

(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2月15日(15.02.2024)



(10) 国際公開番号
WO 2024/034232 A1

(51) 国際特許分類:
G06F 16/33 (2019.01) **G06F 40/295** (2020.01)

(21) 国際出願番号: PCT/JP2023/019895

(22) 国際出願日: 2023年5月29日(29.05.2023)

(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 2022-127014 2022年8月9日(09.08.2022) JP

(71) 出願人: コニカミノルタ株式会社 (**KONICA MINOLTA, INC.**) [JP/JP]; 〒1007015 東京都千代田区丸の内二丁目7番2号 Tokyo (JP).

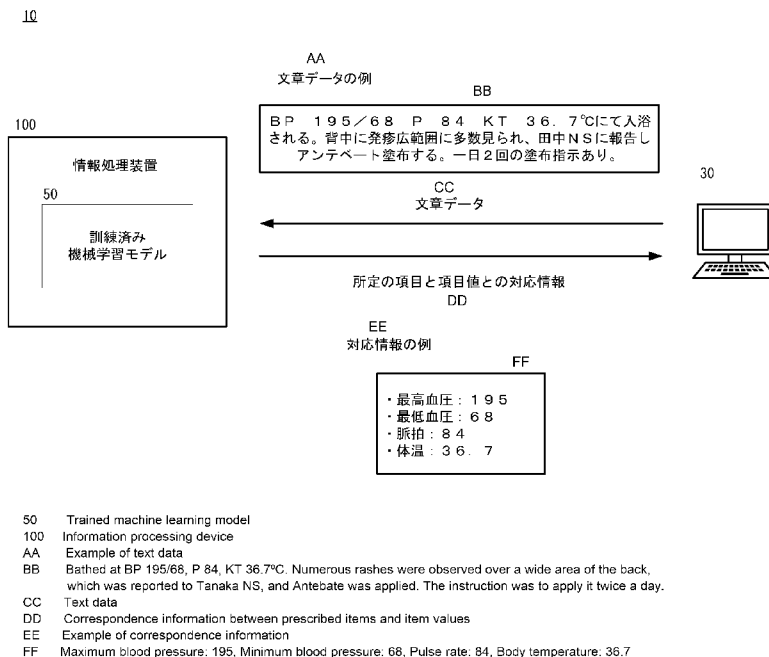
(72) 発明者: 寺中 駿人 (**TERANAKA, Hayato**).

(74) 代理人: 弁理士法人 鷲田国際特許事務所 (**WASHIDA & ASSOCIATES**); 〒1600023 東京都新宿区西新宿1-23-7 新宿ファーストウェスト8階 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, 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,

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

(54) 発明の名称: 情報処理装置、情報処理方法及びプログラム



(57) **Abstract:** Disclosed is a technique for extracting desired information from text data. An embodiment of the present disclosure relates to an information processing device comprising: an input unit that receives text data; an extraction unit that extracts item values corresponding to prescribed items from the received text data using a trained machine learning model that has been trained using training data in which item values in text data are labeled with items corresponding to the item values; and an output unit that associates and outputs the prescribed items and the extracted item values.

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, 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))

(57) 要約: 文章データから所望の情報を抽出するための技術が開示される。本開示の一態様は、文章データを受け付ける入力部と、文章データの中の項目値に対して、項目値に対応する項目をラベル付けした訓練データを用いて訓練された訓練済み機械学習モデルによって、所定の項目に対応する項目値を前記受け付けた文章データから抽出する抽出部と、前記項目と前記項目値とを関連付けて出力する出力部と、を有する、情報処理装置に関する。

明 細 書

発明の名称： 情報処理装置、情報処理方法及びプログラム

技術分野

[0001] 本開示は、情報処理装置、情報処理方法及びプログラムに関する。

背景技術

[0002] 情報技術の進展によって、様々なデータが、紙媒体の代わりにデジタル情報として記憶及び処理される機会が増大している。しかしながら、ユーザが常時パーソナルコンピュータなどにアクセス可能であるとは限らない。例えば、医療現場、製造現場などでは、測定結果、検査結果などのデータは、医療スタッフ、作業員などによって手元のメモに一旦書き留められ、書き留められたメモに基づいて以降にコンピュータ端末に入力されうる。例えば、医療スタッフ、作業員などは、作業終了後などにコンピュータ端末を操作し、手元のメモに記録された内容をデータベースに転記している。

先行技術文献

特許文献

[0003] 特許文献1：特開2019-8656号公報

特許文献2：特開2020-52902号公報

発明の概要

発明が解決しようとする課題

[0004] このようにして医療スタッフ、作業員などによってメモに書き留められた内容は、典型的には、短時間で書き留められる必要がある。このため、完全な文章形式というより、略語、短縮文などの不完全な文章形式で記載されるケースが多い。例えば、「最高血圧は120であり、最低血圧は80である。」という内容が、「BP 120/80」などのような略語と短縮文とを含む不完全な文章として記載されうる。入力される文章がこのような不完全な形式によって記載されている場合であっても、項目「最高血圧」及び「最低血圧」がそれぞれ“120”及び“80”の項目値を有することが検出さ

れ、「最高血圧：１２０」及び「最低血圧：８０」がデータベースに登録される必要がある。

[0005] 本開示の１つの課題は、文章データから所望の情報を抽出するための技術を提供することである。

課題を解決するための手段

[0006] 本開示の一態様は、文章データを受け付ける入力部と、文章データの中の項目値に対して、項目値に対応する項目をラベル付けした訓練データを用いて訓練された訓練済み機械学習モデルによって、所定の項目に対応する項目値を前記受け付けた文章データから抽出する抽出部と、前記項目と前記項目値とを関連付けて出力する出力部と、を有する、情報処理装置に関する。

発明の効果

[0007] 本開示によると、文章データから所望の情報を抽出することができる。

図面の簡単な説明

[0008] [図1]図１は、本開示の一実施例による情報処理装置を示す概略図である。

[図2]図２は、本開示の一実施例による機械学習モデルの訓練処理を示す概略図である。

[図3]図３は、本開示の一実施例による情報処理装置のハードウェア構成を示すブロック図である。

[図4]図４は、本開示の一実施例による情報処理装置の機能構成を示すブロック図である。

[図5]図５Ａ及び５Ｂは、本開示の一実施例による項目値抽出処理を示す図である。

[図6]図６Ａ及び６Ｂは、本開示の一実施例による項目値抽出処理を示す図である。

[図7]図７Ａ及び７Ｂは、本開示の一実施例による項目値抽出処理を示す図である。

[図8]図８Ａ及び８Ｂは、本開示の一実施例による項目値抽出処理を示す図である。

[図9]図9A～9Cは、本開示の一実施例による項目値抽出処理を示す図である。

[図10]図10A～10Cは、本開示の一実施例による項目値抽出処理を示す図である。

[図11]図11A及び11Bは、本開示の一実施例による表示処理を示す図である。

[図12]図12は、本開示の一実施例による項目抽出処理を示すフローチャートである。

発明を実施するための形態

[0009] 以下、図面を参照して本開示の実施の形態を説明する。

[0010] 以下の実施例では、機械学習モデルを利用して、略語、短縮文などを含む不完全な文章データから所定の項目に対応する項目値を抽出する情報処理装置が開示される。

[0011] [概略]

後述される実施例による情報処理装置は、機械学習モデルを利用して、略語、短縮文などの不完全な表記を含む文章データから、所定の項目に対応する項目値を抽出し、各項目と抽出した項目値とを関連付けて出力する。例えば、機械学習モデルとして、固有表現抽出（Named Entity Recognition）モデルが利用されてもよく、固有表現抽出モデルは、BERT（Bidirectional Encoder Representations from Transformers）に基づいて構成されてもよい。

