

(19) 日本国特許庁(JP)

(12) 特許公報(B2)

(11) 特許番号

特許第4924950号
(P4924950)

(45) 発行日 平成24年4月25日(2012.4.25)

(24) 登録日 平成24年2月17日(2012.2.17)

(51) Int.Cl.

G06F 17/30 (2006.01)

F 1

G06F 17/30 180A
G06F 17/30 340B
G06F 17/30 350C

請求項の数 32 (全 30 頁)

(21) 出願番号 特願2007-502677 (P2007-502677)
 (86) (22) 出願日 平成18年2月8日 (2006.2.8)
 (86) 国際出願番号 PCT/JP2006/302570
 (87) 国際公開番号 WO2006/085661
 (87) 国際公開日 平成18年8月17日 (2006.8.17)
 審査請求日 平成21年1月15日 (2009.1.15)
 (31) 優先権主張番号 特願2005-32379 (P2005-32379)
 (32) 優先日 平成17年2月8日 (2005.2.8)
 (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 100093595
 弁理士 松本 正夫
 (72) 発明者 中澤 聰
 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 池田 崇博
 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 佐藤 研治
 東京都港区芝五丁目7番1号 日本電気株
 式会社内

最終頁に続く

(54) 【発明の名称】質問応答データ編集装置、質問応答データ編集方法、質問応答データ編集プログラム

(57) 【特許請求の範囲】

【請求項 1】

対話内容を編集して質問応答データを作成する質問応答データ編集装置であって、前記対話内容に対して、既存の質問応答データと類似する箇所を検出する手段と、検出された前記類似する箇所の近傍の対話内容から、該対話内容がなされた状況を抽出し、抽出された前記状況を新たな質問応答データとして登録する、または前記質問応答データの索引情報として登録する手段とを備えたことを特徴とする質問応答データ編集装置。

【請求項 2】

前記状況の抽出において、前記状況が表現パターンを利用して抽出されることを特徴とする請求項 1 に記載の質問応答データ編集装置。 10

【請求項 3】

対話内容を編集して質問応答データを作成する質問応答データ編集装置であって、前記対話内容から、既存の質問応答データと類似する対話内容を検出する手段と、検出された前記対話内容の近傍から、前記対話のなされた状況や条件を含む表現パターンを抽出し、抽出された前記表現パターンを、前記質問応答データの索引情報として登録する手段とを備えたことを特徴とする質問応答データ編集装置。

【請求項 4】

前記抽出された前記表現パターンを、前記質問対応データと対応付けて編集可能に提示する手段を備えることを特徴とする請求項 3 に記載の質問応答データ編集装置。 20

【請求項 5】

対話内容を編集して、質問部分と応答部分からなる質問応答データを作成する質問応答データ編集装置であって、

前記質問部分または前記応答部分のいずれかが類似する、既存の質問応答データと、対話内容中の質問応答データの組合せを抽出する組合せ抽出手段と、

前記抽出された組合せで相違する部分を検出する相違部分検出手段と、

前記対話内容中の前記相違する部分を、前記既存の質問応答データのバリエーションとして登録する手段とを備えたことを特徴とする質問応答データ編集装置。

【請求項 6】

前記抽出された対話内容中の質問応答データまたはその一部を、前記抽出された既存の質問応答データと対応づけて編集可能に提示する手段を備えることを特徴とする請求項 5 に記載の質問応答データ編集装置。

【請求項 7】

対話内容を編集して質問応答データを作成する質問応答データ編集装置であって、

前記対話内容から、既存の質問応答データと類似する対話内容を検出する手段と、

検出された前記対話内容が複数存在し、かつ複数の前記対話内容が互いに近接している場合、近接している前記対話内容を互いに関連する関連質問応答データとして抽出し、抽出された前記関連質問応答データを、前記質問応答データに関連するデータとして登録する手段とを備えたことを特徴とする質問応答データ編集装置。

【請求項 8】

前記抽出された前記関連質問応答データを、前記質問対応データと対応付けて編集可能に提示する手段を備えることを特徴とする請求項 7 に記載の質問応答データ編集装置。

【請求項 9】

対話内容を編集して質問応答データを作成する質問応答データ編集装置であって、

前記対話内容から、既存の質問応答データと類似する対話内容を検出する手段と、

検出された前記対話内容の近傍から、前記対話のなされた状況や条件を含む表現パターンを抽出し、抽出された前記表現パターンを、前記質問応答データの索引情報として登録する手段と、

検出された前記対話内容の近傍から、前記質問応答データに類似しない質問部分や応答部分を抽出し、抽出された前記質問部分や応答部分を、前記質問応答データのバリエーションとして登録する手段と、

検出された前記対話内容が複数存在し、かつ複数の前記対話内容が互いに近接している場合、近接している前記対話内容を互いに関連する関連質問応答として抽出し、抽出された前記関連質問応答を、前記質問応答データに関連するデータとして登録する手段とを備えたことを特徴とする質問応答データ編集装置。

【請求項 10】

前記対話内容の履歴データを記録する対話履歴データベースと、既存の前記質問応答データを記録する質問応答データベースと、

処理対象とする前記対話内容及び前記質問応答データの条件を入力する条件入力部とを備え、

入力された条件に一致する前記対話内容及び前記質問応答データを、前記対話履歴データベースと前記質問応答データベースから読み込んで、処理対象とすることを特徴とする請求項 1 から請求項 9 の何れか 1 項に記載の質問応答データ編集装置。

【請求項 11】

請求項 1 から請求項 10 に記載の質問応答データ編集装置で作成された前記新たな質問応答データまたは前記索引情報を用いて対話を支援する対話支援装置。

【請求項 12】

対話内容を編集して質問応答データを作成する質問応答データ編集方法であって、

前記対話内容に対して、既存の質問応答データと類似する箇所を検出するステップと、

検出された前記類似する箇所の近傍の対話内容から、該対話内容がなされた状況を抽出

10

20

30

40

50

し、抽出された前記状況を新たな質問応答データとして登録する、または前記質問応答データの索引情報として登録するステップと
を有することを特徴とする質問応答データ編集方法。

【請求項13】

前記状況の抽出において、前記状況が表現パターンを利用して抽出されることを特徴とする請求項12に記載の質問応答データ編集方法。

【請求項14】

対話内容を編集して質問応答データを作成する質問応答データ編集方法であって、
前記対話内容から、既存の質問応答データと類似する対話内容を検出するステップと、
検出された前記対話内容の近傍から、前記対話のなされた状況や条件を含む表現パターンを抽出するステップと、
10

抽出された前記表現パターンを、前記質問応答データの索引情報として登録するステップを有することを特徴とする質問応答データ編集方法。

【請求項15】

前記抽出された前記表現パターンを、前記質問対応データと対応付けて編集可能に提示するステップをさらに有することを特徴とする請求項14に記載の質問応答データ編集方法。

【請求項16】

対話内容を編集して、質問部分と応答部分からなる質問応答データを作成する質問応答データ編集方法であって、
20

質問部分または応答部分のいずれかが類似する、既存の質問応答データと、対話内容中の質問応答データの組合せを抽出する組合せ抽出ステップと、

前記抽出された組合せで相違する部分を検出する相違部分検出ステップと、
前記対話内容中の前記相違する部分を、前記既存の質問応答データのバリエーションとして登録するステップとを有することを特徴とする質問応答データ編集方法。

【請求項17】

対話内容を編集して質問応答データを作成する質問応答データ編集方法であって、
前記対話内容から、既存の質問応答データと類似する対話内容を検出するステップと、
検出された前記対話内容が複数存在し、かつ複数の前記対話内容が互いに近接している場合、近接している前記対話内容を互いに関連する関連質問応答データとして抽出するステップと、
30

抽出された前記関連質問応答データを、前記質問応答データに関連するデータとして登録するステップを有することを特徴とする質問応答データ編集方法。

【請求項18】

前記抽出された前記関連質問応答データを、前記質問対応データと対応付けて編集可能に提示するステップをさらに有することを特徴とする請求項17に記載の質問応答データ編集方法。

【請求項19】

対話内容を編集して質問応答データを作成する質問応答データ編集方法であって、
前記対話内容の履歴データから、既存の質問応答データと類似する対話内容を検出するステップと、
40

検出された前記対話内容の近傍から、前記対話のなされた状況や条件を含む表現パターンを抽出し、抽出された前記表現パターンを、前記質問応答データの索引情報として登録するステップと、

検出された前記対話内容の近傍から、前記質問応答データに類似しない質問部分や応答部分を抽出し、抽出された前記質問部分や応答部分を、前記質問応答データのバリエーションとして登録するステップと、

検出された前記対話内容が複数存在し、かつ複数の前記対話内容が互いに近接している場合、近接している前記対話内容を互いに関連する関連質問応答として抽出し、抽出された前記関連質問応答を、前記質問応答データに関連するデータとして登録するステップと
50

を有することを特徴とする質問応答データ編集方法。

【請求項 2 0】

前記対話内容の履歴データを記録する対話履歴データベースと、既存の前記質問応答データを記録する質問応答データベースから読み込んで処理対象とするための、前記対話内容及び前記質問応答データの条件を入力するステップと、

入力された条件に一致する前記対話内容及び前記質問応答データを、前記対話履歴データベースと前記質問応答データベースから処理対象として読み込むステップを有することを特徴とする請求項 1 2 から請求項 1 9 の何れか 1 項に記載の質問応答データ編集方法。

【請求項 2 1】

対話内容を編集して質問部分と応答部分を組とする質問応答データを作成する質問応答データ編集方法であって、

既存の質問応答データの質問部分または応答部分のいずれかの部分が類似する類似部分を対話内容から抽出し、

前記類似部分と組となる部分を前記既存の質問応答データと前記対話内容の双方から検出し、

前記検出された既存の質問応答データの部分と前記検出された対話内容の部分が異なる場合に前記検出された対話内容の部分を前記類似部分が検出された既存の質問応答データのバリエーションとして登録することを特徴とする質問応答データ編集方法。

【請求項 2 2】

前記抽出された対話内容中の質問応答データまたはその一部を、前記抽出された既存の質問応答データと対応づけて編集可能に提示するステップを有することを特徴とする請求項 1 2 又は請求項 2 1 に記載の質問応答データ編集方法。

【請求項 2 3】

請求項 1 2 から請求項 2 0 に記載の質問応答データ編集方法で作成された前記新たな質問応答データまたは前記索引情報を用いて対話を支援する対話支援方法。

【請求項 2 4】

コンピュータを制御することにより、対話内容を編集して質問応答データを作成する質問応答データ編集プログラムであって、

前記対話内容に対して、既存の質問応答データと類似する箇所を検出する機能と、
検出された前記類似する箇所の近傍の対話内容から、該対話内容がなされた状況を抽出し、抽出された前記状況を新たな質問応答データとして登録する、または前記質問応答データの索引情報を登録する機能と

を前記コンピュータに実行させることを特徴とする質問応答データ編集プログラム。

【請求項 2 5】

前記状況の抽出において、前記状況が表現パターンを利用して抽出されることを特徴とする請求項 2 4 に記載の質問応答データ編集プログラム。

【請求項 2 6】

コンピュータを制御することにより、対話内容を編集して質問応答データを作成する質問応答データ編集プログラムであって、

前記対話内容から、既存の質問応答データと類似する対話内容を検出する機能と、
検出された前記対話内容の近傍から、前記対話のなされた状況や条件を含む表現パターンを抽出する機能と、

抽出された前記表現パターンを、前記質問応答データの索引情報として登録する機能を前記コンピュータに実行させることを特徴とする質問応答データ編集プログラム。

【請求項 2 7】

コンピュータを制御することにより、対話内容を編集して、質問部分と応答部分からなる質問応答データを作成する質問応答データ編集プログラムであって、

質問部分または応答部分のいずれかが類似する、既存の質問応答データと、対話内容中の質問応答データの組合せを抽出する組合せ抽出機能と、

前記抽出された組合せで相違する部分を検出する相違部分検出機能と、

10

20

30

40

50

前記対話内容中の前記相違する部分を、前記既存の質問応答データのバリエーションとして登録する機能とを前記コンピュータに実行させることを特徴とする質問応答データ編集プログラム。

【請求項 28】

コンピュータを制御することにより、対話内容を編集して質問応答データを作成する質問応答データ編集プログラムであって、

前記対話内容から、既存の質問応答データと類似する対話内容を検出する機能と、
検出された前記対話内容が複数存在し、かつ複数の前記対話内容が互いに近接している場合、近接している前記対話内容を互いに関連する関連質問応答データとして抽出する機能と、

抽出された前記関連質問応答データを、前記質問応答データに関連するデータとして登録する機能を前記コンピュータに実行させることを特徴とする質問応答データ編集プログラム。

