(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2014-112316
(P2014-112316A)
(43)公開日　平成26年6月19日(2014.6.19)

| (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 | 210D | |

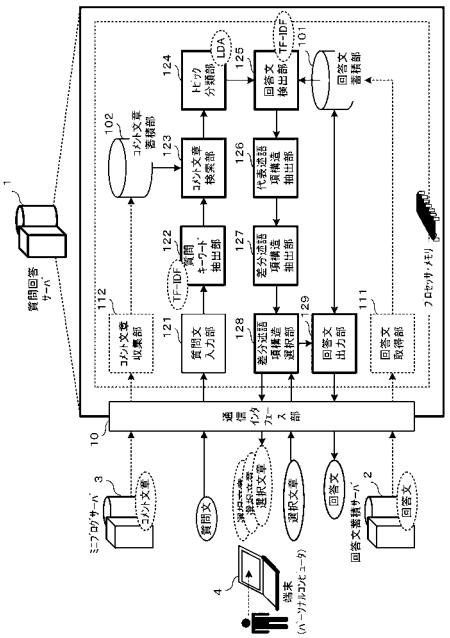
審査請求 未請求　請求項の数 10　ＯＬ　（全 17 頁）

(21)出願番号　特願2012-266589 (P2012-266589)
(22)出願日　平成24年12月5日 (2012.12.5)

(71)出願人　000208891
ＫＤＤＩ株式会社
東京都新宿区西新宿二丁目３番２号
(74)代理人　100135068
弁理士　早原　茂樹
(72)発明者　松本　一則
埼玉県ふじみ野市大原二丁目１番１５号
株式会社ＫＤＤＩ研究所内
(72)発明者　服部　元
埼玉県ふじみ野市大原二丁目１番１５号
株式会社ＫＤＤＩ研究所内
(72)発明者　小野　智弘
埼玉県ふじみ野市大原二丁目１番１５号
株式会社ＫＤＤＩ研究所内

(54)【発明の名称】大量のコメント文章を用いた質問回答プログラム、サーバ及び方法

(57)【要約】
【課題】ユーザの質問文に対して複数の回答文の候補が存在する場合、ユーザの意図を反映した回答文を明示する（に絞り込む）ことができる質問回答サーバ等を提供する。
【解決手段】質問文から複数の質問キーワードを抽出する質問キーワード抽出手段と、質問キーワードを含むコメント文章を検索するコメント文章検索手段と、検索された複数のコメント文章を、複数個のトピックグループに分類するトピック分類手段と、各トピックと各回答文との類似度が所定閾値以上となる回答文を対応付ける回答文検出手段と、各トピックについて代表述語項構造を抽出する代表述語項構造抽出手段と、当該トピックのみに出現する代表述語項構造を、差分述語項構造として抽出する差分述語項構造抽出手段と、複数の差分述語項構造をユーザに明示し且つ選択させる差分述語項構造選択手段と、その差分述語項構造に対応する回答文をユーザに明示する回答文出力手段とを有する。
【選択図】図２



【特許請求の範囲】
【請求項１】
多数のコメント文章を蓄積したコメント文章蓄積部と、多数の回答文を蓄積した回答文蓄積部とを有し、ユーザからの質問文に対する回答文を抽出するようにコンピュータを機能させる質問回答プログラムであって、
質問文を入力する質問文入力手段と、
前記質問文に含まれる複数の質問キーワードを抽出する質問キーワード抽出手段と、
前記コメント文章蓄積部を用いて、前記質問キーワードを含むコメント文章を検索するコメント文章検索手段と、
検索された複数のコメント文章を、述語項構造解析によって、述語項構造の分布から複数個のトピックグループに分類するトピック分類手段と、
各トピックグループに含まれるコメント文章群と、各回答文に含まれる文章との間の類似度を算出し、各トピックグループに前記類似度が所定閾値以上となる回答文を対応付ける回答文検出手段と、
各トピックグループについて、対応付けられた回答文に含まれる述語項構造の中で、当該トピックグループを特徴付ける代表述語項構造を抽出する代表述語項構造抽出手段と、
各トピックグループについて、当該トピックグループのみに出現する代表述語項構造を、差分述語項構造として抽出する差分述語項構造抽出手段と、
前記回答文検出手段によって検出された前記回答文を、対応する１つ以上の前記差分述語項構造に基づく文章と共に明示する回答文出力手段と
してコンピュータを機能させることを特徴とする質問回答プログラム。
【請求項２】
複数の前記差分述語項構造に基づく文章を、ユーザインタフェースを介してユーザに明示すると共に、ユーザ操作に応じていずれか１つの差分述語項構造を選択させる差分述語項構造選択手段を更に有し、
前記回答文出力手段は、選択された文章の差分述語項構造に対応する回答文を、ユーザインタフェースを介して明示する
ようにコンピュータを機能させることを特徴とする請求項１に記載の質問回答プログラム。
【請求項３】
前記トピック分類手段は、当該コメント文章を、分類された各トピックグループに属する確からしさ（トピック比率）を算出するＬＤＡ(Latent Dirichlet Allocation)アルゴリズムを用いて、いずれか１つのトピックグループに分類するようにコンピュータを機能させることを特徴とする請求項１又は２に記載の質問回答プログラム。
【請求項４】
前記回答文検出手段は、
各トピックグループに含まれるコメント文章群から、述語項構造解析によって述語項構造を抽出すると共に、前記トピックグループにおける第１の特徴ベクトルを算出し、
前記回答文蓄積部に蓄積された各回答文から、述語項構造解析によって述語項構造を抽出すると共に当該回答文における第２の特徴ベクトルとを算出し、
前記トピックグループの第１のベクトルと、前記回答文の第２のベクトルとの間のコサイン距離に基づいて類似度を算出する
ようにコンピュータを機能させることを特徴とする請求項１から３のいずれか１項に記載の質問回答プログラム。
【請求項５】
前記代表述語項構造抽出手段は、各トピックグループの代表述語項構造を、赤池情報量基準に応じて優先順に並べるようにコンピュータを機能させることを特徴とする請求項１から４のいずれか１項に記載の質問回答プログラム。
【請求項６】
前記コメント文章は、不特定多数の第三者によって投稿されたものであって、

前記コメント文章蓄積部は、ミニブログ(mini Web log)サーバに投稿されたコメント文章を収集し蓄積したものであるようにコンピュータを機能させることを特徴とする請求項１から５のいずれか１項に記載の質問回答プログラム。

【請求項７】

多数のコメント文章を蓄積したコメント文章蓄積部と、多数の回答文を蓄積した回答文蓄積部とを有し、ユーザからの質問文に対する回答文を抽出する質問回答サーバであって、

端末から、質問文を入力する質問文入力手段と、

前記質問文に含まれる複数の質問キーワードを抽出する質問キーワード抽出手段と、

前記コメント文章蓄積部を用いて、前記質問キーワードを含むコメント文章を検索するコメント文章検索手段と、

検索された複数のコメント文章を、述語項構造解析によって、述語項構造の分布から複数個のトピックグループに分類するトピック分類手段と、

各トピックグループに含まれるコメント文章群と、各回答文との間の類似度を算出し、各トピックグループに前記類似度が所定閾値以上となる回答文を対応付ける回答文検出手段と、

