

(19) 日本国特許庁(JP)

(12) 特許公報(B2)

(11) 特許番号

特許第3981734号

(P3981734)

(45) 発行日 平成19年9月26日(2007.9.26)

(24) 登録日 平成19年7月13日(2007.7.13)

(51) Int. Cl.

G06F 17/30 (2006.01)

F I

G06F 17/30 330C

G06F 17/30 180A

請求項の数 8 (全 17 頁)

(21) 出願番号	特願2003-391938 (P2003-391938)	(73) 特許権者	301022471
(22) 出願日	平成15年11月21日(2003.11.21)		独立行政法人情報通信研究機構
(65) 公開番号	特開2005-157524 (P2005-157524A)		東京都小金井市貫井北町4-2-1
(43) 公開日	平成17年6月16日(2005.6.16)	(74) 代理人	100119161
審査請求日	平成15年11月21日(2003.11.21)		弁理士 重久 啓子
		(72) 発明者	熊本 忠彦
			東京都小金井市貫井北町4-2-1 独立 行政法人通信総合研究所内
		(72) 発明者	村田 真樹
			東京都小金井市貫井北町4-2-1 独立 行政法人通信総合研究所内
		審査官	紀田 馨

最終頁に続く

(54) 【発明の名称】 質問応答システムおよび質問応答処理方法

(57) 【特許請求の範囲】

【請求項1】

自然言語で表現された質問文データを入力し、解答の検索対象である文書データ群から質問文データに対する解答を出力する質問応答システムにおいて、

入力された質問文データからキーワードを抽出し、前記文書データ群から前記キーワードを含む文書データを検索抽出する文書検索手段と、

前記文書データから解答となりうる言語表現を解答候補として抽出する解答候補抽出手段と、

解答候補の言語表現についての表記的または意味的な類型を示す2以上の解答タイプを記憶しておき、前記質問文データに対する解答の解答タイプとしてユーザによって選択入力された解答タイプを取得する解答タイプ入力手段と、

前記解答候補各々が、前記2以上の解答タイプのどの解答タイプであるかを判定する解答タイプ判定手段と、

前記判定された解答タイプをもとに前記解答候補を分類し、前記解答タイプごとに分類された解答候補の全部または一部を、前記2以上の解答タイプを見出し項目として構成し、かつ前記入力された解答タイプの見出しを先頭項目として配置した表形式データとして出力する解答表出力手段とを備える

ことを特徴とする質問応答システム。

【請求項2】

自然言語で表現された質問文データを入力し、解答の検索対象である文書データ群から

10

20

質問文データに対する解答を出力する質問応答システムにおいて、

入力された質問文データからキーワードを抽出し、前記文書データ群から前記キーワードを含む文書データを検索抽出する文書検索手段と、

前記文書データから解答となりうる言語表現を解答候補として抽出する解答候補抽出手段と、

解答候補の言語表現についての表記的または意味的な類型を示す2以上の解答タイプを記憶しておき、前記質問文データの言語表現を解析し、確率にもとづいた機械学習法であって順位付けが可能な数値を算出できるアルゴリズムを用いて、前記質問文データに対する解答が前記解答タイプであることの確信度を推定する解答タイプ推定手段と、

前記解答候補各々が、前記2以上の解答タイプのどの解答タイプであるかを判定する解答タイプ判定手段と、 10

前記判定された解答タイプをもとに前記解答候補を分類し、前記解答タイプごとに分類された解答候補の全部または一部を、前記2以上の解答タイプを見出し項目として構成し、かつ前記解答タイプ推定手段で算出された確信度の高い順に対応する解答タイプの見出し項目を配置した表形式データとして出力する解答表出力手段とを備える

ことを特徴とする質問応答システム。

【請求項3】

前記解答タイプ判定手段は、解答候補の言語表現と前記2以上の解答タイプとの対応関係を示す学習データが与えられた機械学習法によるアルゴリズムを用いて、前記解答候補各々が前記2以上の解答タイプのどの解答タイプになりやすいかを判定する 20

ことを特徴とする請求項1または請求項2のいずれか一項に記載の質問応答システム。

【請求項4】

前記解答表出力手段は、前記解答タイプごとに前記解答タイプ推定手段によって推定された確信度を表示した前記表形式データを作成する

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5】

文書検索手段、解答候補抽出手段、解答タイプ入力手段、解答タイプ判定手段、および解答表出力手段を備えるコンピュータが、自然言語で表現された質問文データを入力し、解答の検索対象である文書データ群から質問文データに対する解答を出力する質問応答処理方法において、 30

前記文書検索手段が、入力された質問文データからキーワードを抽出し、前記文書データ群から前記キーワードを含む文書データを検索抽出する処理過程と、

前記解答候補抽出手段が、前記文書データから解答となりうる言語表現を解答候補として抽出する処理過程と、

前記解答タイプ入力手段が、解答候補の言語表現についての表記的または意味的な類型を示す2以上の解答タイプを記憶しておき、前記質問文データに対する解答の解答タイプとしてユーザによって選択入力された解答タイプを取得する処理過程と、

前記解答タイプ判定手段が、前記解答候補各々が前記2以上の解答タイプのどの解答タイプであるかを判定する処理過程と、

前記解答表出力手段が、前記判定された解答タイプをもとに前記解答候補を分類し、前記解答タイプごとに分類された解答候補の全部または一部を、前記2以上の解答タイプを見出し項目として構成し、かつ前記入力された解答タイプの見出しを先頭項目として配置する表形式データとして出力する処理過程とを備える 40

ことを特徴とする質問応答処理方法。

【請求項6】

文書検索手段、解答候補抽出手段、解答タイプ推定手段、解答タイプ判定手段、および解答表出力手段を備えるコンピュータが、自然言語で表現された質問文データを入力し、解答の検索対象である文書データ群から質問文データに対する解答を出力する質問応答処理方法において、

