

(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
2024年12月12日(12.12.2024)



(10) 国際公開番号

WO 2024/252609 A1

(51) 国際特許分類:

G06F 40/56 (2020.01)

(21) 国際出願番号:

PCT/JP2023/021319

(22) 国際出願日:

2023年6月8日(08.06.2023)

(25) 国際出願の言語:

日本語

(26) 国際公開の言語:

日本語

(71) 出願人: 日本電気株式会社 (NEC CORPORATION) [JP/JP]; 〒1088001 東京都港区芝五丁目7番1号 Tokyo (JP).

(72) 発明者: 秋元 康佑 (AKIMOTO, Kosuke); 〒1088001 東京都港区芝五丁目7番1号 日本電気株式会社内 Tokyo (JP). 竹岡 邦紘 (TAKEOKA, Kunihiro); 〒1088001 東京都港区芝五丁目7番1号 日本電気株式会社内 Tokyo (JP). 小山田 昌史 (OYAMADA, Masafumi); 〒1088001 東京都港区芝五丁目7番1号 日本電気株式会社内 Tokyo (JP).

(74) 代理人: 弁理士法人 H A R A K E N Z O W O R L D P A T E N T & T

R A D E M A R K (HARAKENZO WORLD PATENT & TRADEMARK); 〒5300041 大阪府大阪市北区天神橋2丁目北2番6号 大和南森町ビル Osaka (JP).

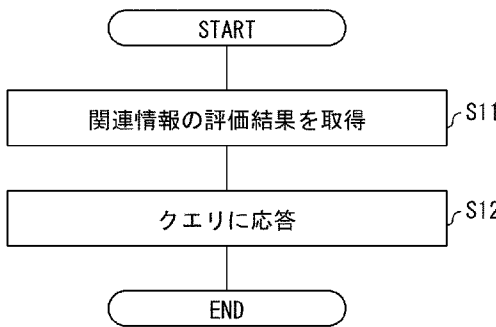
(81) 指定国(表示のない限り、全ての種類の国内保護が可能): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CV, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IQ, IR, IS, IT, JM, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, MG, MK, MN, MU, 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, CV, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SC, 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,

(54) Title: INFORMATION PROCESSING DEVICE, RESPONSE METHOD, AND RESPONSE PROGRAM

(54) 発明の名称: 情報処理装置、応答方法、および応答プログラム

[図2]



S11... ACQUIRE RESULT OF EVALUATING ASSOCIATED INFORMATION
S12... RESPOND TO QUERY

(57) Abstract: In order to increase the accuracy of a response to a query, an information processing device (1) comprises: an evaluation result acquisition unit (11) that acquires an evaluation result obtained by evaluating information associated with a query written in a natural language, the information being inputted together with the query to a machine-learning-trained language processing model that processes the query; and a response means that responds to the query using the language processing model on the basis of the evaluation result.

(57) 要約: クエリに対する応答の精度を高めるために、情報処理装置(1)は、自然言語で記述されたクエリを処理する機械学習済みの言語処理モデルに当該クエリと共に入力される当該クエリの関連情報を評価した評価結果を取得する評価結果取得部(11)と、当該評価結果に基づき、上記言語処理モデルを用いてクエリに応答する応答手段と、を備える。

[続葉有]

DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS,
IT, LT, LU, LV, MC, ME, 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、応答方法、および応答プログラム
技術分野

[0001] 自然言語で記述されたクエリに応答する情報処理装置等に関する。

背景技術

[0002] 自然言語で記述されたクエリに応答する技術が知られている。例えば、下記の特許文献1には、人工知能及び自然言語処理方法をマルチラウンド対話の検索方法に応用して、質問に対して回答の返信を行う技術が開示されている。また、特許文献1には、マルチラウンド対話における対話内容をコンテキスト情報として利用することが記載されている。コンテキスト情報を用いることにより、対話のコンテキストに沿った高精度な回答を生成することができる。

先行技術文献

特許文献

[0003] 特許文献1：日本国特開2021-108178号公報

発明の概要

発明が解決しようとする課題

[0004] また、近年では、自然言語で記述されたクエリを処理する機械学習済みの言語処理モデルに対し、クエリと共に当該クエリの関連情報を入力することにより、クエリに対する応答の精度を高める技術も開発されている。ただし、クエリと共に関連情報を入力することが、必ずしも応答の精度向上に結び付かない場合がある。これは、関連情報には、ユーザの求める応答を導くために有効なもののみが含まれているとは限らず、クエリに関係のないものや、応答をミスリードするようなものが含まれていることもあるためである。

[0005] 本発明の目的の一例は、クエリと共に入力される関連情報に応答の精度向上に結び付かない情報が含まれている場合であっても、応答の精度を高めることを可能にする情報処理装置等を提供することである。

課題を解決するための手段

[0006] 本発明の一側面に係る情報処理装置は、自然言語で記述されたクエリを処理する機械学習済みの言語処理モデルに当該クエリと共に入力される当該クエリの関連情報を評価した評価結果を取得する評価結果取得手段と、前記評価結果に基づき、前記言語処理モデルを用いて前記クエリに応答する応答手段と、を備える。

[0007] 本発明の一側面に係る応答方法は、少なくとも1つのプロセッサが、自然言語で記述されたクエリを処理する機械学習済みの言語処理モデルに当該クエリと共に入力される当該クエリの関連情報を評価した評価結果を取得することと、前記評価結果に基づき、前記言語処理モデルを用いて前記クエリに応答することと、を含む。

[0008] 本発明の一側面に係る応答プログラムは、コンピュータを、自然言語で記述されたクエリを処理する機械学習済みの言語処理モデルに当該クエリと共に入力される当該クエリの関連情報を評価した評価結果を取得する評価結果取得手段、および前記評価結果に基づき、前記言語処理モデルを用いて前記クエリに応答する応答手段、として機能させる。

発明の効果

[0009] 本発明の一態様によれば、クエリと共に入力される関連情報に応答の精度向上に結び付かない情報が含まれている場合であっても、応答の精度を高めることが可能になる。

図面の簡単な説明

[0010] [図1]本発明の例示的实施形態1に係る情報処理装置の構成を示すブロック図である。

[図2]本発明の例示的实施形態1に係る応答方法の流れを示すフロー図である。

[図3]本発明の例示的实施形態2に係る応答システムの構成を示す図である。

[図4]本発明の例示的实施形態2に係る情報処理装置の構成を示すブロック図である。

[図5]言語処理モデルにおける注目度の調整例を示す図である。

[図6]複数の訓練データセットを用いて複数の言語処理モデルを生成した例を示す図である。

[図7]本発明の例示的实施形態2に係る情報処理装置が実行する処理の一例を示すフロー図である。

[図8]クエリへの応答に対するフィードバックの受付を説明する図である。

[図9]関連情報の評価結果の提示例を示す図である。

[図10]各例示的实施形態に係る応答方法の応用例を示す図である。

[図11]本発明の各例示的实施形態に係る各装置の各機能を実現するソフトウェアであるプログラムの命令を実行するコンピュータの一例を示す図である。

発明を実施するための形態

[0011] 〔例示的实施形態1〕

本発明の第1の例示的实施形態について、図面を参照して詳細に説明する。本例示的实施形態は、後述する例示的实施形態の基本となる形態である。

[0012] (情報処理装置の構成)

本例示的实施形態に係る情報処理装置1の構成について、図1を参照して説明する。図1は、情報処理装置1の構成を示すブロック図である。図示のように、情報処理装置1は、評価結果取得部11と応答部12とを備えている。

[0013] 評価結果取得部11は、クエリの関連情報を評価した評価結果を取得する。当該クエリは、自然言語で記述されている。また、上記関連情報は、上記クエリを処理する機械学習済みの言語処理モデルに当該クエリと共に入力される情報である。

[0014] 応答部12は、評価結果取得部11が取得した評価結果に基づき、上記言語処理モデルを用いて上記クエリに応答する。

[0015] 以上のように、本例示的实施形態に係る情報処理装置1は、自然言語で記述されたクエリを処理する機械学習済みの言語処理モデルに当該クエリと共

に入力される当該クエリの関連情報を評価した評価結果を取得する評価結果取得部 11 と、評価結果取得部 11 が取得した評価結果に基づき、上記言語処理モデルを用いて上記クエリに応答する応答部 12 と、を備える。このため、本例示的实施形態に係る情報処理装置 1 によれば、クエリと共に入力する関連情報に応答の精度向上に結び付かない情報が含まれている場合であっても、応答の精度を高めることが可能になるという効果が得られる。

[0016] (応答プログラム)

上述の情報処理装置 1 の機能は、プログラムによって実現することもできる。本例示的实施形態に係る応答プログラムは、コンピュータを、自然言語で記述されたクエリを処理する機械学習済みの言語処理モデルに当該クエリと共に入力される当該クエリの関連情報を評価した評価結果を取得する評価結果取得手段、および、当該評価結果に基づき、上記言語処理モデルを用いて上記クエリに応答する応答手段として機能させる、という構成が採用されている。このため、本例示的实施形態に係る応答プログラムによれば、クエリと共に入力する関連情報に応答の精度向上に結び付かない情報が含まれている場合であっても、応答の精度を高めることが可能になるという効果が得られる。

[0017] (応答方法の流れ)

本例示的实施形態に係る応答方法の流れについて、図 2 を参照して説明する。図 2 は、応答方法の流れを示すフロー図である。なお、この応答方法における各ステップの実行主体は、情報処理装置 1 が備えるプロセッサであってもよいし、他の装置が備えるプロセッサであってもよく、各ステップの実行主体がそれぞれ異なる装置に設けられたプロセッサであってもよい。

[0018] S11 では、少なくとも 1 つのプロセッサが、自然言語で記述されたクエリを処理する機械学習済みの言語処理モデルに当該クエリと共に入力される当該クエリの関連情報を評価した評価結果を取得する。

[0019] S12 では、少なくとも 1 つのプロセッサが、S11 で取得された評価結果に基づき、上記言語処理モデルを用いて上記クエリに応答する。

[0020] 以上のように、本例示的实施形態に係る応答方法は、少なくとも1つのプロセッサが、自然言語で記述されたクエリを処理する機械学習済みの言語処理モデルに当該クエリと共に入力される当該クエリの関連情報を評価した評価結果を取得することと、当該評価結果に基づき、上記言語処理モデルを用いて上記クエリに応答することと、を含む。このため、本例示的实施形態に係る応答方法によれば、クエリと共に入力する関連情報に応答の精度向上に結び付かない情報が含まれている場合であっても、応答の精度を高めることが可能になるという効果が得られる。

[0021] 〔例示的实施形態2〕

(応答システムの概要)

本例示的实施形態に係る応答システム5の概要を図3に基づいて説明する。図3は、応答システム5の構成を示す図である。応答システム5は、応答システム5のユーザが入力するクエリに応答するシステムである。図示のように、応答システム5には、情報処理装置1Aと端末装置2とが含まれている。情報処理装置1Aと端末装置2は、ネットワーク(NW)を介して双方向に通信可能となっている。

[0022] 情報処理装置1Aは、クエリに応答する装置である。クエリとは、タスクの実行を指示するために入力される文である。本例示的实施形態では、自然言語で記載された質問文をクエリとして言語処理モデルに入力し、言語処理モデルの出力に基づいて上記クエリに対する回答文を生成し、この回答文をクエリに対する応答としてユーザに提示する例を説明する。つまり、本例示的实施形態では、上記タスクが質問文に対する回答文を出力するタスクである例を説明する。なお、クエリにより実行を指示するタスクは任意である。例えば、文あるいは文章の一部を補完するタスクや、文または文章の要約を生成するタスク、文または文章を翻訳するタスク等の実行を指示するクエリを情報処理装置1Aに入力してもよい。

[0023] 詳細は後述するが、情報処理装置1Aは、言語処理モデルにクエリと共に入力される関連情報を評価し、その評価結果に基づいてクエリに応答する。

これにより、関連情報に応答の精度向上に結び付かない情報が含まれている場合であっても、応答の精度を高めることが可能になる。なお、情報処理装置 1 A は、クエリに対する応答内容の決定を主な機能とする装置であってもよいし、他の機能も備えた汎用的な装置であってもよい。

[0024] 端末装置 2 は、応答システム 5 のユーザが使用する装置である。図 3 には、端末装置 2 がタブレット型の装置（例えば、スマートフォンまたはタブレット P C（Personal Computer））である例を記載している。端末装置 2 は、ユーザによるクエリの入力を受け付け、情報処理装置 1 A が生成する応答をユーザに提示することができるものであればよい。クエリの入力はタッチパネル等の入力デバイスを介して受け付けてもよいし、クエリを音声入力できるようにしてもよい。

[0025] 応答システム 5 のユーザは、端末装置 2 にクエリを入力することにより、端末装置 2 にクエリに対する応答を出力させることができる。例えば、図 3 の例では、端末装置 2 の表示部に質問を入力するための入力欄と、入力した質問に対する回答を表示する回答欄とが表示されている。このうち入力欄にはユーザが入力した「日本の首都は？」という質問文が表示されている。この質問文はクエリとして情報処理装置 1 A に送信される。

[0026] クエリを受信した情報処理装置 1 A は、受信したクエリの関連情報を取得する。図 3 の例では、（１）「日本の首都は法律上定められておらず、京都とする意見もある。」および（２）「東京（日本）、韓国（ソウル）、北京（中国）、…」という２つの文が関連情報として取得されている。関連情報は、例えば、クエリの全体またはクエリに含まれる一部の語を用いたキーワード検索等により取得することが可能である。

[0027] 次に、情報処理装置 1 A は、取得した関連情報の評価を行う。図 3 の例では、情報処理装置 1 A は、関連情報に含まれる２つの文のうち、（１）の文を低評価、（２）の文を高評価とし、これらの評価結果を総合した関連情報全体の評価値を 0.5 としている。

