

(19) 日本国特許庁(JP)

(12) 特許公報(B2)

(11) 特許番号

特許第4635585号
(P4635585)

(45) 発行日 平成23年2月23日(2011.2.23)

(24) 登録日 平成22年12月3日(2010.12.3)

(51) Int.Cl.

G06F 17/30 (2006.01)

F 1

G06F 17/30 170A
G06F 17/30 330C
G06F 17/30 180A

請求項の数 7 (全 16 頁)

(21) 出願番号 特願2004-349796 (P2004-349796)
 (22) 出願日 平成16年12月2日 (2004.12.2)
 (65) 公開番号 特開2006-163491 (P2006-163491A)
 (43) 公開日 平成18年6月22日 (2006.6.22)
 審査請求日 平成19年11月20日 (2007.11.20)

(73) 特許権者 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
 (74) 代理人 100098497
 弁理士 片寄 恒三
 (72) 発明者 吉岡 健
 神奈川県足柄上郡中井町境430 グリー
 ニテクなかい 富士ゼロックス株式会社内
 (72) 発明者 古木 真
 神奈川県足柄上郡中井町境430 グリー
 ニテクなかい 富士ゼロックス株式会社内

最終頁に続く

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

(57) 【特許請求の範囲】

【請求項 1】

電子化文書を含む情報源から検索質問文に応答して情報を検索する質問応答システムであつて、

検索質問文の構文構造と該検索質問文に対応する検索式とを対応付けた質問・回答パターン情報を保持する質問・回答パターン情報保持手段と、

入力検索質問文に対して構文解析を行い、該入力検索質問文の構文構造を生成する構文解析手段と、

質問・回答パターン情報保持手段に保持された質問・回答パターン情報の中から前記構文解析手段により生成された入力検索質問文の構文構造と適合するものを特定する質問・回答パターン情報特定手段と、

前記質問・回答パターン情報特定手段により特定された質問・回答パターン情報における検索式に基づいて、前記情報源を検索し、回答候補を抽出する第1の検索実行手段と、

前記入力検索質問文に含まれる単語に対応する検索式、及び、前記入力検索質問文の構文構造に対応する検索式の少なくともいずれかを用いて前記情報源を検索し、回答候補を抽出する第2の検索実行手段と、

前記入力検索質問文に基づく統計ベースでの前記情報源の検索を行い、回答候補を抽出する第3の検索実行手段とを有し、

前記構文解析手段は、疑問文である入力検索質問文を肯定文に変換し、当該肯定文の所定の単語を変数化した穴埋め肯定文を生成し、かつ当該穴埋め肯定文の構文構造を生成す

るものであり、

前記質問・回答パターン情報保持手段は、前記構文解析手段によって生成された穴埋め肯定文の所定の単語を変数化した構文構造と当該穴埋め肯定文の構文構造に基づく文法ベースの検索式とを保持するものであり、

前記第1の検索実行手段は、前記第2の検索実行手段により回答候補が抽出されなかつた場合に前記情報源を検索し、

前記第3の検索実行手段は、前記第1および第2の検索実行手段により回答候補が抽出されなかつた場合に前記情報源を検索することを特徴とする質問応答システム。

【請求項2】

前記第3の検索実行手段により抽出された回答候補を検索するための検索式を生成する検索式生成手段と、 10

前記入力検索質問文の構文構造と、前記検索式生成手段により生成された検索式とを対応付けた質問・回答パターン情報を生成し、前記質問・回答パターン情報保持手段へ登録する質問・回答パターン情報生成・登録手段とを有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。

【請求項3】

前記検索式生成手段は、前記第3の検索実行手段により抽出された回答候補のうち、所定の選択操作によって選択された回答候補を検索するための検索式を生成することを特徴とする請求項2に記載の質問応答システム。

【請求項4】

コンピュータを用いて行われる電子化文書を含む情報源から検索質問文に応答して情報を検索する質問応答方法であつて、 20

入力検索質問文に対して構文解析を行い、該入力検索質問文の構文構造を生成する構文解析ステップと、

予め質問・回答パターン情報保持手段に保持された、検索質問文の構文構造と該検索質問文に対応する検索式とを対応付けた質問・回答パターン情報の中から前記構文解析ステップにおいて生成された入力検索質問文の構文構造と適合するものを特定する質問・回答パターン情報特定ステップと、

前記質問・回答パターン情報特定ステップにおいて特定された質問・回答パターン情報における検索式に基づいて、前記情報源を検索し、回答候補を抽出する第1の検索実行ステップと、 30

前記入力検索質問文に含まれる単語に対応する検索式、及び、前記入力検索質問文の構文構造に対応する検索式の少なくともいずれかを用いて前記情報源を検索し、回答候補を抽出する第2の検索実行ステップと、

前記入力検索質問文に基づく統計ベースでの前記情報源の検索を行い、回答候補を抽出する第3の検索実行ステップとを有し、

前記構文解析ステップは、疑問文である入力検索質問文を肯定文に変換し、当該肯定文の所定の単語を変数化した穴埋め肯定文を生成し、かつ当該穴埋め肯定文の構文構造を生成するものであり、

前記質問・回答パターン情報保持手段は、前記構文解析ステップによって生成された穴埋め肯定文の所定の単語を変数化した構文構造と当該穴埋め肯定文の構文構造に基づく文法ベースの検索式とを保持するものあり、 40

前記第1の検索実行ステップは、前記第2の検索実行ステップにより回答候補が抽出されなかつた場合に前記情報源を検索し、

前記第3の検索実行ステップは、前記第1および第2の検索実行ステップにより回答候補が抽出されなかつた場合に前記情報源を検索することを特徴とする質問応答方法。

【請求項5】

コンピュータに電子化文書を含む情報源から検索質問文に応答して情報を検索させる質問応答プログラムであつて、

入力検索質問文に対して構文解析を行い、該入力検索質問文の構文構造を生成する構文 50

解析ステップと、

予め質問・回答パターン情報保持手段に保持された、検索質問文の構文構造と該検索質問文に対応する検索式とを対応付けた質問・回答パターン情報の中から前記構文解析ステップにおいて生成された入力検索質問文の構文構造と適合するものを特定する質問・回答パターン情報特定ステップと、