前記文書検索手段が、入力された質問文データからキーワードを抽出し、前記文書デー 50

タ群から前記キーワードを含む文書データを検索抽出する処理過程と、

前記解答候補抽出手段が、前記文書データから解答となりうる言語表現を解答候補として抽出する処理過程と、

前記解答タイプ推定手段が、解答候補の言語表現についての表記的または意味的な類型を示す2以上の解答タイプを記憶しておき、前記質問文データの言語表現を解析し、確率にもとづいた機械学習法によるアルゴリズムを用いて、前記質問文データに対する解答が前記解答タイプであることの確信度を推定する処理過程と、

前記解答タイプ判定手段が、前記解答候補各々が前記2以上の解答タイプのどの解答タイプであるかを判定する処理過程と、

前記解答表出力手段が、前記判定された解答タイプをもとに前記解答候補を分類し、前記解答タイプごとに分類された解答候補の全部または一部を、前記2以上の解答タイプを見出し項目として構成し、かつ前記解答タイプ推定手段によって算出された確信度の高い順に対応する解答タイプの見出し項目を配置した表形式データとして出力する処理過程とを備える

ことを特徴とする質問応答処理方法。

【請求項7】

前記コンピュータは、解答タイプ判定手段を備え、

前記解答タイプ判定手段は、解答候補の言語表現と前記2以上の解答タイプとの対応関係を示す学習データが与えられた機械学習法によるアルゴリズムを用いて、前記解答候補各々が前記2以上の解答タイプのどの解答タイプになりやすいかを判定する

ことを特徴とする請求項5または請求項6のいずれか一項に記載の質問応答処理方法。

【請求項8】

前記解答表出力手段が行う処理過程では、

前記解答タイプごとに前記解答タイプ推定手段によって推定された確信度を表示した前記表形式データを作成する

ことを特徴とする請求項6に記載の質問応答処理方法。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、コンピュータによる自然言語処理システムとして、自然言語で表現された質問文に対する解答を出力する質問応答システムに関する。

【0002】

質問応答システムとは、自然言語による質問文を入力すると、その解答そのものを出力するシステムである。例えば、「パーキンソン病の兆候は脳のどの部分にある細胞の死が関係していますか。」という質問を入力すると、Web、新聞記事、事典などのデータを含む大量の電子化テキストから「パーキンソン病は、中脳の黒質にあるメラニン細胞が変性し、黒質細胞内で作られる神経伝達物質のドーパミンがなくなり発病する、とされている。」といった文を探し出し、「黒質」と的確に解答を出力する。

【0003】

質問応答システムは、論理式やデータベースからではなく、自然言語で記述された普通の文(テキストデータ)から解答を取り出すことができるため、大量の既存の文書データを利用することができる。また、質問応答システムは、キーワードで検索された記事から使用者自らが解答を探す必要がある情報検索システムなどと異なり、解答自体を出力する。そのため、使用者は、より早く解答の情報を得ることができる。このように質問応答システムは有用であるため、より使いやすい実用的なシステムの実現が期待されている。

【背景技術】

【0004】

一般的な質問応答システムは、おおまかに、解答表現推定処理、文書検索処理、解答抽出処理という3つの処理手段で構成されている(非特許文献1および非特許文献2参照)

。

10

20

30

40

50

【0005】

解答表現推定処理は、入力した質問文中の疑問代名詞の表現などにもとづいて解答表現を推定する処理である。解答表現とは、所望される解答の言語表現の類型であって、解答となる言語表現の意味にもとづいた類型（解答タイプ）、解答となる言語表現の表記にもとづいた類型（解答表現タイプ）などがある。質問応答システムは、どのような質問文の言語表現がどのような解答表現を要求しているかという対応関係を参照して、入力した質問文の解答の解答タイプを推定する。質問応答システムは、例えば、入力した質問文が「日本の面積はどのくらいですか」である場合には、所定の対応関係を参照して、質問文中の「どのくらい」という表現から解答タイプは「数値表現」であると推定する。また、質問文が「日本の首相はだれですか」という場合には、質問文中の「だれ」という表現から

10

【0006】

文書検索処理は、質問文からキーワードを取り出し、このキーワードを用いて解答を検索する対象となっている文書データ群を検索し、解答が記述されていると考えられる文書データを抽出する処理である。質問応答システムは、例えば、入力の質問文が「日本の首都はどこですか」である場合に、質問文から「日本」および「首都」をキーワードとして抽出し、検索対象の文書データ群から、キーワード「日本」および「首都」を含む文書データを検索する。

【0007】

解答抽出処理は、文書検索処理で抽出されたキーワードを含む文書データから、推定した解答タイプに適合する言語表現を抽出し、解答として出力する処理である。質問応答システムは、例えば、文書検索処理において検索されたキーワード「日本」および「首都」を含む文書データから、解答表現推定処理において推定した解答タイプ「固有名詞（地名）」に適合する言語表現「東京」を抽出して解答とする。

20

【0008】

前記のような処理を行うことにより、質問応答システムは、質問文「日本の首都はどこですか」に対して解答「東京」を出力する。

【非特許文献1】前田英作、「パターン認識/統計的学習からみた質問応答」、言語理解とコミュニケーション研究会 講習会資料、電子情報通信学会 言語理解とコミュニケーション研究会（NLIC）、2003年1月27日、p.29-64

30

【非特許文献2】Masaki Murata, Masao Utiyama, and Hitoshi Isahara, "A Question-Answering System Using Unit Estimation and Probabilistic Near-Terms IR", National Institute of Informatics NTCIR Workshop 3 Meeting QAC1, 2002年10月8日

【発明の開示】

【発明が解決しようとする課題】

【0009】

上記のように、従来の質問応答システムでは、検索された文書データから解答となりうる言語表現を解答候補として抽出し、抽出した解答候補それぞれの解答タイプを判定する。そして、質問文から推定した解答タイプと同じか類似する解答タイプと判定した解答候補の評価を高くし、原則的には、解答タイプが同じ解答候補であって所定の評価を得たものを解答として出力する。

40

【0010】

しかし、解答表現推定処理で推定された解答タイプが常に正しいものであるとは限らない。そのため、解答タイプが誤って推定された場合には、解答抽出処理において解答候補を評価する際の基準に誤りを含むことになり、結果的に解答抽出処理の精度を低下させてしまう。

