(19)日本国特許庁(JP)　(12)特　許　公　報(B2)　(11)特許番号
特許第5126068号
(P5126068)
(45)発行日　平成25年1月23日(2013.1.23)　(24)登録日　平成24年11月9日(2012.11.9)

(51)Int.Cl.　　F I
G06F 17/21 (2006.01)　　G06F 17/21 550A
G06F 17/27 (2006.01)　　G06F 17/27 Z

請求項の数 17　(全 22 頁)

(21)出願番号　特願2008-551075 (P2008-551075)
(86)(22)出願日　平成19年12月21日 (2007.12.21)
(86)国際出願番号　PCT/JP2007/074628
(87)国際公開番号　WO2008/078670
(87)国際公開日　平成20年7月3日 (2008.7.3)
審査請求日　平成22年11月4日 (2010.11.4)
(31)優先権主張番号　特願2006-345757 (P2006-345757)
(32)優先日　平成18年12月22日 (2006.12.22)
(33)優先権主張国　日本国(JP)

(73)特許権者　000004237
日本電気株式会社
東京都港区芝五丁目７番１号
(74)代理人　100080816
弁理士　加藤　朝道
(72)発明者　石川　開
東京都港区芝五丁目７番１号　日本電気株式会社内
(72)発明者　赤峯　享
東京都港区芝五丁目７番１号　日本電気株式会社内

審査官　成瀬　博之

最終頁に続く

(54)【発明の名称】文言い換え方法、プログラムおよびシステム

(57)【特許請求の範囲】
【請求項１】
ユーザによって入力された質問文書に対して、同じ内容の異なる表現の質問文書を生成する文言い換え方法であって、
前記質問文書の入力を受け付けたコンピュータが、
記憶手段に格納されている同一の回答文書を有する質問文書のなかから、設定された閾値以上の類似度を有するセンテンスの組み合わせを抽出する第１のセンテンス抽出ステップと、
前記第１のセンテンス抽出ステップで抽出したセンテンスの組み合わせから、同じ内容の異なる表現のセンテンスに変換するための変換規則を算出する変換規則算出ステップと、
前記変換規則算出ステップで算出した変換規則を用いて、前記閾値以上の類似度を有するセンテンスの組み合わせを前記記憶手段から再度抽出する第２のセンテンス抽出ステップと、
第１および第２のセンテンス抽出ステップで抽出したセンテンスの組み合わせがすべて一致しているかどうかを判定する判定ステップと、
前記変換規則算出ステップで算出した変換規則を用いて、入力された質問文書の言い換え文書を生成する文書生成ステップとを含み、
前記コンピュータが、第１および第２のセンテンス抽出ステップで抽出したセンテンスの組み合わせがすべて一致していると判定されるまで、変換規則算出ステップと、第２の

センテンス抽出ステップとを繰り返し実行する
ことを特徴とする文言い換え方法。
【請求項２】
前記コンピュータが、さらに、
生成した言い換え文書をユーザに対して表示するステップを含み、
前記コンピュータが、前記生成した言い換え文書を、入力された質問文に対する類似度を示す値とともに表示する請求項１記載の文言い換え方法。
【請求項３】
前記文書生成ステップは、
前記コンピュータが、
入力された質問文書を単語列に分解するステップと、
前記変換規則算出ステップで算出した変換規則と前記単語列とを比較し、類似度の高い単語列を複数選択する選択ステップと、
前記選択ステップで選択した単語列を並び替える並び換えステップとを含む
請求項１または請求項２記載の文言い換え方法。
【請求項４】
ユーザによって入力された質問文書に対して、同じ内容の異なる表現の質問文書を生成するための変換規則を算出する変換規則算出方法であって、
前記質問文書の入力を受け付けたコンピュータが、
記憶手段に格納されている同一の回答文書を有する質問文書のなかから、設定された閾値以上の類似度を有するセンテンスの組み合わせを抽出する第１のセンテンス抽出ステップと、
前記第１のセンテンス抽出ステップで抽出したセンテンスの組み合わせから、同じ内容の異なる表現のセンテンスに変換するための変換規則を算出する変換規則算出ステップと、
前記変換規則算出ステップで算出した変換規則を用いて、前記閾値以上の類似度を有するセンテンスの組み合わせを前記記憶手段から再度抽出する第２のセンテンス抽出ステップと、
第１および第２のセンテンス抽出ステップで抽出したセンテンスの組み合わせがすべて一致しているかどうかを判定する判定ステップとを含み、
前記コンピュータが、第１および第２のセンテンス抽出ステップで抽出したセンテンスの組み合わせがすべて一致していると判定されるまで、変換規則算出ステップと、第２のセンテンス抽出ステップとを繰り返し実行する
ことを特徴とする変換規則算出方法。
【請求項５】
前記コンピュータが、
記憶手段に、入力手段から入力した質問文書と、前記質問文書に対する回答文書と、前記質問文書および前記回答文書における対応関係を示した情報とを記憶しており、
同じ内容の回答文書をまとめた１つのグループに対して、前記情報を適用し、同じ内容の回答文書に対応した質問文書のグループを生成する
請求項４記載の変換規則算出方法。
【請求項６】
前記コンピュータが、さらに、
前記記憶手段に格納されている情報を更新するステップを含み、前記コンピュータが、前記記憶手段の情報が更新されたことを契機に、第１のセンテンス抽出ステップにおける処理を開始する請求項４または請求項５記載の変換規則算出方法。
【請求項７】
ユーザによって入力された質問文書に対して、同じ内容の異なる表現の質問文書を生成する文言い換えシステムであって、
質問文書を入力するための入力手段と、

前記入力手段に入力された質問文書と、前記質問文書に対する回答文書とを記憶する記憶手段と、
前記記憶手段に格納されている同一の回答文書を有する質問文書のなかから、設定された閾値以上の類似度を有するセンテンスの組み合わせを抽出するセンテンス抽出手段と、
前記センテンス抽出手段で抽出したセンテンスの組み合わせから、同じ内容の異なる表現のセンテンスに変換するための変換規則を算出する変換規則算出手段と、
前記変換規則算出手段が算出した変換規則を用いて、ユーザによって前記入力手段に入力された質問文書の言い換え文書を生成する文書生成手段とを備えた
ことを特徴とする文言い換えシステム。
【請求項８】
記憶手段は、
入力手段に入力された質問文書と、前記質問文書に対する回答文書とを記憶する第１の記憶手段と、
前記センテンス抽出手段が抽出したセンテンスの組み合わせを記憶する第２の記憶手段と、
前記変換規則算出手段が算出した変換規則を記憶する第３の記憶手段とを含む
請求項７記載の文言い換えシステム。
【請求項９】
前記変換規則は、少なくとも翻訳モデルと、言語モデルと、単語列とを含む請求項７または請求項８記載の文言い換えシステム。
【請求項１０】
前記文書生成手段が生成した言い換え文書を表示する表示手段を備え、
前記表示手段は、前記文書生成手段が生成した言い換え文書を、ユーザによって入力手段に入力された質問文との類似度を示す値とともに表示する
請求項７から請求項９のうちのいずれか１項に記載の文言い換えシステム。
【請求項１１】
入力手段に入力されたコマンドにしたがって、記憶手段における情報を更新する情報更新手段を備え、
前記情報更新手段は、前記記憶手段における情報を更新した場合に、センテンスの組み合わせの抽出を開始するための信号を前記センテンス抽出手段に送信する
請求項７から請求項１０のうちのいずれか１項に記載の文言い換えシステム。
【請求項１２】
ユーザによって入力された質問文書に対して、同じ内容の異なる表現の質問文書を生成するための変換規則を算出する変換規則算出用プログラムであって、
コンピュータに、
記憶手段に格納されている同一の回答文書を有する質問文書のなかから、設定された閾値以上の類似度を有するセンテンスの組み合わせを抽出する第１のセンテンス抽出処理と、
前記第１のセンテンス抽出ステップで抽出したセンテンスの組み合わせから、同じ内容の異なる表現のセンテンスに変換するための変換規則を算出する変換規則算出処理と、
前記変換規則算出処理で算出した変換規則を用いて、前記閾値以上の類似度を有するセンテンスの組み合わせを前記記憶手段から再度抽出する第２のセンテンス抽出処理と、
前記第１および第２のセンテンス抽出処理で抽出したセンテンスの組み合わせがすべて一致しているかどうかを判定する判定処理とを実行させ、
前記第１および第２のセンテンス抽出処理で抽出したセンテンスの組み合わせがすべて一致していると判定されるまで、変換規則算出処理と、第２のセンテンス抽出処理とを繰り返させるための
変換規則算出用プログラム。
【請求項１３】
コンピュータに、