[0028] そして、情報処理装置 1 A は、算出した評価値すなわち関連情報の評価結

果に基づき、自然言語で記述されたクエリを処理する機械学習済みの言語処理モデルを用いて、クエリに応答する。クエリに対する応答の内容については、端末装置 2 を介してユーザに提示される。

[0029] このように、応答システム 5 は、自然言語で記述されたクエリの入力を受け付ける端末装置 2 と、機械学習済みの言語処理モデルに上記クエリと共に入力される当該クエリの関連情報を評価した評価結果を取得し、取得した評価結果に基づき、上記言語処理モデルを用いて上記クエリに応答する情報処理装置 1 A とを含む。応答システム 5 では、関連情報を評価した評価結果に基づいてクエリに応答するため、関連情報に応答の精度向上に結び付かない情報が含まれている場合であっても、応答の精度を高めることが可能になっている。

[0030] 例えば、図 3 の例における (1) の文は、「日本の首都は？」というクエリに対する回答が「京都」であることを示唆している。このため、この文は当該クエリに対する回答を「京都」という誤った回答にミスリードするものとなっている。一方、(2) の文は、各国の国名とその首都とをペアにして列挙したものであり、上記クエリに対する回答を「東京」という正しい回答に導くものとなっている。なお、検索により取得した関連情報に限られず、例えばユーザが入力した関連情報であっても、応答の精度向上に結び付かない情報が含まれることはあり得る。情報処理装置 1 A によれば、このように応答の精度向上に結び付かない情報が含まれた関連情報が取得された場合であっても、「東京です」という正しい応答を生成し、これを端末装置 2 に出力させることができる。

[0031] (情報処理装置の構成)

本例示的实施形態に係る情報処理装置 1 A の構成について、図 4 を参照して説明する。図 4 は、情報処理装置 1 A の構成を示すブロック図である。図示のように、情報処理装置 1 A は、情報処理装置 1 A の各部を統括して制御する制御部 10 A と、情報処理装置 1 A が使用する各種データを記憶する記憶部 11 A を備えている。また、情報処理装置 1 A は、情報処理装置 1 A が

他の装置と通信するための通信部 12 A、情報処理装置 1 A に対する各種データの入力を受け付ける入力部 13 A、および情報処理装置 1 A が各種データを出力するための出力部 14 A を備えている。なお、記憶部 11 A から出力部 14 A までの各構成は、情報処理装置 1 A に内蔵されたものであってもよいし、情報処理装置 1 A に外付けされた情報処理装置 1 A の外部の装置であってもよい。

[0032] また、図示のように、情報処理装置 1 A の制御部 10 A には、クエリ取得部 101、関連情報取得部 102、評価結果取得部 103、応答部 104、および情報提示部 105 が含まれている。また、記憶部 11 A には、言語処理モデル 111 が記憶されている。

[0033] クエリ取得部 101 は、クエリを取得する。クエリの取得方法は特に限定されない。例えば、クエリ取得部 101 は、図 3 の例のように端末装置 2 に入力されたクエリを、端末装置 2 および通信部 12 A 経由で取得してもよい。また、例えば、クエリ取得部 101 は、入力部 13 A を介して入力されるクエリを取得してもよい。

[0034] 関連情報取得部 102 は、クエリ取得部 101 が取得するクエリと共に言語処理モデル 111 に入力される、当該クエリの関連情報を取得する。関連情報の取得方法は任意である。例えば、関連情報取得部 102 は、ユーザが入力した関連情報を取得してもよいし、クエリの全体またはクエリに含まれる一部の語を用いたキーワード検索等により取得してもよい。また、関連情報取得部 102 は、所定のデータベースからクエリの内容およびクエリを入力したユーザ（クエリに対する回答の提示対象のユーザであってもよい）の少なくとも何れかに適合する情報を抽出し、それを関連情報として取得してもよい。

[0035] また、関連情報は、応答の生成に利用できるものであればよく、文あるいは文章に限られない。例えば数値、画像、または音声データを関連情報として利用することもできる。テキスト以外の形式の関連情報については、テキスト形式に変換して言語処理モデル 111 に入力すればよい。また、テキス

ト以外の形式の関連情報を入力できるように学習した言語処理モデル 111 を用いてもよく、この場合には上記のような変換は不要である。

[0036] 評価結果取得部 103 は、自然言語で記述されたクエリを処理する機械学習済みの言語処理モデルである言語処理モデル 111 に、当該クエリと共に入力される当該クエリの関連情報を評価した評価結果を取得する。評価結果取得部 103 は、関連情報を評価することにより評価結果を取得してもよいし、通信部 12A または入力部 13A を介して評価結果を取得してもよい。なお、関連情報の評価方法については後記「関連情報の評価方法」の項目で説明する。

[0037] 応答部 104 は、評価結果取得部 103 により取得された評価結果に基づき、言語処理モデル 111 を用いて、クエリ取得部 101 が取得するクエリに応答する。例えば、クエリ取得部 101 が取得したクエリが質問文であれば、応答部 104 は、言語処理モデル 111 にそのクエリを入力することにより回答文を生成する。応答に評価結果を反映させる方法については後記「関連情報の評価結果に基づく応答方法 1」～「関連情報の評価結果に基づく応答方法 4」の項目で説明する。

[0038] 言語処理モデル 111 は、自然言語で記述されたクエリを処理する機械学習済みの言語処理モデルである。例えば、言語処理モデル 111 は、自然言語で記述された質問文をクエリとして入力したときに、その質問文に対する自然言語で記述された回答文を出力するように学習された言語処理モデルであってもよい。あるいは、言語処理モデル 111 は、自然言語で記述された文に対して、所定のラベルを出力する予測モデルであってもよい。ラベルは任意に設定することができる。例えば、言語処理モデル 111 として、文の分類を示すラベルを出力する予測モデルを用いてもよい。

[0039] また、言語処理モデル 111 の出力は、自然言語に限定されず、人工言語であってもよい（数式またはソースコード等）。例えば、言語処理モデル 111 は、クエリとして、「データベースから特定の文字列を含むデータを取得するには？」という質問の入力を受け付ける。言語モデル 111 は、当該

質問に対応するデータベース処理を行うためのプログラムソースコードを出力してもよい。

[0040] また、言語処理モデル 111 が生成するコンテンツは言語で表されるもの（例えばテキストデータ）に限定されない。言語モデル 111 は、例えば、入力されたクエリに対応する画像データ、動画データ、音声データ、又はその他のデータ形式のデータを生成してもよい。

[0041] 情報提示部 105 は、ユーザに対して情報の提示を行う。例えば、情報提示部 105 は、応答部 104 が生成した回答文を、クエリを入力したユーザに提示する。情報の提示態様は任意である。例えば、情報提示部 105 は、図 3 の例のように表示部に情報を表示させることにより情報を提示してもよいし、印字装置に印字出力させる、あるいは、音声出力装置に音声出力させる等の態様で情報を提示してもよい。

[0042] 以上のように、情報処理装置 1A は、自然言語で記述されたクエリを処理する機械学習済みの言語処理モデル 111 に当該クエリと共に入力される当該クエリの関連情報を評価した評価結果を取得する評価結果取得部 103 と、取得された評価結果に基づき、言語処理モデル 111 を用いてクエリに回答する応答部 104 と、を備える。よって、クエリと共に入力する関連情報に回答の精度向上に結び付かない情報が含まれている場合であっても、回答の精度を高めることが可能になる。

[0043] （関連情報の評価方法）

上述のように、評価結果取得部 103 は、通信部 12A または入力部 13A を介して関連情報の評価結果を取得してもよい。例えば、端末装置 2 のユーザが関連情報を評価して、その評価結果を情報処理装置 1A に送信するようにしてもよいし、他の装置に関連情報を評価させて、その評価結果を情報処理装置 1A に送信させてもよい。これらの場合、評価結果取得部 103 は、通信部 12A を介した通信により、上記の評価結果を取得する。

[0044] また、評価結果取得部 103 は、関連情報を所定の評価方法で評価して評価結果を取得してもよい。これにより、例示的实施形態 1 に係る情報処理装

置 1 の奏する効果に加えて、関連情報の評価結果を自動で生成し、クエリに対する応答の精度を高めることが可能になるという効果が得られる。

[0045] 評価結果取得部 103 が適用する評価方法は、応答の精度が高まることが期待できる関連情報またはその要素が高評価、応答の精度が高まることが期待できない関連情報またはその要素が低評価となるような方法であればよい。なお、関連情報またはその要素の評価結果は、高評価と低評価のような段階的な評価結果であってもよいし、数値（つまり評価値）によって表されてもよい。

[0046] 例えば、クエリと関連性の高い関連情報またはその要素は、応答の精度を高め得ると考えられるから、評価結果取得部 103 は、このような関連情報またはその要素を高評価としてもよい。具体例を挙げれば、評価結果取得部 103 は、クエリに含まれる文字列を含む関連情報またはその要素を高評価とし、クエリに含まれる文字列を含まない関連情報またはその要素を低評価としてもよい。このように、評価結果取得部 103 は、関連情報とクエリとの関連度合いを、その関連情報の評価結果としてもよい。

[0047] また、例えば、クエリをその内容に応じて複数のカテゴリに分類すると共に、関連情報についてもその内容や取得元（ウェブページであればそのドメイン等）等に応じて複数のカテゴリに分類しておいてもよい。そして、クエリのカテゴリと、関連情報のカテゴリとの組み合わせのそれぞれについて、関連情報の平均的な評価結果を記憶部 11A 等に記憶させておいてもよい。これにより、評価結果取得部 103 は、クエリ取得部 101 が取得したクエリのカテゴリと、関連情報取得部 102 が取得した関連情報のカテゴリから、当該関連情報の評価結果を特定し、取得することができる。

[0048] また、例えば、評価結果取得部 103 は、関連情報に含まれる複数の要素の一部をクエリと共に言語処理モデル 111 に入力してクエリに対する応答を出力させる処理を、入力する要素を変えながら繰り返し行い、出力された応答のばらつきを評価してもよい。この場合、ばらつきの大きさを示す値を関連情報の評価値とすることができる。これは、関連情報に含まれる各要素

が応答の精度向上に寄与していれば（つまり関連情報が高品質であれば）応答のばらつきは小さくなると考えられるためである。また、関連情報に含まれる要素に応答をミスリードするものが多く含まれていれば（つまり関連情報が低品質であれば）応答のばらつきは大きくなると考えられるためである。

[0049] ここで、言語処理モデル 111 は、関連情報に含まれる各要素（例えば文やそれを構成する単語などの文字列）に対する注目度を算出する機能を備えているものであってもよい。注目度は、クエリに対する応答を出力する上で、各要素を重視する度合いを示すものであり、重みと言い換えることもできる。

[0050] 例えば、Transformerモデルは、アテンションメカニズムを備えており、関連情報の構成要素である各トークンのアテンション（注目度と言い換えることもできる）を算出する。このアテンションを関連情報の評価に利用することができる。具体的には、評価結果取得部 103 は、言語処理モデル 111 が関連情報に含まれる各要素について算出したアテンションの分布を評価することにより評価結果を取得してもよい。

[0051] 例えば、評価結果取得部 103 は、関連情報に含まれる複数の文の一部にアテンションが集中している場合に、その関連情報を低評価としてもよい。これは、関連情報に含まれる複数の文の一部にアテンションが集中している場合には、その関連情報にはクエリとの関連性が低く、応答の精度を高めることに寄与しない文が多く含まれていると考えられるためである。

[0052] また、評価結果取得部 103 は、関連情報に含まれる各文にアテンションが分散している場合に、その関連情報を高評価としてもよい。これは、関連情報に含まれる各文にアテンションが分散している場合には、その関連情報にはクエリとの関連性が高く、応答の精度を高めるために有用な文が多く含まれていると考えられるためである。

[0053] また、関連情報に含まれる各要素に対する注目度は、X A I（Explainable Artificial Intelligence）により算出することもできる。例えば、関連情

報が複数の文を含む場合、評価結果取得部 103 は、入力した文の極性（肯定的な内容であるか、否定的な内容であるか）を出力するように学習した言語処理モデルに、関連情報に含まれる各文を入力してもよい。ここで、評価結果取得部 103 は、XAI を用いることにより、関連情報に含まれる文または文を構成する文字列の何れに注目して出力が生成されたかを特定することができる。そして、評価結果取得部 103 は、上記の特定結果に基づき、関連情報の評価結果を生成することができる。

[0054] また、言語処理モデル 111 の学習に用いた訓練データセットに含まれる関連情報の平均的な品質が既知であるとする。この場合、評価結果取得部 103 は、その訓練データセットから抽出された関連情報については、その品質を既知の品質であるとみなしてもよい。例えば、言語処理モデル 111 を生成する際には、訓練データセットの一部を用いて言語処理モデル 111 の学習を行い、訓練データセットの残りをテストデータとすることがある。この場合、学習に用いた訓練データセットに含まれる関連情報の平均的な品質を予め推定して推定結果を記憶部 11A 等に記憶させておけばよい。これにより、評価結果取得部 103 は、テストデータに含まれる関連情報については、予め記憶された品質の推定結果を適用することができる。

[0055] また、クエリと関連情報の組に対し、その関連情報の品質（応答の精度を高めるために寄与した度合いと言い換えることもできる）を示す情報を対応付けた訓練データを用いた機械学習により、関連情報の品質を予測する品質予測モデルを生成することもできる。評価結果取得部 103 は、このような品質予測モデルにクエリと関連情報を入力することにより、当該関連情報の品質を予測し、その予測結果を評価結果として取得してもよい。