【0011】

50

また、質問応答システムの使用者にとって、質問応答システムが出力した解答タイプが正しくない場合に、別の解答タイプと判定された解答候補を参照することができるような形態で解答が出力されると便利である。特に実際の使用を考慮すると、複数の解答タイプの解答候補が出力される質問応答システムは使用者にとって使いやすい。

【0012】

本発明の目的は、使用者が質問応答システムが出力した解答を解答タイプごとに見比べることができるように、解答タイプごとに分類した解答を表形式で出力することができる質問応答処理システムおよびその処理方法を提供することである。

【課題を解決するための手段】

【0013】

上記の目的を達成するため、本発明は、自然言語で表現された質問文データを入力し、解答の検索対象である文書データ群から質問文データに対する解答を出力する質問応答システムにおいて、所定の解答タイプをもとに分類した解答を解答タイプを見出し項目とした表形式のデータで出力することを特徴とするものである。

【0014】

本発明は、自然言語で表現された質問文データを入力し、解答の検索対象である文書データ群から質問文データに対する解答を出力する質問応答システムにおいて以下の処理を行う。入力された質問文データからキーワードを抽出し、前記文書データ群から前記キーワードを含む文書データを検索抽出する。さらに、前記文書データから解答となりうる言語表現を解答候補として抽出する。そして、解答候補の言語表現についての表記的または意味的な類型を示す2以上の解答タイプを記憶しておき、前記質問文データに対する解答の解答タイプとしてユーザによって選択入力された解答タイプを取得する。解答候補各々が、前記2以上の解答タイプのどの解答タイプであるかを判定し、前記判定された解答タイプをもとに前記解答候補を分類し、前記解答タイプごとに分類された解答候補の全部または一部を、前記2以上の解答タイプを見出し項目として構成し、かつ前記入力された解答タイプの見出しを先頭項目として配置した表形式データとして出力することを特徴とする。

【0015】

これにより、解答表データではユーザが入力した解答タイプの項目が先頭に配置されることになるため、使用者は、解答表の先頭の解答タイプの項目をみて解答を知ることができるとともに、他の解答タイプの解答をも参照することができる。

【0016】

また、本発明は、入力された質問文データからキーワードを抽出し、前記文書データ群から前記キーワードを含む文書データを検索抽出する。前記文書データから解答となりうる言語表現を解答候補として抽出する。そして、解答候補の言語表現についての表記的または意味的な類型を示す2以上の解答タイプを記憶しておき、前記質問文データの言語表現を解析し、確率にもとづいた機械学習法であって順位付けが可能な数値を算出できるアルゴリズムを用いて、前記質問文データに対する解答が前記解答タイプであることの確信度を推定する。前記解答候補各々が、前記2以上の解答タイプのどの解答タイプであるかを判定する。前記判定された解答タイプをもとに前記解答候補を分類し、前記解答タイプごとに分類された解答候補の全部または一部を、前記2以上の解答タイプを見出し項目として構成し、かつ前記解答タイプ推定手段で算出された確信度の高い順に対応する解答タイプの見出し項目を配置した表形式データとして出力することを特徴とする。

【0017】

これにより、解答表データでは最もなりやすいと推定された解答タイプの項目が先頭に配置されることになるため、使用者は、解答表の先頭の解答タイプの項目をみて解答を知ることができるとともに、他の解答タイプの解答をも参照することができる。

【0018】

さらに、本発明は、前記解答タイプを判定する場合に、解答候補の言語表現と前記2以上の解答タイプとの対応関係を示す学習データが与えられた機械学習法によるアルゴリズ

10

20

30

40

50

ムを用いて、前記解答候補各々が前記2以上の解答タイプのどの解答タイプになりやすいかを判定する。

【0019】

さらに、本発明は、前記解答表データを出力する場合に、前記解答タイプ推定手段によって推定された確信度が高い順に前記2以上の解答タイプを配置した前記表形式データを作成する。

【0020】

もしくは、本発明は、前記解答表データを出力する場合に、前記解答タイプごとに前記解答タイプ推定手段によって推定された確信度を表示した前記表形式データを作成する。

【発明の効果】

【0021】

本発明は、文書検索処理で検索された文書データから抽出した解答候補の解答タイプを所定の規則などをもとに判定し、解答候補を解答タイプごとに分類し、所定の順番で配置した解答タイプごとに解答候補の一覧を表形式で示す解答表を出力する。

【0022】

これにより、解答タイプの推定処理を行わないような質問応答システムにおいても、使用者は、質問文に対する解答を解答タイプごとに把握することができ、正しい解答を容易に得ることができる。

【0023】

また、質問応答システムにある事項に関して複数の質問文を与える必要がある場合に、一つの質問文を質問応答システムに与えるだけで複数の解答タイプの解答が出力されるため、使用者は、質問文に対応する解答タイプを見て、その解答タイプごとに解答を得ることができ、複数の質問文を与えることによる作業労力および処理負担を軽減することができる。

【0024】

また、本発明は、質問文に対する解答の解答タイプを推定する質問応答システムにおいて、所定の解答タイプについて、解答候補がその解答タイプである確信度を算出し、解答候補を解答タイプごとに分類し、解答タイプの確信度が高い順に解答タイプを配置して、解答タイプごとの解答候補の一覧を表形式で示す解答表を出力する。

【0025】

これにより、質問応答システムでは解答として確からしい解答タイプの順で解答を見やすい状態で出力することができる。よって、使用者は、最も確信度が高い解答タイプの解答を直ちに得ることができ、さらに他の解答タイプの解答を容易に参照することができる。

【0026】

また、本発明は、ユーザが指定した解答タイプを入力する質問応答システムにおいて、解答候補を解答タイプごとに分類し、入力された解答タイプを先頭項目として配置して、さらに所定の順番で配置した解答タイプごとに解答候補の一覧を表形式で示す解答表を出力する。

