

(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2016-133919

(P2016-133919A)

(43) 公開日 平成28年7月25日(2016.7.25)

(51) Int.Cl. F I テーマコード (参考)  
**G06F 17/30 (2006.01)** G06F 17/30 180A  
 G06F 17/30 220A

審査請求 未請求 請求項の数 7 O L (全 18 頁)

(21) 出願番号 特願2015-7130 (P2015-7130)  
 (22) 出願日 平成27年1月16日 (2015.1.16)

(71) 出願人 000004226  
 日本電信電話株式会社  
 東京都千代田区大手町一丁目5番1号  
 (74) 代理人 110001519  
 特許業務法人太陽国際特許事務所  
 (72) 発明者 西川 仁  
 東京都千代田区大手町一丁目5番1号 日  
 本電信電話株式会社内  
 (72) 発明者 貞光 九月  
 東京都千代田区大手町一丁目5番1号 日  
 本電信電話株式会社内  
 (72) 発明者 宮崎 千明  
 東京都千代田区大手町一丁目5番1号 日  
 本電信電話株式会社内

最終頁に続く

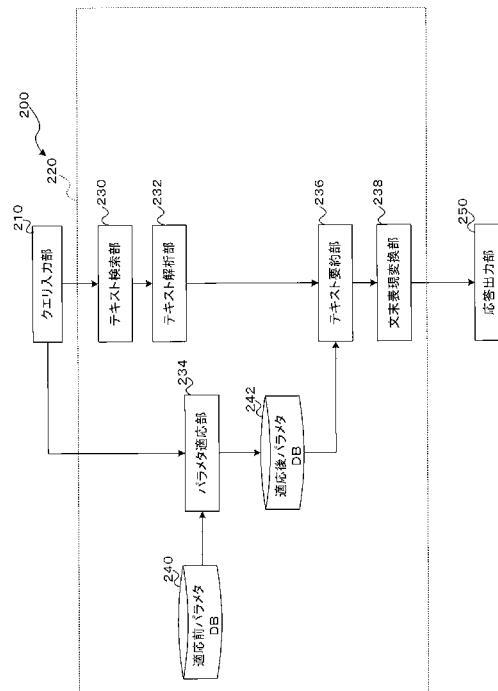
(54) 【発明の名称】 質問応答方法、装置、及びプログラム

(57) 【要約】

【課題】クエリに対する応答として適切な関連情報を提示できる。

【解決手段】クエリ入力部210が、クラスが<実体、属性、値>と定められた3つの要素の組み合わせからなるクエリを受け付け、テキスト検索部230が、受け付けたクエリに含まれる要素に関連するテキストを検索し、テキスト解析部が検索されたテキストを解析し、パラメタ適応部234が、クエリと、3つの要素の各々のクラスに関する特徴、及びその他の単語各々の特徴に対するパラメタとに基づいて、置き換えた適応後パラメタを得て、テキスト要約部236が、テキスト解析部232で解析されたテキストと、パラメタ適応部234で得られた適応後パラメタデータベース242とに基づいて、テキストについての要約文を生成し、文末表現変換部238で要約文に含まれる文末表現を、応答を表現するための文末表現に変換して、応答出力部250が変換された要約文を応答として出力する。

【選択図】図6



**【特許請求の範囲】****【請求項 1】**

ユーザから与えられたクエリに対する応答を生成する質問応答装置における質問応答方法であって、

クエリ入力部が、クラスが定められた複数の要素の組み合わせからなるクエリを受け付けるステップと、

テキスト検索部が、予め用意されたテキストデータベースから、前記クエリ入力部で受け付けた前記クエリに含まれる要素に関連するテキストを検索するステップと、

テキスト解析部が、前記テキスト検索部で検索されたテキストの構文を解析するステップと、

テキスト要約部が、前記テキスト解析部で解析されたテキストと、前記クエリの前記複数の要素の組み合わせとに基づいて、前記テキストについての要約文を生成するステップと、

応答出力部が、前記テキスト要約部で生成された前記要約文を応答として出力するステップと、

を含む質問応答方法。

**【請求項 2】**

パラメタ適応部が、前記クエリ入力部で受け付けた前記クエリと、予め学習された、前記複数の要素の各々のクラスに関する特徴を含む複数の特徴に対するパラメタとに基づいて、前記複数の要素の各々のクラスに関する特徴に対するパラメタを、前記クエリの前記複数の要素の各々に関する特徴に対するパラメタに置き換えた適応後パラメタを得るステップを更に含み、

前記テキスト要約部が、前記テキストについての要約文を生成するステップは、前記テキスト解析部で解析されたテキストと、前記パラメタ適応部で得られた適応後パラメタとに基づいて、前記テキストについての要約文を生成する請求項 1 に記載の質問応答方法。

**【請求項 3】**

文末表現変換部が、前記テキスト要約部で生成された要約文について、前記要約文に含まれる文末表現を、前記応答を表現するための文末表現に変換するステップを更に含み、

応答出力部が、要約文を応答として出力するステップは、前記文末表現変換部で変換された前記要約文を応答として出力する請求項 1 又は請求項 2 に質問応答方法。

**【請求項 4】**

ユーザから与えられたクエリに対する応答を生成する質問応答装置であって、

クラスが定められた複数の要素の組み合わせからなるクエリを受け付けるクエリ入力部と、

予め用意されたテキストデータベースから、前記クエリ入力部で受け付けた前記クエリに含まれる要素に関連するテキストを検索するテキスト検索部と、

前記テキスト検索部で検索されたテキストの構文を解析するテキスト解析部と、

前記テキスト解析部で解析されたテキストと、前記クエリの前記複数の要素の組み合わせとに基づいて、前記テキストについての要約文を生成するテキスト要約部と、

前記テキスト要約部で生成された前記要約文を応答として出力する応答出力部と、

を含む質問応答装置。

**【請求項 5】**

前記クエリ入力部で受け付けた前記クエリと、予め学習された、前記複数の要素の各々のクラスに関する特徴を含む複数の特徴に対するパラメタとに基づいて、前記複数の要素の各々のクラスに関する特徴に対するパラメタを、前記クエリの前記複数の要素の各々に関する特徴に対するパラメタに置き換えた適応後パラメタを得るパラメタ適応部を更に含み、