前記質問・回答パターン情報特定ステップにおいて特定された質問・回答パターン情報における検索式に基づいて、前記情報源を検索し、回答候補を抽出する第1の検索実行ステップと、

前記入力検索質問文に含まれる単語に対応する検索式、及び、前記入力検索質問文の構文構造に対応する検索式の少なくともいずれかを用いて前記情報源を検索し、回答候補を抽出する第2の検索実行ステップと、

前記入力検索質問文に基づく統計ベースでの前記情報源の検索を行い、回答候補を抽出する第3の検索実行ステップとを有し、

前記構文解析ステップは、疑問文である入力検索質問文を肯定文に変換し、当該肯定文の所定の単語を変数化した穴埋め肯定文を生成し、かつ当該穴埋め肯定文の構文構造を生成するものであり、

前記質問・回答パターン情報保持手段は、前記構文解析ステップによって生成された穴埋め肯定文の所定の単語を変数化した構文構造と当該穴埋め肯定文の構文構造に基づく文法ベースの検索式とを保持するものであり、

前記第1の検索実行ステップは、前記第2の検索実行ステップにより回答候補が抽出されなかった場合に前記情報源を検索し、

前記第3の検索実行ステップは、前記第1および第2の検索実行ステップにより回答候補が抽出されなかった場合に前記情報源を検索することを特徴とする質問応答プログラム。

【請求項6】

前記第3の検索実行ステップにおいて抽出された回答候補を検索するための検索式を生成する検索式生成ステップと、

前記入力検索質問文の構文構造と、前記検索式生成ステップにおいて生成された検索式とを対応付けた質問・回答パターン情報を生成し、前記質問・回答パターン情報保持手段へ登録する質問・回答パターン情報生成・登録ステップとを有することを特徴とする請求項5に記載の質問応答プログラム。

【請求項7】

前記検索式生成ステップは、前記第3の検索実行ステップにおいて抽出された回答候補のうち、所定の選択操作によって選択された回答候補を検索するための検索式を生成することを特徴とする請求項6に記載の質問応答プログラム。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、電子化文書を含む情報源から検索要求に対応する文を検索する質問応答システム、質問応答方法及び質問応答プログラムに関する。

【背景技術】

【0002】

近年、文書作成や情報収集を目的として、ウェブブラウザ上で入力された検索質問文に対して、テキスト形式の回答候補を提示する質問応答システムが研究されている。このようなシステムは、例えば特許文献1乃至3に開示されている。

特許文献1は、形態素解析技術と構文解析技術と意味解析技術を用いて、自然言語データベースから解答を導出する質問応答方式を開示している。特許文献2は、質問文を解析し、検索キーを検出して検索し、一次文書を作成し、これを解析して入力列にない付加的句を一次文書内に検出するものである。特許文献3は、検索語の近傍性を利用した文書検索部を持つ質問応答システムを開示している。いわゆるサーチエンジン等において実現され

10

20

30

40

50

ている検索システムは、入力されたキーワードを含む文を情報源から抽出して利用者に提示する。これに対し、質問応答システムは、検索質問文を解析して、キーワードと、人名に対する質問である等の質問のタイプとを特定し、この解析結果に基づいて情報源を検索して回答候補を抽出し、利用者に対して提示する。

【0003】

【特許文献1】特開昭61-221873号公報

【特許文献2】特開平7-56954号公報

【特許文献3】特開2004-127131号公報

【発明の開示】

【発明が解決しようとする課題】

10

【0004】

上述した従来の質問応答システムは、検索結果から回答候補を抽出する場合、情報源から回答候補を含む文を複数抽出し、キーワードの出現頻度等の統計ベースでの処理を行うとともに、その統計ベースでの処理の結果に基づいて回答候補の順位付けを行い、上位の回答候補を提示する。しかし、統計ベースでの処理の結果のみに基づいて回答候補が特定されると、利用者が意図した使われ方をしていないキーワードを含んだ回答候補が上位に提示されてしまうこともあり、検索質問文に対応する適切な回答候補が提示されない場合がある。このため、質問応答システムが出力する回答候補の精度が低いという課題がある。

【0005】

20

このような問題を解決すべく、検索質問文と回答候補を含む文の双方について構文解析を行って構文構造を表す構文木を作成し、この構文木を用いて情報源の検索を行うことにより回答候補の抽出を精度良く行う質問応答システムが提案されている。しかしながら、この質問応答システムでは、情報源に含まれる全ての文について、予め構文解析を行っておく必要があり、更には、構文構造が検索質問文とは異なるものの意味として類似する文が、回答候補を含む文であると判断されない場合がある。このため、統計ベースの質問応答システムと同様に、質問応答システムが出力する回答候補の精度が低いという課題がある。

【0006】

30

本発明の目的は、上述した問題を解決するものであり、検索質問文に対する適切な回答候補の抽出が可能な質問応答システム、質問応答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】

【0007】

本発明に係る、電子化文書を含む情報源から検索質問文に応答して情報を検索する質問応答システムは、検索質問文の構文構造と該検索質問文に対応する検索式とを対応付けた質問・回答パターン情報を保持する質問・回答パターン情報保持手段と、入力検索質問文に対して構文解析を行い、該入力検索質問文の構文構造を生成する構文解析手段と、質問・回答パターン情報保持手段に保持された質問・回答パターン情報の中から前記構文解析手段により生成された入力検索質問文の構文構造と適合するものを特定する質問・回答パターン情報特定手段と、前記質問・回答パターン情報特定手段により特定された質問・回答パターン情報における検索式に基づいて、前記情報源を検索し、回答候補を抽出する第1の検索実行手段とを有する。

40

【0008】

この構成により、検索質問文の構文構造と該検索質問文に対応する検索式とを対応付けた質問・回答パターン情報から入力検索質問文の構文構造と適合するものが特定され、その特定された質問・回答パターン情報における検索式に基づいて情報源が検索されて回答候補が抽出されるため、過去の検索結果を利用した検索質問文に対する適切な回答候補の抽出が可能となる。

