

(19) 日本国特許庁(JP)

(12) 公開特許公報(A)

(11) 特許出願公開番号

特開2017-204018

(P2017-204018A)

(43) 公開日 平成29年11月16日(2017.11.16)

(51) Int.Cl.  
G06F 17/30 (2006.01)F I  
G06F 17/30 180A

テーマコード (参考)

審査請求 未請求 請求項の数 8 O L (全 17 頁)

(21) 出願番号 特願2016-93659 (P2016-93659)  
(22) 出願日 平成28年5月9日(2016.5.9)(71) 出願人 000005223  
富士通株式会社  
神奈川県川崎市中原区上小田中4丁目1番  
1号  
(74) 代理人 100103528  
弁理士 原田 一男  
(72) 発明者 牧野 拓哉  
神奈川県川崎市中原区上小田中4丁目1番  
1号 富士通株式会社内

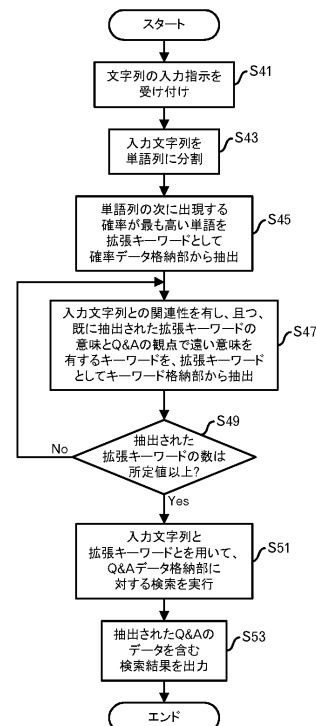
(54) 【発明の名称】 検索処理方法、検索処理プログラムおよび情報処理装置

(57) 【要約】 (修正有)

【課題】 入力された文字列に基づき適切なQ & Aを抽出する検索処理方法、処理プログラム及び情報処理装置を提供する。

【解決手段】 文字列の入力を受け付け、問合せについてのデータを含む問合せデータから、第1の単語を、問合せデータにおいて文字列の次に第1の単語が出現する確率に基づき特定し、問合せデータから、同一の質問回答データを正解とする1又は複数の問合せをそれぞれ含む複数の問合せ集合を抽出し、複数の問合せ集合のうち第1の単語が出現する問合せ集合とは異なる問合せ集合に出現する第2の単語を、複数の問合せ集合の各々に第1の単語が出現する確率と第2の単語が出現する確率との比に基づき特定し、文字列と第1の単語と第2の単語とに基づき、質問回答データを格納する第1データ格納部に対して検索を行う処理を含む。

【選択図】 図14



**【特許請求の範囲】****【請求項 1】**

コンピュータに、

文字列の入力を受け付け、

問合せについてのデータを含む問合せデータから、第 1 の単語を、前記問合せデータにおいて前記文字列の次に前記第 1 の単語が出現する確率に基づき特定し、

前記問合せデータから、同一の質問回答データを正解とする 1 又は複数の問合せをそれぞれ含む複数の問合せ集合を抽出し、

前記複数の問合せ集合のうち前記第 1 の単語が出現する問合せ集合とは異なる問合せ集合に出現する第 2 の単語を、前記複数の問合せ集合の各々に前記第 1 の単語が出現する確率と前記第 2 の単語が出現する確率との比に基づき特定し、

前記文字列と前記第 1 の単語と前記第 2 の単語とに基づき、質問回答データを格納する第 1 データ格納部に対して検索を行う、

処理を実行させる検索処理プログラム。

**【請求項 2】**

前記コンピュータに、

前記複数の問合せ集合に含まれる単語の各々について、当該単語が前記複数の問合せ集合の各々に出現する確率を算出し、

前記複数の問合せ集合の各々について、当該問合せ集合に出現する確率が所定値以上である単語を特定し、第 2 データ格納部に格納する

処理をさらに実行させ、

前記第 2 の単語を特定する処理において、

前記複数の問合せ集合の各々に前記第 1 の単語が出現する確率と前記第 2 の単語が出現する確率との比に基づき、前記第 2 データ格納部に格納されている単語から前記第 2 の単語を特定する、

請求項 1 記載の検索処理プログラム。

**【請求項 3】**

前記コンピュータに、

前記問合せデータに出現し且つ 2 つの単語を含む単語列の各々について、当該単語列が出現する確率を算出し、算出した当該確率を第 3 データ格納部に格納する、

処理をさらに実行させ、

前記第 1 の単語を特定する処理において、

前記第 3 データ格納部に格納された前記確率に基づき、前記第 1 の単語を特定する、

請求項 1 又は 2 記載の検索処理プログラム。

**【請求項 4】**

前記コンピュータに、

前記複数の問合せ集合のうち前記第 1 の単語が出現する問合せ集合及び前記第 2 の単語が出現する問合せ集合とは異なる問合せ集合に出現する第 3 の単語を、前記複数の問合せ集合の各々に前記第 1 の単語及び前記第 2 の単語が出現する確率と前記第 3 の単語が出現する確率との比に基づき特定する、

処理をさらに実行させ、

前記検索を行う処理において、

前記文字列と前記第 1 の単語と前記第 2 の単語と前記第 3 の単語とに基づき、前記第 1 データ格納部に対して検索を行う、

請求項 1 記載の検索処理プログラム。

**【請求項 5】**

前記第 2 の単語を特定する処理において、

前記第 2 の単語が前記文字列の次に出現する確率にさらに基づき、前記第 2 の単語を特定する、

請求項 1 乃至 4 のいずれか 1 つ記載の検索処理プログラム。

10

20

30

40

50

## 【請求項 6】

前記コンピュータに、  
前記第 1 データ格納部に対する前記検索の結果を出力する、  
処理をさらに実行させる請求項 1 乃至 5 のいずれか 1 つ記載の検索処理プログラム。