[0012] [情報処理システム]

まず、図1を参照して、本開示の一実施例による情報処理システム10を説明する。図1は、本開示の一実施例による情報処理システム10を示す概略図である。図1に示されるように、情報処理システム10は、端末30及び情報処理装置100を有する。端末30は、通信ネットワーク（図示せず）を介し情報処理装置100にアクセス可能である。

[0013] 本実施例による情報処理システム 10 では、ユーザが端末 30 に文章データを入力すると、端末 30 は、取得した文章データを情報処理装置 100 に送信する。例えば、ユーザによって入力される文章データは、「BP 195/68 P 84 KT 36.7℃にて入浴される。背中に発疹広範囲に多数見られ、田中NSに報告しアンテベート塗布する。一日2回の塗布指示あり。」などであってもよい。ここで、ユーザは、何れか適切な入力方法によって文章データを端末 30 に入力してもよい。例えば、キーボードを操作することによって文章データを入力してもよい。あるいは、ユーザは、現場で書き留めたメモを撮像することによって画像データを生成し、画像データに対して文字認識を実行することによって文章データを入力してもよい。あるいは、ユーザは、文章データの内容を読み上げることによって音声データを生成し、音声データに対して音声認識を実行することによって文章データを入力してもよい。

[0014] 情報処理装置 100 は、訓練済み機械学習モデル 50 を利用して、端末 30 から取得した文章データから 1 つ以上の所定の項目と各項目の項目値とを抽出し、各項目と項目値とを関連付けて対応情報として端末 30 に返す。例えば、上述した文章データに対する対応情報は、「最高血圧：195」、「最低血圧：68」、「脈拍：84」及び「体温：36.7」を示すものであってもよい。すなわち、当該機械学習モデルは、文章データのデータ部分「BP 195/68」から「最高血圧：195」及び「最低血圧：68」を抽出し、データ部分「P 84」から「脈拍：84」を抽出し、データ部分「KT 36.7℃」から「体温：36.7」を抽出する。換言すると、本開示による機械学習モデル 50 は、このような略語及び短縮形などの不完全な表記によって抽出対象のデータ項目が記載されている場合であっても、データ項目と対応する項目値との間の関連付けを推定するよう訓練されている。

[0015] 情報処理装置 100 から対応情報を取得すると、端末 30 は、取得した対応情報を画面上に表示し、ユーザによる確認を得た後、データベース（図示

せず)に登録してもよい。なお、図示された実施例では、情報処理装置100は、訓練済み機械学習モデル50を保持しているが、これに限定されることはない。例えば、訓練済み機械学習モデル50は、情報処理装置100から通信アクセス可能な外部装置、クラウドなどに格納されてもよい。

[0016] [機械学習モデル]

以下の実施例による機械学習モデルは、文章データから固有表現により記載された所定の項目に対応する項目値を抽出するよう訓練されている。特に、本実施例による機械学習モデルは、所定の項目が略語によって記載されたり、あるいは、文章内容が完全な形式でなく短縮されている不完全な形式の文章データから、所定の項目と対応する項目値とを関連付けて出力する。

[0017] 図2は、本開示の一実施例による機械学習モデル50の訓練処理を示す概略図である。図2に示されるように、例えば、機械学習モデル50は、BERTなど、インターネット上などで入手可能な何れか適切な(双方向性の)固有表現抽出モデルに基づいて構築されてもよい。入手された何れかの固有表現抽出モデルは、日本語Wikipediaなどの大規模な日本語データによって事前学習され、日本語に適応するよう訓練されてもよい。

[0018] このように事前学習された固有表現抽出モデルは、次に、処理対象の文章データの適用分野に適応するようファインチューニングされてもよい。例えば、情報処理装置100が、医用文章に対して適用される場合、固有表現抽出モデルは、医用データを訓練データとして使用してファインチューニングされてもよい。また、情報処理装置100が、介護文章に対して適用される場合、固有表現抽出モデルは、介護用データを訓練データとして使用してファインチューニングされてもよい。

[0019] 例えば、医用データ及び／又は介護用データとして、「血圧」が「BP」として略語表記されることがあること、短縮形「BP XX／YY」が「最高血圧：XX」及び「最低血圧：YY」であることなどが訓練データとして提供され、固有表現抽出モデルを訓練するのに使用されてもよい。これにより、訓練済み機械学習モデル50は、各適用分野における略語及び短縮形を

適切に認識することが可能になる。

[0020] ここで、情報処理装置100は、サーバ、パーソナルコンピュータ（PC）、スマートフォン、タブレット等の計算装置によって実現されてもよく、例えば、図3に示されるようなハードウェア構成を有してもよい。すなわち、情報処理装置100のそれぞれは、バスBを介し相互接続されるドライブ装置101、ストレージ装置102、メモリ装置103、プロセッサ104、ユーザインタフェース（UI）装置105及び通信装置106を有する。

[0021] 情報処理装置100における後述される各種機能及び処理を実現するプログラム又は指示は、CD-ROM（Compact Disk-Read Only Memory）、フラッシュメモリ等の着脱可能な記憶媒体に格納されてもよい。当該記憶媒体がドライブ装置101にセットされると、プログラム又は指示が記憶媒体からドライブ装置101を介しストレージ装置102又はメモリ装置103にインストールされる。ただし、プログラム又は指示は、必ずしも記憶媒体からインストールされる必要はなく、ネットワークなどを介し何れかの外部装置からダウンロードされてもよい。

[0022] ストレージ装置102は、ハードディスクドライブなどによって実現され、インストールされたプログラム又は指示と共に、プログラム又は指示の実行に用いられるファイル、データ等を格納する。

[0023] メモリ装置103は、ランダムアクセスメモリ、スタティックメモリ等によって実現され、プログラム又は指示が起動されると、ストレージ装置102からプログラム又は指示、データ等を読み出して格納する。ストレージ装置102、メモリ装置103及び着脱可能な記憶媒体は、非一時的な記憶媒体（non-transitory storage medium）として総称されてもよい。

[0024] プロセッサ104は、1つ以上のプロセッサコアから構成されうる1つ以上のCPU（Central Processing Unit）、GPU（Graphics Processing Unit）、処理回路（processing circuitry）等によって実現されてもよく、メ

メモリ装置 103 に格納されたプログラム、指示、当該プログラム若しくは指示を実行するのに必要なパラメータなどのデータ等に従って、後述される情報処理装置 100 の各種機能及び処理を実行する。

[0025] ユーザインタフェース (UI) 装置 105 は、キーボード、マウス、カメラ、マイクロフォン等の入力装置、ディスプレイ、スピーカ、ヘッドセット、プリンタ等の出力装置、タッチパネル等の入出力装置から構成されてもよく、ユーザと情報処理装置 100 との間のインタフェースを実現する。例えば、ユーザは、ディスプレイ又はタッチパネルに表示された GUI (Graphical User Interface) をキーボード、マウス等を操作し、情報処理装置 100 を操作する。

[0026] 通信装置 106 は、外部装置、インターネット、LAN (Local Area Network)、セルラーネットワーク等の通信ネットワークとの有線及び／又は無線通信処理を実行する各種通信回路により実現される。

[0027] しかしながら、上述したハードウェア構成は単なる一例であり、本開示による情報処理装置 100 は、他の何れか適切なハードウェア構成により実現されてもよい。

[0028] [情報処理装置]