10

【請求項 29】

コンピュータを制御することにより、対話内容を編集して質問応答データを作成する質問応答データ編集プログラムであって、

前記対話内容の履歴データから、既存の質問応答データと類似する対話内容を検出する機能と、

検出された前記対話内容の近傍から、前記対話のなされた状況や条件を含む表現パターンを抽出する機能と、

20

抽出された前記表現パターンを、前記質問応答データの索引情報として登録する機能と、

検出された前記対話内容の近傍から、前記質問応答データに類似しない質問部分や応答部分を抽出する機能と、

抽出された前記質問部分や応答部分を、前記質問応答データのバリエーションとして登録する機能と、

検出された前記対話内容が複数存在し、かつ複数の前記対話内容が互いに近接している場合、近接している前記対話内容を互いに関連する関連質問応答として抽出する機能と、

抽出された前記関連質問応答を、前記質問応答データに関連するデータとして登録する機能を前記コンピュータに実行させることを特徴とする質問応答データ編集プログラム。

30

【請求項 30】

前記対話内容の履歴データを記録する対話履歴データベースと、既存の前記質問応答データを記録する質問応答データベースから読み込んで処理対象とするための、前記対話内容及び前記質問応答データの条件を入力する機能と、

入力された条件に一致する前記対話内容及び前記質問応答データを、前記対話履歴データベースと前記質問応答データベースから読み込む機能を有することを特徴とする請求項24から請求項29の何れか1項に記載の質問応答データ編集プログラム。

【請求項 31】

前記抽出された対話内容中の質問応答データまたはその一部を、前記抽出された既存の質問応答データと対応づけて編集可能に提示する機能を前記コンピュータに実行させることを特徴とする請求項27に記載の質問応答データ編集プログラム。

40

【請求項 32】

請求項24から請求項30に記載の質問応答データ編集プログラムで作成された前記新たな質問応答データまたは前記索引情報を用いて対話を支援する機能を前記コンピュータに実現させる対話支援プログラム。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、いわゆるFAQやQA集と呼ばれる質問・応答集や、セールス・受付などの応対マニュアルなどの、予め想定して用意しておく質問応答データを編集・追加する、

50

質問応答データ編集装置、その編集方法及びプログラムに関し、特に蓄積された過去の対話履歴を用いて、追加・編集する質問応答データの情報や構造を定める質問応答データ編集装置、その編集方法及びプログラムに関するものである。

【背景技術】

【0002】

製品やサービスに関して顧客から企業に寄せられる様々な問い合わせや苦情に対応する、いわゆるコンタクトセンターやコールセンターの業務においては、製品やサービスの見積もり・発注、アフターサービス、修理などに関する問い合わせが日々多数寄せられる。オペレータがこうした問い合わせに応答する際、一般にFAQやQA集などと呼ばれる質問・応答集を予め準備しておくことは大変有用である。

10

問い合わせに応答するオペレータは、顧客からの問い合わせ内容に応じて適宜、こうした質問・応答集を参照することで、効率よく適切な返答を行うことができる。さらに企業のホームページなどに、こうしたFAQやQA集などの質問・応答集を掲示しておくことで、顧客が自分で質問・応答集を調べて問題解決を図ることができるので、電話やメール等による直接的な問い合わせの件数そのものが減少することも期待できる。

また、セールスや受付など実際に顧客と対面して応対する業務においても、想定される問い合わせに対する返答事例や、ある種の質問に対する注意点などをまとめた応対マニュアルなどを準備しておくことは重要である。こうした応対マニュアルは、実際の接客時だけでなく、応対するオペレータの教育にも利用することができる。

ここで、コンタクトセンターやコールセンターなどで利用される質問・応答集と、セールスや受付などで利用される応対マニュアルは、基本的には同種の構造を持つデータであるといえる。こうしたデータは、質問・問い合わせの内容や状況を表す質問部分と、それに対する応答方法や留意点などを表す応答部分からなる。本明細書では、再利用することを前提に編集されたこれらのデータを質問応答データと記す。

20

上述したような質問応答データを作成する場合、様々な状況に対応した豊富な内容の質問応答内容を準備することに加えて、作成した質問応答データに対して、容易に検索や参照が可能となるように検索のためのインデックスを作成することも非常に重要となる。

これら質問応答データの作成に関連する従来技術としては、例えば、特開平11-272584号公報（文献1）、特開2003-30224号公報（文献2）、特開2003-223460号公報（文献3）に開示される技術が存在する。

30

文献1では、インターネット上のホームページに対する電子メールでの質問を対象にした技術が公開されている。この従来技術では、ホームページの閲覧者の質問と共に、質問が発生した状況を示す状況データと、閲覧者の属性を示すユーザデータを収集し、データベースに蓄積する。これにより例えば、どのような閲覧者はどのページを見て質問を発したのか、などの情報が分かり、ホームページの改善に利用できるようにしている。また、QA集を作成する際にも、蓄積された質問内容を、ページや日付毎にソートし、それに対してオペレータが回答を作成することで、ページや日付毎のQA集とする方法が示されている。

文献2では、蓄積された質問文書と回答文書の対の集合を、非階層的ないくつかのクラスタに分類する技術が公開されている。この技術によれば、多数の質問・回答文書から類似の文書をまとめ上げてクラスタを作成し、各クラスタから典型的な質問・回答文書を選択することで、FAQ作成に利用できる。また、質問・回答文書の検索の際に、各クラスタの典型文書だけ出力することで、検索結果の一覧性向上に役立つ。

40

文献3では、顧客の問い合わせ中に用いられることが予測される表現と、それが関連する検索用キーワードを予めリンクして保持しておく技術が公開されている。例えばプリンタの「目詰まり」という検索用キーワードに対して、「インクが出ない」「線が細い」などの表現をリンクしておく。これにより、実際に顧客から「インクが出ない」という内容の問い合わせが来たときにすぐにプリンタの「目詰まり」に関するQA集を検索することができる。

しかし、上記従来技術には、以下のような問題点がある。

50

第 1 の問題点は、F A Q などの質問応答データを作成する際に、実際に顧客が使用した表現や、個々の質問・問い合わせが生じた状況・条件が詳細には分からぬことである。

一般に、質問応答データを作成する際には、実際にオペレータが顧客からの問い合わせに対して回答した内容の報告や、経験内容の記録が利用されるが、多くの場合、こうした報告・記録は実際の問い合わせ内容と回答内容を簡潔にまとめたものになっており、生のデータではない。そのため、内容的には同種の問い合わせであって、顧客が使用する表現によって多くのバリエーションがある場合、報告・記録からはそのことが分からぬ。また、質問・問い合わせには、必ずそれが生じた状況・条件があり、こうした状況・条件に応じて、同じ問い合わせに対しても回答が異なってくる場合や、回答に対する顧客のその後の反応や選択が異なってくる場合があるが、報告・記録からはこうした状況・条件を判断することが難しい。10

文献 1 で公開されている、ホームページの質問受付メールの表示方法では、問い合わせが生じた状況が問い合わせと組にして蓄積される。しかし、この技術で対象としているのは、ホームページを閲覧していた際に生じた問い合わせメールのみであり、問い合わせが生じた状況についても、ユーザプロファイルと問い合わせの日時、問い合わせを発した際に閲覧していたページの情報に限定される。

文献 2 では、蓄積された質問文書と回答文書を類似性の高い複数のクラスタにまとめ上げる技術が公開されている。従って、このクラスタリングがうまく働いていれば、同じクラスタに属する文書を調べることで、同種の内容の表現バリエーションを調べることができる。しかし通常、クラスタリング技術は文書の形式や長さなどの条件が揃っていないと有効に働かない。このため、ある質問文書は非常に短いが、ある質問文書は長くしかも複数の質問を行っている場合など、たとえ同種の質問を含んでいても、同一のクラスタにまとめ上げられるとは限らない。質問が起きた状況・条件についても、この文献 2 の技術からでは分からぬ。20

文献 3 で公開されている情報提供支援システムでは、実際に顧客が使用する表現「インクが出ない」「線が細い」などと、検索用キーワード「目詰まり」を予め結びつけて、Q A 集を検索する際に用いる。これは、検索用キーワード「目詰まり」の表現バリエーションを取り扱っていることができる。しかし、実際に顧客が使用する表現と、Q A 集で用いられている表現（検索用キーワード）とを、どのように収集し、どの表現とどの検索用キーワードをリンクさせれば適切であると判断するのかが未解決のままである。30

第 2 の問題点は、複数の質問応答データ間の関係が分からぬことである。

顧客が一度に複数の問い合わせを行ったり、質問をしてその回答を受け取り、さらにそこから質問を行う、など複数の質問がセットやシリーズとして尋ねられることがよく発生し得る。この場合、こうした質問や問い合わせから、F A Q のような質問応答データのデータベースを作成するには、個別の質問応答データとして蓄積するのではなく、それぞれの質問応答データを関連質問応答としてインデックスしておくことが望ましい。関連質問応答としてインデックスされていれば、実際に顧客と対応している際に、一方の質問応答で検索した後、関連質問応答をオペレータが見て、実際に顧客から尋ねられる前に準備しておくことや、関連質問応答の情報をオペレータ側から先に顧客に提示することもできる。また、ホームページ上などで質問応答データを F A Q として掲示する際にも、関連のある質問応答データをまとめたり、リンクして表示することで、閲覧性が向上する。40

しかし、従来技術では、こうした個々の質問応答データ間の関係を効率よく抽出するための方法は未だ提案されていない。

本発明はこのような問題点を解消するためになされたものであり、過去になされたオペレータと顧客との対話履歴データから、質問応答データに関する状況・条件を含む表現パターンや質問応答データの表現のバリエーションを抽出することで、質問応答データやインデックス情報を作成することのできる、質問応答データ編集装置、その編集方法及びプログラムを提供することを目的とする。

また本発明の第 2 の目的は、過去になされたオペレータと顧客との対話履歴データから、質問応答データに類似する対話内容を検出し、元の質問応答データとの関連付けを行う50

ことが可能な質問応答データ編集装置、その編集方法及びプログラムを提供することである。

【発明の開示】

【0003】

本発明の顧客との対話内容を編集して質問応答データを作成する質問応答データ編集装置は、前記顧客との前記対話内容の履歴データから、記憶されている既存の質問応答データと類似する対話内容を検出する手段と、検出された前記対話内容の近傍から、前記対話のなされた状況や条件を含む表現パターンを抽出し、抽出された前記表現パターンを、前記質問応答データの索引情報として登録する手段とを備える。

上記目的を達成するための本発明は、顧客との対話内容を編集して質問応答データを作成する質問応答データ編集装置であって、前記顧客との前記対話内容の履歴データから、記憶されている既存の質問応答データと類似する対話内容を検出する手段と、検出された前記対話内容の近傍から、前記質問応答データに類似しない質問部分や応答部分を抽出し、抽出された前記質問部分や応答部分を、前記質問応答データのバリエーションとして登録する手段とを備えたことを特徴とする。 10

また、本発明は、顧客との対話内容を編集して質問応答データを作成する質問応答データ編集装置であって、前記顧客との前記対話内容の履歴データから、記憶されている既存の質問応答データと類似する対話内容を検出する手段と、検出された前記対話内容が複数存在し、かつ複数の前記対話内容が互いに近接している場合、近接している前記対話内容を互いに関連する関連質問応答データとして抽出し、抽出された前記関連質問応答データを、前記質問応答データに関連するデータとして登録する手段とを備えたことを特徴とする。 20

さらに、本発明は、顧客との対話内容を編集して質問応答データを作成する質問応答データ編集装置であって、前記顧客との前記対話内容の履歴データから、記憶されている既存の質問応答データと類似する対話内容を検出する手段と、前記対話内容の履歴データの、類似する対話内容として検出されなかった前記対話内容から、質問部分と応答部分を組みにして抽出し、抽出された前記質問部分と応答部分を、新規の質問応答データとして登録する手段とを備えたことを特徴とする。

さらにまた、本発明は、顧客との対話内容を編集して質問応答データを作成する質問応答データ編集装置であって、前記顧客との前記対話内容の履歴データから、記憶されている既存の質問応答データと類似する対話内容を検出する手段と、検出された前記対話内容の近傍から、前記対話のなされた状況や条件を含む表現パターンを抽出し、抽出された前記表現パターンを、前記質問応答データの索引情報として登録する手段と、検出された前記対話内容の近傍から、前記質問応答データに類似しない質問部分や応答部分を抽出し、抽出された前記質問部分や応答部分を、前記質問応答データのバリエーションとして登録する手段と、検出された前記対話内容が複数存在し、かつ複数の前記対話内容が互いに近接している場合、近接している前記対話内容を互いに関連する関連質問応答として抽出し、抽出された前記関連質問応答を、前記質問応答データに関連するデータとして登録する手段と、前記対話内容の履歴データの、類似する対話内容として検出されなかった前記対話内容から、質問部分と応答部分を組みにして抽出し、抽出された前記質問部分と応答部分を、新規の質問応答データとして登録する手段とを備えたことを特徴とする。 30

