

(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2007-199876

(P2007-199876A)

(43) 公開日 平成19年8月9日(2007.8.9)

(51) Int. Cl. F I テーマコード (参考)  
**G06F 17/30 (2006.01)** G06F 17/30 350C 5B075  
 G06F 17/30 180A

審査請求 未請求 請求項の数 15 O L (全 14 頁)

(21) 出願番号 特願2006-15719 (P2006-15719)  
 (22) 出願日 平成18年1月24日 (2006.1.24)

(71) 出願人 000005496  
 富士ゼロックス株式会社  
 東京都港区赤坂九丁目7番3号  
 (74) 代理人 100087480  
 弁理士 片山 修平  
 (72) 発明者 山根 洋平  
 神奈川県足柄上郡中井町境430 グリー  
 ンテクなかい 富士ゼロックス株式会社内  
 Fターム(参考) 5B075 ND03 NK02 NR12 PR06 QM10

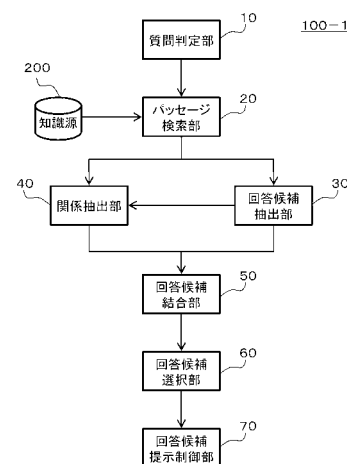
(54) 【発明の名称】 質問応答システム、質問応答処理方法及び質問応答プログラム

(57) 【要約】

【課題】 回答候補間の関係を考慮した適切な回答候補の提示が可能な質問応答システム、質問応答処理方法及び質問応答プログラムを提供する。

【解決手段】 質問応答システム100-1は、検索質問文から検索語を抽出する質問判定部10と、検索語に基づいて知識源200からパッセージを取得するパッセージ検索部20と、パッセージから回答候補を抽出して、回答候補の尤度を表す得点を付与する回答候補抽出部30と、回答候補が表す事物の同一性についての回答候補間の関係を示す関係規則情報を取得する関係抽出部40と、関係規則情報に基づいて同一の事物を表す回答候補を結合して回答候補集合を生成し、該回答候補集合に含まれる回答候補に付与された得点を集計する回答候補結合部50と、得点が最も大きい回答候補集合に含まれる回答候補を優先的に提示する制御を行う回答候補選択部60及び回答候補提示制御部70とを有する。

【選択図】 図3



## 【特許請求の範囲】

## 【請求項 1】

入力した検索質問文に応じて、検索対象文を含む第 1 の知識源を検索し、前記検索質問文の回答となり得る回答候補を得る質問応答システムであって、

前記検索質問文から検索語を抽出する検索語抽出手段と、

前記検索語に基づいて前記第 1 の知識源に含まれる検索対象文の中から検索結果の文を取得する検索手段と、

前記検索結果の文から前記回答候補を抽出し、前記回答候補毎に該回答候補の尤度を表す数値を付与する回答候補抽出手段と、

前記回答候補が表す事物の同一性についての回答候補間の関係を示す関係情報を生成する関係情報生成手段と、 10

前記抽出された回答候補から、前記関係情報に基づいて、同一の事物を表す回答候補からなる回答候補集合を生成し、前記回答候補集合毎に、該回答候補集合に含まれる回答候補に付与された尤度の数値を集計する回答候補結合手段と、

前記尤度の集計値が最も大きい前記回答候補集合に含まれる回答候補のいずれかを優先的に提示する制御を行う提示制御手段とを有することを特徴とする質問応答システム。

## 【請求項 2】

前記関係情報生成手段は、1 の回答候補を含む検索結果の文における単語のベクトルと、他の回答候補を含む検索結果の文における単語のベクトルとの内積を算出し、該内積が所定値以上である場合に、前記 1 及び他の回答候補が同一の事物を表すものであることを示す関係情報を生成することを特徴とする請求項 1 に記載の質問応答システム。 20

## 【請求項 3】

前記関係情報生成手段は、同一の事物を表す回答候補のいずれかがこれら回答候補の代表となる回答候補であることを示す関係情報を生成し、

前記回答候補結合手段は、前記同一の事物を表す回答候補からなり、前記代表となる回答候補に他の回答候補を置き換えた回答候補集合を生成し、

前記提示制御手段は、前記代表となる回答候補を優先的に提示する制御を行うことを特徴とする請求項 1 又は 2 に記載の質問応答システム。

## 【請求項 4】

前記関係情報生成手段は、前記第 1 の知識源とは異なる第 2 の知識源に含まれる文に基づいて、前記回答候補が表す事物の同一性についての回答候補間の関係を示す関係情報を生成することを特徴とする請求項 1 乃至 3 のいずれかに記載の質問応答システム。 30

## 【請求項 5】

前記関係情報生成手段により生成された関係情報を保持する関係情報保持手段を有することを特徴とする請求項 1 乃至 4 のいずれかに記載の質問応答システム。

## 【請求項 6】

入力した検索質問文に応じて、検索対象文を含む第 1 の知識源を検索し、前記検索質問文の回答となり得る回答候補を得る質問応答システムでの処理方法であって、

前記検索質問文から検索語を抽出する検索語抽出ステップと、

前記検索語に基づいて前記第 1 の知識源に含まれる検索対象文の中から検索結果の文を取得する検索ステップと、 40

前記検索結果の文から前記回答候補を抽出し、前記回答候補毎に該回答候補の尤度を表す数値を付与する回答候補抽出ステップと、

前記回答候補が表す事物の同一性についての回答候補間の関係を示す関係情報を生成する関係情報生成ステップと、

前記抽出された回答候補から、前記関係情報に基づいて、同一の事物を表す回答候補からなる回答候補集合を生成し、前記回答候補集合毎に、該回答候補集合に含まれる回答候補に付与された尤度の数値を集計する回答候補結合ステップと、