## 【請求項 7】

コンピュータが、  
文字列の入力を受け付け、  
問合せについてのデータを含む問合せデータから、第 1 の単語を、前記問合せデータにおいて前記文字列の次に前記第 1 の単語が出現する確率に基づき特定し、  
前記問合せデータから、同一の質問回答データを正解とする 1 又は複数の問合せをそれぞれ含む複数の問合せ集合を抽出し、  
前記複数の問合せ集合のうち前記第 1 の単語が出現する問合せ集合とは異なる問合せ集合に出現する第 2 の単語を、前記複数の問合せ集合の各々に前記第 1 の単語が出現する確率と前記第 2 の単語が出現する確率との比に基づき特定し、  
前記文字列と前記第 1 の単語と前記第 2 の単語とに基づき、質問回答データを格納する第 1 データ格納部に対して検索を行う、  
処理を実行する検索処理方法。

## 【請求項 8】

文字列の入力を受け付け、問合せについてのデータを含む問合せデータから、第 1 の単語を、前記問合せデータにおいて前記文字列の次に前記第 1 の単語が出現する確率に基づき特定する第 1 特定部と、  
前記問合せデータから、同一の質問回答データを正解とする 1 又は複数の問合せをそれぞれ含む複数の問合せ集合を抽出し、前記複数の問合せ集合のうち前記第 1 の単語が出現する問合せ集合とは異なる問合せ集合に出現する第 2 の単語を、前記複数の問合せ集合の各々に前記第 1 の単語が出現する確率と前記第 2 の単語が出現する確率との比に基づき特定する第 2 特定部と、  
前記文字列と前記第 1 の単語と前記第 2 の単語とに基づき、質問回答データを格納する第 1 データ格納部に対して検索を行う検索部と、  
を有する情報処理装置。

## 【発明の詳細な説明】

## 【技術分野】

## 【0001】

本発明は、検索処理技術に関する。

## 【背景技術】

## 【0002】

コールセンタ等においては、顧客からの問合せに応じるため、Q & A (Question and Answer) 集の検索システムが利用される。検索システムを利用するオペレータは、顧客が話したことを基に文字列の入力操作（例えば打鍵）を行うことで検索システムに検索を実行させ、正解の Q & A を提示させる。

## 【0003】

但し、従来技術を利用したとしても正解の Q & A が提示されないことがある。

## 【先行技術文献】

## 【特許文献】

## 【0004】

【特許文献 1】特開 2007-157006 号公報

【特許文献 2】特開 2014-120053 号公報

【特許文献 3】特開 2006-39881 号公報

【特許文献 4】特開 2014-134871 号公報

【特許文献 5】特開 2012-242966 号公報

【非特許文献】

## 【0005】

【非特許文献1】Steffen Bickel, Peter Haider, and Tobias Scheffer, "Learning to Complete Sentences", European Conference on Machine Learning, 2005, pp.497-504

## 【発明の概要】

## 【発明が解決しようとする課題】

## 【0006】

本発明の目的は、1つの側面では、入力された文字列に基づき適切なQ & A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】

## 【0007】

本発明に係る検索処理方法は、文字列の入力を受け付け、問合せについてのデータを含む問合せデータから、第1の単語を、問合せデータにおいて文字列の次に第1の単語が出現する確率に基づき特定し、問合せデータから、同一の質問回答データを正解とする1又は複数の問合せをそれぞれ含む複数の問合せ集合を抽出し、複数の問合せ集合のうち第1の単語が出現する問合せ集合とは異なる問合せ集合に出現する第2の単語を、複数の問合せ集合の各々に第1の単語が出現する確率と第2の単語が出現する確率との比に基づき特定し、文字列と第1の単語と第2の単語とに基づき、質問回答データを格納する第1データ格納部に対して検索を行う処理を含む。

## 【発明の効果】

## 【0008】

1つの側面では、入力された文字列に基づき適切なQ & Aを抽出できるようになる。

## 【図面の簡単な説明】

## 【0009】

【図1】図1は、文字列の入力と検索結果の表示とについて説明するための図である。

【図2A】図2Aは、検索処理装置の機能ブロック図である。

【図2B】図2Bは、検索処理部の機能ブロック図である。

【図3】図3は、問合せデータ格納部に格納されるデータの一例を示す図である。

【図4】図4は、Q & Aデータ格納部に格納されるデータの一例を示す図である。

【図5】図5は、第1算出処理部が実行する処理の処理フローを示す図である。

【図6】図6は、問合せデータ格納部に格納される問合せのデータの一例を示す図である

10

20

30

【図7】図7は、文データ格納部に格納されるデータの一例を示す図である。

【図8】図8は、単語列データ格納部に格納されるデータの一例を示す図である。

【図9】図9は、cnt(w)の一例及びcnt(u, w)の一例を示す図である。

【図10】図10は、確率データ格納部に格納されるデータの一例を示す図である。

【図11】図11は、第1算出部が処理を実行した後に第2算出部が実行する処理の処理フローを示す図である。

【図12】図12は、確率分布データ格納部に格納されるデータの一例を示す図である。

【図13】図13は、キーワード格納部に格納されるデータの一例を示す図である。

【図14】図14は、検索処理部が実行する処理の処理フローを示す図である。

【図15】図15は、抽出される拡張キーワードの一例を示す図である。

【図16】図16は、言語モデルについて説明するための図である。

【図17】図17は、第2の実施の形態のシステムの概要を示す図である。

【図18】図18は、コンピュータの機能ブロック図である。

## 【発明を実施するための形態】

## 【0010】

## [実施の形態1]

入力された文字列に基づいて検索を行う場合、文字列に含まれる文字の数が増えるほど検索の手がかりが増えるため、正解のQ & Aを抽出する可能性が高くなるが、ユーザの負担は多くなる。例えば図1に示すように、ユーザが入力しようと考えた文字列の一部が

40

50

入力欄 1001 に入力された段階で、検索結果の表示欄 1002 に正解の Q & A（図 1 においては、太枠 1003 で囲まれた部分）が表示されることが好ましい。

#### 【0011】

また、図 1 の例のように、入力された文字列が正解の Q & A の文には含まれていない場合においても正解の Q & A が抽出されることが好ましい。但し、入力された文字列のみを手がかりとして検索を行う方法を利用すると、図 1 の例における正解の Q & A は表示されず、正解ではない Q & A が表示される。また、入力された文字列と一緒に出現する傾向がある文字列を使用して検索を行う場合も、検索結果が多様な Q & A を含むとは限らず、正解の Q & A が表示されないことがある。

