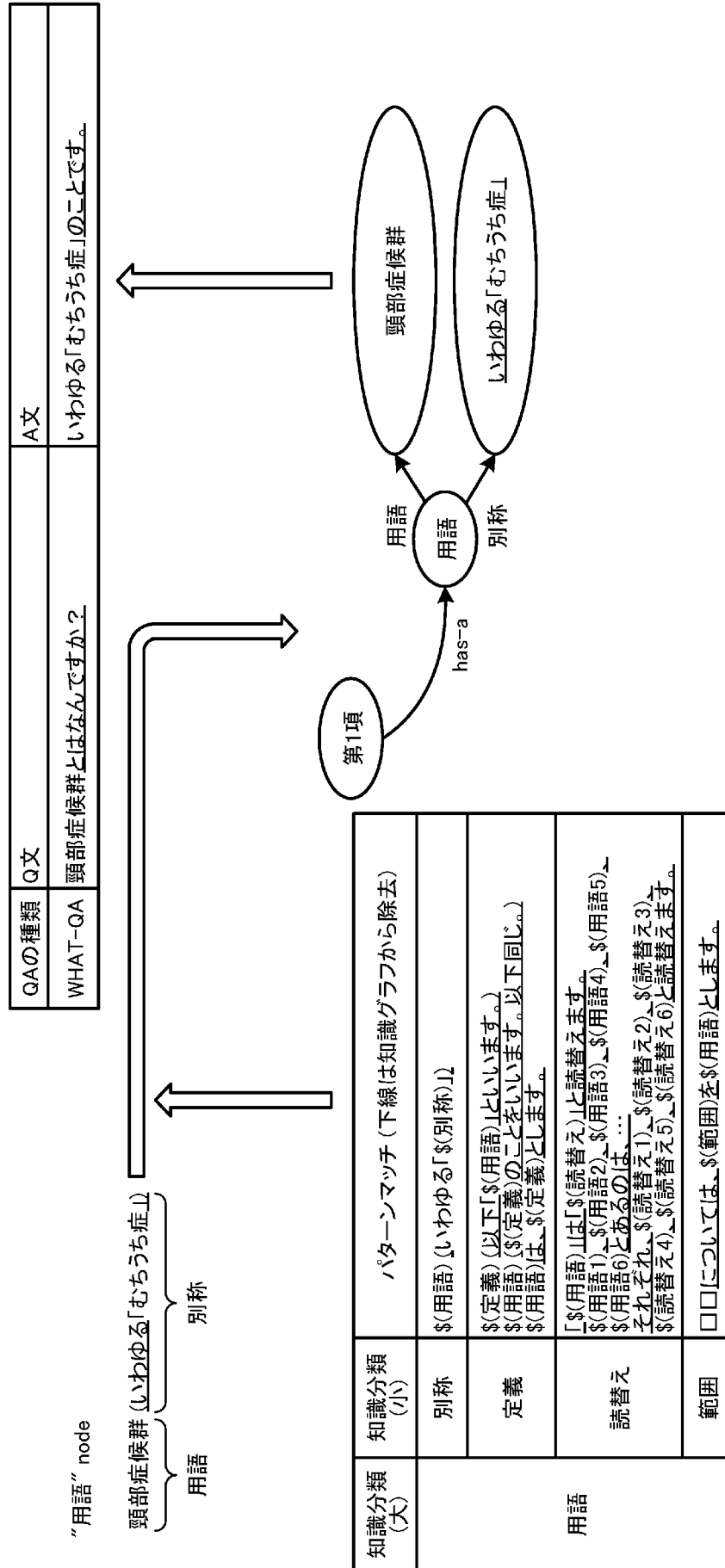
[図8]



【表9】

知識分類(大)	知識分類(小)	QAの種類(例)	本知識から生成される代表的Q文例(下線はテンプレート)	本知識から生成される代表的A文例(下線はテンプレート)
用語	別称、定義、読替え	WHAT-QA	頸部症候群とはなんですか？	いわゆる「むちうち症」のことです。
概要	約款概要、条概要、項概要	WHAT-QA	その第2条つてのはなんですか？	責任開始期についての取決めになります。
指示	払込指示、提出指示、変更指示、通知指示、指定指示、請求指示	HOW-QA	保険金の支払事由が生じたときは、どうしたらよいでしょうか？	もしお客様が保険契約のまたはは保険金受取人であれば、遅滞なく当社に通知してください。
支払義務	支払あり、支払なし、支払可能性あり	IS-QA	死亡保険金を支払う前に高度障害保険金の支払い要求を受け、高度障害保険金が支払われるときは、死亡保険金は支払われるのでしょうか？	その場合、会社は死亡保険金を支払いません。
権利	請求権、選択権、変更権、解除権、解約権、取消権、前納権、返還権、復活権、貸付件、復旧権、照会権、承諾許可、返済権	CAN-QA	死亡保険金の受取人は、会社の定める取扱いに従い、会社所定の金額を上限として、死亡保険金の一部または全部を請求できますか？	はい、死亡保険金の受取人は、会社の定める約款に従い、会社所定の金額を上限として、死亡保険金の一部または全部を請求することができます。
禁止	復活禁止、解除禁止、変更禁止、対抗禁止、使用禁止、公開禁止	CAN-QA	保険契約者が解約返戻金を請求したときは、保険契約を復活できますか？	残念ながら、保険契約者が解約返戻金を請求した場合には、保険契約を復活することはできません。
行為	確認、交付、割引、継続、通知、取扱い、処理、払戻、登録、設定、返済、判断、付加、積増し	HOW-QA	〇〇のとき、どうなりますか？	：
参照	次のとおり		次の規定について、くわしく教えてください？	：
範囲	含む、限る	IS-QA	この場合、〇〇は含まれますか？	：
期日	効力失効日、責任発生時、猶予期間、登録期間、更改日、基準日、効力発生日	WHEN-QA	〇〇が開始されるのは、いつからですか？	：
人間	受取人、契約者、代理人	WHO-QA	〇〇できるのは誰になりますか？	：
計算式	金匱定義、計算方法、最低金額、差引、繰入、充当、割合、切捨て、減額	HOW MUCH-QA	〇〇のとき、保険金はいくらですか？	：
条件	金額、払込、告知	WHAT-QA	その場合、条件などありますか？	：
適用	適用、準用、適用外			：
その他	責任の所在、管轄裁判所、効力無効、消滅、配当なし、払戻なし			：

[図10]

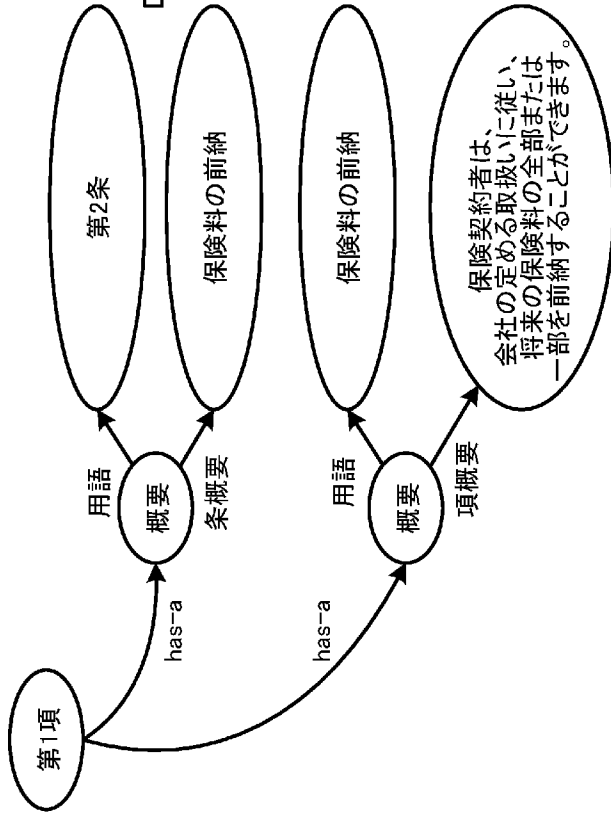


[図11]