次に、図 4 から図 11 を参照して、本開示の一実施例による情報処理装置 100 を説明する。本実施例による情報処理装置 100 は、訓練済み機械学習モデル 50 を利用して、略語、短縮形などの不完全な表記を含む不完全な文章データから、所定の項目と対応する項目値とを抽出する。

[0029] 図 4 は、本開示の一実施例による情報処理装置 100 の機能構成を示すブロック図である。図 4 に示されるように、情報処理装置 100 は、入力部 110、抽出部 120 及び出力部 130 を有する。また、抽出部 120 は、判定部 121 を有してもよい。例えば、入力部 110、抽出部 120、判定部 121 及び出力部 130 の 1 つ以上の機能部は、1 つ以上のプロセッサ 104 が 1 つ以上のプログラム又は指示を実行することによって実現されてもよい。

[0030] 入力部 110 は、文章データを受け付ける。具体的には、入力部 110 は、端末 30 を介しユーザから処理対象の文章データを受け付ける。例えば、本実施例による文章データは、完全な形式の文章（例えば、「最高血圧は 195 であり、最低血圧は 68 である。」）に加えて又は代わりに、略語、短縮形の表記などの不完全な表記を含む不完全な形式（例えば、「BP 195 / 68」）の文章として記載されたものであってもよい。入力部 110 は、受け付けた文章データを抽出部 120 にわたす。なお、略語及び短縮形は、不完全な表記の一例に過ぎず、本開示による不完全な表記は、完全な表記以外の他の任意の表記を含んでもよい。

[0031] 抽出部 120 は、訓練済み機械学習モデル 50 によって、文章データから所定の項目に対応する項目値を抽出する。具体的には、入力部 110 から文章データを取得すると、抽出部 120 は、取得した文章データを訓練済み機械学習モデル 50 に入力し、文章データにおける所定の項目に対応する項目値を抽出する。ここで、訓練済み機械学習モデル 50 は、文章データの中の項目値に対して、項目値に対応する項目をラベル付けした訓練データを用いて訓練されている。例えば、入力部 110 が、図 5A に示されるように、「今朝から悪寒あり、120 / 80 P80 KT38.0」の文章データを受け付けた場合、抽出部 120 は、受け付けた当該文章データを訓練済み機械学習モデル 50 に入力し、図 5B に示されるように、所定の項目「最高血圧」、「最低血圧」、「脈拍」及び「体温」と、それぞれに対応する項目値「120」、「80」、「80」及び「38.0」とを抽出する。

[0032] 医用データ又は介護用データを使用して訓練された場合、訓練済み機械学習モデル 50 は、「120 / 80」を「最高血圧：120 及び最低血圧：80」として抽出しうる。同様に、訓練済み機械学習モデル 50 は、「P80」及び「KT38.0」をそれぞれ「脈拍：80」及び「体温：38.0℃」として抽出しうる。すなわち、「最高血圧が 120 であり、最低血圧が 80 である」という完全な形式の表記は、文章データになく、「120 / 80」という短縮形の表記とされている。また、「脈拍が 80 である」及び「体

温が38.0℃である」という完全な形式の表記も、文章データになく、「P80」及び「KT38.0」という略語の表記とされている。機械学習モデル50は、このような短縮形、略語などの不完全な形式の表記の訓練データを使用してファインチューニングされており、不完全な形式の文章データから所定の項目に対応する項目値を抽出することができる。

[0033] 出力部130は、項目と項目値とを関連付けて出力する。具体的には、抽出部120によって文章データから所定の項目に対応する項目値が抽出されると、出力部130は、抽出された各項目と対応する項目値とのペアを示す対応情報を出力する。

[0034] 一例として、抽出部120が、図6Aに示されるように、文章データ「入浴前準備、BP=110/70 P=65、体温36.5」から、「最高血圧：110」、「最低血圧：70」、「脈拍：65」及び「体温：36.5」を抽出したとする。このとき、出力部130は、図6Bに示されるように、「最高血圧：110 [mmHg]」、「最低血圧：70 [mmHg]」、「脈拍：65 [bpm]」及び「体温：36.5 [℃]」の所定の項目とその項目値とのペアから構成される対応情報を出力してもよい。

[0035] また、一例として、抽出部120が、図7Aに示されるように、文章データ「KT35.3 P68 Bp118/83 しっかり覚醒されているが、活気なく目もうつろ」から、「最高血圧：118」、「最低血圧：83」、「脈拍：68」及び「体温：35.3」を抽出したとする。このとき、出力部130は、図7Bに示されるように、「最高血圧：118 [mmHg]」、「最低血圧：83 [mmHg]」、「脈拍：68 [bpm]」及び「体温：35.3 [℃]」の所定の項目とその項目値とのペアから構成される対応情報を出力してもよい。図6A及び6Bに示される文章データでは、「最高血圧」、「最低血圧」、「脈拍」及び「体温」の4つのデータ項目が当該順序で含まれているが、図7A及び7Bに示される文章データでは、「最高血圧」、「最低血圧」、「脈拍」及び「体温」の4つのデータ項目が、「体温」、「脈拍」、「最高血圧」及び「最低血圧」の順序で含まれている。こ

の場合であっても、訓練済み機械学習モデル50は、各項目と対応する項目値とを適切に抽出することが可能であり、出力部130は、抽出された各項目と項目値とのペアに従って対応情報を出力することができる。

[0036] なお、出力部130が出力する対応情報は、図示された形式に限定されず、項目と項目値との関連付けを示す他の何れかの形式を有してもよい。例えば、対応情報は、ユーザに表示及び確認された後、データベースに登録されるが、ユーザが確認しやすいような形式で作成されてもよい。また、出力部130は、経時的に取得された対応情報を時系列表示してもよい。

[0037] また、一例として、抽出部120は、図8Aに示されるように、文章データ「KT37.5 体感熱あり、クーリング開始する。」から、「最高血圧：なし」、「最低血圧：なし」、「脈拍：なし」及び「体温：37.5」を抽出してもよい。すなわち、訓練済み機械学習モデル50は、所定の項目のうち「体温」とこれに対応する項目値「37.5」とのペアのみが文章データに含まれており、項目「最高血圧」、「最低血圧」及び「脈拍」とこれらに対応する項目値とは含まれていないと判定してもよい。このとき、出力部130は、図8Bに示されるように、「最高血圧：なし」、「最低血圧：なし」、「脈拍：なし」及び「体温：37.5 [°C]」の所定の項目とその項目値とから構成される対応情報を出力してもよい。

[0038] 一実施例では、抽出部120は、文章データにおける所定の項目に対応する項目値の適否を判定する判定部121を有してもよい。一例として、抽出部120は、図9Aに示されるように、文章データ「KT36.3 P612 BP108/73 体感熱なく、顔色良好」から、「最高血圧：108」、「最低血圧：73」、「脈拍：612」及び「体温：36.3」を抽出してもよい。判定部121は、各項目に対して項目値のとりうる範囲を示す範囲情報を予め保持し、抽出した項目値が当該範囲情報のとりうる範囲外にある場合、検出エラーであると判定してもよい。検出エラーであると判定された項目について、出力部130は、項目値がエラーである旨を示すメッセージを付記してもよい。例えば、図9Aに示された例において、判定部12

1 は、予め保持する範囲情報に基づいて項目「脈拍」の項目値「612」が範囲外にあると判定し、出力部130は、図9Bに示されるように、「最高血圧：108 [mmHg]」、「最低血圧：73 [mmHg]」、「脈拍：612 [bpm] エラー」及び「体温：36.3 [℃]」の所定の項目とその項目値とから構成される対応情報を出力してもよい。