前記尤度の集計値が最も大きい前記回答候補集合に含まれる回答候補のいずれかを優先的に提示する制御を行う提示制御ステップとを有することを特徴とする処理方法。 50

**【請求項 7】**

前記関係情報生成ステップは、1の回答候補を含む検索結果の文の集合における単語のベクトルと、他の回答候補を含む検索結果の文の集合における単語のベクトルとの内積を算出し、該内積が所定値以上である場合に、前記1及び他の回答候補が同一の事物を表すものであることを示す関係情報を生成することを特徴とする請求項6に記載の処理方法。

**【請求項 8】**

前記関係情報生成手段は、同一の事物を表す回答候補のいずれかがこれら回答候補の代表となる回答候補であることを示す関係情報を生成し、

前記回答候補結合ステップは、前記同一の事物を表す回答候補からなり、前記代表となる回答候補に他の回答候補を置き換えた回答候補集合を生成し、

前記提示制御ステップは、前記代表となる回答候補を優先的に提示する制御を行うことを特徴とする請求項6又は7に記載の処理方法。

**【請求項 9】**

前記関係情報生成ステップは、前記第1の知識源とは異なる第2の知識源に含まれる文に基づいて、前記回答候補が表す事物の同一性についての回答候補間の関係を示す関係情報を生成することを特徴とする請求項6乃至8のいずれかに記載の処理方法。

**【請求項 10】**

前記関係情報生成ステップにより生成された関係情報を記憶装置に保持させる関係情報保持ステップを有することを特徴とする請求項6乃至9のいずれかに記載の処理方法。

**【請求項 11】**

入力した検索質問文に応じて、検索対象文を含む第1の知識源を検索し、前記検索質問文の回答となり得る回答候補を得る質問応答システムにて実行可能なプログラムであって、

前記検索質問文から検索語を抽出する検索語抽出ステップと、

前記検索語に基づいて前記第1の知識源に含まれる検索対象文の中から検索結果の文を取得する検索ステップと、

前記検索結果の文から前記回答候補を抽出し、前記回答候補毎に該回答候補の尤度を表す数値を付与する回答候補抽出ステップと、

前記回答候補が表す事物の同一性についての回答候補間の関係を示す関係情報を生成する関係情報生成ステップと、

前記抽出された回答候補から、前記関係情報に基づいて、同一の事物を表す回答候補からなる回答候補集合を生成し、前記回答候補集合毎に、該回答候補集合に含まれる回答候補に付与された尤度の数値を集計する回答候補結合ステップと、

前記尤度の集計値が最も大きい前記回答候補集合に含まれる回答候補のいずれかを優先的に提示する制御を行う提示制御ステップとを有することを特徴とするプログラム。

**【請求項 12】**

前記関係情報生成ステップは、1の回答候補を含む検索結果の文の集合における単語のベクトルと、他の回答候補を含む検索結果の文の集合における単語のベクトルとの内積を算出し、該内積が所定値以上である場合に、前記1及び他の回答候補が同一の事物を表すものであることを示す関係情報を生成することを特徴とする請求項11に記載のプログラム。

**【請求項 13】**

前記関係情報生成手段は、同一の事物を表す回答候補のいずれかがこれら回答候補の代表となる回答候補であることを示す関係情報を生成し、

前記回答候補結合ステップは、前記同一の事物を表す回答候補からなり、前記代表となる回答候補に他の回答候補を置き換えた回答候補集合を生成し、

前記提示制御ステップは、前記代表となる回答候補を優先的に提示する制御を行うことを特徴とする請求項11又は12に記載のプログラム。

**【請求項 14】**

前記関係情報生成ステップは、前記第1の知識源とは異なる第2の知識源に含まれる文に基づいて、前記回答候補が表す事物の同一性についての回答候補間の関係を示す関係情報

10

20

30

40

50

を生成することを特徴とする請求項 1 1 乃至 1 3 のいずれかに記載のプログラム。

【請求項 1 5】

前記関係情報生成ステップにより生成された関係情報を記憶装置に保持させる関係情報保持ステップを有することを特徴とする請求項 1 1 乃至 1 4 のいずれかに記載のプログラム。

【発明の詳細な説明】

【技術分野】

【0 0 0 1】

本発明は、入力した検索質問文に応じた応答を行う質問応答システム、該質問応答システムにおける質問応答処理方法及びコンピュータにおいて実行される質問応答プログラム 10 に関する。

【背景技術】

【0 0 0 2】

一般的な質問応答の技術として、特許文献 1 に記載されたものがある。この質問応答システムは、検索質問文から検索語を抽出するとともに、質問種別の判定を行う質問解析部と、検索語に基づいて知識源を検索し、検索結果文であるパッセージを抽出する情報検査部と、パッセージからいくつかの回答候補を抽出、提示する回答抽出部と、その回答候補の根拠を提示する根拠提示部とにより構成される。

【0 0 0 3】

ところで、上述した質問応答システムにおける回答抽出部は、高い精度で正解となる回答候補を抽出することが困難であるため、一般に複数の回答候補を提示する。この提示された回答候補の確からしさをユーザが判断可能とするために、根拠提示部が回答候補の根拠文である回答候補の抽出元文や、重要文を提示することもなされる（たとえば、特許文献 2 及び 3 参照）。 20