[0056] また、評価結果取得部 103 は、上述した以外にも、関連情報の信頼性や、関連情報に示される内容の的確さ等、様々な観点から関連情報を評価することが可能である。例えば、関連情報の情報源ごとに予め信頼度を定めておいてもよい。これにより、評価結果取得部 103 は、上記の信頼度を評価結果に反映させることができる。

[0057] また、ドメイン（情報源と言い換えることもできる）ごとに、そのドメインで取得された様々な関連情報の品質を評価することにより、各ドメインにおける関連情報の平均的な品質を予め特定することができる。このようなドメインごとの品質を記憶部 11A 等に記憶させておけば、評価結果取得部 103 は、評価対象となる関連情報が取得されたドメインに基づき、その関連情報の品質の評価結果を取得することができる。なお、関連情報は、情報源以外の基準で分類することも可能である。関連情報を分類すれば、その分類の関連情報の平均的な品質を予め求めることができるから、評価結果取得部 103 は、評価対象となる関連情報を分類し、その分類における平均的な品質を、当該関連情報の評価結果として取得することができる。

[0058] ここで、上述のように、関連情報には複数の要素が含まれていてもよい。関連情報に複数の要素が含まれている場合、評価結果取得部 103 は、関連情報に含まれる各要素を評価し、それらの評価結果に基づいて当該関連情報の評価結果を生成してもよい。例えば、評価結果取得部 103 は、関連情報に含まれる複数の要素のうち、高評価となった要素の数と、低評価となった要素の数との比を、当該関連情報の評価結果としてもよい。また、評価結果取得部 103 は、関連情報に含まれる各要素の評価値を算出した場合、それらの評価値の平均値を当該関連情報の評価結果としてもよい。

[0059] また、評価結果取得部 103 は、情報処理装置 1A のユーザおよびクエリの内容の少なくとも何れかに基づいて関連情報を評価してもよい。これにより、例示的实施形態 1 に係る情報処理装置 1 の奏する効果に加えて、ユーザおよびクエリの内容の少なくとも何れかに応じた応答が可能になるという効果が得られる。

[0060] 例えば、ユーザをその属性（年齢、性別、職業等）に応じて分類し、ユーザの分類毎に評価方法を定めておいてもよい。これにより、例えば、評価結果取得部 103 は、ユーザが内科医である場合には内科医にとって信頼のける情報源（例えば、内科の診療記録）から取得した関連情報を高評価にすることができる。一方、評価結果取得部 103 は、ユーザが外科医である場

合には外科医にとって信頼のおける情報源から取得した関連情報を高評価にすることができる。これにより、診療科に適合した応答を誘導することができる。

[0061] 同様に、クエリもその内容に応じて分類することが可能であり、クエリの分類毎に評価方法を予め定めておくことができる。例えば、クエリが体調不良の対処法を問うものである場合には、信頼できる対処法を発信している情報源から取得した関連情報を高評価にし、民間療法などの根拠が希薄な対処法を発信している情報源から取得した関連情報を低評価にする、といった評価方法を定めておくこともできる。

[0062] (関連情報の品質と応答の精度について)

クエリおよびその関連情報と、正しい出力との関係を学習することにより、関連情報を加味した応答を可能にする言語処理モデル 1 1 1 を生成することができる。具体的には、この学習には、クエリと関連情報の組に対し、正しい出力を示す情報を対応付けた訓練データを複数含む訓練データセットが用いられる。そして、学習中のモデルの出力と、訓練データに示される正しい出力との乖離度合いを示す損失が小さくなるように学習中のモデルのパラメータを更新する処理を、訓練データを変えながら繰り返し行うことにより、言語処理モデル 1 1 1 が生成される。

[0063] このような学習により生成された言語処理モデル 1 1 1 は、訓練データセットに含まれる関連情報を過学習したものとなり得る。例えば、正しい出力との関連性が高い要素の割合が小さい（つまり品質が低い）関連情報を用いて学習した言語処理モデル 1 1 1 は、学習時と同様の品質の関連情報に対して高精度な応答を出力し得る。しかし、このような言語処理モデル 1 1 1 に、正しい出力との関連性が高い要素の割合が大きい（つまり品質が高い）関連情報を入力したときには応答の精度が低下することがある。

[0064] このような、関連情報の品質のミスマッチによる応答精度の低下は、本願の発明者らの実験の結果によって裏付けられている。当該実験では、学習時に用いた関連情報に含まれる各文章の評価結果と、学習済みの言語処理モデ

ルに入力した関連情報に含まれる各文章の評価結果の一致度が、その言語処理モデルが出力する回答文の精度に影響を及ぼすことが確認されている。

[0065] 情報処理装置 1 A では、評価結果取得部 1 0 3 が関連情報の評価結果を取得するから、クエリと共に入力する関連情報の品質に応じた処理を行うことができ、これにより応答の精度を高めることが可能になっている。なお、関連情報の評価結果は、必ずしも処理の対象となる関連情報の個々の要素の品質を正確に反映している必要はない。

[0066] (関連情報の評価結果に基づく応答方法 1)

応答部 1 0 4 は、関連情報に含まれる複数の要素のそれぞれに対する言語処理モデル 1 1 1 の注目度を、評価結果取得部 1 0 3 が取得した評価結果に応じて調整（最適化）してもよい。これにより、言語処理モデル 1 1 1 によるクエリおよび関連情報の処理において、評価結果に応じた度合いで各要素に注目させることができるから、クエリに対する応答の精度を高めることが可能になる。

[0067] 図 5 は、言語処理モデル 1 1 1 における注目度の調整例を示す図である。図 5 には、言語処理モデル 1 1 1 に対し、関連情報として文 1 1 2 a と文 1 1 2 b が入力された例を示している。そして、言語処理モデル 1 1 1 は、文 1 1 2 a の注目度 s_1 を 0.7、文 1 1 2 b の注目度 s_2 を 0.3 と算出している。なお、1つのクエリに対して関連情報を1つ入力する言語処理モデル 1 1 1 を用いる場合、その言語処理モデル 1 1 1 への入力は、例えば「クエリ+文 1 1 2 a < s e p > クエリ+文 1 1 2 b < s e p > 答えは : 」のような形式としてもよい。< s e p > はセパレータである。

[0068] 言語処理モデル 1 1 1 として、T 5 (Text-To-Text Transfer Transformer) や G P T (Generative Pretrained Transformer) などのTransformerモデルや、それらを応用したFiDやFiEなどのモデルを用いる場合、これらのモデルにおいて「アテンション」と呼ばれる数値を注目度として利用することができる。FiDについては、Izacard, Gautier, and Edouard Grave. "Leveraging passage retrieval with generative models for open domain question

answering.” arXiv preprint arXiv:2007.01282 (2021)に記載されている。

また、FiEについてはKedia, Akhil, Mohd Abbas Zaidi, and Haejun Lee. ”FiE: Building a Global Probability Space by Leveraging Early Fusion in Encoder for Open-Domain Question Answering.” arXiv preprint arXiv:2211.10147 (2022)に記載されている。

[0069] また、言語処理モデル111として、RAG (Retrieval-Augmented Generation) などの、関連情報中の個々の文（またはその一部）に対して重みを付与し、重みが大きいほど処理結果に大きく影響を及ぼすようなモデルを用いてもよい。この場合、応答部104は、言語処理モデル111により算出された文またはその一部に対する重みを、文またはその一部に対する注目度として利用することができる。RAGについてはLewis, Patrick, et al. ”Retrieval-augmented generation for knowledge-intensive nlp tasks.” Advances in Neural Information Processing Systems 33 (2020): 9459-9474に記載されている。

[0070] 図5に示す調整例1では、文112aの注目度 s_1 が $s_1' = 0.5$ に調整され、文112bの注目度 s_2 が $s_2' = 0.5$ に調整されている。つまり、評価結果取得部103が文112aおよび112bの評価結果として言語処理モデル111から取得した注目度 s_1 、 s_2 が、応答部104により s_1' 、 s_2' に調整されて言語処理モデル111に戻される。戻された $s_1' = s_2'$ であるから言語処理モデル111は文112aと文112bに均等に注目して回答文を生成する。応答部104は、関連情報の品質が全体として高い（言い換えれば関連情報に含まれる要素に占める、正しい出力との関連性が高い要素の割合が大きい）場合に、このような調整（つまり関連情報の複数の要素の注目度を一様にする調整）を行えばよい。これにより、関連情報に含まれる正しい出力との関連性が高い要素の考慮漏れが生じる可能性を低減することができる。

[0071] 一方、図5に示す調整例2では、文112aの注目度 s_1 が $s_1'' = 0.9$ に調整され、文112bの注目度 s_2 が $s_2'' = 0.1$ に調整されている。応

応答部 104 が注目度をこのようにして調整することにより、言語処理モデル 111 は文 112 b よりも文 112 a に注目して回答文を生成する。応答部 104 は、関連情報の品質が全体として低い（言い換えれば関連情報に含まれる要素に占める、正しい出力との関連性が高い要素の割合が小さい）場合に、このような調整（つまり関連情報の要素のうち注目度が高い要素の注目度をさらに高くする調整）を行えばよい。これにより、関連情報に含まれる正しい出力との関連性が低い要素を考慮することにより応答精度が低下する可能性を低減することができる。

[0072] なお、厳密には、Transformerモデルは、文をトークンと呼ばれる当該文の構成要素に分割して処理する。このため、アテンションはトークンごとに算出される。応答部 104 は、文に含まれる各要素に対応するトークンのアテンションから、その文全体の注目度を算出することができる。また、文およびその構成要素はベクトルで表現される。図 5 においては、1つのトークンに対応するベクトルを1つの矩形で示している。なお、言語処理モデル 111 によっては、その言語処理モデル 111 が備えるデコーダのレイヤごとに異なるアテンションが設定される場合もある。この場合、応答部 104 は、少なくとも何れか一つのレイヤにおけるアテンションを注目度として取得し、その値（あるいはそれらの値）を調整すればよい。

[0073] 応答部 104 は、調整後の注目度（アテンションあるいは重み値と言い換えることもできる）を計算で求めることができる。例えば、関連情報に含まれる 1〜N（Nは2以上の整数）の要素（例えば文）について言語処理モデル 111 が算出した注目度が $\{s_1, s_2, \dots, s_N\}$ であったとする。このとき、応答部 104 は、各要素の注目度を下記の数式を用いて調整してもよい。なお、「 $\text{softmax}(\cdot)$ 」はソフトマックス関数であり、正規化された注目度を得るために用いている。

$$\{s_1', s_2', \dots, s_N'\} = \{\text{softmax}(\log(s_1)/T, \log(s_2)/T, \dots, \log(s_N)/T)\}$$

具体的には、応答部 104 は、上記数式における T（Tは $T > 0$ のスカラ

値)を、評価結果取得部103の評価結果に応じた値にすることにより、各要素の調整後の注目度 $\{s_1', s_2', \dots, s_N'\}$ を算出する。この調整法を適用した場合、 T を大きい値にするほど各要素の調整後の注目度は一様に近づく。一方、 T を小さい値にするほど、注目度が大きい要素の注目度がより大きい値となり、かつ、注目度が小さい要素の注目度がより小さい値となる。このため、応答部104は、例えば、関連情報の品質が全体として高いほど T の値を大きくし、関連情報の品質が全体として低いほど T の値を小さくしてもよい。

[0074] また、応答部104は、調整後の注目度 $\{s_1', s_2', \dots, s_N'\}$ のエントロピーが所定の値となるような T を算出してもよい。この場合、評価結果取得部103が取得する評価結果を目標とするエントロピーの値に変換する変換方法を予め定めておけばよい。これにより、応答部104は、評価結果取得部103が取得する評価結果から目標とするエントロピーの値を求め、求めたエントロピーの値から T を算出し、算出した T から各要素の調整後の注目度を算出することができる。

[0075] なお、上述のように、言語処理モデル111には、各トークンの注目度（アテンションまたは重み値と言い換えることもできる）を算出するものがある。このような言語処理モデル111を用いる場合、応答部104は、1つの文における各トークンの注目度の比を調整後も維持するようにしてもよい。

[0076] 例えば、関連情報に2つの文が含まれており、1つ目の文は3つのトークン、2つ目の文は2つのトークンで構成されていたとする。そして、言語処理モデル111は、1つ目の文の3つのトークンの注目度を、先頭から順に、0.2、0.2、0.1と算出し、2つ目の文の2つのトークンの注目度を、先頭から順に、0.4、0.1と算出したとする。これら2つの文の注目度は何れも全体として0.5となる。

[0077] ここで、応答部104は、1つ目の文の注目度を0.5から0に調整し、2つ目の文の注目度を0.5から1に調整することを決定したとする。この

場合、応答部 104 は、1 つ目の文の 3 つのトークンの注目度を全てゼロとし、2 つ目の文の 2 つのトークンの注目度を先頭から順に、0.8、0.2 とする。これにより、2 つ目の文の 2 つのトークンにおける注目度の比は、調整後も調整前も 4 : 1 で維持される。

[0078] (関連情報の評価結果に基づく応答方法 2)

上述のように、言語処理モデル 111 は、クエリと当該クエリの関連情報を入力して得られる出力と、与えられた正しい出力との乖離度が低くなるように学習することにより生成されたものであってもよい。この場合、応答部 104 は、評価結果がそれぞれ異なる関連情報を用いて学習された複数の言語処理モデル 111 のうち、評価結果取得部 103 が取得した評価結果に応じた評価結果の関連情報を用いて学習された言語処理モデル 111 を用いてクエリに応答する。