入力手段から入力された質問文書と、前記質問文書に対する回答文書と、前記質問文書および前記回答文書における対応関係を示した情報とを記憶した記憶手段において、
同じ内容の回答文書をまとめた１つのグループに対して、前記情報を適用し、同じ内容の回答文書に対応した質問文書のグループを生成する処理を実行させるための
請求項１２記載の変換規則算出用プログラム。
【請求項１４】
コンピュータに、
前記記憶手段に格納されている情報を更新する処理と、
前記記憶手段の情報が更新されたことを契機に、第１のセンテンス抽出処理を開始する処理とを実行させるための
請求項１２または請求項１３記載の変換規則算出用プログラム。
【請求項１５】
ユーザによって入力された質問文書に対して、同じ内容の異なる表現の質問文書を生成する文言い換え用プログラムであって、
コンピュータに、
記憶手段に格納されている同一の回答文書を有する質問文書のなかから、設定された閾値以上の類似度を有するセンテンスの組み合わせを抽出する第１のセンテンス抽出処理と、
前記第１のセンテンス抽出処理で抽出したセンテンスの組み合わせから、同じ内容の異なる表現のセンテンスに変換するための変換規則を算出する変換規則算出処理と、
前記変換規則算出処理で算出した変換規則を用いて、前記閾値以上の類似度を有するセンテンスの組み合わせを前記記憶手段から再度抽出する第２のセンテンス抽出処理と、
前記第１および第２のセンテンス抽出処理で抽出したセンテンスの組み合わせがすべて一致しているかどうかを判定する判定処理と、
前記第１および第２のセンテンス抽出処理で抽出したセンテンスの組み合わせがすべて一致していると判定されるまで、変換規則算出処理と、第２のセンテンス抽出処理とを繰り返す処理と、
前記変換規則算出処理で算出した変換規則を用いて、入力された質問文書の言い換え文書を生成する文書生成処理とを
実行させるための文言い換え用プログラム。
【請求項１６】
コンピュータに、入力された質問文に対する類似度を示す値とともに、生成した言い換え文書を表示する処理を実行させるための請求項１５記載の文言い換え用プログラム。
【請求項１７】
コンピュータに、
文書生成処理で、
入力された質問文書を単語列に分解する処理と、
前記変換規則算出処理で算出した変換規則と前記単語列とを比較し、類似度の高い単語列を複数選択する選択処理と、
前記選択処理で選択した単語列を並び替える並び換え処理とを実行させるための
請求項１５または請求項１６記載の文言い換え用プログラム。
【発明の詳細な説明】
【技術分野】
【０００１】
（関連出願）本願は、先の日本特許出願２００６－３４５７５７号（２００６年１２月２２日出願）の優先権を主張するものであり、前記先の出願の全記載内容は、本書に引用をもって繰込み記載されているものとみなされる。
本発明は、文言い換え方法、変換規則算出方法、文言い換えシステム、変換規則算出用プログラムおよび文言い換え用プログラムに関し、特に文の言い換え規則を学習する文言い換え方法、変換規則算出方法、文言い換えシステム、変換規則算出用プログラムおよび

文言い換え用プログラムに関する。
【背景技術】
【０００２】
人間によって入力された文章をコンピュータが他の国の言語の文章に変換する機械翻訳が広く用いられているが、同一の言語間を対象とした場合には、文章表現を変換するための文の言い換え方法が用いられている（非特許文献１参照）。
【０００３】
非特許文献１には、従来の文の言い換え表現作成方法の一例が記載されている。非特許文献１に記載されている文の言い換え表現作成方法は、統語的な素性（性質）で表現された言い換え規則をあらかじめ人間が作成する方法である。そして、得られた言い換え表現に対しては、最終的に人手による確認が必要である。
【０００４】
また、文の言い換え方法は、変換した質問文に対して自動的に回答を出力する質疑応答システム等にも用いられている（特許文献１参照）。例えば、特許文献１には、言い換えコーパス（corpus）から機械学習によって文の言い換えモデルを獲得する文の言い換え表現作成方法の一例が記載されている。特許文献１に記載されている文の言い換え表現作成方法では、複数の、辞書や新聞等の記事における同じ見出しに対する説明文を、記述内容が一致する異なった文書対であるとみなす。すなわち、同じ見出しに対する説明文を、言い換えの関係にある文書対であるとみなして言い換えコーパスを作成し、それに対して機械学習を実行する。
【０００５】
【非特許文献１】藤田　篤、他１名、「A Class-oriented Approach to Building a Paraphrase Corpus 」, Proceedings of the 3rd International Workshop on Paraphrasing (IWP) , ２００５年１０月, ｐ. ２５－３２
【特許文献１】特開２００２－０１４９９０号公報（段落００５８－００５９、図４）
【発明の開示】
【発明が解決しようとする課題】
【０００６】
上記非特許文献１及び特許文献１の全開示内容はその引用をもって本書に繰込み記載されているものとする。
【０００７】
しかし、従来の文の言い換え表現作成方法を、多数のユーザが入力する質問文に対して適用する場合、言い換え規則を人手で作成する方法（例えば、非特許文献１に記載されている文の言い換え表現作成方法。）では、膨大な量の規則作成を行う必要がある。多様な質問文の内容や表現に対して、十分な網羅性を実現しなければならないからである。さらに、最終的な言い換え表現を得る際に人手による確認作業を必要とするため、ユーザによる質問文入力をリアルタイムで処理するシステムには適用できない場合が多い。
【０００８】
また、特許文献１に記載されている文の言い換え表現作成方法では、既存の辞書の見出しや新聞等の記事で網羅されない分野に関しては、学習データの入手が困難である。さらに、学習データの中に、表現上の言い換えの関係にない文対が含まれていると、学習の際のノイズ（不必要な情報）となる。その場合、学習データの品質によっては適切な文の言い換えモデル（言い換え規則。以下、言い換えモデルという。）が得られない場合がある。すなわち、特許文献１に記載されているような既存の辞書や新聞等の記事データから機械学習する方法では、網羅される質問文が既存の辞書や新聞等の記事データに関連する内容に限定されるため、ユーザの多様な質問文に対して十分な網羅性を実現することができない場合が多い。なお、文対とは、２つの文（センテンス）の組み合わせ（２文１組）をいう。
【０００９】
そこで、本発明は、ユーザによる質問文入力に対して、リアルタイムで言い換え表現を

生成する文言い換え方法、変換規則算出方法、文言い換えシステム、変換規則算出用プログラムおよび文言い換え用プログラムを提供することを目的とし、特に、ユーザにより入力された質問文の言い換えモデルを学習できる文言い換え方法、変換規則算出方法、文言い換えシステム、変換規則算出用プログラムおよび文言い換え用プログラム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】

本発明の第１の視点によれば、ユーザによって入力された質問文書に対して、同じ内容の異なる表現の質問文書を生成する文言い換え方法であって、前記質問文書の入力を受け付けたコンピュータが、記憶手段に格納されている同一の回答文書を有する質問文書のなかから、設定された閾値以上の類似度を有するセンテンスの組み合わせを抽出する第１のセンテンス抽出ステップと、第１のセンテンス抽出ステップで抽出したセンテンスの組み合わせから、同じ内容の異なる表現のセンテンスに変換するための変換規則を算出する変換規則算出ステップと、変換規則算出ステップで算出した変換規則を用いて、閾値以上の類似度を有するセンテンスの組み合わせを記憶手段から再度抽出する第２のセンテンス抽出ステップと、第１および第２のセンテンス抽出ステップで抽出したセンテンスの組み合わせがすべて一致しているかどうかを判定する判定ステップと、変換規則算出ステップで算出した変換規則を用いて、入力された質問文書の言い換え文書を生成する文書生成ステップとを含み、前記コンピュータが、第１および第２のセンテンス抽出ステップで抽出したセンテンスの組み合わせがすべて一致していると判定されるまで、変換規則算出ステップと、第２のセンテンス抽出ステップとを繰り返し実行する文言い換え方法が提供される。

【００１１】

上記文言い換え方法において、生成した言い換え文書をユーザに対して表示するステップを含み、生成した言い換え文書を、入力された質問文に対する類似度を示す値とともに表示することが好ましい。

【００１２】

また、上記文言い換え方法における文書生成ステップに、入力された質問文書を単語列に分解するステップと、変換規則算出ステップで算出した変換規則と単語列とを比較し、類似度の高い単語列を複数選択する選択ステップと、選択ステップで選択した単語列を並び替える並び換えステップとを含めることができる。

【００１３】

本発明の第２の視点によれば、ユーザによって入力された質問文書に対して、同じ内容の異なる表現の質問文書を生成するための変換規則を算出する変換規則算出方法であって、前記質問文書の入力を受け付けたコンピュータが、記憶手段に格納されている同一の回答文書を有する質問文書のなかから、設定された閾値以上の類似度を有するセンテンスの組み合わせを抽出する第１のセンテンス抽出ステップと、第１のセンテンス抽出ステップで抽出したセンテンスの組み合わせから、同じ内容の異なる表現のセンテンスに変換するための変換規則を算出する変換規則算出ステップと、変換規則算出ステップで算出した変換規則を用いて、閾値以上の類似度を有するセンテンスの組み合わせを記憶手段から再度抽出する第２のセンテンス抽出ステップと、第１および第２のセンテンス抽出ステップで抽出したセンテンスの組み合わせがすべて一致しているかどうかを判定する判定ステップとを含み、前記コンピュータが、第１および第２のセンテンス抽出ステップで抽出したセンテンスの組み合わせがすべて一致していると判定されるまで、変換規則算出ステップと、第２のセンテンス抽出ステップとを繰り返し実行する変換規則算出方法が提供される。