【特許文献 1】特開 2 0 0 2 - 1 3 2 8 1 1 号公報

【特許文献 2】特開 2 0 0 2 - 1 3 2 8 1 2 号公報

【特許文献 3】特開 2 0 0 2 - 1 2 5 9 3 7 1 号公報

【発明の開示】

【発明が解決しようとする課題】

【0 0 0 4】

しかし、上述した特許文献 2 及び 3 に記載された技術は、提示される複数の回答候補のそれぞれが適切であるか否かをユーザが判断可能とするものであり、回答候補間の関係については何ら考慮されていない。このため、回答候補を適切に提示することができない場合がある。 30

【0 0 0 5】

例えば、検索質問文「大阪でおきた大地震の名前はなんですか？」が質問応答システムに入力され、回答候補として順次「大地震」、「元禄大地震」、「新潟県中越地震」、「阪神・淡路大震災」、「関西大震災」、「阪神大震災」、「兵庫県南部地震」が得られた場合を考える。この場合、これら回答候補のうち、検索質問文に対する正解は、「阪神・淡路大震災」、「関西大震災」、「阪神大震災」、「兵庫県南部地震」の 4 つであるが、 40 単に回答候補が得られた順に提示されると、これらの正解はいずれも提示順序が下位になってしまう。すなわち、同一の事物を表す回答候補であっても、異なる単語であるために別々に扱われて提示されており、ユーザがどの回答候補が適切であるのかを判断しにくい。

【0 0 0 6】

本発明の目的は、上述した問題を解決するものであり、回答候補間の関係を考慮した適切な回答候補の提示が可能な質問応答システム、質問応答処理方法及び質問応答プログラムを提供するものである。

【課題を解決するための手段】

【0 0 0 7】

本発明に係る、入力した検索質問文に応じて、検索対象文を含む第1の知識源を検索し、前記検索質問文の回答となり得る回答候補を得る質問応答システムは、前記検索質問文から検索語を抽出する検索語抽出手段と、前記検索語に基づいて前記第1の知識源に含まれる検索対象文の中から検索結果の文を取得する検索手段と、前記検索結果の文から前記回答候補を抽出し、前記回答候補毎に該回答候補の尤度を表す数値を付与する回答候補抽出手段と、前記回答候補が表す事物の同一性についての回答候補間の関係を示す関係情報を生成する関係情報生成手段と、前記抽出された回答候補から、前記関係情報に基づいて、同一の事物を表す回答候補からなる回答候補集合を生成し、前記回答候補集合毎に、該回答候補集合に含まれる回答候補に付与された尤度の数値を集計する回答候補結合手段と、記尤度の集計値が最も大きい前記回答候補集合に含まれる回答候補のいずれかを優先的に提示する制御を行う提示制御手段とを有することを特徴とする。 10

【0008】

この構成によれば、同一の事物を表す回答候補からなる回答候補集合が生成され、各回答候補集合のうち、回答候補に付与された尤度の集計値が最も大きい回答候補集合に含まれる回答候補のいずれかが回答に相応しい回答候補として優先的に提示される。

【0009】

また、質問応答システムは、前記関係情報生成手段が、1の回答候補を含む検索結果の文における単語のベクトルと、他の回答候補を含む検索結果の文における単語のベクトルとの内積を算出し、該内積が所定値以上である場合に、前記1及び他の回答候補が同一の事物を表すものであることを示す関係情報を生成するようにしてもよい。 20

【0010】

この構成によれば、検索結果の文における単語のベクトルの内積値により、1及び他の回答候補の関係を適切に表すことができる。

【0011】

また、質問応答システムは、前記関係情報生成手段は、同一の事物を表す回答候補のいずれかがこれら回答候補の代表となる回答候補であることを示す関係情報を生成し、前記回答候補結合手段が、前記同一の事物を表す回答候補からなり、前記代表となる回答候補に他の回答候補を置き換えた回答候補集合を生成し、前記提示制御手段が、前記代表となる回答候補を優先的に提示する制御を行うようにしてもよい。

【0012】

この構成によれば、回答候補に付与された尤度の集計値が最も大きい回答候補集合に含まれる回答候補のうち、これら回答候補を代表する回答候補が優先的に提示されることが可能となる。 30

【0013】

また、質問応答システムは、前記関係情報生成手段が、前記第1の知識源とは異なる第2の知識源に含まれる文に基づいて、前記回答候補が表す事物の同一性についての回答候補間の関係を示す関係情報を生成するようにしてもよい。

【0014】

この構成によれば、検索対象以外の知識源に含まれる文を考慮することにより、回答候補が表す事物の同一性についての回答候補間の関係をより適切に示すことが可能となる。 40

【0015】

また、質問応答システムは、前記関係情報生成手段により生成された関係情報を保持する関係情報保持手段を有するようにしてもよい。

【0016】

この構成によれば、過去の回答候補提示の際に用いられた関係情報をも用いて、より迅速に回答候補の提示を行うことが可能となる。

【0017】

本発明に係る、入力した検索質問文に応じて、検索対象文を含む第1の知識源を検索し、前記検索質問文の回答となり得る回答候補を得る質問応答システムでの処理方法は、前記検索質問文から検索語を抽出する検索語抽出ステップと、前記検索語に基づいて前記第 50

1の知識源に含まれる検索対象文の中から検索結果の文を取得する検索ステップと、前記検索結果の文から前記回答候補を抽出し、前記回答候補毎に該回答候補の尤度を表す数値を付与する回答候補抽出ステップと、前記回答候補が表す事物の同一性についての回答候補間の関係を示す関係情報を生成する関係情報生成ステップと、前記抽出された回答候補から、前記関係情報に基づいて、同一の事物を表す回答候補からなる回答候補集合を生成し、前記回答候補集合毎に、該回答候補集合に含まれる回答候補に付与された尤度の数値を集計する回答候補結合ステップと、前記尤度の集計値が最も大きい前記回答候補集合に含まれる回答候補のいずれかを優先的に提示する制御を行う提示制御ステップとを有することを特徴とする。