QAの種類	Q文	A文
WHAT-QA	その第2条ってのはなんですか？	責任開始期についての取決めになります。

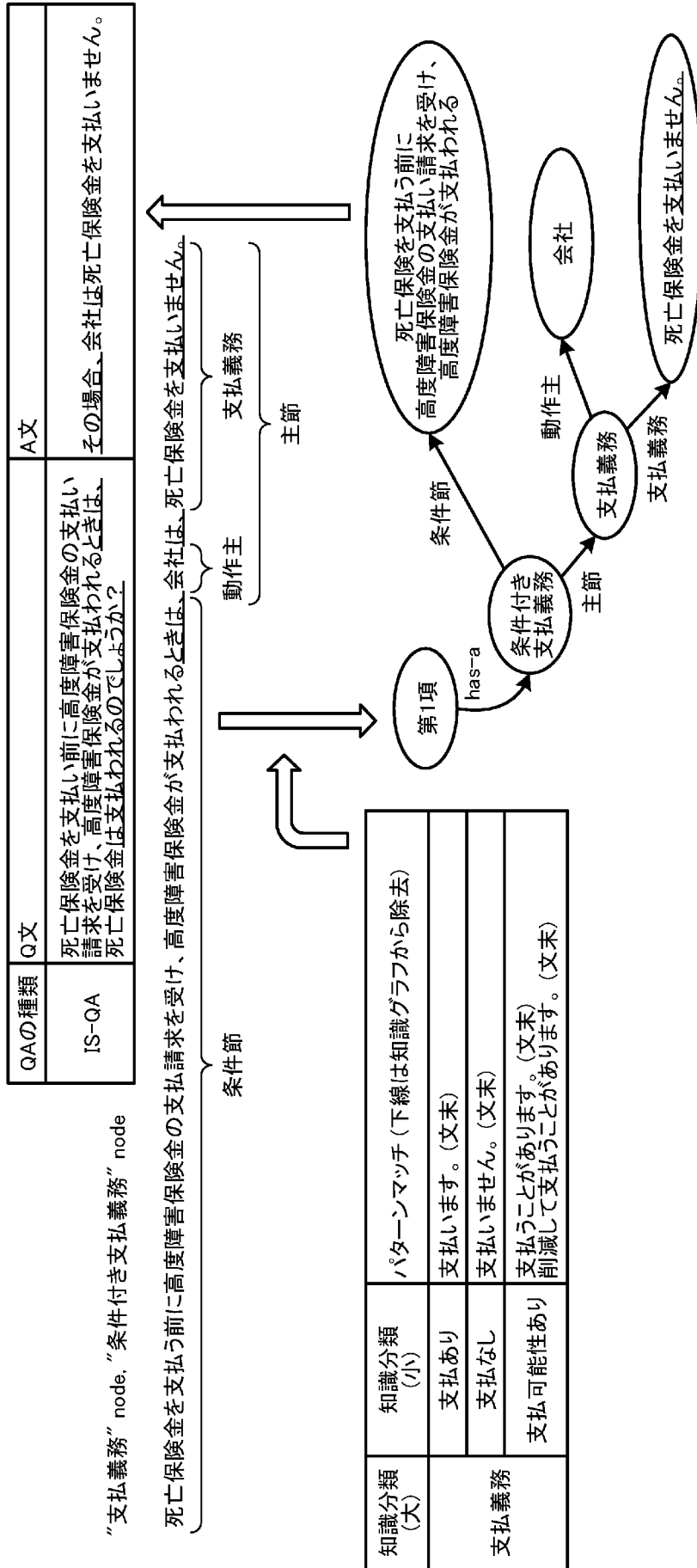
“概要” node

条番号・項番号	原文 (下線は知識グラフから除去)
第1条	保険料の前納
第1条第1項	保険契約者は、会社の定める取扱いに従い、将来の保険料の全部または一部を前納することができます。 この場合には、会社所定の利率で割引きます。

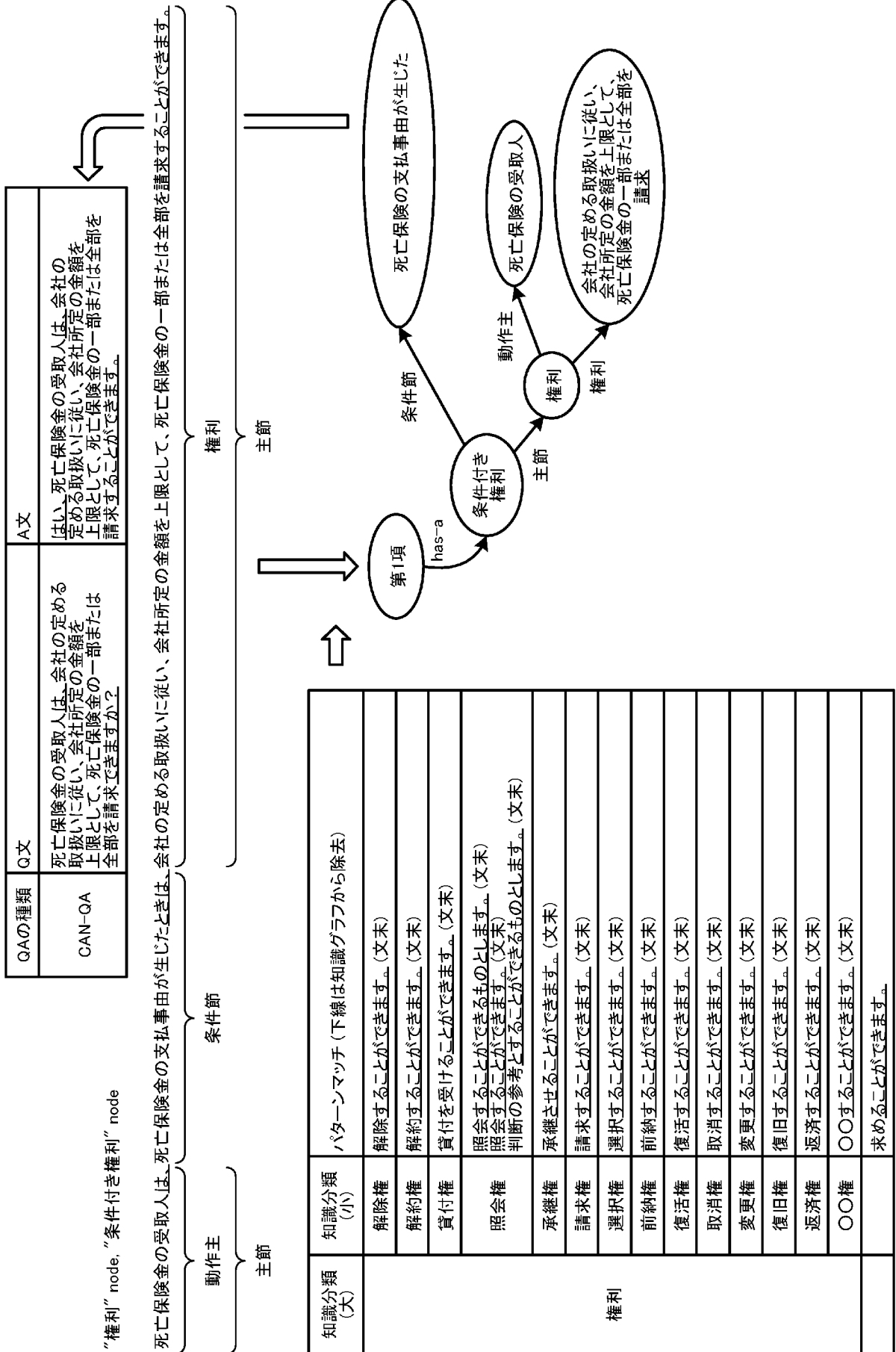


知識分類 (大)	知識分類 (小)	抽出対象
概要	約款概要	\$(用語) = 約款タイトル(約款, (は除く)) \$(概要) = この保険の趣旨
	条概要	条の原文を対象 \$(用語) = 条番号, \$(概要) = 条の原文
	項概要	条の第1項の原文を対象 \$(用語) = 条の原文, \$(概要) = 最初の項の原文

