(19)日本国特許庁(JP)

(12)特許公報(B2)

(11)特許番号
特許第7305077号
(P7305077)

(45)発行日　令和5年7月7日(2023.7.7)

(24)登録日　令和5年6月29日(2023.6.29)

(51)国際特許分類
Ｇ０６Ｆ　16/332(2019.01)

ＦＩ
Ｇ０６Ｆ　16/332

請求項の数　15　（全24頁）

(21)出願番号　特願2023-501746(P2023-501746)
(86)(22)出願日　令和3年2月25日(2021.2.25)
(86)国際出願番号　PCT/JP2021/007002
(87)国際公開番号　WO2022/180721
(87)国際公開日　令和4年9月1日(2022.9.1)
審査請求日　令和5年1月17日(2023.1.17)
早期審査対象出願

(73)特許権者　000006013
三菱電機株式会社
東京都千代田区丸の内二丁目７番３号
(74)代理人　100116964
弁理士　山形 洋一
(74)代理人　100120477
弁理士　佐藤 賢改
(74)代理人　100135921
弁理士　篠原 昌彦
(74)代理人
半田 淳一
(74)代理人　100203677
弁理士　山口 力
(72)発明者　斉藤 辰彦
東京都千代田区丸の内二丁目７番３号
最終頁に続く

(54)【発明の名称】　情報処理装置、要約文出力方法、及び要約文出力プログラム

(57)【特許請求の範囲】
【請求項１】
対話内容を示すテキストデータと、過去の報告書に基づいて作成された複数の文を含む知識情報とを取得する取得部と、
形態素解析を用いて、前記テキストデータを解析する解析部と、
前記形態素解析の結果を用いて、前記テキストデータに含まれている複数の文の中から、１つの文を、第１の重要文として、抽出する重要文抽出部と、
前記第１の重要文に基づいて、クエリを作成し、前記知識情報の中から、前記クエリにより得られる文を検索し、予め設定された方法で、検索された複数の文のそれぞれのスコアを算出する作成検索算出部と、
前記第１の重要文と、検索された複数の文のそれぞれとに基づいて、前記第１の重要文と、検索された複数の文のそれぞれとが類似している度合である複数の類似度を算出し、前記複数の類似度に基づいて、検索された複数の文のそれぞれのスコアを更新する算出更新部と、
複数の文のそれぞれのスコアに基づいて、検索された複数の文のうちの１つの文を、要約文として、選択する選択部と、
前記要約文を出力する出力部と、
を有する情報処理装置。
【請求項２】
前記重要文抽出部は、前記テキストデータに含まれている複数の文の中から、複数の文

を、複数の重要文として、抽出し、
　前記作成検索算出部は、前記複数の重要文のうちの前記第１の重要文に基づいて、前記クエリを作成する、
　請求項１に記載の情報処理装置。
【請求項３】
　予め設定された方法で、前記テキストデータの中から不要語を削除する不要語削除部をさらに有する、
　請求項１又は２に記載の情報処理装置。
【請求項４】
　前記作成検索算出部は、前記第１の重要文の中の連続する単語を語彙化することにより、前記クエリを作成する、
　請求項１から３のいずれか１項に記載の情報処理装置。
【請求項５】
　前記作成検索算出部は、前記第１の重要文の中の否定語と、前記否定語の直前の動詞とを語彙化することにより、前記クエリを作成する、
　請求項４に記載の情報処理装置。
【請求項６】
　前記作成検索算出部は、前記第１の重要文の中の数詞と、前記数詞の後の単位とを語彙化することにより、前記クエリを作成する、
　請求項４又は５に記載の情報処理装置。
【請求項７】
　前記作成検索算出部は、前記第１の重要文と、前記テキストデータの中の前記第１の重要文の前文と後文とのうちの少なくとも１つとに基づいて、前記クエリを作成する、
　請求項１から６のいずれか１項に記載の情報処理装置。
【請求項８】
　前記作成検索算出部は、前記第１の重要文に含まれている単語の類義語を前記クエリに含める、
　請求項１から７のいずれか１項に記載の情報処理装置。
【請求項９】
　前記算出更新部は、前記複数の類似度のうちの１つの類似度を算出する場合、検索された複数の文のうちの１つの文に含まれる単語の集合から前記第１の重要文に含まれる単語の集合を引いた差集合の要素数が大きい場合にペナルティを与えることにより、集合同士の類似度を算出する、
　請求項１から８のいずれか１項に記載の情報処理装置。
【請求項１０】
　前記知識情報は、前記過去の報告書の中から非文が削除されることにより作成された情報である、
　請求項１から９のいずれか１項に記載の情報処理装置。
【請求項１１】
　カテゴリ推定部をさらに有し、
　前記知識情報に含まれる複数の文のそれぞれには、カテゴリが対応付けられており、
　前記取得部は、単語に基づいて、カテゴリを推定するカテゴリ推定モデルを取得し、
　前記カテゴリ推定部は、前記形態素解析により得られた単語と、前記カテゴリ推定モデルとを用いて、前記対話内容のカテゴリを推定し、
　前記作成検索算出部は、推定されたカテゴリと前記クエリを用いて、前記知識情報に対して検索を行う、
　請求項１から１０のいずれか１項に記載の情報処理装置。
【請求項１２】
　前記取得部は、複数の述語である複数の単語のそれぞれと、複数の単語のそれぞれの関係性を示す複数の関係情報のそれぞれとの対応関係を示す情報である補助情報を取得し、