[0079] 上述のように、本願の発明者らの実験により、学習時に用いた関連情報に含まれる各要素の評価結果と、学習済みの言語処理モデルに入力した関連情報に含まれる各要素の評価結果との一致度が応答の精度に影響を及ぼすことが確認された。よって、関連情報に含まれる各要素の評価結果に応じた言語処理モデル 111 を用いて応答する上記の構成によれば、クエリに対する応答の精度を高めることが可能になる。

[0080] 図 6 は、評価結果がそれぞれ異なる関連情報を用いて複数の言語処理モデル 111 を生成した例を示す図である。具体的には、図 6 には、訓練データセット 113 a ~ 113 c を用いて言語処理モデル 111 a ~ 111 c を生成した例を示している。

[0081] 訓練データセット 113 a には、クエリと関連情報と正答（与えられた正しい出力と言い換えることもできる）とが対応付けられた訓練データが複数含まれている。訓練データセット 113 a に含まれている関連情報は、全体として低品質である。このような訓練データセット 113 a を生成する場合、関連情報の候補となる文または文章を複数用意し、各候補のそれぞれをクエリと共に言語処理モデル 111 に入力し、当該候補が応答の精度向上に寄

与するかを判定すればよい。そして、訓練データセットの全体において、精度向上に寄与しない関連情報の割合を、精度向上に寄与する関連情報よりも多くすることにより、低品質な関連情報を含む訓練データセット 1 1 3 a を生成することができる。また、同様にして、中品質な関連情報を含む訓練データセット 1 1 3 b および高品質な関連情報を含む訓練データセット 1 1 3 c を生成することもできる。

[0082] なお、関連情報が精度向上に寄与するか否かは、その関連情報とクエリと言語処理モデル 1 1 1 に入力して得られる出力と正答との乖離度により判定することができる。乖離度は、例えば公知の類似度指標等で表すことができる。

[0083] また、訓練データセットにおける関連情報の評価方法は特に限定されない。例えば、正答の文字列を含む関連情報を高品質、正答の文字列を含まない関連情報を低品質とみなしてもよい。この場合、複数の関連情報における、高品質な関連情報と低品質な関連情報との比を、それら関連情報の全体としての品質としてもよい。つまり、関連情報の全体としての品質が所定の値となるように、正答の文字列を含む関連情報の数と、正答の文字列を含まない関連情報の数とを調整することにより、所定の品質の関連情報を含む訓練データセットを生成することができる。

[0084] 訓練データセット 1 1 3 a を用いた学習により生成された言語処理モデル 1 1 1 a は、低品質の関連情報を用いた応答に使用される。つまり、関連情報取得部 1 0 2 が取得した関連情報について評価結果取得部 1 0 3 が取得した評価結果が、当該関連情報の品質が低いことを示していた場合、応答部 1 0 4 は、言語処理モデル 1 1 1 a を用いてクエリに応答する。例えば、応答部 1 0 4 は、評価結果取得部 1 0 3 により取得された評価値が所定の下限値未満であったときに、言語処理モデル 1 1 1 a を用いてクエリに応答してもよい。

[0085] また、訓練データセット 1 1 3 b および 1 1 3 c にも、クエリと関連情報と正答とが対応付けられた訓練データが複数含まれている。訓練データセッ

ト 1 1 3 b に含まれている関連情報は全体として中程度の品質であり、訓練データセット 1 1 3 c に含まれている関連情報は全体として高品質である。

[0086] 訓練データセット 1 1 3 b を用いた学習により生成された言語処理モデル 1 1 1 b は、中品質の関連情報を用いた応答に使用される。つまり、関連情報取得部 1 0 2 が取得した関連情報について評価結果取得部 1 0 3 が取得した評価結果が、当該関連情報の品質が中程度であることを示していた場合、応答部 1 0 4 は、言語処理モデル 1 1 1 b を用いてクエリに応答する。例えば、応答部 1 0 4 は、評価結果取得部 1 0 3 により取得された評価値が所定の上限値以下、所定の下限値以上であったときに、言語処理モデル 1 1 1 b を用いてクエリに応答してもよい。

[0087] また、訓練データセット 1 1 3 c を用いた学習により生成された言語処理モデル 1 1 1 c は、高品質の関連情報を用いた応答に使用される。つまり、関連情報取得部 1 0 2 が取得した関連情報について評価結果取得部 1 0 3 が取得した評価結果が、当該関連情報の品質が高いことを示していた場合、応答部 1 0 4 は、言語処理モデル 1 1 1 c を用いてクエリに応答する。例えば、応答部 1 0 4 は、評価結果取得部 1 0 3 により取得された評価値が所定の上限値を越えているときに、言語処理モデル 1 1 1 c を用いてクエリに応答してもよい。

[0088] このように、応答部 1 0 4 は、言語処理モデル 1 1 1 a ~ 1 1 1 c のうち、評価結果取得部 1 0 3 が取得した評価結果に応じた評価結果の関連情報を用いて学習されたモデルを用いて応答を生成してもよい。これにより、学習時と推論時における関連情報の質の乖離を要因とする応答精度の低下を抑えることが可能になる。

[0089] なお、評価結果取得部 1 0 3 は、評価値ではなく、関連情報の品質が複数の分類（例えば、高品質、中品質、低品質）の何れに該当するかを示す評価結果を取得してもよい。この場合、応答部 1 0 4 は、言語処理モデル 1 1 1 a ~ 1 1 1 c のうち、評価結果が示す分類に対応するものを用いて応答すればよい。

[0090] (関連情報の評価結果に基づく応答方法3)

また、応答部104は、関連情報に含まれる複数の要素のうち、評価結果取得部103が取得した評価結果に応じて選択した要素をクエリと共に言語処理モデル111に入力することにより得られる出力に基づきクエリに応答してもよい。

[0091] 上述のように、関連情報の評価結果がどのようなものであるかは、言語処理モデル111を用いた応答の精度に影響を及ぼし得る。よって、評価結果に応じて選択した要素を言語処理モデル111に入力する上記の構成によれば、クエリに対する応答の精度を高めることが可能になる。

[0092] 例えば、言語処理モデル111を生成する際に用いた訓練データセットに含まれる関連情報のうち応答精度の向上に寄与するものと、応答精度の向上に寄与しないものとの比が3:7であったとする。この場合、応答部104は、関連情報に含まれる複数の要素から、応答精度の向上に寄与するものと寄与しないものとを3:7の割合で抽出し、それらをクエリと共に言語処理モデル111に入力してもよい。

[0093] このように、応答部104は、学習に用いた関連情報に占める応答精度の向上に寄与するものの割合と、言語処理モデル111に入力する関連情報の複数の要素に占める応答精度の向上に寄与するものの割合とを一致させてもよい。これにより、クエリに対する応答の精度を高めることが可能になる。なお、上記の割合は完全に一致している必要はなく、応答部104は、応答精度の向上が期待できる程度に一致していればよい。

[0094] また、図6の例のように、言語処理モデル111を生成する際に用いた訓練データセットに含まれる関連情報の品質を、例えば、高品質、低品質等に予め分類しておいてもよい。この場合、応答部104は、言語処理モデル111を生成する際に用いた訓練データセットに含まれる関連情報の品質と同じ分類になるように、関連情報取得部102が取得した関連情報から要素を選択すればよい。

[0095] (関連情報の評価結果に基づく応答方法4)

言語処理モデル１１１は、クエリと当該クエリの関連情報を入力して得られる出力と、与えられた正しい出力との乖離度が低くなるように学習することにより生成されたものであってもよいことは上述のとおりである。この学習において使用される関連情報には、当該関連情報の評価結果を示す情報が付与されていてもよい。つまり、言語処理モデル１１１は、評価結果を示す情報が付与された関連情報を用いて学習することにより生成されたモデルであってもよい。

[0096] このような言語処理モデル１１１を用いる場合、応答部１０４は、クエリと、評価結果取得部１０３が取得した評価結果を示す情報を付与した、当該クエリの関連情報とを言語処理モデル１１１に入力することにより得られる出力に基づきクエリに応答してもよい。このような構成を採用した場合にも、クエリに対する応答の精度を高めることが可能になる。

[0097] 例えば、言語処理モデル１１１の学習に用いる訓練データセットに含まれる関連情報のうち応答精度の向上に寄与するものには、関連情報の評価結果を示す情報として「G o o d」との文字列を付与してもよい。また、言語処理モデル１１１の学習に用いる訓練データセットに含まれる関連情報のうち応答精度の向上に寄与しないものには、関連情報の評価結果を示す情報として「B a d」との文字列を付与してもよい。このような訓練データセットを用いて学習を行うことにより、「G o o d」との文字列を含む関連情報またはその要素の注目度を高く、「B a d」との文字列を含む関連情報またはその要素の注目度は低くするような言語処理モデル１１１を生成することができる。

[0098] そして、この場合、応答部１０４は、評価結果取得部１０３が取得した評価結果が、応答精度の向上に寄与することを示す関連情報またはその要素には「G o o d」との文字列を付与する。また、応答部１０４は、評価結果取得部１０３が取得した評価結果が、応答精度の向上に寄与しないことを示す関連情報またはその要素には「B a d」との文字列を付与する。応答部１０４は、このようにして評価結果を示す情報を付与した関連情報をクエリと共

に言語処理モデル 111 に入力する。これにより、応答精度の向上に寄与する関連情報またはその要素の注目度が高く、応答精度の向上に寄与しない関連情報またはその要素の注目度が低くなるように誘導することができるから、応答の精度を高めることが可能になる。

[0099] (処理の流れ)

情報処理装置 1A が実行する処理の流れを図 7 に基づいて説明する。図 7 は、情報処理装置 1A が実行する処理の一例を示すフロー図である。なお、以下では、「関連情報の評価結果に基づく応答方法 1」で説明した応答方法を適用する例を説明する。

[0100] S21 では、クエリ取得部 101 がクエリを取得する。例えば、クエリ取得部 101 は、図 3 の例のように、端末装置 2 を介して入力されたクエリを、通信部 12A を介した通信により取得してもよい。

[0101] S22 では、関連情報取得部 102 が関連情報を取得する。以下では、複数の文を要素として含む関連情報が取得された例を説明する。

[0102] S23 では、応答部 104 が、S21 で取得されたクエリと、S22 で取得された関連情報を言語処理モデル 111 に入力し、言語処理モデル 111 が関連情報に含まれる各要素（具体的には文または単語）について算出した注目度（例えばアテンション）を特定する。なお、S23 の処理は、S25 より先に行われればよく、S24 の後または S24 と並行で行われてもよい。

[0103] S24 では、評価結果取得部 103 が、S22 で取得された関連情報の評価結果を取得する。例えば、評価結果取得部 103 は、S22 で取得された関連情報に含まれる要素の 1 つである文と S21 で取得されたクエリとを言語処理モデル 111 に入力して出力値を得るという処理を、当該関連情報に含まれる各文について行ってもよい。そして、評価結果取得部 103 は、得られた出力値のばらつきを評価して評価結果を取得してもよい。また、例えば、評価結果取得部 103 は、S23 で特定された注目度が、関連情報に含まれる複数の文の一部に集中しているか否かにより当該関連情報を評価して

評価結果を取得してもよい。また、評価結果取得部 103 は、ユーザが入力する評価結果を取得してもよい。

[0104] S25 では、応答部 104 が、S25 で取得された評価結果に応じて、S22 で取得された関連情報に含まれる各要素（具体的には文または単語）に対する言語処理モデル 111 の注目度を調整する。

[0105] S26 では、応答部 104 は、S25 で注目度が調整された言語処理モデル 111 を用いて S21 で取得されたクエリに応答する。具体的には、応答部 104 は、注目度が調整された言語処理モデル 111 に、S21 で取得されたクエリと、S22 で取得された関連情報とを入力して言語処理モデル 111 の出力値を得る。例えば、クエリが質問文である場合、出力値はその質問文に対する回答文であってもよい。そして、応答部 104 は、情報提示部 105 に指示して上記の出力値をユーザに提示させる。これにより、図 7 の処理は終了となる。

[0106] なお、「関連情報の評価結果に基づく応答方法 2」で説明した応答方法を適用する場合、S25 の処理は行われず、S26 では S24 の評価結果に応じた言語処理モデル 111 を用いて応答がなされる。

[0107] また、「関連情報の評価結果に基づく応答方法 3」で説明した応答方法を適用する場合も、S25 の処理は行われず、その代わりに、S24 の評価結果に応じて、関連情報に含まれる複数の要素の一部を選択する処理が行われる。そして、S26 では、選択された要素とクエリとを言語処理モデル 111 に入力することにより応答がなされる。

[0108] また、「関連情報の評価結果に基づく応答方法 4」で説明した応答方法を適用する場合も、S25 の処理は行われず、その代わりに、S24 の評価結果を示す情報を関連情報に付与する処理が行われる。そして、S26 では、S24 の評価結果を示す情報が付与された関連情報とクエリとを言語処理モデル 111 に入力することにより応答がなされる。

[0109] 以上説明した図 7 のフローには、本例示的实施形態に係る応答方法が含まれている。当該応答方法は、自然言語で記述されたクエリを処理する機械学

習済みの言語処理モデル 111 に当該クエリと共に入力される当該クエリの関連情報を評価した評価結果を取得すること（S24）と、当該評価結果に基づき、言語処理モデル 111 を用いてクエリに応答すること（S26）と、を含む。この応答方法によれば、クエリと共に入力する関連情報に応答の精度向上に結び付かない情報が含まれている場合であっても、応答の精度を高めることが可能になる。

[0110] （フィードバックの受付について）

情報処理装置 1A は、上述のようにしてクエリに応答した後、その応答に対するユーザのフィードバックを受け付けてもよい。これにより、情報処理装置 1A は、フィードバックを反映させたより好ましい応答を行うことが可能になる。本項目ではフィードバックの受付について図 8 に基づいて説明する。図 8 は、クエリへの応答に対するフィードバックの受付を説明する図である。図 8 には、フィードバックを受け付ける際の処理の流れを示すフロー図と、フィードバックを受け付ける際に表示させる画像 IMG1 とを示している。