【0018】

10

また、質問応答処理方法は、前記関係情報生成ステップが、1の回答候補を含む検索結果の文の集合における単語のベクトルと、他の回答候補を含む検索結果の文の集合における単語のベクトルとの内積を算出し、該内積が所定値以上である場合に、前記1及び他の回答候補が同一の事物を表すものであることを示す関係情報を生成するようにしてもよい。

【0019】

また、質問応答処理方法は、前記関係情報生成手段が、同一の事物を表す回答候補のいずれかがこれら回答候補の代表となる回答候補であることを示す関係情報を生成し、前記回答候補結合ステップが、前記同一の事物を表す回答候補からなり、前記代表となる回答候補に他の回答候補を置き換えた回答候補集合を生成し、前記提示制御ステップが、前記代表となる回答候補を優先的に提示する制御を行うようにしてもよい。

20

【0020】

また、質問応答処理方法は、前記関係情報生成ステップが、前記第1の知識源とは異なる第2の知識源に含まれる文に基づいて、前記回答候補が表す事物の同一性についての回答候補間の関係を示す関係情報を生成するようにしてもよい。

【0021】

また、質問応答処理方法は、前記関係情報生成ステップにより生成された関係情報を記憶装置に保持させる関係情報保持ステップを有するようにしてもよい。

【0022】

本発明に係る、入力した検索質問文に応じて、検索対象文を含む第1の知識源を検索し、前記検索質問文の回答となり得る回答候補を得る質問応答システムにて実行可能なプログラムは、前記検索質問文から検索語を抽出する検索語抽出ステップと、前記検索語に基づいて前記第1の知識源に含まれる検索対象文の中から検索結果の文を取得する検索ステップと、前記検索結果の文から前記回答候補を抽出し、前記回答候補毎に該回答候補の尤度を表す数値を付与する回答候補抽出ステップと、前記回答候補が表す事物の同一性についての回答候補間の関係を示す関係情報を生成する関係情報生成ステップと、前記抽出された回答候補から、前記関係情報に基づいて、同一の事物を表す回答候補からなる回答候補集合を生成し、前記回答候補集合毎に、該回答候補集合に含まれる回答候補に付与された尤度の数値を集計する回答候補結合ステップと、前記尤度の集計値が最も大きい前記回答候補集合に含まれる回答候補のいずれかを優先的に提示する制御を行う提示制御ステップとを有することを特徴とする。

30

40

【0023】

また、質問応答プログラムは、前記関係情報生成ステップが、1の回答候補を含む検索結果の文の集合における単語のベクトルと、他の回答候補を含む検索結果の文の集合における単語のベクトルとの内積を算出し、該内積が所定値以上である場合に、前記1及び他の回答候補が同一の事物を表すものであることを示す関係情報を生成するようにしてもよい。

【0024】

また、質問応答プログラムは、前記関係情報生成手段が、同一の事物を表す回答候補のいずれかがこれら回答候補の代表となる回答候補であることを示す関係情報を生成し、前

50

記回答候補結合ステップが、前記同一の事物を表す回答候補からなり、前記代表となる回答候補に他の回答候補を置き換えた回答候補集合を生成し、前記提示制御ステップが、前記代表となる回答候補を優先的に提示する制御を行うようにしてもよい。

【0025】

また、質問応答プログラムは、前記関係情報生成ステップが、前記第1の知識源とは異なる第2の知識源に含まれる文に基づいて、前記回答候補が表す事物の同一性についての回答候補間の関係を示す関係情報を生成するようにしてもよい。

【0026】

また、質問応答プログラムは、前記関係情報生成ステップにより生成された関係情報を記憶装置に保持させる関係情報保持ステップを有するようにしてもよい。

10

【発明の効果】

【0027】

本発明によれば、同一の事物を表す回答候補からなる回答候補集合が生成され、回答候補に付与された尤度の集計値が最も大きい回答候補集合に含まれる回答候補のいずれかが回答に相応しい回答候補として優先的に提示されるため、異なる単語の回答候補によって同一の事物が表される場合であっても、回答候補間の関連性を考慮した適切な回答候補の提示が可能となる。

【発明を実施するための最良の形態】

【0028】

本発明の実施の形態について、図面を参照して具体的に説明する。図1及び図2は、第1の質問応答システムのハードウェア構成の一例を示す図である。図1に示す第1の質問応答システム100-1は、パーソナルコンピュータ(PC)であり、内部バス107に接続されたCPU101、メモリ102、ハードディスクドライブ(HDD)103、通信インターフェース(I/F)部104、操作部105及びモニタ106によって構成され、通信I/F部104が外部の通信ネットワーク250を介して知識源200に接続されている。

20

【0029】

一方、図2に示す質問応答システム100-1は、LAN等のバス113に接続されたサーバ111及びPC112により構成され、サーバ111が外部の通信ネットワーク250を介して知識源200に接続される。

30

【0030】

図3は、第1の質問応答システム100-1の機能ブロック図である。同図に示す質問応答システム100-1は、電子化された検索質問文に応じて、電子化された検索対象文を含む外部の知識源200を検索して回答を得るものである。

【0031】

この質問応答システム100-1は、質問判定部10、パッセージ検索部20、回答候補抽出部30、関係抽出部40、回答候補結合部50、回答候補選択部60及び回答候補提示制御部70により構成される。これら各機能ブロックは、図1のハードウェア構成においては、CPU101がHDD103から読み出してメモリ102に記憶させた所定のプログラムを実行し、通信I/F部104を介して知識源200にアクセスすることによって実現される。一方、これら各機能ブロックは、図2のハードウェア構成においては、サーバ111がPC112の要求に応じて所定のプログラムを実行し、知識源200にアクセスすることによって実現される。