[図13]



[図14]



[図15]

QAの種類	Q文	A文
CAN-QA	保険契約者が解約返戻金を請求したときは、保険契約を復活できますか？	残念ながら、保険契約者が解約返戻金を請求した場合には、保険契約を復活することはできません。

“禁止” node, “条件付き禁止” node

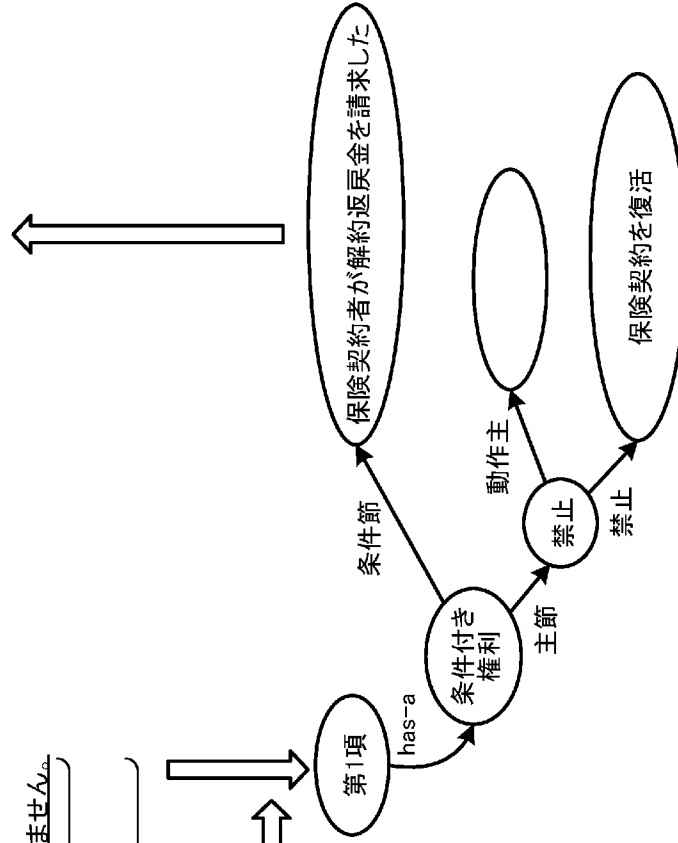
ただし、保険契約者が解約返戻金を請求した場合には、保険契約金を復活することはできません。

条件節

禁止

主節

知識分類 (大)	知識分類 (小)	パターンマッチ (下線は知識グラフから除去)
禁止	復活禁止	復活することはできません。(文末)
	解除禁止	解除をすることができません。(文末)
	変更禁止	変更することができません。(文末)
	対抗禁止	対抗することができません。(文末)
	使用禁止	用いないものとします。(用いない→使用しない)(文末)
	公開禁止	公開しないものとします。(文末)
〇〇禁止	〇〇することはできません。(文末)	
	〇〇しないものとします。(文末)	
条件節		しなければ、

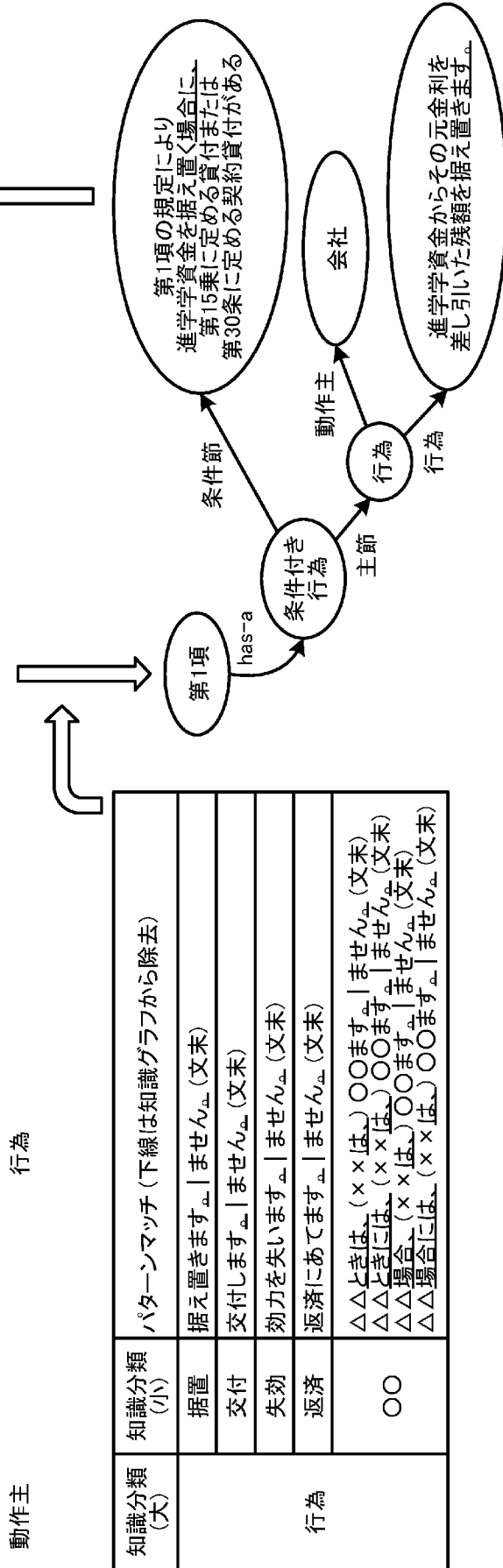


[図16]

QAの種類	Q文	A文
IS-QA	学資金を受け取る場合には、貸付金は差し引かれるのでしょうか？	はい、会社は、進学学資金からその元金利を差し引いた残額を据え置きます。

“行為” node, “条件付き行為” node

第1項の規定により進学学資金を据え置く場合に、第15条に定める貸付または第30条に定める契約者貸付があるときは、会社は、進学学資金からその元金利を差し引いた残額を据え置きます。



[図17]

イレギュラーなケース		追加で生成されるQ文
言い回し バリエーションの追加.	約款に載っているのは「入院給付金日額」だが、FAQでは「保障額」と言っている。	保障額を減額するには、どのように手続きをすればよいですか？
	約款に載っているのは「本条の請求をするとき」だが、FAQでは「解約するとき」と言っている。	解約するにはどうすればよいですか？
ネガティブな ケースの追加.	約款に載っているのは「支払われるケース」だが、質問では「支払われないケース」を聞いている。	入院や手術などの給付金の請求をしても、支払われないことはありますか？
	約款に載っているのは給付金が支払われる「手術の種類」であって、「日帰りかどうか」は記載がない。	日帰りで手術を受けました。給付金は支払われますか？
	約款に載っているのは「新しい保険証券が送られる」ケースだが、質問では「送られてこないケース」を聞いている。	契約内容の更新・変更をした場合、新しい保険証券が送られてきますか？