[0111] 図 8 のフロー図における S31 では、応答部 104 が、直近に生成した応答に対するユーザのフィードバックを受け付ける。S32 では、応答部 104 は、S31 で受け付けたフィードバックの内容に応じて、応答を再生成するか否かを判定する。S32 で再生成すると判定された場合には S33 に進み、S32 で再生成しないと判定された場合には図 8 に示すフロー図の処理は終了する。

[0112] S31 において、例えば、応答部 104 は、情報提示部 105 に指示して図 8 に示す画像 IMG1 を端末装置 2 に表示させることにより、フィードバックを受け付けてもよい。画像 IMG1 には、応答内容の適否を問うメッセージ M1 が含まれている。また、画像 IMG1 には、ユーザが選択可能なオブジェクトとして、応答内容の適否を「良い」または「悪い」で評価するためのオブジェクト Ob1、Ob2 が含まれている。

[0113] ユーザは、オブジェクト Ob1、Ob2 のうち、自身による応答内容の適

否の評価に近いものを選ぶことにより、応答内容を評価することができる。
また、オブジェクトO b 2は、応答を再生成させるためのオブジェクトでもある。オブジェクトO b 2が選択された場合、S 3 2ではY E Sと判定され、応答の再生成が行われる。一方、オブジェクトO b 1が選択された場合、S 3 2ではN Oと判定され、処理は終了となる。

[0114] S 3 3では、応答部1 0 4は、関連情報の各要素に対する言語処理モデル1 1 1の注目度を再調整する。S 3 4では、応答部1 0 4は、S 3 3で注目度が再調整された言語処理モデル1 1 1を用いてクエリに応答する。この後、処理はS 3 1に戻る。

[0115] S 3 3における再調整の方法は予め定めておけばよい。例えば、応答部1 0 4は、関連情報の各要素に対する言語処理モデル1 1 1の注目度を、直近の調整が行われる前の値に戻してもよい。また、例えば、応答部1 0 4は、注目度をスカラ値Tで調整する場合、Tの値を所定幅で増加または減少させることにより注目度を調整してもよい。

[0116] なお、応答を再生成する方法は上述の例に限られない。例えば、「関連情報の評価結果に基づく応答方法2」で説明した応答方法を適用する場合、応答部1 0 4は、先の応答生成時とは異なる言語処理モデル1 1 1を用いて応答を再生成してもよい。また、例えば、評価結果取得部1 0 3が関連情報の評価結果を更新し、応答部1 0 4が更新後の評価結果に基づいて応答を再生成するようにしてもよい。

[0117] (関連情報の評価結果の提示)

情報提示部1 0 5は、評価結果取得部1 0 3が取得した関連情報の評価結果をユーザに提示してもよい。図9は、関連情報の評価結果の提示例を示す図である。図示の例では、端末装置2の表示部に、関連情報の評価結果を示すメッセージM 2と、評価結果に応じた対処法を示すメッセージM 3が表示されている。メッセージM 2に示される「0. 1」という評価値は、評価結果取得部1 0 3が取得した関連情報の評価結果を示している。

[0118] メッセージM 3については、その候補となる文章を予め複数パターン用意

しておいてもよい。例えば、評価結果が悪い場合のメッセージとして他の関連情報の入力を勧める文章を記憶部 11A 等に記憶させておくと共に、評価結果が良い場合の文章としてクエリの手直しを勧めるメッセージを記憶部 11A 等に記憶させておいてもよい。そして、情報提示部 105 は、記憶部 11A 等に記憶された文章の中から、評価結果取得部 103 が取得した関連情報の評価結果に応じたものを選択して、端末装置 2 等に表示させてもよい。

[0119] (応用例)

クエリの入力を受け付け、入力されたクエリに対する応答をブラウザや、アプリケーションソフトウェアの表示画面等に出力させるサービスが知られている。このようなサービスにおいて、上述の例示的实施形態に係る応答方法を適用することにより、関連情報の評価結果を反映させて応答の精度を高めることもできる。この場合、当該サービスを提供するための情報処理装置（例えばサーバ）が、情報処理装置 1 または情報処理装置 1A ということになる。

[0120] また、入力されたクエリについて、それぞれ異なる複数の言語処理モデルにより応答を生成し、それらの応答を提示するサービスも知られている。このようなサービスにおいて、上述の例示的实施形態に係る応答方法を適用することにより、各応答に関連情報の評価結果を反映させてそれら応答の精度を高めることも可能である。以下、上述の例示的实施形態に係る応答方法を、医療・ヘルスケア関連サービスに適応した例を説明する。

[0121] 図 10 は、各例示的实施形態に係る応答方法の応用例を示す図である。図 10 には、クエリの入力受付および応答出力のために表示される画像 IMG 2 と、画像 IMG 2 に示される回答を生成するための関連情報として用いた医療データ 112c とを示している。

[0122] 画像 IMG 2 は、情報提示部 105 により例えば端末装置 2 の表示部等に表示される。画像 IMG 2 は、例えば、ブラウザの表示画面に表示されてもよいし、入力したクエリに応答を返すアプリケーションソフトウェアの表示画面に表示されてもよい。

- [0123] 医療データ 112c は関連情報取得部 102 により取得される。例えば、関連情報取得部 102 は、ユーザが端末装置 2 を介して入力する医療データ 112c を取得してもよいし、医療データを記憶している各種データベースにアクセスする等して医療データ 112c を取得してもよい。
- [0124] 画像 IMG 2 には、ユーザが入力したクエリ Q 1 が表示されていると共に、クエリ Q 1 について、第 1 の言語処理モデル 111 を用いて生成した回答 A 1 と、第 2 の言語処理モデル 111 を用いて生成した回答 A 2 とが表示されている。第 1 の言語処理モデル 111 と第 2 の言語処理モデル 111 は、それぞれ異なるモデルであるため、同じクエリ Q 1 および同じ関連情報に対して、回答 A 1 と A 2 はそれぞれ異なる内容になっている。ただし、何れの回答も上述の例示的实施形態に係る応答方法により生成されたものであるから、関連情報の品質を考慮しない場合と比べて回答内容の精度は高まっていると期待できる。
- [0125] このように 2 つの言語処理モデルのそれぞれに応答を生成させる場合、情報処理装置 1A の制御部 10A には、第 1 の言語処理モデル 111 を用いてクエリに応答する第 1 の応答部 104 と、第 2 の言語処理モデル 111 を用いてクエリに応答する第 2 の応答部 104 とを設ければよい。第 1 の応答部 104 は、評価結果取得部 103 が取得する評価結果に基づいて、第 1 の言語処理モデル 111 を用いてクエリに応答する。また、第 2 の応答部 104 は、評価結果取得部 103 が取得する評価結果に基づいて、第 2 の言語処理モデル 111 を用いてクエリに応答する。そして、情報提示部 105 は、第 1 の応答部 104 による応答の内容と、第 2 の応答部 104 による応答の内容とをユーザに提示する。
- [0126] なお、第 1 および第 2 の言語処理モデル 111 のそれぞれについて、個別の評価結果取得部 103 を設けてもよい。この場合、使用する言語処理モデル 111 のそれぞれに適した評価結果を取得し、その評価結果に基づいて応答することができる。また、3 つ以上の言語処理モデルを併用することも可能である。

[0127] 医療データ 112c は、医療機関で収集された各種情報を示すデータである。具体的には、医療データ 112c は、収集された医療機関を示す情報と、その医療機関で診察が行われた日にちと、患者の症状を示す情報と、医師等の診断結果を示す情報と、処方された医薬品を示す情報とが対応付けられたものである。医療データ 112c は、患者の症状に応じた診断結果あるいは処方薬を問うクエリに対する妥当な回答を導くために有用なものであり、関連情報として好適である。

[0128] ここで、上述のように、評価結果取得部 103 は、ユーザおよびクエリの内容の少なくとも何れかに応じた評価方法で関連情報を評価した評価結果を取得してもよい。例えば、図 10 に示すクエリ Q1 には「患者」との語が含まれている。この場合、評価結果取得部 103 は、医療関係の信頼できる情報源から得られた関連情報についての評価結果を肯定的なものとしてもよい。

[0129] また、例えば、クエリ Q1 を入力したユーザが、特定の診療科を受診したことが分かっている場合、評価結果取得部 103 は、当該診療科のデータを、他の診療科のデータよりも高評価としてもよい。例えば、クエリ Q1 の一行目が「内科を受診した患者さんです。」という文であった場合、評価結果取得部 103 は、医療データ 112c に含まれるデータのうち「内科」のデータを、他の診療科のデータよりも高評価としてもよい。これにより、処方薬として「薬剤 a」を勧める回答を誘導することができる。

[0130] [変形例]

上述の実施形態で説明した各処理の実行主体は任意であり、上述の例に限られない。つまり、相互に通信可能な複数の装置（プロセッサということもできる）により、情報処理装置 1 および 1A の機能を実現することができる。例えば、図 2、図 7、および図 8 の各フローチャートに記載されている各処理を複数のプロセッサに分担で実行させることもできる。つまり、上述の実施形態における応答方法の実行主体は、1つのプロセッサであってもよいし、複数のプロセッサであってもよい。

[0131] 〔ソフトウェアによる実現例〕

情報処理装置 1, 1 A の一部又は全部の機能は、集積回路（ＩＣチップ）等のハードウェアによって実現してもよいし、ソフトウェアによって実現してもよい。

[0132] 後者の場合、情報処理装置 1, 1 A は、例えば、各機能を実現するソフトウェアであるプログラムの命令を実行するコンピュータによって実現される。このようなコンピュータの一例（以下、コンピュータ C と記載する）を図 11 に示す。コンピュータ C は、少なくとも 1 つのプロセッサ C 1 と、少なくとも 1 つのメモリ C 2 と、を備えている。メモリ C 2 には、コンピュータ C を情報処理装置 1, 1 A として動作させるためのプログラム（応答プログラム）P が記録されている。コンピュータ C において、プロセッサ C 1 は、プログラム P をメモリ C 2 から読み取って実行することにより、情報処理装置 1, 1 A の各機能が実現される。例えば、図 3 に示した端末装置 2 にプログラム P を実行させることもできる。この場合、端末装置 2 単体でクエリの入力の受付から、クエリに対する応答内容の決定、および応答の出力まで行うことができる。

[0133] プロセッサ C 1 としては、例えば、C P U（Central Processing Unit）、G P U（Graphic Processing Unit）、D S P（Digital Signal Processor）、M P U（Micro Processing Unit）、F P U（Floating point number Processing Unit）、P P U（Physics Processing Unit）、T P U（Tensor Processing Unit）、量子プロセッサ、マイクロコントローラ、又は、これらの組み合わせなどを用いることができる。メモリ C 2 としては、例えば、フラッシュメモリ、H D D（Hard Disk Drive）、S S D（Solid State Drive）、又は、これらの組み合わせなどを用いることができる。

[0134] なお、コンピュータ C は、プログラム P を実行時に展開したり、各種データを一時的に記憶したりするための R A M（Random Access Memory）を更に備えていてもよい。また、コンピュータ C は、他の装置との間でデータを送受信するための通信インタフェースを更に備えていてもよい。また、コンピ

ユータCは、キーボードやマウス、ディスプレイやプリンタなどの入出力機器を接続するための入出力インタフェースを更に備えていてもよい。

[0135] また、プログラムPは、コンピュータCが読み取り可能な、一時的でない有形の記録媒体Mに記録することができる。このような記録媒体Mとしては、例えば、テープ、ディスク、カード、半導体メモリ、又はプログラマブルな論理回路などを用いることができる。コンピュータCは、このような記録媒体Mを介してプログラムPを取得することができる。また、プログラムPは、伝送媒体を介して伝送することができる。このような伝送媒体としては、例えば、通信ネットワーク、又は放送波などを用いることができる。コンピュータCは、このような伝送媒体を介してプログラムPを取得することもできる。

[0136] 〔付記事項1〕

本発明は、上述した実施形態に限定されるものでなく、請求項に示した範囲で種々の変更が可能である。例えば、上述した実施形態に開示された技術的手段を適宜組み合わせ得られる実施形態についても、本発明の技術的範囲に含まれる。

[0137] 〔付記事項2〕

上述した実施形態の一部又は全部は、以下のようにも記載され得る。ただし、本発明は、以下の記載する態様に限定されるものではない。

[0138] （付記1）

自然言語で記述されたクエリを処理する機械学習済みの言語処理モデルに当該クエリと共に入力される当該クエリの関連情報を評価した評価結果を取得する評価結果取得手段と、前記評価結果に基づき、前記言語処理モデルを用いて前記クエリに応答する応答手段と、を備える情報処理装置。

[0139] （付記2）

前記評価結果取得手段は、前記関連情報を所定の評価方法で評価して前記評価結果を取得する、付記1に記載の情報処理装置。

[0140] （付記3）

前記評価結果取得手段は、前記情報処理装置のユーザおよび前記クエリの内容の少なくとも何れかに基づいて前記関連情報を評価する、付記 2 に記載の情報処理装置。

[0141] (付記 4)

前記応答手段は、前記関連情報に含まれる複数の要素のそれぞれに対する前記言語処理モデルの注目度を、前記評価結果に応じて調整する、付記 1 に記載の情報処理装置。

[0142] (付記 5)

前記言語処理モデルは、クエリと当該クエリの関連情報を入力して得られる出力と、与えられた正しい出力との乖離度が低くなるように学習することにより生成されたものであり、