40

【0032】

質問判定部10は、ユーザによる図1の操作部105の操作等に応じて入力した検索質問文を解析し、当該検索質問文に含まれる検索語を抽出する。パッセージ検索部20は、質問判定部10によって抽出された検索語を用いて、知識源200に含まれる検索対象文を検索し、検索結果の文であるパッセージを取得する。回答候補抽出部30は、パッセージ検索部20によって取得されたパッセージから検索質問文の回答となり得る回答候補として抽出する。この際、回答候補抽出部30は、各回答候補に当該回答候補の確からしさ

50

(尤度)を表す得点を付与する。

#### 【0033】

関係抽出部40は、回答候補抽出部30によって抽出された回答候補が表す事物の同一性についての回答候補間の関係を示す関係規則情報を生成する。回答候補結合部50は、関係抽出部40によって生成された関係規則情報に基づいて、回答候補抽出部30により抽出された回答候補について、同一の事物を表すものを結合して回答候補集合を生成する。回答候補選択部60は、回答候補結合部50によって生成された各回答候補集合毎に、当該回答候補集合に含まれる回答候補に付与された得点を集計する。回答候補提示制御部70は、回答候補選択部60によって集計された、対応する得点が高い所定件数の回答候補集合に含まれる回答候補を、図1のモニタ106等を用いてユーザに提示する制御を行う。 10

#### 【0034】

以下、フローチャートを参照しつつ、質問応答システム100-1の動作を説明する。図4は、質問応答システム100-1の動作を示すフローチャートである。なお、以下においては、「大阪で起きた大地震の名前は何ですか？」という検索質問文に対する回答候補を抽出、提示する場合を適宜、例にとって説明する。

#### 【0035】

ユーザが、図1の操作部105等を用いて検索質問文の入力操作を行うと、質問判定部10には、その検索質問文の文字列が入力される。質問判定部10は、定期的に検索質問文を入力したか否かを判定する(S101)。検索質問文を入力した場合、質問判定部10は、公知の検索語の抽出方法であるTF(Term Frequency)／IDF法等により、その検索質問文から検索語を抽出する(S102)。例えば、検索質問文が「大阪で起きた大地震の名前は何ですか？」である場合、検索語として、「起きる」、「大」、「大阪」、「地震」、「名前」が抽出される。抽出された検索語は、パッセージ検索部20へ送られる。 20

#### 【0036】

パッセージ検索部20は、質問判定部10からの検索語を入力すると、知識源200に含まれる検索対象文を検索し、検索語が含まれる検索対象文をパッセージとして取得する(S103)。取得されたパッセージは、回答候補抽出部30及び関係抽出部40へ送られる。 30

#### 【0037】

回答候補抽出部30は、パッセージ検索部20からのパッセージを入力すると、当該パッセージに含まれる名詞やそれに準ずる品詞を回答候補として抽出する。更に、回答候補抽出部30は、抽出した各回答候補に対して、当該回答候補の尤度を表す得点を付与する(S104)。得点は、例えば、パッセージにおける回答候補の出現回数等に応じて予め定められたものが付与される。例えば、図5に示すように、検索質問文「大阪で起きた大地震の名前は何ですか？」に対して、「大地震」、「元禄大地震」、「新潟県中越地震」、「阪神・淡路大震災」、「関西大震災」、「阪神大震災」、「兵庫県南部地震」が回答候補として抽出され、これら各回答候補に得点が付与される。回答候補とパッセージは、関係抽出部40へ送られる。また、回答候補と当該回答候補に付与された得点との組み合わせは、回答候補結合部50へ送られる。 40

#### 【0038】

関係抽出部40は、回答候補抽出部30からの各回答候補について、当該回答候補が表す事物の同一性についての回答候補間の関係を示す関係規則情報を生成する(S105)。具体的には、関係抽出部40は、1の回答候補を含んでいるパッセージの集合の単語のベクトルと、他の回答候補を含んでいるパッセージの集合の単語のベクトルとの内積を算出する。更に、関係抽出部40は、算出した内積値が所定値以上である場合に、1の回答候補と他の回答候補とが同一の事物を表すものであると判定する。

#### 【0039】

図6は、回答候補間の関係の一例を示す図である。同図は、検索質問文「大阪で起きた 50

大地震の名前は何か？」に対して、「大地震」、「元禄大地震」、「新潟県中越地震」、「阪神・淡路大震災」、「関西大震災」、「阪神大震災」、「兵庫県南部地震」が回答候補として抽出される場合の例を示している。図6において、2つの回答候補を結ぶ実線は、当該実線で結ばれた1の回答候補を含んでいるパッセージの集合の単語のベクトルと他の回答候補を含んでいるパッセージの集合の単語のベクトルとの内積値が所定値以上であり、これら2つの回答候補が同一の事物を表すものであることを示している。例えば、回答候補「阪神・淡路大震災」と回答候補「阪神大震災」及び「兵庫県南部地震」とは、実線で結ばれており、同一の事物を表している。また、回答候補「阪神大震災」と回答候補「関西大震災」とは、実線で結ばれており、同一の事物を表している。更には、回答候補「阪神・淡路大震災」と回答候補「関西大震災」とは、直接、実線で結ばれてはいないものの、回答候補「阪神大震災」を介して実線で結ばれており、同一の事物を表している。