【0009】

50

好ましくは、本発明に係る質問応答システムは、前記入力検索質問文に含まれる単語に対応する検索式、及び、前記入力検索質問文の構文構造に対応する検索式の少なくともいずれかを用いて前記情報源を検索し、回答候補を抽出する第2の検索実行手段を有し、前記第1の検索実行手段が、前記第2の検索実行手段により回答候補が抽出されなかった場合に、前記情報源を検索する。

【0010】

この構成により、入力検索質問文に含まれる単語に対応する検索式や入力検索質問文の構文構造に対応する検索式を用いた通常の検索において、回答候補が抽出されなかった場合に、質問・回答パターン情報における検索式に基づいて情報源が検索されるため、情報検索の柔軟性を高め、適切な回答候補の抽出が可能となる。

10

【0011】

好ましくは、本発明に係る質問応答システムは、前記入力検索質問文に基づく統計ベースでの前記情報源の検索を行い、回答候補を抽出する第3の検索実行手段と、前記第3の検索実行手段により抽出された回答候補を検索するための検索式を生成する検索式生成手段と、前記入力検索質問文の構文構造と、前記検索式生成手段により生成された検索式とを対応付けた質問・回答パターン情報を生成し、前記質問・回答パターン情報保持手段へ登録する質問・回答パターン情報生成・登録手段とを有する。

【0012】

好ましくは、本発明に係る質問応答システムは、前記第3の検索実行手段が、前記第1の検索実行手段により回答候補が抽出されなかった場合、及び、前記質問・回答パターン情報特定手段により、前記質問・回答パターン情報保持手段に保持された質問・回答パターン情報の中から前記入力検索質問文の構文構造と適合するものが特定されなかった場合の少なくともいずれかにおいて、前記情報源を検索する。

20

【0013】

好ましくは、本発明に係る質問応答システムは、前記検索式生成手段が、前記第3の検索実行手段により抽出された回答候補のうち、所定の選択操作によって選択された回答候補を検索するための検索式を生成する。

【0014】

入力検索質問文の構文構造が考慮されない統計ベースでの検索によって抽出された回答候補であっても、利用者等により選択された回答候補であれば、入力検索質問文の構文構造との間に何らかの関係があると考えられるため、選択された回答候補を検索するための検索式が生成され、入力検索質問文の構文構造と対応付けられて質問・回答パターン情報として登録されることにより、その後の情報検索において、質問・回答パターン情報に基づく適切な回答候補の抽出が可能となる。

30

【0015】

好ましくは、本発明に係る質問応答システムは、前記質問・回答パターン情報における検索質問文の構文構造及び検索式が、所定の単語が変数に置換されたものである。この構成により、変数化された検索質問文の構文構造及び検索式に基づくより柔軟な情報検索が可能となる。

【0016】

本発明に係る、電子化文書を含む情報源から検索質問文に応答して情報を検索する質問応答方法は、入力検索質問文に対して構文解析を行い、該入力検索質問文の構文構造を生成する構文解析ステップと、予め質問・回答パターン情報保持手段に保持された、検索質問文の構文構造と該検索質問文に対応する検索式とを対応付けた質問・回答パターン情報の中から前記構文解析ステップにおいて生成された入力検索質問文の構文構造と適合するものを特定する質問・回答パターン情報特定ステップと、前記質問・回答パターン情報特定ステップにおいて特定された質問・回答パターン情報における検索式に基づいて、前記情報源を検索し、回答候補を抽出する第1の検索実行ステップとを有する。

40

【0017】

好ましくは、本発明に係る質問応答方法は、前記入力検索質問文に含まれる単語に対応

50

する検索式、及び、前記入力検索質問文の構文構造に対応する検索式の少なくともいずれかを用いて前記情報源を検索し、回答候補を抽出する第2の検索実行ステップを有し、前記第1の検索実行ステップが、前記第2の検索実行ステップにより回答候補が抽出されなかった場合に、前記情報源を検索する。

好ましくは、本発明に係る質問応答方法は、前記入力検索質問文に基づく統計ベースでの前記情報源の検索を行い、回答候補を抽出する第3の検索実行ステップと、前記第3の検索実行ステップにおいて抽出された回答候補を検索するための検索式を生成する検索式生成ステップと、前記入力検索質問文の構文構造と、前記検索式生成ステップにおいて生成された検索式とを対応付けた質問・回答パターン情報を生成し、前記質問・回答パターン情報保持手段へ登録する質問・回答パターン情報生成・登録ステップとを有する。

10

【0018】

好ましくは、本発明に係る質問応答方法は、前記第3の検索実行ステップが、前記第1の検索実行ステップにおいて回答候補が抽出されなかった場合、及び、前記質問・回答パターン情報特定ステップにおいて前記質問・回答パターン情報保持手段に保持された質問・回答パターン情報の中から前記入力検索質問文の構文構造と適合するものが特定されなかった場合の少なくともいずれかにおいて、前記情報源を検索する。

【0019】

好ましくは、本発明に係る質問応答方法は、前記検索式生成ステップが、前記第3の検索実行ステップにおいて抽出された回答候補のうち、所定の選択操作によって選択された回答候補を検索するための検索式を生成する。

20

【0020】

好ましくは、本発明に係る質問応答方法は、前記質問・回答パターン情報における検索質問文の構文構造及び検索式が、所定の単語が変数に置換されたものである。

【0021】

本発明に係る、コンピュータに電子化文書を含む情報源から検索質問文に応答して情報を検索させる質問応答プログラムは、入力検索質問文に対して構文解析を行い、該入力検索質問文の構文構造を生成する構文解析ステップと、予め質問・回答パターン情報保持手段に保持された、検索質問文の構文構造と該検索質問文に対応する検索式とを対応付けた質問・回答パターン情報の中から前記構文解析ステップにおいて生成された入力検索質問文の構文構造と適合するものを特定する質問・回答パターン情報特定ステップと、前記質問・回答パターン情報特定ステップにおいて特定された質問・回答パターン情報における検索式に基づいて、前記情報源を検索し、回答候補を抽出する第1の検索実行ステップとを有する。