前記テキスト要約部は、前記テキスト解析部で解析されたテキストと、前記パラメタ適応部で得られた適応後パラメタとに基づいて、前記テキストについての要約文を生成する請求項 4 に記載の質問応答装置。

10

20

30

40

50

## 【請求項 6】

前記テキスト要約部で生成された要約文について、前記要約文に含まれる文末表現を、前記応答を表現するための文末表現に変換する文末表現変換部を更に含み、

応答出力部は、前記文末表現変換部で変換された前記要約文を応答として出力する請求項 4 又は請求項 5 に質問応答装置。

## 【請求項 7】

コンピュータに、請求項 1 ~ 請求項 3 に記載の質問応答方法を構成する各ステップを実行させるためのプログラム。

## 【発明の詳細な説明】

## 【技術分野】

10

## 【0001】

本発明は質問応答方法、装置、及びプログラムに係り、特に、質問に対する応答を生成するための質問応答方法、装置、及びプログラムに関する。

## 【背景技術】

## 【0002】

インターネット上には機械が再利用しやすい形でデータが蓄積されつつあり、それらのデータを用いることで人間に対して知的な応答を行う試みがなされている。これらの情報は様々な形で格納されているが、典型的には<実体、属性、値>の3つ組の形で保存されていることが多い。例えば、「鶴岡八幡宮の創建はいつですか?」というような情報は、<鶴岡八幡宮、創建、康平6年(1063年)>というような3つ組として表現することができる。このような情報として、既存のデータベースを用いることもできるし、事前に何らかの情報抽出などを行って得ることもできる。これら3つ組の情報が格納されたデータベースを利用する質問応答を、以降、データベース型質問応答と呼ぶ。

20

## 【0003】

これらのデータベースを用いることによって、例えば「鶴岡八幡宮の創建はいつですか?」というような質問から、<鶴岡八幡宮、創建、?>というような検索クエリを生成し、実体が「鶴岡八幡宮」、属性が「創建」のレコード<鶴岡八幡宮、創建、康平6年(1063年)>をデータベースから探してくることによって、「康平6年(1063年)」というような回答を生成することができる。

## 【先行技術文献】

30

## 【特許文献】

## 【0004】

【特許文献 1】特開 2013 - 250926 号公報

【特許文献 2】特開 2013 - 097723 号公報

## 【非特許文献】

## 【0005】

【非特許文献 1】徳永健伸、情報検索と言語処理、東京大学出版会、1999.

【非特許文献 2】Takeshi Fuchi and Shinichiro Takagi, "Japanese Morphological Analyzer using Word Co-occurrence - JTAG", In Proceedings of the 36th Annual Meeting of the Association for Computational Linguistics and 17th International Conference on Computational Linguistics (Volume 1), pp. 409-413, 1998.

40

【非特許文献 3】Kenji Imamura, Genichiro Kikui and Norihito Yasuda, "Japanese Dependency Parsing Using Sequential Labeling for Semi-spoken Language", In Proceedings of the 45th Annual Meeting of the Association for Computational Linguistics Companion Volume Proceedings of the Demo and Poster Sessions, pp. 225-228, 2007.

【非特許文献 4】Hitoshi Nishikawa, Kazuho Arita, Katsumi Tanaka, Tsutomu Hirao, Toshiro Makino and Yoshihiro Matsuo, "Learning to Generate Coherent Summary with Discriminative Hidden Semi-Markov Model", In Proceedings of the 25th International Conference on Computational Linguistics: Technical Papers, pp.1648-1659, 20

50

14.

## 【発明の概要】

## 【発明が解決しようとする課題】

## 【0006】

データベース型質問応答の方法は、ある特定の山の具体的な標高や、人名などの、典型的には固有名詞や数量表現などの、情報抽出の対象となる固有表現を回答することができる。その一方、例えば、どのような経緯で鶴岡八幡宮が創建されたかというような、付加的な情報を質問応答システムが回答することを考えると、そのような情報が格納されているデータベースが必要になる。しかし、一般に既存のデータベースに格納されている情報は上述のように固有名詞や数量表現などであり、例えば「康平6年(1063年)8月に河内国(大阪府羽曳野市)を本拠地とする河内源氏2代目の源頼義が、前九年の役での戦勝を祈願した京都の石清水八幡宮護国寺(あるいは河内源氏氏神の壺井八幡宮)を鎌倉の由比郷鶴岡(現材木座1丁目)に鶴岡若宮として勧請したのが始まりである。」(非特許文献5:日本語版 Wikipedia「鶴岡八幡宮」,[2014年10月24日検索],インターネット<URL:http://ja.wikipedia.org/wiki/%E9%B6%B4%E5%B2%A1%E5%85%AB%E5%B9%A1%E5%AE%AE>)というような自然文での値はデータベースには用意されていないことが多く、そのままでは適切な回答を生成することはできない。

10

## 【0007】

本発明は、上記問題点を解決するために成されたものであり、クエリに対する応答として適切な関連情報を提示できる質問応答装置、方法、及びプログラムを提供することを目的とする。

20

## 【課題を解決するための手段】

## 【0008】

上記目的を達成するために、第1の発明に係る質問応答方法は、ユーザから与えられたクエリに対する応答を生成する質問応答装置における質問応答方法であって、クエリ入力部が、クラスが定められた複数の要素の組み合わせからなるクエリを受け付けるステップと、テキスト検索部が、予め用意されたテキストデータベースから、前記クエリ入力部で受け付けた前記クエリに含まれる要素に関連するテキストを検索するステップと、テキスト解析部が、前記テキスト検索部で検索されたテキストの構文を解析するステップと、テキスト要約部が、前記テキスト解析部で解析されたテキストと、前記クエリの前記複数の要素の組み合わせとに基づいて、前記テキストについての要約文を生成するステップと、応答出力部が、前記テキスト要約部で生成された前記要約文を応答として出力するステップと、を含んで実行することを特徴とする。