【００１４】

上記変換規則算出方法において、記憶手段に、入力手段から入力した質問文書と、質問文書に対する回答文書と、質問文書および回答文書における対応関係を示した情報とを記憶させ、同じ内容の回答文書をまとめた１つのグループに対して、情報を適用し、同じ内容の回答文書に対応した質問文書のグループを生成することができる。

【００１５】
上記変換規則算出方法に、記憶手段に格納されている情報を更新するステップを含ませることができる。この場合、前記記憶手段の情報が更新されたことを契機に、第１のセンテンス抽出ステップにおける処理を開始してもよい。
【００１６】
本発明の第３の視点によれば、ユーザによって入力された質問文書に対して、同じ内容の異なる表現の質問文書を生成する文言い換えシステムであって、入力手段と、入力手段に入力された質問文書と、質問文書に対する回答文書とを記憶する記憶手段と、記憶手段に格納されている同一の回答文書を有する質問文書のなかから、設定された閾値以上の類似度を有するセンテンスの組み合わせを抽出するセンテンス抽出手段と、センテンス抽出手段で抽出したセンテンスの組み合わせから、同じ内容の異なる表現のセンテンスに変換するための変換規則を算出する変換規則算出手段と、変換規則算出手段が算出した変換規則を用いて、ユーザによって入力手段に入力された質問文書の言い換え文書を生成する文書生成手段とを備える文言い換えシステムが提供される。
【００１７】
上記文言い換えシステムにおける記憶手段は、入力手段に入力された質問文書と、質問文書に対する回答文書とを記憶する第１の記憶手段と、センテンス抽出手段が抽出したセンテンスの組み合わせを記憶する第２の記憶手段と、変換規則算出手段が算出した変換規則を記憶する第３の記憶手段とを含むことが好ましい。
【００１８】
また、上記文言い換えシステムにおける変換規則は、少なくとも翻訳モデルと、言語モデルと、単語列とを含むものとすることができる。
【００１９】
また、上記文言い換えシステムに、文書生成手段が生成した言い換え文書を表示する表示手段を備えることができる。この表示手段に、文書生成手段が生成した言い換え文書を、ユーザによって入力手段に入力された質問文との類似度を示す値とともに表示してもよい。
【００２０】
また、上記文言い換えシステムに、入力手段に入力されたコマンドにしたがって、記憶手段における情報を更新する情報更新手段を備えることができる。この情報更新手段は、記憶手段における情報を更新した場合に、センテンスの組み合わせの抽出を開始するための信号をセンテンス抽出手段に送信するよう構成することができる。
【００２１】
また、上記文言い換えシステムは、過去にユーザが入力した質問文書と、各質問文書に対する回答文書との対応関係から、同一の回答文書に対応する質問文書集合を作成し、文の言い換えの学習データとして利用する文言い換えシステムとすることができる。同一の回答文書に対応する質問文書であっても、それぞれの文書を作成したユーザの分野における知識や問題の理解レベル、意図や戦略の違いなどによって、必ずしも文書同士が言い換えの関係にある保証はないため、この文言い換えシステムは、質問文集合のすべての文対の中から、言い換えの関係にあるものだけを篩い分けるために、言い換えモデル（例えば、後述する式（１）等）によって言い換えの関係にある文対の抽出を行う。そして、この抽出された言い換え文対（例えば、図８に示されている文対）を学習データとしてさらに言い換えモデルの学習を行う。この言い換え文対における、言い換えモデルによる言い換え文対の抽出と、抽出された言い換え文対における言い換えモデルの学習との２つの過程を、抽出文セット（抽出された言い換え文対）が収束する（新たに学習した言い換えモデルを用いて抽出した言い換え文対が、それ以前に学習した言い換えモデルを用いて抽出した言い換え文対と同一になる）まで交互に繰り返し、最終的な言い換えモデルを得る。上記のような逐次処理によって、同一の回答文書に対応する質問文書集合の中から、言い換え文の関係にある文対を抽出することができるため、これを学習データとして適切な言い換えモデルを学習することができる。

【００２２】
上記のような学習を行うことにより、上記質問文言い換えシステムは、例えばメールで受信した質問文書に対する回答文書を自動的に生成することが可能になる。すなわち、顧客がさまざまな表現で問い合わせを行っていても、その表現に対して、さまざまな言い換えを生成して照合することができる。したがって、質問文書と回答文書の対をあらかじめ持っていれば、問い合わせの表現の違いによらずに、自動的に回答文書を割り振ることができる。
【００２３】
また、上記した文言い換え方法は、例えば応対記録の自動生成に用いることができる。すなわち、回答文書（回答のメールや、オペレータが電話で回答した音声を音声認識等によってテキスト化したもの）と、その応対記録（オペレータが対応内容を要約して記録するもの）とが文言い換えシステムに蓄積されている場合、その回答文書から応対記録への言い換えのモデルを学習することによって、応対記録の自動生成を実現できる。
【００２４】
本発明の第４の視点によれば、ユーザによって入力された質問文書に対して、同じ内容の異なる表現の質問文書を生成するための変換規則を算出する変換規則算出プログラムであって、コンピュータに、記憶手段に格納されている同一の回答文書を有する質問文書のなかから、設定された閾値以上の類似度を有するセンテンスの組み合わせを抽出する第１のセンテンス抽出処理と、第１のセンテンス抽出ステップで抽出したセンテンスの組み合わせから、同じ内容の異なる表現のセンテンスに変換するための変換規則を算出する変換規則算出処理と、変換規則算出処理で算出した変換規則を用いて、閾値以上の類似度を有するセンテンスの組み合わせを記憶手段から再度抽出する第２のセンテンス抽出処理と、第１および第２のセンテンス抽出処理で抽出したセンテンスの組み合わせがすべて一致しているかどうかを判定する判定処理とを実行させ、第１および第２のセンテンス抽出処理で抽出したセンテンスの組み合わせがすべて一致していると判定されるまで、第１のセンテンス抽出処理と、変換規則算出処理と、第２のセンテンス抽出処理とを繰り返させる変換規則算出用プログラムが提供される。
【００２５】
上記変換規則算出用プログラムは、コンピュータに、入力手段から入力された質問文書と、質問文書に対する回答文書と、質問文書および回答文書における対応関係を示した情報とを記憶した記憶手段において、同じ内容の回答文書をまとめた１つのグループに対して、情報を適用し、同じ内容の回答文書に対応した質問文書のグループを生成する処理を実行させるものとすることができる。
【００２６】
また、上記変換規則算出用プログラムは、コンピュータに、記憶手段に格納されている情報を更新する処理と、記憶手段の情報が更新されたことを契機に、第１のセンテンス抽出処理を開始する処理とを実行させるものであってもよい。
【００２７】
本発明の第５の視点によれば、ユーザによって入力された質問文書に対して、同じ内容の異なる表現の質問文書を生成する文言い換え用プログラムであって、コンピュータに、記憶手段に格納されている同一の回答文書を有する質問文書のなかから、設定された閾値以上の類似度を有するセンテンスの組み合わせを抽出する第１のセンテンス抽出処理と、第１のセンテンス抽出処理で抽出したセンテンスの組み合わせから、同じ内容の異なる表現のセンテンスに変換するための変換規則を算出する変換規則算出処理と、変換規則算出処理で算出した変換規則を用いて、閾値以上の類似度を有するセンテンスの組み合わせを記憶手段から再度抽出する第２のセンテンス抽出処理と、第１および第２のセンテンス抽出処理で抽出したセンテンスの組み合わせがすべて一致しているかどうかを判定する判定処理と、第１および第２のセンテンス抽出処理で抽出したセンテンスの組み合わせがすべて一致していると判定されるまで、第１のセンテンス抽出処理と、変換規則算出処理と、第２のセンテンス抽出処理とを繰り返す処理と、変換規則算出処理で算出した変換規則を

用いて、入力された質問文書の言い換え文書を生成する文書生成処理とを実行させる文言い換え用プログラムが提供される。

【００２８】

上記文言い換え用プログラムは、コンピュータに、入力された質問文に対する類似度を示す値とともに、生成した言い換え文書を表示する処理を実行させるものとすることができる。

【００２９】

また、上記文言い換え用プログラムは、コンピュータに、文書生成処理で、入力された質問文書を単語列に分解する処理と、変換規則算出処理で算出した変換規則と単語列とを比較し、類似度の高い単語列を複数選択する選択処理と、選択処理で選択した単語列を並び替える並び換え処理とを実行させるものとすることができる。