#### 【0012】

そこで、本実施の形態においては以下のような方法で検索処理が実行される。

#### 【0013】

図 2 A に、本実施の形態における検索処理装置 1 の機能ブロック図を示す。検索処理装置 1 は、問合せデータ格納部 101 と、文データ格納部 102 と、単語列データ格納部 103 と、Q & A データ格納部 104 と、確率データ格納部 105 と、確率分布データ格納部 106 と、キーワード格納部 107 と、出力データ格納部 108 と、第 1 算出部 111 と、第 2 算出部 112 と、検索処理部 113 とを含む。図 2 B に、検索処理部 113 の機能ブロック図を示す。検索処理部 113、第 1 処理部 1131 と、第 2 処理部 1132 と、第 3 処理部 1133 とを含む。

#### 【0014】

第 1 算出部 111 は、問合せデータ格納部 101 に格納されているデータに基づき処理を実行し、処理結果を文データ格納部 102、単語列データ格納部 103 及び確率データ格納部 105 に格納する。第 2 算出部 112 は、単語列データ格納部 103 に格納されているデータ、Q & A データ格納部 104 に格納されているデータ及び確率データ格納部 105 に格納されているデータに基づき処理を実行し、処理結果を確率分布データ格納部 106 及びキーワード格納部 107 に格納する。検索処理部 113 は、確率データ格納部 105 に格納されているデータ、確率分布データ格納部 106 に格納されているデータ及びキーワード格納部 107 に格納されているデータに基づき処理を実行し、処理結果を出力データ格納部 108 に格納する。特に、第 1 処理部 1131 は、拡張キーワードのうち最初に追加される拡張キーワードを抽出する処理を実行する。第 2 処理部 1132 は、拡張キーワードのうち 2 番目以降に追加される拡張キーワードを抽出する処理を実行する。第 3 処理部 1133 は、入力された文字列と拡張キーワードとに基づき検索を実行する。

#### 【0015】

図 3 に、問合せデータ格納部 101 に格納されるデータの一例を示す。図 3 の例では、問合せの ID (Identifier) と、問合せに関する自然言語のデータと、問合せの正解として適切な Q & A（すなわち、問合せに対して提示される応答として適切な Q & A）の ID とが格納される。問合せデータ格納部 101 に格納される問合せのデータは、実際に過去に受け付けた問合せのデータである。

#### 【0016】

図 4 に、Q & A データ格納部 104 に格納されるデータの一例を示す。図 4 の例では、Q & A の ID と、質問のデータと、回答のデータとが格納される。Q & A データ格納部 104 に格納される質問のデータ及び回答のデータは、Q & A のモデルとして管理者等により入力されたデータ（例えば F A Q (Frequently Asked Questions) のデータ）である。

#### 【0017】

次に、図 5 乃至図 16 を用いて、検索処理装置 1 の動作を説明する。

#### 【0018】

まず、図 5 乃至図 10 を用いて、第 1 算出部 111 が実行する処理について説明する。検索処理装置 1 の第 1 算出部 111 は、問合せデータ格納部 101 に格納されている問合せのデータを文単位に分割して文データを生成する。そして、第 1 算出部 111 は、生成した文データを文データ格納部 102 に格納する（図 5：ステップ S1）。

## 【0019】

図6に、問合せデータ格納部101に格納される問合せのデータの一例を示す。問合せのデータは、問合せごとに、1又は複数の文のデータを含む。ステップS1の処理によって、例えば図7に示すように、各文について文データが生成されて文データ格納部102に格納される。

## 【0020】

第1算出部111は、文データ格納部102に格納された文データに対して単語分割（品詞分解とも呼ばれる）を実行して単語列データを生成する。そして、第1算出部111は、生成した単語列データを単語列データ格納部103に格納する（ステップS3）。

## 【0021】

図8に、単語列データ格納部103に格納されるデータの一例を示す。図8の例では、文データが単語単位に分割されているが、単語の出現順序は維持される。

## 【0022】

第1算出部111は、単語列データ格納部103に格納されている単語のうち未処理の単語を1つ特定する（ステップS5）。ステップS5において特定した単語をwとする。

## 【0023】

第1算出部111は、ステップS5において特定した単語wが単語列データ格納部103に格納された単語列データにおいて出現する回数を計数する（ステップS7）。ステップS7において計数される回数をcnt(w)とする。図9(a)に、ステップS7において計数されるcnt(w)の一例を示す。

## 【0024】

第1算出部111は、単語列データ格納部103に格納された単語列データにおいて単語uの次に単語wが出現する回数を単語uごとに計数する（ステップS9）。ステップS9において計数される回数をcnt(u, w)とする。図9(b)に、ステップS9において計数されるcnt(u, w)の一例を示す。

## 【0025】

第1算出部111は、単語uの次に単語wが出現する確率を単語uごとに算出し、算出した確率を確率データ格納部105に格納する（ステップS11）。ステップS11においては、以下の式に従って単語uごとに確率が算出される。

## 【0026】

【数1】

$$P(w|u) = \frac{cnt(u, w)}{cnt(w)}$$

## 【0027】

図10に、確率データ格納部105に格納されるデータの一例を示す。図10の例では、単語uと単語wとの組合せごとに、P(w|u)が格納される。

## 【0028】

第1算出部111は、未処理の単語が有るか判定する（ステップS13）。未処理の単語が有る場合（ステップS13：Yesルート）、ステップS5の処理に戻る。一方、未処理の単語が無い場合（ステップS13：Noルート）、処理は終了する。

## 【0029】

以上のような処理を実行すれば、事前に単語列の出現確率が算出されるので、検索に要する時間が長くなることを抑制できるようになる。

## 【0030】

次に、図11乃至図13を用いて、第1算出部111が処理を実行した後に第2算出部112が実行する処理について説明する。

## 【0031】

まず、第2算出部112は、単語列データ格納部103に格納されている単語列データ

10

20

30

40

50

から未処理の内容語（名詞、動詞、形容詞など）を1つ特定する（図11：ステップS21）。ステップS21において特定された内容語を処理対象の内容語と呼ぶ。

【0032】

第2算出部112は、Q & Aデータ格納部104にIDが格納されているQ & Aのうち未処理のQ & AのIDを1つ特定する（ステップS23）。

【0033】

第2算出部112は、ステップS23において特定したQ & AのIDに対応する問合せ集合（すなわち、ステップS23において特定したQ & Aを正解とする問合せの集合）を問合せデータ格納部101から特定する（ステップS25）。