【0027】

これにより、質問応答システムでは入力された解答タイプを先頭項目とし解答を見やすい状態で出力することができる。よって、使用者は、指定した解答タイプの解答を簡単に得ることができ、さらに他の解答タイプの解答を容易に参照することができる。

【発明を実施するための最良の形態】

【0028】

以下に、本発明を実施するための形態を図を用いて説明する。

〔第1の実施例〕

第1の実施例として、本発明を、解答の解答タイプを推定しない質問応答システムに適用する場合を説明する。

【0029】

10

20

30

40

50

図 1 に、第 1 の実施例における本発明の構成例を示す。質問応答システム 1 は、質問文入力部 1 1、文書検索部 1 3、解答候補抽出部 1 4、解答タイプ判定部 1 5、解答表出力部 1 6、および文書データベース 2 0 を備える。

【 0 0 3 0 】

質問文入力部 1 1 は、自然言語で表現された質問文データ（質問文）を入力する手段である。

【 0 0 3 1 】

文書検索部 1 3 は、質問文入力部 1 1 によって入力された質問文から抽出したキーワードを用いて、解答探索の対象である文書データベース 2 0 から、キーワードを含む文書データを検索・抽出する手段である。文書検索部 1 3 は、既知の一般的な文書検索手法により検索処理を行う。なお、文書データベース 2 0 として、新聞記事、百科事典、英和辞書、Web ページなどの文書データを利用する。

10

【 0 0 3 2 】

解答候補抽出部 1 4 は、文書検索部 1 3 によって検索された文書データから解答となりうる言語表現を抽出して解答候補とし、解答候補に評価点を付与する手段である。例えば、解答候補抽出部 1 4 は、文書検索部 1 3 によって検索された文書データから解答となりうる言語表現（解答候補）を抽出し、抽出元の文書データ内における解答候補とキーワードとの近接性を確率的に評価し、近接性にもとづく評価点を解答候補に付与する。

【 0 0 3 3 】

解答タイプ判定部 1 5 は、固有表現抽出処理により解答候補の固有表現を特定し、所定の解答タイプ判定規則を参照して解答候補の解答タイプを判定する手段である。

20

【 0 0 3 4 】

固有表現抽出処理は、人名、地名、組織名、人工名（小説の題名、賞の名称など）などの固有名詞や、時間、距離、金額などの数値表現のように特定の事物・数量を意味する言語表現を特定する処理である。解答タイプ判定規則は、固有表現抽出処理により抽出された言語表現（解答候補）に対応する解答タイプが何であるかを判定するヒューリスティックな規則である。

【 0 0 3 5 】

解答表出力部 1 6 は、解答候補抽出部 1 4 によって抽出された解答候補を解答タイプをもとに分類し、解答タイプごとの解答候補の中から所定の評価の解答候補を解答として抽出し、抽出した解答を解答タイプごとに項目分けして表形式で示す表データ（解答表）を作成して出力する手段である。

30

【 0 0 3 6 】

図 2 に、第 1 の実施例における本発明の処理の流れを示す。

【 0 0 3 7 】

質問応答システム 1 の質問文入力部 1 1 は、質問文を入力する（ステップ S 1 0）。そして、文書検索部 1 3 は、質問文からキーワードを抽出し（ステップ S 1 1）、抽出したキーワードを用いて文書データベース 2 0 を検索し、キーワードを含む文書データを抽出する（ステップ S 1 2）。具体的には、文書検索部 1 3 は、入力された質問文「日本の首都はどこですか」である場合に、質問文を形態素解析して質問文から「日本、首都」という名詞を切り出してキーワードとする。そして、キーワード「日本、首都」を用いて文書データベース 2 0 を検索し、キーワード「日本、首都」を含む文書データを抽出する。検索の結果、以下のような文書データが抽出され、質問文に対する解答を抽出する対象となる。

40

「1999年に初めて日本の首都東京で国際会議 A が B 学会のもとで開催される。約 800 人の参加が見込まれている。前会長の C 氏は現会長の D 氏の苦勞をねぎらった。」

次に、解答候補抽出部 1 4 は、抽出された文書データから解答となりうる言語表現（解答候補）を抽出する（ステップ S 1 3）。解答候補抽出部 1 4 は、抽出された文書データから n - g r a m の文字列を切り出して生成した名詞、名詞句などの言語表現を解答候補として抽出する。

50

「1999年、東京、国際会議A、B学会、約800人、参加、前会長、C氏、現会長、D氏、苦労」

さらに、解答候補抽出部14は、各解答候補に評価点を付与する(ステップS14)。解答候補抽出部14は、抽出された文書データにおいて抽出した解答候補とキーワードとの出現箇所の近接性を判定し、解答候補がキーワードと近接して出現するほど良い評価となるような所定の式を用いて評価点を算出する。ここでは、文書データにおいて解答候補とキーワードとが狭い範囲内に出現するほど、解答候補とキーワードとは相互に関連性が高く、キーワードと関連性が高い解答候補が質問文に対する解答としてより良いものであることを前提としている。

【0038】

10

解答タイプ判定部15は、解答タイプ判定規則を参照して、解答候補の解答タイプを判定する(ステップS15)。解答タイプ判定部15は、固有表現抽出処理により、人名、地名、数値表現などの名詞、名詞句の固有表現を特定し、さらに特定した固有表現をもとに、以下のような解答タイプ判定規則を参照して解答候補の解答タイプを判定する。

【0039】

- (1) 解答候補の固有表現が「人名」であれば、解答タイプは「人名」である、
- (2) 解答候補の固有表現が「地名」であれば、解答タイプは「地名」である、
- (3) 解答候補の固有表現が「人工的に命名されたもの」であれば、解答タイプは「人工名」である、
- (4) 解答候補の固有表現が「時間を示すもの」であれば、解答タイプは「時間」である、
- (5) 解答候補の固有表現が「数値を示すもの」であれば、解答タイプは「数値表現」である、
- (6) 解答候補の固有表現が上記の(1)～(5)のいずれにも該当しないものであれば、解答タイプは「その他」である。