[図18]

ユーザー発話	約款知識から自動生成した応答文 (従来)	ペルソナ (契約例)を用いて自動生成した応答文
<p>健康還付金はいくら支払われますか？</p>	<p>被保険者が健康還付金給付支払日に生存しているときに、 既払込保険料相当額から、 この保険契約の締結の際の責任開始期から、 健康還付金給付支払日の前日までの入院、手術、 放射線治療、骨髄幹細胞採取手術および 骨髄移植術に対して支払われる入院給付金等の 金額の合計を差し引いた残額が、 保険契約者に支払われます。</p>	<p>既払込保険料から、入院給付金等の合計額を差し引いた残額が、健康還付金として支払われます。</p> <p>たとえば、お客様が、日々4610円の保険料を5年間支払っていた場合、支払った保険料の総額は4610円×12ヶ月×5年で、28万円になります。この28万円が既払込保険料です。</p> <p>また、この5年間でいちど病気になるって10日間入院していた場合、お客様に支払われる入院給付金は1日5000円×10日で、5万円になります。この5万円が、入院給付金等の合計額です。</p> <p>この28万円の既払込保険料から、5万円の入院給付金等の合計額を差し引いた残額、つまり23万円が、健康還付金として支払われます。</p>
<p>だれに支払われますか？</p>	<p>保険契約者に支払われます。</p>	<p>保険契約者に支払われます。</p> <p>たとえば、お客様ご本人が保険契約者の場合、お客様に支払われます。</p>
<p>入院給付金等には他にどんなものがありますか？</p>	<p>入院、手術、放射線治療、骨髄幹細胞採取手術および骨髄移植術に対して支払われる入院給付金、が含まれます。</p>	<p>入院、手術、放射線治療、骨髄幹細胞採取手術および骨髄移植術に対して支払われる入院給付金、が含まれます。</p>
<p>なにか条件はありますか？</p>	<p>被保険者が健康還付金支払日に生存しているときに、になります。</p>	<p>被保険者が健康還付金支払日に生存しているときに、になります。</p> <p>たとえば、お客様が健康還付金支払日の前日まで死亡したとき、支払われません。</p>

[図19]

ペルソナ

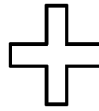
項目事例	パラメタ
被保険者	35歳
性別	男性
1入院の支払限度	120日型
入院給付金日額	5,000円
保険期間	終身
保険料払込期間	終身
個別扱月払保険料	4,610円 (男性)

予め生成される文章

たとえば、お客様が、月々4,610円の保険料を5年間支払っていた場合、支払った保険料の総額は4610円×12ヶ月×5年で、28万円になります。この28万円が、既払込保険料です。



既払込保険料



約款の演算式

号	用語	用語の定義
(1)	既払込保険料相当額	次の算式により計算した金額をいいます。 月払保険料相当額 × 12 × 契約日から健康還付給付金支払日の前日までの年数

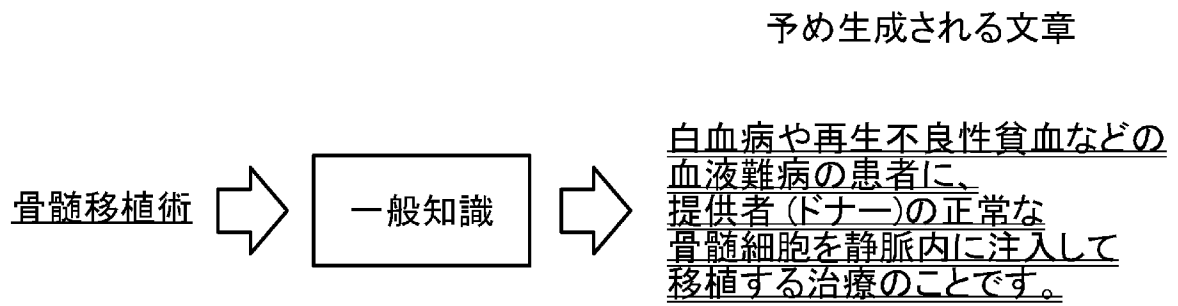
(5)	月払込保険料相当額	保険料の払込方法(回数)を月払とし、個別扱の保険料率を適用した場合のこの保険契約の1か月分の保険料をいい、この保険契約に付加された特約があるときは、その特約の保険料は含みません。

総合医療保険(無解約返戻金型)普通保険約款 第54条

[20]

ユーザー発話	約款知識から自動生成した応答文 (従来)	一般知識を用いて自動生成した応答文
<p>骨髓移植給付金とはなんですか？</p>	<p>公的医療保険制度の輸血料の算定対象となる骨髓移植を受けたとき、入院給付金日額×20日分の金額が支払われるものです。</p>	<p>公的医療保険制度の輸血料の算定対象となる骨髓移植を受けたとき、入院給付金日額×20日分の金額が支払われるものです。</p>
<p>公的医療保険制度とはなんですか？</p>	<p>(約款にかかれていない医療用語なので答えられない)</p>	<p><u>公的医療保険は、病気やケガをしたときに医療費の一部が負担軽減される重要な制度です。</u></p>
<p>骨髓移植とはなんですか？</p>	<p>(約款にかかれていない医療用語なので答えられない)</p>	<p><u>白血病や再生不良性貧血などの血液難病の患者に、提供者(ドナー)の正常な骨髓細胞を静脈内に注入して移植することです。</u></p>
<p>ドナーはすぐみつかれるものですか？</p>	<p>(約款にかかれていない医療用語なので答えられない)</p>	<p><u>骨髓移植を行う場合、ドナーと移植対象患者との間で、白血球の血液型であるHLAが適合しなければなりません。HLA型が適合して移植が可能になる確率は、同父母の兄弟姉妹間で25%、非血縁者間では数百分の1といわれています。</u></p>
<p>後遺症はあるのですか？</p>	<p>(約款にかかれていない医療用語なので答えられない)</p>	<p><u>移植の前処置で致死量を超える大量の抗がん剤投与と放射線照射を行う為、後遺症として、小児の場合は成長障害、また男女とも極めて高い確率で永久的な不妊となることが知られています。</u></p>

[図21]



[図22]

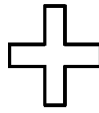
一つ前の約款 (契約時のもの)

号	用語	用語の定義
(1)	既払込保険料相当額	次の算式により計算した金額をいいます。 月払保険料相当額 × 12 × 契約日から健康還付給付金支払日の前日までの年数

(5)	月払込保険料相当額	保険料の払込方法 (回数) を月払とし、個別扱の保険料率を適用した場合の保険契約の1か月分の保険料をいい、この保険契約に付加された特約があるときは、その特約の保険料は含みません。



既払込保険料



最新の約款



号	用語	用語の定義
(1)	既払込保険料相当額	次の算式により計算した金額をいいます。 月払保険料相当額 × 11 × 契約日から健康還付給付金支払日の前日までの年数