【発明の効果】

【００３０】

本発明によれば、過去にユーザが入力した質問文書と、各質問文書に対する回答文書との対応関係から質問文言い換えモデルを学習することができる。その理由は、言い換え文の関係にない質問文対を段階的に除去しながら、同時に言い換えモデルを学習して、言い換えモデルの精度を高めていく逐次的な処理過程を導入したからである。そうすることにより、同一の回答文書に対応する質問文書集合の中から、言い換え文の関係にある文対を精度よく抽出することができる。

【図面の簡単な説明】

【００３１】

【図１】本発明による文言い換えシステムの一例を示したシステム構成図である。

【図２】文言い換えシステムの動作を説明するためのフローチャートである。

【図３】文言い換えシステムの動作を説明するためのフローチャートである。

【図４】文書集合記憶手段に記憶されている質問文集合の一例を示した説明図である。

【図５】図４に示されている各質問文書に対する回答文書の一例を示した説明図である。

【図６】図４における各質問文書と図５における回答文書との対応関係を示した説明図である。

【図７】図６の対応関係から作成される、同一の回答文書に対応する質問文書集合の関係を示した説明図である。

【図８】文対抽出手段が抽出した言い換え文対の一例を示した説明図である。

【図９】図８に示した言い換え文対に対応した各質問文のテキストの一例を示した説明図である。

【図１０】モデル学習手段が言い換え文対（Q1_S1，Q1_S2）に対して実行したアライメントの結果を示す説明図である。

【図１１】モデル学習手段が言い換え文対（Q1_S1，Q10_S2 ）に対して実行したアライメントの結果を示す説明図である。

【図１２】モデル学習手段が言い換え文対（Q2_S1，Q6_S1）に対して実行したアライメントの結果を示す説明図である。

【図１３】モデル学習手段が言い換え文対（Q2_S1，Q9_S2）に対して実行したアライメントの結果を示す説明図である。

【図１４】モデル学習手段が言い換え文対（Q2_S1，Q10_S2 ）に対して実行したアライメントの結果を示す説明図である。

【図１５】モデル学習手段が言い換え文対（Q2_S1，Q12_S2 ）に対して実行したアライメントの結果を示す説明図である。

【図１６】文対抽出手段が抽出した言い換え文対の一例を示した説明図である。

【図１７】新たに追加された文対（Q4_S1，Q8_S1）に対応した質問文のテキストの一例を示した説明図である。

【図１８】モデル学習手段が言い換え文対（Q4_S1，Q8_S1）に対して実行したアライメントの結果を示す説明図である。

【図１９】入力手段に入力された質問文の一例を示した説明図である。
【図２０】候補生成手段が言い換えモデルを用いて生成した言い換え候補の一例を示した説明図である。
【符号の説明】
【００３２】
１００　入力手段
２００　出力手段
３００　データ処理装置
３０１　文対抽出手段
３０２　モデル学習手段
３０３　文書登録手段
３０４　候補生成手段
４００　記憶手段
４０１　文書集合記憶手段
４０２　モデル記憶手段
４０３　文対記憶手段
【発明を実施するための最良の形態】
【００３３】
以下、本発明の実施の形態を図面を参照して説明する。図１は、本発明による文言い換えシステムの一例を示したシステム構成図である。文言い換えシステムは、入力手段１００と、出力手段２００と、プログラム制御により動作するデータ処理装置３００と、記憶手段４００とを備えている。なお、本実施の形態における文言い換えシステムは、入力手段１００に入力された質問文の言い換え候補を、記憶手段４００に記憶されている質問文言い換えモデルを用いてリアルタイムに生成する。以下、入力手段１００に入力された文章が質問文である場合について説明する。
【００３４】
入力手段１００は、例えば文章等を入力するためのキーボードやマウス等によって実現される。入力手段１００は、ユーザの操作に応じた入力信号をデータ処理装置３００に出力する機能を有する。
【００３５】
出力手段２００は、液晶表示装置等によって実現される。出力手段２００には、例えばデータ処理装置３００が生成した質問文の言い換え候補が表示される。
【００３６】
データ処理装置３００は、文対抽出手段３０１と、モデル学習手段３０２と、文書登録手段３０３と、候補生成手段３０４とを含む。データ処理装置３００は、例えばＣＰＵと図示しない周辺回路とで実現される。データ処理装置３００は、プログラム制御によって動作し、入力手段１００に入力された質問文の言い換え候補を生成する。
【００３７】
文対抽出手段３０１は、後述する記憶手段４００に記憶されている質問文書の集合におけるすべての文対に対して、各々の文対が言い換え文対の関係にあるかどうかを判定する機能を有する。具体的には、文対抽出手段３０１は、文書集合記憶手段４０１に記憶されている質問文書集合におけるすべての文対に対して、モデル記憶手段４０２に記憶されている質問文言い換えモデルを用いて言い換え文対であるかどうかを判定する。なお、本実施の形態における言い換え文対とは、言い換えの関係にある（すなわち、内容が同一である）センテンスの組み合わせ（２文１組）のことをいう。例えば、質問文が複数のセンテンスを含む場合には、その質問文における２つのセンテンスが言い換え文対と判定される場合があってもよい。文対抽出手段３０１は、言い換え文対と判定した文対を出力し、文対記憶手段４０３に記憶する。
【００３８】
モデル学習手段３０２は、文対記憶手段４０３に記憶されている言い換え文対を学習デ

ータとして利用し、質問文言い換えモデルを学習する機能を有する。モデル学習手段３０２は、学習した質問文言い換えモデルをモデル記憶手段４０２に記憶する。なお、本実施の形態における学習とは、従来有している質問文言い換えモデルと異なる、新たな質問文言い換えモデルを取得するまでの処理全体を指す。

【００３９】

また、本実施の形態では、質問文言い換えモデルは、各処理工程毎に最適な形式で適用される。例えば、文対抽出手段３０１は、言い換え文対を抽出する際に、質問文言い換えモデルを言い換え文対であるかどうかを判定するための翻訳確率等として使用する。また、例えばモデル学習手段３０２は、抽出された言い換え文対から質問文言語モデル等を学習する。本実施の形態における質問文言い換えモデルには、翻訳確率や質問文言い換え翻訳モデル、質問文言語モデル、および単語列等が含まれる。

【００４０】

文書登録手段３０３は、入力手段１００から入力されたコマンドにしたがって文書集合記憶手段４０１の情報を更新する機能を有する。具体的には、文書登録手段３０３は、入力手段１００に入力された、同一の回答文書を有する質問文書集合を文書集合記憶手段４０１に記憶する。また、文書登録手段３０３は、入力手段１００に入力されたコマンドにしたがって、文書集合記憶手段４０１における質問文書集合を削除する機能を有する。なお、文書登録手段３０３が文対抽出手段３０１に出力した信号を契機として、後述する言い換え文対の抽出処理が開始される。

【００４１】

候補生成手段３０４は、入力手段１００から入力された質問文に対し、言い換え候補を生成して出力手段２００に出力する機能を有する。具体的には、候補生成手段３０４は、モデル記憶手段４０２に記憶された質問文言い換えモデルを適用し、入力された質問文に対する言い換え候補をリアルタイムに生成する機能を有する。そして、候補生成手段３０４は、生成した言い換え候補を出力手段２００に出力する。

【００４２】

記憶手段４００は、文書集合記憶手段４０１と、モデル記憶手段４０２と、文対記憶手段４０３とを含む。記憶手段４００は、例えばＲＡＭ等の記憶媒体によって実現され、データ処理装置３００が実行する制御プログラム等の各種のデータを格納する。なお、記憶手段４００の一部がＲＯＭによって実現されていてもよく、ＲＯＭ部分に制御プログラムが格納されていてもよい。

【００４３】

記憶手段４００は、ユーザにより入力された質問文やその質問文に対する回答文、質問文および回答文の対応関係、および質問文言い換えモデル等を記憶する機能を有している。

【００４４】

文書集合記憶手段４０１は、入力手段１００から入力された質問文書の集合と、その各質問文書に対する回答文書とから作成された、同一の回答文書に対応する質問文書の集合等を記憶する。

【００４５】

モデル記憶手段４０２は、質問文言い換えモデルを記憶する機能を有する。具体的には、モデル記憶手段４０２は、モデル学習手段３０２が学習した、質問文言語モデルや単語列等の新たな質問文言い換えモデルを記憶する。モデル記憶手段４０２が記憶している質問文言い換えモデルは、文対抽出手段３０１や候補生成手段３０４によって読み出される。

【００４６】

文対記憶手段４０３は、文対抽出手段３０１が言い換え文対であると判定した文対（センテンスの組み合わせ）やそれぞれの文対のテキストデータ等を記憶する機能を有する。文対記憶手段４０３が記憶している言い換え文対は、文対抽出手段３０１およびモデル学習手段３０２によって読み出される。