また、本発明は、前記対話内容の履歴データを記録する対話履歴データベースと、既存の質問応答データを記録する質問応答データベースと、処理対象とする前記対話内容及び前記質問応答データの条件を入力する条件入力部とを備え、入力された条件に一致する前記対話内容及び前記質問応答データを、前記対話履歴データベースと前記質問応答データベースから読み込んで、類似する対話内容を検出することを特徴とする。

さらに、本発明の好ましい態様によれば、前記抽出された前記表現パターン、前記抽出された前記質問部分や応答部分、前記抽出された前記関連質問応答データ、前記新規の質問応答データをそれぞれ、編集可能に提示する手段を備える。

本発明は、このような構成を備え、既存の対話履歴データから質問応答データに対応し 50

た対話が発せられた状況・条件を含む表現パターンや、質問応答データの表現バリエーションを抽出し、もとの質問応答データの追加情報とすることで、本発明の第1の目的を達成することができる。

また本発明は、このような構成において、1つの対話履歴データ中に検出された複数の質問応答データを、関連質問応答として抽出することで、本発明の第2の目的を達成することができる。

本発明の質問応答データ編集装置、その編集方法及びプログラムによれば、以下の効果が得られる。

第1に、質問応答データが実際の対話中で使用される際の状況や条件等を含む表現パターンを質問応答データの索引情報として登録することができる。 10

第1の効果が得られる理由は、蓄積された対話履歴データから質問応答データに類似した箇所を検出し、その箇所の近傍から状況や条件等を含む表現パターンの記述を抽出して、質問応答データのインデックス情報として記録するためである。

第2に、質問応答データが実際の対話中で使用される際の質問部分や応答部分の表現のバリエーションを抽出することができるというものである。

第2の効果が得られる理由は、蓄積された対話履歴データから質問応答データの質問部分または応答部分に類似した箇所を検出し、その箇所の近傍からもとの質問応答データとは類似しない、応答または質問の記述を抽出して、質問応答データのインデックス情報として記録するためである。

第3に、質問応答データが実際の対話中で使用される際に、同一の対話中に連続して現れやすい関連の質問応答データを抽出することができる。 20

第3の効果が得られる理由は、蓄積された対話履歴データの中に連続して出現する2つの質問応答データの類似箇所を検出し、それらの質問応答データを関連質問応答データとしてインデックス情報に記録するためである。

第4に、蓄積された対話履歴データから、既存の質問応答データとは異なる新規の質問応答データを抽出することができる。新規の質問応答データを、新たに質問応答データベースに編集・追加することで、質問応答データの内容を充実させることができる。

第4の効果が得られる理由は、蓄積された対話履歴データの中から、既存の質問応答データには類似しない質問と応答の連続した記述を抽出し、その質問と応答の記述を新規の質問応答データとして記録するためである。 30

【図面の簡単な説明】

【0004】

図1は、本発明の実施の形態による質問応答データ編集装置の全体構成を示すブロック図である。

図2は、本発明の実施の形態の質問応答データ編集装置におけるマッチング部の構成を示すブロック図である。

図3は、本発明の実施の形態の質問応答データ編集装置における追加情報抽出部の構成を示すブロック図である。

図4は、本発明の実施の形態の質問応答データ編集装置における追加情報蓄積動作を示すフローチャートである。 40

図5は、本発明の実施の形態の質問応答データ編集装置におけるマッチング動作を示すフローチャートである。

図6は、本発明の実施の形態の質問応答データ編集装置における追加情報抽出動作を示すフローチャートである。

図7は、本発明の実施の形態における質問応答データ編集動作を示すフローチャートである。

図8は、本発明の実施の形態である質問応答データ編集装置のハードウェア構成を示すブロック図である。

図9は、対話履歴データ例1（携帯電話操作に関する問い合わせ）を示す図である。

図10は、対話履歴データ例2（電子掲示板での質問）を示す図である。 50

図11は、質問応答データの例3（携帯電話の操作に関する質問応答集）を示す図である。

図12は、質問応答データ例3のインデックス例を示す図である。

図13は、対話履歴データ例1における類似箇所検出結果を示す図である。

図14は、状況・条件等の記述抽出用の表現パターン例を示す図である。

図15は、対話履歴データ例1における表現パターンの記述の抽出結果を示す図である。

図16は、質問応答データ例3のインデックス例（追加情報蓄積後）を示す図である。

図17は、質問応答データ例4（パソコンに関するFAQの一部）を示す図である。

図18は、対話履歴データ例2における質問応答バリエーションの抽出結果を示す図である。
10

図19は、対話履歴データ例3（電気店エアコン売り場での顧客対応）を示す図である。

図20は、質問応答データ例5（電気店エアコン売り場での応対マニュアルの一部）を示す図である。

図21は、対話履歴データ例3における追加情報抽出部の処理結果を示す図である。

図22は、質問応答データ編集部によって画面上に提示される処理結果の一覧の例を示す図である。

10 : マッチング条件入力部、 20 : 対話履歴データベース、 30 : 質問応答データベース、 31 : 質問応答データ記録領域、 32 : インデックス記録領域、 40 : マッチング部、 41 : 対話履歴読み込み部、 42 : 対話履歴分割部、 43 : 質問応答データ読み込み部、 44 : 類似箇所検出部、 50 : 追加情報抽出部、 51 : 追加対象質問応答データ決定部、 52 : 表現パターン抽出部、 53 : 質問応答バリエーション抽出部、 54 : 関連質問応答抽出部、 55 : 新規質問応答抽出部、 56 : 追加情報出力部、 60 : 追加情報蓄積部、 70 : 質問応答データ編集部
20

【発明を実施するための最良の形態】

【0005】

次に、発明を実施するための最良の形態について図面を参照して詳細に説明する。

図1を参照すると、本発明の実施の形態による質問応答データ編集装置100は、処理対象とする対話履歴データの条件と質問応答データの条件を入力するマッチング条件入力部10と、既存の対話内容を対話履歴データとして蓄積する対話履歴データベース20と、編集作成された既存の質問応答データとそのインデックスを蓄積する質問応答データベース30と、対話履歴データと質問応答データのマッチング処理を行うマッチング部40と、マッチング結果から質問応答データおよびそのインデックスに追加すべき情報を抽出する追加情報抽出部50と、抽出された追加情報を一時的に記録・蓄積する追加情報蓄積部60と、蓄積された追加情報を編集する質問応答データ編集部70とからなる。
30

図8は、本発明の実施の形態である質問応答データ編集装置100のハードウェア構成を示すブロック図である。

図8を参照すると、質問応答データ編集装置100は、そのハードウェア構成として、通信部101と、キーボード102及びマウス103と、プログラム制御によりマッチング部40、追加情報抽出部50及び質問対応データ編集70の機能を実現するCPU100-1と、質問対応データ編集プログラムをロードして保持するメインメモリ100-2と、追加情報蓄積部60として機能する一時記憶部100-3と、ディスプレイ200と、対話履歴データベース20及び質問応答データベース30として機能するハードディスク装置等の記憶装置300を備えている。
40

通信部101は、マッチング条件入力部10として機能し、通信回線400を介して他の機器（端末装置等）から送信されるマッチング条件を受け付ける。また、キーボード102及びマウス103も、マッチング条件入力部10として機能し、ユーザの操作によってマッチング条件が直接入力される。

CPU100-1は、不揮発メモリやDRAM等に格納される各機能を提供する質問対

10

20

30

40

50

応データ編集プログラム（アプリケーション）を実行することにより、マッチング部40、追加情報抽出部50及び質問対応データ編集70の機能を提供する。

マッチング条件入力部10は、処理対象とする対話履歴データの条件と同じく処理対象とする質問応答データの条件とをマッチング条件として指定して入力する機能を有する。このマッチング条件入力部10は、ユーザ（編集を行う管理者）が直接入力するキーボードやマウスなどの通常使用される入力装置によって構成してもよいし、ネットワークなどを介して他の機器（端末装置等）から送信されるマッチング条件を受け付けて入力する通信手段等によって構成することも可能である。マッチング条件入力部10として具体的にどのような形態の装置を用いるかは、質問応答データ編集装置100を使用する環境によって異なってくる。10

またここで、対話履歴データの条件とは、対話履歴データベース20に蓄積されている個々の対話履歴データを読み出すために指定する任意の条件を指す。例えば、各対話履歴データにその対話がいつなされたものか時間情報が付与されているならば、ある日時以降の対話履歴データのみを処理対象とするといった条件の指定でもよい。また対話履歴データが記録された場面や状況、場所といった情報が付与されているならば、ある場所で記録された対話履歴データのみを対象とするといった条件の指定でも良い。さらには、本発明の質問応答データ編集装置100を前回使用した日時を記録しておく場合には、その後追加された対話履歴データのみを処理対象とする、といった条件の指定も可能である。

同様に、質問応答データの条件とは、質問応答データベース30に蓄積されている個々の質問応答データを読み出すために指定する任意の条件を指す。質問応答データが分野や目的別に分類されているのならば、その分野や目的を条件として指定することで、ある目的の質問応答データのみを処理対象とすることができます。また個々の質問応答データの参照回数などが記録されているのならば、参照回数を条件として指定することで、参照回数の少ないもの、あるいは多いもののみを処理対象とすることができます。任意の質問応答データを個別に指定して、指定された質問応答データのみ処理対象とするといった条件指定も可能である。20

さらには、本発明の実施の形態による質問応答データ編集装置100を使用する度に、処理対象とする対話履歴データや質問応答データの条件を詳細に指定する必要がない場合には、予めデフォルトの条件を定めておき、自動的にデフォルトの条件で処理が進められるようにしてもよい。30

対話履歴データベース20は、既存の対話履歴データを保存・蓄積するデータベースである。ここで対話履歴とは、2者以上の主体がやりとりした対話の記録のことである。主体には、通常の人間の他に応答エージェントなども含まれる。よって、人間と自動応答プログラムによる対話もこれに含まれる。また対話は、電話など音声でなされたものであっても、ファックスやメールなどテキストを介してなされたものであってもよい。掲示板やチャットシステムを介してのやりとりもこれに含まれる。対話履歴データには、質問応答データ編集装置100を使用する際の目的や用途に応じて、各対話の発言者や発言日時、場所、状況、条件など、その対話に関する情報が記録されている。

図9は、携帯電話の操作に関する問い合わせ電話による対話履歴データの例1を示している。図中、オペレータの発話内容を「O P :」で、顧客の発話内容を「C S :」で示している。この例では、問い合わせ電話で交わされた対話内容を録音し、その録音内容をテキストデータに変換することで対話履歴データとしている。また、各発話内容に対してそれが顧客の発話なのか、それともオペレータの発話なのかを記録している。40

図10は、インターネット上の電子掲示板における対話履歴データの例2を示している。この電子掲示板では、各記事の投稿毎に、投稿者名、投稿年月日、投稿のタイトル、投稿本文が記録される。また、各投稿者が投稿時にその記事の分類を「質問」「返答」「コメント」「その他」から選択するようになっている（図10では、各記事の左上にこの分類が示されている）。さらに、どの記事がどの記事に対して書かれたものなのか記事間の関連がスレッドによる階層構造として記録されている。このような例では、こうした記録のログがそのまま対話履歴データとして利用できる。50

対話履歴データの記録方法としては、テキストデータの他に、音声データ、映像データ等が考えられる。電子メールなど元々の対話内容がテキストで構成されている場合には、それをそのままテキストデータとして記録すればよい。電話や店頭での応対の様子など元々の対話が音声でなされている場合は、音声データとして記録しても良いし、音声から人手もしくは音声認識技術を用いてテキストを作成し、それをテキストデータとして記録しても良い。また、対話内容の音声データから、各対話の概要・キーワードのみや、オペレータの発話分だけなど、対話内容の一部をテキストデータとして書き起こし、詳細な内容を音声データとして記録する方式のように、複数のデータ形式を組み合わせて記録するようにもよい。

ただし、対話履歴データを記録するためにどのようなデータ形式を用いる場合でも、各対話履歴データは後述する類似箇所検出部44や追加情報抽出部50にて、テキストで記述された質問応答データとマッチング処理を行うことから、テキスト化する必要がある。10

従って、全ての対話履歴データは、類似箇所検出部44にてマッチングを取る前に、マッチング処理や情報抽出処理などでテキストによる処理が可能なよう、少なくとも対話の一部または全部をテキスト化しておく必要がある。

通常は、各対話履歴データを対話履歴データベース20に蓄積する前に、対話内容の一部または全部のテキスト化を行っておくことが望ましい。なぜならば、同じ対話履歴データに対して複数回マッチング処理を行う可能性があるが、対話履歴データベース20に蓄積された時点で既にテキスト化が終了していれば、マッチング等のたびにテキスト化を行う必要がないためである。20

しかし、本発明の実施の形態による質問応答データ編集装置100の用途や目的に応じて、後述する対話履歴読み込み部41や対話履歴分割部42にて、各対話履歴データが読み込まれるたびにテキスト化を行う方式であっても構わない。マッチングためのテキスト化をどの段階に行うかは、本発明の構成の他の部分や手順には影響しない。もちろん、元々の対話内容がテキストで記録されている場合には、あらためてテキスト化する必要ない。