[0039] また、文章データにおいて所定の項目に対応する項目値が適切でないと判定されると、出力部130は、当該項目値を修正するよう促してもよい。具体的には、抽出部120によって検出された所定の項目の項目値が範囲情報の範囲外にある場合、判定部121は、文章データに転記ミスがあった可能性があると判断し、出力部130は、端末30を介しユーザに不適切な項目値を提示し、正しい項目値を入力するよう促してもよい。具体的には、出力部130は、不適切な項目値をユーザに修正させるためのメッセージ、画面等を端末30に送信してもよい。あるいは、出力部130は、不適切な項目値を強調表示（例えば、色付け、フォント変更等）し、ユーザが認識しやすいようにしてもよい。なお、本開示による入力の促しは、これらに限定されず、不適切な項目値をユーザに認識及び修正させるための何れの形態により実現されてもよい。この場合、ユーザは、図9Cに示されるように、対応情報における不適切な項目値を修正するようにしてもよいし、あるいは、端末30に修正された文章データを再入力してもよい。

[0040] また、一実施例では、抽出部120は、所定の項目に対応する項目値が文章データにあるか否かを判定する判定部121を有してもよい。一例として、抽出部120は、図10Aに示されるように、文章データ「KT36.2 P61 BP110/70 体感熱なし。様子見る。倦怠感なし。」から、「最高血圧：110」、「最低血圧：70」、「脈拍：なし」及び「体温：36.2」を抽出してもよい。すなわち、判定部121は、所定の項目のうち「最高血圧」、「最低血圧」及び「体温」と対応する項目値「110」、「70」及び「37.5」とのペアが文章データに含まれており、項目「脈拍」と対応する項目値とは含まれていないと判定してもよい。このとき、出

力部130は、図10Bに示されるように、「最高血圧：110 [mmHg]」、「最低血圧：70 [mmHg]」、「脈拍：なし」及び「体温：36.2 [℃]」の所定の項目とその項目値とから構成される対応情報を入力してもよい。

[0041] また、所定の項目に対応する項目値が文章データにないと判定されると、出力部130は、当該項目値を入力するよう促してもよい。具体的には、抽出部120が所定の項目の1つ以上について対応する項目値が文章データにおいて検出できなかった場合、判定部121は、文章データに転記漏れがあった可能性があるとして判断し、出力部130は、端末30を介しユーザに未検出の項目を提示し、対応する項目値を入力するよう促してもよい。具体的には、出力部130は、未検出の項目値をユーザに修正させるためのメッセージ、画面等を端末30に送信してもよい。あるいは、出力部130は、未検出の項目を強調表示（例えば、色付け、フォント変更等）し、ユーザが認識しやすいようにしてもよい。なお、本開示による入力の促しは、これらに限定されず、未検出の項目をユーザに認識及び修正させるための何れの形態により実現されてもよい。この場合、ユーザは、図10Cに示されるように、対応情報における未検出の項目「脈拍」に項目値を追記するようにしてもよいし、あるいは、端末30に追記された文章データを再入力してもよい。ここでは、項目「脈拍」に項目値「61」が入力されている。

[0042] また、一実施例では、出力部130は、対応情報を時系列表示してもよい。具体的には、出力部130は、経時的に取得した対応情報におけるデータ項目の項目値を時系列表示してもよい。例えば、出力部130は、図11Aに示されるように、項目「最高血圧」、「最低血圧」、「脈拍」及び「体温」の各時点での項目値を端末30上にテーブル形式で時系列表示してもよい。また、出力部130は、図11Bに示されるように、項目「最高血圧」、「最低血圧」、「脈拍」及び「体温」の各時点での項目値を端末30上にグラフ形式で時系列表示してもよい。

[0043] 本実施例によると、機械学習モデルを利用して、略語、短縮形などの不完

全な表記を含む不完全な形式の文章データから、所定の項目と項目値とを抽出し、項目と項目値とのペアをデータベースに登録することができる。また、本実施例による機械学習モデルは、インターネット上で入手可能な公開されたBERTなどの固有表現抽出モデルを事前学習及びファインチューニングすることによって、取得できる。

[0044] [項目値抽出処理]

次に、図12を参照して、本開示の一実施例による項目値抽出処理を説明する。当該項目値抽出処理は、上述した情報処理装置100によって実行され、より詳細には、情報処理装置100の1つ以上のプロセッサ104が1つ以上のメモリ装置103に格納された1つ以上のプログラム又は指示を実行することによって実現されてもよい。

[0045] 本実施例による情報処理装置100は、文章データから所定の項目とその項目値とのペアを抽出するよう訓練された機械学習モデル50を利用してもよい。また、機械学習モデル50は、適用分野において使用される略語及び／又は短縮形などの不完全な表記を含む不完全な形式の文章データと、当該不完全な形式の文章データにおける所定の項目に対応する項目値を訓練データとして使用して訓練されてもよい。ここで、文章データは、医用文章であってもよく、機械学習モデル50は、医用文章から所定の医用項目に対応する数値を抽出してもよい。また、文章データは、介護記録文章であってもよく、機械学習モデル50は、介護記録文章から所定の介護項目に対応する数値を抽出してもよい。

[0046] 図12は、本開示の一実施例による項目値抽出処理を示すフローチャートである。図12に示されるように、ステップS101において、情報処理装置100は、文章データを受け付ける。具体的には、情報処理装置100は、端末30を介しユーザから処理対象の文章データを受け付ける。例えば、本実施例による文章データは、完全な形式の文章に加えて又は代わりに、略語、短縮形の表記などの不完全な表記を含む不完全な形式の文章として記載されたものであってもよい。

- [0047] ステップS 1 0 2において、情報処理装置 1 0 0は、訓練済み機械学習モデル 5 0によって、文章データから所定の項目に対応する項目値を抽出する。具体的には、情報処理装置 1 0 0は、取得した文章データを訓練済み機械学習モデル 5 0に入力し、文章データにおける所定のデータ項目に対応する項目値を抽出する。
- [0048] ステップS 1 0 3において、情報処理装置 1 0 0は、各項目と項目値とを関連付けて出力してもよい。具体的には、情報処理装置 1 0 0は、文章データから所定の項目に対応する項目値が抽出されると、各項目と対応する項目値とのペアを示す対応情報を出力してもよい。
- [0049] 本実施例によると、機械学習モデルを利用して、略語、短縮形などの不完全な表記を含む不完全な形式の文章データから、所定の項目に対応する項目値を抽出し、項目と項目値とのペアをデータベースに登録することができる。また、本実施例による機械学習モデルは、インターネット上で入手可能な公開されたB E R Tなどの固有表現抽出モデルを事前学習及びファインチューニングすることによって、比較的容易に取得できる。
- [0050] なお、以上の説明に関して更に以下の付記を開示する。
- (付記 1)
- 文章データを受け付ける入力部と、
- 文章データの中の項目値に対して、項目値に対応する項目をラベル付けした訓練データを用いて訓練された訓練済み機械学習モデルによって、所定の項目に対応する項目値を前記文章データから抽出する抽出部と、
- 前記項目と前記項目値とを関連付けて出力する出力部と、
- を有する、情報処理装置。
- (付記 2)
- 前記文章データは、不完全な表記を含む、付記 1 に記載の情報処理装置。
- (付記 3)
- 前記不完全な表記は、略語又は短縮形を含む、付記 2 に記載の情報処理装置。

(付記 4)

前記機械学習モデルは、双方向性の固有表現抽出モデルに基づいて訓練されている、付記 1 から 3 の何れか 1 つに記載の情報処理装置。

(付記 5)