30

## 【0009】

また、第1の発明に係る質問応答方法において、パラメタ適応部が、前記クエリ入力部で受け付けた前記クエリと、予め学習された、前記複数の要素の各々のクラスに関する特徴を含む複数の特徴に対するパラメタとに基づいて、前記複数の要素の各々のクラスに関する特徴に対するパラメタを、前記クエリの前記複数の要素の各々に関する特徴に対するパラメタに置き換えた適応後パラメタを得るステップを更に含み、前記テキスト要約部が、前記テキストについての要約文を生成するステップは、前記テキスト解析部で解析されたテキストと、前記パラメタ適応部で得られた適応後パラメタとに基づいて、前記テキストについての要約文を生成するようにしてもよい。

40

## 【0010】

また、第1の発明に係る質問応答方法において、文末表現変換部が、前記テキスト要約部で生成された要約文について、前記要約文に含まれる文末表現を、前記応答を表現するための文末表現に変換するステップを更に含み、応答出力部が、要約文を応答として出力するステップは、前記文末表現変換部で変換された前記要約文を応答として出力するようにしてもよい。

## 【0011】

第1の発明に係る質問応答装置は、ユーザから与えられたクエリに対する応答を生成す

50

る質問応答装置であって、クラスが定められた複数の要素の組み合わせからなるクエリを受け付けるクエリ入力部と、予め用意されたテキストデータベースから、前記クエリ入力部で受け付けた前記クエリに含まれる要素に関連するテキストを検索するテキスト検索部と、前記テキスト検索部で検索されたテキストの構文を解析するテキスト解析部と、前記テキスト解析部で解析されたテキストと、前記クエリの前記複数の要素の組み合わせとに基づいて、前記テキストについての要約文を生成するテキスト要約部と、前記テキスト要約部で生成された前記要約文を応答として出力する応答出力部と、を含んで構成されている。

【0012】

また、第1の発明に係る質問応答装置において、前記クエリ入力部で受け付けた前記クエリと、予め学習された、前記複数の要素の各々のクラスに関する特徴を含む複数の特徴に対するパラメタとに基づいて、前記複数の要素の各々のクラスに関する特徴に対するパラメタを、前記クエリの前記複数の要素の各々に関する特徴に対するパラメタに置き換えた適応後パラメタを得るパラメタ適応部を更に含み、前記テキスト要約部は、前記テキスト解析部で解析されたテキストと、前記パラメタ適応部で得られた適応後パラメタとに基づいて、前記テキストについての要約文を生成してもよい。

10

【0013】

また、第1の発明に係る質問応答装置において、前記テキスト要約部で生成された要約文について、前記要約文に含まれる文末表現を、前記応答を表現するための文末表現に変換する文末表現変換部を更に含み、応答出力部は、前記文末表現変換部で変換された前記要約文を応答として出力してもよい。

20

【0014】

第1の発明に係るプログラムは、コンピュータに、上記の質問応答方法を構成する各ステップを実行させるためのプログラムである。

【発明の効果】

【0015】

本発明の質問応答方法、装置、及びプログラムによれば、クラスが定められた複数の要素の組み合わせからなるクエリを受け付け、予め用意されたテキストデータベースから、クエリに含まれる要素に関連するテキストを検索し、検索されたテキストの構文を解析し、解析されたテキストと、クエリの複数の要素の組み合わせとに基づいて、テキストにつ

30

【図面の簡単な説明】

【0016】

【図1】本実施の形態に係る質問応答パラメタ学習装置の機能的な構成例を示すブロック図である。

【図2】訓練事例データベース8の一例を示す図である。

【図3】ある訓練事例の一例を示す図である。

【図4】特徴ベクトルの一例を示す図である。

【図5】適応前パラメタデータベース40の一例を示す図である。

40

【図6】本実施の形態に係る質問応答装置の機能的な構成例を示すブロック図である。

【図7】検索されたテキストの一例を示す図である。

【図8】解析されたテキストの一例を示す図である。

【図9】適応後パラメタデータベース40の一例を示す図である。

【図10】文末表現抽出装置の機能的な構成例を示すブロック図である。

【図11】文末表現変換部の機能的な構成例を示すブロック図である。

【図12】本実施の形態における質問応答パラメタ学習処理ルーチンを示すフローチャートである。

【図13】本実施の形態における質問応答処理ルーチン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】

50

【0017】

以下、図面を参照して本発明の実施の形態を詳細に説明する。

【0018】

<本発明の実施の形態に係る概要>

【0019】

まず、本発明の実施の形態における概要を説明する。

【0020】

本発明の実施の形態では、データベース型質問応答において、クラスが定められた複数の要素の組み合わせとして、3つ組<実体、属性、値>が与えられた際に、それに付随する応答を生成する方法を提案する。これによって、例えば「鶴岡八幡宮の創建はいつですか?」という質問がなされ、それに対して3つ組のクラスの各々に関する要素の組み合わせ<鶴岡八幡宮、創建、康平6年(1063年)>が回答として得られたときに、「康平6年(1063年)8月に河内国(大阪府羽曳野市)を本拠地とする河内源氏2代目の源頼義が、前9年の役での戦勝を祈願した京都の石清水八幡宮護国寺(あるいは河内源氏氏神の壺井八幡宮)を鎌倉の由比郷鶴岡(現材木座1丁目)に鶴岡若宮として勧請したのが始まりである。」(上記非特許文献5参照)という関連情報を生成することができるようになる。また、これによって、質問応答システムの利用者に対して、単なる端的な回答だけでなく、関連する情報も合わせて提示することができるようになる。なお、本実施の形態では、クラスが定められた3つの要素の組み合わせを用いた場合を例に説明するが、これに限定されるものではなく、クラスが定められているものであれば3つ以上の要素の組み合わせを用いて、応答を生成するようにしてもよい。

10

20

【0021】

<本発明の実施の形態に係る質問応答パラメタ学習装置の構成>

【0022】

次に、本発明の実施の形態に係る質問応答パラメタ学習装置の構成について説明する。図1に示すように、本発明の実施の形態に係る質問応答パラメタ学習装置100は、CPUと、RAMと、後述する質問応答パラメタ学習処理ルーチンを実行するためのプログラムや各種データを記憶したROMと、を含むコンピュータで構成することが出来る。この質問応答パラメタ学習装置100は、機能的には図1に示すように入力部10と、演算部20とを備えている。