10

#### 【0040】

更に、関係抽出部40は、同一の事物を表す回答候補について、いずれかを代表の回答候補としての置き換え先とし、他の回答候補を置き換え元とすることを示す関係規則情報を生成する。例えば、回答候補間の関係が図6のように示される場合、関係抽出部40は、同一の事物を表す回答候補「阪神・淡路大震災」、「阪神大震災」、「兵庫県南部地震」、「関西大震災」のうち、実線で結ばれる他の回答候補の数が最も多い回答候補「阪神・淡路大震災」又は「阪神大震災」を選択する。更に、関係抽出部40は、回答候補「阪神・淡路大震災」を選択した場合には、図7に示すように、選択した回答候補「阪神・淡路大震災」を置き換え先とし、他の回答候補「関西大震災」、「阪神大震災」、「兵庫県南部地震」を置き換え元とすることを示す関係規則情報を生成する。生成された関係規則情報は、回答候補結合部50へ送られる。

20

#### 【0041】

回答候補結合部50は、回答候補抽出部30からの各回答候補から、関係抽出部40からの関係規則情報に基づいて、同一の事物を表す回答候補からなる回答候補集合を生成するとともに、各回答候補集合毎に、当該回答候補集合に含まれる回答候補に付与された得点を集計する(S106)。

#### 【0042】

具体的には、回答候補結合部50は、回答候補抽出部30からの回答候補を1つずつ選択し、その回答候補が、関係規則情報で示される置き換え先の回答候補、及び、置き換え元の回答候補のいずれかと一致するか否かを判定する。選択した回答候補が関係規則情報で示される置き換え先の回答候補、及び、置き換え元の回答候補のいずれとも一致しない場合には、回答候補結合部50は、その選択した回答候補を単独で回答候補集合とする。一方、選択した回答候補が関係規則情報で示される置き換え先の回答候補と一致する場合には、回答候補結合部50は、その選択した回答候補を、回答候補集合を代表する回答候補(代表回答候補)とする。また、選択した回答候補が関係規則情報で示される置き換え元の回答候補と一致する場合には、回答候補結合部50は、その選択した回答候補を、置き換え先の回答候補と一致する代表回答候補に置き換える。回答候補抽出部30からの回答候補の全てについて、このような処理が行われることによって、同一の事物を表す回答候補からなる回答候補集合が生成される。例えば、関係規則情報が図6で示される場合には、回答候補結合部50は、回答候補「阪神・淡路大震災」を代表回答候補とするとともに、回答候補「関西大震災」、「阪神大震災」、「兵庫県南部地震」を、代表回答候補「阪神・淡路大震災」に置き換えて、回答候補「阪神・淡路大震災」を代表とする回答候補集合を生成する。

30

40

#### 【0043】

更に、回答候補結合部50は、回答候補抽出部30からの各回答候補に付与された得点に基づいて、各回答候補集合について、当該回答候補集合に含まれる回答候補に付与された得点を集計する。例えば、図5に示すように各回答候補に得点が付与され、更に、回答候補「大地震」からなる回答候補集合と、回答候補「元禄大地震」からなる回答候補集合

50

と、回答候補「新潟県中越地震」からなる回答候補集合と、回答候補「阪神・淡路大震災」、「関西大震災」、「阪神大震災」、「兵庫県南部地震」からなり、回答候補「阪神・淡路大震災」を代表とする回答候補集合とが生成される場合を考える。この場合、回答候補結合部50は、図8に示すように、回答候補「大地震」からなる回答候補集合と、回答候補「元禄大地震」からなる回答候補集合と、回答候補「新潟県中越地震」からなる回答候補集合については、その回答候補に付与された得点を、回答候補集合に対応する得点とするとともに、回答候補「阪神・淡路大震災」を代表とする回答候補集合については、回答候補「阪神・淡路大震災」、「関西大震災」、「阪神大震災」、「兵庫県南部地震」のそれぞれに付与された得点0.3、0.3、0.2、0.1を集計した0.9を回答候補集合の得点とする。各回答候補集合と、当該回答候補集合に対応する得点とは、回答候補選択部60へ送られる。 10

#### 【0044】

回答候補選択部60は、回答候補集合と、当該回答候補集合に対応する得点とを入力すると、得点の高い順に回答候補集合を並べ替える。例えば、回答候補集合に対応する得点が図8に示すものである場合、回答候補選択部60は、図9に示すように、回答候補「阪神・淡路大震災」を代表とする得点0.9の回答候補集合、回答候補「大地震」からなる得点0.7の回答候補集合、回答候補「元禄大地震」からなる得点0.6の回答候補集合、回答候補「新潟県中越地震」からなる得点0.5の回答候補集合の順に並べ替える。

#### 【0045】

更に、回答候補選択部60は、対応する得点の高い所定件数の回答候補集合に属する回答候補を、提示すべき回答候補として選択する(S107)。例えば、回答候補集合が図9に示すように並び替えられ、得点の最も高い1件の回答候補集合に含まれる回答候補が選択される場合には、得点0.9の回答候補集合に含まれる代表の回答候補「阪神・淡路大震災」が選択される。提示すべきものとして選択された回答候補は、回答候補提示制御部70へ送られる。回答候補提示制御部70は、入力した回答候補を、図1のモニタ106等を用いてユーザに提示する制御を行う(S108)。 20

#### 【0046】

このように、本実施形態の質問応答システム100-1では、関係抽出部40が、回答候補抽出部30によって抽出された回答候補が表す事物の同一性についての回答候補間の関係を示す関係規則情報を生成し、回答候補結合部50が、回答候補抽出部30により抽出された回答候補から、この関係規則情報に基づいて、同一の事物を表す回答候補からなる回答候補集合を生成し、回答候補選択部60が、各回答候補集合毎に、当該回答候補集合に含まれる回答候補に付与された得点を集計し、回答候補提示制御部70によって、対応する得点が高い所定件数の回答候補集合に含まれる回答候補が提示される。すなわち、異なる単語の回答候補によって同一の事物が表される場合であっても、それらの回答候補はまとめられ、回答として相応しい場合には優先的に提示されるため、適切な回答候補の提示が可能となる。 30