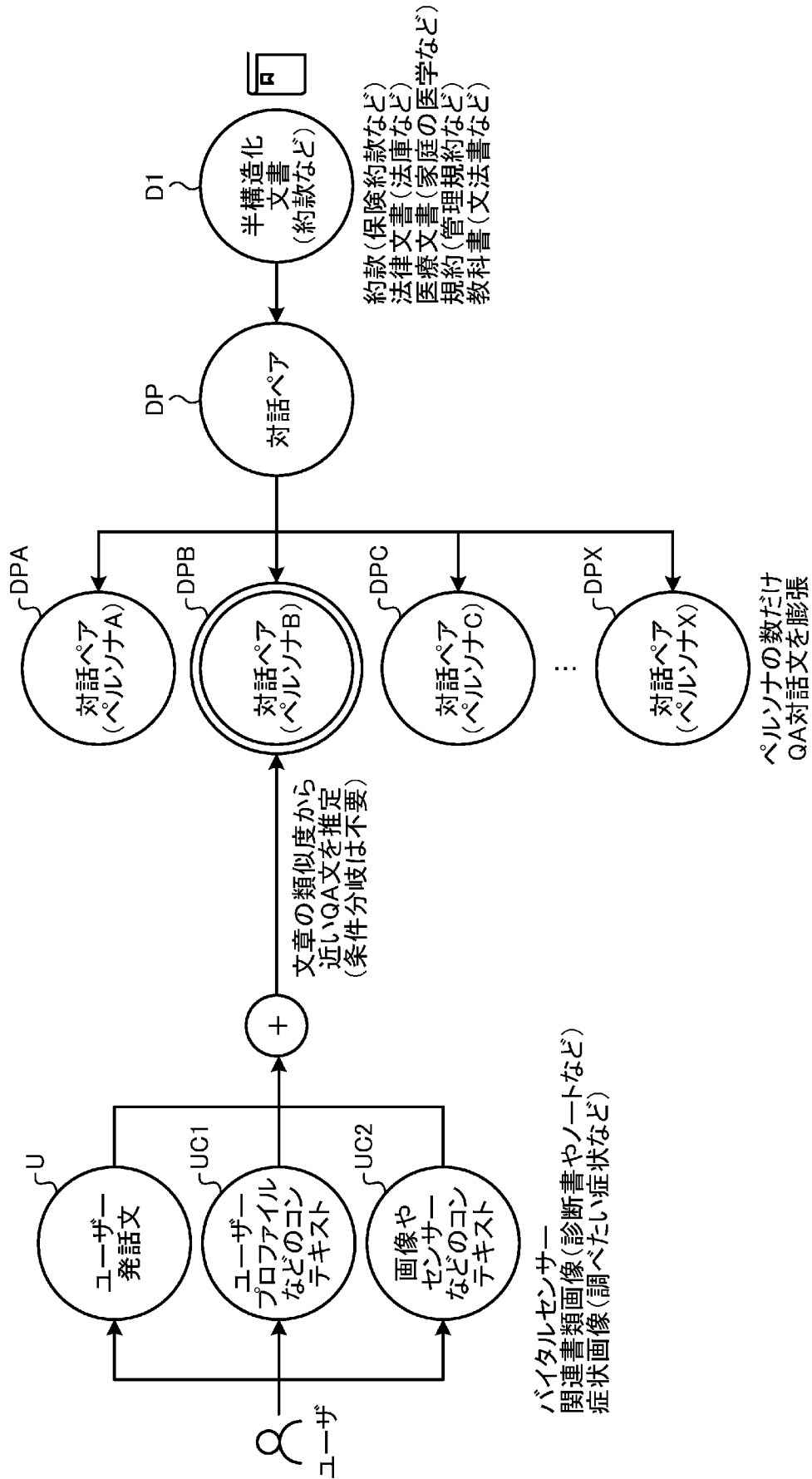
(5)	月払込保険料相当額	保険料の払込方法 (回数) を月払とし、個別扱の保険料率を適用した場合の保険契約の1か月分の保険料をいい、この保険契約に付加された特約があるときは、その特約の保険料は含みません。

予め生成される文章

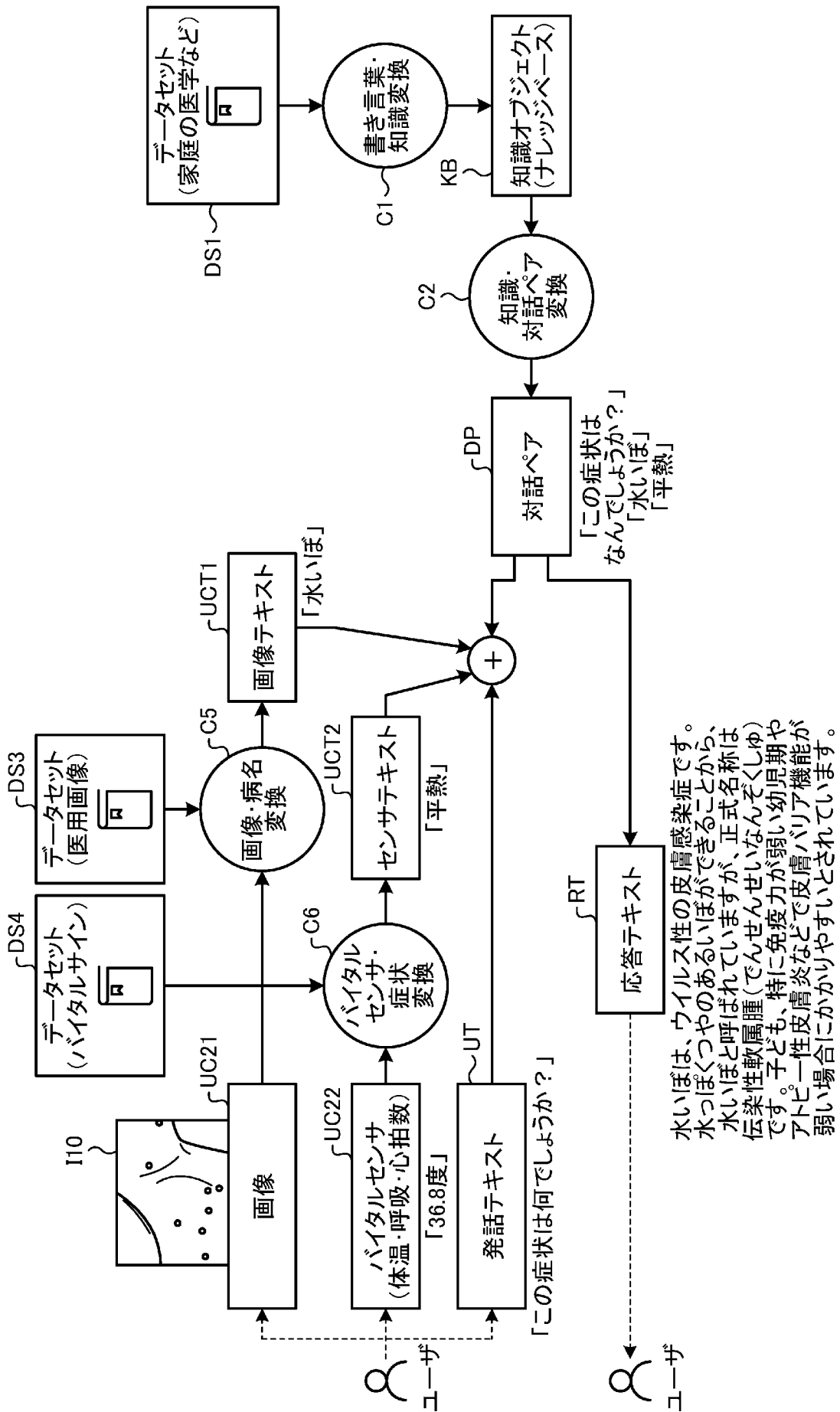
たとえば、お客様が、月々4,610円の保険料を5年間支払っていた場合、支払った保険料の総額は4610円 × 12ヶ月 × 5年で、28万円になります。この28万円が、既払込保険料です。