各トピックグループについて、対応付けられた回答文に含まれる述語項構造の中で、当該トピックグループを特徴付ける代表述語項構造を抽出する代表述語項構造抽出手段と、

各トピックグループについて、当該トピックグループのみに出現する代表述語項構造を、差分述語項構造として抽出する差分述語項構造抽出手段と、

前記回答文検出手段によって検出された前記回答文を、対応する１つ以上の前記差分述語項構造に基づく文章と共に明示する回答文出力手段と

を有することを特徴とする質問回答サーバ。

【請求項８】

複数の前記差分述語項構造に基づく文章を、ユーザインタフェースを介してユーザに明示すると共に、ユーザ操作に応じていずれか１つの差分述語項構造を選択させる差分述語項構造選択手段を更に有し、

前記回答文出力手段は、選択された文章の差分述語項構造に対応する回答文を、ユーザインタフェースを介して明示する

ことを特徴とする請求項７に記載の質問回答サーバ。

【請求項９】

多数のコメント文章を蓄積したコメント文章蓄積部と、多数の回答文を蓄積した回答文蓄積部とを有し、ユーザからの質問文に対する回答文を抽出する装置における質問回答方法であって、

質問文を入力する第１のステップと、

前記質問文に含まれる複数の質問キーワードを抽出する第２のステップと、

前記コメント文章蓄積部を用いて、前記質問キーワードを含むコメント文章を検索する第３のステップと、

検索された複数のコメント文章を、述語項構造解析によって、述語項構造の分布から複数個のトピックグループに分類する第４のステップと、

各トピックグループに含まれるコメント文章群と、各回答文との間の類似度を算出し、各トピックグループに前記類似度が所定閾値以上となる回答文を対応付ける第５のステップと、

各トピックグループについて、対応付けられた回答文に含まれる述語項構造の中で、当該トピックグループを特徴付ける代表述語項構造を抽出する第６のステップと、

各トピックグループについて、当該トピックグループのみに出現する代表述語項構造を、差分述語項構造として抽出する第７のステップと、

第５のステップによって検出された前記回答文を、対応する１つ以上の前記差分述語項構造に基づく文章と共に明示する第８のステップと

を有することを特徴とする質問回答方法。

【請求項１０】
第８のステップについて、
複数の前記差分述語項構造に基づく文章を、ユーザインタフェースを介してユーザに明示すると共に、ユーザ操作に応じていずれか１つの差分述語項構造を選択させ、
選択された文章の差分述語項構造に対応する回答文を、ユーザインタフェースを介して明示することを特徴とする請求項９に記載の質問回答方法。
【発明の詳細な説明】
【技術分野】
【０００１】
本発明は、質問文の入力に対して最適な回答文を出力する質問回答プログラムの技術に関する。
【背景技術】
【０００２】
近年、ＦＡＱ(Frequently Asked Questions)に基づく質問回答システムが構築されている。「ＦＡＱ」とは、多数の人が共通して頻繁に尋ねる質問に対する回答をまとめた問答集をいう。質問回答システムは、特定種類の情報に関する質問文をユーザから自然言語で入力し、その回答文を出力するソフトウェアをいう。一般に、質問回答システムは、仮想質問文とそれに紐づけられた回答候補文とを予めデータベースに記憶する。その上で、質問回答システムは、以下のようなステップで処理を実行する。
（１）ユーザから入力された質問文から、特徴的な単語をクエリとして抽出する。
（２）検索エンジンを用いて、複数のクエリの出現頻度が高い仮想質問文を選択する。
（３）選択された仮想選択文に対する回答文を選択する。
（４）選択された回答文をユーザに提示する。
【０００３】
このような質問回答システムは、ユーザに対して単体装置として存在するものもあれば、インターネット上に質問回答サーバとして接続されたものものある。この質問回答サーバは、ユーザ操作の端末からネットワークを介して質問文を受信し、回答文をその端末へ送信する（例えば非特許文献１参照）。
【０００４】
また、他の技術として、インターネット上に、ブログ(Web log)サーバやミニブログ(mini Web log)（例えばtwitter（登録商標））サーバが接続されている。このようなブログサーバは、不特定多数の第三者からのコメント文章を受信し、他の第三者へ公開する。このようなコメント文章は、様々な話題について公開されており、勿論、前述した質問回答システムに入出力される質問文及び回答文に関連するコメント文章も多く議論されている。
【先行技術文献】
【特許文献】
【０００５】
【特許文献１】特開２０１１－８１６２６号公報
【特許文献２】特開２００５－１４１４２８号公報
【特許文献３】特開２００５－２８４２０９号公報
【非特許文献】
【０００６】
【非特許文献１】ＫＤＤＩ、「au one NETコンシェルジュ」、[online]、［平成２４年１０月７日検索］、インターネット＜URL: http://concierge.auone-net.jp/inagoNetPeople/BrowserClient/GUI/kddi_missConcie3/help/help.html＞
【非特許文献２】坪坂正志、「Latent Dirichlet Allocation入門」、[online]、［平成２４年１０月７日検索］、インターネット＜URL:http://www.slideshare.net/tsubosaka/tokyotextmining＞
【非特許文献３】榊博史、松本一則、黒岩眞吾、橋本和夫、「再起演算を用いた自然言語

変換方式」、電子情報通信学会論文誌(D-II), Vol.J72-D-II, No.12, pp.2080-2093, Dec. 1989、[online]、［平成２４年１０月７日検索］、インターネット＜URL:http://jglobal.jst.go.jp/public/20090422/200902065122383276＞

【発明の概要】

【発明が解決しようとする課題】

【０００７】

しかしながら、同じ質問文であっても、そのユーザの質問の意図が複数あり得る場合がある。このような場合、ユーザに対して、適切な回答文が返答されない場合が多い。

【０００８】

ユーザの質問文の例

Q「携帯電話機の紛失」

この質問文に対して、質問回答システムは、以下の２つキーワードを抽出する。

「携帯電話機」「紛失」

これらキーワードをクエリとして回答文を検索すると、複数の回答の選択肢がある。

A「携帯探せて安心サービスの申込方法」に関する回答文

A「携帯探せて安心サービスの利用方法」に関する回答文

この場合、ユーザとしては、紛失した携帯電話機を遠隔からロックする「利用方法」を問い合わせたつもりであるにも拘わらず、質問回答システムは、「申込方法」について回答してしまう場合もある。

【０００９】

そこで、本発明は、ユーザの質問文に対して複数の回答文の候補が存在する場合、ユーザの意図を反映した回答文を明示する（に絞り込む）ことができる質問回答プログラム、サーバ及び方法を提供することを目的とする。

【課題を解決するための手段】

【００１０】

本発明によれば、多数のコメント文章を蓄積したコメント文章蓄積部と、多数の回答文を蓄積した回答文蓄積部とを有し、ユーザからの質問文に対する回答文を抽出するようにコンピュータを機能させる質問回答プログラムであって、

質問文を入力する質問文入力手段と、

質問文に含まれる複数の質問キーワードを抽出する質問キーワード抽出手段と、

コメント文章蓄積部を用いて、質問キーワードを含むコメント文章を検索するコメント文章検索手段と、

検索された複数のコメント文章を、述語項構造解析によって、述語項構造の分布から複数個のトピックグループに分類するトピック分類手段と、

各トピックグループに含まれるコメント文章群と、各回答文に含まれる文章との間の類似度を算出し、各トピックグループに類似度が所定閾値以上となる回答文を対応付ける回答文検出手段と、