20

【0040】

例えば、解答候補「1999年」の固有表現が「時間」であると特定された場合に、解答タイプ判定規則(4)により、解答タイプは「時間、数値表現」であると判定される。また、解答候補「東京」の固有表現が「地名」であると特定された場合に、解答タイプ判定規則(2)により、解答タイプは「地名」であると判定される。

30

【0041】

なお、解答タイプ判定部15は、固有表現抽出処理として、名詞句以外の品詞句(動詞句、形容詞句など)を抽出することも可能である。

【0042】

次に、解答表出力部16は、解答候補を解答タイプごとに分類し、所定の値以上の評価点が付与された解答候補を解答とし、解答タイプごとに解答を示す解答表を作成して出力する(ステップS16)。解答表出力部16は、見出し項目として解答タイプを所定の順序で配置し、解答タイプの項目ごとに解答を評価のよい順に並べた一覧を示す解答表を作成する。

【0043】

40

以下に示すように、解答候補は解答タイプごとに分類され、所定の評価点を得て選択された解答は、各解答タイプにおいて評価点の高い順に並び替えられる。

【0044】

人名：C氏、D氏、
 地名：東京、
 組織名：B学会、
 時間：1999年、
 人工名：国際会議A、
 数値表現：1999年、約800人
 その他：参加、前会長、現会長、苦労

50

図3に、出力される解答表の例を示す。図3に示す解答表において、解答タイプの項目は所定の順序で配置され、解答は各解答タイプ内で評価点の高い順に先頭から配置される。使用者は、解答タイプが「地名」であることを知っているから、図3の解答表の解答タイプ「地名」の項目を見て、解答が「東京」であることがすぐにわかる。

【0045】

本実施例に示すように、本発明によれば、質問文から解答タイプを推定する処理を行わないような質問応答システムにおいても、解答を解答タイプごとに示す表形式で出力できる。これにより、使用者は、解答表から該当する解答タイプの項目を参照して容易に正しい解答を得ることができる。

【0046】

また、使用者が、関連する事項に関する複数の解答タイプの解答を得たいと考える場合に、一つの質問文を質問応答システムに与えるだけで複数の解答タイプの解答を一時に取得することができる。例えば、使用者が、以下のような質問文を続けて入力して解答を得たいとする。

【0047】

質問文Q1：「国際会議Aは、どこで開催されましたか。」

質問文Q2：「国際会議Aは、いつ開催されましたか。」

質問文Q3：「国際会議Aは、どの学会によって開催されましたか。」

本発明によれば、質問応答システム1は、質問文Q1を入力すると、上記の処理を行って、その質問文Q1に対する解答と共に同時に取得した他の解答タイプの解答も併せて取得し、図3に示すような解答表を出力する。使用者は質問文Q1～Q3の解答タイプを知っているから、図3の解答表を見て、質問文Q1の解答「東京」、質問文Q2の解答「1999年」、質問文Q3の解答「B学会」のように、複数の質問文それぞれに対応する解答を知ることができる。

〔第2の実施例〕

第2の実施例として、本発明を、解答の解答タイプを推定する質問応答システムに適用する場合を説明する。

【0048】

図4に、第2の実施例における本発明の構成例を示す。質問応答システム2は、質問文入力部21、解答タイプ推定部22、文書検索部23、解答候補抽出部24、解答タイプ判定部25、解答表出力部26、および文書データベース20を備える。

【0049】

質問文入力部21、文書検索部23、解答候補抽出部24、解答タイプ判定部25および解答表出力部26は、それぞれ、質問応答システム1の質問文入力部11、文書検索部13、解答候補抽出部14、解答タイプ判定部15、解答表出力部16と同じ処理を行う処理手段である。

【0050】

解答タイプ推定部22は、入力された質問文から、確率にもとづいた機械学習法であって順位付けが可能な数値を算出できるものを用いて、所定の解答タイプについて、解答がその解答タイプであることの確かさ（確信度）を推定する手段である。

【0051】

解答タイプ推定部22は、確率にもとづいた機械学習法として、最大エントロピー法などの手法を用いる。最大エントロピー法は、学習データにおける推定に用いられる情報の細かい単位である素性の出現の期待値と、未知データにおける素性の出現の期待値とが等しいという条件で確率分布のエントロピーが最大となる場合の確率分布を求め、求めた確率分布にもとづき素性の各出現パターンに対して各分類になる確率を求め、最大の確率を持つ分類先を求める分類とする処理手法である。

【0052】

最大エントロピー法によれば、所定の解答タイプの確からしさを確率値で算出することができるので、算出した確率値をもとに解答タイプの表示の順序を決定することができる

10

20

30

40

50

【 0 0 5 3 】

図 5 に、第 2 の実施例における本発明の処理の流れを示す。

【 0 0 5 4 】

質問応答システム 2 の質問文入力部 2 1 は、質問文を入力し（ステップ S 2 0 ）、解答タイプ推定部 2 2 は、機械学習法を用いた推定処理により、質問文の表現から解答タイプの確信度を推定する（ステップ S 2 1 ）。解答タイプ推定部 2 2 は、入力された質問文を形態素解析し、解析した疑問代名詞などの表現を手がかりに、最大エントロピー法などの機械学習法を用いて質問文に対する解答の解答タイプを推定する。例えば、入力した質問文が「日本の首都はどこですか」である場合に、質問文の「どこ」という表現などを手がかりに解答タイプは「地名」とであると推定する。

10

【 0 0 5 5 】

そして、文書検索部 2 3 は、質問文からキーワードを抽出し（ステップ S 2 2 ）、抽出したキーワードを用いて文書データベース 2 0 を検索し、キーワードを含む文書データを抽出する（ステップ S 2 3 ）。解答候補抽出部 2 4 は、固有表現抽出処理により、抽出された文書データから解答となりうる言語表現（解答候補）を抽出する（ステップ S 2 4 ）。さらに、解答候補抽出部 2 4 は、抽出された文書データにおける抽出した解答候補とキーワードとの出現箇所の近接性を判定し、その解答候補に評価点を付与する（ステップ S 2 5 ）。そして、解答タイプ判定部 2 5 は、所定の解答タイプ判定規則を参照して解答候補の解答タイプを判定する（ステップ S 2 6 ）。