【００４７】
なお、本発明による文言い換えシステムでは、後述するように、文対抽出手段３０１と、モデル学習手段３０２と、文書登録手段３０３と、文書集合記憶手段４０１と、モデル学習手段４０２と、文対記憶手段４０３とによって、質問文言い換えモデルの学習が実現される。また、本発明による文言い換えシステムでは、入力手段１００と、出力手段２００と、候補生成手段３０４と、モデル記憶手段４０２とによって、質問文言い換えモデルを用いた言い換え候補の生成が実現される。
【００４８】
次に、本実施の形態における文言い換えシステムの動作を図面を参照して説明する。図２は、本発明による文言い換えシステムの動作を説明するためのフローチャートである。本実施の形態における文言い換えシステムでは、候補生成手段３０４がモデル記憶手段４０２から質問文言い換えモデルを読み出し、入力された質問文の言い換え候補をリアルタイムに生成する。なお、本実施の形態では、ユーザが質問文を入力する際には、質問文言い換えモデルがモデル記憶手段４０２に既に存在しているものとする。以下、文言い換えシステムによる質問文言い換えモデルの学習方法を説明する。
【００４９】
ユーザによって入力手段１００にコマンドが入力されると、文書登録手段３０３は、文書集合記憶手段４０１に含まれる情報を更新する（ステップＳ１０１）。例えば登録コマンドとともに質問文書集合が入力されると、文書登録手段３０３は、入力された質問文書集合を文書集合記憶手段４０１に記録する。また、例えば文書集合記憶手段４０１における質問文書を選択して削除コマンドが入力されると、文書登録手段３０３は、選択された質問文書を文書集合記憶手段４０１から削除する。文書登録手段３０３は、文書集合記憶手段４０１における情報が更新されたことを契機に、言い換え文対の抽出処理を開始するための信号（トリガ）を文対抽出手段３０１に送信する。
【００５０】
文対抽出手段３０１は、文書集合記憶手段４０１に記憶されている質問文書集合におけるすべての文対に対して、それぞれの文対が言い換え文対の関係にあるかどうかを判定する。その際、文対抽出手段３０１は、モデル記憶手段４０２に記憶されている質問文言い換えモデルを用いて言い換え文対であるかどうかを判定する。文対抽出手段３０１は、言い換え文対と判定した文対を出力し、文対記憶手段４０３に記憶する（ステップＳ１０２）。
【００５１】
次に、モデル学習手段３０２は、文対記憶手段４０３に記憶されている言い換え文対を学習データとして利用し、質問文言い換えモデルを学習する（ステップＳ１０３）。モデル学習手段３０２は、学習した質問文言い換えモデルをモデル記憶手段４０２に記憶する。なお、質問文言い換えモデルとは、翻訳確率や質問文言い換え翻訳モデル、質問文言語モデル、単語列等をいう。
【００５２】
文対抽出手段３０１は、文書集合記憶手段４０１に記憶されている質問文書集合におけるすべての文対に対して、それぞれの文対が言い換え文対の関係にあるかどうかを再度判定する。具体的には、文対抽出手段３０１は、文書集合記憶手段４０１におけるすべての文対に対して、モデル学習手段３０２が学習した質問文言い換えモデルを用いて言い換え文対であるかどうかを判定する。その場合、同一の質問文書から言い換え文対が抽出されてもよい。文対抽出手段３０１は、言い換え文対と判断した文対を出力し、文対記憶手段４０３に再度記憶する（ステップＳ１０４）。
【００５３】
文対抽出手段３０１は、抽出した言い換え文対が収束したかどうかを判定する（ステップＳ１０５）。抽出した言い換え文対が収束していない場合、文対抽出手段３０１およびモデル学習手段３０２は、上述したステップＳ１０３、ステップＳ１０４における処理を、抽出される言い換え文対が収束するまで繰り返し実行する。なお、ステップＳ１０２の

処理、第１巡目のステップＳ１０４の処理は、それぞれ、第１のセンテンス抽出ステップ、第２のセンテンス抽出ステップに相当する。さらに、第ｎ巡目のステップＳ１０４の処理、第（ｎ＋１）巡目のステップＳ１０４の処理は、それぞれ、第１のセンテンス抽出ステップ、第２のセンテンス抽出ステップに相当する。ｎは１以上の整数である。また、言い換え文対が収束するとは、文対抽出手段３０１が最新の質問文言い換えモデルを用いて抽出した言い換え文対が、それ以前の質問文言い換えモデルを用いて抽出した言い換え文対とすべて同一となった場合のことをいう。

【００５４】

抽出した言い換え文対が収束している場合（ステップＳ１０５：ＹＥＳ）、文言い換えシステムは、質問文言い換えモデルの生成処理を終了する。

【００５５】

次に、入力手段１００に入力された質問文の言い換え候補の生成方法を説明する。図３は、本発明による文言い換えシステムの動作を説明するためのフローチャートである。

【００５６】

入力手段１００に質問文が入力されると（ステップ２０１）、候補生成手段３０４は、モデル記憶手段４０２から質問文言い換えモデルを読み出す（ステップＳ２０２）。候補生成手段３０４は、入力手段１００から入力された質問文に対して、モデル記憶手段４０２に記憶されている質問文言い換えモデルを適用し、入力された質問文に対する言い換え候補をリアルタイムに生成する。

【００５７】

候補生成手段３０４は、入力手段１００に入力された質問文を単語単位に分解し、モデル記憶手段４０２から読み出した単語列と比較する（ステップＳ２０３）。そして、後述する単語の言い換え確率が大きい単語列を複数選択する（ステップＳ２０４）。

【００５８】

次に、候補生成手段３０４は、選択した単語列を並び替え、後述する並び替え確率を求める（ステップＳ２０５）。候補生成手段３０４は、並び替え確率の大きい文章を複数選択し、質問文言い換え候補として出力手段２００に出力する（ステップＳ２０６）。なお、候補生成手段３０４が出力する言い換え候補は、１つであってもよい。

【００５９】

次に、上述した動作を図面を参照してより具体的に説明する。

【００６０】

図４は、文書集合記憶手段４０１に記憶されている質問文書集合の一例を示した説明図である。また、図５は、図４に示されている各質問文書に対する回答文書の一例を示した説明図である。さらに、図６は、図４における各質問文書と図５における回答文書との対応関係を示した説明図である。

【００６１】

図６に示されている対応関係では、例えば図４における質問文書Ｑ１に対し、図５における回答文書Ａ１によって回答されることを示している。また、図６に示されている対応関係では、例えば図４における質問文書Ｑ４に対しては、図５における回答文書Ａ１とＡ２との２つの文書をつないだ文書で回答されることを示している。なお、図６において、質問文書に対して複数の回答文書が対応付けられている場合、その質問文書に対する回答には複数の回答文書が含まれることを示している。例えば、上述した質問文書Ｑ４に対する回答文書の場合、回答文書にはＡ１とＡ２のいずれか一方の回答文書が含まれるという意味ではなく、Ａ１とＡ２の両方の回答文書が含まれることを示している。

【００６２】

図７は、図６の対応関係から作成される、同一の回答文書に対応する質問文書集合の関係を示した説明図である。文書集合記憶手段４０１は、図７に示すような質問文書集合の関係と、図４に示すような質問文書（テキストデータ）とを記憶している。

【００６３】

文書集合記憶手段４０１に記憶されている、同一の回答文書に対応した質問文書集合に

おけるすべての文対に対して、文対抽出手段３０１は、モデル記憶手段４０２に記憶されている質問文言い換えモデルを用いて言い換え文対であるかどうかを判定する。例えば、図７において、文対抽出手段３０１は、回答文書Ａ１に対応した質問文書Ｑ１およびＱ２に含まれるすべての文対に対して、それぞれのセンテンスが言い換え文対の関係にあるかどうかを判定する。同様に、文対抽出手段３０１は、回答文書Ａ１に対応した質問文書Ｑ１およびＱ３にふくまれるすべてのセンテンスについても、それぞれが言い換え文対の関係にあるかどうかを判定する。なお、１つの質問文書に複数のセンテンスが含まれる場合（例えば、質問文書Ｑ１が３つのセンテンスから構成されている場合）、文対抽出手段３０１は、同一の質問文書におけるそれぞれのセンテンスに対しても、言い換え文対であるかどうかを判定する。

【００６４】

次に、文対抽出手段３０１による言い換え文対の判定方法を説明する。ここで、共通の回答文書Ａを持つ言い換え文対、質問文１（センテンス１）および質問文２（センテンス２）があるとき、文対抽出手段３０１は、質問文１の単語列$W_1 \cdots W_k$から質問文２の単語列$W''_1 \cdots W''_m$への翻訳確率を、例えば次のモデル$P_n = {}_1$を用いて推定する。そして、その値（翻訳確率）が閾値より高い場合に、質問文１と質問文２とが言い換え文対の関係にあると判定する。