【0034】

第2算出部112は、ステップS23において特定したQ & Aを正解とする問合せに処理対象の内容語が出現する回数を計数する（ステップS27）。

【0035】

第2算出部112は、問合せデータ格納部101にIDが格納されている全問合せに処理対象の内容語が出現する回数を計数する（ステップS29）。なお、ステップS29の処理が既に実行されている場合にはステップS29の処理を省略してもよいため、図11においてステップS29のブロックは破線で示されている。

【0036】

第2算出部112は、ステップS23において特定したQ & Aを正解とする問合せ集合に処理対象の内容語が出現する確率を算出し、算出した確率を確率分布データ格納部106に格納する（ステップS31）。

【0037】

ステップS31においては、以下の式に従って計算が行われる。

【0038】

【数2】

$$P_i(w) = \frac{cnt(w, F_i)}{\sum_k cnt(w, F_k)}$$

【0039】

ここで、 $i$ はQ & AのIDを表す変数であり、 $w$ はステップS21において特定した内容語である。 $cnt(w, F_i)$ は識別子が $i$ であるQ & Aを正解とする問合せ集合に内容語 $w$ が出現する回数であり、 $\sum_k cnt(w, F_k)$ は全問合せに内容語 $w$ が出現する回数を表す。

【0040】

図12に、確率分布データ格納部106に格納されるデータの一例を示す。図12の例では、内容語ごとに、その内容語が各Q & Aを正解とする問合せ集合に出現する確率が格納されている。

【0041】

第2算出部112は、ステップS31において算出した確率が0ではない場合に、処理対象の内容語を拡張キーワードの候補としてQ & AのIDに対応付けてキーワード格納部107に登録する（ステップS33）。

【0042】

図13に、キーワード格納部107に格納されるデータの一例を示す。図13の例では、Q & Aの識別子と、そのQ & Aを正解とする問合せ集合に出現する確率が0ではないキーワードとが格納される。

【0043】

第2算出部112は、未処理のQ & Aが有るか判定する（ステップS35）。未処理のQ & Aが有る場合（ステップS35：Yesルート）、ステップS23の処理に戻る。

【0044】

10

20

30

40

50

一方、未処理の Q & A が無い場合（ステップ S 3 5：N o ルート）、第 2 算出部 1 1 2 は、未処理の内容語が有るか判定する（ステップ S 3 7）。

#### 【0045】

未処理の内容語が有る場合（ステップ S 3 7：Y e s ルート）、ステップ S 2 1 の処理に戻る。一方、未処理の内容語が無い場合（ステップ S 3 7：N o ルート）、処理は終了する。

#### 【0046】

以上のような処理を実行すれば、各内容語が各問合せ集合（ここでは、同一の Q & A を正解とする問合せ集合）に出現する確率が事前に算出されるので、検索に要する時間が長くなることを抑制できるようになる。

#### 【0047】

次に、図 1 4 乃至図 1 6 を用いて、検索処理部 1 1 3 が実行する処理について説明する。

#### 【0048】

まず、検索処理部 1 1 3 は、文字列の入力指示を検索処理装置 1 の操作者から受け付ける（図 1 4：ステップ S 4 1）。ステップ S 4 1 における文字列は、例えば、特許請求の範囲における文字列に相当する。

#### 【0049】

検索処理部 1 1 3 は、入力された文字列を単語列に分割する（ステップ S 4 3）。

#### 【0050】

検索処理部 1 1 3 における第 1 処理部 1 1 3 1 は、入力された文字列から生成された単語列の次に出現する確率が最も高い単語を、拡張キーワードとして確率データ格納部 1 0 5 から抽出する（ステップ S 4 5）。例えば「こどもが」という文字列が入力された場合、文字列は「こども／が」という単語列に分割されるので、「こども」の次に「が」が出現する確率と「が」の次に或る単語が出現する確率とに基づき、「こどもが」の次に或る単語が出現する確率を求めることができる。ここでは、図 1 5（a）に示すように、「風邪」という単語が抽出されたとする。ステップ S 4 5 において特定される単語は、例えば、特許請求の範囲における第 1 の単語に相当する。

#### 【0051】

なお、単語列のつながりのよさを計算する言語モデルが知られており、ステップ S 4 5 の処理における計算にもその技術を利用することができる。例えば図 1 6 に示すように、「こどもがインフルエンザにかかった」という文が入力された場合、入力された文は「こども／が／インフルエンザ／に／かかっ／た」という単語列に分割することができる。ここで、「こどもがインフルエンザにかかった」という文が出現する確率は、 $P（が | こども） * P（インフルエンザ | が） * P（に | インフルエンザ） * P（かかっ | に） * P（た | かかっ）$  によって算出される。このような言語モデルについては、非特許文献 1 にも記述がある。

#### 【0052】

検索処理部 1 1 3 における第 2 処理部 1 1 3 2 は、入力された文字列との関連性を有し、且つ、既に抽出された拡張キーワードの意味と Q & A の観点で遠い意味を有する単語を、拡張キーワードとしてキーワード格納部 1 0 7 から抽出する（ステップ S 4 7）。ステップ S 4 7 において特定される単語は、例えば、特許請求の範囲における第 2 の単語に相当する。

#### 【0053】

ステップ S 4 7 においては、以下の式によってキーワードが抽出される。

#### 【0054】

#### 【数 3】

$$\arg \max_{w_i \in V \setminus S} \lambda \text{sim}_1(w_i, Q) - (1 - \lambda) \max_{q_j \in S} \text{sim}_2(w_i, q_j)$$

10

20

30

40

50

## 【0055】

ここで、Qは入力された文字列から生成された単語列  $t_1, t_2, \dots$  である。Vは拡張キーワードの候補の集合である。 $w_i$ はVに含まれる拡張キーワードの候補である。Sは計算時点までに選択された拡張キーワードの集合である。 $q_j$ はSに含まれる拡張キーワードである。 $\lambda$ はハイパーパラメータである。