20

【 0 0 5 6 】

その後、解答表出力部 2 6 は、解答候補を解答タイプごとに分類し、所定の評価以上の解答候補を解答として、解答タイプごとに解答を示す解答表を作成して出力する（ステップ S 2 7 ）。解答表出力部 2 6 は、見出し項目として解答タイプを確信度の高い順に配置し、解答タイプの項目ごとに解答を評価のよい順に並べた一覧を示す解答表を作成する。

【 0 0 5 7 】

図 6 および図 7 に、出力される解答表の例を示す。図 6 に示す解答表において、解答タイプの項目は、ステップ S 2 1 の処理において推定された確信度の高い順に先頭（左）から「地名、組織名、その他、人工名、...」のように配置される。また解答タイプごとに分類された解答は、それぞれの解答タイプ内で評価点の高い順に先頭から配置される。

30

【 0 0 5 8 】

また、図 7 に示すように、解答タイプの項目が、図 6 と同様に推定された確信度の高い順に先頭（上）から「地名、組織名、その他、人工名、...」のように配置される。

【 0 0 5 9 】

また、解答表出力部 2 6 は、図 6 および図 7 の解答表の解答タイプの項目内に、解答タイプ推定部 2 2 で算出した確信度を「X%」のように表示してもよい。

【 0 0 6 0 】

本実施例において、使用者は、質問応答システムが出力した解答表で解答タイプの確からしさが高い順に並べられた項目をみて正しい解答を見つけることができる。さらに、質問応答システムが解答タイプの推定を誤った場合でも、解答表には全ての解答タイプの解答が表示されているため、使用者は、解答表から正しい解答を選択することができる。

40

〔 第 3 の実施例 〕

第 3 の実施例として、本発明を、解答の解答タイプを入力する質問応答システムに適用する場合を説明する。

【 0 0 6 1 】

図 8 に、第 3 の実施例における本発明の構成例を示す。質問応答システム 3 は、質問文入力部 3 1、解答タイプ入力部 3 2、文書検索部 3 3、解答候補抽出部 3 4、解答タイプ判定部 3 5、解答表出力部 3 6、および文書データベース 2 0 を備える。

【 0 0 6 2 】

質問文入力部 3 1、文書検索部 3 3、解答候補抽出部 3 4、解答タイプ判定部 3 5、お

50

よび解答表出力部 36 は、それぞれ、質問応答システム 1 の質問文入力部 11、文書検索部 13、解答候補抽出部 14、解答タイプ判定部 15、解答表出力部 16 と同様の処理を行う処理手段である。

【0063】

解答タイプ入力部 32 は、使用者が選択または入力の指示をした解答タイプを入力する手段である。

【0064】

図 9 に、第 3 の実施例における本発明の処理の流れを示す。

【0065】

質問応答システム 3 の質問文入力部 31 は、質問文を入力し（ステップ S30）、解答タイプ入力部 32 は、解答タイプを入力する（ステップ S31）。ここで、入力された解答タイプが「地名」とする。

10

【0066】

そして、文書検索部 33 は、質問文からキーワードを抽出し（ステップ S32）、抽出したキーワードを用いて文書データベース 20 を検索し、キーワードを含む文書データを抽出する（ステップ S33）。解答候補抽出部 34 は、固有表現抽出処理により、抽出された文書データから解答となりうる言語表現（解答候補）を抽出する（ステップ S34）。さらに、解答候補抽出部 34 は、抽出された文書データにおける抽出した解答候補とキーワードとの出現箇所の近接性を判定し、その解答候補に評価点を付与する（ステップ S35）。また、解答タイプ判定部 35 は、所定の解答タイプ判定規則を参照して、解答候補の解答タイプを判定する（ステップ S36）。

20

【0067】

次に、解答表出力部 36 は、解答候補を解答タイプごとに分類し、所定の評価以上の解答候補を解答として、解答タイプごとに解答を示す解答表を作成して出力する（ステップ S37）。解答表出力部 36 は、見出し項目として入力された解答タイプを先頭項目として配置し、続いて入力された解答タイプ以外の解答タイプを所定の順に配置し、解答タイプの項目ごとに解答を評価のよい順に並べた一覧を示す解答表を作成する。

【0068】

図 10 に、出力される解答表の例を示す。図 10 に示す解答表において、入力された解答タイプ「地名」が先頭（最左側）に配置され、続いて入力された解答タイプ以外の解答タイプが、所定の順序で配置される。また解答タイプごとに分類された解答は、それぞれの解答タイプ内で評価点の高い順に先頭から配置される。

30

【0069】

これにより、使用者は、質問応答システムが出力した解答表で入力した解答タイプの解答を確実に見つけることができるうえ、他の解答タイプの解答を容易に参照することができる。また、質問応答システム 3 では解答タイプの推定処理を行わないため、解答タイプの推定処理を行う質問応答システムに比べて高い精度の処理を行うことができる。

【0070】

上記の第 1 の実施例～第 3 の実施例では、解答となりうる言語表現の類型として、地名、人名、人工名などの言語表現の意味にもとづく類型（解答タイプ）を用いたが、解答タイプの代わりに解答表現タイプを用いてもよい。解答表現タイプは、解答となる言語表現の表記にもとづく類型である。「ひらがなのみからなる表現、カタカナのみからなる表現、漢字のみからなる表現、英文字のみからなる表現、英語記号・数字のみからなる表現、漢字カタカナのみからなる表現、数字表現を含む表現」などの解答表現タイプを予め定義しておく。

40

【0071】

この場合、解答候補抽出部 14、24、34 は、検索した文書データ内の文字列の字種（ひらがな、カタカナ、漢字、英字など）を利用して解答候補を抽出する。そして、解答タイプ判定部 15、25、35 は、解答候補の字種から解答表現タイプを判定する。