【００６５】

【数１】

$$P_n(W''_1 \cdots W''_m \mid W_1 \cdots W_k, A)$$
$$= \sum_{w'_1 \cdots w'_l} \prod_i P_n(W'_j \mid W_i, A) * P_n(W''_1 \cdots W''_m \mid W'_1 \cdots W'_l, A)$$
$$\geq P_0 \quad \cdots\cdots (1)$$

【００６６】

なお、式（１）におけるＰθは閾値を表している。また、式（１）における単語列W_iに関する積「$\Pi_i P_n$（Ｗ’$_j$｜W_i，Ａ）」は、質問文言い換え翻訳モデルに相当する。

【００６７】

初回（ｎ＝１）では、単語の言い換え確率P_n（Ｗ’$_j$｜W_i，Ａ）、および単語の並べ替え確率P_n（Ｗ”$_1$・・・Ｗ”$_m$｜Ｗ’$_1$・・・Ｗ’$_l$，Ａ）が未学習であるため、

【００６８】

【数２】

$$P_n(W'_j \mid W_i, A) = (1-\varepsilon)\delta_{w'_j, w_i}\alpha + \varepsilon(1-\delta_{w'_j, w_i}) \quad \cdots\cdots (2)$$
$$P_n(W''_1 \cdots W''_m \mid W'_1 \cdots W'_l, A) = \delta_{m,l}\delta_{\{w''_1, \cdots, w''_m\}, l\{w'_1, \cdots, w'_l\}} \quad \cdots (3)$$

【００６９】

などのようなアドホック（Ad Hoc）な初期翻訳モデルを用いる。すなわち、文対抽出手段３０１は、最初に文書集合記憶手段４０１から言い換え文対を抽出する際に、式（２）、（３）によって表される質問文言い換えモデルを使用する。

【００７０】

上記のような質問文言い換えモデルを用いて判定することにより、文対抽出手段３０１は、ステップＳ１０２の処理によって図８および図９に示すような言い換え文対および質問文（センテンス）を得る。なお、図８は、文対抽出手段３０１が抽出した言い換え文対の一例を示した説明図である。また、図９は、図８に示した言い換え文対に対応した各質問文のテキストの一例を示した説明図である。文対抽出手段３０１は、抽出した言い換え

文対と、その言い換え文対に対応した各質問文のテキストとを文対記憶手段４０３に記憶する（ステップＳ１０２）。

【００７１】

次に、モデル学習手段３０２は、文対記憶手段４０３に記憶されている言い換え文対（図８および図９に示されているような言い換え文対および質問文）を学習データとして利用し、質問文言い換えモデル$P_{n=2}$を学習する。

【００７２】

その際、モデル学習手段３０２は、単語の言い換え確率$P_n(W'_j|W_i, A)$（すなわち、質問文言い換えモデル）を求めるために、言い換え文対の単語に関する対応関係の推定（アライメント）を行う。具体的には、モデル学習手段３０２は、図８に示された言い換え文対の関係を用いて、図９に示された質問文における単語の対応関係を推定する。

【００７３】

図１０は、モデル学習手段３０２が言い換え文対（Q1_S1，Q1_S2）に対して実行した推定の結果を示す説明図である。同様に、図１１は言い換え文対（Q1_S1，Q10_S2 ）、図１２は言い換え文対（Q2_S1，Q6_S1）、図１３は言い換え文対（Q2_S1，Q9_S2）、図１４は言い換え文対（Q2_S1，Q10_S2 ）、および、図１５は言い換え文対（Q2_S1，Q12_S2 ）に対して実行した推定の結果を示す説明図である。なお、図１０～１５における符号「φ」は、対応する単語が存在しないことを示している。

【００７４】

モデル学習手段３０２は、図１０～１５に示す言い換え文対における、対応した単語間の言い換え回数にもとづいて、単語の言い換え確率$P_n(W'_j|W_i, A)$を推定する。また、モデル学習手段３０２は、単語の並べ替え確率$Pn(W''_1 \cdots W''_m|W'_1 \cdots W'_l, A)$を推定する。例えば、モデル学習手段３０２は、単語の置き換えを行った単語列$W'_1 \cdots W'_l$から、言い換え文における単語列$W''_1 \cdots W''_m$への並び替えを、言い換え後の単語列の出現確率で代替する。

【００７５】

【数３】

$$P_n(W''_1 \cdots W''_m \mid W'_1 \cdots W'_l, A) = \delta_{m,l}\,\delta_{\{W''_1,\cdots,W''_m\},\{W'_1,\cdots,W'_l\}}\,P_{N\text{-}gram}(W''_1 \cdots W''_m) \quad \cdots\cdots(4)$$

【００７６】

なお、式（４）における単語列の出現確率$P_{N-gram}(W''_1 \cdots W''_m)$は、質問文言語モデルに相当する。モデル学習手段３０２は、式（４）に示すような質問文言語モデル等を質問文言い換えモデルとして学習し、モデル記憶手段４０２に記憶する（ステップＳ１０３）。

【００７７】

文対抽出手段３０１は、新たに学習した質問文言い換えモデル$P_{n=2}$を用いて、文書集合記憶手段４０１から再度言い換え文対を抽出する（ステップＳ１０４）。そして、文対抽出手段３０１は、言い換え文対と判定した文対を文対記憶手段４０３に記録する。

【００７８】

図１６は、文対抽出手段３０１が質問文言い換えモデル$P_{n=2}$を用いて抽出した言い換え文対の一例を示した説明図である。例えば、図１６における言い換え文対と図８における言い換え文対とを比較すると、図１６に示す言い換え文対には、文対（Q4_S1，Q8_S1 ）が新たな言い換え文対として追加されている。文対抽出手段３０１は、図１６に示されている言い換え文対を文対記憶手段４０３に記録する。

【００７９】

なお、図１７は、図１６における、新たに追加された言い換え文対（Q4_S1，Q8_S1）に

対応した質問文のテキストの一例を示した説明図である。また、図１８は、モデル学習手段３０２が言い換え文対（Q4_S1，Q8_S1）に対して実行した推定の結果を示す説明図である。

【００８０】

文対記憶手段４０３に文対（Q4_S1，Q8_S1）が新たな言い換え文対として追加されたため、文対抽出手段３０１は、言い換え文対が収束していないと判定する（ステップＳ１０５：ＮＯ）。なお、文対抽出手段３０１は、質問文言い換えモデル$P_n = {}_2$を用いて抽出した言い換え文対を文対記憶手段４０３に記録する際に、質問文言い換えモデル$P_n = {}_1$を用いて抽出した言い換え文対を文対記憶手段４０３から読み出し、言い換え文対が収束したかどうかを判定する。

【００８１】

モデル学習手段３０２および文対抽出手段３０１は、再度、ステップＳ１０３およびステップＳ１０４における処理を実行する。そうすることによって、モデル学習手段３０２は、新たな質問文言い換えモデル$P_n = {}_3$を得る（ステップＳ１０３）。

【００８２】

文対抽出手段３０１は、質問文言い換えモデル$P_n = {}_3$を用いて文書集合記憶手段４０１から再度言い換え文対を抽出する（ステップＳ１０４）。そして、文対抽出手段３０１は、抽出した言い換え文対が収束したかどうかを判定する（ステップＳ１０５）。質問文言い換えモデル$P_n = {}_3$を用いて抽出した言い換え文対が前回抽出した言い換え文対（質問文言い換えモデル$P_n = {}_2$を用いて抽出した言い換え文対）と変わらない場合、文対抽出手段３０１は、言い換え文対が収束したと判定する（ステップＳ１０５：ＹＥＳ）。質問文言い換え文対が収束したと判定されると、文言い換えシステムは、質問文言い換えモデルの生成を終了する。

【００８３】

そして、入力手段１００に質問文が入力されると（ステップＳ２０１）、候補生成手段３０４は、モデル記憶手段４０２から質問文言い換えモデルを読み出す（ステップＳ２０２）。候補生成手段３０４は、入力手段１００から入力された質問文に対して、モデル記憶手段４０２に記憶されている質問文言い換えモデルを適用し、入力された質問文に対する言い換え候補をリアルタイムに生成する。

【００８４】

候補生成手段３０４は、入力手段１００に入力された質問文を単語単位に分解し、モデル記憶手段４０２から読み出した単語列と比較する（ステップＳ２０３）。そして、単語の言い換え確率が大きい単語列を複数選択する（ステップＳ２０４）。

【００８５】

候補生成手段３０４は、選択した単語列の並び替えを実行し、単語の並び替え確率を求める（ステップＳ２０５）。候補生成手段３０４は、並び替え確率の大きい文章を複数選択し、質問文言い換え候補として出力手段２００に出力する（ステップＳ２０５）。なお、候補生成手段３０４が出力する言い換え候補は、１つであってもよい。

【００８６】

以上のように、候補生成手段３０４は、入力手段１００から入力された質問文に対し、モデル記憶手段４０２に記憶された質問文言い換えモデルを適用して言い換え候補を生成する。