30

【0023】

入力部10は、訓練事例データベース8の入力を受け付ける。訓練事例データベース8は3つの要素の組み合わせからなるクエリと、テキストと、当該テキストを要約して作成された、クエリに付随する説明の3つを組とした1つ以上のレコードからなるデータベースである。図2に訓練事例データベース8の一例を示す。図2の事例番号1は、3つの要素の組み合わせからなるクエリ「<建長寺、創建、建長5年(1253年)>」、「建長寺」に関するテキスト、及び、クエリに付随する説明の3つからなる。

【0024】

演算部20は、パラメタ推定部30と、適応前パラメタデータベース40とを含んで構成されている。

40

【0025】

パラメタ推定部30は、入力部10で受け付けた訓練事例データベース8に基づいて、パラメタを学習し、適応前パラメタデータベース40として記憶する。パラメタの学習には、一例として上記特許文献2記載のテキスト要約方法を用いることができる。

【0026】

パラメタ推定部30は、具体的には、まず、入力された訓練事例データベース8の訓練事例の各々、すなわち3つの要素の組み合わせからなるクエリ(以下、qとする)、テキスト(以下、xとする)、及び説明(以下、yとする)に対してそれぞれ形態素解析および係り受け解析を行う。なお、形態素解析に非特許文献2記載の方法を、係り受け解析に非特許文献3記載の方法を利用することができる。

50

## 【0027】

次に、パラメタ推定部30は、訓練事例の各々の解析の結果から、クエリに含まれる3つの要素の各々のクラスに関する特徴を要素として含む特徴ベクトルを構築する。特徴ベクトル $f(x; y, q)$ は、一例として、テキスト $x$ 及びクエリ $q$ に含まれている単語の各々に対する、当該単語が、説明 $y$ に含まれている場合に1、そうでない場合に0となる要素から構成される。なお、ベクトルを表す記号には、記号の前に「 $\wedge$ 」を付して表現する。

## 【0028】

ここで、パラメタ推定部30では、特徴ベクトル $f(x; y, q)$ を構築する際に、クエリの<実体、属性、値>を構成する単語については、単に単語として扱うのではなく、クエリの<実体、属性、値>といったクラスとして扱う。例を図3及び図4に示す。図3に示す訓練事例では、テキスト $x$ に含まれる「京都」という単語は説明 $y$ に含まれないため、図4に示す特徴ベクトルでは、対応する特徴ベクトルの要素は0となっている。一方、単語「落慶」はいずれにも含まれるため、対応する特徴ベクトルの要素は1となっている。また、クエリに含まれる<実体、属性、値>を構成する単語については、それぞれ一例として「単語-<実体>」、「単語-<属性>」、「単語-<値>」という要素を用意し、これを説明 $y$ が含むか否かに基づいて、ベクトルの値を決定する。例えば、クエリの実体である単語「建長寺」は説明 $y$ に含まれるため、対応する「単語-<実体>」のベクトルの要素は1となる。同様に、「単語-<属性>」である「創建」、「単語-<値>」である「建長5年(1253年)」も説明 $y$ に含まれるため、対応するベクトルの要素はそれぞれ1となる。なお、ここで、例えば「建長5年(1253年)」などの複数の単語から構成される表現がクエリの実体や属性、値であった場合は、表現が完全一致している場合だけでなく、この例では「建長」などのクエリの実体や属性、値を構成する単語が、部分的に説明 $y$ に含まれている場合でも、ベクトルの要素を1としてもよい。

## 【0029】

次に、パラメタ推定部30は、訓練事例の各々について構築された特徴ベクトルを用いて、クエリに含まれる3つの要素(単語)の各々のクラスに関する特徴(「単語-<実体>」、「単語-<属性>」、「単語-<値>」が説明 $y$ に含まれている否か)、及びその他の単語の特徴(単語が説明 $y$ に含まれている否か)の各々に対するパラメタの学習を行う。パラメタの学習についても、上記特許文献2記載の方法を用いて行うことができる。学習の際に用いられる要約方法としては、上記非特許文献4記載の方法を用いることができる。そして、学習したパラメタを適応前パラメタデータベース40として記憶する。

## 【0030】

適応前パラメタデータベース40は、パラメタ推定部30において訓練事例データベース4を用いて機械学習を行った結果として得られるものである。図5に例を示す。図5では、一例として、クエリの「単語-<実体>」、「単語-<属性>」、「単語-<値>」の各々に対する重み、及びその他の単語の各々に対する重みが、パラメタとして格納されている。

## 【0031】

<本発明の実施の形態に係る質問応答装置の構成>

## 【0032】

次に、本発明の実施の形態に係る質問応答装置の構成について説明する。図1に示すように、本発明の実施の形態に係る質問応答装置200は、CPUと、RAMと、後述する質問応答処理ルーチンを実行するためのプログラムや各種データを記憶したROMと、を含むコンピュータで構成することが出来る。この質問応答装置200は、機能的には図6に示すようにクエリ入力部210と、演算部220と、応答出力部250とを備えている。

## 【0033】

クエリ入力部210は、クラスが<実体、属性、値>と定められた3つの要素の組み合わせからなるクエリを受け付ける。このクエリはデータベース型質問応答システムに対し

10

20

30

40

50

て入力された何らかの質問に対して、データベース型質問応答システムが出力した応答である。本実施の形態では、例えば、「鶴岡八幡宮の創建はいつ？」というような質問に対して、データベース型質問応答システムが生成した3つの要素の組み合わせとして、＜鶴岡八幡宮、創建、康平6年（1063年）＞からなるクエリを入力として受け付ける。なお、データベース型質問応答システムとしては、例えば上記特許文献1記載の方法（対象抽出と属性判定に基づく質問応答装置）を用いることができる。

#### 【0034】

演算部220は、テキスト検索部230と、テキスト解析部232と、パラメタ適応部234と、テキスト要約部236と、文末表現変換部238と、適応前パラメタデータベース240と、適応後パラメタデータベース242とを含んで構成されている。