30

【0022】

好ましくは、本発明に係る質問応答プログラムは、前記入力検索質問文に含まれる単語に対応する検索式、及び、前記入力検索質問文の構文構造に対応する検索式の少なくともいずれかを用いて前記情報源を検索し、回答候補を抽出する第2の検索実行ステップを有し、前記第1の検索実行ステップが、前記第2の検索実行ステップにより回答候補が抽出されなかった場合に、前記情報源を検索する。

【0023】

好ましくは、本発明に係る質問応答プログラムは、前記入力検索質問文に基づく統計ベースでの前記情報源の検索を行い、回答候補を抽出する第3の検索実行ステップと、前記第3の検索実行ステップにおいて抽出された回答候補を検索するための検索式を生成する検索式生成ステップと、前記入力検索質問文の構文構造と、前記検索式生成ステップにおいて生成された検索式とを対応付けた質問・回答パターン情報を生成し、前記質問・回答パターン情報保持手段へ登録する質問・回答パターン情報生成・登録ステップとを有する。

40

【0024】

好ましくは、本発明に係る質問応答プログラムは、前記第3の検索実行ステップが、前記第1の検索実行ステップにおいて回答候補が抽出されなかった場合、及び、前記質問・

50

回答パターン情報特定ステップにおいて前記質問・回答パターン情報保持手段に保持された質問・回答パターン情報の中から前記入力検索質問文の構文構造と適合するものが特定されなかった場合の少なくともいずれかにおいて、前記情報源を検索する。

【0025】

好ましくは、本発明に係る質問応答プログラムは、前記検索式生成ステップが、前記第3の検索実行ステップにおいて抽出された回答候補のうち、所定の選択操作によって選択された回答候補を検索するための検索式を生成する。

【0026】

好ましくは、本発明に係る質問応答プログラムは、前記質問・回答パターン情報における検索質問文の構文構造及び検索式が、所定の単語が変数に置換されたものである。

10

【発明の効果】

【0027】

本発明によれば、検索質問文の構文構造と該検索質問文に対応する検索式とを対応付けた質問・回答パターン情報から入力検索質問文の構文構造と適合するものが特定され、その特定された質問・回答パターン情報における検索式に基づいて情報源が検索されて回答候補が抽出されるため、過去の検索結果を利用した検索質問文に対する適切な回答候補の抽出が可能となる。

【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は、質問応答システムの構成例を示す図である。同図に示す質問応答システム100は、検索対象文書である電子化文書を含む情報源である検索対象文書データベース(DB)200から検索質問文に応答して情報を検索するものであり、検索質問文解析部110、情報検索部120、情報抽出部130、回答特定部140、質問・回答パターン情報DB150及び回答提示部160を有する。これらのうち、情報検索部120は、検索式生成部122及び文書検索部124を有し、情報抽出部130は、固有表現抽出部132を有する。また、回答特定部140は、回答候補抽出部142及びパターン情報生成部144を有し、回答提示部160は、文書要約部162を有する。

20

【0029】

次に、質問応答システム100による動作を、フローチャートを参照しつつ説明する。図2及び図3は、質問応答システムの動作のフローチャートである。利用者の操作等に応じた検索質問文が入力されると(S101)、検索質問文解析部110は、その入力された検索質問文(入力検索質問文)について、形態素解析を行い、その形態素解析の結果に基づいて、入力検索質問文がどのようなタイプの質問であるのかを解析する(S102)。質問タイプは、例えば、人名についての質問、場所についての質問、数量についての質問、日時についての質問等を挙げることができる。質問タイプの解析方法としては、人手で記述された解析ルールを用いる方法と機械学習アルゴリズムで解析する方法とに分類できる。

30

【0030】

また、検索質問文解析部110は、質問タイプの解析(S102)とともに、入力検索質問文に対する形態素解析の結果に基づいて、単語ベースでの入力検索質問文の解析を行い、質問に関係する単語を抽出する(S103)。情報検索部120内の検索式生成部122は、S103において抽出された単語を含んだ単語ベースの検索式を生成する(S104)。そして、情報検索部120内の文書検索部124は、S104において生成された単語ベースの検索式を用いて、検索対象文書DB200を検索し、単語ベースの検索式に含まれる単語を含んだ電子化文書を抽出する。更に、文書検索部124は、抽出した電子化文書から検索式に含まれる単語が存在する文と、その前後の複数の文との集合(パッセージ)を生成する(S105)

40

【0031】

図4は、入力検索質問文、検索式に含まれる単語、及び、パッセージの対応関係の一例を示す図である。入力検索質問文が「特別自治区で生まれた大統領は誰ですか？」というものである場合、質問に関する単語として「特別自治区」及び「大統領」が抽出され、これらを含む単語ベースの検索式が生成される。そして、この単語ベースの検索式を用いた文書検索が行われることにより、「特別自治区」及び「大統領」を含む文と、その前後の文との集合（パッセージ）が生成される。

【0032】

再び、図2に戻って説明する。情報抽出部130内の固有表現抽出部132は、S105において生成されたパッセージから固有表現を抽出し、この固有表現を、回答候補を含む文（解答候補文）とする（S106）。

10

【0033】

更に、回答特定部140内の回答候補抽出部142は、S106においてパッセージから抽出された各回答候補文について、LFG（Lexical Functional Grammar）理論に基づく構文解析を行い、解析結果としての文の格構造、時制、様相、話法等を表現したfunctional structure(f-structure (fs))を生成する（S107）。