前記作成検索算出部は、前記第１の重要文に対して形態素解析を実行することで得られた複数の品詞付単語の中から、述語になれる名詞の単語又は述語の単語を特定し、名詞の単語が述語に変換された単語又は特定された述語の単語と、前記複数の品詞付単語の中の第１の単語と、前記補助情報とに基づいて、前記第１の単語と関係がある単語である関係単語を特定し、前記第１の重要文と前記関係単語とに基づいて前記クエリを作成する、
請求項１から１１のいずれか１項に記載の情報処理装置。
【請求項１３】
抽出更新部をさらに有し、
前記補助情報は、前記関係性の度合を示す関係度を含み、
前記取得部は、前記要約文の修正情報を取得し、
前記抽出更新部は、前記要約文と、前記修正情報とを比較して差分を抽出し、前記第１の重要文と前記要約文の差分とに基づいて、前記第１の重要文の単語と前記要約文の差分に対応する単語との前記関係度を前記補助情報から特定し、特定された前記関係度を、現状の値よりも低くし、前記第１の重要文と前記修正情報の差分とに基づいて、前記第１の重要文の単語と前記修正情報の差分に対応する単語との前記関係度を前記補助情報から特定し、特定された前記関係度を、現状の値よりも高くする、
請求項１２に記載の情報処理装置。
【請求項１４】
情報処理装置が、
対話内容を示すテキストデータと、過去の報告書に基づいて作成された複数の文を含む知識情報とを取得し、
形態素解析を用いて、前記テキストデータを解析し、
前記形態素解析の結果を用いて、前記テキストデータに含まれている複数の文の中から、１つの文を、第１の重要文として、抽出し、
前記第１の重要文に基づいて、クエリを作成し、
前記知識情報の中から、前記クエリにより得られる文を検索し、
予め設定された方法で、検索された複数の文のそれぞれのスコアを算出し、
前記第１の重要文と、検索された複数の文のそれぞれとに基づいて、前記第１の重要文と、検索された複数の文のそれぞれとが類似している度合である複数の類似度を算出し、
前記複数の類似度に基づいて、検索された複数の文のそれぞれのスコアを更新し、
複数の文のそれぞれのスコアに基づいて、検索された複数の文のうちの１つの文を、要約文として、選択し、
前記要約文を出力する、
要約文出力方法。
【請求項１５】
情報処理装置に、
対話内容を示すテキストデータと、過去の報告書に基づいて作成された複数の文を含む知識情報とを取得し、
形態素解析を用いて、前記テキストデータを解析し、
前記形態素解析の結果を用いて、前記テキストデータに含まれている複数の文の中から、１つの文を、第１の重要文として、抽出し、
前記第１の重要文に基づいて、クエリを作成し、
前記知識情報の中から、前記クエリにより得られる文を検索し、
予め設定された方法で、検索された複数の文のそれぞれのスコアを算出し、
前記第１の重要文と、検索された複数の文のそれぞれとに基づいて、前記第１の重要文と、検索された複数の文のそれぞれとが類似している度合である複数の類似度を算出し、
前記複数の類似度に基づいて、検索された複数の文のそれぞれのスコアを更新し、
複数の文のそれぞれのスコアに基づいて、検索された複数の文のうちの１つの文を、要約文として、選択し、
前記要約文を出力する、

処理を実行させる要約文出力プログラム。
【発明の詳細な説明】
【技術分野】
【０００１】
本開示は、情報処理装置、要約文出力方法、及び要約文出力プログラムに関する。
【背景技術】
【０００２】
コールセンターでは、オペレータは、顧客とオペレータとの対話内容をまとめて、報告書を作成する。報告書を作成する作業は、アフターコールワークとも言う。オペレータがアフターコールワークを行うことは、オペレータの作業負担が大きい。そこで、対話内容を自動的に要約することが考えられる。ここで、要約を作成する技術が提案されている（特許文献１を参照）。特許文献１の要約装置は、入力されたデータから抽出された重要な文を結合した要約を作成する。なお、当該データは、音声認識結果を示すデータなどである。
【先行技術文献】
【特許文献】
【０００３】
【文献】特許５５６２２１９号公報
国際公開第２０２２／０４９６６８号
【発明の概要】
【発明が解決しようとする課題】
【０００４】
上記の技術では、音声認識結果を示すデータを用いて、要約文が作成される。そのため、例えば、音声認識に誤りがあった場合、要約文には、誤りが含まれる。また、例えば、要約文の内容は、話し言葉で表される。このように、上記の技術では、要約文は、データの文体に依存される。そのため、上記の技術に基づく要約文は、適切と言えない場合がある。
【０００５】
本開示の目的は、適切な要約文を出力することである。
【課題を解決するための手段】
【０００６】
本開示の一態様に係る情報処理装置が提供される。情報処理装置は、対話内容を示すテキストデータと、過去の報告書に基づいて作成された複数の文を含む知識情報とを取得する取得部と、形態素解析を用いて、前記テキストデータを解析する解析部と、前記形態素解析の結果を用いて、前記テキストデータに含まれている複数の文の中から、１つの文を、第１の重要文として、抽出する重要文抽出部と、前記第１の重要文に基づいて、クエリを作成し、前記知識情報の中から、前記クエリにより得られる文を検索し、予め設定された方法で、検索された複数の文のそれぞれのスコアを算出する作成検索算出部と、前記第１の重要文と、検索された複数の文のそれぞれとに基づいて、前記第１の重要文と、検索された複数の文のそれぞれとが類似している度合である複数の類似度を算出し、前記複数の類似度に基づいて、検索された複数の文のそれぞれのスコアを更新する算出更新部と、複数の文のそれぞれのスコアに基づいて、検索された複数の文のうちの１つの文を、要約文として、選択する選択部と、前記要約文を出力する出力部と、を有する。
【発明の効果】
【０００７】
本開示によれば、適切な要約文を出力することができる。
【図面の簡単な説明】
【０００８】
【図１】実施の形態１の情報処理装置の機能を示すブロック図である。
【図２】実施の形態１の情報処理装置が有するハードウェアを示す図である。