前記抽出部は、前記所定の項目に対応する項目値が前記文章データにあるか否かを判定する判定部を備える、付記 1 から 4 の何れか 1 つに記載の情報処理装置。

(付記 6)

前記所定の項目に対応する項目値が前記文章データにないと判定されると、前記出力部は、前記項目値を入力するよう促す、付記 5 に記載の情報処理装置。

(付記 7)

前記抽出部は、前記文章データにおける前記所定の項目に対応する項目値の適否を判定する判定部を備える、付記 1 から 6 の何れか 1 つに記載の情報処理装置。

(付記 8)

前記文章データにおいて前記所定の項目に対応する項目値が適切でないと判定されると、前記出力部は、前記項目値を修正するよう促す、付記 7 に記載の情報処理装置。

(付記 9)

前記文章データは、医用文章であり、

前記機械学習モデルは、前記医用文章から所定の医用項目に対応する数値を抽出する、付記 1 から 8 の何れか 1 つに記載の情報処理装置。

(付記 10)

前記文章データは、介護記録文章であり、

前記機械学習モデルは、前記介護記録文章から所定の介護項目に対応する数値を抽出する、付記 1 から 9 の何れか 1 つに記載の情報処理装置。

(付記 11)

文章データを受け付けることと、

文章データの中の項目値に対して、項目値に対応する項目をラベル付けした訓練データを用いて訓練された訓練済み機械学習モデルによって、所定の項目に対応する項目値を前記文章データから抽出することと、

前記項目と前記項目値とを関連付けて出力することと、
を有する、コンピュータが実行する情報処理方法。

(付記 1 2)

文章データを受け付けることと、

文章データの中の項目値に対して、項目値に対応する項目をラベル付けした訓練データを用いて訓練された訓練済み機械学習モデルによって、所定の項目に対応する項目値を前記文章データから抽出することと、

前記項目と前記項目値とを関連付けて出力することと、
をコンピュータに実行させるプログラム。

[0051] 以上、本開示の実施例について詳述したが、本開示は上述した特定の実施形態に限定されるものではなく、特許請求の範囲に記載された本開示の要旨の範囲内において、種々の変形・変更が可能である。

[0052] 2022年8月9日出願の特願2022-127014号の日本出願に含まれる明細書、図面および要約書の開示内容は、すべて本願に援用される。

符号の説明

[0053] 10 情報処理システム
30 端末
50 機械学習モデル
100 情報処理装置
110 入力部
120 抽出部
121 判定部
130 出力部

請求の範囲

- [請求項1] 文章データを受け付ける入力部と、
文章データの中の項目値に対して、項目値に対応する項目をラベル付けした訓練データを用いて訓練された訓練済み機械学習モデルによって、所定の項目に対応する項目値を前記受け付けた文章データから抽出する抽出部と、
前記項目と前記項目値とを関連付けて出力する出力部と、
を有する、情報処理装置。
- [請求項2] 前記文章データは、不完全な表記を含む、請求項1に記載の情報処理装置。
- [請求項3] 前記不完全な表記は、略語又は短縮形を含む、請求項2に記載の情報処理装置。
- [請求項4] 前記訓練済み機械学習モデルは、双方向性の固有表現抽出モデルに基づいて訓練されている、請求項1に記載の情報処理装置。
- [請求項5] 前記抽出部は、前記所定の項目に対応する項目値が前記文章データにあるか否かを判定する判定部を備える、請求項1に記載の情報処理装置。
- [請求項6] 前記所定の項目に対応する項目値が前記文章データにないと判定されると、前記出力部は、前記項目値を入力するよう促す、請求項5に記載の情報処理装置。
- [請求項7] 前記抽出部は、前記文章データにおける前記所定の項目に対応する項目値の適否を判定する判定部を備える、請求項1に記載の情報処理装置。
- [請求項8] 前記文章データにおいて前記所定の項目に対応する項目値が適切でないと判定されると、前記出力部は、前記項目値を修正するよう促す、請求項7に記載の情報処理装置。
- [請求項9] 前記文章データは、医用文章であり、
前記機械学習モデルは、前記医用文章から所定の医用項目に対応す

る数値を抽出する、請求項 1 に記載の情報処理装置。

[請求項10]

前記文章データは、介護記録文章であり、

前記機械学習モデルは、前記介護記録文章から所定の介護項目に対応する数値を抽出する、請求項 1 に記載の情報処理装置。

[請求項11]

文章データを受け付けることと、

文章データの中の項目値に対して、項目値に対応する項目をラベル付けした訓練データを用いて訓練された訓練済み機械学習モデルによって、所定の項目に対応する項目値を前記受け付けた文章データから抽出することと、

前記項目と前記項目値とを関連付けて出力することと、

を有する、コンピュータが実行する情報処理方法。

[請求項12]

文章データを受け付けることと、

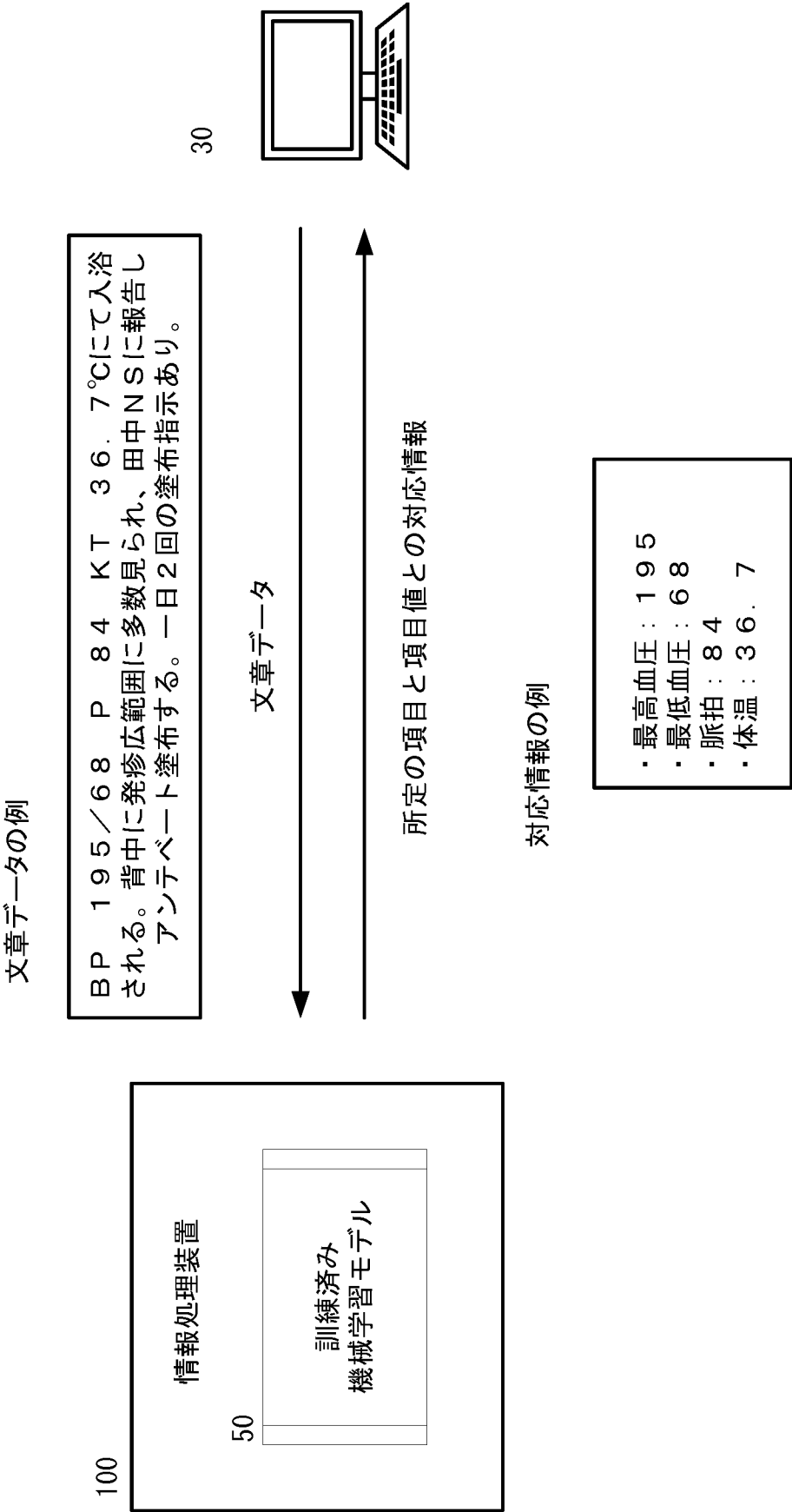
文章データの中の項目値に対して、項目値に対応する項目をラベル付けした訓練データを用いて訓練された訓練済み機械学習モデルによって、所定の項目に対応する項目値を前記受け付けた文章データから抽出することと、