なお、2019年4月に更新された最新の約款では、12ヶ月から11ヶ月に変更になりますので、この場合、支払った保険料の総額は4610円 × 11ヶ月 × 5年で、約25万円になります。

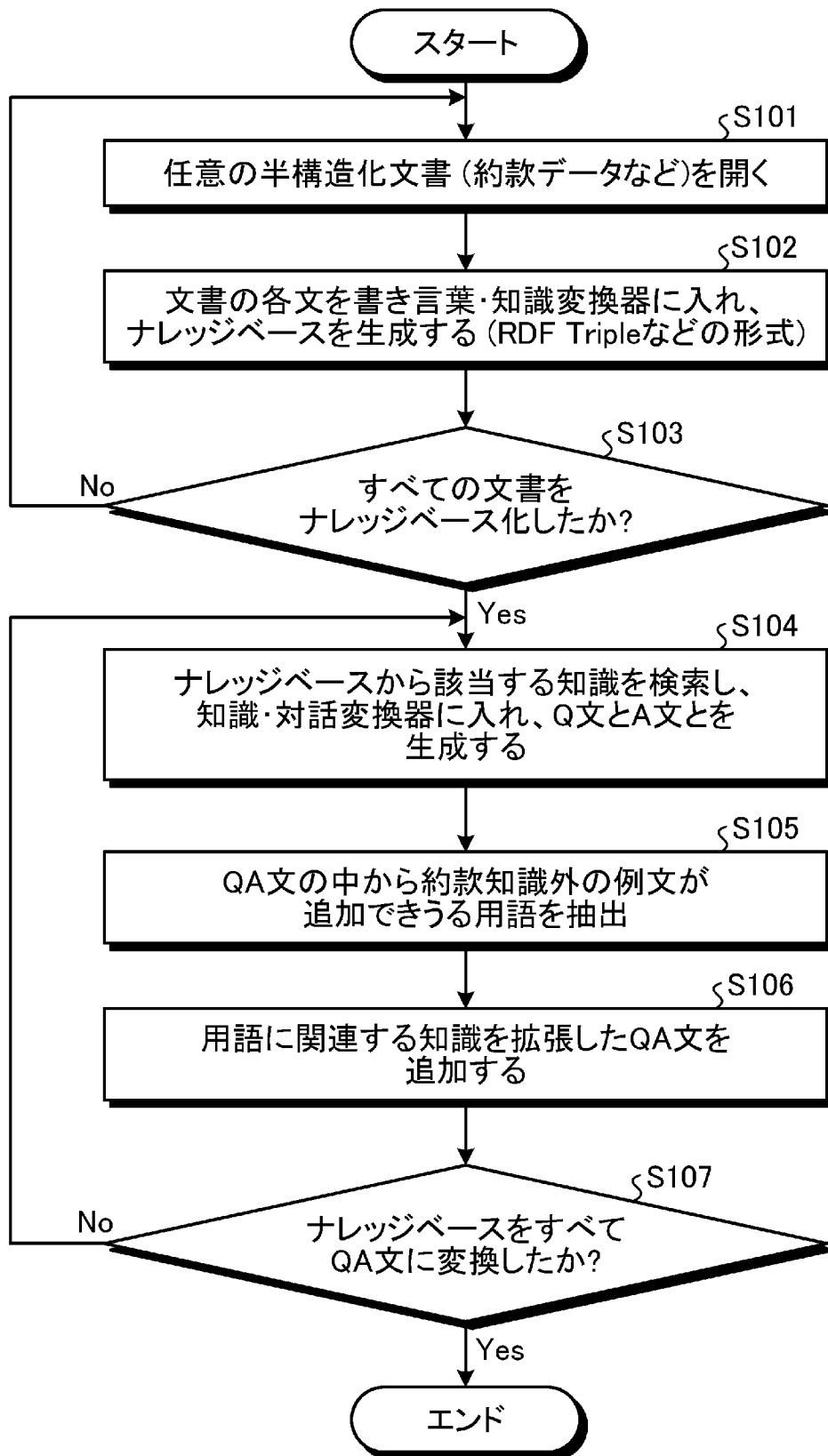
[図23]



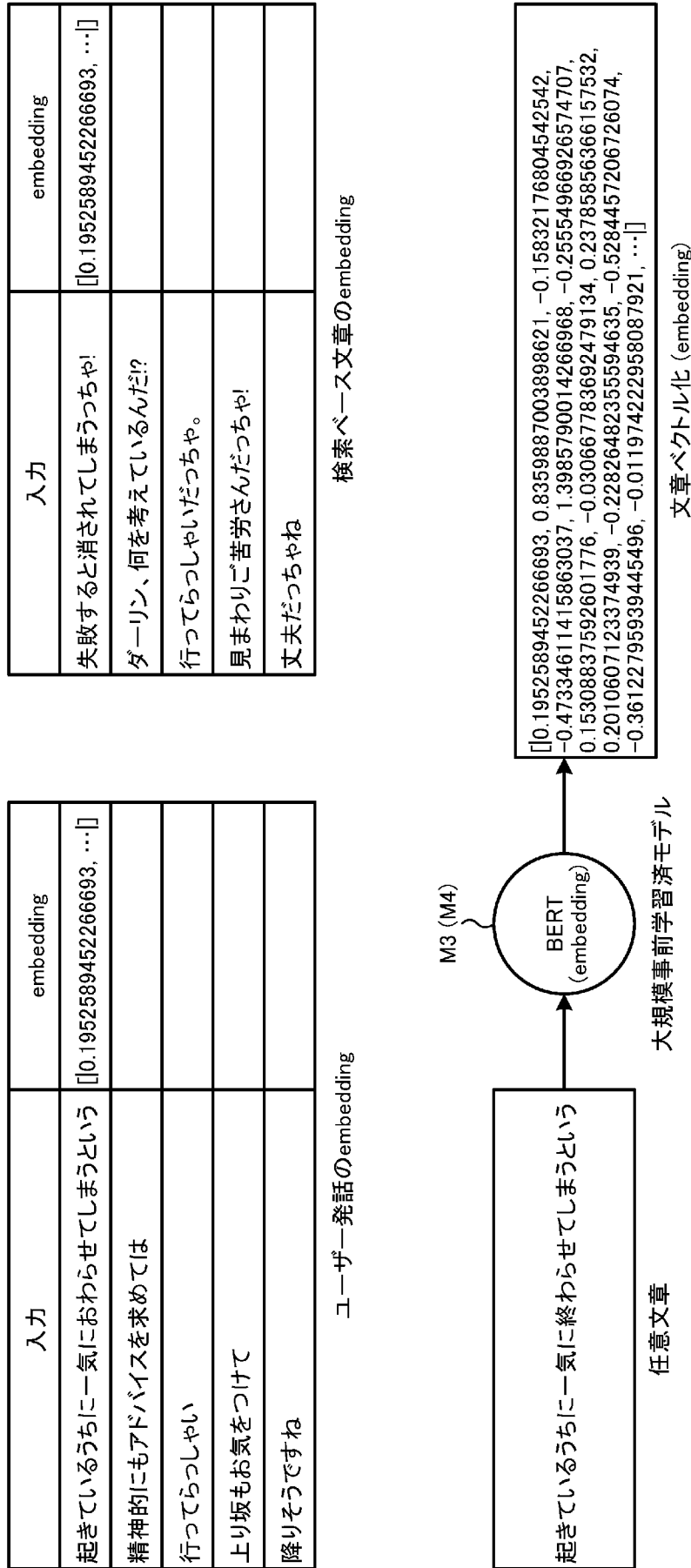
[図24]