【図３】実施の形態１の不要語辞書の例を示す図である。
【図４】実施の形態１の文分割辞書の例を示す図である。
【図５】実施の形態１の重要文の抽出の例を示す図である。
【図６】実施の形態１のクエリの例を示す図である。
【図７】実施の形態１の知識データベースの例を示す図である。
【図８】実施の形態１の検索結果の例を示す図である。
【図９】実施の形態１の更新されたスコアの例を示す図である。
【図１０】実施の形態１の情報処理装置が実行する処理の例を示すフローチャートである。
【図１１】実施の形態１の作成装置の機能を示すブロック図である。
【図１２】実施の形態１の対話データベースの例を示す図である。
【図１３】実施の形態１の報告書データベースの例を示す図である。
【図１４】実施の形態１の非文の削除の例を示す図である。
【図１５】実施の形態１の作成装置が実行する処理の例を示すフローチャートである。
【図１６】実施の形態２の情報処理装置の機能を示すブロック図である。
【図１７】実施の形態２のカテゴリが推定されない場合の例を示す図である。
【図１８】実施の形態２の情報処理装置が実行する処理の例を示すフローチャートである。
【図１９】実施の形態２の作成装置の機能を示すブロック図である。
【図２０】実施の形態２の作成装置が実行する処理の例を示すフローチャートである。
【図２１】実施の形態３の情報処理装置の機能を示すブロック図である。
【図２２】実施の形態３の補助情報の例を示す図である。
【図２３】実施の形態３の情報処理装置が実行する処理の例を示すフローチャートである。
【図２４】実施の形態３の作成装置の機能を示すブロック図である。
【図２５】実施の形態３の作成装置が実行する処理の例を示すフローチャートである。
【図２６】実施の形態４の情報処理装置の機能を示すブロック図である。
【図２７】実施の形態４の情報処理装置が実行する処理の例を示すフローチャートである。
【発明を実施するための形態】
【０００９】
以下、図面を参照しながら実施の形態を説明する。以下の実施の形態は、例にすぎず、本開示の範囲内で種々の変更が可能である。
【００１０】
実施の形態１．
図１は、実施の形態１の情報処理装置の機能を示すブロック図である。情報処理装置１００は、要約文出力方法を実行する装置である。情報処理装置１００は、対話要約生成装置と呼んでもよい。
【００１１】
まず、情報処理装置１００が有するハードウェアを説明する。
図２は、実施の形態１の情報処理装置が有するハードウェアを示す図である。情報処理装置１００は、プロセッサ１０１、揮発性記憶装置１０２、不揮発性記憶装置１０３、及びインタフェース１０４を有する。
【００１２】
プロセッサ１０１は、情報処理装置１００全体を制御する。例えば、プロセッサ１０１は、ＣＰＵ（Ｃｅｎｔｒａｌ Ｐｒｏｃｅｓｓｉｎｇ Ｕｎｉｔ）、ＦＰＧＡ（Ｆｉｅｌｄ Ｐｒｏｇｒａｍｍａｂｌｅ Ｇａｔｅ Ａｒｒａｙ）などである。プロセッサ１０１は、マルチプロセッサでもよい。また、情報処理装置１００は、処理回路を有してもよい。処理回路は、単一回路又は複合回路でもよい。
【００１３】
揮発性記憶装置１０２は、情報処理装置１００の主記憶装置である。例えば、揮発性記憶装置１０２は、ＲＡＭ（Ｒａｎｄｏｍ Ａｃｃｅｓｓ Ｍｅｍｏｒｙ）である。不揮発性記憶装置１０３は、情報処理装置１００の補助記憶装置である。例えば、不揮発性記憶装置１０３は、ＨＤＤ（Ｈａｒｄ Ｄｉｓｋ Ｄｒｉｖｅ）、又はＳＳＤ（Ｓｏｌｉｄ Ｓｔ

ａｔｅ　Ｄｒｉｖｅ）である。
【００１４】
インタフェース１０４は、他の装置と通信する。また、インタフェース１０４は、顧客とオペレータとの音声を示す音声信号を受信してもよい。さらに、インタフェース１０４は、オペレータが入力したテキストデータを受信してもよい。
【００１５】
図１に戻って、情報処理装置１００が有する機能を説明する。
情報処理装置１００は、記憶部１１０、取得部１２０、解析部１３０、不要語削除部１４０、重要文抽出部１５０、作成検索算出部１６０、算出更新部１７０、選択部１８０、及び出力部１９０を有する。
【００１６】
記憶部１１０は、揮発性記憶装置１０２又は不揮発性記憶装置１０３に確保した記憶領域として実現してもよい。
取得部１２０、解析部１３０、不要語削除部１４０、重要文抽出部１５０、作成検索算出部１６０、算出更新部１７０、選択部１８０、及び出力部１９０の一部又は全部は、処理回路によって実現してもよい。また、取得部１２０、解析部１３０、不要語削除部１４０、重要文抽出部１５０、作成検索算出部１６０、算出更新部１７０、選択部１８０、及び出力部１９０の一部又は全部は、プロセッサ１０１が実行するプログラムのモジュールとして実現してもよい。例えば、プロセッサ１０１が実行するプログラムは、要約文出力プログラムとも言う。例えば、要約文出力プログラムは、記録媒体に記録されている。
【００１７】
記憶部１１０は、不要語辞書１１１、文分割辞書１１２、単語重要度モデル１１３、及び知識データベース１１４を記憶してもよい。不要語辞書１１１、文分割辞書１１２、単語重要度モデル１１３、及び知識データベース１１４については、後で説明する。
【００１８】
取得部１２０は、テキストデータを取得する。例えば、取得部１２０は、テキストデータを外部装置（例えば、クラウドサーバ）から取得する。また、例えば、取得部１２０は、テキストデータを記憶部１１０から取得する。テキストデータは、対話内容を示すデータである。例えば、対話内容は、顧客とオペレータとの対話内容、チャットボットとユーザとの対話内容などである。なお、顧客とオペレータとの対話内容を示すテキストデータは、音声認識技術を用いることで、生成される。また、テキストデータには、対話内容を示す複数の文が含まれている。
【００１９】
取得部１２０は、知識データベース１１４を取得する。例えば、取得部１２０は、知識データベース１１４を記憶部１１０から取得する。また、例えば、取得部１２０は、知識データベース１１４を外部装置から取得する。
【００２０】
解析部１３０は、形態素解析を用いて、テキストデータを解析する。これにより、テキストデータに含まれる単語及び品詞が、抽出される。また、解析部１３０は、構文解析を用いて、テキストデータ内の文節及び文節間の関係を解析してもよい。
解析部１３０は、後述する不要語削除処理が実行された後のテキストデータを、解析してもよい。
【００２１】
不要語削除部１４０は、予め設定された方法で、テキストデータの中から不要語を削除する。例えば、不要語削除部１４０は、不要語辞書１１１を用いて、不要語を削除する。ここで、不要語辞書１１１を例示する。
【００２２】
図３は、実施の形態１の不要語辞書の例を示す図である。例えば、不要語辞書１１１は、記憶部１１０に格納されている。不要語辞書１１１には、“あー”、“えー”などのフィラー、及び“お待たせ致しました。”のような定型句が登録される。また、不要語辞書１１１