質問応答データベース30は、既に編集作成されている既存の質問応答データを保存・蓄積するデータベースである。ここで質問応答データとは、背景技術の欄で述べたように、コンタクトセンターやコールセンター等で顧客との対応に際して利用される質問・応答集や、セールス・受付等で利用される応対マニュアルなど、質問・問い合わせの内容や状況を表す質問部分と、それに対する応答方法と留意点などを表す応答部分からなるテキストデータのことを示している。30

質問応答データは、便宜上「質問」、「応答」と表現しているが、そのデータの内容としては、必ずしも、通常の日本語の厳密な意味での質問や応答でなくともよい。以下に、質問応答データの例を示す。

(質問応答データ例1)

質問：「お客様が大きな声で激高された場合」

応答：「まず様子を見て激高の理由を判断すること。また、他のお客様が不安に感じられないよう注意する。必要に応じて他の場所に移り、とにかくお客様の話を真剣に聞く。」40

上記の質問応答データ例1のように、「質問」では単なる状況を記述していて、「応答」では注意点・対処法のみ記されている。このように、厳密な意味での質疑応答の形式になっていない、顧客に対する応対内容も質問応答データに含まれる。

もちろん、次の質問応答データ例2に示すように、本来の意味での質問と応答の組も、質問応答データに含まれる。

(質問応答データ例2)

質問：「パソコンがウイルスに感染したらしいのですが、どうすればよいですか？」

応答：「まずそのパソコンがネットワークに接続している場合は、ただちにネットワークの線を抜いて下さい。ただし、電源は消さないで下さい。ついで、ウイルス対策ソフトをインストールし、マイコンピュータのスキャンを行って下さい。」50

質問応答データベース30は、さらに詳細には、質問応答データを記録する質問応答データ記録領域31と質問応答データを検索するためのインデックスを記録するインデックス記録領域32からなる。質問応答データ記録領域31は上記した質問応答データ自体を記録する部分である。インデックス記録領域32は、質問応答データを検索するための検索キーワードや表現バリエーションからなる検索用インデックス、複数の質問応答データ間の関係を表す関連インデックス、質問応答データが有効に成立する状況や条件を含む表現パターンを表す表現パターン（状況・条件）インデックスなど、各質問応答データのメタ情報を保持する。具体的にどのような種類のインデックスを用いるかは、本発明の実施の形態による質問応答データ編集装置100の用途や目的、質問応答データの形式などによって異なる。

10

図11は、携帯電話の操作に関する質問応答データの例3を示している。図11に挙げられている質問応答データ3-1や質問応答データ3-2のような、質問応答データ自体が質問応答データ記録領域31に記録される。

一方、図12は、図11の質問応答データ3-1や質問応答データ3-2に対するインデックスの例を示している。質問応答データの作成者が、質問応答データ3-1、質問応答データ3-2を作成したときに、各質問応答データに含まれている単語や説明内容から作成したインデックスとなっている。こうしたインデックスが、それぞれの質問応答データと対応する形で、インデックス記録領域32に記録されている。

マッチング部40は、対話履歴データベース20から読み込んだ対話履歴データと、質問応答データベース30から読み込んだ質問応答データとのマッチングを取り、対話履歴データの中から質問応答データに類似する箇所を検出する機能を有する。

20

このマッチング部40の詳細を図2に示す。マッチング部40は、対話履歴読み込み部41と、対話履歴分割部42と、質問応答データ読み込み部43と、類似箇所検出部44とから構成される。

対話履歴読み込み部41は、マッチング条件入力部10で入力された対話履歴データの条件を満たす対話履歴データを全て、対話履歴データベース20から読み込んで、対話履歴分割部42に渡す。

対話履歴分割部42では、対話履歴読み込み部41から受け取った対話履歴データを解析し、各対話内容を複数のマッチング処理単位（マッチング処理を行う対話内容の範囲）に分割して、類似箇所検出部44に渡す。そして、類似箇所検出部44では、ここで分割されたマッチング処理単位毎に、類似箇所の判定を行う。

30

各対話内容をどのようなマッチング処理単位で分割するかは、本発明の実施の形態による質問応答データ編集装置100を使用する際の用途や目的、蓄積されている対話履歴データの形式や種類、類似箇所検出部44で用いられる類似箇所検出方式などによって異なる。

最も単純には、各対話内容を発話毎に区切って、個々の発話をマッチング処理単位としてもよい。発話毎ではなく、話者が切り替わる毎に区切って、マッチング処理単位とすることも可能である。また、おのののの発話が短く、1人以上の話者による複数の発話で1つの話題やまとまりとなっているような対話内容の場合は、自然言語処理による話題検出技術を用いて、話題の区切りを検出し、各話題によって対話内容を区切って、マッチング処理単位としてもよい。例えば、コールセンターでの対話で顧客が発話した内容を、オペレータが必ず内容をまとめた上で繰り返し、確認を取っているような場合では、このオペレータの確認のための発話部分を表現のパターンマッチング手法などで切り出し、マッチング処理単位とすることができます。

40

質問応答データ読み込み部43は、マッチング条件入力部10で入力された質問応答データの条件を満たす質問応答データを全て、質問応答データベース30から読み込んで、類似箇所検出部44に渡す。

類似箇所検出部44は、対話履歴分割部42から受け取った全ての対話履歴データを、マッチング処理単位毎に調べ、対話履歴データについて、質問応答データ読み込み部43から受け取った質問応答データの質問部分または応答部分に類似する箇所がないか判定す

50

る。

この類似箇所の判定は、受け取った全ての対話履歴データの全マッチング処理単位と、受け取った全ての質問応答データとの組み合わせに対して行い、対話履歴データ中に類似箇所が検出された場合には、そのマッチング処理単位がどの質問応答データの質問部分または応答部分に類似しているのか、後からわかるようマッチング処理単位にマーク付けをしておく。すなわち、マッチング処理単位と、そのマッチング処理単位に類似する質問応答データの質問部分または応答部分とをマーク付けによって対応付ける。

もし1つのマッチング処理単位が、複数の異なる質問応答データに類似していると判定された場合には、もっとも類似性の高い質問応答データを1つ選択して、その質問応答データだけをマーク付けするようにしてもよい。また、類似性がある一定値以上の質問応答データ全てを選択して、そのマッチング処理単位にマーク付けするようにしてもよい。10

類似箇所検出部44による類似箇所の判定には、自然言語処理における類似文書の判定技術を用いることができる。類似文書の判定技術としては、例えば、類似度を比較する2つの文書における単語出現ベクトルの一致度を見る手法や、表現パターンの一致を見る手法、単語間の接続を比較する手法、複数の手法を組み合わせたハイブリッド手法などが既に知られている。本発明の実施の形態による質問応答データ編集装置100を使用する際の用途や目的、さらには蓄積されている対話履歴データの形式や種類、既存の質問応答データの書式などに応じて適切な類似文書判定技術を利用することができる。これら類似文書の判定技術は既存の技術であるのでその詳細は省略する。

これら自然言語処理における類似文書の判定技術については、例えば、文献4（東京大学出版会；ISBN：4130654055「言語と計算（5）情報検索と言語処理」徳永 健伸（著），辻井 潤一（編集））に記載されている技術が存在する。20

図13は、図9の対話履歴データ例1と、図11の質問応答データ例3とに対して類似箇所検出部44による類似箇所検出処理を行った結果の例を示す。この例では、図9の対話履歴データ例1に含まれる1つ1つの発話それぞれを、単純にマッチング処理単位として分割している。図13に示す類似箇所検出の結果において、左側にある四角い吹き出しが、マッチング処理単位と類似する質問応答データの質問部分または応答部分がマーク付けされていることを示している。この場合の類似箇所は、対話履歴データのマッチング処理単位と、質問応答データの質問部分または応答部分とで、出現単語ベクトルの一致度がある閾値を超えたかどうかによって判定している。30

追加情報抽出部50は、マッチング部40で処理された、対話履歴データ中の既存の質問応答データとの類似箇所の検出結果から、新規に追加すべき質問応答データや、質問応答データのインデックス情報を抽出する。

この追加情報抽出部50の詳細を図3に示す。追加情報抽出部50は、追加対象質問応答データ決定部51と、表現パターン抽出部52と、質問応答バリエーション抽出部53と、関連質問応答抽出部54と、新規質問応答抽出部55と、追加情報出力部56とから構成される。

追加対象質問応答データ決定部51は、マッチング部40から受け取った対話履歴データの1つ1つに対して、マッチング処理単位毎に、類似箇所検出部44で付けられたマークがないかどうかを調べる。40

マッチング処理単位にマークが付けられている場合には、追加対象質問応答データ決定部51は、そのマッチング処理単位と類似していると判定された質問応答データを、新規にインデックス情報を抽出・追加する追加対象として決定し、その対話履歴データとマークされたマッチング処理単位の位置を表現パターン抽出部52と、質問応答バリエーション抽出部53と、関連質問応答抽出部54とに送る。なお、1つのマッチング処理単位に複数の異なる質問応答データがマークされている場合には、マークされている質問応答データ全てを新規にインデックス情報を抽出・追加する追加対象とする。

マッチング処理単位にマークが見つからない場合には、現在調べている対話履歴データとマッチング処理単位の位置を新規質問応答抽出部55に送る。

表現パターン抽出部52は、マークが付けられたマッチング処理単位の位置を基準にし50

て、その対話履歴データの近傍に、マッチング処理単位で述べられている内容の状況や条件等を含む表現パターンを述べた記述が存在しないか探索し、存在する場合は、その表現パターンを述べた記述部分を抽出する。

ここで、状況や条件等を含む表現パターンを述べた記述としては、例えば、「～だから」「～なので」のような理由を述べる表現と、「～したい」「できれば」「～を探している」のように目的・願望を述べる表現と、「××が～なった」「××を～した」のように状況を述べる表現、「～場合～必要」「ただし～必要」「～なければならない」のように条件を述べる表現、「××がよい」「～が気に入った」のように意見・評価を述べる表現などがある。これらの表現を記述した箇所を対話履歴データから見つけるためには、予め上記のような表現パターンのリストを用意しておき、対話履歴データ中で表現パターンに適合する記述部分を抽出する手法のような、既に知られているテキストからの情報抽出技術を利用することができる。10

図14は、表現パターン抽出部52において対話履歴データから状況・条件等を述べた記述部分を抽出するための表現パターンの例を示している。表現パターン抽出部52は、図14に示すような表現パターンにマッチした記述を含むマッチング処理単位を対話履歴データから抽出する。

表現パターンとの一致を取る際には、単純に文字列の一致を調べる手法であってもよいが、表現パターンに動詞や形容詞などの活用語が含まれている場合には、そうした表現パターン中の活用語の語尾を活用展開して、対話履歴データ中の表現との一致を調べる手法を用いてもよい。20

また、ある1つのマッチング処理単位が、複数種類の状況・条件等を述べた記述部分を含む場合には、表現パターンとの適合度の度合いに基づいて、何れか1つの記述部分を選択してもよいし、予め記述の種類に優先度を設定しておき、その優先度に従って、1つの記述部分を選択してもよい。また、1つに選択することなく、マッチング処理単位から、複数種類の状況・条件等を述べた表現パターンの記述部分を抽出するようにしてもよい。

図14に示すような表現パターンのリストを用意しておく以外の情報抽出技術としては、予めどの文が理由で、どの文が意見・評価なのか、など正解を付与した文書を多数学習用に準備し、単語nグラムなどの統計量をそうした学習用の文書に対して調べて機械学習を行い、その学習結果から新規の文書の種類も判定する手法がある。

これら情報抽出技術については、例えば、文献5（情報処理，Vol.45, No.4, June 2004, p563-568「情報抽出 - 情報を整理して提示する - 」関根聰（著））に記載されている技術が存在する。30

実際にどのような記述を、表現パターンを述べた記述として抽出するかは、本発明の実施の形態による質問応答データ編集装置100を使用する際の用途や目的、データベースに記録されている既存の対話履歴データや質問応答データの性質などに応じて予め定めておく。

また、基準の位置としたマッチング処理単位からどこまでの範囲をその近傍として、表現パターンを述べた記述を探索するかについても、本発明の実施の形態による質問応答データ編集装置100を使用する際の用途や目的、データベースに記録されている既存の対話履歴データや質問応答データの性質などに応じて予め定めておく。探索する近傍の範囲の例としては、基準位置としたマッチング処理単位から、対話履歴データを遡って、基準位置より以前にマークが付けられたマッチング処理単位の手前まで（マークが見つからない場合は対話履歴データの先頭まで）とする設定方法や、基準位置としたマッチング処理単位の前後の数発話（例えば5発話）の範囲を探索する、といった一定範囲を近傍とする設定方法などが考えられる。40