## 【0056】

第1項の  $\text{sim}_1(w_i, Q)$  は以下のように表される。

## 【0057】

## 【数4】

$$\text{sim}_1(w_i, Q) = P(w_i | Q) = P(w_i | t_1, t_2, \dots)$$

10

## 【0058】

第1項は、単語列  $t_1, t_2, \dots$  とのつながりのよさ（すなわち、単語列  $t_1, t_2, \dots$  の次に出現する確率の高さ）を表す。

## 【0059】

第2項の  $\text{sim}_2(w_i, q_j)$  は以下のように表される。

## 【0060】

## 【数5】

$$\text{sim}_2(w_i, q_j) = \left\{ \sum_k P_k(w_i) \log \frac{P_k(w_i)}{P_k(q_j)} \right\}^{-1}$$

20

## 【0061】

第2項は、既に選択された拡張キーワードとの、Q & Aの観点での語義の近さを表す。出現確率の比  $p_k(w_i) / p_k(q_j)$  が大きいほど第2の項の値は小さくなる。例えば、或る問合せ集合に  $w_i$  が出現する確率が高く且つ  $q_j$  が出現する確率が低い場合、第2項の値は小さくなる。また、或る問合せ集合に  $w_i$  が出現する確率が低く且つ  $q_j$  が出現する確率が高い場合も、第2項の値は小さくなる。

## 【0062】

30

例えば図15(b)の例に示すように、「こどもが」という文字列が入力され且つ「風邪」という拡張キーワードが既に選択された場合、「こどもが」の次に出現する確率が比較的高く且つ「風邪」とはQ & Aの観点で意味が近くない「扶養」が選択される。

## 【0063】

また、例えば図15(c)の例に示すように、「こどもが」という文字列が入力され、「風邪」という拡張キーワードが既に選択され、且つ「扶養」という拡張キーワードが既に選択された場合、「こどもが」の次に出現する確率が比較的高く且つ「風邪」とはQ & Aの観点で意味が近くない「産まれる」が選択される。

## 【0064】

検索処理部113は、ステップS45及びS47において抽出された拡張キーワードの数が所定値以上であるか判定する（ステップS49）。ステップS45及びS47において抽出された拡張キーワードの数が所定値以上ではない場合（ステップS49：Nルート）、ステップS47の処理に戻る。

40

## 【0065】

一方、ステップS45及びS47において抽出された拡張キーワードの数が所定値以上である場合（ステップS49：Yesルート）、検索処理部113における第3処理部1133は、入力された文字列と抽出された拡張キーワードとを用いて、Q & Aデータ格納部104に対する検索を実行する（ステップS51）。例えば、（入力文字列）AND（拡張キーワード OR 拡張キーワード OR  $\dots$  OR 拡張キーワード）のような検索式によって検索が実行される。

50

## 【0066】

検索処理部113は、検索によって抽出されたQ & Aのデータを含む検索結果のデータを生成し、出力データ格納部108に格納する。そして、検索処理部113は、出力データ格納部108に格納された検索結果のデータを出力する（ステップS53）。例えば、検索処理部113は、検索結果のデータを検索処理装置1の表示装置に表示させる。そして処理は終了する。

## 【0067】

以上のような処理を実行すれば、多様な観点で特定された拡張キーワードに基づく検索が実行されるので、観点が偏った検索結果が抽出されることを回避できるようになる。

## 【0068】

また、入力された文字列の次に出現する確率を使用するので、入力された文字列に関連を有する拡張キーワードを抽出できるようになり、正解のQ & Aが抽出されやすくなる。

## 【0069】

また、打鍵などの入力操作の負担を減らすことができるようになる。

## 【0070】

## 〔実施の形態2〕

図17に、第2の実施の形態におけるシステムの概要を示す。第2の実施の形態においては、インターネット等のネットワーク5に、検索処理装置1と、ユーザ端末3a及び3bとが接続される。図17においてユーザ端末の数は2であるが、数に限定は無い。

## 【0071】

ユーザ端末3a及び3bは、ユーザから文字列の入力指示を受け付け、入力された文字列を検索処理装置1に送信する。検索処理装置1は、受信した文字列に基づき検索を実行し、検索結果をユーザ端末3a及び3bに送信する。

## 【0072】

このようにすれば、検索処理装置1を直接操作しないユーザが検索処理装置1によるQ & Aデータの検索を利用することができるようになる。

## 【0073】

以上本発明の一実施の形態を説明したが、本発明はこれに限定されるものではない。例えば、上で説明した検索処理装置1の機能ブロック構成は実際のプログラムモジュール構成に一致しない場合もある。

## 【0074】

また、上で説明した各テーブルの構成は一例であって、上記のような構成でなければならないわけではない。さらに、処理フローにおいても、処理結果が変わらなければ処理の順番を入れ替えることも可能である。さらに、並列に実行させるようにしても良い。

## 【0075】

なお、上で述べた検索処理装置1は、コンピュータ装置であって、図18に示すように、メモリ2501とCPU（Central Processing Unit）2503とハードディスク・ドライブ（HDD：Hard Disk Drive）2505と表示装置2509に接続される表示制御部2507とリムーバブル・ディスク2511用のドライブ装置2513と入力装置2515とネットワークに接続するための通信制御部2517とがバス2519で接続されている。オペレーティング・システム（OS：Operating System）及び本実施例における処理を実施するためのアプリケーション・プログラムは、HDD2505に格納されており、CPU2503により実行される際にはHDD2505からメモリ2501に読み出される。CPU2503は、アプリケーション・プログラムの処理内容に応じて表示制御部2507、通信制御部2517、ドライブ装置2513を制御して、所定の動作を行わせる。また、処理途中のデータについては、主としてメモリ2501に格納されるが、HDD2505に格納されるようにしてもよい。本発明の実施例では、上で述べた処理を実施するためのアプリケーション・プログラムはコンピュータ読み取り可能なリムーバブル・ディスク2511に格納されて頒布され、ドライブ装置2513からHDD2505にインストールされる。インターネットなどのネットワーク及び通信制御部2517を経由し