には、単語と品詞との対応関係を示す情報が登録されてもよい。
【００２３】
　また、不要語削除部１４０は、機械学習により得られた学習済モデルを用いて、テキストデータの中から不要語を削除してもよい。なお、テキストデータの中に不要語が含まれていない場合、不要語削除部１４０は、処理を実行しない。
【００２４】
　重要文抽出部１５０は、形態素解析の結果を用いて、テキストデータ（例えば、不要語が削除されたテキストデータ）に含まれている複数の文の中から、予め設定された数の文を、重要文として抽出する。なお、予め設定された数は、１つでもよいし、２つ以上でもよい。予め設定された数が１つである場合、抽出された重要文は、第１の重要文とも言う。以下の説明では、予め設定された数は、２つ以上とする。
【００２５】
　詳細に、重要文の抽出処理を説明する。まず、重要文抽出部１５０は、テキストデータに含まれている複数の文を分割する。複数の文を分割する方法としては、音声認識の区切れ目で分割する方法、テキストデータに含まれる句点又は読点で分割する方法、文分割辞書１１２を用いて分割する方法、学習済モデルを用いて分割する方法などが挙げられる。ここで、文分割辞書１１２を例示する。
【００２６】
　図４は、実施の形態１の文分割辞書の例を示す図である。例えば、文分割辞書１１２は、記憶部１１０に格納されている。重要文抽出部１５０は、文分割辞書１１２を用いて、複数の文を分割してもよい。
【００２７】
　重要文抽出部１５０は、複数の文のそれぞれに対して、重要度を算出する。まず、１つの文に対応する重要度の算出方法を説明する。例えば、重要文抽出部１５０は、形態素解析の結果により得られた、当該１つの文に含まれている複数の単語のそれぞれの重要度を、ＴＦ－ＩＤＦ、Ｏｋａｐｉ　ＢＭ２５などを用いて、算出する。重要文抽出部１５０は、複数の単語のそれぞれの重要度を加算することで、当該１つの文に対応する重要度を算出する。また、重要文抽出部１５０は、形態素解析の結果により得られた、当該１つの文に含まれている複数の単語のそれぞれの重要度を、学習済モデルである単語重要度モデル１１３を用いて、算出してもよい。そして、重要文抽出部１５０は、複数の単語のそれぞれの重要度を加算することで、当該１つの文に対応する重要度を算出する。また、重要文抽出部１５０は、複数の単語のそれぞれの重要度の平均値を、当該１つの文に対応する重要度として、算出してもよい。
【００２８】
　このように、当該１つの文に対応する重要度が、算出される。同様に、重要文抽出部１５０は、複数の文のそれぞれに対して、重要度を算出する。これにより、複数の文に対応する複数の重要度が、算出される。
【００２９】
　重要文抽出部１５０は、重要度が高い順に、複数の文を並べる。重要文抽出部１５０は、上位の予め設定された数の文を、重要文として抽出する。具体的に、重要文の抽出を例示する。
【００３０】
　図５は、実施の形態１の重要文の抽出の例を示す図である。図５は、テキストデータに含まれている複数の文（すなわち、複数の分割文）を示している。重要文抽出部１５０は、複数の文の中から、重要文を抽出する。
【００３１】
　上述したように、重要文抽出部１５０は、上位の予め設定された数の文を、重要文として抽出する。これにより、重要度の低い文が除かれる。後述するように、重要文に基づいて要約文が選択されるので、重要文のみが抽出されることは、適切な要約文が選択される。
【００３２】

作成検索算出部１６０は、重要文に基づいて、クエリを作成する。クエリは、文単位、又は単語（すなわち、形態素）単位で作成されてもよい。クエリは、文又は単語のｎ－ｇｒａｍ連鎖により作成されてもよい。ここで、クエリを例示する。

【００３３】

図６は、実施の形態１のクエリの例を示す図である。図６は、重要文に基づいて作成されたクエリを示している。図６は、２－ｇｒａｍ連鎖を用いて作成されたクエリを示している。

【００３４】

ここで、作成検索算出部１６０は、重要文の中の連続する単語を語彙化することにより、クエリを作成してもよい。例えば、作成検索算出部１６０は、重要文の中の否定語と、否定語の直前の動詞とを語彙化することにより、クエリを作成する。例えば、図６は、否定語と直前の動詞とが連結された“居る＿ない”を示している。