10

#### 【0035】

適応前パラメタデータベース240には、適応前パラメタデータベース40と同じものが記憶されている。

#### 【0036】

テキスト検索部230は、予め用意されたテキストデータベースから、クエリ入力部210で受け付けたクエリに含まれる特定の要素に関連するテキストを検索する。

#### 【0037】

テキスト検索部230は、具体的には、クエリ入力部210で受け付けたクエリを用いて、テキスト検索部230に予め用意された何らかのデータベースに格納されている1つ以上のテキストを検索する。一例として、日本語版ウィキペディアの全ての記事が格納されたデータベースを予め用意しておき、その中から、入力された3つの要素の組み合わせのうち、＜実体＞の要素と完全一致する見出しを持つテキストを検索することができる。例えば、3つの要素の組み合わせとして、＜鶴岡八幡宮、創建、康平6年（1063年）＞というクエリが入力された場合、データベースの中から「鶴岡八幡宮」の見出しを持つテキストを検索する。検索には、一例として上記非特許文献1記載の、既知の検索方法を利用することができる。図7に、テキスト検索部230で検索されたテキストの一例を示す。

20

#### 【0038】

テキスト解析部232は、テキスト検索部230で検索されたテキストの構文を解析する。テキスト解析部232は、具体的には、テキスト検索部230で検索され出力されたテキストを受け取り、当該テキストに対して形態素解析および係り受け解析の処理を加える。それぞれ、形態素解析に非特許文献2記載の方法を、係り受け解析に上記非特許文献3記載の方法を利用することができる。図8に、テキスト解析部で解析されたテキストの一例を示す。図8では、テキスト検索部230で検索されたテキストに対して解析が行われている。図8のmから始まる行は形態素解析の結果として得られた、わかち書き及び形態素の品詞情報、cから始まる行は、係り受け解析の結果として得られた文節及び分節の係り受けの情報である。この例では、係り受け解析器は「康平6年（1063年）」という文節が次の「8月に」という文節にかかり、また「8月に」という文節は「8月」という形態素と「に」という形態素に分かれることが示されている。

30

#### 【0039】

パラメタ適応部234は、クエリ入力部210で受け付けた3つの要素の組み合わせからなるクエリと、適応前パラメタデータベース240に記憶されている3つの要素の各々のクラスに関する特徴、及びその他の単語各々の特徴に対するパラメタとに基づいて、3つの要素の各々のクラスに関する特徴に対するパラメタを、クエリの3つの要素の各々に関する特徴に対するパラメタに置き換えた適応後パラメタを得て、適応後パラメタデータベース242として記憶する。

40

#### 【0040】

パラメタ適応部234は、具体的には、適応前パラメタデータベース240に「単語-＜実体＞」、「単語-＜属性＞」、「単語-＜値＞」と抽象化された形で記憶されている特徴を、クエリ入力部210で受け付けたクエリに合わせて書き換える。例えば、3つの要素の組み合わせからなるクエリが＜鶴岡八幡宮、創建、康平6年（1063年）＞であった場

50



合、適応前パラメータデータベース240の「単語-＜実体＞」は「鶴岡八幡宮」に、「単語-＜属性＞」は「創建」に、「単語-＜値＞」は「康平6年（1063年）」にそれぞれ置き換えられる。このように置き換えることで、クエリ入力部210で受け付けたクエリの3つの要素の組み合わせに合わせて、パラメータデータベースを適応させる。

【0041】

適応後パラメータデータベース242は、パラメータ適応部234から出力されたパラメータを格納するデータベースである。図9に示す例では、クエリの要素のクラス「実体」、「属性」、「値」の重みが、クエリに適応するように書き換えられて記憶されている。例えば、「単語-＜実体＞」は「鶴岡八幡宮」に、「単語-＜属性＞」は「創建」に、「単語-＜値＞」は「康平6年（1063年）」にそれぞれ置き換えられている。

10

【0042】

テキスト要約部236は、テキスト解析部232で解析されたテキストと、パラメータ適応部234で得られた適応後パラメータデータベース242に記憶された適応後パラメータに基づいて、テキストについての要約文を生成する。要約文の生成方法は、一例として、非特許文献4記載の方法を用いて、テキストから一文を選択し、選択した一文を要約文として生成すればよい。

【0043】

文末表現変換部238は、以下に説明するように、テキスト要約部236で生成された要約文について、要約文に含まれる文末表現を、応答を表現するための文末表現に変換する。

20

【0044】

文末表現変換部238における文末表現の変換の一例としては、要約文の文末を敬体に変換することである。例えば、「康平6年（1063年）8月に河内国（大阪府羽曳野市）を本拠地とする河内源氏2代目の源頼義が、前九年の役での戦勝を祈願した京都の石清水八幡宮護国寺（あるいは河内源氏氏神の壺井八幡宮）を鎌倉の由比郷鶴岡（現材木座1丁目）に鶴岡若宮として勧請したのが始まりである。」という応答の文末は常体であり、口語的ではないため、質問応答システムの応答としては、場合によっては不適當である。そのような場合は、「康平6年（1063年）8月に河内国（大阪府羽曳野市）を本拠地とする河内源氏2代目の源頼義が、前九年の役での戦勝を祈願した京都の石清水八幡宮護国寺（あるいは河内源氏氏神の壺井八幡宮）を鎌倉の由比郷鶴岡（現材木座1丁目）に鶴岡若宮として勧請したのが始まりです。」のように文末を敬体に変換する。文末表現の変換には具体的には、以下の方法を用いることができる。

30

【0045】

文末表現変換部238は、例えば、文末表現抽出装置において抽出した特徴的文末表現リストデータベースに基づいて、テキスト要約部236で生成された要約文の文末表現を、応答を表現する文末表現に変換することができる。

【0046】

[文末表現抽出装置の処理]

【0047】

文末表現抽出装置1000は、機能的には、図10に示すように、リスト生成用入力部1001と、リスト生成用演算部1002と、対話行為毎の特徴的文末表現リストデータベース1003とを含んだ構成で表すことができる。以下、各部について詳述する。