各トピックグループについて、対応付けられた回答文に含まれる述語項構造の中で、当該トピックグループを特徴付ける代表述語項構造を抽出する代表述語項構造抽出手段と、

各トピックグループについて、当該トピックグループのみに出現する代表述語項構造を、差分述語項構造として抽出する差分述語項構造抽出手段と、

回答文検出手段によって検出された回答文を、対応する１つ以上の差分述語項構造に基づく文章と共に明示する回答文出力手段と

してコンピュータを機能させることを特徴とする。

【００１１】

本発明の質問回答プログラムにおける他の実施形態によれば、

複数の差分述語項構造に基づく文章を、ユーザインタフェースを介してユーザに明示すると共に、ユーザ操作に応じていずれか１つの差分述語項構造を選択させる差分述語項構造選択手段を更に有し、

回答文出力手段は、選択された文章の差分述語項構造に対応する回答文を、ユーザイン

タフェースを介して明示する
ようにコンピュータを機能させることも好ましい。
【００１２】
本発明の質問回答プログラムにおける他の実施形態によれば、トピック分類手段は、当該コメント文章を、分類された各トピックグループに属する確からしさ（トピック比率）を算出するＬＤＡ(Latent Dirichlet Allocation)アルゴリズムを用いて、いずれか１つのトピックグループに分類するようにコンピュータを機能させることも好ましい。
【００１３】
本発明の質問回答プログラムにおける他の実施形態によれば、
回答文検出手段は、
各トピックグループに含まれるコメント文章群から、述語項構造解析によって述語項構造を抽出すると共に、トピックグループにおける第１の特徴ベクトルを算出し、
回答文蓄積部に蓄積された各回答文から、述語項構造解析によって述語項構造を抽出すると共に当該回答文における第２の特徴ベクトルとを算出し、
トピックグループの第１のベクトルと、回答文の第２のベクトルとの間のコサイン距離に基づいて類似度を算出する
ようにコンピュータを機能させることも好ましい。
【００１４】
本発明の質問回答プログラムにおける他の実施形態によれば、代表述語項構造抽出手段は、各トピックグループの代表述語項構造を、赤池情報量基準に応じて優先順に並べるようにコンピュータを機能させることも好ましい。
【００１５】
本発明の質問回答プログラムにおける他の実施形態によれば、
コメント文章は、不特定多数の第三者によって投稿されたものであって、
コメント文章蓄積部は、ミニブログ(mini Web log)サーバに投稿されたコメント文章を収集し蓄積したものであるようにコンピュータを機能させることも好ましい。
【００１６】
本発明によれば、多数のコメント文章を蓄積したコメント文章蓄積部と、多数の回答文を蓄積した回答文蓄積部とを有し、ユーザからの質問文に対する回答文を抽出する質問回答サーバであって、
端末から、質問文を入力する質問文入力手段と、
質問文に含まれる複数の質問キーワードを抽出する質問キーワード抽出手段と、
コメント文章蓄積部を用いて、質問キーワードを含むコメント文章を検索するコメント文章検索手段と、
検索された複数のコメント文章を、述語項構造解析によって、述語項構造の分布から複数個のトピックグループに分類するトピック分類手段と、
各トピックグループに含まれるコメント文章群と、各回答文との間の類似度を算出し、各トピックグループに類似度が所定閾値以上となる回答文を対応付ける回答文検出手段と、
各トピックグループについて、対応付けられた回答文に含まれる述語項構造の中で、当該トピックグループを特徴付ける代表述語項構造を抽出する代表述語項構造抽出手段と、
各トピックグループについて、当該トピックグループのみに出現する代表述語項構造を、差分述語項構造として抽出する差分述語項構造抽出手段と、
回答文検出手段によって検出された回答文を、対応する１つ以上の差分述語項構造に基づく文章と共に明示する回答文出力手段と
を有することを特徴とする。
【００１７】
本発明の質問回答サーバにおける他の実施形態によれば、
複数の差分述語項構造に基づく文章を、ユーザインタフェースを介してユーザに明示すると共に、ユーザ操作に応じていずれか１つの差分述語項構造を選択させる差分述語項構

造選択手段を更に有し、
回答文出力手段は、選択された文章の差分述語項構造に対応する回答文を、ユーザインタフェースを介して明示することも好ましい。
【００１８】
本発明によれば、多数のコメント文章を蓄積したコメント文章蓄積部と、多数の回答文を蓄積した回答文蓄積部とを有し、ユーザからの質問文に対する回答文を抽出する装置における質問回答方法であって、
質問文を入力する第１のステップと、
質問文に含まれる複数の質問キーワードを抽出する第２のステップと、
コメント文章蓄積部を用いて、質問キーワードを含むコメント文章を検索する第３のステップと、
検索された複数のコメント文章を、述語項構造解析によって、述語項構造の分布から複数個のトピックグループに分類する第４のステップと、
各トピックグループに含まれるコメント文章群と、各回答文との間の類似度を算出し、各トピックグループに類似度が所定閾値以上となる回答文を対応付ける第５のステップと、
各トピックグループについて、対応付けられた回答文に含まれる述語項構造の中で、当該トピックグループを特徴付ける代表述語項構造を抽出する第６のステップと、
各トピックグループについて、当該トピックグループのみに出現する代表述語項構造を、差分述語項構造として抽出する第７のステップと、
第５のステップによって検出された回答文を、対応する１つ以上の差分述語項構造に基づく文章と共に明示する第８のステップと
を有することを特徴とする。
【００１９】
本発明の質問回答方法における他の実施形態によれば、
第８のステップについて、
複数の差分述語項構造に基づく文章を、ユーザインタフェースを介してユーザに明示すると共に、ユーザ操作に応じていずれか１つの差分述語項構造を選択させ、
選択された文章の差分述語項構造に対応する回答文を、ユーザインタフェースを介して明示することも好ましい。
【発明の効果】
【００２０】
本発明の質問回答プログラム、サーバ及び方法によれば、ユーザの質問文に対して複数の回答文の候補が存在する場合、ユーザの意図を反映した回答文を明示する（に絞り込む）ことができる。
【図面の簡単な説明】
【００２１】
【図１】本発明におけるシステム構成図である。
【図２】本発明における質問回答サーバの機能構成図である。
【図３】質問キーワード抽出部及びコメント文章検索部の処理を表す説明図である。
【図４】トピック分類部の処理を表す説明図である。
【図５】回答文検出部の処理を表す説明図である。
【図６】代表述語項構造抽出部、差分述語項構造抽出部、差分述語項構造選択部及び回答文出力部の処理を表す説明図である。
【図７】本発明におけるシーケンス図である。
【発明を実施するための形態】
【００２２】
以下、本発明の実施の形態について、図面を用いて詳細に説明する。
【００２３】
図１は、本発明におけるシステム構成図である。