【００３５】

ここで、数詞と、数詞の後の単位とのそれぞれが、クエリに含まれる場合が考えられる。例えば、重要文が“エアコンを２５℃に設定した”である場合、クエリは、“エアコン　２５　℃　設定”である。しかし、当該クエリを用いて、検索が行われた場合、意味が異なる文が検索されることがある。例えば、“冷蔵庫を２５日に購入”が、検索される。そこで、作成検索算出部１６０は、重要文の中の数詞と、当該数詞の後の単位とを語彙化することにより、クエリを作成してもよい。例えば、語彙化されたクエリは、“２５＿℃”である。これにより、上記のクエリは、“エアコン　２５＿℃　設定”に変わる。このように、数詞と単位とを１つの語彙にすることで、意味が異なる文が検索されることが、防止される。

【００３６】

作成検索算出部１６０は、知識データベース１１４の中から、クエリにより得られる文を検索する。言い換えれば、作成検索算出部１６０は、クエリを用いて、知識データベース１１４の中から、重要文の意味に近い文を検索する。ここで、知識データベース１１４を例示する。

【００３７】

図７は、実施の形態１の知識データベースの例を示す図である。知識データベース１１４は、知識情報とも言う。知識データベース１１４は、過去の報告書に基づいて作成された複数の文を含む。知識データベース１１４は、文、インデックス登録クエリ、ステップ、及びカテゴリの項目を有する。

【００３８】

文の項目には、過去の報告書に基づいて作成された文が登録される。また、文の項目には、文節、連続した複数の文節、連続した複数の文が登録されてもよい。インデックス登録クエリの項目には、クエリが登録される。ステップの項目には、業務の名称が登録される。カテゴリの項目には、文の項目に登録されている文の内容が示すカテゴリが登録される。このように、文の項目に登録されている複数の文のそれぞれには、カテゴリが対応付けられている。

また、知識データベース１１４は、グラフ形式の情報でもよい。

【００３９】

このように、作成検索算出部１６０は、クエリを用いて、知識データベース１１４の中から、重要文の意味に近い文を検索する。検索結果として、複数の文が検索される。

【００４０】

また、作成検索算出部１６０は、重要文と、テキストデータの中の当該重要文の前文と後文とのうちの少なくとも１つとに基づいて、クエリを作成してもよい。前文と後文とうちの少なくとも１つに含まれている単語が、クエリの中に含まれることで、作成検索算出部１６０は、重要文が短い場合でも、重要文と関係のある文を検索できる。

【００４１】

作成検索算出部１６０は、重要文に含まれている単語の類義語をクエリに含めてもよい。重要文に含まれている単語が“点かない”である場合、作成検索算出部１６０は、“点かな

い”の類義語である“消える”をクエリに含める。なお、作成検索算出部１６０は、ｗｏｒｄ２ｖｅｃを用いて、重要文に含まれている単語の類義語を得ることができる。このように、作成検索算出部１６０は、類義語をクエリに含めることで、類義語が含まれている文を検索できる。

【００４２】

また、作成検索算出部１６０は、検索対象を、名詞、動詞、形容詞、形状詞などの品詞でフィルタリングしてもよい。作成検索算出部１６０は、重要文と知識データベース１１４とを用いて、文ベクトルの類似度を算出してもよい。

【００４３】

作成検索算出部１６０は、予め設定された方法で、検索された複数の文のそれぞれのスコアを算出する。例えば、予め設定された方法は、次のような方法である。例えば、クエリは、“エアコン　２５℃　設定”とする。検索された文は、“エアコンは２５℃”と“エアコンが動かない”とする。作成検索算出部１６０は、“エアコンは２５℃”のスコアを２と算出する。作成検索算出部１６０は、“エアコンが動かない”のスコアを、１と算出する。すなわち、作成検索算出部１６０は、検索された文とクエリに含まれる単語とのマッチ数をスコアとして、算出する。このように、スコアが、算出される。また、作成検索算出部１６０は、Ｅｌａｓｔｉｃｓｅａｒｃｈのような検索エンジンで用いられている算出方法を用いて、スコアを算出してもよい。

【００４４】

作成検索算出部１６０は、複数の文のそれぞれのスコアを検索結果に含めてもよい。ここで、検索結果を例示する。

図８は、実施の形態１の検索結果の例を示す図である。図８が示すように、検索された文と、当該検索された文のスコアとが、検索結果として出力される。

【００４５】

算出更新部１７０は、重要文と、検索された複数の文のそれぞれとに基づいて、複数の類似度を算出する。まず、重要文と、検索された複数の文のうちの１つの文とに基づいて、類似度が算出される場合を説明する。なお、当該類似度は、当該重要文と、当該１つの文とが類似している度合である。よって、上記の複数の類似度とは、重要文と、検索された複数の文のそれぞれとが類似している度合である。

【００４６】

ここで、類似度を算出する方法として、Ｊａｃｃａｒｄ係数などを用いる方法が考えられる。しかし、音声対話が対象である場合、Ｊａｃｃａｒｄ係数などの方法を用いることは、望ましくない。理由は、差集合の要素数が大きいほど値が小さくなるというＪａｃｃａｒｄ係数の特徴にある。音声対話において、発話が冗長になることが多いため、重要文に含まれる単語の集合から検索された文に含まれる単語の集合を引いた差集合の要素数が大きいことは、許容されるべきである。一方で、検索された文に含まれる単語の集合から重要文に含まれる単語の集合を引いた差集合の要素数が大きい場合は、発話していない余計な内容が検索された文に含まれる可能性があるためペナルティを与えたい。そこで、算出更新部１７０は、式（１）を用いて、類似度を算出する。Ｉは、重要文に含まれる単語の集合である。Ｋは、検索された文に含まれる単語の集合である。