#### 【0047】

ところで、上述した実施形態では、関係抽出部40は、回答候補抽出部30からの各回答候補について、検索対象の知識源200から抽出された、回答候補を含んでいるパッセージの集合の単語のベクトル間の内積値に基づいて、関係規則情報を生成したが、更に、検索対象でない他の知識源から回答候補を含むパッセージを抽出し、そのパッセージの集合の単語のベクトル間の内積値に基づいて、関係規則情報を生成するようにしてもよい。 40

#### 【0048】

図10は、他の知識源を利用する質問応答システムの機能ブロック図である。同図に示す第2の質問応答システム100-2は、図3に示す質問応答システム100-1と比較すると、関係抽出部40は、質問判定部10に接続されるとともに、外部の知識源300に接続されている。なお、質問応答システム100-2のハードウェア構成は、図1や図2に示す質問応答システム100-1と同様である。但し、質問応答システム100-2は、通信ネットワーク250を介して知識源200のみならず知識源300にも接続され 50

ている。

#### 【0049】

この質問応答システム100-2において、質問判定部10は、検索質問文から抽出した検索語をパッセージ検索部20のみならず、関係抽出部40にも出力する。関係抽出部40は、パッセージ検索部20からのパッセージを取得するとともに、質問判定部10からの検索語を用いて、知識源300に含まれる検索対象文を検索し、検索結果の文であるパッセージを取得する。そして、関係抽出部40は、回答候補抽出部30からの各回答候補について、1の回答候補を含んでいるパッセージの集合の単語のベクトルと、他の回答候補を含んでいるパッセージの集合の単語のベクトルとの内積を算出する。更に、関係抽出部40は、算出した内積値が所定値以上である場合に、1の回答候補と他の回答候補とが同一の事物を表すものであると判定し、同一の事物を表す回答候補について、いずれかを代表回答候補としての置き換え先とし、他の回答候補を置き換え元とすることを示す関係規則情報を生成する。このように、検索対象でない知識源300に含まれる文を考慮することにより、回答候補が表す事物の同一性についての回答候補間の関係をより適切に示すことが可能となる。

10

#### 【0050】

また、関係抽出部40が生成した関係規則情報を保持するようにしてもよい。図11は、関係規則情報を保持する質問応答システムの機能ブロック図である。同図に示す質問応答システム100-3は、図3に示す質問応答システム100-1と比較すると、関係規則情報を保持する関係規則データベース(DB)80を有する。なお、質問応答システム100-3のハードウェア構成は、図1や図2に示す質問応答システム100-1と同様である。但し、質問応答システム100-3は、図1におけるメモリ106や図2におけるサーバ111内のメモリ(図示せず)に関係規則DB80を保持している。

20

#### 【0051】

関係抽出部40は、関係規則情報を生成する毎に、当該関係規則情報を関係規則DB80に保持させる。そして、その後の回答候補の提示処理においては、関係抽出部40は、回答候補抽出部30からの回答候補が関係規則DB80内の関係規則情報における置き換え元又は置き換え先と一致する場合には、その関係規則情報をそのまま回答候補結合部50に出力する。これにより、過去の回答候補提示の際に用いられた関係規則情報をも用いて、より迅速に回答候補の提示を行うことが可能となる。

30

#### 【0052】

また、上述した実施形態では、回答候補集合を構成する回答候補のうち、代表となる回答候補、換言すれば、置き換え先となる回答候補のみが提示されたが、他の回答候補、換言すれば、置き換え元の回答候補が提示されるようにしてもよい。

#### 【産業上の利用可能性】

#### 【0053】

以上、説明したように、本発明に係る質問応答システム、質問応答処理方法及び質問応答プログラムによれば、回答候補間の関係を考慮した適切な回答候補の提示が可能であり、質問応答システム等として有用である。