【0072】

50

図 1 1 に、出力される解答表の例を示す。図 1 1 に示す解答表では、所定の順として解答表現タイプ「漢字のみ、数値表現を含む、…」が配置される。また解答タイプごとに分類された解答は、それぞれの解答タイプ内で評価点の高い順に先頭から配置される。なお、解答表現タイプの確信度が推定される場合には、推定された確信度の順に解答表現タイプが配置される。

【 0 0 7 3 】

また、第 1 の実施例～第 3 の実施例では、解答表出力部 1 6、2 6、3 6 は、解答候補が存在しない解答タイプの項目を省略した解答表を作成するようにしてもよい。

【 0 0 7 4 】

特に第 2 の実施例では、解答表出力部 2 6 は、解答タイプ推定部 2 2 で算出された解答タイプの確信度が所定の評価以上のもののみを項目として示す解答表、または解答タイプの確信度が高い順に所定の数以内のもののみを項目として示す解答表を作成するようにしてもよい。

10

【 0 0 7 5 】

以上、本発明を実施例にもとづいて説明したが、本発明はその主旨の範囲において種々の変形が可能であることは当然である。

【 0 0 7 6 】

例えば、第 1 の実施例～第 3 の実施例において、質問応答システム 1、2、3 は、所定のヒューリスティックな解答タイプ判定規則を参照して解答タイプを判定する解答タイプ判定部 1 5、2 5、3 5 で構成されるものとして説明した。

20

【 0 0 7 7 】

しかし、質問応答システム 1、2、3 は、ヒューリスティックな規則を用いて処理を行う代わりに、最大エントロピー法、サポートベクトルマシン法などの教師あり機械学習法を利用して解答タイプを推定もしくは判定する解答タイプ判定部 1 5'、2 5'、3 5' で構成されるものであってもよい。

【 0 0 7 8 】

この場合に、解答タイプ判定部 1 5'、2 5'、3 5' は、問題ごとに正しい入力（言語表現）と出力（判定する解答タイプ）の対のパターンをあらかじめ学習データとして人手で作成して、どのような言語表現の場合にどのような解答タイプとなりやすいかを学習しておく。そして、抽出した言語表現（解答候補）が、どういう解答タイプになりやすいかを判定する。

30

【 0 0 7 9 】

なお、サポートベクトルマシン法は、空間を超平面で分割することにより 2 つの分類からなるデータを分類する手法であって、学習データにおける 2 つの分類の事例群と超平面との間隔（マージン）が大きいものほど未知データにおいて誤った分類をする可能性が低いとの考え方を前提にして、このマージンを最大にする超平面を求め、それを用いて分類を行なう。3 つ以上の分類からなるデータを分類する場合には、複数のサポートベクトルマシンを組み合わせる。

【 0 0 8 0 】

また、質問応答システム 2 において、解答タイプ推定部 2 2 は、質問文とその解答の解答タイプとの対応関係を定義したヒューリスティックな解答タイプ推定規則を用いた処理を行う処理手段であってよい。この場合、解答タイプ推定規則に、質問文と解答の解答タイプとの対応関係と共に、“if then”規則などを用いて、どのような質問文に対してどの解答タイプであるかの確信度を定義しておく。

40

【 0 0 8 1 】

また、本発明は、コンピュータにより読み取られ実行される処理プログラムとして実施することができる。本発明を実現する処理プログラムは、コンピュータが読み取り可能な、可搬媒体メモリ、半導体メモリ、ハードディスクなどの適当な記録媒体に格納することができ、これらの記録媒体に記録して提供され、または、通信インタフェースを介して種々の通信網を利用した送受信により提供されるものである。

50

【図面の簡単な説明】

【0082】

- 【図1】第1の実施例における本発明の構成例を示す図である。
- 【図2】第1の実施例における本発明の処理の流れを示す図である。
- 【図3】出力される解答表の例を示す図である。
- 【図4】第2の実施例における本発明の構成例を示す図である。
- 【図5】第2の実施例における本発明の処理の流れを示す図である。
- 【図6】出力される解答表の例を示す図である。
- 【図7】出力される解答表の例を示す図である。
- 【図8】第3の実施例における本発明の構成例を示す図である。
- 【図9】第3の実施例における本発明の処理の流れを示す図である。
- 【図10】出力される解答表の例を示す図である。
- 【図11】出力される解答表の例を示す図である。

10

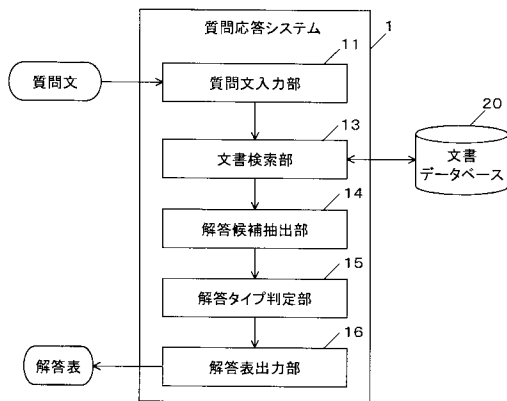
【符号の説明】

【0083】

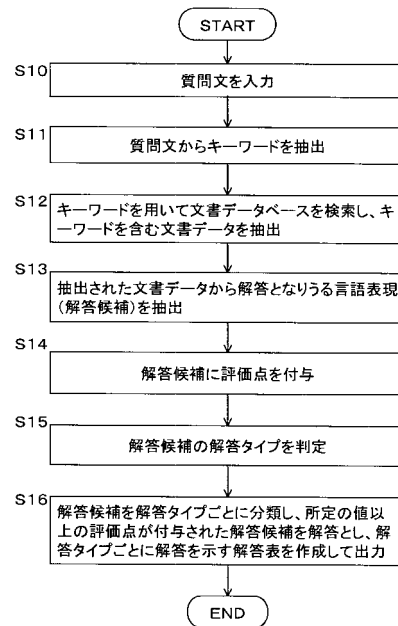
- 1、2、3 質問応答システム
- 11、21、31 質問文入力部
- 13、23、33 文書検索部
- 14、24、34 解答候補抽出部
- 15、25、35 解答タイプ判定部
- 16、26、36 解答表出力部
- 20 文書データベース
- 22 解答タイプ推定部
- 32 解答タイプ入力部