【００４７】

【数１】

$$\text{類似度} = \mathrm{Distance}(\mathrm{I},\mathrm{K}) = \frac{\frac{|\mathrm{I}\cap\mathrm{K}|}{|\mathrm{I}\cup\mathrm{K}|}}{\frac{|\mathrm{K}-\mathrm{I}|}{|\mathrm{K}|}} \cdots (1)$$

【００４８】

このように、算出更新部１７０は、検索された文に含まれる単語の集合から重要文に含まれる単語の集合を引いた差集合の要素数が大きい場合にペナルティを与えることにより

、集合同士の類似度を算出する。これにより、冗長な発話を吸収しつつ、発話していない余計な内容を含まない文が検索できるようになる。

【００４９】

上記したように、算出更新部１７０は、重要文と、当該１つの文とに基づいて、類似度を算出する。同様に、算出更新部１７０は、複数の文のそれぞれに対応する類似度を算出する。これにより、複数の類似度が算出される。算出更新部１７０は、検索された複数の文のうち、上位Ｎ個の文のそれぞれを用いて、類似度を算出してもよい。

算出更新部１７０は、複数の類似度に基づいて、検索された複数の文のそれぞれのスコアを更新する。ここで、更新されたスコアを例示する。

【００５０】

図９は、実施の形態１の更新されたスコアの例を示す図である。図９のスコアは、類似度を示している。

【００５１】

選択部１８０は、更新された、複数の文のそれぞれのスコアに基づいて、検索された複数の文のうちの１つの文を、要約文として、選択する。例えば、選択部１８０は、最も高いスコアに対応する文を、要約文として、選択する。以下の説明では、最も高いスコアに対応する文が、要約文として、特定されるものとする。

出力部１９０は、要約文を出力する。

【００５２】

次に、情報処理装置１００が実行する処理を、フローチャートを用いて、説明する。

図１０は、実施の形態１の情報処理装置が実行する処理の例を示すフローチャートである。

（ステップＳ１１）取得部１２０は、テキストデータを取得する。

（ステップＳ１２）解析部１３０は、形態素解析を用いて、テキストデータを解析する。

（ステップＳ１３）不要語削除部１４０は、テキストデータの中から不要語を削除する。

【００５３】

（ステップＳ１４）重要文抽出部１５０は、テキストデータに含まれている複数の文を分割する。重要文抽出部１５０は、複数の文のそれぞれに対して、重要度を算出する。これにより、複数の文に対応する複数の重要度が算出される。重要文抽出部１５０は、複数の重要度に基づいて、上位の予め設定された数の文を、重要文として抽出する。これにより、複数の重要文が抽出される。

【００５４】

（ステップＳ１５）作成検索算出部１６０は、未処理の重要文があるか否かを判定する。未処理の重要文がある場合、処理は、ステップＳ１６に進む。全ての重要文が処理された場合、処理は、ステップＳ２０に進む。

（ステップＳ１６）作成検索算出部１６０は、未処理の重要文の中から、１つの重要文を選択する。なお、選択された重要文は、第１の重要文と呼んでもよい。

【００５５】

（ステップＳ１７）作成検索算出部１６０は、重要文に基づいて、クエリを作成し、知識データベース１１４の中から、クエリにより得られる文を検索する。これにより、複数の文が検索される。作成検索算出部１６０は、検索された複数の文のそれぞれのスコアを算出する。これにより、複数の文に対応する複数のスコアが、算出される。

【００５６】

（ステップＳ１８）算出更新部１７０は、重要文と、上位Ｎ個のスコアに対応する複数の文のそれぞれとに基づいて、複数の類似度を算出する。算出更新部１７０は、複数の類似度に基づいて、上位Ｎ個のスコアに対応する複数の文のそれぞれのスコアを更新する。

（ステップＳ１９）選択部１８０は、最も高いスコアに対応する文を、要約文として、選択する。そして、処理は、ステップＳ１５に進む。

（ステップＳ２０）出力部１９０は、特定された複数の要約文をまとめて、要約テキストとして、出力する。出力部１９０は、報告書形式に作成された要約テキストを出力して

もよい。

【００５７】

実施の形態１によれば、要約テキストに含まれる要約文は、音声認識結果に基づく文でない。当該要約文は、過去の報告書に基づく文である。そのため、当該要約文には、誤りが含まれていない可能性が高い。また、当該要約文の内容は、話し言葉で表されていない。よって、情報処理装置１００は、適切な要約文を出力することができる。

【００５８】

ここで、作成装置を説明する。作成装置は、単語重要度モデル１１３と知識データベース１１４とを作成する。具体的に、作成装置を説明する。

図１１は、実施の形態１の作成装置の機能を示すブロック図である。作成装置２００は、記憶部２１０、単語重要度学習部２２０、及びデータベース作成部２３０を有する。

【００５９】

記憶部２１０は、作成装置２００が有する揮発性記憶装置又は不揮発性記憶装置に確保した記憶領域として実現してもよい。

単語重要度学習部２２０及びデータベース作成部２３０の一部又は全部は、作成装置２００が有する処理回路によって実現してもよい。また、単語重要度学習部２２０及びデータベース作成部２３０の一部又は全部は、作成装置２００が有するプロセッサが実行するプログラムのモジュールとして実現してもよい。

【００６０】

記憶部２１０は、対話データベース２１１を記憶する。ここで、対話データベース２１１を例示する。

図１２は、実施の形態１の対話データベースの例を示す図である。対話データベース２１１は、記憶部２１０に格納されている。対話データベース２１１には、過去の対話履歴が登録されている。具体的には、対話データベース２１１は、対話ＩＤ（ｉｄｅｎｔｉｆｉｅｒ）、音声認識結果、カテゴリ、及び受付日時の項目を有する。