【００８７】

図１９は、入力手段１００に入力された質問文の一例を示した説明図である。また、図２０は、図１９に示した質問文に対して、候補生成手段３０４が質問文言い換えモデルを用いて生成した言い換え候補の一例を示した説明図である。図１９に示すような入力質問文Q_in に対し、候補生成手段３０４は、質問文言い換えモデル$P_n = {}_3$を用いて、図２０に示すような言い換え候補を生成する。候補生成手段３０４は、生成した言い換え候補を、質問文言い換えモデルの確率値（スコア）とともに出力手段２００に出力する（ステップＳ１０５）。なお、質問文言い換えモデルの確率値（スコア）は、質問文言い換えモ

デル（単語の並び替え確率）の対数で表現される。また、スコアの値が大きい（すなわち、スコアの絶対値が小さい）言い換え候補ほど、入力された質問文に対して類似度が高いことを示している。

【００８８】

以上のように、本実施の形態における文言い換えシステムでは、同一の回答文書に対応する質問文書集合から、言い換えの関係にあるセンテンスのみが篩い分けられる。すなわち、質問文言い換えモデルを用いることによって、言い換えの関係にある文対のみを抽出できる。

【００８９】

また、本発明による文言い換えシステムは、質問文言い換えモデルを用いた言い換え文対の抽出と、抽出された言い換え文対による質問文言い換えモデルの学習との２つの過程を抽出文セット（抽出される言い換え文対）が収束するまで交互に繰り返し、最終的な質問文言い換えモデルを得る。このような逐次処理によって、適切な言い換えモデルを学習することができ、同一の回答文書に対応する質問文書集合の中から、言い換え文の関係にある文対を精度良く抽出することができる。

【００９０】

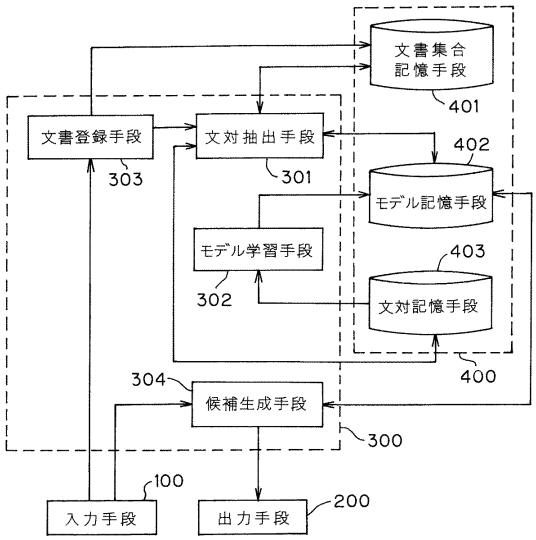
以上本発明の好適な実施形態について説明したが、本発明の全開示（請求の範囲を含む）の枠内において、さらにその基本的技術思想に基づいて、実施形態ないし実施例の変更・調整が可能である。また、本発明の請求の範囲の枠内において種々の開示要素の多様な組み合わせないし選択が可能である。例えば、上記した実施形態では、質問文言い換えモデルがモデル記憶手段４０２に既に存在している場合を説明したが、文言い換えシステムは、入力手段１００に質問文が入力されたことを契機として、質問文言い換えモデルの生成を開始してもよい。

【産業上の利用可能性】

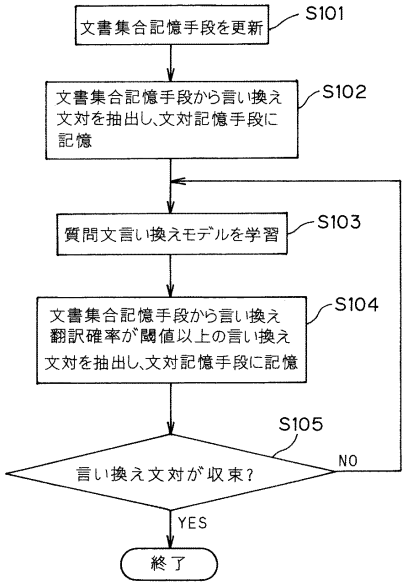
【００９１】

本発明は、入力された文章を異なる表現の文章に変換する文の言い換えシステムに適用できる。特に本発明は、質問文の表現不一致を解消できるため、メール応対事例の検索精度向上に適用できる。また、本発明は、メールで受信した質問文書に対する回答文書の自動生成や応対記録（要約）の自動生成に適用できる。

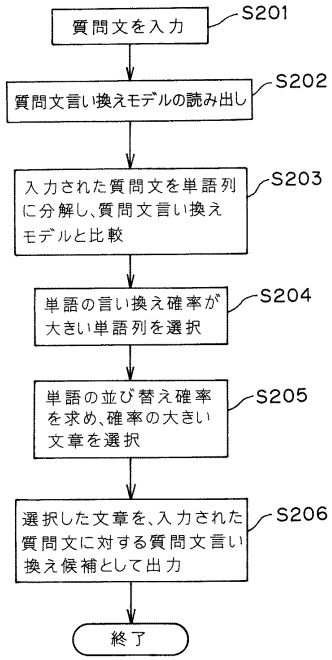
【図１】



【図２】



【図３】



【図４】

Q_ID	文書内容
Q1	パスワードを忘れました。ブラウザに記憶させていましたが、クッキーを誤って消去してしまったために、パスワードを忘れてしまいました。教えてくださるか、ホームページのURLの再発行をお願いします。
Q2	FTPパスワードが分からないのですが…？調べることはできますか？
Q3	ホームページIDとHPパスワードがわかりません。再発行して頂けますか？
Q4	転送にしたときFTPパスワードが正しくないといって拒否されました。
Q5	今回初めてhpを転送するのですが、何度やっても、「サーバーXXXXXXXX のユーザーIDまたはパスワードが正しくありません」と出るのですがどのように対処すればよいのでしょうか？
Q6	ホームページを作成したのですがFTPパスワードがわからないのです。前に頂いた下の項目のほかないのです。教えて下さい。 【ファイル転送時のご注意】 ★トップページのファイル名はindex.htmlまたは、index.htmにして下さい。 ★アップロードする際のファイル名は半角英数字をご利用下さい。
Q7	お手軽ホームページというのを「購入」したら、次のページで「設定」というのがでてきて、「パスワード」とかを入力するところがありますが、何を入力したらいいのかわかりません。教えてください。
Q8	ホームページのIDまたはパスワードが違いますと出るので転送が出来ません。できれば正しいのをお教え願えませんか。お忙しいところ申し訳ありません。
Q9	現在あるホームページを編集したいのに、FTPパスワードが一致しませんと何度も出てしまうんですけどどうしたらいいのでしょうか。ホームページIDとFTPパスワードが分からないです。(メモから入力してるんですが何が違うのか分からない。)
Q10	こんばんわ！お気に入りギャラリーセットを購入したのですがFTPパスワードを忘れてしまったので環境設定が出来ません。教えてください。
Q11	お願いいたします。「お手軽ホームページ」申し込みました。設定に入りましたが、ホームページIDは「XXXXXXXX」ですがホームページパスワードがよく解りません。何処を見ればよいのでしょうか。
Q12	ホームページの新規開設手続きをしたところパスワードが送られてきませんでした。そのため、パスワードがわかりません！しかし、パスワードの再発行をしようとしても、なぜか出来ませんでした。どうすればよいのですか？
Q13	お手軽ホームページを作成しようとしているのですが、パスワードを設定しても、新規作成をして保存をしたあとのパスワードを記入するとあいません。どうしたらいいのでしょうか…。なるべくはやく返答していただけると嬉しいです。
Q14	メールパスワードを変更しました。そうしたらHPの転送がうまく行かなくなりました。転送はnextFTPで行なっています。nextFTPのホストの設定のパスワードは新しいメールパスワードにするのでしょうか。

【図５】

A_ID	文書内容
A1	FTPアカウント、FTPパスワードとは、ホームページ新規開設の時に発行されたものです。FTPパスワードがエラーとなってしまう場合や、FTPパスワードをお忘れの場合は、下記URLから、「FTPパスワード再発行」手続きを行って下さい。 <FTPパスワード再発行> http://XXXXXXXX/
A2	ファイルが転送できない場合は、下記事項を必ず明記の上、再度お問い合わせ下さいますようお願いいたします。 ・お客様のmicrobe ID(会員証記載の、英字3字・数字5字のIDです) ・ホームページURL ・ご使用になったFTPソフト名 ・実際に入力された項目(但し、パスワードの記載は不要です。) ・エラーメッセージが出る場合は、その内容 ・Internetへの接続形態(アナログ回線、ISDN、LAN経由・・・・など) ・その他詳しい状況(ご希望)
A3	サーバ名は、XXXXXXXXになりますのでつづりにご注意ください。転送に必要な設定をまとめると以下の様になります。今一度ご確認の上、転送をお試しください。 <設定例> 転送先サーバー名:XXXXXXXX ユーザーID:FTPアカウント パスワード:FTPパスワード 転送先フォルダの設定:public_html
A4	下記手順にてお手軽ホームページの環境設定を行って頂き、パスワードの認証をご確認ください。下記URLより「作成／編集」をクリックします。 http://XXXXXXXX/ 「環境設定」をクリックします。 FTPアカウント(ホームページID):FTPアカウントを入力 FTPパスワード(HPパスワード):FTPパスワードを入力 パスワードを記憶「する」にチェックを入れましたら「設定」をクリックして下さい。上記の手順でエラーが発生しない場合は、ページの作成・保存へお進みいただき、同じFTPパスワードで保存をお試しください。
A5	FTPの際に使用するパスワードは、「ホームページパスワード(FTPパスワード)」です。Internet接続用パスワードとは異なります。