【００２４】
図１によれば、インターネット上に、本発明における質問回答サーバ１が接続されている。質問回答サーバ１は、回答文を予め蓄積しているものであってもよいし、他の回答文蓄積サーバ２から回答文を受信するものであってもよい。尚、本発明によれば、ＦＡＱのような質問文候補と回答文候補とを予め紐付けて記憶しておく必要はない。あくまで、回答文候補のみを予め蓄積している。
【００２５】
質問者が操作する端末４は、アクセスネットワーク及びインターネットを介して、質問回答サーバ１へアクセスする。そして、端末４は、質問文を質問回答サーバ１へ送信し、これに対し、質問回答サーバ１から回答文を受信する。以下の実施形態の中では、質問者が自然言語のテキストで端末４へ入力することを想定しているが、質問者が音声で入力しテキストに変換されたものであってもよい。
【００２６】
また、図１によれば、不特定多数の第三者から投稿されたコメント文章を公開するブログサーバ３が、インターネットに更に接続されている。ブログサーバ３は、例えばtwitter（登録商標）サーバのようなミニブログサーバである。不特定多数の第三者は、自ら所持する端末５を用いて、ミニブログサーバ３へコメント文章を自由に投稿することができる。
【００２７】
本発明における質問回答サーバ１は、ミニブログサーバ３から大量のコメント文章を収集する。そして、質問回答サーバ１は、ユーザの質問文に対して複数の回答文の候補が存在する場合、収集したコメント文章を用いて、ユーザの意図を反映した回答文を明示する（に絞り込む）。
【００２８】
図２は、本発明における質問回答サーバの機能構成図である。
【００２９】
図２によれば、質問回答サーバ１は、通信インタフェース部１０と、回答文蓄積部１０１と、回答文取得部１１１と、コメント文章蓄積部１０２と、コメント文章収集部１１２とを有する。
【００３０】
回答文蓄積部１０１は、多数の回答文を蓄積する。回答文取得部１１１が、これら回答文を、ネットワークを介して回答文蓄積サーバ２から受信し、回答文蓄積部１０１へ蓄積するものであってもよい。
【００３１】
コメント文章蓄積部１０２は、不特定多数の第三者によって投稿された多数のコメント文章を蓄積する。コメント文章収集部１１２が、これらコメント文章を、ネットワークを介してブログサーバ３から受信し、コメント文章蓄積部１０２へ蓄積するものであってもよい。
【００３２】
「コメント文章」とは、例えばtwitter（登録商標）で発信された、日本語の「つぶやき」（最大文字数：１４０文字）のようなものである。コメント文章は、例えば、ユーザid(from_user_id)、つぶやきＩＤ(id_str)、発信時間(created_at)、つぶやき(texts)を含む。ここで、コメント文章収集部１１２は、予め指定した複数のキーワードを含むコメント文章のみを収集することもできる。
【００３３】
また、図２によれば、質問回答サーバ１は、質問文入力部１２１と、質問キーワード抽出部１２２と、コメント文章検索部１２３と、トピック分類部１２４と、回答文検出部１２５と、代表述語項構造抽出部１２６と、差分述語項構造抽出部１２７と、差分述語項構造選択部１２８と、回答文出力部１２９とを有する。これら機能構成部は、サーバに搭載されたコンピュータを機能させるプログラムを実行することによって実現される。

【００３４】
［質問文入力部１２１］
質問文入力部１２１は、質問者の端末４から、ネットワークを介して質問文を受信する。例えばユーザの質問文は、以下のようなものである。
Ｑ「携帯電話機の紛失」
その質問文は、質問キーワード抽出部１２２へ出力される。
【００３５】
図３は、質問キーワード抽出部及びコメント文章検索部の処理を表す説明図である。
【００３６】
［質問キーワード抽出部１２２］
質問キーワード抽出部１２２は、質問文に含まれる複数の質問キーワードを抽出する。ここで、質問キーワード抽出部１２２は、質問文から形態素解析によってキーワードを抽出すると共に、ＴＦ－ＩＤＦ（Term Frequency - Inverse Document Frequency：単語の出現頻度－逆出現頻度）によって特徴的な単語を、質問キーワードとして抽出する。
【００３７】
質問キーワード抽出部１２２は、最初に、質問文から形態素解析によって単語を抽出する。「形態素解析」とは、文章を、意味のある単語に区切り、辞書を利用して品詞や内容を判別する技術をいう。「形態素」とは、文章の要素のうち、意味を持つ最小の単位を意味する。形態素解析のように単語単位で検索することなく、文字単位で分解し、後続のN-1文字を含めた状態で出現頻度を求める「N-gram」によって解析するものであってもよい。
【００３８】
次に、ＴＦ－ＩＤＦによって特徴的なキーワードを、質問キーワードとして抽出する。ＴＦ－ＩＤＦとは、各単語に重みを付けて、クエリから文章をベクトル空間で表し、文章とクエリの類似度でランク付けをする技術である。ランク付けられた値が高いほど、重要キーワードと認識される。
【００３９】
図３の例によれば、以下のように抽出される。
質問文　　「携帯電話機の紛失」
質問キーワード「携帯電話機」「紛失」
【００４０】
［コメント文章検索部１２３］
コメント文章検索部１２３は、コメント文章蓄積部１０２を用いて、質問キーワードを含むコメント文章を検索する。具体的には、質問キーワードをクエリとして、各コメント文章からＴＦ(Term Frequency)値やＤＦ(Document Frequency)値を抽出し、これら値が所定閾値以上となる複数のコメント文章を検索する。ＴＦ値は、文章における検索語の出現頻度をいい、ＤＦ値は、索引語が現れる相対文章頻度をいう。コメント文章検索部１２３は、ソーシャルメディア検索機能であって、投稿された大量のつぶやきの中から、質問キーワードに関するつぶやきのみを検索するようなものである。
【００４１】
図３によれば、例えば４つのコメント文章が検索されている。これらコメント文章には、少なくとも「携帯電話機」又は「紛失」が含まれている。
【００４２】
図４は、トピック分類部の処理を表す説明図である。
【００４３】
［トピック分類部１２４］
トピック分類部１２４は、検索された複数のコメント文章を、述語項構造解析によって、述語項構造の分布から複数個のトピックグループに分類する。トピック分類部１２４は、当該コメント文章を、分類された各トピックグループに属する確からしさ（トピック比率）を算出するＬＤＡ(Latent Dirichlet Allocation)アルゴリズムを用いて、いずれか

１つのトピックグループに分類する。特に、トピック分類部１２４のＬＤＡは、キーワードによる分類でなく、述語項構造による分類である。

【００４４】

ＬＤＡは、単語文書行列を次元圧縮する技術（ＬＳＩ(latent Semantic Indexin)）に対して、単語の特徴ベクトルに揺らぎに基づく確率的な枠組みを導入したものである（例えば非特許文献２参照）。その圧縮した次元の集合をトピックという。

【００４５】

また、「述語項構造」とは、文章中の述語に対して「項」となる名詞句等を当てたものである。述語項構造を用いることによって、文章の意味の骨格を把握することができる。述語項構造解析として、例えばフリーソフトであるSyncha等の述語項構造解析器を用いることができる。

【００４６】

述語項構造は、「述語」に対する「目的語」とその格とから構成される。図４によれば、例えば「携帯を探す」の述語項構造は、述語「探す」に対して目的語「携帯」及び格「ヲ」からなる。また、例えば「サービスに申し込む」の述語項構造は、述語「申し込む」に対して目的語「サービス」及び格「ニ」からなる。

【００４７】

トピック分類部１２４は、以下のステップで処理を実行する。

（Ｓ４１）質問キーワードに関する多数のコメント文章から、述語項構造毎の出現頻度（出現回数）をＬＤＡ処理へ入力する。そして、コメント文章毎に、各述語項構造の出現頻度を計数する。

（Ｓ４２）次に、本件でのＬＤＡ処理では、トピック毎の述語項構造分布や、コメント文章（ネット側意見）毎のトピック比率を取得する。このトピック比率によって、コメント文章が属するトピックグループに分類する。そして、トピックグループ毎に、全てのコメント文章に含まれる各述語項構造の出現頻度を計数する。