前記項目と前記項目値とを関連付けて出力することと、

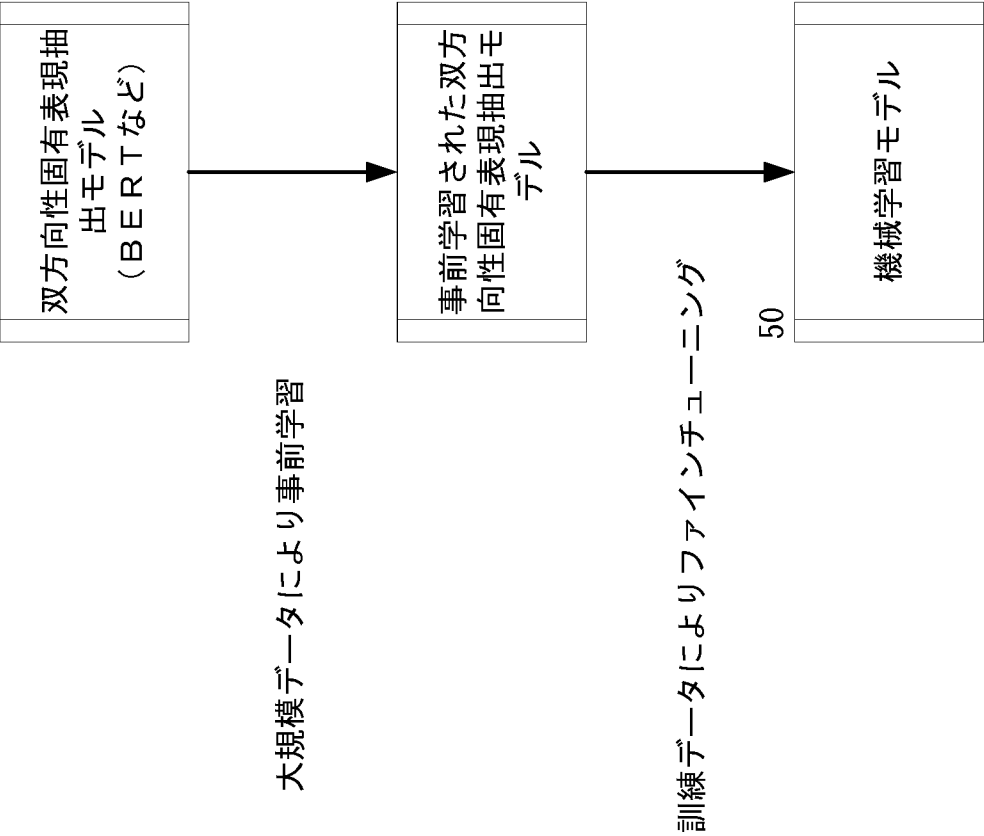
をコンピュータに実行させるプログラム。

[図1]

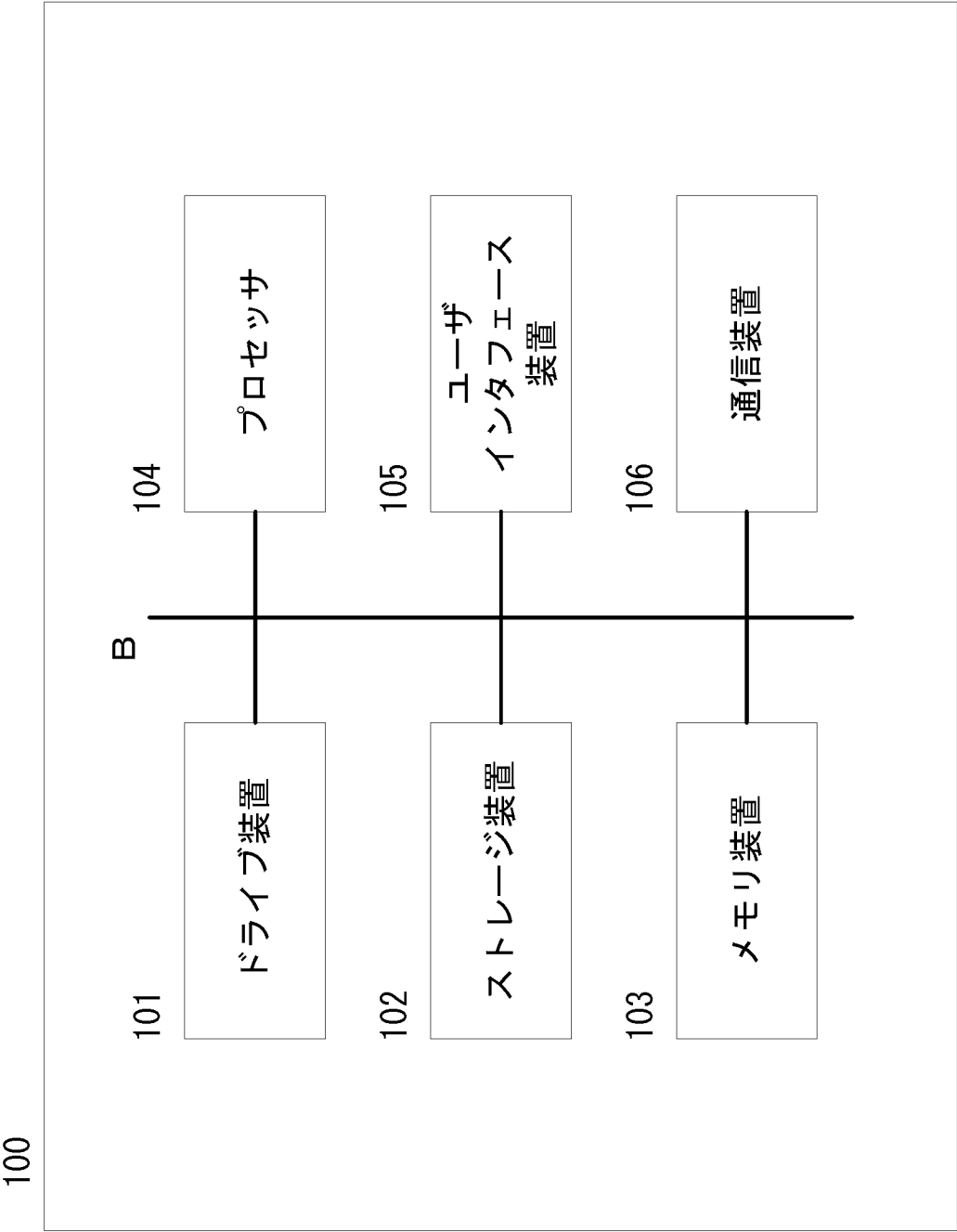
10



[図2]