【図６】

質問文書	質問文書に対する回答文書(文書の構成要素)
Q1	A1
Q2	A1
Q3	A1
Q4	A1,A2
Q5	A1,A3
Q6	A1
Q7	A1
Q8	A1,A2
Q9	A1
Q10	A1
Q11	A1,A4
Q12	A1
Q13	A1
Q14	A1,A5

【図７】

質問文書集合	質問文書集合における共通の回答文書
Q1,Q2,Q3,Q6,Q7,Q9,Q10,Q12,Q13	A1
Q4,Q8	A1,A2
Q5	A1,A3
Q11	A1,A4
Q14	A1,A5

【図８】

言い換え文対
(Q1_S1,Q1_S2)
(Q1_S1,Q10_S2)
(Q1_S2,Q10_S2)
(Q2_S1,Q3_S1)
(Q2_S1,Q6_S1)
(Q2_S1,Q9_S2)
(Q2_S1,Q10_S2)
(Q2_S1,Q12_S2)
(Q3_S1,Q6_S1)
(Q3_S1,Q9_S2)
(Q3_S1,Q12_S2)
(Q6_S1,Q9_S2)
(Q6_S1,Q10_S2)
(Q6_S1,Q12_S2)
(Q9_S2,Q10_S2)
(Q9_S2,Q12_S2)

【図９】

質問文
(Q1_S1)パスワードを忘れました。
(Q1_S2)ブラウザに記憶させていましたが、クッキーを誤って消去してしまったために、パスワードを忘れてしまいました。
(Q2_S1)FTPパスワードが分からないのですが…?
(Q3_S1)ホームページIDとHPパスワードがわかりません。
(Q6_S1)ホームページを作成したのですがFTPパスワードがわからないのです。
(Q9_S2)ホームページIDとFTPパスワードが分からないです。
(Q10_S2)お気に入りギャラリーセットを購入したのですがFTPパスワードを忘れてしまったので環境設定が出来ません。
(Q12_S2)そのため、パスワードがわかりません!

【図１０】

Q1_S1	⇔	Q1_S2
φ	⇔	／ブラウザ／に／記憶／させて／いました／が／、／クッキー／を／誤っ／て／消去／して／しまった／ため／に／、／
／パスワード／	⇔	／パスワード／
／を／	⇔	／を／
／忘れ／	⇔	／忘れ／て／しまい／
／まし／	⇔	／まし／
／た／	⇔	／た／
／。／	⇔	／。／

【図１１】

Q1_S1	⇔	Q10_S2
φ	⇔	/お気に入り/ギャラリーセット/を/購入/した/の/です/が/
/パスワード/	⇔	/FTP/パスワード/
/を/	⇔	/を/
/忘れ/	⇔	/忘れ/
/まし/	⇔	/て/しまっ/
/た/	⇔	/た/
φ	⇔	/ので/環境設定/が/出来/ません/
/。/	⇔	/。/

【図１２】

Q2_S1	⇔	Q6_S1
φ	⇔	/ホームページ/を/作成/した/の/です/が/
/FTP/パスワード/	⇔	/FTP/パスワード/
/が/	⇔	/が/
/分から/ない/	⇔	/わから/ない/
/の/	⇔	/の/
/です/	⇔	/です/
/が/…/?/	⇔	/。/

【図１３】

Q2_S1	⇔	Q9_S2
φ	⇔	/ホームページ/ID/と/
/FTP/パスワード/	⇔	/FTP/パスワード/
/が/	⇔	/が/
/分から/	⇔	/分から/
/ない/	⇔	/ない/
/の/です/	⇔	/です/
/が/…/?/	⇔	/。/

【図１４】

Q2_S1	⇔	Q10_S2
φ	⇔	/お気に入り/ギャラリーセット/を/購入/した/の/です/が/
/FTP/	⇔	/FTP/
/パスワード/	⇔	/パスワード/
/が/	⇔	/を/
/分から/ない/	⇔	/忘れ/て/しまっ/た/
/の/	⇔	/の/
/です/	⇔	/で/
φ	⇔	/環境設定/が/出来/ません/
/が/…/?/	⇔	/。/

【図１５】

Q2_S1	⇔	Q12_S2
φ	⇔	/そのため/、/
/FTP/パスワード/	⇔	/パスワード/
/が/	⇔	/が/
/分から/ない/の/です/	⇔	/わかり/ません/
/が/…/?/	⇔	/!/

【図１６】

言い換え文対
(Q1_S1,Q1_S2)
(Q1_S1,Q10_S2)
(Q1_S2,Q10_S2)
(Q2_S1,Q3_S1)
(Q2_S1,Q6_S1)
(Q2_S1,Q9_S2)
(Q2_S1,Q10_S2)
(Q2_S1,Q12_S2)
(Q3_S1,Q6_S1)
(Q3_S1,Q9_S2)
(Q3_S1,Q12_S2)
(Q6_S1,Q9_S2)
(Q6_S1,Q10_S2)
(Q6_S1,Q12_S2)
(Q9_S2,Q10_S2)
(Q9_S2,Q12_S2)
(Q4_S1,Q8_S1)

【図１７】

質問文
(Q4_S1) 転送にしたときFTPパスワードが正しくないといって拒否されました。
(Q8_S1) ホームページのIDまたはパスワードが違いますと出るので転送が出来ません。

【図１８】

Q4_S1	⇔	Q8_S1
/転送/に/した/とき/	⇔	/転送/が/
φ	⇔	/ホームページ/の/ID/または/
/FTP/パスワード/	⇔	/パスワード/
/が/	⇔	/が/
/正しくない/	⇔	/違います/
/と/	⇔	/と/
/いっ/て/	⇔	/出る/ので/
/拒否/され/ました/	⇔	/出来/ません/
/。/	⇔	/。/

【図１９】

入力質問文Q_in
転送するときにパスワードを入れましたが「パスワードが違います」のエラーメッセージが出るので出来ません。

【図２０】

スコア	入力質問文Q_inに対する言い換え候補
−0.134	転送するときにFTPパスワードを入れましたが「パスワードが違います」のエラーメッセージが出るので出来ません。
−0.134	転送するときにパスワードを入れましたが「FTPパスワードが違います」のエラーメッセージが出るので出来ません。
−0.243	転送するときにFTPパスワードを入れましたが「FTPパスワードが違います」のエラーメッセージが出るので出来ません。
−0.329	転送するときにFTPパスワードを入れましたが「パスワードが違います」のエラーメッセージが出るので出来ませんが…?
−0.329	転送するときにパスワードを入れましたが「FTPパスワードが違います」のエラーメッセージが出るので出来ませんが…?
−0.472	転送するときにFTPパスワードを入れましたが「FTPパスワードが違います」のエラーメッセージが出るので出来ませんが…?
−0.249	パスワードを入れましたが「パスワードが違います」のエラーメッセージが出るので転送が出来ません。
−0.393	FTPパスワードを入れましたが「パスワードが違います」のエラーメッセージが出るので出来ません。
−0.393	パスワードを入れましたが「FTPパスワードが違います」のエラーメッセージが出るので転送が出来ません。
−0.430	FTPパスワードを入れましたが「FTPパスワードが違います」のエラーメッセージが出るので転送が出来ません。
−0.487	パスワードを入れましたが「パスワードが違います」のエラーメッセージが出るので転送が出来ませんが…?
−0.537	FTPパスワードを入れましたが「パスワードが違います」のエラーメッセージが出るので転送が出来ませんが…?
−0.537	パスワードを入れましたが「FTPパスワードが違います」のエラーメッセージが出るので転送が出来ませんが…?
−0.595	FTPパスワードを入れましたが「FTPパスワードが違います」のエラーメッセージが出るので転送が出来ませんが…?

フロントページの続き

(56)参考文献 特開２００５－０９２２７１（ＪＰ，Ａ）
特開２００４－１７１３５４（ＪＰ，Ａ）
石川開 他１名，質問タイプとトピックを用いるＦＡＱ検索手法の提案，ＦＩＴ２００６ 第５回情報科学技術フォーラム 一般講演論文集，日本，電子情報通信学会，２００６年 ８月２１日，第２分冊，171-172頁

(58)調査した分野(Int.Cl.，ＤＢ名)
G06F 17/20-17/28