（Ｓ４３）次に、コメント文章毎に、各トピックグループに属する述語項構造を計数する。そして、コメント文章を計数値の高いトピックグループに分類する。

【００４８】

図５は、回答文検出部の処理を表す説明図である。

【００４９】

［回答文検出部１２５］

回答文検出部１２５は、各トピックグループに含まれるコメント文章群と、各回答文に含まれる文章との間の類似度を算出し、各トピックグループに類似度が所定閾値以上となる回答文を対応付ける。

【００５０】

類似度の算出方法は、例えば以下のようにする。

（Ｓ５１）回答文検出部１２５は、各トピックグループに含まれるコメント文章群から述語項構造解析によって述語項構造を抽出すると共に、トピックグループにおける第１の特徴ベクトルを算出する。

各トピックグループ：Ｃi(i=1,2,・・・)

トピックグループiに含まれるコメント文章：Ｔij(j=1,2,・・・)

（Ｓ５２）回答文蓄積部１０１に蓄積された各回答文から述語項構造解析によって述語項構造を抽出すると共に、当該回答文における第２の特徴ベクトルとを算出する。

回答文：Ak(k=1,2,・・・)

（Ｓ５３）トピックグループの第１のベクトルと、回答文の第２のベクトルとの間のコサイン距離に基づいて類似度を算出する。具体的には、各コメント文章Ｔi1, Ｔi2,・・・を含むトピックグループＣiと、回答文Ａjとの類似度Dist（Ｃi,Ａj）を算出する。

Dist（Ｃi,Ａj）＝cosin距離Ｄ(Ｔi1,Ａj), Ｄ(Ｔi2, Ａj), ・・・の平均値

＝ａｒｇｉ ｍａｘ（Dist(Ｃi,Ａj)）

【００５１】

図５によれば、回答文蓄積部１０１には、多数の回答文が蓄積されている。
回答文１「・・・」
回答文２「携帯探せて安心サービスの申込方法」
述語項構造＝述語「探す」、目的語「携帯」、ヲ格
述語項構造＝述語「申し込む」、目的語「サービス」、ニ格
回答文３「・・・」
回答文４「・・・」
回答文５「携帯探せて安心サービスの利用方法」
述語項構造＝述語「探す」、目的語「携帯」、ヲ格
述語項構造＝述語「利用する」、目的語「サービス」、ヲ格
回答文６「・・・」

【００５２】
図５によれば、トピックグループ１と回答文２との類似度が、所定閾値δよりも高い場合、両者は類似していると判定されている。また、トピックグループ２と回答文５との類似度が、所定閾値δよりも高い場合、両者は類似していると判定されている。これによって、トピックグループＣ1,Ｃ2,・・・毎に、０個以上の回答文が割り当てられる。

【００５３】
図６は、代表述語項構造抽出部、差分述語項構造抽出部、差分述語項構造選択部及び回答文出力部の処理を表す説明図である。

【００５４】
［代表述語項構造抽出部１２６］
代表述語項構造抽出部１２６は、各トピックグループについて、対応付けられた回答文に含まれる述語項構造の中で、当該トピックグループを特徴付ける代表述語項構造を抽出する。

【００５５】
図６によれば、トピックグループ１に対応する回答文２からは、以下の表のような述語項構造が抽出される。
回答文２「携帯探せて安心サービスの申し込み方法は以下のようになります・・・」
述語項構造＝述語「探す」、目的語「携帯」、ヲ格
述語項構造＝述語「申し込む」、目的語「サービス」、ニ格
回答文５「携帯探せて安心サービスの利用方法は以下のようになります・・・」
述語項構造＝述語「探す」、目的語「携帯」、ヲ格
述語項構造＝述語「利用する」、目的語「サービス」、ヲ格

【００５６】
ここで、述語項構造抽出部１２６は、各トピックグループの代表述語項構造を、赤池情報量基準に応じて優先順に並べることも好ましい。トピックグループＣ1,Ｃ2,・・・に割り当てられた回答文のいずれかに出現する述語項構造を、s1,s2,・・・とする。ここでは、述語項構造E(i)が、トピックグループＣjの判別に役立つかどうかの指標を与える。

【００５７】
以下では、述語項構造sが、トピックグループＣの判別に役立つかどうかの指標Ｅ(s,Ｃ)の算出方法を表す。

【００５８】
（Ｓ１）トピックグループ含まれるコメント文章（つぶやき）の集合Ｕから、以下の４種類の頻度を得る。
ｎ11＝トピックグループＣに類似し、述語項構造sが出現するコメント文章の数
ｎ12＝トピックグループＣ以外に類似し、述語項構造sが出現するコメント文章の数
ｎ21＝トピックグループＣに類似し、述語項構造sが出現しないコメント文章の数
ｎ22＝トピックグループＣ以外に類似し、述語項構造sが出現しないコメント文章の数

【００５９】
（Ｓ２）次に、ｎ11,ｎ12,ｎ21,ｎ22に対して、赤池情報量規準(ＡＩＣ：Akaike's Infor

mation Criterion)を用いて、独立モデルに対する値MLL_IM(s,C)及び従属モデルに対する値MLL_DM(s,C)を算出する。これは、述語項構造とトピックグループとの組毎の不当割合を算出する。

MLL_IM(s,C)＝(n11+n12) log(n11+n12)
＋(n11+n21) log(n11+n21)
＋(n21+n22) log(n21+n22)
＋(n12+n22) log(n12+n22)－2N log N

MLL_DM(s,C)＝n11 log n11＋n12 log n12＋n21 log n21＋n22 log n22－N log N

但し、N＝n11＋n12＋n21＋n22

【００６０】

（Ｓ３）前述のMLL_IM(s,C)及びMLL_DM(s,C)から、以下のＥ(s,C)を算出する。

AIC_IM(s,C)＝-2 × MLL_IM(s,C) ＋ 2×2

AIC_DM(s,C)＝-2 × MLL_DM(s,C) ＋ 2×3

Ｅ(s,C)＝AIC_IM(s, C) － AIC_DM(s,C)

【００６１】

前述で算出されたＥ(s,Ｃ)は、述語項構造sがトピックグループＣに偏って出現する不当割合を表す。Ｅ(s,Ｃ)は、赤池情報量基準に従って、トピックグループＣの判別に役立つ述語項構造ほど、Ｅ(s,Ｃ)の値が高くなる。本発明によれば、各トピックグループＣiに対し、Ｅ(s,C)の値が大きい順に、ｍ個の述語項構造 Ｃi,1、Ｃi,2、Ｃi,3、・・・Ｃi,m を抽出し、トピックグループＣiの代表述語項構造とする。

【００６２】

［差分述語項構造抽出部１２７］

差分述語項構造抽出部１２７は、各トピックグループについて、当該トピックグループのみに出現する代表述語項構造を、差分述語項構造として抽出する。回答文２及び５について、［述語項構造＝述語「探す」、目的語「携帯」、ヲ格］は共通する。そこで、図６によれば、以下の差分述語項構造が抽出される。