【0034】

LFG理論では、自然言語文を解析することにより、解析結果として文の意味内容が属性-属性値ペアの入れ子構造（マトリックス構造）であるf-structureによって表現される。属性には、SUBJ（主語）、OBJ（目的語）、PRED（述語）、COMP（文が入れ子となっていることを表す）、XCOMP（文が入れ子になっており、且つ、入れ子の文が主語に係っている）、OBL（主語、目的語以外の格要素）、ADJUNCT（助詞を介さない修飾節）等の文の構造を表すものや、ADDRESS（モダリティなど）、TENSE（時制など）等の文の意味を分類するもの等があり、これら属性が階層構造を形成している。また、属性の属性値には、例えばADDRESS属性の場合における、possible（可能表現）、approach（現在完了に相当する表現）、expectation（予期表現）等の値が表される。このf-structureに格納されている言語に関する情報は、一般的の係り受け解析レベルの構文木に比べて情報量が豊富である。LFG理論を応用した実用的な自然言語アプリケーションの研究としては、Anette Frank著、「From Parallel Grammar Development towards Machine Translation」、In Proceedings of MT Summit VII、pp. 134-142、1999年があり、この文献では、f-structureを中間言語とみなした機械翻訳手法が提案されている。

20

【0035】

また、検索質問文解析部110は、質問タイプの解析（S102）及び単語ベースでの入力検索質問文の解析（S103）とともに、疑問文である入力検索質問文を肯定文に変換して、穴埋め肯定文を生成する（S108）。例えば、入力検索質問文が「特別自治区で生まれた大統領は誰ですか？」である場合、穴埋め肯定文は「特別自治区で生まれた大統領はAです。」となる。

【0036】

次に、検索質問文解析部110は、S108において生成した穴埋め肯定文について、LFG理論に基づく構文解析を行い、解析結果としてのf-structureを生成する（S109）。図5は、穴埋め肯定文「特別自治区で生まれた大統領はAです。」に対応するf-structureである。

30

【0037】

再び、図2に戻って説明する。情報検索部120内の検索式生成部122は、S109において生成された、穴埋め肯定文に対応するf-structureをXML文書に変換し（S110）、当該XML文書に変換後のf-structureに基づいて、文法ベースでの検索式であるXPathを生成する（S111）。図6は、穴埋め肯定文「特別自治区で生まれた大統領はAです。」に対応するf-structureが図5に示すように生成された場合におけるXPathを示す図である。図6に示すXPathは、fsノード下に子孫ノードとしてSUBJが存在しているかを調べるために、軸要素"descendant"とノードチェックの":SUBJ"を指定している。更に、指定された範囲において、要素名がそれぞれ"PRED"、"ADJUNCT"が並列に存在してい

40

50

るようなXML文書が適合する。SUBJ属性の下にPRED属性とADJUNCT属性（修飾節）が並列に存在するということは、「主語が修飾されている」ということを意味し、図6に示すXPathは「主語が修飾されているような文」の検索に用いられる検索式となる。

【0038】

S102、S107及びS111の動作終了後、図3に示す動作に移行し、回答特定部140内の回答候補抽出部142は、文法ベースでの回答候補の抽出を行う（S201）。具体的には、回答候補抽出部142は、S107において生成された各回答候補文に対応するf-structureの中からS111において生成されたXPathに適合するものを抽出する。例えば、XPathが図6に示すものである場合、「特別自治区で生まれた大統領はです。」との回答候補文に対応するf-structureは、図7の抽出部分に示すように、SUBJ属性の下にPRED属性及びADJUNCT属性が存在するため、図6に示すXPathに適合するが、「大統領は説明した。」との回答候補文に対応するf-structureは、図8の抽出部分に示すように、SUBJ属性の下にADJUNCT属性が存在しないため、図6に示すXPathに適合しない。更に、回答候補抽出部142は、f-structureがXPathに適合する各回答候補文からS102で得られた質問タイプに適合する単語を抽出して回答候補とする。

【0039】

次に、回答候補抽出部142は、S201において抽出された回答候補が存在するか否かを判定する（S202）。回答候補が存在する場合には、回答提示部150は、その抽出された回答候補を提示する（S203）。この際、回答提示部150内の文書要約部152を用いて、回答候補が含まれていた回答候補文の要約を生成し、回答候補とともに提示してもよい。要約の生成方法に関しては、回答候補中に含まれる単語を含む文を提示するなど、従来技術を用いることができる。

【0040】

次に、回答特定部140内のパターン情報生成部144は、S203において提示された回答候補のいずれかが利用者によって選択されたか否かを判定する（S204）。

【0041】

利用者によって回答候補が選択された場合、パターン情報生成部144は、S108において生成された穴埋め肯定文について所定の単語を変数化したf-structureを生成し、この所定の単語を変数化したf-structureと、S111において生成されたXPathとを対応付けた質問・回答パターン情報を生成し、質問・回答パターン情報DB150へ登録する（S205）。

【0042】

例えば、S108において、穴埋め肯定文「特別自治区で生まれた大統領はAです。」が生成された場合、S205では、「特別自治区」及び「大統領」を変数化した穴埋め肯定文「Xで生まれたYはAです。」に対応するf-structureが生成され、更に、S111において生成されたXPathと対応付けられて質問・回答パターン情報DB150へ登録される。なお、この場合、「特別自治区」を変数化した「X」については、場所を表す名詞である等の属性に関する情報が付加されるようにしてもよい。これにより、その後の情報検索において、例えば、入力検索質問文が「東京で生まれた政治家は誰ですか？」であった場合には、穴埋め肯定文「Xで生まれたYはAです。」に対応するf-structureに対応付けられたXPathが検索に用いられることが可能となる。

【0043】

一方、S202において抽出した回答候補が存在しないと判定された場合、回答特定部140内の回答候補抽出部142は、質問・回答パターン情報DB150を検索する（S206）。次に、回答候補抽出部142は、検索の結果、S101において入力された入力検索質問文に対応する質問・回答パターン情報、換言すれば、S109において生成された穴埋め肯定文に対応するf-structureと同一構造のf-structureを有する質問・回答パターン情報が存在するか否かを判定する（S207）。

【0044】

S109において生成された穴埋め肯定文に対応するf-structureと同一構造のf-structure

10

20

30

40

50

tureを有する質問・回答パターン情報が存在する場合、回答候補抽出部142は、文法ベースでの回答候補の抽出を行う(S208)。具体的には、S201とほぼ同様の動作であり、回答候補抽出部142は、S107において生成された各回答候補文に対応するf-structureの中から、S109において生成された穴埋め肯定文に対応するf-structureと同一構造のf-structureを有する質問・回答パターン情報に含まれるXPathに適合するものを抽出する。更に、回答候補抽出部142は、f-structureがXPathに適合する各回答候補文からS102で得られた質問タイプに適合する単語を抽出して回答候補とする。