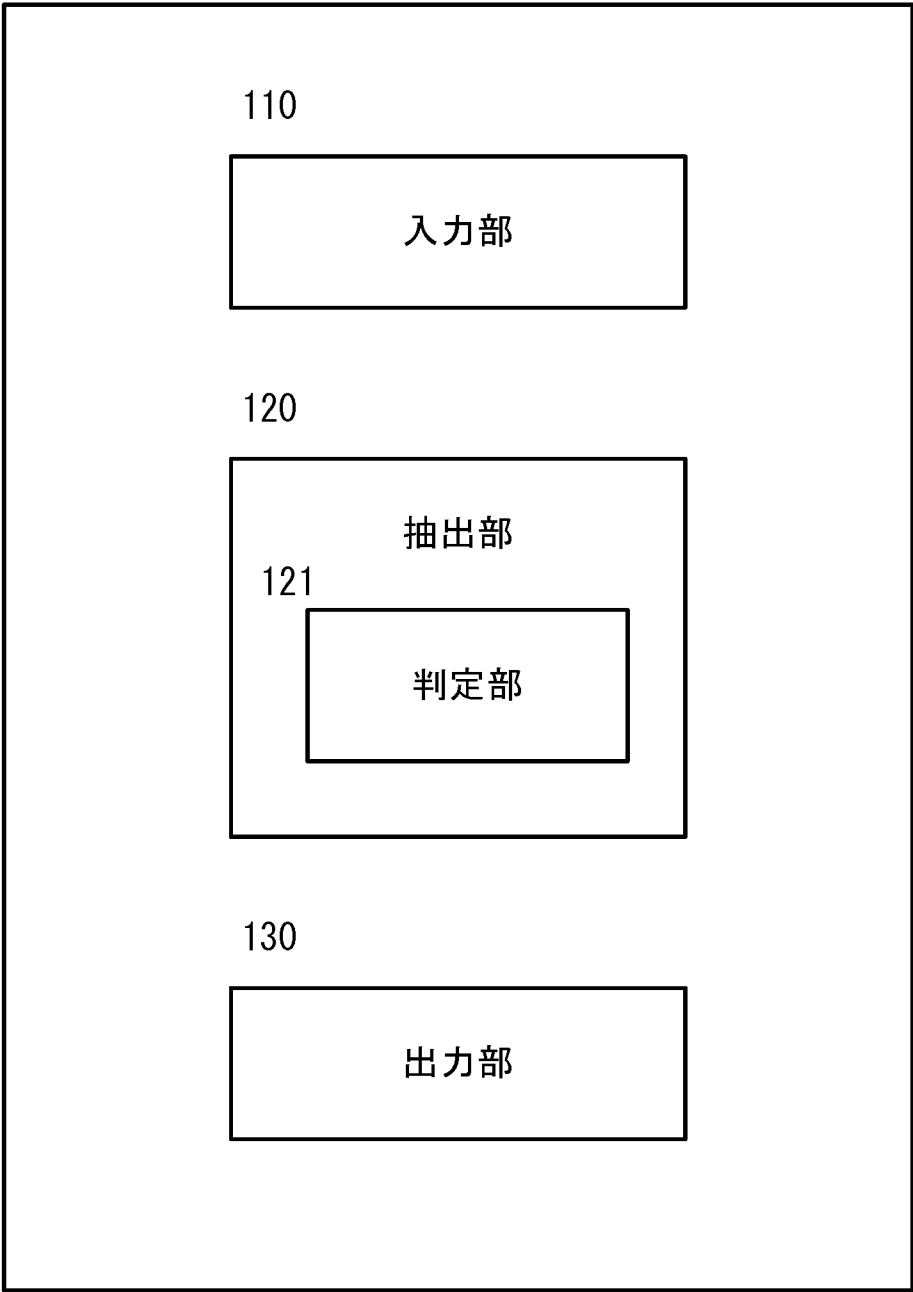


[図3]



[図4]

100



[図5]

ラベル付与

前

次

文章 (32/100)

今朝から悪寒あり, 120/80 P80

KT38.0

図 5 A

ラベル付与

前

次

文章 (32/100)

今朝から悪寒あり, 120/80 P80

KT38.0

最高血圧

脈拍

体温

最低血圧

図 5 B

[図6]

入浴前準備、BP = 110 / 70 P = 6
5, 体温 36. 5

図 6 A

入浴前準備、BP = 110 / 70 P = 6
5, 体温 36. 5

- ・最高血圧: 110 [mmHg]
- ・最低血圧: 70 [mmHg]
- ・脈拍: 65 [bpm]
- ・体温: 36. 5 [°C]

図 6 B

[図7]

KT35.3 P68 Bp118/83 しっかり覚醒されているが、活気なく目もうつろ

図7A

KT35.3 P68 Bp118/83 しっかり覚醒されているが、活気なく目もうつろ

- ・最高血圧:118[mmHg]
- ・最低血圧:83[mmHg]
- ・脈拍:68[bpm]
- ・体温:35.3[°C]

図7B

[図8]

KT 37.5 体感熱あり、クーリング開始する。

図 8 A

KT 37.5 体感熱あり、クーリング開始する。

- ・最高血圧:なし
- ・最低血圧:なし
- ・脈拍:なし
- ・体温:37.5[°C]

図 8 B

[図9]

K T 3 6 . 3 P 6 1 2 B P 1 0 8 / 7 3 体
熱感なく、顔色良好

図 9 A

K T 3 6 . 3 P 6 1 2 B P 1 0 8 / 7 3 体
熱感なく、顔色良好

- ・最高血圧: 108 [mmHg]
- ・最低血圧: 73 [mmHg]
- ・脈拍: 612 [bpm] エラー
- ・体温: 36. 3 [°C]

図 9 B

K T 3 6 . 3 P 6 2 B P 1 0 8 / 7 3 体熱
感なく、顔色良好

- ・最高血圧: 108 [mmHg]
- ・最低血圧: 73 [mmHg]
- ・脈拍: 62 [bpm] エラー
- ・体温: 36. 3 [°C]

図 9 C

[図10]

KT36.2 P61 BP110/70 体熱
感なし。様子見る。倦怠感なし。

図10A

KT36.2 P61 BP110/70 体熱
感なし。様子見る。倦怠感なし。

- ・最高血圧:110[mmHg]
- ・最低血圧:70[mmHg]
- ・脈拍:なし
- ・体温:36.2[°C]

図10B

KT36.2 P61 BP110/70 体熱
感なし。様子見る。倦怠感なし。

- ・最高血圧:110[mmHg]
- ・最低血圧:70[mmHg]
- ・脈拍:61[bpm]
- ・体温:36.2[°C]

図10C

[図11]