回答文２「携帯探せて安心サービスの申し込み方法は以下のようになります・・・」
述語項構造＝述語「申し込む」、目的語「サービス」、ニ格

回答文５「携帯探せて安心サービスの利用方法は以下のようになります・・・」
述語項構造＝述語「利用する」、目的語「サービス」、ヲ格

【００６３】

［差分述語項構造選択部１２８］

差分述語項構造選択部１２８は、複数の差分述語項構造に基づく文章を、ユーザインタフェースを介してユーザに明示する。

【００６４】

差分述語項構造からの日本語文章を生成するために、例えば以下のようなルールが設定される。

（ルール１）ヲ格のみからなる述語項構造Ｓの場合
->Ｗ（Ｓ，ヲ格）＋「を」＋Ｗ（Ｓ，述語）

（ルール２）ヲ格とデ格からなる述語項構造Ｓの場合
->Ｗ（Ｓ，デ格）＋「で」＋Ｗ（Ｓ，ヲ格）＋「を」＋Ｓ（述語）

（ルール３）差分述語項構造Ｓ１のヲ格と、共通述語項構造Ｂのデ格とが一致する場合、
->Ａから生成した日本語のヲ格の前方修飾語として、Ｂの日本語を埋め込む

Ｗ（Ｓ，ヲ格）は、述語項構造Ｓのヲ格の単語を表す。

Ｗ（Ｓ，デ格）は、述語項構造Ｓのデ格の単語を表す。

尚、このようなルール基づく日本語の生成については、機械翻訳システムの技術が適用できる（例えば非特許文献３参照）。

【００６５】

例えば、ルール１～３を用いて、図６によれば、トピックグループ毎に、以下の２つの日本語文章が生成される。

「携帯を探すサービスを申し込む」
「携帯を探すサービスを利用する」

【００６６】

これに対し、端末４は、ユーザ操作に応じていずれか１つの差分述語項構造の文章を選択させる。ユーザから見ると、例えば、質問文をキーボードで入力した後、トピックグループ毎の差分述語項構造に基づいた自然な日本語文章がディスプレイに表示される。そして、ユーザは、いずれかの文章を選択することができる。ここで、図６によれば、ユーザは、「携帯を探すサービスを利用する」を選択している。ユーザに選択された文章の差分述語項構造は、回答文出力部１２９へ出力される。

【００６７】

［回答文出力部１２９］

回答文出力部１２９は、回答文検出部１２５によって検出された回答文を、対応する１つ以上の差分述語項構造の文章と共に明示する。本発明によれば、ユーザの質問に曖昧性があり、コメント文章群が複数のトピックグループに分類され、各トピックグループに対応付けられた回答文を得ることができる。ここで、この得られた回答文の数が少ない場合、差分述語項構造に基づく文章は、提示される回答文の傾向をユーザが認識するために有益な情報となる。

【００６８】

また、回答文の数が多い場合、ユーザとインタラクション（やりとり）をすることによって、回答文を絞り込むことが好ましい。そこで、回答文出力部１２９は、選択された差分述語項構造に対応する回答文を、ユーザインタフェースを介して明示する。例えば、その回答文を、ユーザが視認するディスプレイに表示する。図６によれば、「携帯探せて安心サービスの利用方法」の回答文が、ユーザへ表示される。これによって、ユーザは、質問文に対する回答文を認識することができる。

【００６９】

図７は、本発明におけるシーケンス図である。

【００７０】

（Ｓ７１）質問者が操作する端末４から、質問回答サーバ１へ、ユーザの質問文が送信される（図２の質問文入力部１２１参照）。

（Ｓ７２）質問回答サーバ１は、質問文に含まれる複数の質問キーワードを抽出する（図２の質問キーワード抽出部１２２参照）。

（Ｓ７３）質問回答サーバ１は、コメント文章蓄積部１０２を用いて、質問キーワードを含むコメント文章を検索する（図２のコメント文章検索部１２３参照）。

（Ｓ７４）質問回答サーバ１は、検索された複数のコメント文章を、述語項構造解析によって、述語項構造の分布から複数個のトピックグループに分類する（図２のトピック分類部１２４参照）。

（Ｓ７５）質問回答サーバ１は、各トピックグループに含まれるコメント文章群と、各回答文との間の類似度を算出し、各トピックグループに類似度が所定閾値以上となる回答文を対応付ける（図２の回答文検出部１２５参照）。

（Ｓ７６）質問回答サーバ１は、各トピックグループについて、対応付けられた回答文に含まれる述語項構造の中で、当該トピックグループを特徴付ける代表述語項構造を抽出する（図２の代表述語項構造抽出部１２６参照）。

（Ｓ７７）質問回答サーバ１は、各トピックグループについて、当該トピックグループのみに出現する代表述語項構造を、差分述語項構造として抽出する（図２の差分述語項構造抽出部１２７参照）。

（Ｓ７８）質問回答サーバ１は、複数の差分述語項構造に基づく文章を、ユーザ操作の端末４へ送信する（図２の差分述語項構造選択部１２８参照）。そして、端末４では、ユーザ操作に応じていずれか１つの文章が選択させる。選択された文章の差分述語項構造は、端末４から質問回答サーバ１へ送信される。

（Ｓ７９）質問回答サーバ１は、選択された差分述語項構造に対応する回答文を、ユーザ

の端末４へ送信する（図２の回答文出力部１２９参照）。
【００７１】
前述したように本発明の質問回答サーバによれば、例えばtwitterのような大量のコメント文章から、質問文の意図を表す代表的な述語項構造を抽出し、質問文を補完することによって、回答文を高精度に検索することができる。具体的には、最初に、質問文に含まれるキーワードを抽出してソーシャルメディアを検索し、大量の検索結果を複数のトピックグループ（トピック毎に１つの検索意図に対応）に高速に分類し、各トピックに類似する回答文を回答文蓄積部から検索する。次に、各トピックグループに特有の単語（差分述語項構造）を自動的に抽出してユーザに提示し、ユーザの選択結果に従った回答文に絞り込んで、ユーザとの対話形式を繰り返し実行することができる。
【００７２】
以上、詳細に説明したように、本発明の質問回答プログラム、サーバ及び方法によれば、ユーザの質問文に対して複数の回答文の候補が存在する場合、ユーザの意図を反映した回答文を明示する（に絞り込む）ことができる。
【００７３】
最後に、本発明が、キーワード検索ではなく、述語項構造検索を用いた効果について詳述する。
【００７４】
一般に、例えば、「携帯電話が紛失したらどうしよう」というユーザからの短い質問の場合、「携帯電話紛失に備えたサービス申し込みの要望」なのか、又は、「端末の紛失への対応法に関する問い合わせ」なのかといった曖昧性が存在する。これに対し、コンテキストに依存した曖昧性を検出し、対話形式で回答候補を絞り込みながらＴＩＰＳ等を返答する技術が提案されている（本願と同一出願人及び同一発明者によって出願された平成２４年１月１２日付け特許出願、以下「先の出願に係る発明」と称す）。この技術によれば、コンテキストの曖昧性を高速かつ適切に検出し、「安心、申し込み」や「端末、発見」といったキーワードを対話の選択肢として提示することができる。
【００７５】
しかしながら、先の出願に係る発明によれば、第１の課題として、コンテキストを絞り込む際に「携帯紛失、サービス、申し込み」又は「携帯紛失、サービス、利用」といったキーワードが提示されるだけであって、、ユーザにとっては、コンテキストの差異を理解しづらい。
【００７６】
また、第２の課題として、コンテキストの絞り込みが終了しても、非特許文献３に記載された技術のようなキーワードによる検索によれば、検索条件としての情報が不足することがある。例えば、「携帯＋発見」といったキーワードで検索する場合、宝探しゲームのように携帯電話機で何かを発見するサービスや、携帯電話を発見するサービスを発見するサービスの情報の両方がキーワード検索結果に現れる。このため、回答精度を現状以上に向上させることが難しい。
【００７７】
第１の課題に対して、本発明によれば、差分キーワードの単純な提示ではなく、動詞を中心に主語や目的語等の関係をリンクで表す「述語項構造」と呼ばれるデータから自然な応対の文章を生成し、それを利用者に提示する手法を用いることで理解度が深まることが期待できる。例えば、「携帯＋紛失」のユーザクエリに対して従来システムが「安心＋申し込み」もしくは、「発見」といったキーワードを提示していたのに対し、「端末を紛失した際に安心できる申し込みに関する情報」もしくは「携帯電話を紛失した際に端末を発見すること」といった自然な文章をユーザに提示することでユーザの利便性を向上させる手段を提供する。
【００７８】
第２の課題に対して、本発明によれば、述語項構造を検索パラメータとして知識源のテキストを検索することにより、コンテキスト絞り込み後の検索精度を大きく向上させる。