10

20

30

40

50

て、HDD 2505にインストールされる場合もある。このようなコンピュータ装置は、上で述べたCPU 2503、メモリ2501などのハードウェアとOS及びアプリケーション・プログラムなどのプログラムとが有機的に協働することにより、上で述べたような各種機能を実現する。

【0076】

以上述べた本発明の実施の形態をまとめると、以下のようになる。

【0077】

本実施の形態に係る検索処理方法は、(A)文字列(例えば実施の形態のステップS41の文字列)の入力を受け付け、(B)問合せについてのデータを含む問合せデータ(例えば実施の形態における問合せデータ格納部101に格納されているデータ)から、第1の単語(例えば実施の形態のステップS45において抽出される単語)を、問合せデータにおいて文字列の次に第1の単語が出現する確率に基づき特定し、(C)問合せデータから、同一の質問回答データを正解とする1又は複数の問合せをそれぞれ含む複数の問合せ集合を抽出し、(D)複数の問合せ集合のうち第1の単語が出現する問合せ集合とは異なる問合せ集合に出現する第2の単語(例えば実施の形態のステップS47において抽出される単語)を、複数の問合せ集合の各々に第1の単語が出現する確率と第2の単語が出現する確率との比に基づき特定し、(E)文字列と第1の単語と第2の単語とに基づき、質問回答データを格納する第1データ格納部(例えば実施の形態におけるQ&Aデータ格納部104)に対して検索を行う処理を含む。

【0078】

入力された文字列からだけではユーザの真の意図をくみ取るのは難しい。しかし、上で述べたようにすれば、多様な観点で特定された単語に基づく検索が実行されるので、観点が偏った検索結果が抽出されることを回避し、正解の質問回答データを抽出することができるようになる。

【0079】

また、本検索処理方法は、(F)複数の問合せ集合に含まれる単語の各々について、当該単語が複数の問合せ集合の各々に出現する確率を算出し、(G)複数の問合せ集合の各々について、当該問合せ集合に出現する確率が所定値以上である単語を特定し、第2データ格納部に格納する処理をさらに含んでもよい。そして、第2の単語を特定する処理において、(d1)複数の問合せ集合の各々に第1の単語が出現する確率と第2の単語が出現する確率との比に基づき、第2データ格納部に格納されている単語から第2の単語を特定してもよい。

【0080】

正解の質問回答データが同じである単語が選ばれることを抑制できるようになる。また、予め確率を計算しておけば、文字列の入力時に迅速に検索を行えるようになる。

【0081】

また、本検索処理方法は、(H)問合せデータに出現し且つ2つの単語を含む単語列の各々について、当該単語列が出現する確率を算出し、算出した当該確率を第3データ格納部に格納してもよい。そして、第1の単語を特定する処理において、(b1)第3データ格納部に格納された確率に基づき、第1の単語を特定してもよい。

【0082】

予め確率を算出しておけば、文字列の入力時に迅速に検索を行えるようになる。

【0083】

また、本検索処理方法は、(I)複数の問合せ集合のうち第1の単語が出現する問合せ集合及び第2の単語が出現する問合せ集合とは異なる問合せ集合に出現する第3の単語を、複数の問合せ集合の各々に第1の単語及び第2の単語が出現する確率と第3の単語が出現する確率との比に基づき特定する処理をさらに含んでもよい。そして、検索を行う処理において、(e1)文字列と第1の単語と第2の単語と第3の単語とに基づき、第1データ格納部に対して検索を行ってもよい。

【0084】

さらに異なる観点から得られた単語に基づく検索を行うことができるようになる。

【0085】

また、第2の単語を特定する処理において、(d2)第2の単語が文字列の次に出現する確率にさらに基づき、第2の単語を特定してもよい。

【0086】

より適切な第2の単語を特定できるようになる。

【0087】

また、本検索処理方法は、(J)第1データ格納部に対する検索の結果を出力する処理をさらに含んでもよい。

【0088】

検索の結果を文字列を入力したユーザ等が確認できるようになる。

【0089】

また、第1の単語は、文字列の次に出現する確率が最も高い単語であってもよい。

【0090】

また、第2の単語は内容語であってもよい。

【0091】

なお、上記方法による処理をコンピュータに実行させるためのプログラムを作成することができ、当該プログラムは、例えばフレキシブルディスク、CD-ROM、光磁気ディスク、半導体メモリ、ハードディスク等のコンピュータ読み取り可能な記憶媒体又は記憶装置に格納される。尚、中間的な処理結果はメインメモリ等の記憶装置に一時保管される。

【0092】

以上の実施例を含む実施形態に関し、さらに以下の付記を開示する。

【0093】

(付記1)

コンピュータに、

文字列の入力を受け付け、

問合せについてのデータを含む問合せデータから、第1の単語を、前記問合せデータにおいて前記文字列の次に前記第1の単語が出現する確率に基づき特定し、

前記問合せデータから、同一の質問回答データを正解とする1又は複数の問合せをそれぞれ含む複数の問合せ集合を抽出し、

前記複数の問合せ集合のうち前記第1の単語が出現する問合せ集合とは異なる問合せ集合に出現する第2の単語を、前記複数の問合せ集合の各々に前記第1の単語が出現する確率と前記第2の単語が出現する確率との比に基づき特定し、

前記文字列と前記第1の単語と前記第2の単語とに基づき、質問回答データを格納する第1データ格納部に対して検索を行う、

処理を実行させる検索処理プログラム。

【0094】

(付記2)

前記コンピュータに、

前記複数の問合せ集合に含まれる単語の各々について、当該単語が前記複数の問合せ集合の各々に出現する確率を算出し、

前記複数の問合せ集合の各々について、当該問合せ集合に出現する確率が所定値以上である単語を特定し、第2データ格納部に格納する

処理をさらに実行させ、

前記第2の単語を特定する処理において、

前記複数の問合せ集合の各々に前記第1の単語が出現する確率と前記第2の単語が出現する確率との比に基づき、前記第2データ格納部に格納されている単語から前記第2の単語を特定する、

付記1記載の検索処理プログラム。

10

20

30

40

50

## 【0095】

(付記3)