#### 【図面の簡単な説明】

40

#### 【0054】

【図1】質問応答システムのハードウェア構成の第1の例を示す図である。

【図2】質問応答システムのハードウェア構成の第2の例を示す図である。

【図3】第1の質問応答システムの機能ブロック図である。

【図4】質問応答システムの動作を示すフローチャートである。

【図5】回答候補に付与される得点の一例である。

【図6】回答候補間の関係の一例を示す図である。

【図7】関係規則情報の一例を示す図である。

【図8】回答候補集合に対応する得点の一例を示す図である。

【図9】得点順に並び替えられた回答候補集合の一例を示す図である。

50

【図 1 0】 第 2 の質問応答システムの機能ブロック図である。

【図 1 1】 第 3 の質問応答システムの機能ブロック図である。

【符号の説明】

【 0 0 5 5 】

1 0 質問判定部

2 0 パッセージ検索部

3 0 回答候補抽出部

4 0 関係抽出部

5 0 回答候補結合部

6 0 回答候補選択部

7 0 回答候補提示制御部

8 0 関係規則 D B

1 0 0 - 1、1 0 0 - 2、1 0 0 - 3 質問応答システム

1 0 1 C P U

1 0 2 メモリ

1 0 3 H D D

1 0 4 通信 I / F 部

1 0 5 操作部

1 0 6 モニタ

1 0 7 内部バス

1 1 1 サーバ

1 1 2 P C

1 1 3 バス

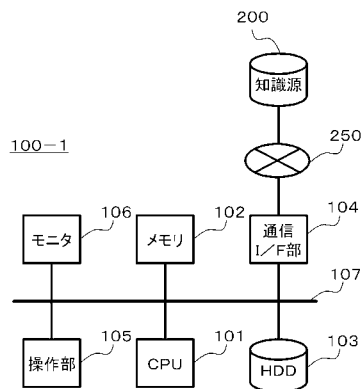
2 0 0、3 0 0 知識源

2 5 0 通信ネットワーク

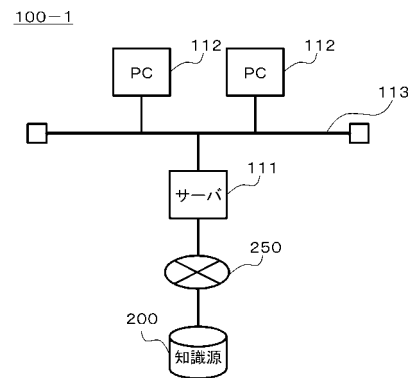
10

20

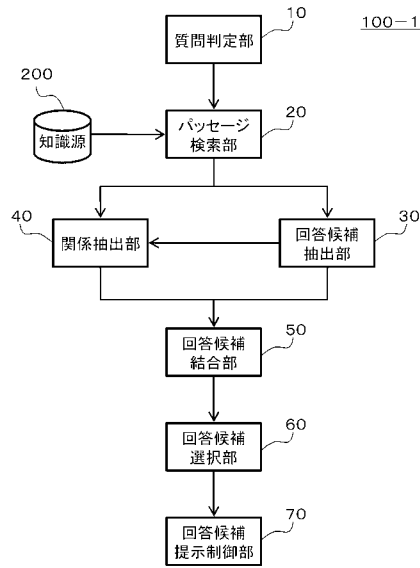
【図 1】



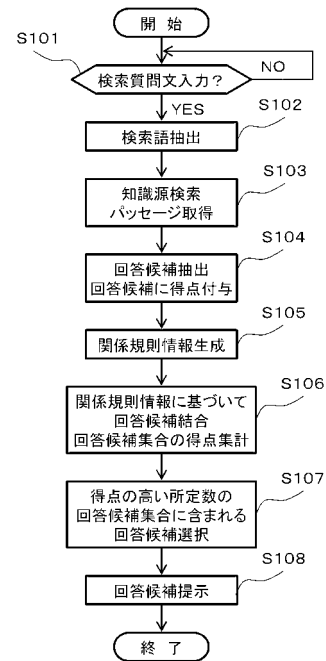
【図 2】



【図 3】



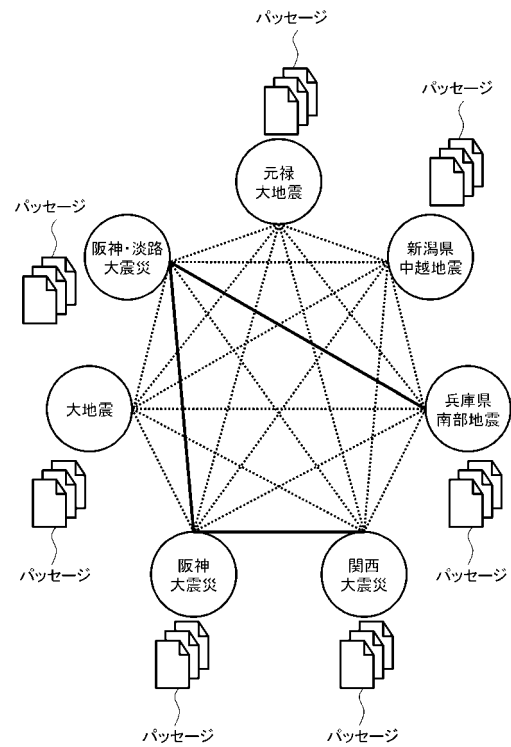
【図 4】



【図 5】

|          |     |
|----------|-----|
| 大地震      | 0.7 |
| 元禄大地震    | 0.6 |
| 新潟県中越地震  | 0.5 |
| 阪神・淡路大震災 | 0.3 |
| 関西大震災    | 0.3 |
| 阪神大震災    | 0.2 |
| 兵庫県南部地震  | 0.1 |

【図 6】



【図 7】

| 置き換え元   | 置き換え先    |
|---------|----------|
| 関西大震災   | 阪神・淡路大震災 |
| 阪神大震災   | 阪神・淡路大震災 |
| 兵庫県南部地震 | 阪神・淡路大震災 |

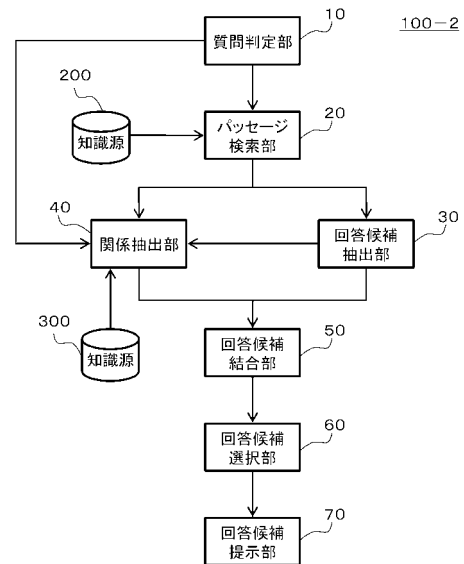
【図 8】

|          |     |
|----------|-----|
| 大地震      | 0.7 |
| 元禄大地震    | 0.6 |
| 新潟県中越地震  | 0.5 |
| 阪神・淡路大震災 | 0.9 |

【図 9】

|          |     |
|----------|-----|
| 阪神・淡路大震災 | 0.9 |
| 大地震      | 0.7 |
| 元禄大地震    | 0.6 |
| 新潟県中越地震  | 0.5 |

【図 10】



【図 11】