基準の位置としたマッチング処理単位の近傍で、表現パターンを述べた記述が見つかった場合は、それが「理由」なのか「目的・願望」なのか、など記述の種類を記録し、その記述の位置や記述内容、基準位置としたマッチング処理単位などの情報を組にして追加情報出力部56に送る。

図15は、表現パターン抽出部52によって、図14に示した表現パターンを用いて、

50

図9に示した対話履歴データ例1から表現パターン（状況・条件等）の記述を抽出した結果の例を示している。図中、丸い吹き出しで示している発話が、抽出された表現パターン（状況・条件等）の記述である。

図15で最初に示されている目的・願望の記述は、質問応答データ3-1の質問に類似しているマッチング処理単位を基準として前後2発話と設定された近傍範囲を探索した場合に「～したい」という表現パターンとの一致により抽出されたものである。基準としたマッチング処理単位に「質問応答データ3-1の質問に類似」というマークが付いているため、追加対象質問応答データ決定部51で決定された質問応答データは「質問応答データ3-1」であるから、この目的・願望の記述も、後述する追加情報出力部56にて、「質問応答データ3-1」用の追加情報としてまとめられる。

図15に示されている次の条件の記述は、質問応答データ3-2の応答に類似しているマッチング処理単位を基準として探索した際に「ただし～必要」という表現パターンとの一致により抽出されたものである。基準としたマッチング処理単位に「質問応答データ3-2の応答に類似」というマークが付いているため、追加対象質問応答データ決定部51で決定された質問応答データは「質問応答データ3-2」であるから、条件の記述も、後述する追加情報出力部56にて、「質問応答データ3-2」用の追加情報としてまとめられる。

質問応答バリエーション抽出部53は、マークが付けられたマッチング処理単位が、類似箇所検出部44で類似していると判定された質問応答データ（以降、この質問応答データを、基準質問応答データと称する）の質問部分に類似しているのか、それとも、応答部分に類似しているのかをまず調べる。

以下、質問部分に類似している場合と、応答部分に類似している場合とに分けて説明する。

(1) 質問部分に類似している場合

質問部分に類似している場合は、マークが付けられたマッチング処理単位の位置を基準にして、その対話履歴データの後方に（基準位置より後でなされた対話の方向に）向かって、マッチング処理単位毎に対話履歴データの内容を調べていく。

最初に、基準質問応答データの応答部分に類似しているマッチング処理単位が見つかった場合は、質問応答バリエーションが無い、として追加情報出力部56に送る。また、最初に基準質問応答データ以外の質問応答データの質問部分に類似しているマッチング処理単位が見つかった場合や、対話履歴データの最後まで何も見つからなかった場合も同様に、質問応答バリエーションが無い、として追加情報出力部56に送る。

最初に、基準質問応答データ以外の質問応答データの応答部分に類似しているマッチング処理単位が見つかった場合や、どの質問応答データとも類似していない（マークが付いていない）が何らかの応答を記述していると判断できるマッチング処理単位が見つかった場合には、それが基準質問応答データの応答データのバリエーションであると判断し、基準位置のマッチング処理単位と、見つかったマッチング処理単位の位置や応答内容を組にして、追加情報出力部56に送る。

ここで、あるマッチング処理単位が応答を記述しているかどうか判断するためには、「～して下さい」「～していただけようお願い」「～手順をお伝えします」「～を試してみて下さい」「～をお薦めします」など、予め応答によく使用される表現パターンのリストを用意しておき、こうした表現パターンと一定数以上文字列が一致したところを抽出する手法のような、前述したテキストからの情報抽出技術を利用することができる。

実際にどのような表現を、応答を記述した箇所として抽出するかは、本発明の実施の形態による質問応答データ編集装置100を使用する際の用途や目的、データベースに記録されている既存の対話履歴データや質問応答データの性質などに応じて予め定めておく。

(2) 応答部分に類似している場合

一方、基準位置のマッチング処理単位が基準質問応答データの応答部分に類似している場合には、その対話履歴データの前方に（基準位置より前でなされた対話の方向に）向かって、マッチング処理単位毎に対話履歴データの内容を調べていく。

10

20

30

40

50

最初に、基準質問応答データの質問部分に類似しているマッチング処理単位が見つかった場合は、質問応答バリエーションが無い、として追加情報出力部 56 に送る。また、最初に基準質問応答データ以外の質問応答データの応答部分に類似しているマッチング処理単位が見つかった場合や、対話履歴データの最初まで何も見つからなかった場合も同様に、質問応答バリエーションが無い、として追加情報出力部 56 に送る。

最初に、基準質問応答データ以外の質問応答データの質問部分に類似しているマッチング処理単位が見つかった場合や、どの質問応答データとも類似していない（マークが付いていない）が何らかの質問を記述していると判断できるマッチング処理単位が見つかった場合には、それが質問応答データ例 3 の質問データのバリエーションであると判断し、基準位置のマッチング処理単位と、見つかったマッチング処理単位の位置や質問内容を組にして、追加情報出力部 56 に送る。10

ここで、あるマッチング処理単位が質問を記述しているかどうか判断するためには、「～どうしたらいい」「～が分からない」「～を教えて」「どこに～」「どこへ～」「何を～すれば」「何が問題なの」「いくら～」など、予め質問によく使用される疑問の表現パターンのリストを用意しておき、こうした表現パターンと一定数以上文字列が一致したところを抽出する、前述したテキストからの情報抽出技術が利用できる。

実際にどのような表現を、質問を記述した箇所として抽出するかは、本発明の実施の形態による質問応答データ編集装置 100 を使用する際の用途や目的、データベースに記録されている既存の対話履歴データや質問応答データの性質などに応じて予め定めておく。

このように、質問応答バリエーション抽出部 53 では、基準位置のマッチング処理単位が類似している質問応答データの質問データのバリエーション、または、応答データのバリエーションを、基準位置の近傍の対話履歴データから抽出し、追加情報出力部 56 に送る。基準位置のマッチング処理単位が類似している質問応答データが複数存在している場合には、それぞれの質問応答データに対して上記の処理を行う。20

図 18 は、図 10 の対話履歴データ例 2 に対して図 17 のような質問応答データ例 4 とマッチングを取った際の、質問応答バリエーション抽出結果を示している。この例では、電子掲示板での 1 つの記事を 1 つのマッチング処理単位としている。また、電子掲示板では記事がスレッドなどに沿ってツリー上に管理されることがあるが、この例では記事が投稿された時間順に沿って、1 次元の順序をもつ対話履歴データと見なしている。

図 18 では、図中の上から 3 番目の記事が、使用されている単語の一致から類似箇所検出部 44 によって「質問応答データ 4 - 2 の応答に類似」とのマークが付けられている。よって、追加対象質問応答データ決定部 51 で、この 3 番目の記事のマッチング処理単位が調べられた際には、新規にインデックス情報を抽出・追加する追加対象の質問応答データとして質問応答データ 4 - 2 が決定され、この対話履歴データとマークされたマッチング処理単位の位置が、表現パターン抽出部 52 と、質問応答バリエーション抽出部 53 と、関連質問応答抽出部 54 とに送られる。30

質問応答バリエーション抽出部 53 では、基準としたマッチング処理単位に質問応答データの応答に類似したマークが付いているので、対話履歴データの前方を探索した結果、1 番目の記事が「質問の記述」であると判定され、質問応答データ 4 - 2 の質問データのバリエーションとして抽出される。なお、この例では 1 番目の記事中の「何が問題」「教えて」といった表現パターンから、この記事が「質問の記述」であると判定しているが、図 17 の対話履歴データ例 2 のように各記事に予めそれが「質問」であるのか「応答」であるのか情報が付与されている場合には、その情報を用いて各記事が「質問の記述」であるのかそれとも「応答の記述」であるかの判定を行ってもよい。40

図 17 の質問応答データ例 4 の質問応答データ 4 - 2 では、「マザーボードの BIOS アップデートが必要」になる事態として、質問部分の「内蔵ハードディスクを増設」した場合のみが記述されていたが、図 18 のような質問応答バリエーションの抽出結果から、実は「パソコンの内蔵ハードディスク（C ドライブ）を 160 GB に換装」した場合でも、BIOS アップデートが必要になることが分かる。よって、質問応答データ編集装置 100 の利用者は、後述する質問応答データ編集部 70 によってこの抽出結果を確認する50

とにより、質問応答データ4-2を更新することができる。

同様に、図21は、図19の対話履歴データ例3に対して図20のような質問応答データ例5とマッチングを取った際の、質問応答バリエーション抽出結果を示している。対話履歴データ例3は、電気店でのエアコン売り場での顧客と店員との対話を、店員が身に付けているマイクで自動的に記録して書き起こした対話内容を想定している。図19で、顧客と先頭に書かれている文が、顧客の発話部分であり、店員と先頭に書かれている文が、それに対する店員の発話部分である。この対話履歴データ例3では、対話履歴分割部42で各発話毎にマッチング処理単位に分割されているものとする。

図21では、図中の上から3番目の発話が、使用されている単語の一致から類似箇所検出部44によって「質問応答データ5-1の質問に類似」とのマークが付けられている。
10 よって、追加対象質問応答データ決定部51で、この3番目の発話のマッチング処理単位が調べられた際には、新規にインデックス情報を抽出・追加する追加対象の質問応答データとして質問応答データ5-1が決定され、この対話履歴データとマークされたマッチング処理単位の位置が、表現パターン抽出部52と、質問応答バリエーション抽出部53と、関連質問応答抽出部54とに送られる。

質問応答バリエーション抽出部53では、基準としたマッチング処理単位に質問応答データの質問に類似したマークが付いているので、対話履歴データの後方を探索した結果、上から6番目の店員の発話が「応答の記述」だと判定され、質問応答データ5-1の応答データのバリエーションとして抽出される。なお、この例では6番目の発話中の「お薦めします」という表現パターンから、この発話が「応答の記述」であると判定している。
20

図20の質問応答データ例5の質問応答データ5-1では、「エアコン購入時に、室外機の工事依頼があった場合」、「特別料金で、エアコン搬送時に設置工事が可能」であるとしか記されていない。しかし、図21のような質問応答バリエーションの抽出結果から、実は「専門の業者への発注」が必要となる場合があることが分かる。よって、質問応答データ編集装置100の利用者は、後述する質問応答データ編集部70によってこの抽出結果を確認することにより、質問応答データ5-1を更新することができる。以後は、店員が本実施の形態によって編集作成された質問応答データとしての応対マニュアルをチェックしていれば、間違って地下室用エアコンの室外機の工事依頼を受けてしまうミスを防ぐことができる。

関連質問応答抽出部54は、対話履歴データ中、ある質問応答データ（の類似箇所）の次に別の質問応答データ（の類似箇所）が現れる、という関係を抽出し、連続して出現する質問応答データを関連質問応答データとして追加情報出力部56に送る。
30

具体的には、マークが付けられたマッチング処理単位が類似している質問応答データ（基準質問応答データとする）が、その質問応答データの応答部分に類似していた場合、マークが付けられたマッチング処理単位を基準位置にして、対話履歴データの後方を調べる。後方に、基準質問応答データ以外の質問応答データ（質問応答データ2とする）の質問部分に類似しているマッチング処理単位が最初に見つかったとき、基準質問応答データと質問応答データ2とを関連質問応答データとして、追加情報出力部56に送る。それ以外の場合は、何もしない。基準質問応答データや質問応答データ2にあたる質問応答データが
40

複数ある場合には、それぞれの質問応答データの組み合わせを全て関連質問応答データとして追加情報出力部56に送る。

図13に示す例では、「質問応答データ3-1の応答に類似」しているマッチング処理単位の後方に、「質問応答データ3-2の質問に類似」しているマッチング処理単位が見つかる。よって、質問応答データ3-1と質問応答データ3-2との組み合わせが関連質問応答データとして、追加情報出力部56に送られる。

新規質問応答抽出部55は、追加対象質問応答データ決定部51から受け取った、マークの付いていないマッチング処理単位が、既存の質問応答データではなく、新規の質問応答データであるかどうかを判定する。

新規質問応答抽出部55は、まず受け取ったマークの付いていないマッチング処理単位
50

が、質問を記述しているかどうか判断する。質問を記述している場合には、さらに対話履歴データの後方を調べ、マークの付いているマッチング処理単位が見つかるまでに、マークの付いていないマッチング処理単位で、応答を記述しているものが見つかった場合、最初のマッチング処理単位と応答記述が見つかったマッチング処理単位との、位置と記述内容を組を、新規の質問応答データの候補として、追加情報出力部 5 6 に送る。それ以外の場合は、何も追加情報出力部 5 6 に送らない。

質問や応答を記述しているマッチング処理単位であるかどうかを判断するためには、質問応答バリエーション抽出部 5 3 で使用した、既に知られているテキストからの情報抽出技術を利用することができる。