【0045】

次に、回答候補抽出部142は、S208において抽出された回答候補が存在するか否かを判定する(S209)。回答候補が存在する場合には、回答提示部150は、その抽出された回答候補を提示する(S203)。

【0046】

次に、回答特定部140内のパターン情報生成部144は、S203において提示された回答候補のいずれかが利用者によって選択されたか否かを判定する(S204)。

【0047】

利用者によって回答候補が選択された場合、パターン情報生成部144は、S109において生成された穴埋め肯定文に対応するf-structureについても所定の単語を変数化し、この所定の単語を変数化したf-structureと、S208において用いられたXPathとを対応付けて質問・回答パターン情報を生成し、質問・回答パターン情報DB150へ登録する(S205)。

【0048】

一方、S209において、回答候補が存在しないと判定した場合、回答候補抽出部142は、統計ベースでの回答候補の抽出を行う(S210)。具体的には、回答候補抽出部142は、S105において生成された各メッセージの中から、S103において抽出された単語の出現頻度の多い所定数を選択し、その選択したメッセージから、S102で得られた質問タイプに適合する単語を抽出して回答候補とする。

【0049】

次に、回答候補抽出部142は、S210において抽出された回答候補が存在するか否かを判定する(S211)。回答候補が存在しない場合には、一連の動作が終了する。一方、回答候補が存在する場合には、回答提示部150は、その抽出された回答候補を提示する(S203)。

【0050】

次に、回答特定部140内のパターン情報生成部144は、S203において提示された回答候補のいずれかが利用者によって選択されたか否かを判定する(S204)。

【0051】

利用者によって回答候補が選択された場合、パターン情報生成部144は、選択された回答候補を含んでいたメッセージ内の文について、構文解析を行ってf-structureを生成する。次に、パターン情報生成部144は、このf-structureから利用者によって選択された回答候補を検索するためのXPathを生成し、その生成したXPathについて所定の単語を変数化する。更に、パターン情報生成部144は、S109において生成された穴埋め肯定文に対応するf-structureについても所定の単語を変数化し、これら所定の単語が変数化されたXPath及びf-structureを対応付けた質問・回答パターン情報を、質問・回答パターン情報DB150へ登録する(S205)。

【0052】

例えば、図4に示す入力検索質問文が入力されてメッセージが生成された場合を考える。この場合、回答候補抽出部142は、S210における統計ベースでの回答候補の抽出において、質問タイプが人名についての質問であることを考慮し、メッセージ1から「ジェームズ」、メッセージ2から「トーマス」、メッセージ3から「ジェファーソン」をそれぞれ回答候補として抽出する。そして、これら回答候補が提示され、利用者が「ジェームズ」を選択すると、パターン情報生成部144は、S205において、メッセージ1に

10

20

30

40

50

含まれる文「特別自治区であるセカンドサウスで生まれたジェームズ氏は、後に大統領となった。」に対応するf-structure(図9参照)を生成し、このf-structureから選択された回答候補である「ジェームズ」を、「特別自治区」、「大統領」の単語や、述語である「生まれる」等の構文関係から検索するためのXPath(図10参照)を生成する。更に、パターン情報生成部144は、生成したXPathにおける単語「特別自治区」を変数「X」に置き換えたXPath(図12参照)を生成する。なお、この場合、「特別自治区」を変数化した「X」については、場所を表す名詞である等の属性に関する情報が付加されるようにしてもよい。

【0053】

そして、パターン情報生成部144は、穴埋め肯定文「特別自治区で生まれた大統領はAです。」について「特別自治区」及び「大統領」を変数化した穴埋め肯定文「Xで生まれたYはAです。」に対応するf-structureを生成し、XPathと対応付けて質問・回答パターン情報として、質問・回答パターン情報D B 150へ登録する。

10

【0054】

ここで、図10と図12のXPathを比較すると、図10のXPathが検索に用いられた場合には、図11に示すように、「特別自治区で生まれた大統領はAです。」や「特別自治区であるセカンドサウスで生まれたジェームズ氏は、後に大統領になった。」がXPathに適合する文となる。これに対し、図12のXPathが検索に用いられた場合には、「Xで生まれたYはA(回答候補)です。」や「XであるZで生まれたA(回答候補)は、後にYになった。」がXPathに適合する文となる。従って、その後の検索では、例えば、入力検索質問文が「東京で生まれた政治家は誰ですか？」であった場合には、穴埋め肯定文「Xで生まれたYはAです。」と「XであるZで生まれたAは、後にYになった。」に対応するf-structureに対応付けられたXPathが検索に用いられることが可能となる。

20

【0055】

このように、本実施形態の質問応答システム100は、単語ベースの検索が行われ、更に文法ベースでの回答候補の抽出が行われた場合において回答候補が抽出されなかった際に、過去の検索結果に基づく質問・回答パターン情報から入力検索質問文を変換した穴埋め肯定文に対応するf-structureと同一構造のf-structureを有する質問・回答パターン情報を特定し、その特定した質問・回答パターン情報におけるXPathに基づいて文法ベースでの回答候補の抽出を行い、単語ベースの検索によって得られた回答候補から適切な回答候補を抽出する。従って、過去の検索結果を利用した柔軟性のある適切な回答候補の抽出が可能となる。

30

【0056】

また、質問応答システム100は、質問・回答パターン情報に入力検索質問文を変換した穴埋め肯定文に対応するf-structureと同一構造のf-structureを有するものが存在しない場合や、質問・回答パターン情報におけるXPathに基づく回答候補の抽出ができなかつた場合には、統計ベースでの回答候補の抽出を行い、更に、利用者によって回答候補が選択された場合には、その選択された回答候補を検索するためのXPathを生成し、入力検索質問文を変換した穴埋め肯定文に対応するf-structureと対応付けた質問・回答パターン情報を質問・回答パターン情報D B 150へ登録する。従って、入力検索質問文を変換した穴埋め肯定文に対応するf-structureが考慮されない統計ベースでの回答候補の抽出であっても、利用者等により選択された回答候補であれば、入力検索質問文との間に何らかの関係があるとみなされて、質問・回答パターン情報が生成、登録されることになり、その後の情報検索において、質問・回答パターン情報に基づく適切な回答候補の抽出が可能となる。更には、質問応答システム100では、質問・回答パターン情報におけるf-structure及びXPathにおいて、所定の単語が変数に置換されており、より柔軟な情報検索が可能となる。