40

【0048】

リスト生成用入力部1001は、発話の意図を表す対話行為の種類が各々付与された複数の対話テキストを受け付ける。対話テキストは、1つ以上の文を含む文書である。また、対話行為の種類（タグセット）は、（1）情報提供、（2）自己開示（事実、経験、欲求等をサブタグとして持つ）、（3）質問（事実、経験、欲求、自問等をサブタグとして持つ）、（4）共感・同意、非共感・非同意であるが、本実施の形態では、（1）情報提供に係る特徴的文末表現リストのみを抽出する。なお、本実施の形態では、応答を表現する文末表現として、情報提供を表現する文末表現を用いる場合を例に説明する。

50

## 【0049】

リスト生成用演算部1002は、対話行為の種類が情報提供である場合に関する特徴的文末表現リストを生成する。リスト生成用演算部1002は、対話行為付与済テキストデータベース1020と、リスト生成部1022とを備えている。

## 【0050】

対話行為付与済テキストデータベース1020には、リスト生成用入力部1001により受け付けた複数の対話テキストが格納される。

## 【0051】

リスト生成部1022は、文末表現抽出部1024と、特徴的文末表現抽出部1026と、文末表現拡張部1028と、不要文末表現フィルタ部1030とを備えている。

10

## 【0052】

文末表現抽出部1024は、対話行為付与済テキストデータベース1020に格納された複数の対話テキストの各々について、当該対話テキストから、文末表現を抽出し、文末表現リストを生成する。また、文末表現抽出部1024は、対話行為付与済テキストデータベース1020に格納された複数の対話テキストに基づいて、各文末表現に対応する対話行為の種類に関する対話行為リストを生成する。

## 【0053】

特徴的文末表現抽出部1026は、文末表現抽出部1024によって、複数の対話テキストから抽出された文末表現から、当該情報提供に関する対話テキストにおける文末表現の出現頻度と、対話行為付与済テキストデータベース1020に格納されている、当該情報提供とは異なる対話行為の種類が付与された対話テキストにおける文末表現の出現頻度とに基づいて、対話における情報提供を表現するための文末表現を抽出する。

20

## 【0054】

文末表現拡張部1028は、特徴的文末表現抽出部1026によって生成された特徴的文末表現候補リストと、予め用意された動詞・形容詞の活用語尾リストとに基づいて、特徴的文末表現抽出部1026によって抽出された対話における情報提供を表現するための文末表現の各々に対して、文末表現の活用語尾を変化させて、対話における情報提供を表現するための文末表現として、活用型が異なる複数の文末表現を生成する。

## 【0055】

不要文末表現フィルタ部1030は、文末表現拡張部1028によって生成された活用型拡張文末表現リストに基づいて、対話における情報提供を表現するための文末表現から、予め定められた言いさし表現、及び予め定められた文脈に依存する表現の少なくとも一方を含む文末表現を削除して、情報提供に関する文末表現のリストを、特徴的文末表現リストデータベース1003に記憶する。

30

## 【0056】

[文末表現変換部の処理]

## 【0057】

文末表現変換部238は、機能的には、図11に示すように、入力部2004と、対話行為毎の特徴的文末表現リストデータベース2005と、演算部2006と、出力部2007とを含んだ構成で表すことができる。以下、各部について詳述する。

40

## 【0058】

入力部2004は、文末表現変換対象のテキストとして、要約文を受け付ける。

## 【0059】

特徴的文末表現リストデータベース2005には、文末表現抽出装置1000の特徴的文末表現リストデータベース1003に格納されている特徴的文末表現リストと同じ情報提供に関する特徴的文末表現リストが格納される。

## 【0060】

演算部2006は、文末表現抽出部2060と、文末表現候補生成部2062と、文末表現変換部2064とを備えている。

50

## 【0061】

文末表現抽出部2060は、入力部2004により受け付けた要約文から、文末表現及び当該文末表現の直前に出現する形態素の品詞を抽出する。

## 【0062】

文末表現候補生成部2062は、特徴的文末表現リストデータベース2005に格納された、情報提供に関する特徴的文末表現リストにおける、文末表現の直前に出現する形態素の品詞と、文末表現抽出部2060によって抽出された文末表現の直前の形態素の品詞とを照合する。そして、文末表現候補生成部2062は、情報提供に関する特徴的文末表現リストから、文末表現抽出部2060によって抽出された文末表現の直前の形態素の品詞に対応する、情報提供を表現するための文末表現を、文末表現の変換候補として抽出し、抽出された文末表現からなる変換候補リストを生成する。

10

## 【0063】

文末表現変換部2064は、文末表現候補生成部2062によって生成された変換候補リストを用いて、要約文に含まれる文末表現を、対話における情報提供を表現するための文末表現に変換することにより、応答を表現するための文末表現に変換する。

## 【0064】

## 【0065】

そして、応答出力部250は、文末表現変換部238で変換された要約文を応答として出力する。

## 【0066】

<本発明の実施の形態に係る質問応答パラメタ学習装置の作用>

20

## 【0067】

次に、本発明の実施の形態に係る質問応答パラメタ学習装置100の作用について説明する。入力部10において、3つの要素の組み合わせからなるクエリと、テキストと、当該テキストを要約して作成された、クエリに付随する説明の3つを組とした1つ以上のレコードを受け付けると、訓練事例データベース8に格納し、質問応答パラメタ学習装置100は、図12に示す質問応答パラメタ学習処理ルーチンを実行する。

## 【0068】

まず、ステップS100では、訓練事例データベース8から訓練事例として1つのレコードを選択する。

30

## 【0069】

次に、ステップS102では、ステップS100で選択した訓練事例について、クエリ $q$ 、テキスト $x$ 、及び説明 $y$ に対してそれぞれ形態素解析および係り受け解析を行う。

## 【0070】

ステップS104では、ステップS102で解析したクエリ $q$ 、テキスト $x$ 、及び説明 $y$ に基づいて、特徴ベクトル $f(x; y, q)$ を構築する。

## 【0071】

ステップS106では、全ての訓練事例について、ステップS102～ステップS104の処理を終了したかを判定し、終了していなければステップS100へ戻って処理を繰り返し、終了していればステップS108へ移行する。