前記応答手段は、評価結果がそれぞれ異なる関連情報を用いて学習された複数の前記言語処理モデルのうち、前記評価結果取得手段が取得した評価結果に応じた評価結果の関連情報を用いて学習された言語処理モデルを用いて前記クエリに応答する、付記 1 に記載の情報処理装置。

[0143] (付記 6)

前記応答手段は、前記関連情報に含まれる複数の要素のうち、前記評価結果取得手段が取得した前記評価結果に応じて選択した要素を前記クエリと共に前記言語処理モデルに入力することにより得られる出力に基づき前記クエリに応答する、付記 1 に記載の情報処理装置。

[0144] (付記 7)

前記言語処理モデルは、クエリと当該クエリの関連情報を入力して得られる出力と、与えられた正しい出力との乖離度が低くなるように学習することにより生成されたものであって、学習に用いた前記関連情報には当該関連情報の評価結果を示す情報が付与されており、前記応答手段は、前記クエリと、前記評価結果取得手段が取得した評価結果を示す情報を付与した、当該クエリの関連情報とを前記言語処理モデルに入力することにより、当該クエリに応答する、付記 1 に記載の情報処理装置。

[0145] (付記 8)

少なくとも 1 つのプロセッサが、自然言語で記述されたクエリを処理する機械学習済みの言語処理モデルに当該クエリと共に入力される当該クエリの関連情報を評価した評価結果を取得することと、前記評価結果に基づき、前記言語処理モデルを用いて前記クエリに応答することと、を含む応答方法。

[0146] (付記 9)

コンピュータを、自然言語で記述されたクエリを処理する機械学習済みの言語処理モデルに当該クエリと共に入力される当該クエリの関連情報を評価した評価結果を取得する評価結果取得手段、および前記評価結果に基づき、前記言語処理モデルを用いて前記クエリに応答する応答手段、として機能させる応答プログラム。

[0147] [付記事項 3]

上述した実施形態の一部又は全部は、更に、以下のように表現することもできる。

[0148] (付記 1)

少なくとも 1 つのプロセッサを備え、前記少なくとも 1 つのプロセッサは、自然言語で記述されたクエリを処理する機械学習済みの言語処理モデルに当該クエリと共に入力される当該クエリの関連情報を評価した評価結果を取得する評価結果取得処理と、前記評価結果に基づき、前記言語処理モデルを用いて前記クエリに応答する応答処理と、を実行する情報処理装置。

[0149] なお、前記情報処理装置は、更にメモリを備えていてもよく、このメモリには、前記評価結果取得処理と、前記応答処理とを前記プロセッサに実行させるための応答プログラムが記憶されていてもよい。また、この応答プログラムは、コンピュータ読み取り可能な一時的でない有形の記録媒体に記録されていてもよい。

[0150] (付記 2)

前記プロセッサは、前記評価結果取得処理において、前記関連情報を所定の評価方法で評価して前記評価結果を取得する、付記 1 に記載の情報処理装

置。

[0151] (付記3)

前記プロセッサは、前記評価結果取得処理においては、前記情報処理装置のユーザおよび前記クエリの内容の少なくとも何れかに基づいて前記関連情報を評価する、付記2に記載の情報処理装置。

[0152] (付記4)

前記プロセッサは、前記応答処理において、前記関連情報に含まれる複数の要素のそれぞれに対する前記言語処理モデルの注目度を、前記評価結果に応じて調整する、付記1に記載の情報処理装置。

[0153] (付記5)

前記言語処理モデルは、クエリと当該クエリの関連情報を入力して得られる出力と、与えられた正しい出力との乖離度が低くなるように学習することにより生成されたものであり、前記プロセッサは、前記応答処理において、評価結果がそれぞれ異なる関連情報を用いて学習された複数の前記言語処理モデルのうち、前記評価結果取得手段が取得した評価結果に応じた評価結果の関連情報を用いて学習された言語処理モデルを用いて前記クエリに応答する、付記1に記載の情報処理装置。

[0154] (付記6)

前記プロセッサは、前記応答処理において、前記関連情報に含まれる複数の要素のうち、前記評価結果取得手段が取得した前記評価結果に応じて選択した要素を前記クエリと共に前記言語処理モデルに入力することにより得られる出力に基づき前記クエリに応答する、付記1に記載の情報処理装置。

[0155] (付記7)

前記言語処理モデルは、クエリと当該クエリの関連情報を入力して得られる出力と、与えられた正しい出力との乖離度が低くなるように学習することにより生成されたものであって、学習に用いた前記関連情報には当該関連情報の評価結果を示す情報が付与されており、前記プロセッサは、前記応答処理において、前記クエリと、前記評価結果取得手段が取得した評価結果を示

す情報を付与した、当該クエリの関連情報とを前記言語処理モデルに入力することにより、当該クエリに応答する、付記 1 に記載の情報処理装置。

符号の説明

[0156]	1	情報処理装置
	1 1	評価結果取得部（評価結果取得手段）
	1 2	応答部（応答手段）
	1 A	情報処理装置
	1 0 3	評価結果取得部（評価結果取得手段）
	1 0 4	応答部（応答手段）
	1 1 1	言語処理モデル

請求の範囲

- [請求項1] 自然言語で記述されたクエリを処理する機械学習済みの言語処理モデルに当該クエリと共に入力される当該クエリの関連情報を評価した評価結果を取得する評価結果取得手段と、
- 前記評価結果に基づき、前記言語処理モデルを用いて前記クエリに応答する応答手段と、を備える情報処理装置。
- [請求項2] 前記評価結果取得手段は、前記関連情報を所定の評価方法で評価して前記評価結果を取得する、請求項1に記載の情報処理装置。
- [請求項3] 前記評価結果取得手段は、前記情報処理装置のユーザおよび前記クエリの内容の少なくとも何れかに基づいて前記関連情報を評価する、請求項2に記載の情報処理装置。
- [請求項4] 前記応答手段は、前記関連情報に含まれる複数の要素のそれぞれに対する前記言語処理モデルの注目度を、前記評価結果に応じて最適化する、請求項1に記載の情報処理装置。
- [請求項5] 前記言語処理モデルは、クエリと当該クエリの関連情報を入力して得られる出力と、与えられた正しい出力との乖離度が低くなるように学習することにより生成されたものであり、
- 前記応答手段は、評価結果がそれぞれ異なる関連情報を用いて学習された複数の前記言語処理モデルのうち、前記評価結果取得手段が取得した評価結果に応じた評価結果の関連情報を用いて学習された言語処理モデルを用いて前記クエリに応答する、請求項1に記載の情報処理装置。
- [請求項6] 前記応答手段は、前記関連情報に含まれる複数の要素のうち、前記評価結果取得手段が取得した前記評価結果に応じて選択した要素を前記クエリと共に前記言語処理モデルに入力することにより得られる出力に基づき前記クエリに応答する、請求項1に記載の情報処理装置。
- [請求項7] 前記言語処理モデルは、クエリと当該クエリの関連情報を入力して得られる出力と、与えられた正しい出力との乖離度が低くなるように

学習することにより生成されたものであって、学習に用いた前記関連情報には当該関連情報の評価結果を示す情報が付与されており、

前記応答手段は、前記クエリと、前記評価結果取得手段が取得した評価結果を示す情報を付与した、当該クエリの関連情報とを前記言語処理モデルに入力することにより、当該クエリに応答する、請求項 1 に記載の情報処理装置。

[請求項8] 少なくとも 1 つのプロセッサが、

自然言語で記述されたクエリを処理する機械学習済みの言語処理モデルに当該クエリと共に入力される当該クエリの関連情報を評価した評価結果を取得することと、

前記評価結果に基づき、前記言語処理モデルを用いて前記クエリに応答することと、を含む応答方法。

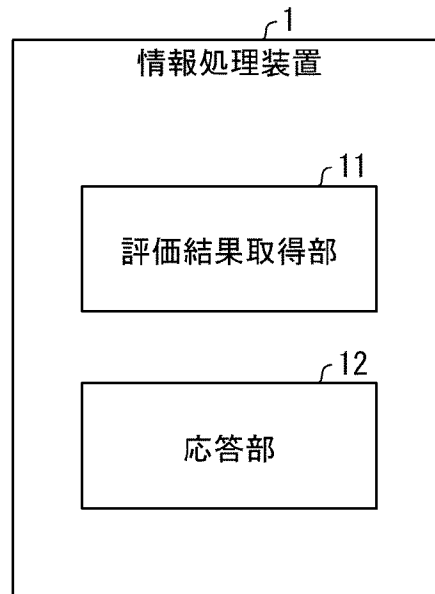
[請求項9] コンピュータを、

自然言語で記述されたクエリを処理する機械学習済みの言語処理モデルに当該クエリと共に入力される当該クエリの関連情報を評価した評価結果を取得する評価結果取得手段、および

前記評価結果に基づき、前記言語処理モデルを用いて前記クエリに応答する応答手段、として機能させる応答プログラム。

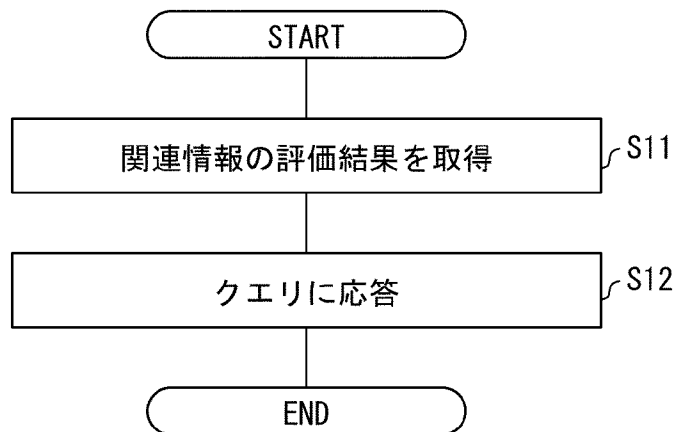
[図1]

図 1



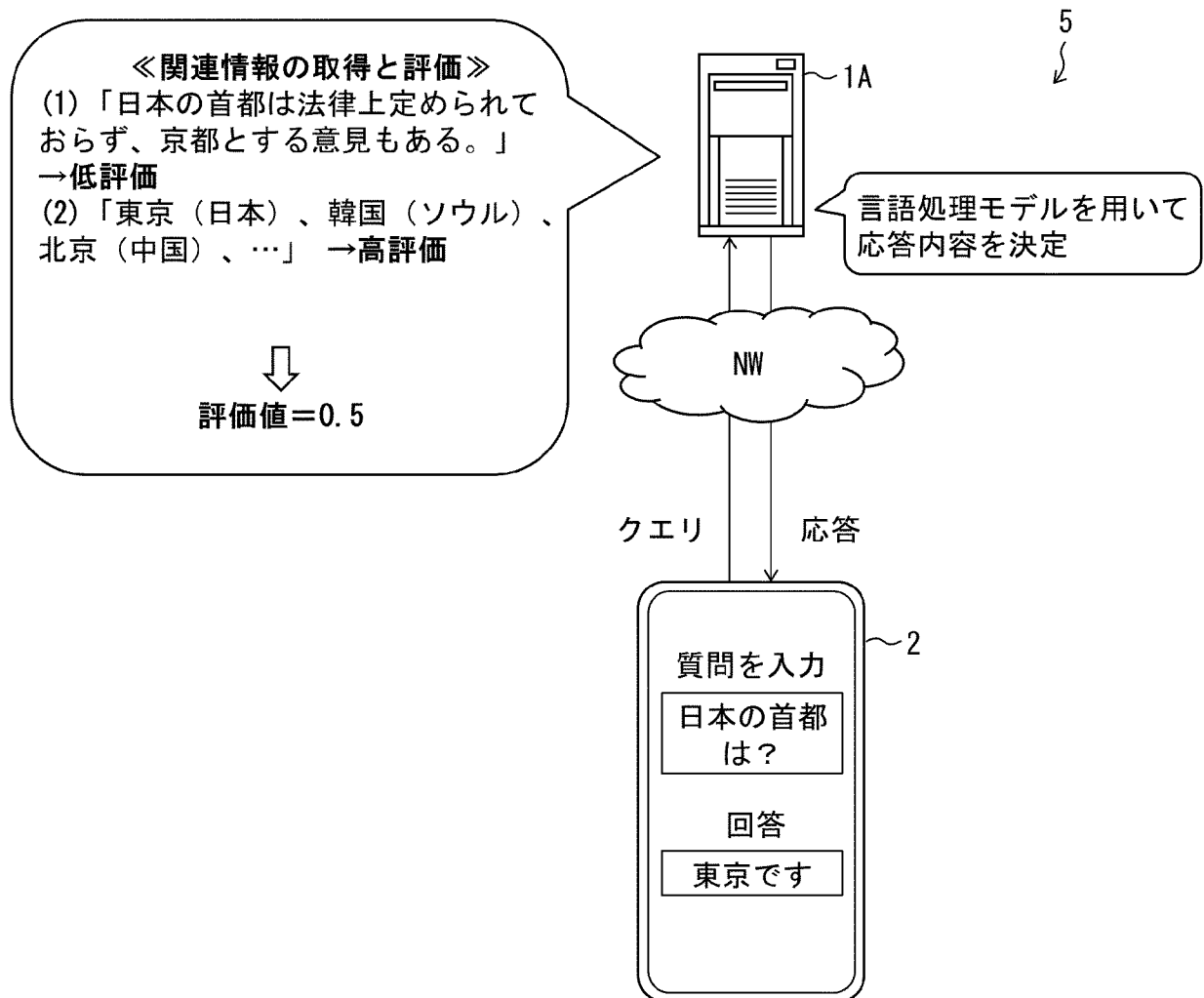
[図2]