前記コンピュータに、

前記問合せデータに出現し且つ2つの単語を含む単語列の各々について、当該単語列が出現する確率を算出し、算出した当該確率を第3データ格納部に格納する、

処理をさらに実行させ、

前記第1の単語を特定する処理において、

前記第3データ格納部に格納された前記確率に基づき、前記第1の単語を特定する、

付記1又は2記載の検索処理プログラム。

## 【0096】

(付記4)

前記コンピュータに、

前記複数の問合せ集合のうち前記第1の単語が出現する問合せ集合及び前記第2の単語が出現する問合せ集合とは異なる問合せ集合に出現する第3の単語を、前記複数の問合せ集合の各々に前記第1の単語及び前記第2の単語が出現する確率と前記第3の単語が出現する確率との比に基づき特定する、

処理をさらに実行させ、

前記検索を行う処理において、

前記文字列と前記第1の単語と前記第2の単語と前記第3の単語とに基づき、前記第1データ格納部に対して検索を行う、

付記1記載の検索処理プログラム。

## 【0097】

(付記5)

前記第2の単語を特定する処理において、

前記第2の単語が前記文字列の次に出現する確率にさらに基づき、前記第2の単語を特定する、

付記1乃至4のいずれか1つ記載の検索処理プログラム。

## 【0098】

(付記6)

前記コンピュータに、

前記第1データ格納部に対する前記検索の結果を出力する、

処理をさらに実行させる付記1乃至5のいずれか1つ記載の検索処理プログラム。

## 【0099】

(付記7)

前記第1の単語は、前記文字列の次に出現する確率が最も高い単語である、

付記1乃至6のいずれか1つ記載の検索処理プログラム。

## 【0100】

(付記8)

前記第2の単語は内容語である、

付記1乃至7のいずれか1つ記載の検索処理プログラム。

## 【0101】

(付記9)

コンピュータが、

文字列の入力を受け付け、

問合せについてのデータを含む問合せデータから、第1の単語を、前記問合せデータにおいて前記文字列の次に前記第1の単語が出現する確率に基づき特定し、

前記問合せデータから、同一の質問回答データを正解とする1又は複数の問合せをそれぞれ含む複数の問合せ集合を抽出し、

前記複数の問合せ集合のうち前記第1の単語が出現する問合せ集合とは異なる問合せ集合に出現する第2の単語を、前記複数の問合せ集合の各々に前記第1の単語が出現する確

10

20

30

40

50

率と前記第 2 の単語が出現する確率との比に基づき特定し、

前記文字列と前記第 1 の単語と前記第 2 の単語とに基づき、質問回答データを格納する第 1 データ格納部に対して検索を行う、

処理を実行する検索処理方法。

【0102】

(付記 10)

文字列の入力を受け付け、問合せについてのデータを含む問合せデータから、第 1 の単語を、前記問合せデータにおいて前記文字列の次に前記第 1 の単語が出現する確率に基づき特定する第 1 特定部と、

前記問合せデータから、同一の質問回答データを正解とする 1 又は複数の問合せをそれぞれ含む複数の問合せ集合を抽出し、前記複数の問合せ集合のうち前記第 1 の単語が出現する問合せ集合とは異なる問合せ集合に出現する第 2 の単語を、前記複数の問合せ集合の各々に前記第 1 の単語が出現する確率と前記第 2 の単語が出現する確率との比に基づき特定する第 2 特定部と、

前記文字列と前記第 1 の単語と前記第 2 の単語とに基づき、質問回答データを格納する第 1 データ格納部に対して検索を行う検索部と、

を有する情報処理装置。

【符号の説明】

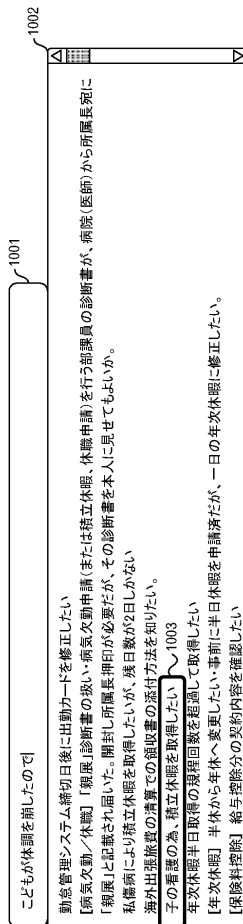
【0103】

1 検索処理装置      101 問合せデータ格納部  
102 文データ格納部      103 単語列データ格納部  
104 Q & A データ格納部      105 確率データ格納部  
106 確率分布データ格納部      107 キーワード格納部  
108 出力データ格納部      111 第 1 算出部  
112 第 2 算出部      113 検索処理部  
3a, 3b ユーザ端末      5 ネットワーク  
1131 第 1 処理部      1132 第 2 処理部  
1133 第 3 処理部

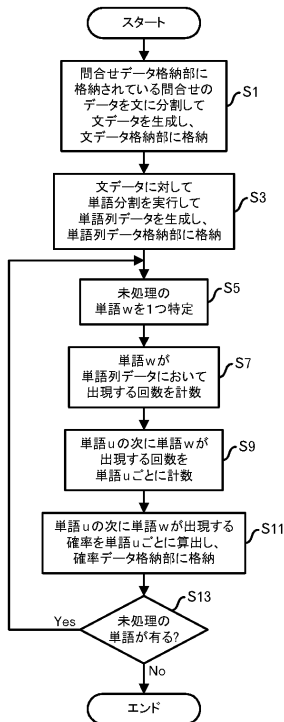
10

20

【図 1】



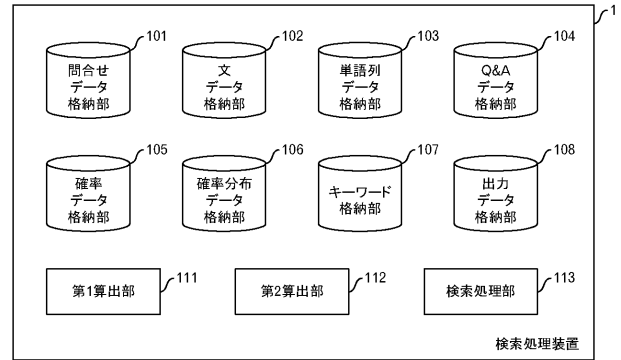
【図 5】



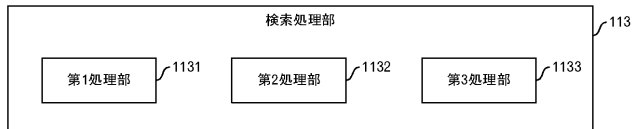
【図 6】

・子供が体調をくずしてしまった。休みを取得するには…  
・定期券を失くしたので再発行方法を教えて…  
・電車が遅延して遅刻になってしまった。どのように申請…

【図 2 A】



【図 2 B】



【図 3】

| 問合せのID | 問合せ                   | 正解のQ&AのID |
|--------|-----------------------|-----------|
| 1      | こどもがインフルエンザにかかったのですが… | 1         |
| ⋮      | ⋮                     | ⋮         |