【００６１】

対話ＩＤの項目には、識別子が登録される。音声認識結果の項目には、対話内容が登録される。カテゴリの項目には、音声認識結果の項目に登録されている文の内容が示すカテゴリが登録される。受付日時の項目には、対話が行われた日時が登録される。

【００６２】

記憶部２１０は、報告書データベース２１２を記憶する。ここで、報告書データベース２１２を例示する。

図１３は、実施の形態１の報告書データベースの例を示す図である。報告書データベース２１２は、記憶部２１０に格納されている。報告書データベース２１２は、過去の報告書に基づいて作成された情報である。報告書データベース２１２は、対話ＩＤ、受付履歴、対応履歴、カテゴリ、及び受付日時の項目を有する。

【００６３】

対話ＩＤの項目には、識別子が登録される。受付履歴の項目には、対話データベース２１１の音声認識結果の項目に登録されている情報の要約文が登録される。対応履歴の項目には、対応内容が登録される。カテゴリの項目には、受付履歴に登録されている内容が示すカテゴリが登録される。受付日時の項目には、対話が行われた日時が登録される。

【００６４】

単語重要度学習部２２０は、対話データベース２１１を用いて機械学習を行うことにより、単語重要度モデル１１３を作成する。なお、単語重要度モデル１１３は、単語が入力された場合、当該単語の重要度を出力する。

【００６５】

データベース作成部２３０は、報告書データベース２１２に基づいて、知識データベース１１４を作成する。例えば、データベース作成部２３０は、報告書データベース２１２の受付履歴及び対応履歴の項目に登録されている情報のうち、意味のある文、又は意味のある文節を抽出することで、知識データベース１１４を作成する。

【００６６】
　データベース作成部２３０は、報告書データベース２１２に登録されている情報のうち、言語として不自然な文（以下、非文という）を削除してもよい。例えば、データベース作成部２３０は、ｎ－ｇｒａｍ尤度等を用いて、文の尤度を算出し、文の尤度に基づいて、非文を削除してもよい。ここで、非文が削除される例を示す。
【００６７】
　図１４は、実施の形態１の非文の削除の例を示す図である。図１４が示すように、データベース作成部２３０は、報告書データベース２１２に登録されている情報の中から、非文を削除する。このように、知識データベース１１４は、過去の報告書の中から非文が削除されることにより作成される。そして、過去の報告書の中から非文が削除されることにより、非文が要約文として選択されることが防止できる。
【００６８】
　次に、作成装置２００が実行する処理を、フローチャートを用いて、説明する。
　図１５は、実施の形態１の作成装置が実行する処理の例を示すフローチャートである。
　（ステップＳ２１）単語重要度学習部２２０は、対話データベース２１１の音声認識結果を参照し、形態素解析を用いて、音声認識結果を解析する。
　（ステップＳ２２）単語重要度学習部２２０は、ＴＦ－ＩＤＦなどを用いて、解析により得られた単語の重要度を算出する。単語重要度学習部２２０は、単語と重要度との対応関係を示す情報を作成する。
【００６９】
　（ステップＳ２３）単語重要度学習部２２０は、全ての音声認識結果に対して処理を行ったか否かを判定する。全ての音声認識結果に対して処理を行った場合、処理は、ステップS２４に進む。未処理の音声認識結果がある場合、処理は、ステップＳ２１に進む。
【００７０】
　このように、ステップＳ２１とステップＳ２２とが繰り返されることで、単語に対応する重要度が変化する。言い換えれば、学習により、単語に対応する重要度が更新される。そして、ステップＳ２１とステップＳ２２とが繰り返されることで得られた、単語と重要度との対応関係を示す情報が、単語重要度モデル１１３になる。
【００７１】
　（ステップＳ２４）データベース作成部２３０は、報告書データベース２１２に基づいて、知識データベース１１４を作成する。
　これにより、単語重要度モデル１１３と知識データベース１１４とが、作成される。
【００７２】
実施の形態２．
　次に、実施の形態２を説明する。実施の形態２では、実施の形態１と相違する事項を主に説明する。そして、実施の形態２では、実施の形態１と共通する事項の説明を省略する。
　図１６は、実施の形態２の情報処理装置の機能を示すブロック図である。情報処理装置１００は、さらに、カテゴリ推定部１９１を有する。
【００７３】
　記憶部１１０は、さらに、カテゴリ推定モデル１１５を記憶してもよい。カテゴリ推定モデル１１５は、単語が入力された場合、当該単語に基づいたカテゴリを出力する。言い換えれば、カテゴリ推定モデル１１５は、当該単語に基づいて、カテゴリを推定する。なお、例えば、カテゴリは、機種である。
【００７４】
　取得部１２０は、カテゴリ推定モデル１１５を取得する。例えば、取得部１２０は、カテゴリ推定モデル１１５を記憶部１１０から取得する。また、例えば、取得部１２０は、カテゴリ推定モデル１１５を外部装置から取得する。
　カテゴリ推定部１９１は、形態素解析により得られた単語と、カテゴリ推定モデル１１５とを用いて、対話内容のカテゴリを推定する。
【００７５】