40

【0057】

以上、本発明の好ましい実施の形態について詳述したが、本発明は、かかる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発明の要旨の範囲内におい

50

て、種々の変形、変更が可能である。

【産業上の利用可能性】

【0058】

以上、説明したように、本発明に係る質問応答システム、質問応答方法及び質問応答プログラムによれば、検索質問文に対する適切な回答候補の抽出が可能という効果を奏し、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】

【0059】

【図1】質問応答システムの構成を示す図である。

【図2】質問応答システムの動作の第1のフローチャートである。

10

【図3】質問応答システムの動作の第2のフローチャートである。

【図4】入力検索質問文、検索式に含まれる単語、及び、パッセージの対応関係の一例を示す図である。

【図5】穴埋め肯定文とf-structureの対応関係の一例を示す図である。

【図6】第1のXPathを示す図である。

【図7】第1のXPathに適合するf-structureを示す図である。

【図8】第1のXPathに適合しないf-structureを示す図である。

【図9】回答候補文とf-structureの対応関係の一例を示す図である。

【図10】第2のXPathを示す図である。

【図11】第2のXPathに適合するf-structureを示す図である。

20

【図12】第3のXPathを示す図である。

【図13】第3のXPathに適合するf-structureを示す図である。

【符号の説明】

【0060】

100 質問応答システム

110 検索質問文解析部

120 情報検索部

122 検索式生成部

124 文書検索部

130 情報抽出部

30

132 固有表現抽出部

140 回答特定部

142 回答候補抽出部

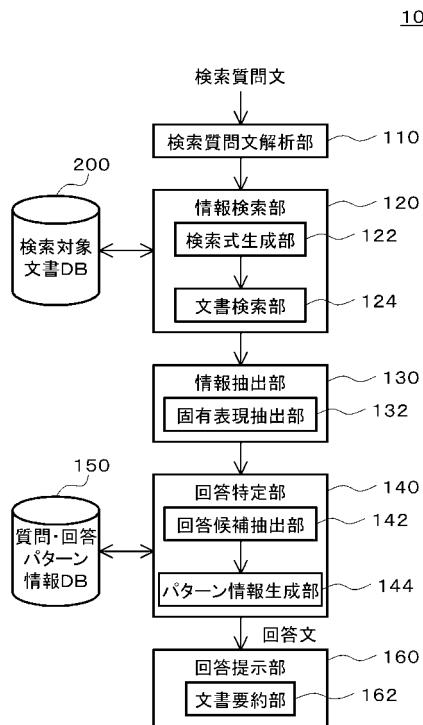
144 パターン情報生成部

150 質問・回答パターン情報D B

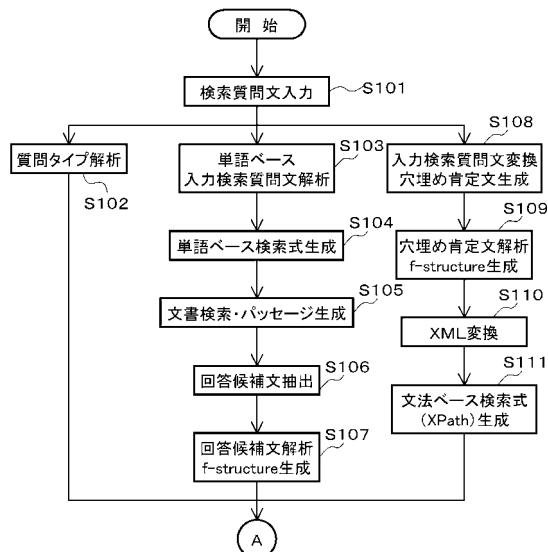
160 回答提示部

162 文書要約部

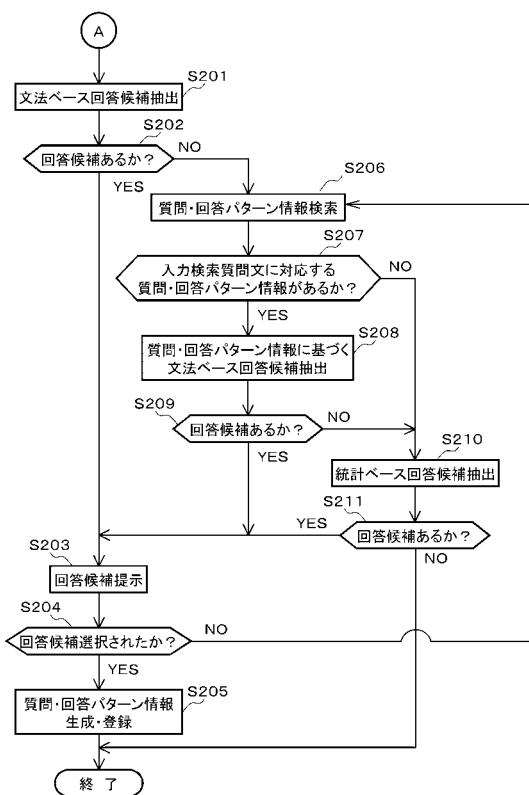
【図1】



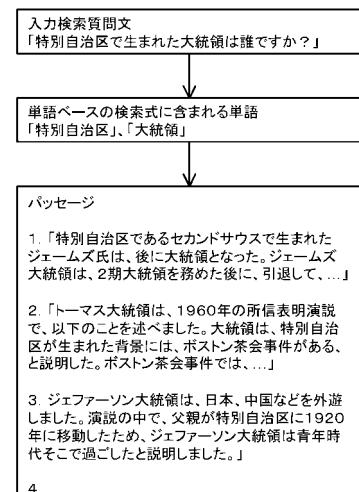
【図2】



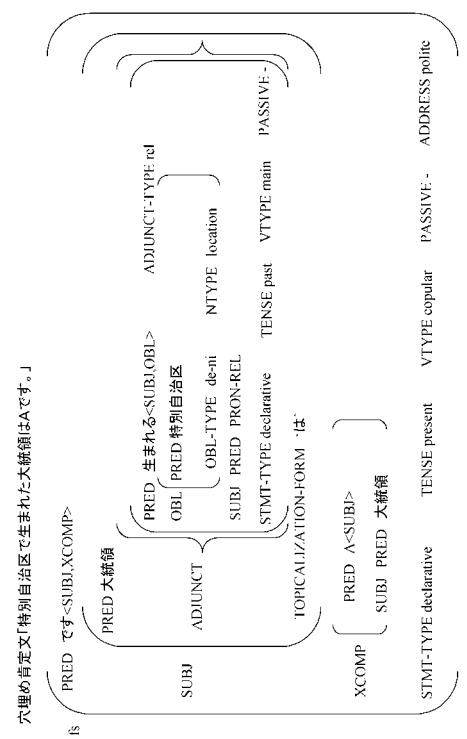
【図3】



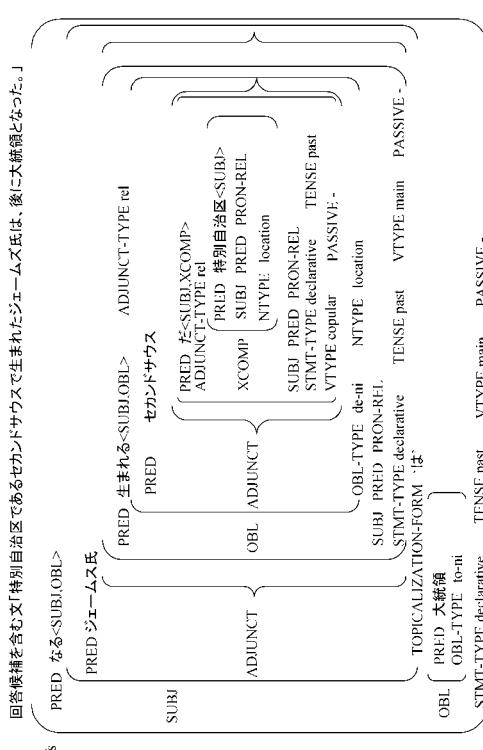
【図4】



【図5】



【図9】



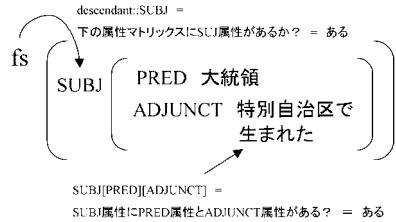