【図 4】

| Q&AのID | 質問           | 回答             |
|--------|--------------|----------------|
| 1      | 子の体調不良による遅刻… | やむをえない事情による場合… |
| ⋮      | ⋮            | ⋮              |

【図 7】

・子供が体調をくずしてしまった。  
・休みを取得するにはどのように…  
・定期券を失くしたので再発行方法を教えて…  
・電車が遅延して遅刻になってしまった。  
・どのように申請…

【図 8】

・子供が/体調/を/くずし/て/しまっ/た/。  
・休み/を/取得/する/に/は/どの/よう/に/…  
・定期券/を/失くした/ので/再発行/方法/を/教えて/…  
・電車/が/遅延/して/遅刻/に/なっ/て/しまっ/た/。  
・どの/よう/に/申請/…

【図 9】

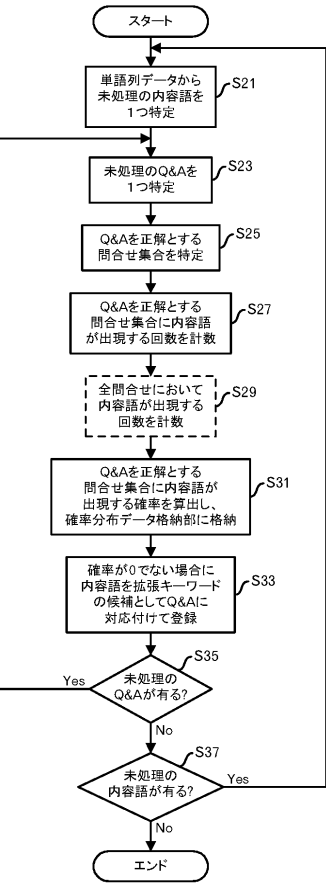
(a)  $\text{cnt}(\text{子供}) = 1, \text{cnt}(\text{しまっ}) = 2, \dots$

(b)  $\text{cnt}(\text{子供,が}) = 1, \text{cnt}(\text{の,子供}) = 1, \dots$

【図 10】

| u   | w  | $P(w u)$ |
|-----|----|----------|
| 再発行 | 方法 | 0.8      |
| 電車  | が  | 1        |
| 子供  | が  | 0.5      |
| ⋮   | ⋮  | ⋮        |

【図 1 1】



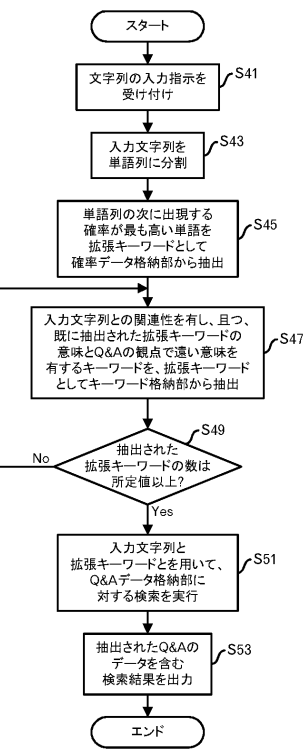
【図 1 2】

|    | Q&A1を正解とする問合せ集合 | Q&A2を正解とする問合せ集合 | Q&A3を正解とする問合せ集合 | ... |
|----|-----------------|-----------------|-----------------|-----|
| 風邪 | 1.0             | 0               | 0               | ... |
| 出産 | 0               | 0               | 1.0             | ... |
| 妻  | 0               | 0.33            | 0.66            | ... |
| ⋮  | ⋮               | ⋮               | ⋮               | ⋮   |

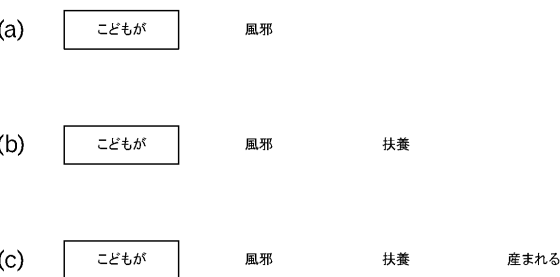
【図 1 3】

| Q&A | キーワード         |
|-----|---------------|
| 1   | こども,風邪,妻...   |
| 2   | こども,扶養,申請...  |
| 3   | こども,産まれる,妻... |
| ⋮   | ⋮             |

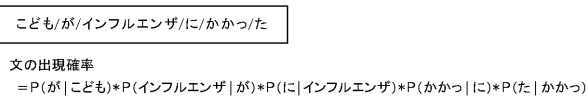
【図 1 4】



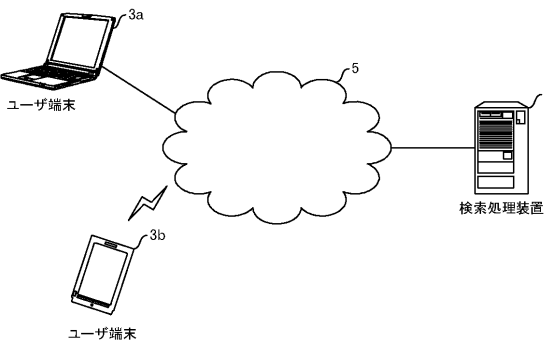
【図 1 5】



【図 1 6】



【図 1 7】



【図 18】