図 2 1 に示す対話履歴データ例 3 では、下から 3 番目の発話がマークが付いていないマッチング処理単位であるため、新規質問応答抽出部 5 5 に送られる。新規質問応答抽出部 5 5 では、まず受け取った下から 3 番目の発話が「どこに」という表現パターンの一一致から、「質問の記述」であると判断される。さらに対話履歴データの後方を調べた結果、マークの付いているマッチング処理単位が見つかるまでに、下から 2 番目の発話が、マークの付いていないマッチング処理単位で、表現パターンの一一致から「応答の記述」であると判定される。よって、見つかった下から 3 番目と下から 2 番目の発話の組が、新規の質問応答データの候補として追加情報出力部 5 6 に送られる。

追加情報出力部 5 6 は、表現パターン抽出部 5 2 と、質問応答バリエーション抽出部 5 3 と、関連質問応答抽出部 5 4 とから送られてきた情報（状況・条件等を含む表現パターンの記述、質問応答バリエーション、関連質問応答データ）を受け取り、追加対象質問応答データ決定部 5 1 で決定した質問応答データのインデックス情報としてまとめて、追加情報蓄積部 6 0 に送る。複数の異なる質問応答データを対象とした情報が送られてきた場合は、それぞれを対象とする質問応答データ毎にまとめて追加情報蓄積部 6 0 に送る。

特に、1 つの対話履歴データの中で、ある 1 つの質問応答データに対して、表現パターン等の記述を抽出した情報と、質問応答バリエーションとが見つかった場合は、既存の質問応答データと異なる質問応答バリエーションが生じた理由が、その表現パターンの記述に述べられている可能性があるため、表現パターンの記述を抽出した情報と質問応答バリエーションとを組み合わせてインデックス情報としてまとめ追加情報蓄積部 6 0 に送るようにしてよい。

例えば、図 2 1 において下から 6 番目の発話が、表現パターン抽出部 5 2 で上から 3 番目の発話の「条件の記述」として抽出されたとする。一方、質問応答バリエーション抽出部 5 3 の説明で述べたように、上から 3 番目の発話に対する上から 6 番目の発話が、質問応答データ 5 - 1 の応答バリエーションとして抽出される。このような場合、下から 6 番目の発話を、この質問応答データ 5 - 1 の応答バリエーションが生じるための条件として、それらを組み合わせてインデックス情報にまとめても良い。

また、新規質問応答抽出部 5 5 から受け取った新規の質問と応答の記述内容は、新しく追加すべき質問応答データの候補として、質問応答の記述内容と、どの対話履歴データのどの位置から抽出されたかの情報を組にして、追加情報蓄積部 6 0 に送る。

追加情報蓄積部 6 0 は、追加情報出力部 5 6 から送られてきた情報を一時的に蓄積・記録する。蓄積する際には、インデックス情報が対象とする質問応答データ毎に分けて記録しても良い。また、インデックス情報の種類毎に分けて記録しても良い。新規に追加すべき質問応答データの候補は、インデックス情報とは別に記録する。

追加情報蓄積部 6 0 に蓄積した情報は、後述する質問応答データ編集部 7 0 からの要求に従って出力される。また、質問応答データ編集部 7 0 からの要求に従って、蓄積した情報の一部ないし全部を消去される。

図 1 6 は、対話履歴データ例 1 が、図 1 5 のように、表現パターン抽出部 5 2 と関連質問応答抽出部 5 4 とで処理された結果、インデックス情報が追加された例を示している。図 1 6 においては、図 1 2 に比べて、質問応答データ 3 - 1 と質問応答データ 3 - 2 とがそれぞれ関連質問応答として結びつけられている。これによって、一方の質問応答データを参照する際に、他方を関連質問応答データとして提示したり、一方から他方への検索が

10

20

30

40

50

可能となる。また、表現パターン抽出部 52 で抽出された表現パターンの記述が、インデックス情報として追加されている。

本発明の実施の形態による質問応答データ編集装置 100 の利用者は、後述する質問応答データ編集部 70 の機能を用いることにより、このインデックス情報を提示して確認し、必要ならば適宜もとの質問応答データに加筆・修正するなどの更新を行うことができる。また、これらの記述に用いられている表現で、以後質問応答データ 3-1 や質問応答データ 3-2 を検索することが可能となる。

質問応答データ編集部 70 は、質問応答データを編集・追加処理(更新)するための機能を有する装置であり、追加情報蓄積部 60 と質問応答データベース 30 にアクセスすることができる。この質問応答データ編集部 70 は、追加情報蓄積部 60 に蓄積されているインデックス情報や新規の質問応答データの候補を画面上に表示出力する機能、表示出力されたインデックス情報や新規の質問応答データの候補の内容を修正する機能、インデックス情報を質問応答データベース 30 の質問応答データインデックス 32 に追加登録し、新規の質問応答データの候補を質問応答データベース 30 の質問応答データ記録領域 31 に追加登録する機能を有する。10

ユーザからの要求に従って、追加情報蓄積部 60 に蓄積されているインデックス情報が画面上に表示出力されると、ユーザは質問応答データ編集部 70 を操作することで、表示されたインデックス情報を確認し、必要に応じて適宜その内容の修正作業や、質問応答データベース 30 のインデックス記録領域 32 に対して追加登録を行う。

また、追加情報蓄積部 60 に蓄積されている新規の質問応答データの候補を確認して、必要に応じて適宜その内容の修正作業や、新規の質問応答データとして質問応答データベース 30 の質問応答データ記録領域 31 に追加登録を行う。20

図 22 によって、質問応答データ編集部 70 によって画面上に提示される処理結果の一覧(インデックス情報や新規の質問応答データの表示)の例を示す。

図 22 に示す表示例においては、追加情報抽出動作において対話履歴データから抽出された追加情報(インデックス情報や新規の質問応答データ)が、質問応答データ毎にリストアップされている。

「対話履歴データ」の欄には、抽出された追加情報がどの対話履歴データから得られたもののかを示している。

この欄に設けられている「参照」ボタンをマウスなどでクリックすることで、元の対話履歴データが他の画面(ウインドウ)に表示される。30

また、各行に設けられている「編集」ボタンをマウスなどでクリックすると、個々の抽出された追加情報に対して修正を加えるための編集画面が表示される。

右端の各行に対応して設けられている「採用」の欄は、対応する追加情報を質問応答データベース 30 に登録するかどうかを指定するための欄であり、ユーザが登録すべきであると判断した追加情報に、チェックを付けることで登録対象であることを指定する。

画面右下の「登録」ボタンを押すと、「採用」の欄にチェックされている追加情報が、質問応答データベース 30 の質問応答データ記録領域 31 又はインデックス記録領域 32 に登録される。

質問応答データベース 30 に登録された内容は、追加情報蓄積部 60 から自動的に削除するよう設定しておいてよい。40

また、ユーザが追加情報蓄積部 60 に蓄積されている情報(インデックス情報や新規の質問応答データの候補)を確認することなく、予め設定した条件を満たす情報である場合には、自動的に質問応答データベース 30 に登録するようにしてもよい。

自動的に質問応答データベース 30 に登録するための条件としては、1つの質問応答データに対する同種のインデックス情報が一定件数を超えて抽出された場合といった、設定が考えられる。

なお、本実施の形態による質問応答データ編集装置 100 を構成する、マッチング条件入力部 10、対話履歴データベース 20、質問応答データベース 30、マッチング部 40、追加情報抽出部 50、追加情報蓄積部 60、質問応答データ編集部 70 については、各50

部の機能を実現するプログラム（アプリケーション）を、CD-ROMやフレキシブルディスクなどの機械読み取り可能な記録媒体や、インターネットなどのネットワークを介して質問応答データ編集装置100を構成するコンピュータ装置に提供し、当該プログラムをコンピュータ装置上で実行することで、コンピュータ装置に上記各部の機能を持たせるように構成することもできる。

次に、本発明の実施の形態による質問応答データ編集装置100の動作について、図4から図7のフローチャートに沿って説明する。

本発明の実施の形態による質問応答データ編集装置100における動作は、大きく分けて、図4に示す追加情報蓄積動作と、図7に示す質問応答データ編集動作との2つの動作からなる。この2つの動作は、それぞれ独立に実行することができる。

10

まず、図4のフローチャートに沿って、追加情報蓄積動作について説明する。

本発明の実施の形態による質問応答データ編集装置100の追加情報蓄積動作では、ユーザがマッチング条件入力部10によって、対話履歴データベース20に記録されている対話履歴データのうち、今回処理対象とする対話履歴データの条件を設定し入力する（ステップA401）。

また質問応答データベース30に記録されている質問応答データのうち、今回処理対象とする質問応答データの条件を入力する（ステップA402）。このステップA401とA402はどちらを先に処理してもよい。

ついで、マッチング部40において、ステップA401とA402で入力された条件を満たす対話履歴データと質問応答データとのマッチング動作を行う（ステップA403）。

20

このマッチング動作については、図5のフローチャートを用いてより詳細に説明する。

マッチング部40のマッチング動作では、まず条件にあった対話履歴データを読み込み（ステップA501）、次に読み込んだ対話履歴データを適切なマッチング処理単位に分割する（ステップA502）。また条件に一致した質問応答データを読み込む（ステップA503）。ここでステップA501とA502とはこの順番で処理しなければならないが、ステップA503は、ステップA501とA502との処理とは独立に実行してよい。

次に、条件を満たす全ての対話履歴データをマッチング処理単位毎に調べ、条件を満たす質問応答データの質問部分または応答部分に類似する箇所がないか判定する。その結果、対話履歴データ中に類似箇所が検出された場合には、そのマッチング処理単位がどの質問応答データの質問部分または応答部分に類似しているのか、後からわかるようマッチング処理単位にマーク付けする（ステップA504）。これでマッチング動作が終了する。

30

マッチング動作の後は、このマッチング動作で得られた対話履歴データ中の既存の質問応答データ類似箇所の検出結果から、追加情報抽出部50で、新規に追加すべき質問応答データや、質問応答データのインデックス情報を抽出する、追加情報抽出動作を行う（ステップA404）。この追加情報抽出動作については、図6のフローチャートを用いてより詳細に説明する。

追加情報抽出部50による追加情報抽出動作では、追加情報抽出動作が未処理の対話履歴データを1つ選択する（ステップA601）。次に選択した対話履歴データのマッチング処理単位を順にチェックし（ステップA602）、そのマッチング処理単位にマッチング動作の結果のマークが付いているかどうか確認する（ステップA603）。

40

マークが付いている場合は、そのマークから抽出する情報の追加対象となる質問応答データを決定する（ステップA604）。

ついで、マークが付いているマッチング処理単位を基準に対話履歴データの近傍に表現パターンの記述があればそれを抽出する（ステップA605）。

また、マークが付いているマッチング処理単位が類似している質問応答データが、質問部分と応答部分と揃って対話履歴データ中のその箇所近傍に存在しているか確認し、どちらか一方だけが別の応答または質問と近接して存在している場合には、その応答または質問の記述をもとの質問応答データのバリエーションとして抽出する（ステップA606）

50

。

さらに対話履歴データ中、マークが付いているマッチング処理単位の箇所で、2つの質問応答データが連続して存在すれば、それらの質問応答データを関連質問応答として抽出する（ステップA607）。

ここでステップA605、ステップA606、ステップA607はそれぞれ独立に実行することができる。本発明の実施の形態の用途や目的に応じては、ステップA605、ステップA606、ステップA607の一部だけを実行するようにしてもよい。

一方、ステップA603でマークが付いていなかった場合には、対話履歴データのそのマッチング処理単位の位置から先に、既存の質問応答データとは類似していない質問の記述と応答の記述が存在しないかを調べて、存在している場合には、新規の質問応答データの候補として抽出する（ステップA608）。

ステップA605～A607の処理、またはステップA608の処理が完了すると、それらの処理で抽出した追加情報（もし存在すれば）を取りまとめる（ステップA609）。

次に、現在選択中の対話履歴データ中にステップA603～A609の手順が未処理のマッチング処理単位が残っているかどうかチェックし、残っている場合はステップA602に戻る。マッチング処理単位が残っていない場合はステップA611に進む（ステップA610）。

現在選択中の対話履歴データに未処理のマッチング処理単位が残っていない場合は、他に追加情報抽出動作が未処理の対話履歴データが残っていないかどうかチェックし、残っている場合はステップA601に戻る。残っていない場合は追加情報抽出動作を終了する（ステップA611）。

追加情報抽出動作の後は、そこで追加された情報を追加情報蓄積部60に蓄積・記録する（ステップA405）。このステップA405の処理が終了すると、本発明の実施の形態の質問応答データ編集装置100における動作のうち、追加情報蓄積動作が終了する。

次に図7のフローチャートに沿って、質問応答データ編集部70による質問応答データ編集動作について説明する。

質問応答データ編集動作では、追加情報蓄積動作で追加情報蓄積部60に蓄積された、既存の質問応答データのインデックス情報や新規の質問応答データの候補を、本発明の実施の形態のユーザに提示する（ステップA701）。このインデックス情報や新規の質問応答データの候補は、図22に示した例のように、画面上に表示されてユーザに提示される。