例えば、「端末を発見する」と「端末で発見する」の意味を区別できる述語項構造を用いるので、従来型のキーワード検索で行われていた過剰検出が減る。

【００７９】

また、述語項構造を使用して文書の類似性を判定する場合、抽出した述語項構造の一致度合いを判定する必要があり、従来のキーワードを利用した場合より多くの計算時間が必要となることも問題となる。この問題に対しては、述語項構造の中で格と呼ばれるデータスロットに注目し、使用頻度が高いスロットの組み合わせを事例から事前に学習しておき、使用頻度の高い組み合わせに対してはハッシュ関数を使って高速に検索できるようにする。

【００８０】

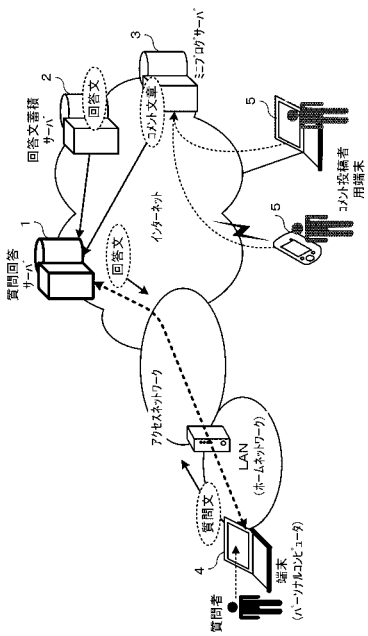
前述した本発明の種々の実施形態について、本発明の技術思想及び見地の範囲の種々の変更、修正及び省略は、当業者によれば容易に行うことができる。前述の説明はあくまで例であって、何ら制約しようとするものではない。本発明は、特許請求の範囲及びその均等物として限定するものにのみ制約される。

【符号の説明】

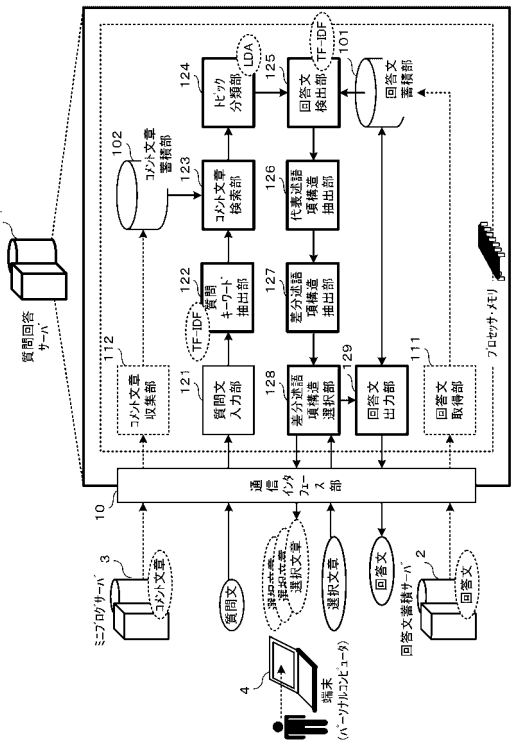
【００８１】

１　質問回答サーバ
１０　通信インタフェース部
１０１　回答文蓄積部
１０２　コメント文章蓄積部
１１１　回答文取得部
１１２　コメント文章収集部
１２１　質問文入力部
１２２　質問キーワード抽出部
１２３　コメント文章検索部
１２４　トピック分類部
１２５　回答文検出部
１２６　代表述語項構造抽出部
１２７　差分述語項構造抽出部
１２８　差分述語項構造選択部
１２９　回答文出力部
２　回答文蓄積サーバ
３　ブログサーバ
４　端末
５　コメント投稿者用の汎用端末

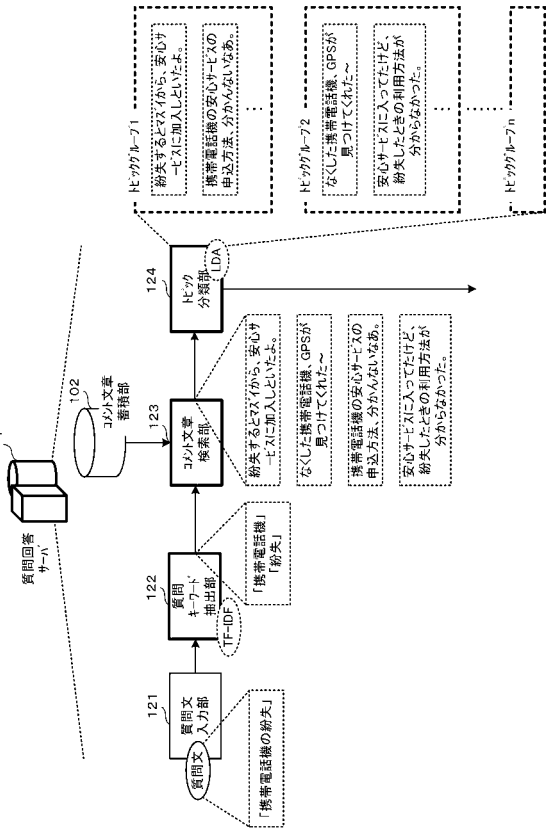
【図１】


質問回答サーバ
回答文蓄積サーバ
ミニブログサーバ
インターネット
アクセスネットワーク
LAN（ホームネットワーク）
質問者
端末（パーソナルコンピュータ）
コメント投稿者用端末
質問文
回答文
コメント文章

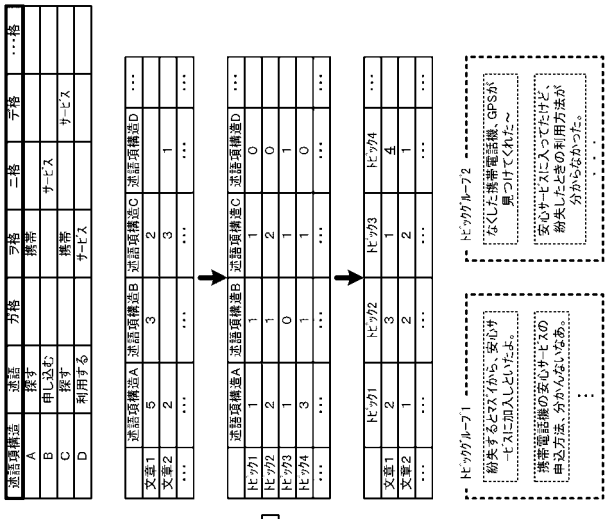
【図２】


質問回答サーバ
コメント文章蓄積部
トピック分類部
LDA
回答文検出部
TF-IDF
回答文蓄積部
コメント文章検索部
代表述語項構造抽出部
差分述語項構造抽出部
質問キーワード抽出部
質問文入力部
差分述語項構造選択部
回答文出力部
回答文取得部
コメント文章収集部
プロセッサ・メモリ
通信インタフェース部
ミニブログサーバ
質問文
選択文章
回答文
回答文蓄積サーバ
端末（パーソナルコンピュータ）