図 2



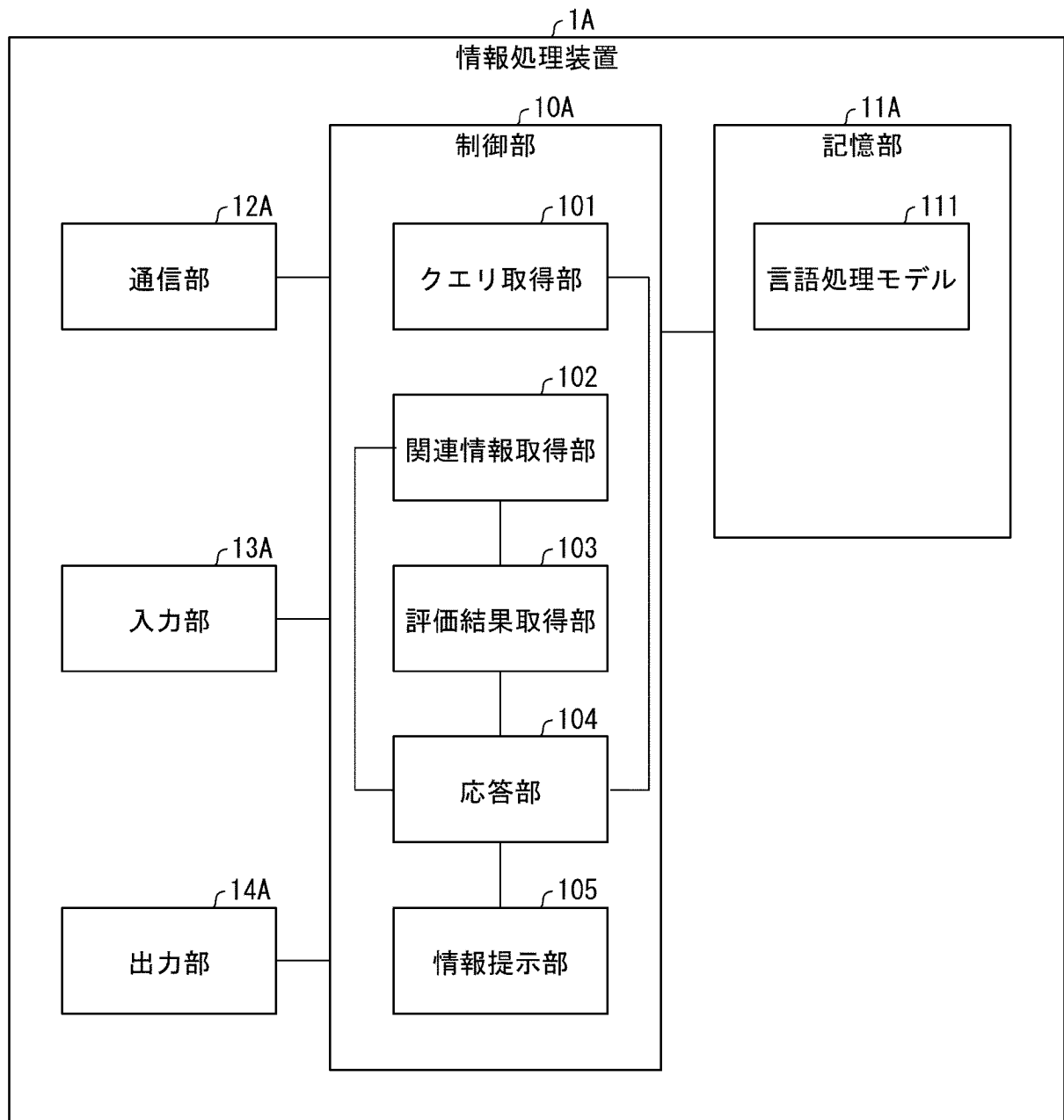
[図3]

図 3



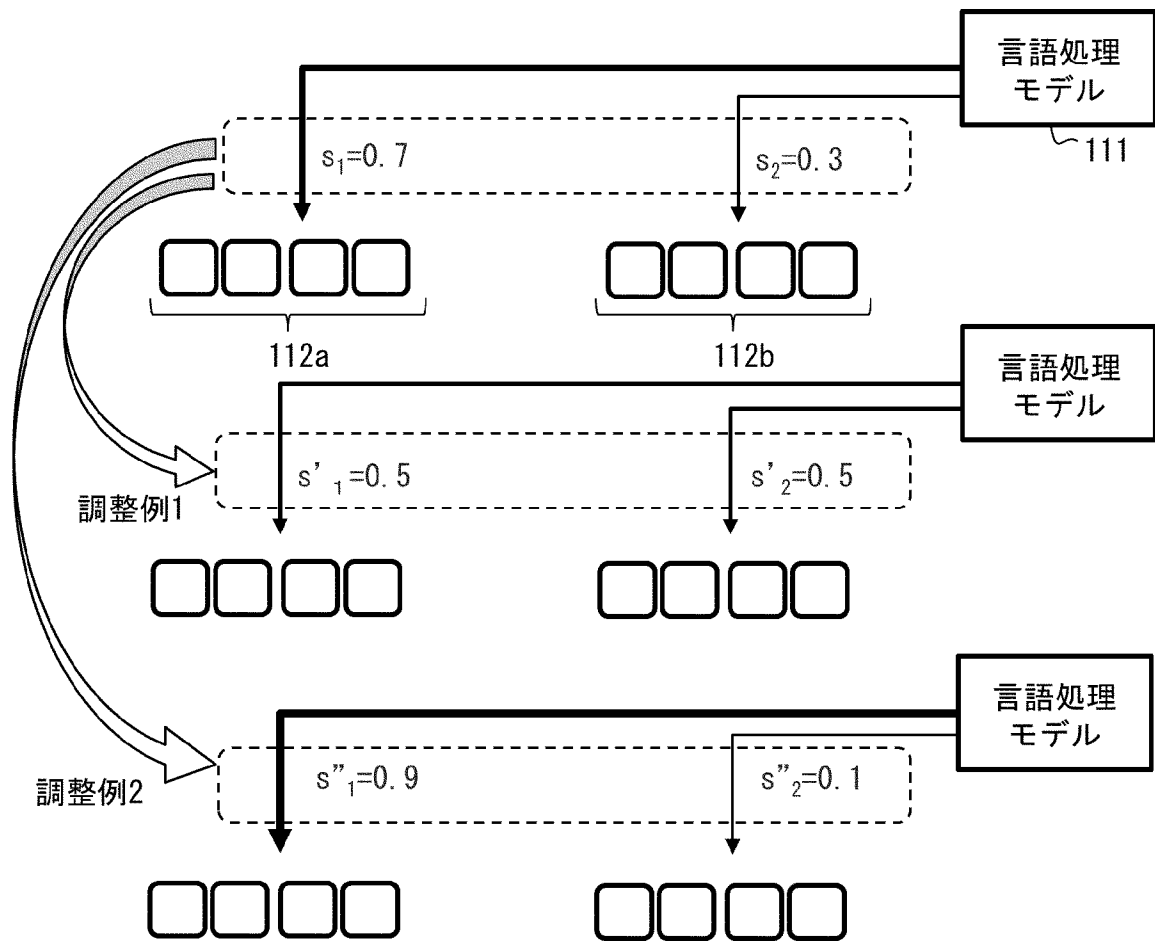
[図4]

図 4



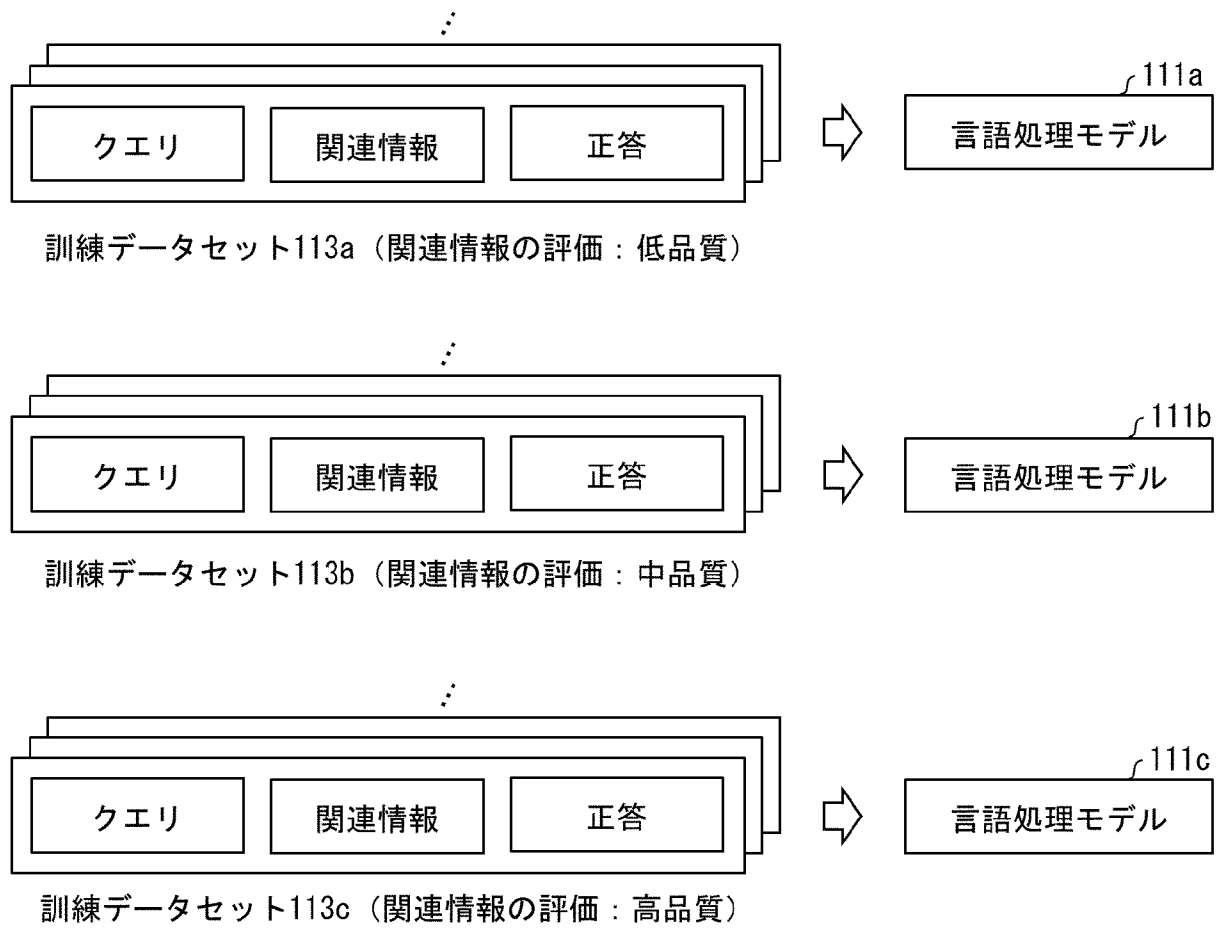
[図5]

図 5



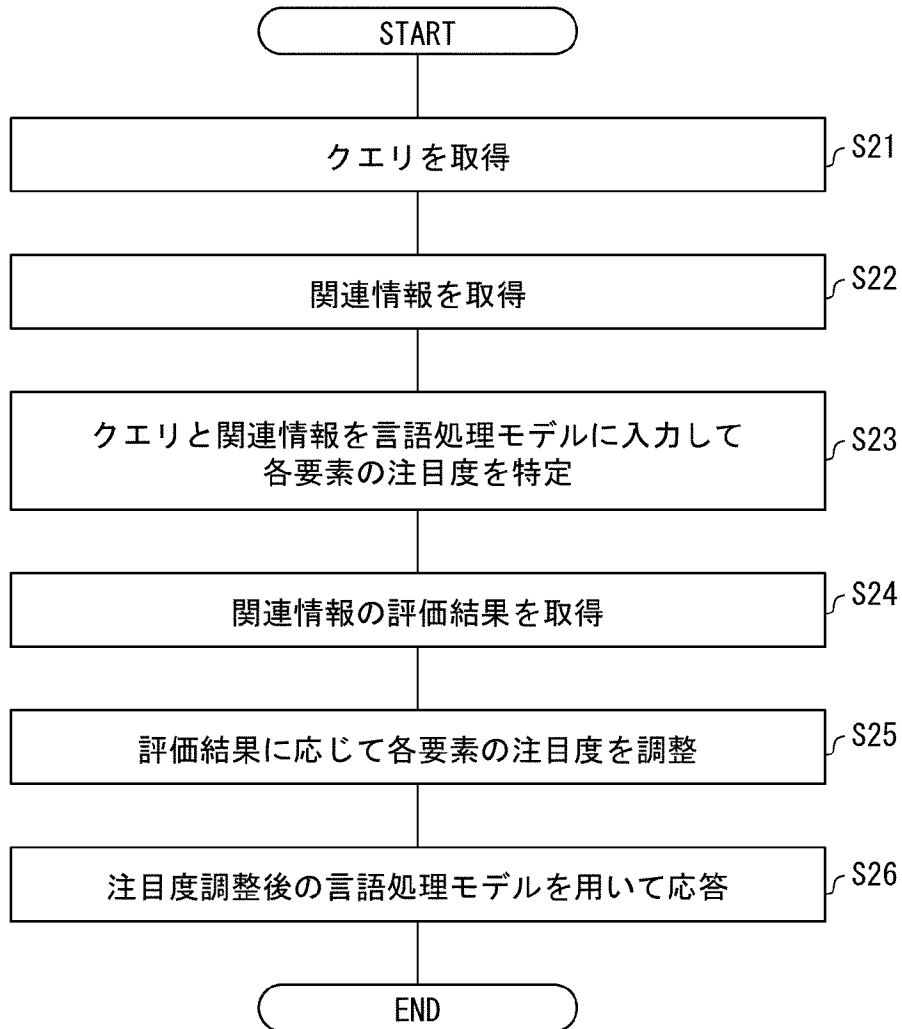
[図6]