[図25]



[図26]

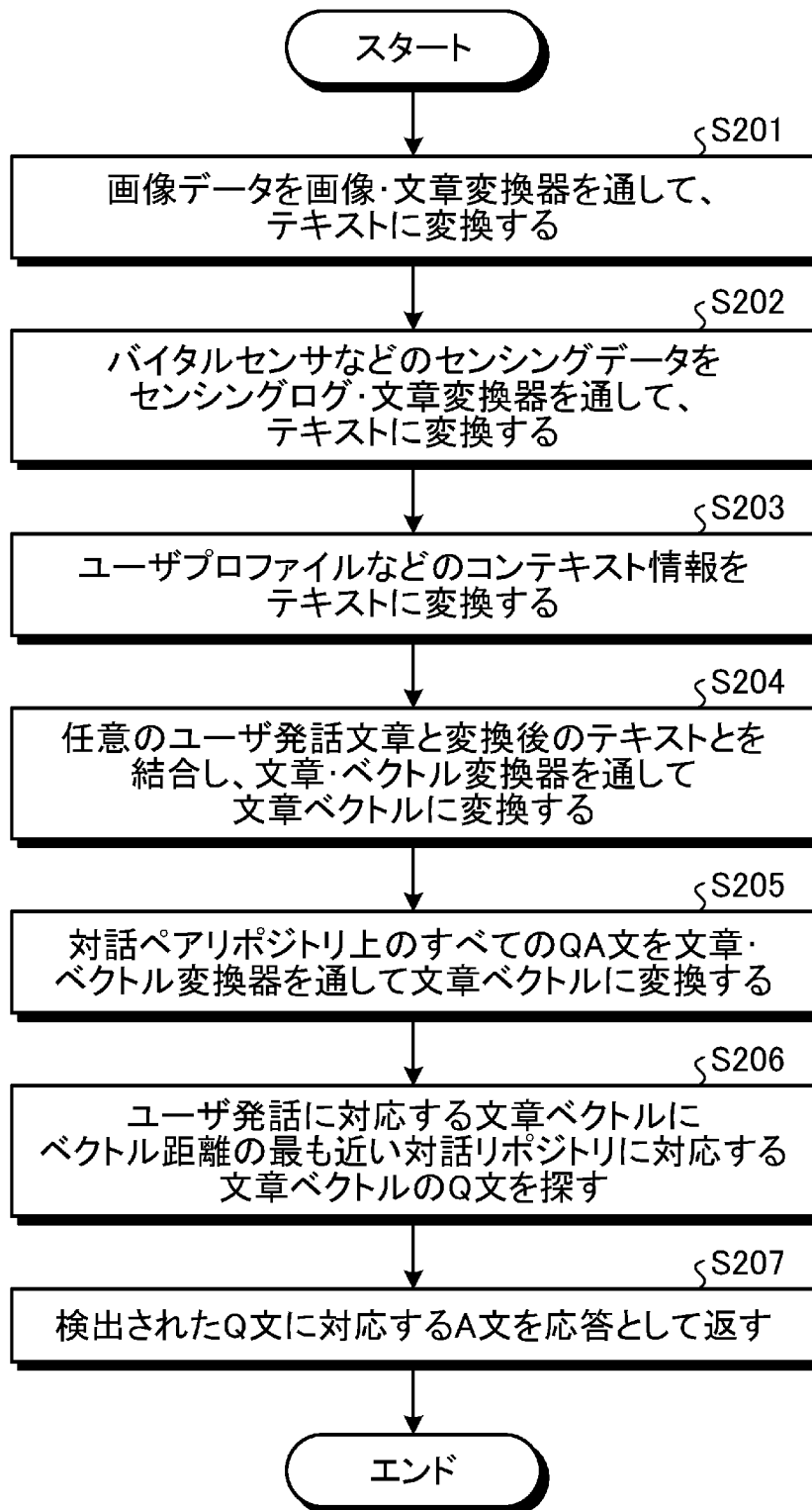


[図27]

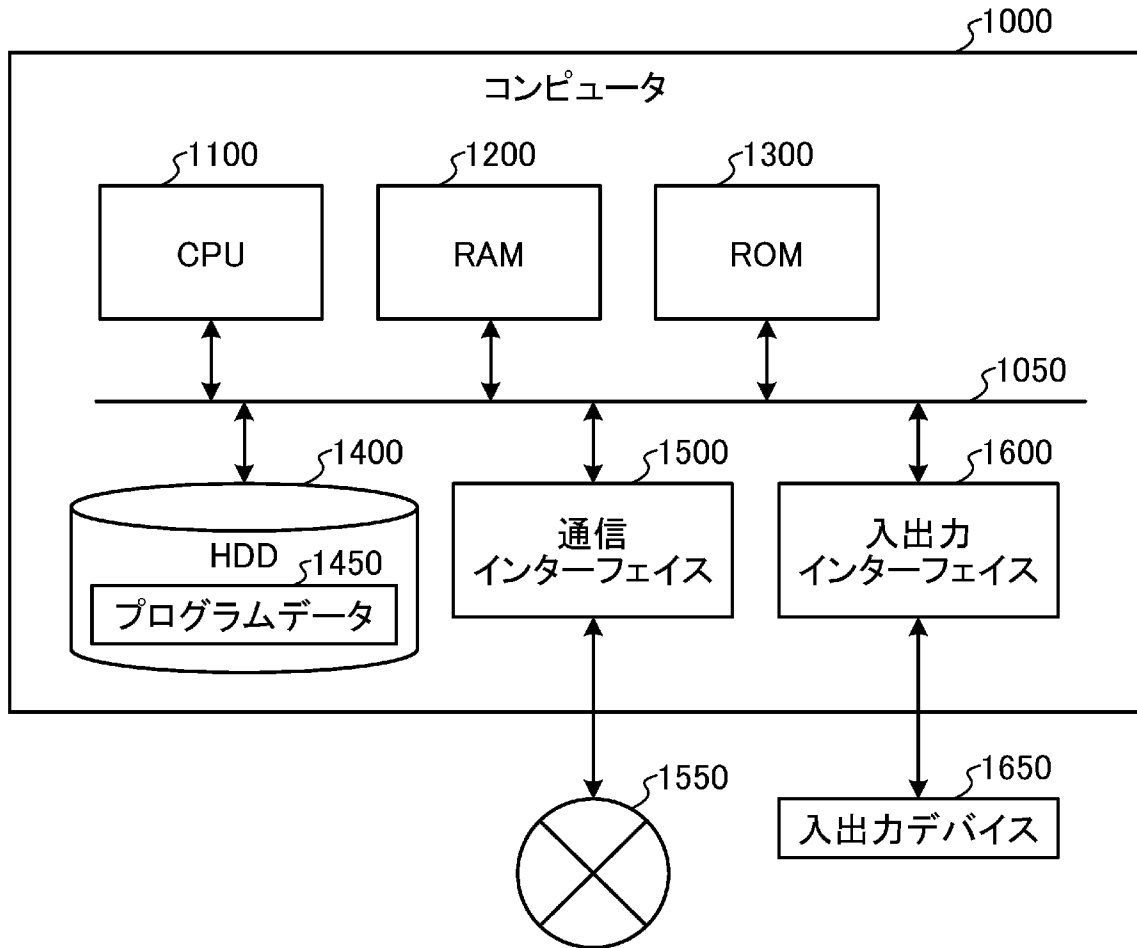
入力 (Q文)	出力 (A文)
<p>保険証券に記載する事項がわからないです</p>	<p>社名、保険契約者の氏名または名称、被保険者の氏名、受取人の氏名、または名称その他の受取人を特定するために必要な事項、保険期間および保険契約の型、基準学資金額、保険料およびその払込方法、契約日、保証証券を作成した年月日です</p>
<p>金融機関等の会社の指定した口座に送金することにより払込む方法は、保険料の払込方法<経路>に含まれますか</p>	<p>はい、含まれます</p>
<p>出生前の保険契約者の死亡について教えてください</p>	<p>被保険者となるべき者の出生前に、保険契約者が死亡したときは、第5条第1項の適用に際しては、次のとおり読替えます。 「保険契約者が死亡した場合は、被保険者となるべき者が出生した時から保険契約上の一切の権利義務を承継します。」</p>
<p>前項各号に定める事項についての研究機関等の専門機関による医学または工学等の科学技術的な特別の調査、分析または鑑定の、日数が知りたいです。</p>	<p>その場合、日数は、180日です</p>
<p>保険契約者の氏名または名称は、保険証券に記載する事項に含まれますか</p>	<p>はい、含まれます</p>