40

## 【0072】

ステップS108では、ステップS104で訓練事例の各々について構築された特徴ベクトルを用いて、クエリに含まれる3つの要素(単語)の各々のクラスに関する特徴、及びその他の単語の特徴の各々に対するパラメタの学習し、適応前パラメタデータベース40として記憶し処理を終了する。

## 【0073】

以上説明したように、本実施の形態に係る質問応答パラメタ学習装置によれば、3つの要素の組み合わせからなるクエリと、テキストと、当該テキストを要約して作成された、クエリに付随する説明の3つを組とした1つ以上のレコードからなる訓練事例データベース8に基づいて、パラメタを学習することにより、クエリに対する応答として適切な関連

50

情報を提示するためのパラメタを学習できる、という効果が得られる。

【0074】

<本発明の実施の形態に係る質問応答装置の作用>

【0075】

次に、本発明の実施の形態に係る質問応答装置200の作用について説明する。クエリ入力部210において3つの要素の組み合わせからなるクエリを受け付けると、質問応答装置200は、図13に示す質問応答処理ルーチンを実行する。

【0076】

まず、ステップS200では、クエリ入力部210において受け付けた3つの要素の組み合わせからなるクエリを取得する。

【0077】

次に、ステップS202では、予め用意されたテキストデータベースから、ステップS200で取得したクエリに含まれる特定の要素に関連するテキストを検索する。

【0078】

ステップS204では、ステップS202で検索されたテキストの構文を解析する。

【0079】

ステップS206では、ステップS200で取得したクエリと、適応前パラメタデータベース240に記憶されているパラメタとに基づいて、3つの要素の各々のクラスに関する特徴に対するパラメタを、ステップS200で取得したクエリの3つの要素の各々に関する特徴に対するパラメタに置き換えた適応後パラメタを得る。

【0080】

ステップS208では、ステップS206で解析されたテキストと、ステップS208で得られた適応後パラメタとに基づいて、テキストについての要約文を生成する。

【0081】

ステップS210では、ステップS208で生成された要約文について、要約文に含まれる文末表現を、応答を表現するための文末表現に変換する。

【0082】

そして、ステップS212では、ステップS210で文末表現が変換された要約文を応答として出力して処理を終了する。

【0083】

以上説明したように、本実施の形態に係る質問応答装置によれば、クラスが<実体、属性、値>と定められた3つの要素の組み合わせからなるクエリを受け付け、受け付けたクエリの特定の要素に関連するテキストを検索し、検索されたテキストを解析し、クエリと、3つの要素の各々のクラスに関する特徴、及びその他の単語各々の特徴に対するパラメタとに基づいて、置き換えた適応後パラメタを得て、テキスト解析部232で解析されたテキストと、適応後パラメタデータベース242とに基づいて、テキストについての要約文を生成し、要約文に含まれる文末表現を、応答を表現するための文末表現に変換して、応答として出力することにより、クエリに対する応答として適切な関連情報を提示できる、という効果が得られる。

【0084】

なお、本発明は、上述した実施の形態に限定されるものではなく、この発明の要旨を逸脱しない範囲内で様々な変形や応用が可能である。

【0085】

例えば、上述した実施の形態では、要約部236において、テキストから一文を選択し、選択した一文を要約文として生成したが、これに限定されるものではなく、選択した一文が一定文字数以下となるように不要な文字列を除去する等して要約文を生成してもよい。また、テキストから、一定文字数以下となるように、1つ以上の文を選択して要約文を生成する等、いかなる方法で要約文を生成してもよい。

【符号の説明】

【0086】

10

20

30

40

50

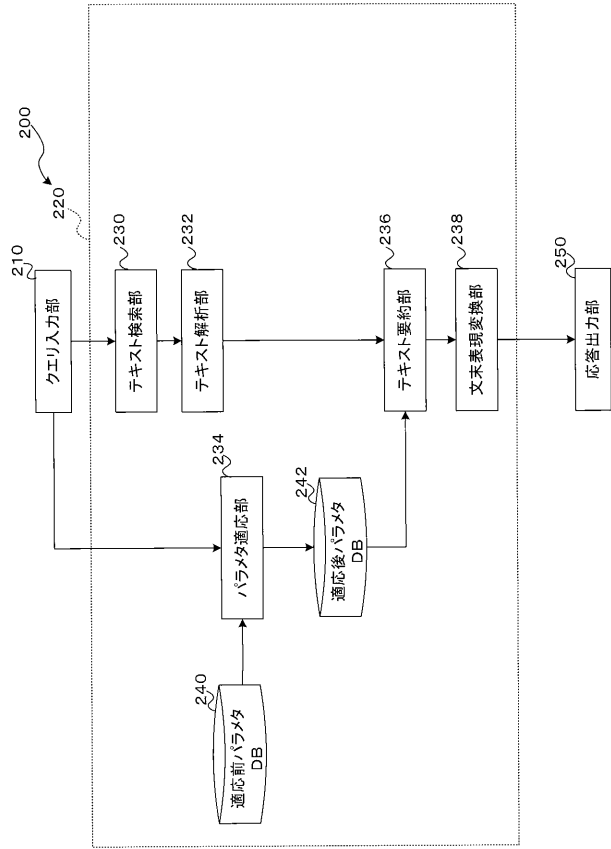
8	訓練事例データベース	
10、2004	入力部	
20、220、2006	演算部	
30	パラメタ推定部	
36	テキスト要約部	
38	文末表現変換部	
40	適応前パラメタデータベース	
42	適応後パラメタデータベース	
100	質問応答パラメタ学習装置	
200	質問応答装置	10
210	クエリ入力部	
230	テキスト検索部	
232	テキスト解析部	
234	パラメタ適応部	
236	テキスト要約部	
238	文末表現変換部	
240	適応前パラメタデータベース	
242	適応後パラメタデータベース	
250	応答出力部	
1000	文末表現抽出装置	20
1001	リスト生成用入力部	
1002	リスト生成用演算部	
1003、2005	特徴的文末表現リストデータベース	
1020	対話行為付与済テキストデータベース	
1022	リスト生成部	
1024、2060	文末表現抽出部	
1026	特徴的文末表現抽出部	
1028	文末表現拡張部	
1030	不要文末表現フィルタ部	
2007	出力部	30
2062	文末表現候補生成部	
2064	文末表現変換部	



【 図 5 】

特徴量	重み
単語<値>	10
単語<属性>	3
単語<実体>	2
当時	-1
日本	0.1
京都	0.05
仏教	0.1
帰依	0.1
...	...