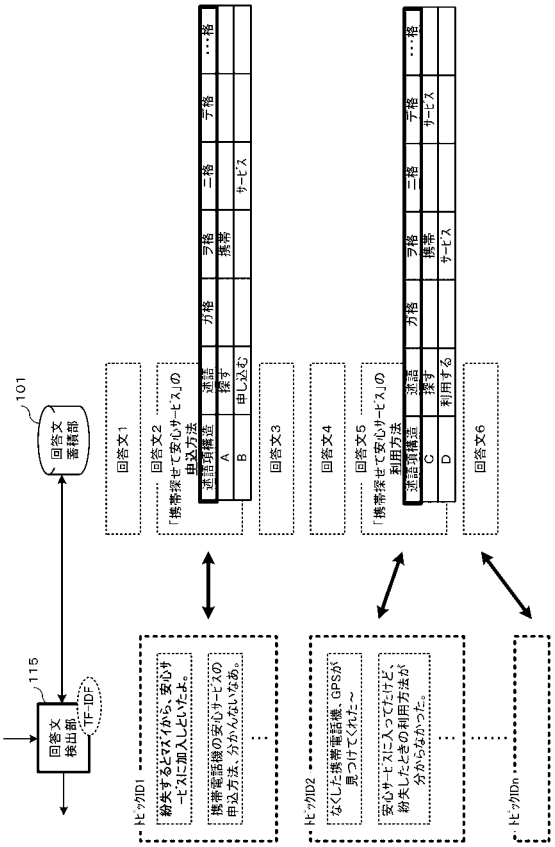
【図３】


質問回答サーバ
コメント文章蓄積部
コメント文章検索部
トピック分類部（LDA）
質問キーワード抽出部
TF-IDF
質問文入力部
質問文
「携帯電話機の紛失」
「携帯電話機」「紛失」
紛失するとマズイから、安心サービスに加入しといたよ。
なくした携帯電話機、GPSが見つけてくれた～
携帯電話機の安心サービスの申込方法、分かんないなあ。
安心サービスに入ってたけど、紛失したときの利用方法が分からなかった。
トピックグループ1
トピックグループ2
トピックグループn

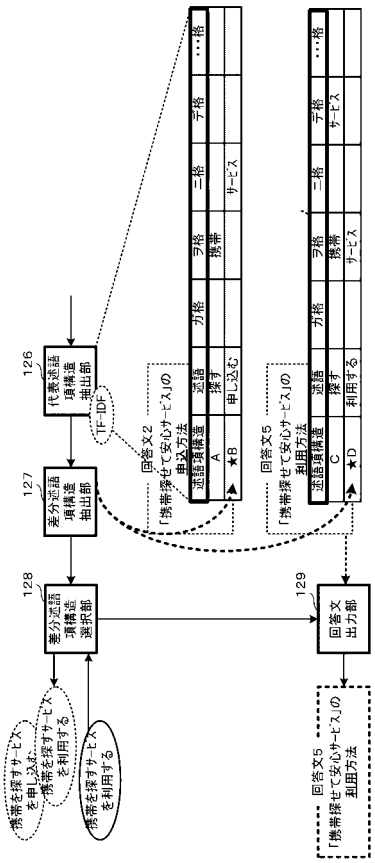
【図４】


述語項構造
訳
二格
デ格
A 携帯 探す
B 申し込む
C 探す サービス
D 利用する
文章1 文章2
トピック1 トピック2 トピック3 トピック4 トピックn
トピックグループ1
トピックグループ2
紛失するとマズイから、安心サービスに加入しといたよ。
携帯電話機の安心サービスの申込方法、分かんないなあ。
なくした携帯電話機、GPSが見つけてくれた～
安心サービスに入ってたけど、紛失したときの利用方法が分からなかった。

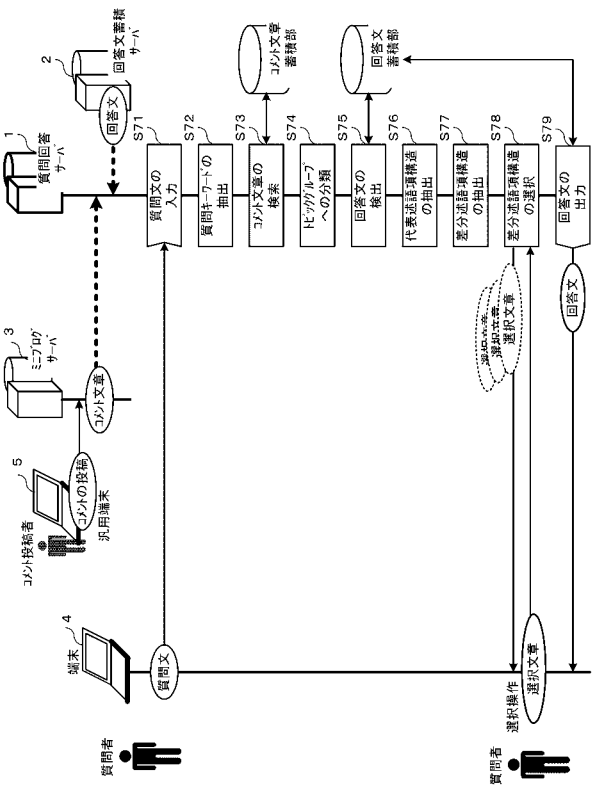
【図５】


101
回答文
蓄積部
115
回答文
検出部
TF-IDF
回答文1
回答文2
「携帯探せて安心サービス」の
申込方法
述語項構造
述語
力格
ヲ格
二格
デ格
…格
A
探す
携帯
サービス
B
申し込む
回答文3
回答文4
回答文5
「携帯探せて安心サービス」の
利用方法
C
D
利用する
回答文6
トピックID1
紛失するとマズイから、安心サ
ービスに加入しといたよ。
携帯電話機の安心サービスの
申込方法、分かんないなあ。
トピックID2
なくした携帯電話機、GPSが
見つけてくれた～
安心サービスに入ってたけど、
紛失したときの利用方法が
分からなかった。
トピックIDn

【図６】


126
代表述語
項構造
抽出部
TF-IDF
127
差分述語
項構造
抽出部
128
差分述語
項構造
選択部
129
回答文
出力部
回答文2
「携帯探せて安心サービス」の
申込方法
回答文5
「携帯探せて安心サービス」の
利用方法
述語項構造
述語
力格
ヲ格
二格
デ格
…格
A
探す
携帯
サービス
★B
申し込む
C
★D
利用する
携帯を探すサービス
を申し込む
携帯を探すサービス
を利用する
携帯を探すサービス
を利用する
回答文5
「携帯探せて安心サービス」の
利用方法

【図７】


1
質問回答
サーバ
2
回答文素積
サーバ
回答文
3
ミニブログ
サーバ
コメント文章
5
コメントの投稿
汎用端末
コメント投稿者
4
端末
質問者
質問文
選択操作
選択文章
コメント文章
蓄積部
回答文
蓄積部
S71
質問文の
入力
S72
質問キーワードの
抽出
S73
コメント文章の
検索
S74
トピックグループ
への分類
S75
回答文の
検出
S76
代表述語項構造
の抽出
S77
差分述語項構造
の抽出
S78
差分述語項構造
の選択
S79
回答文の
出力
回答文
選択候補
選択文章
質問者