20

【図1】



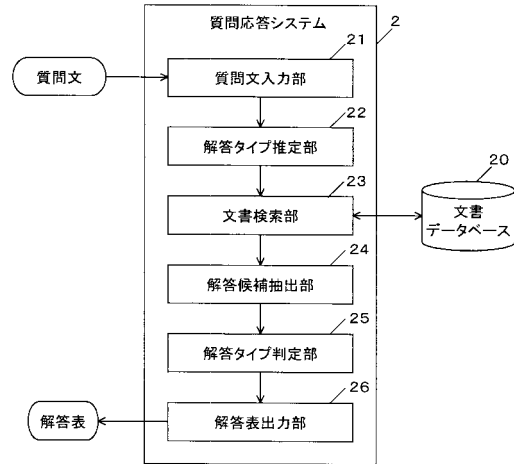
【図2】



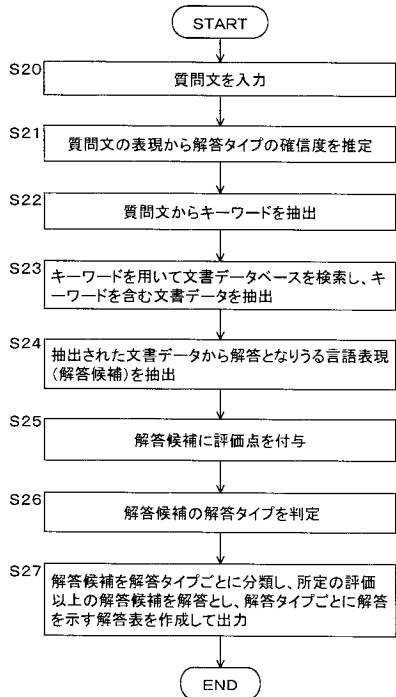
【 図 3 】

人名	C氏 D氏	地名	東京	組織名	B学会	時間	1999年	人工名	国際会議A	数値表現	1999年 約800人	その他	参加 前会長 現会長 苦勞
----	----------	----	----	-----	-----	----	-------	-----	-------	------	----------------	-----	------------------------

【 図 4 】



【 図 5 】



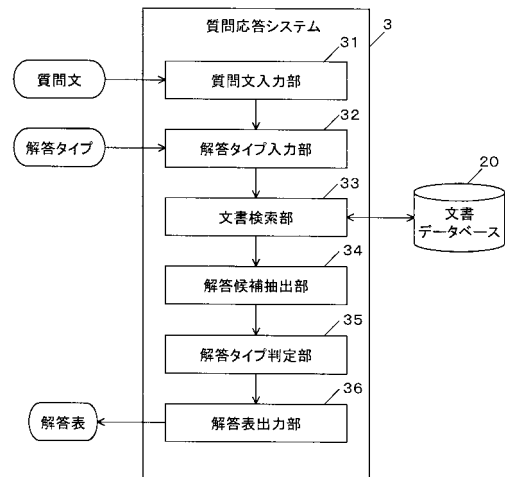
【 図 6 】

地名	東京	組織名	B学会	その他	参加 前会長 現会長 苦勞	人工名	国際会議A	人名	C氏 D氏	時間	1999年	数値表現	1999年 約800人
----	----	-----	-----	-----	------------------------	-----	-------	----	----------	----	-------	------	----------------

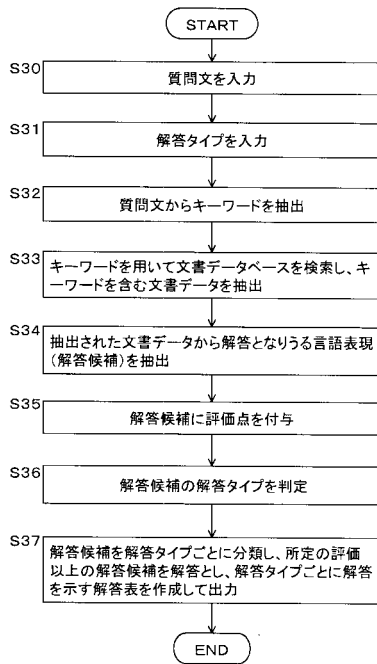
【 図 7 】

地名	東京
組織名	日学会
その他	参加, 前会長, 現会長, 苦勞
人工名	国際会議A
人名	C氏, D氏
時間	1999年
数値表現	1999年, 約800人

【 図 8 】



【 図 9 】



【 図 10 】

地名	組織名	その他	...
東京	日学会	参加 前会長 現会長 苦勞	...

【 図 1 1 】

漢字のみ	数値表現を含む	...
東京 B学会 国際会議A C氏 D氏 参加 前会長 現会長 苦勞	1999年 約800人	...

フロントページの続き

- (56)参考文献 特開2002-132811(JP,A)
特開平11-195041(JP,A)
特開2003-099454(JP,A)
前田 英作,磯崎 秀樹,佐々木 裕,賀沢 秀人,平尾 努,鈴木 潤,高度検索技術 質問
応答システム:SAIQA-何でも答える物知り博士,NTT R&D,日本,社団法人電気通
信協会,2003年 2月10日,第52巻 第2号,第122-133頁
鈴木 潤,佐々木 裕,前田 英作,単語属性N-gramと統計的機械学習による質問タイプ
同定,情報処理学会論文誌,日本,社団法人情報処理学会,2003年11月15日,第44巻
第11号,第2839-2853頁
佐々木 裕,磯崎 秀樹,平 博順,廣田 啓一,賀沢 秀人,平尾 努,加藤 恒昭,質問応
答システムの比較と評価,電子情報通信学会技術研究報告,日本,社団法人電子情報通信学会,
2000年10月20日,Vol.100 No.401,第17-24頁
- (58)調査した分野(Int.Cl.,DB名)
G06F 17/30
JSTPlus(JDream2)
NRIサイバーパテント