ユーザは、ステップA701で提示されたインデックス情報や新規の質問応答データの候補について、質問応答データベース30に登録する情報を選択し、必要に応じてその内容の修正を行う（ステップA702）。

そして、ステップA702で選択されたインデックス情報や新規の質問応答データ又は修正されたインデックス情報や新規の質問応答データが、質問応答データベース30に登録される（ステップA703）。

ついで、ステップA702、A703の作業を全ての追加情報について終了したかどうか確認し、まだ追加情報が残っている場合は、ステップA701に戻る。追加情報が残っていない場合は、質問応答データ編集動作を終了する（ステップA704）。

ここで、ステップA702ではユーザが質問応答データベース30へ登録する追加情報を選択するように説明したが、追加情報蓄積動作で追加情報蓄積部60に蓄積された追加情報のうち、質問応答データベース30に追加・登録する基準を別途定めておき、質問応答データ編集動作の一部または全部を自動的に登録するようにしてもよい。

以上のように、追加情報蓄積動作と質問応答データ編集動作とを、それぞれ独立に1回以上処理することで、本発明の実施の形態の質問応答データ編集装置における動作全体の終了となる。

以上説明した本発明の実施の形態の質問応答データ編集装置100によれば、以下のような効果が達成される。

10

20

30

40

50

第1に、蓄積された対話履歴データから質問応答データに類似した箇所を検出し、その箇所の近傍から状況・条件等を含む表現パターンの記述を抽出して、質問応答データのインデックス情報として記録するため、質問応答データが実際の対話中で使用される際の状況・条件等を含む表現パターンを質問応答データのインデックスとして登録することができるようになる。

抽出した状況・条件等を含む表現パターンは、もとの質問応答データと結びつけて記録することで、質問応答データを検索する際のインデックス情報として利用できる。また質問応答データを編集・追加する際に、参考情報として利用できる。抽出された状況・条件等を含む表現パターンをカウントしておき、例えば、ある質問応答データが、ある状況・条件下でよく使用されると分かった場合には、その状況・条件を盛り込んだ質問応答データを新たに追加する、などの利用法がある。10

第2に、蓄積された対話履歴データから質問応答データの質問部分または応答部分に類似した箇所を検出し、その箇所の近傍からもとの質問応答データとは類似しない、応答または質問の記述を抽出して、質問応答データのインデックス情報として記録するため、質問応答データが実際の対話中で使用される際の質問部分や応答部分の表現のバリエーションを抽出することができるることである。

抽出した表現バリエーションは、もとの質問応答データと結びつけて記録することで、質問応答データを検索する際のインデックス情報として利用できる。また質問応答データのバリエーションを、質問応答データが使用される際の状況・条件等を含む表現パターンと組にして記録することで、質問応答データを状況・条件等で場合分けして細分化する際に利用できる。20

第3に、蓄積された対話履歴データの中に連続して出現する2つの質問応答データの類似箇所を検出し、それらの質問応答データを関連質問応答データとしてインデックス情報に記録するため、質問応答データが実際の対話中で使用される際に、同一の対話中に連続して現れやすい質問応答データを抽出することができるることである。

こうした関連性の高い質問応答データは、それぞれをリンクしてお互いのインデックス情報に記録することで検索性や閲覧性が高まる。また、質問応答データをグループ分けして表示する際などにも、グループ分けの基準情報として利用できる。

第4に、蓄積された対話履歴データの中から、既存の質問応答データには類似しない質問と応答の連続した記述を抽出し、その質問と応答の記述を新規の質問応答データの候補として記録するため、対話履歴データから、既存の質問応答データとは異なる新規の質問応答データを抽出することができる。新規の質問応答データを、新たに質問応答データベースに編集・追加することで、質問応答データを充実させることができる。30

【産業上の利用可能性】

【0006】

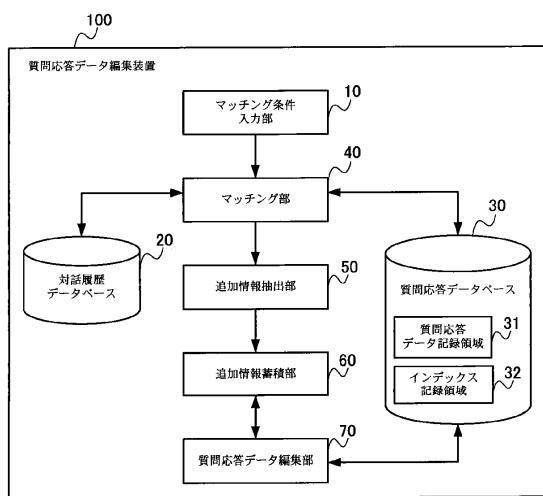
本発明によれば、コンタクトセンターやコールセンターなどの業務で使用するFAQや質問応答集を編集・追加する質問応答データ編集装置や、質問応答データ編集装置をコンピュータに実現するためのプログラムといった用途に適用できる。

またセールスや受付で使用する応対マニュアルを編集・追加する装置や、応対マニュアルの編集・追加機能をコンピュータ上に実現するためのプログラムといった用途にも適用できる。40

さらに、掲示板やチャットなどのやりとりからFAQや質問応答集を編集・追加する質問応答データ編集装置や、質問応答データ編集装置をコンピュータに実現するためのプログラムといった用途にも適用できる。

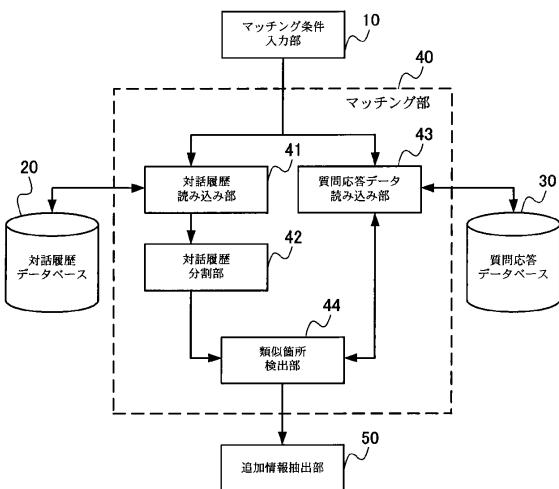
【図1】

図1



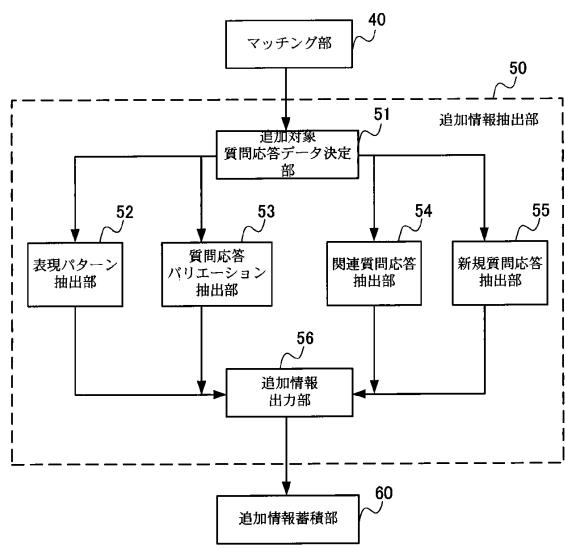
【図2】

図2



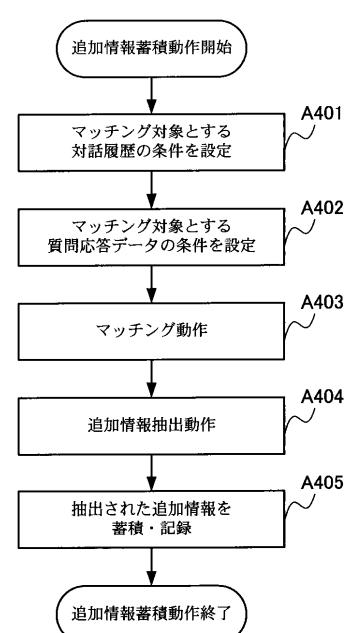
【図3】

図3



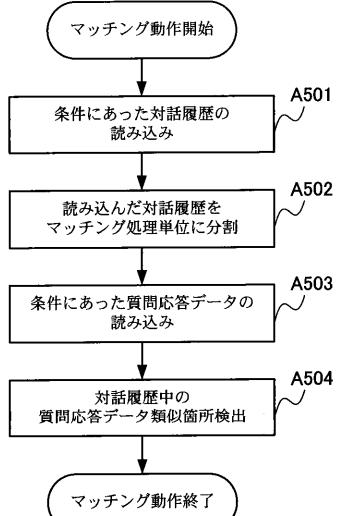
【図4】

図4

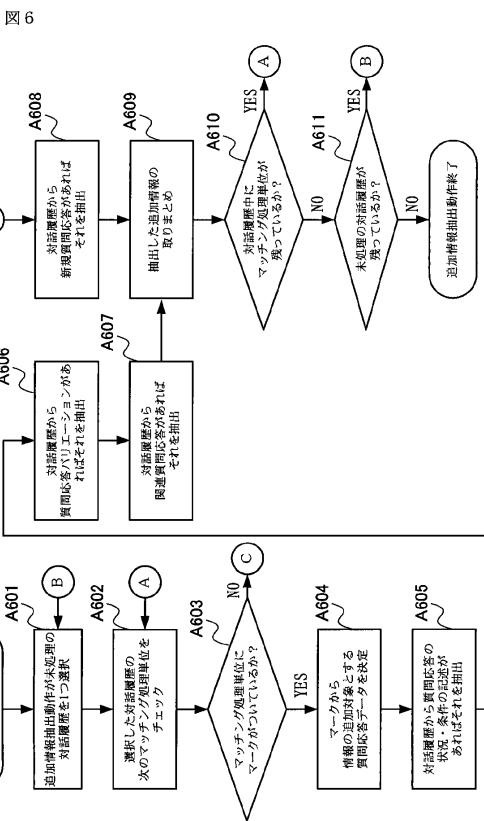


【図5】

図5

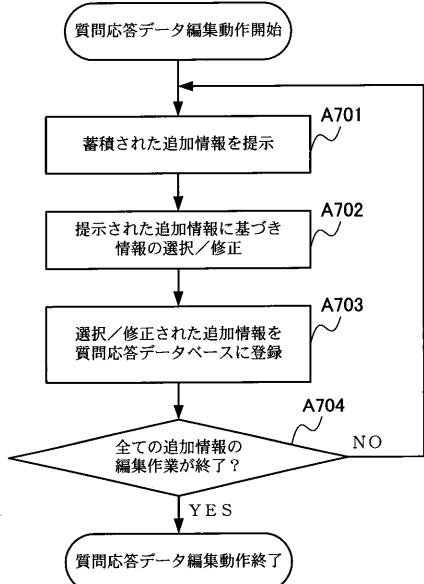


【図6】



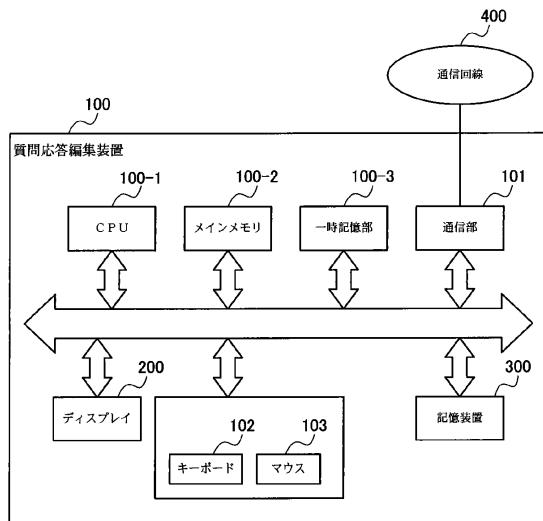
【図7】

図7



【図8】

図8



【図 9】

図 9

対話履歴データ例 1 (携帯電話操作に関する問い合わせ)

質問：お電話ありがとうございます。NACお客様センターの中川です。

OP：お電話ありがとうございます。NACお客様センターの中川です。

CS：すいません、ちょっと携帯電話について聞きたいんですけど。

OP：はい、携帯電話について聞きたいんですね。ご利用の携帯電話の機種名はお分かりですか？

CS：えっと、N1111です。

OP：N1111ですね、どういったことでしょうか？

CS：お待たせいたしまして、ボタンが壊れましたので、ダイヤルロックの操作方法についてです。

OP：携帯電話を防ぐための、ダイヤルロックの操作方法についてですね。少々お待ちください。

CS：はい。

OP：操作を防ぐための、ダイヤルロックの操作についてです。

CS：えっと、携帯電話についているボタンを、3秒以上押し続けると、ダイヤルロックされます。

OP：お待たせしました。

CS：えっと、機のボタン…あれ、ボタンつて2つあるけどどちらですか、口に近いほう？

OP：サイドボタンですね。そうですね、通話口に近いほうのボタンになります。

CS：あ、できました。よかったです。

OP：ありがとうございます。ありがとうございます。

CS：ありがとうございます。

OP：ありがとうございます。3秒以上サイドボタンを押してください。暗証番号が設定されている場合には、さらに4桁の暗証番号を入力してください。

CS：わかりました。

OP：あと、ダイヤルロック中にも電話に出ることはできるのですか？

CS：あ、いい、通話ボタンを押すことで、電話に出ることができます。

OP：はい、どちらから電話をかけるにはダイヤルロックを一度解除する必要があります。

CS：なるほど、わかりました。

OP：お問い合わせは以上でよろしいでしょうか？

CS：はい、助かりました。

OP：お問い合わせありがとうございました。私、NACお客様センターの中川が承りました。

【図 10】

図 10

対話履歴データ例 2 (電子掲示板での質問)