入力・出力の学習データ (例)

[図28]



[図29]



INTERNATIONAL SEARCH REPORT

International application No.

PCT/JP2021/017336

A. CLASSIFICATION OF SUBJECT MATTER
 G06F 16/83(2019_01)i; G06F 16/90(2019_01)i; G06F 40/205(2020.01)i; G06F 40/30(2020.01)i; G06F 40/56(2020.01)i
 FI: G06F16/90 100; G06F40/56; G06F40/205; G06F40/30; G06F16/83
 According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
 G06F16/83; G06F16/90; G06F40/205; G06F40/30; G06F40/56

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Published examined utility model applications of Japan	1922-1996
Published unexamined utility model applications of Japan	1971-2021
Registered utility model specifications of Japan	1996-2021
Published registered utility model applications of Japan	1994-2021

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	JP 2007-11775 A (NIPPON TELEGRAPH AND TELEPHONE CORP.) 18 January 2007 (2007-01-18) paragraphs [0013]-[0025], fig. 6	1-10
A	JP 2019-36210 A (SUMITOMO MITSUI BANKING CORP.) 07 March 2019 (2019-03-07) paragraphs [0065]-[0072]	1-10
A	US 2020/0065389 A1 (TENCENT TECHNOLOGY (SHENZHEN) COMPANY LIMITED) 27 February 2020 (2020-02-27) paragraphs [0065]-[0077]	1-10
A	US 2020/0042649 A1 (INTERNATIONAL BUSINESS MACHINES CORPORATION) 06 February 2020 (2020-02-06) paragraphs [0042]-[0055]	1-10

Further documents are listed in the continuation of Box C. See patent family annex.

* Special categories of cited documents:	“I” later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
“A” document defining the general state of the art which is not considered to be of particular relevance	“X” document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
“E” earlier application or patent but published on or after the international filing date	“Y” document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
“L” document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	“&” document member of the same patent family
“O” document referring to an oral disclosure, use, exhibition or other means	
“P” document published prior to the international filing date but later than the priority date claimed	

Date of the actual completion of the international search 09 June 2021 (09.06.2021)	Date of mailing of the international search report 13 July 2021 (13.07.2021)
--	---

Name and mailing address of the ISA/ Japan Patent Office 3-4-3, Kasumigaseki, Chiyoda-ku, Tokyo 100-8915, Japan	Authorized officer Telephone No.
--	---

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.
PCT/JP2021/017336

Patent Documents referred in the Report	Publication Date	Patent Family	Publication Date
JP 2007-11775 A	18 Jan. 2007	(Family: none)	
JP 2019-36210 A	07 Mar. 2019	(Family: none)	
US 2020/0065389 A1	27 Feb. 2020	JP 2020-522817 A paragraphs [0058]- [0069] CN 108509411 A WO 2019/072166 A1	
US 2020/0042649 A1	06 Feb. 2020	(Family: none)	

A. 発明の属する分野の分類（国際特許分類（IPC）） G06F 16/83(2019.01)i; G06F 16/90(2019.01)i; G06F 40/205(2020.01)i; G06F 40/30(2020.01)i; G06F 40/56(2020.01)i FI: G06F16/90 100; G06F40/56; G06F40/205; G06F40/30; G06F16/83		
B. 調査を行った分野 調査を行った最小限資料（国際特許分類（IPC）） G06F16/83; G06F16/90; G06F40/205; G06F40/30; G06F40/56 最小限資料以外の資料で調査を行った分野に含まれるもの 日本国実用新案公報 1922 - 1996年 日本国公開実用新案公報 1971 - 2021年 日本国実用新案登録公報 1996 - 2021年 日本国登録実用新案公報 1994 - 2021年 国際調査で使用した電子データベース（データベースの名称、調査に使用した用語）		
C. 関連すると認められる文献		
引用文献の カテゴリー*	引用文献名 及び一部の箇所が関連するときは、その関連する箇所の表示	関連する 請求項の番号
A	JP 2007-11775 A（日本電信電話株式会社）18.01.2007（2007 - 01 - 18） [0013]-[0025], [図6]	1-10
A	JP 2019-36210 A（株式会社三井住友銀行）07.03.2019（2019 - 03 - 07） [0065]-[0072]	1-10
A	US 2020/0065389 A1（TENCENT TECHNOLOGY（SHENZHEN）COMPANY LIMITED） 27.02.2020（2020 - 02 - 27） [0065]-[0077]	1-10
A	US 2020/0042649 A1（INTERNATIONAL BUSINESS MACHINES CORPORATION）06.02.2020 （2020 - 02 - 06） [0042]-[0055]	1-10
<input type="checkbox"/> C欄の続きにも文献が列挙されている。 <input checked="" type="checkbox"/> パテントファミリーに関する別紙を参照。		
* 引用文献のカテゴリー “A” 特に関連のある文献ではなく、一般的な技術水準を示すもの “E” 国際出願日前の出願または特許であるが、国際出願日以後に公表されたもの “L” 優先権主張に疑義を提起する文献又は他の文献の発行日若しくは他の特別な理由を確立するために引用する文献（理由を付す） “O” 口頭による開示、使用、展示等に言及する文献 “P” 国際出願日前で、かつ優先権の主張の基礎となる出願の日の後に公表された文献	“T” 国際出願日又は優先日後に公表された文献であって出願と抵触するものではなく、発明の原理又は理論の理解のために引用するもの “X” 特に関連のある文献であって、当該文献のみで発明の新規性又は進歩性がないと考えられるもの “Y” 特に関連のある文献であって、当該文献と他の1以上の文献との、当業者にとって自明である組合せによって進歩性がないと考えられるもの “&” 同一パテントファミリー文献	
国際調査を完了した日 09.06.2021	国際調査報告の発送日 13.07.2021	
名称及びあて先 日本国特許庁(ISA/JP) 〒100-8915 日本国 東京都千代田区霞が関三丁目4番3号	権限のある職員（特許庁審査官） 鹿野 博嗣 5N 4063 電話番号 03-3581-1101 内線 3586	

国際調査報告
 パテントファミリーに関する情報

国際出願番号

PCT/JP2021/017336

引用文献	公表日	パテントファミリー文献	公表日
JP 2007-11775 A	18.01.2007	(ファミリーなし)	
JP 2019-36210 A	07.03.2019	(ファミリーなし)	
US 2020/0065389 A1	27.02.2020	JP 2020-522817 A [0058]-[0069] CN 108509411 A WO 2019/072166 A1	
US 2020/0042649 A1	06.02.2020	(ファミリーなし)	