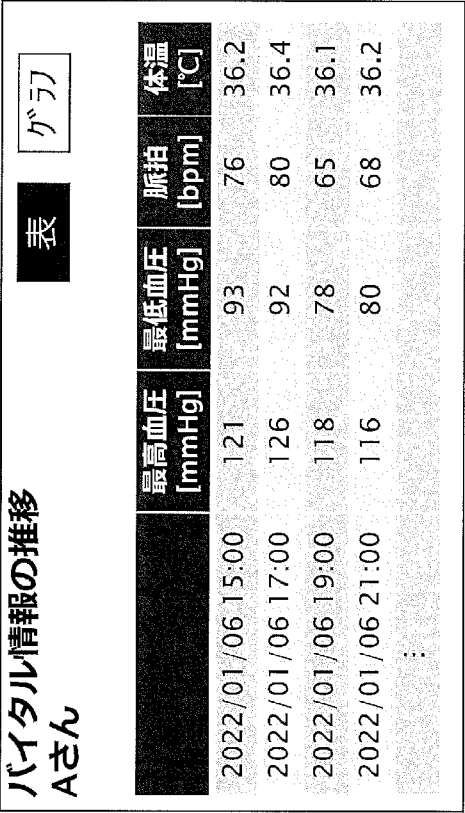
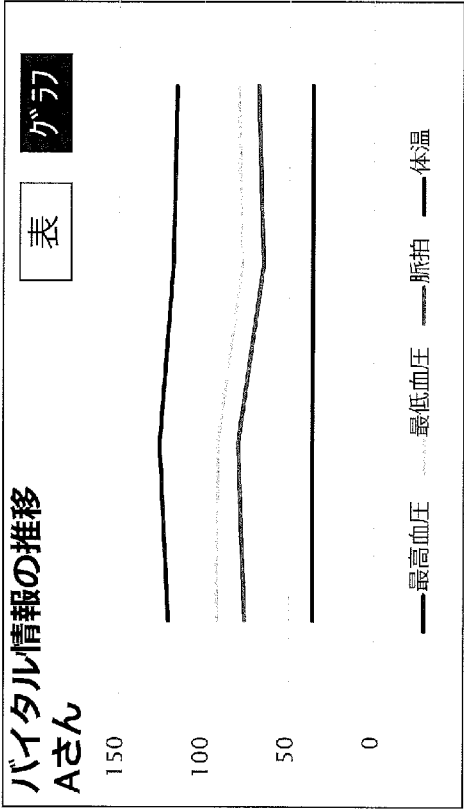


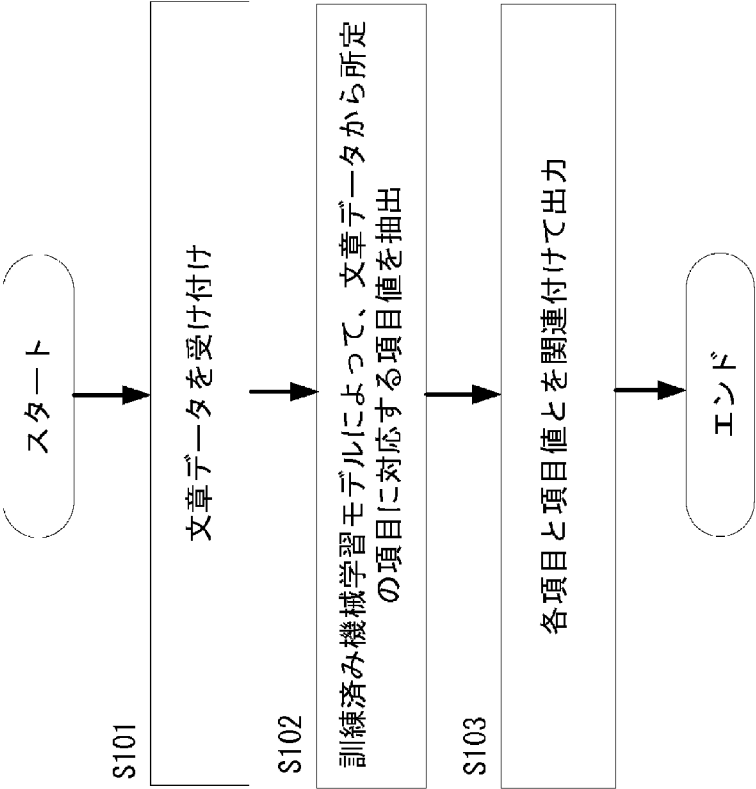
図 1 1 A



最高血圧
最低血圧
脈拍
体温

図 1 1 B

[図12]



INTERNATIONAL SEARCH REPORT

International application No.

PCT/JP2023/019895

A. CLASSIFICATION OF SUBJECT MATTER**G06F 16/33**(2019.01)i; **G06F 40/295**(2020.01)i

FI: G06F16/33; G06F40/295

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)

G06F16/00-16/958; G06F40/20-40/58

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.
Y	JP 2008-257569 A (KONICA MINOLTA MEDICAL & GRAPHIC INC.) 23 October 2008 (2008-10-23) paragraphs [0003], [0052]-[0067], fig. 3, 4	1-12
Y	JP 2017-91463 A (TOSHIBA CORP.) 25 May 2017 (2017-05-25) paragraphs [0002], [0003], [0015], fig. 2	1-12
Y	JP 2016-212533 A (TOTTORI UNIV.) 15 December 2016 (2016-12-15) abstract, paragraph [0063]	5-6



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

08 August 2023

Date of mailing of the international search report

15 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
Information on patent family members

International application No.
PCT/JP2023/019895

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)	Publication date (day/month/year)
JP	2008-257569	A	23 October 2008	US 2008/0250102 A1 paragraphs [0006], [0059]- [0074], fig. 3, 4	
JP	2017-91463	A	25 May 2017	US 2017/0139774 A1 paragraphs [0003], [0004], [0033], fig. 2	
JP	2016-212533	A	15 December 2016	(Family: none)	

A. 発明の属する分野の分類（国際特許分類（IPC）） G06F 16/33(2019.01)i; G06F 40/295(2020.01)i FI: G06F16/33; G06F40/295										
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）） G06F16/00-16/958; G06F40/20-40/58										
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 <table border="0"> <tr> <td>日本国実用新案公報</td> <td>1922-1996年</td> </tr> <tr> <td>日本国公開実用新案公報</td> <td>1971-2023年</td> </tr> <tr> <td>日本国実用新案登録公報</td> <td>1996-2023年</td> </tr> <tr> <td>日本国登録実用新案公報</td> <td>1994-2023年</td> </tr> </table>			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
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								
Y	JP 2008-257569 A（コニカミノルタエムジー株式会社）23.10.2008（2008-10-23） 段落[0003]，段落[0052]－[0067]，図3－4	1－12								
Y	JP 2017-91463 A（株式会社東芝）25.05.2017（2017-05-25） 段落[0002]－[0003]，段落[0015]，図2	1－12								
Y	JP 2016-212533 A（国立大学法人鳥取大学）15.12.2016（2016-12-15） 要約，段落[0063]	5－6								
☐ 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。										
<table border="0"> <tr> <td> * 引用文献のカテゴリー “A” 特に関連のある文献ではなく、一般的技术水準を示すもの “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 </td> <td> “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 </td> </tr> </table>			* 引用文献のカテゴリー “A” 特に関連のある文献ではなく、一般的技术水準を示すもの “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						
* 引用文献のカテゴリー “A” 特に関連のある文献ではなく、一般的技术水準を示すもの “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 08.08.2023	国際調査報告の発送日 15.08.2023									
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） 酒井 恭信 5N 9190 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/JP2023/019895

引用文献			公表日	パテントファミリー文献	公表日
JP	2008-257569	A	23.10.2008	US 2008/0250102 A1 段落[0006], 段落[0059]－ [0074], 図3－4	
JP	2017-91463	A	25.05.2017	US 2017/0139774 A1 段落[0003]－[0004], 段落 [0033], 図2	
JP	2016-212533	A	15.12.2016	(ファミリーなし)	