質問： ハードディスクがよめない
ななさん 2004年9月11日

先日、家のパソコンの内蔵ハードディスク（Cドライブ）を160Gに換装すると、OSが立ち上がらなくなってしまいました。ハードディスクの中身はディスクイメージコピーソフトでコピー済みです。何が問題なのでしょうか。ヒントでも教えてください。

返答： ハードはなし？（Re：ハードディスクがよめない）
イオさん 2004年9月12日

こういう質問には、機種名、環境等を書いてください。

返答： いつものマザーボードの問題じゃないの？
kakashiさん 2004年9月12日

マザーボードには、BigDriveとよばれる大容量のハードディスクに対応していないものがあるから、調べてみたら、BIOSアップデートでうまくいく場合もあるよ。

その他： ありがとうございます。
ななさん 2004年9月11日

kakashiさんのおっしゃるとおりでした。BIOSアップで立ち上がるようになりました。ちなみにマシンはNACのPC-98000です。今度から環境も書くようにします。トイオさん

【図 11】

図 11

質問応答データ例 3 (携帯電話の操作に関する質問応答集)

・質問応答データ 3-1 質動作を防ぐためのダイヤルロックをするにはどうすればよいですか？

応答：携帯電話の横にあるロックとかかれたサイドボタンを、3秒以上押し続けてください。

・質問応答データ 3-2 携帯電話を押すことで、通常通り、電話に出ることができますか？

応答：電話ボタンを押すことでの電話が出ます。

回答：通話ボタンを押すことで、通常通り、電話に出ることができます。

【図 12】

図 12

質問応答データ例 3 のインデックス

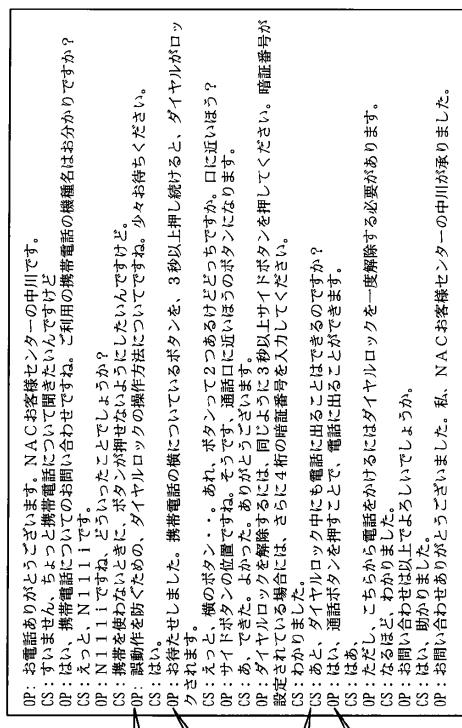
- ・質問応答データ 3-1 のインデックス
検索用キーワード：
関連インデックス：(検索用キーワード)
表現パターン (状況・条件) インデックス：
なし
- ・質問応答データ 3-2 のインデックス
検索用キーワード：
関連インデックス：(検索用キーワード)
表現パターン (状況・条件) インデックス：
なし

(27)

JP 4924950 B2 2012.4.25

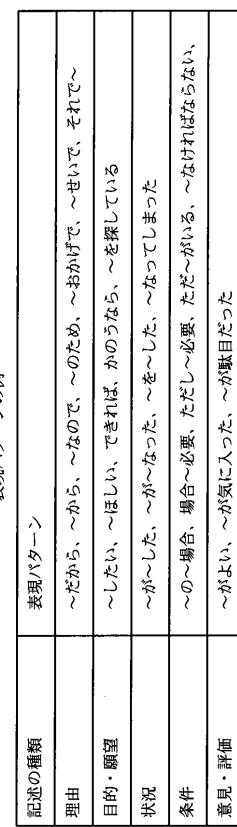
【図13】

図 1 3



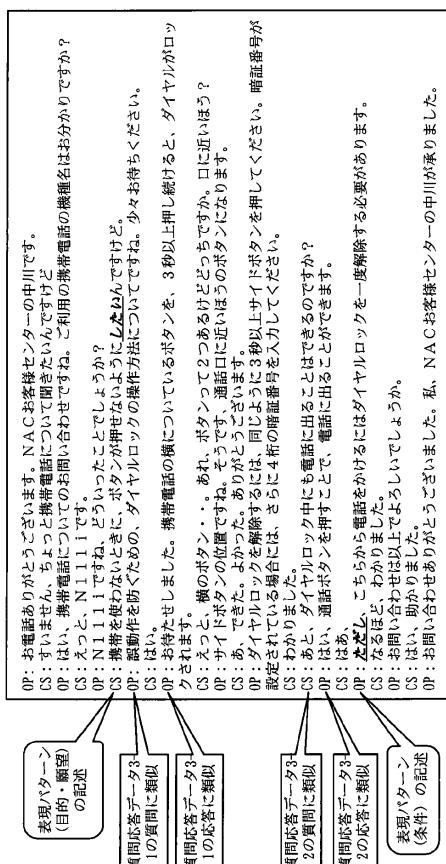
【図 1-4】

図 1



(图 15)

図 1.5



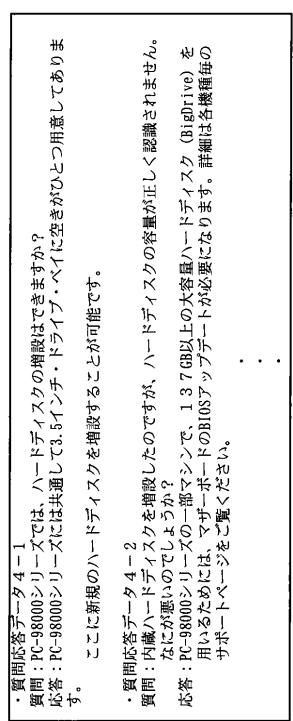
【図 16】

图 1 (a)

- ・質問応答データ3-1のインデックス
検索用インデックス：
(検索用キーワード)
関連インデックス：
(関連質問応答)
- 表現バーナインデックス：
(目的・願望)
- ・質問応答データ3-2のインデックス
検索用インデックス：
(検索用キーワード)
関連インデックス：
(関連質問応答)
- 表現バーナインデックス：
(目的・願望)
- ・質問応答データ3-3のインデックス
検索用インデックス：
(検索用キーワード)
関連インデックス：
(関連質問応答)
- 表現バーナインデックス：
(目的・願望)

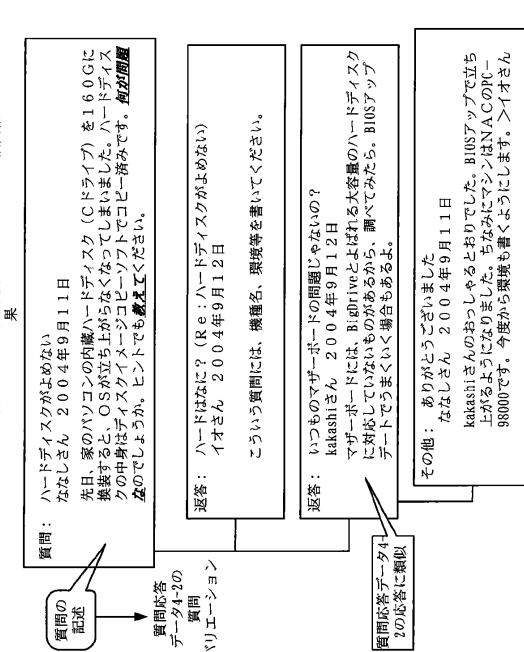
【図 17】

図 17



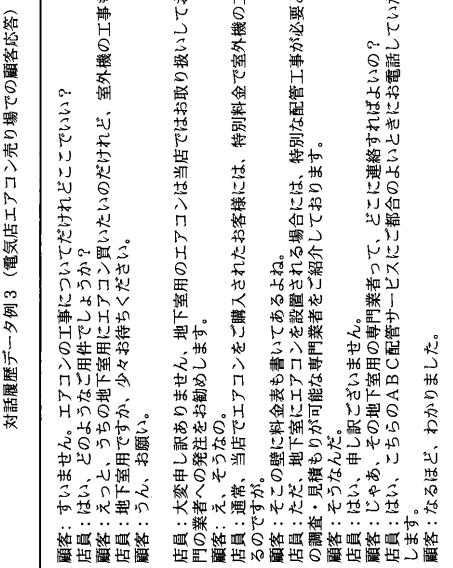
【図 18】

図 18



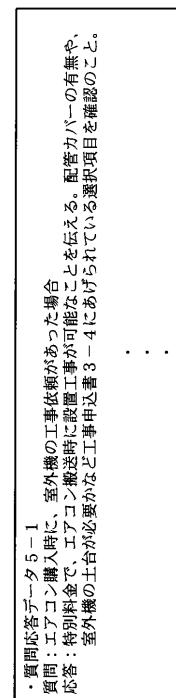
【図 19】

図 19



【図 20】

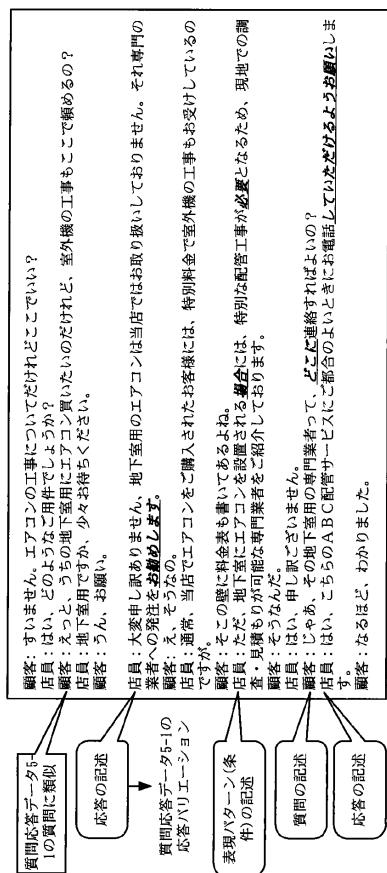
図 20



【図 21】

図 21

对话履歴データ例 3における追加情報抽出部の処理結果



【図 22】

図 22

| <処理結果一覧> | | 対話履歴データ |
|----------------|---------------------------|---|
| 質問応答データ 3-1 | 質問応答データ (目的、範囲) の記述 | 2010/01/24-0p41-13 マッチング処理単位6 [参照] |
| 質問応答データ 3-1 | 質問応答データ 開始 | 2010/01/24-0p41-13 マッチング処理単位15 [参照] |
| 質問応答データ 3-2 | 質問応答データ (条件)の記述 | 2010/01/24-0p41-13 マッチング処理単位18 [参照] |
| 質問応答データ 5-1 | 表現パターン (条件)の記述 | 2010/01/24-0p41-13 マッチング処理単位18 [参照] |
| 質問応答データ 5-1 | 回答 | 2010/01/24-0p41-13 マッチング処理単位6 [参照] |
| 質問応答データ 5-1 | バリエーション | 2010/01/24-0p41-13 マッチング処理単位6 [参照] |
| 質問応答データ 5-1 | 表現パターン (条件)の記述 | 2010/01/24-0p41-13 マッチング処理単位10 [参照] |
| 新規 質問応答データ | 質問応答データの 候補 | 質問：「じゃあ、その地下室用の専門業者って、どこに連絡すればよいの？」 回答：「はい、こちらのABC配管サービスにご都合のよいお電話していただけます。」 |

登録 キャンセル

フロントページの続き

審査官 鈴木 和樹

(56)参考文献 特開2004-178123(JP,A)

特開2003-263443(JP,A)

特開2002-230015(JP,A)

特開2001-134575(JP,A)

特開2000-339314(JP,A)

那須川哲哉, コールセンターにおけるテキストマイニング, 人工知能学会誌, 日本, (社)人工

知能学会, 2001年 3月 1日, 第16巻, 第2号, p. 219-225

太田竜男、外2名, 質問応答システムの構築 - 情報共有 / 再利用の技術 -, FUJITSU, 1

994年12月31日, 第45巻, 第5号, p. 436-442

(58)調査した分野(Int.Cl., DB名)

G06F 17/30