【図6】

fs [descendant::SUBJ[PRED][ADJUNCT]]

= fs下にSUBJ属性があり、且つ、SUBJ属性にPRED属性とADJUNCT属性が存在
= 主語が修飾されている

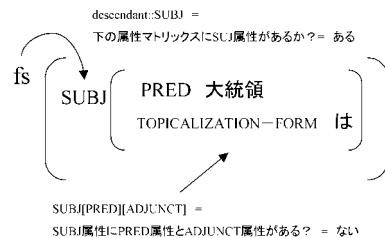
【図7】

マッチする文「特別自治区で生まれた大統領はOOです。」



【図8】

マッチしない文「大統領は説明した。」



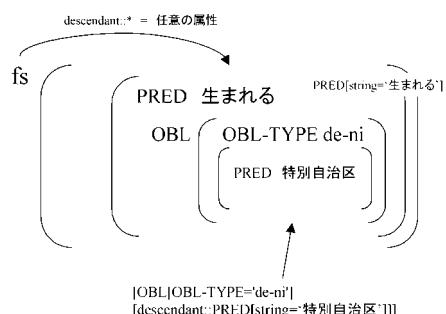
【図10】

fs [descendant::*[PRED[string='生まれる']][OBL [OBL-TYPE='de-ni']][descendant::PRED[string='特別自治区']]])

= 任意の属性の下に「生まれる」という文字列を有するPRED属性と「特別自治区」の文字列を有し、且つ、OBL-TYPEが「de-ni」の値を有するOBL属性が存在
= 「特別自治区」が何らかの形で「で」か「に」を介して「生まれる」に係る文

【図11】

マッチする文「特別自治区で生まれた大統領はAです。」
マッチする文「特別自治区であるセカンドサウスで生まれたジェームズ氏は、後に大統領となつた。」



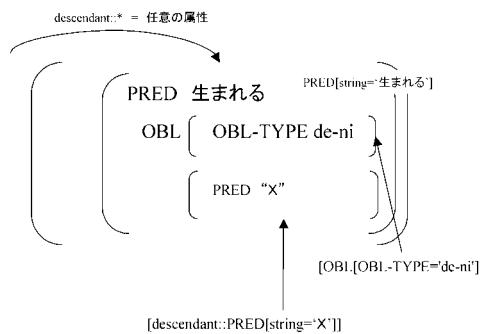
【図12】

fs [descendant::*[PRED[string='生まれる']][OBL [OBL-TYPE='de-ni']][descendant::PRED[string='X']]])

= 任意の属性の下に「生まれる」という文字列を有するPRED属性と「X」の文字列を有する持つ何らかの属性が存在
= 「X」が何らかの形で「生まれる」に係り、「に」か「で」を有する格要素を含む文

【図 1 3】

マッチする文「Xで生まれたYはA(回答候補)です。」
マッチする文「XであるZで生まれたA(回答候補)は、後にYになった。」



フロントページの続き

(72)発明者 増田 佳弘

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

(72)発明者 山田 敏哉

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

(72)発明者 桑原 隆比古

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

審査官 鈴木 和樹

(56)参考文献 特開平08-161354 (JP, A)

特開平07-244669 (JP, A)

特開2002-297592 (JP, A)

特開昭63-216175 (JP, A)

(58)調査した分野(Int.Cl., DB名)

G06F 17/30