【 図 6 】



【 図 7 】

康平6年(1063年)8月に河内国(大阪府羽曳野市)を本拠地とする河内源氏2代目の源頼義が、前九年の役での戦勝を祈願した京都の石清水八幡宮護国寺(あるいは河内源氏神の壺井八幡宮)を鎌倉の由比郷鶴岡(現材木座1丁目)に鶴岡若宮として勧請したのが始まりである。永保元年(1081年)2月には河内源氏3代目の源義家(八幡太郎義家)が修復を加えた。

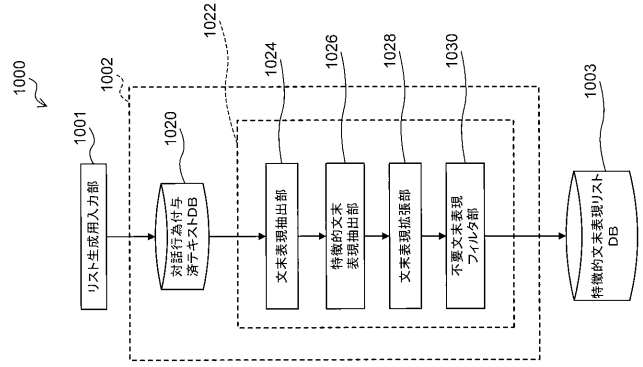
【 図 8 】

- c: 0 1D 4/4
- m: 康平 名詞:名:固有
- m: 6年 名詞:時:連用
- m: ( 括弧:開括弧
- m: 1063 Number
- m: 年 助数詞
- m: ) 括弧:閉括弧
- c: 1 5D 0/1
- m: 8月 名詞:日時:連用
- m: に 格助詞:連用
- c: 2 3D 1/1
- m: 河内 名詞:固有:地
- m: 国 名詞接尾辞:名詞
- c: 3 5D 4/6
- m: ( 括弧:開括弧
- m: 大阪 名詞:固有:地
- m: 府 名詞接尾辞:名詞
- m: 羽曳野 名詞:固有:地
- m: 市 名詞接尾辞:名詞
- m: ) 括弧:閉括弧
- m: を 格助詞:連用
- c: 4 5D 0/1
- m: 本拠地 名詞
- m: と 判定詞:連用
- c: 5 11D 0/0
- m: する 動詞語幹:SUJU
- c: 6 7D 2/2
- m: 河内 名詞:固有:姓
- m: 源 名詞:名:固有
- m: 氏 名詞接尾辞:名詞
- c: 7 8D 1/2
- m: 2 Number
- m: 代目 助数詞
- m: の 格助詞:連体
- c: 8 11D 1/2
- m: 源 名詞:固有:姓
- m: 頼義 名詞:名:固有
- m: が 格助詞:連用
- m: 、 読点
- c: 9 10D 0/1
- m: 前九年の役 名詞:固有
- m: での 格助詞:連体
- .....

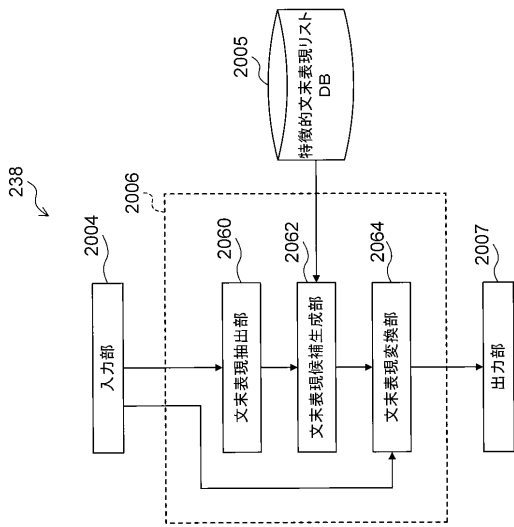
【図9】

特徴量	重み
鶴岡八幡宮	10
創建	3
康平6年(1063年)	2
当時	-1
日本	0.1
京都	0.05
仏教	0.1
帰依	0.1
...	...

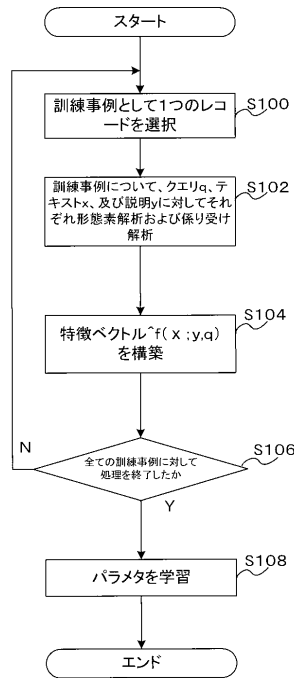
【図10】



【図11】

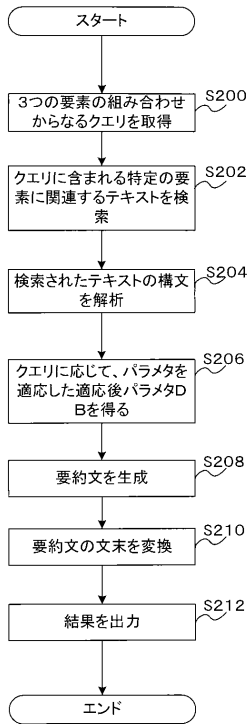


【図12】





【 図 1 3 】



---

フロントページの続き

- (72)発明者 浅野 久子  
東京都千代田区大手町一丁目5番1号 日本電信電話株式会社内
- (72)発明者 松尾 義博  
東京都千代田区大手町一丁目5番1号 日本電信電話株式会社内