図 6



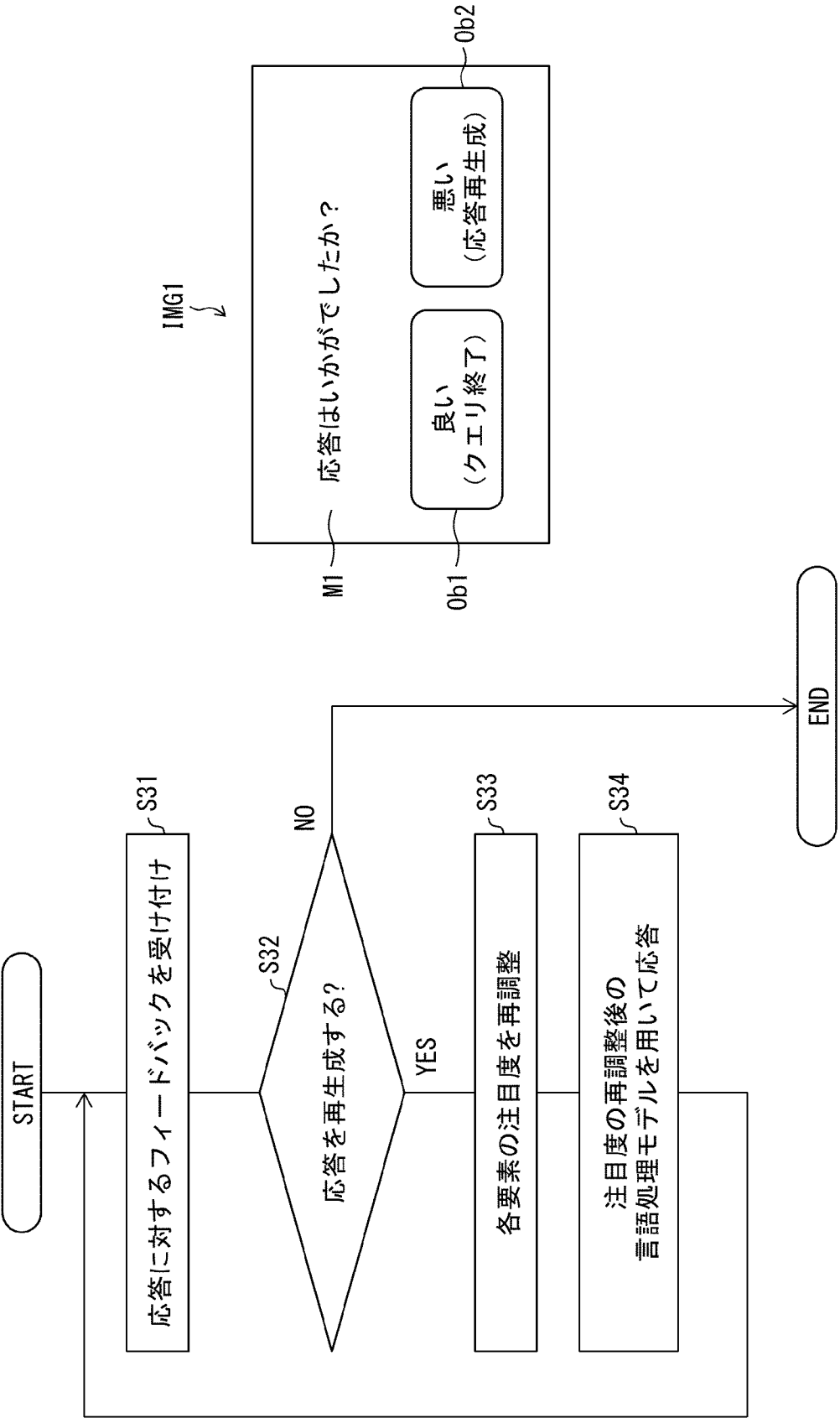
[図7]

図 7



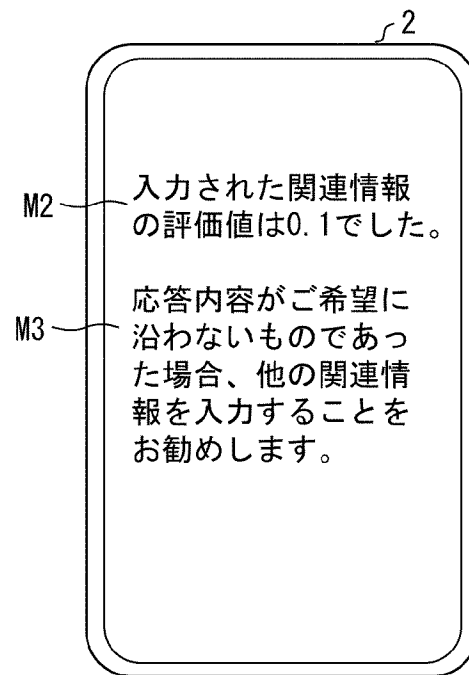
[図8]

図 8



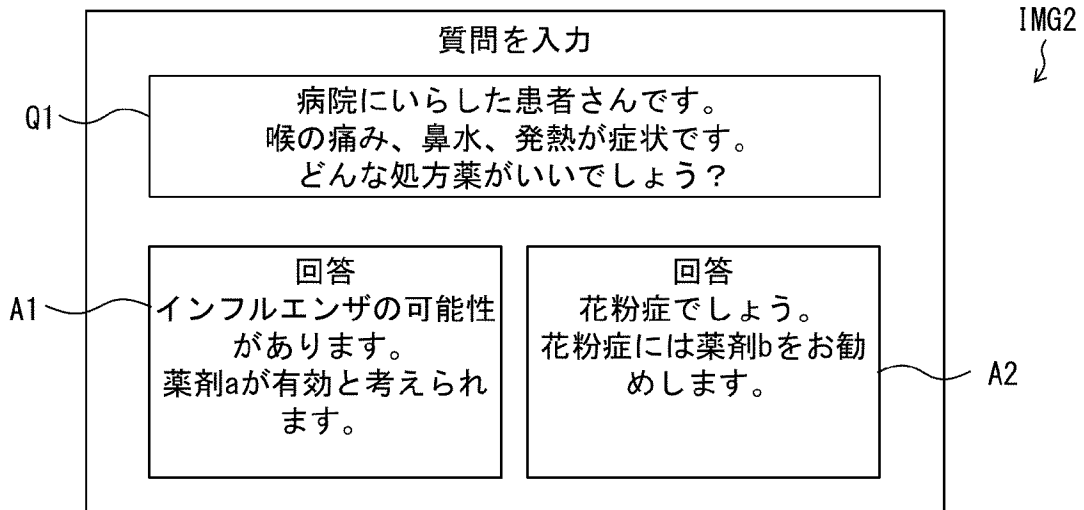
[図9]

図 9



[図10]

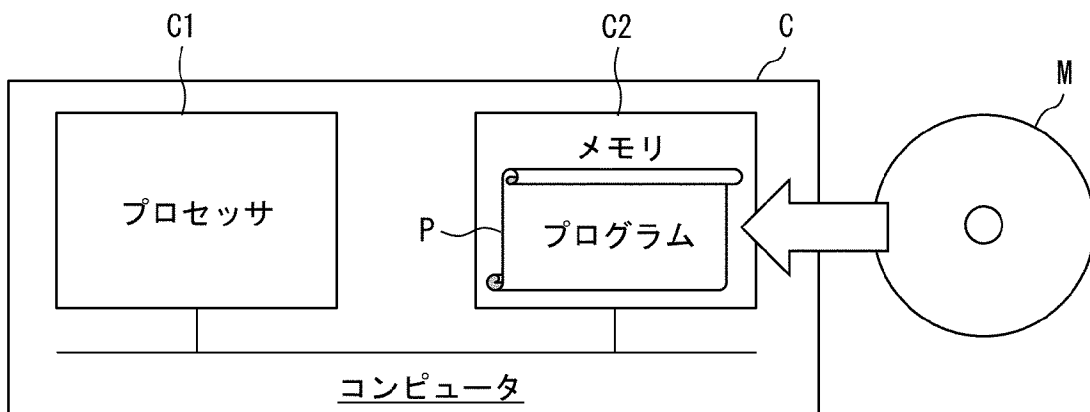
図 10



医療機関	診察日	症状	診断結果	処方薬
A内科	2022/01/01	発熱、喉の痛み	インフルエンザ	薬剤a
B小児科	2022/01/03	咳、鼻水、発熱	インフルエンザ	薬剤a
C耳鼻科	2022/01/05	鼻水	花粉症	薬剤b
...	...	...	...	...

[図11]

図 11



INTERNATIONAL SEARCH REPORT

International application No.

PCT/JP2023/021319

A. CLASSIFICATION OF SUBJECT MATTER**G06F 40/56**(2020.01)i

FI: G06F40/56

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)

G06F40/56, G06F16/90

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 1971-2023
 Registered utility model specifications of Japan 1996-2023
 Published registered utility model applications of Japan 1994-2023

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	水上 雅博 外4名, 話者情報を考慮する注意機構を用いた応答生成手法の検討, 一般社団法人 人工知能学会 第34回全国大会 (2020) [online], 09 June 2020, pp. 1-4, Internet <URL: https://www.jstage.jst.go.jp/article/pjsai/JSAI2020/0/JSAI2020_3O1GS1304/_pdf/-char/ja >, (MIZUKAMI, Masahiro et al. Examination of a response generation method using attention mechanism considering speaker information. The 34th Annual Conference of the Japanese Society for Artificial Intelligence, 2020.) entire text, all drawings	1-9
A	杉山 弘晃 外6名, Transformer encoder-decoderモデルベース雑談システムの学習方法に対する主観評価の変動分析, 一般社団法人 人工知能学会 第35回全国大会 (2021) [online], 08 June 2021, pp. 1-4, Internet <URL: https://www.jstage.jst.go.jp/article/pjsai/JSAI2021/0/JSAI2021_4E1OS11a03/_pdf/-char/ja >, (SUGIYAMA, Hiroaki et al. Analysis of Subjective Evaluation for Fine-tuning Methods of Transformer encoder-decoder based Conversational Systems. The 35th Annual Conference of the Japanese Society for Artificial Intelligence, 2021.) entire text, all drawings	1-9
A	JP 2017-97561 A (CANON KABUSHIKI KAISHA) 01 June 2017 (2017-06-01) particularly, paragraphs [0023]-[0024]	1-9



Further documents are listed in the continuation of Box C.



See patent family annex.

* Special categories of cited documents:

“A” document defining the general state of the art which is not considered to be of particular relevance
 “E” earlier application or patent but published on or after the international filing date
 “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)
 “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

“T” 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
 “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
 “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
 “&” document member of the same patent family

Date of the actual completion of the international search

28 July 2023

Date of mailing of the international search report

08 August 2023

Name and mailing address of the ISA/JP

Japan Patent Office (ISA/JP)
3-4-3 Kasumigaseki, Chiyoda-ku, Tokyo 100-8915
Japan

Authorized officer

Telephone No.

INTERNATIONAL SEARCH REPORT

International application No.

PCT/JP2023/021319

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
E, A	JP 7313757 B1 (SPIRAL.AI INC.) 25 July 2023 (2023-07-25) particularly, paragraphs [0008]-[0011], [0053]-[0054]	1-9

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.
PCT/JP2023/021319

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)	Publication date (day/month/year)
JP	2017-97561	A	01 June 2017	(Family: none)	
JP	7313757	B1	25 July 2023	(Family: none)	

A. 発明の属する分野の分類（国際特許分類（IPC）） G06F 40/56(2020.01)i FI: G06F40/56		
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）） G06F40/56, G06F16/90		
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 1971-2023年 日本国実用新案登録公報 1996-2023年 日本国登録実用新案公報 1994-2023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	水上 雅博 外4名, 話者情報を考慮する注意機構を用いた応答生成手法の検討, 一般社団法人 人工知能学会 第34回全国大会(2020) [online], 2020.06.09, pp.1-4, MIZUKAMI, Masahiro et.al., Internet<URL:https://www.jstage.jst.go.jp/article/pjsai/JSAI2020/0/JSAI2020_301GS1304/_pdf/-char/ja>Examination of a response generation method using attention mechanism considering speaker information 全文、全図	1-9
A	杉山 弘晃 外6名, Transformer encoder-decoderモデルベース雑談システムの学習方法に対する主観評価の変動分析, 一般社団法人 人工知能学会 第35回全国大会(2021) [online], 2021.06.08, pp.1-4, SUGIYAMA, Hiroaki et.al., Internet<URL:https://www.jstage.jst.go.jp/article/pjsai/JSAI2021/0/JSAI2021_4E10S11a03/_pdf/-char/ja>Analysis of Subjective Evaluation for Fine-tuning Methods of Transformer encoder-decoder based Conversational Systems 全文、全図	1-9
A	JP 2017-97561 A (キヤノン株式会社) 01.06.2017 (2017-06-01) 特に、段落0023-0024	1-9
☒ C欄の続きにも文献が列挙されている。 ☒ 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 28.07.2023		国際調査報告の発送日 08.08.2023
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） 長 由紀子 5N 4233 電話番号 03-3581-1101 内線 3586

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
E, A	JP 7313757 B1 (S p i r a l . A I 株式会社) 25.07.2023 (2023 - 07 - 25) 特に、段落0008-0011, 0053-0054	1-9

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/JP2023/021319

引用文献			公表日	パテントファミリー文献	公表日
JP	2017-97561	A	01.06.2017	(ファミリーなし)	
JP	7313757	B1	25.07.2023	(ファミリーなし)	