ここで、カテゴリが推定されない場合の例を示す。
図１７は、実施の形態２のカテゴリが推定されない場合の例を示す図である。テキストデータが示す対話内容のカテゴリは、冷蔵庫である。カテゴリが推定されない場合、単語“庫内灯”に基づいて、オーブンレンジに関する文が、多く検索される。そこで、作成検索算出部１６０は、推定されたカテゴリ“冷蔵庫”を用いて、“冷蔵庫”に関する文を検索する。これにより、対話内容のカテゴリに関する文のみが、検索される。情報処理装置１００は、対話内容のカテゴリに関する文のみを検索することで、適切な要約文を選択することができる。
【００７６】
図１８は、実施の形態２の情報処理装置が実行する処理の例を示すフローチャートである。図１８の処理は、ステップＳ１３ａが実行される点が図１０の処理と異なる。また、図１８の処理では、ステップＳ１７がステップＳ１７ａに変更される。そこで、図１８では、ステップＳ１３ａ，１７ａを説明する。そして、ステップＳ１３ａ，１７ａ以外の処理の説明は、省略する。
【００７７】
（ステップＳ１３ａ）カテゴリ推定部１９１は、形態素解析により得られた単語と、カテゴリ推定モデル１１５とを用いて、対話内容のカテゴリを推定する。
（ステップＳ１７ａ）作成検索算出部１６０は、重要文に基づいて、クエリを作成する。作成検索算出部１６０は、推定されたカテゴリとクエリを用いて、知識データベース１１４に対して検索を行う。すなわち、作成検索算出部１６０は、推定されたカテゴリに情報を絞った状態で、クエリを用いて、知識データベース１１４に対して検索を行う。これにより、推定されたカテゴリと関係のある複数の文が検索される。作成検索算出部１６０は、検索された複数の文のそれぞれのスコアを算出する。これにより、複数の文に対応する複数のスコアが、算出される。
【００７８】
実施の形態２によれば、情報処理装置１００は、対話内容のカテゴリに関する文のみを検索することで、適切な要約文を選択することができる。
【００７９】
ここで、カテゴリ推定モデル１１５は、作成装置２００により、作成される。カテゴリ推定モデル１１５の作成について説明する。
図１９は、実施の形態２の作成装置の機能を示すブロック図である。作成装置２００は、さらに、カテゴリ推定学習部２４０を有する。カテゴリ推定学習部２４０は、カテゴリ推定モデル１１５を作成する。
【００８０】
図２０は、実施の形態２の作成装置が実行する処理の例を示すフローチャートである。図２０の処理は、ステップＳ２４ａが実行される点が図１５の処理と異なる。そこで、図２０では、ステップＳ２４ａを説明する。そして、ステップＳ２４ａ以外の処理の説明は、省略する。
【００８１】
（ステップＳ２４ａ）カテゴリ推定学習部２４０は、対話データベース２１１の音声認識結果に対して、形態素解析を行う。カテゴリ推定学習部２４０は、形態素解析により得られた単語と、対話データベース２１１のカテゴリとにおける自己相互情報量を算出し、自己相互情報量に基づいて、単語とカテゴリとの対応関係を示す情報を、カテゴリ推定モデル１１５として、作成する。
【００８２】
また、カテゴリ推定学習部２４０は、報告書データベース２１２の受付履歴に対して、形態素解析を行ってもよい。カテゴリ推定学習部２４０は、形態素解析により得られた単語と、報告書データベース２１２のカテゴリとにおける自己相互情報量を算出し、自己相互情報量に基づいて、単語とカテゴリとの対応関係を示す情報を、カテゴリ推定モデル１１５として、作成する。

【００８３】
実施の形態３．
次に、実施の形態３を説明する。実施の形態３では、実施の形態１と相違する事項を主に説明する。そして、実施の形態３では、実施の形態１と共通する事項の説明を省略する。
図２１は、実施の形態３の情報処理装置の機能を示すブロック図である。記憶部１１０は、補助情報１１６を記憶してもよい。補助情報１１６は、クエリの生成を補助する情報である。言い換えれば、補助情報１１６は、クエリを生成する際に用いられる情報である。例えば、補助情報１１６は、特許文献２に記載の３次元情報と同じような情報であると考えてもよい。ここで、補助情報１１６を具体的に示す。
【００８４】
図２２は、実施の形態３の補助情報の例を示す図である。補助情報１１６は、複数の述語である複数の単語のそれぞれと、複数の関係情報のそれぞれとの対応関係を示す情報である。関係情報とは、複数の単語のそれぞれの関係性を示す。
【００８５】
例えば、補助情報１１６は、述語ラベルとサブ単語文脈行列の項目を有する。述語ラベルの項目には、述語である単語が登録される。例えば、述語ラベルの項目には、述語である単語“落ちる”が登録される。サブ単語文脈行列の項目には、関係情報が登録される。ここで、関係情報は、２次元のテーブルで表されると考えてもよい。すなわち、関係情報は、２次元情報と考えてもよい。関係情報には、述語である単語が対応付けられる。述語である単語と関係情報との対応関係を示す情報は、３次元情報と考えてもよい。よって、補助情報１１６は、３次元情報と考えてもよい。
【００８６】
上述したように、関係情報は、複数の単語のそれぞれの関係性を示す。当該複数の単語とは、動詞、名詞、形容詞などの単語である。図２２の関係情報では、名詞の単語が例示されている。上述の通り、図２２の関係情報には、名詞以外の品詞の単語が含まれてもよい。
【００８７】
次に、補助情報１１６を具体的に説明する。例えば、述語ラベル“落ちる”には、関係情報が対応付けられている。図２２の関係情報では、“証明”と“照明”との関係性の度合が“１５９”であることが示されている。ここで、“１５９”などの数字は、複数の単語のそれぞれの関係性の度合を示す関係度と呼ぶ。このように、補助情報１１６には、関係度が含まれている。また、関係度は、自己相互情報量と考えてもよい。なお、関係度の上限は、１００に限らない。関係度“１５９”は、予め設定された閾値よりも大きい。よって、図２２の関係情報は、“証明”と“照明”との関係性が強いことを示している。
【００８８】
取得部１２０は、補助情報１１６を取得する。例えば、取得部１２０は、補助情報１１６を記憶部１１０から取得する。また、例えば、取得部１２０は、補助情報１１６を外部装置から取得する。
【００８９】
作成検索算出部１６０は、重要文に対して形態素解析を実行することで得られた複数の品詞付単語の中から、述語になれる名詞の単語又は述語の単語を特定する。作成検索算出部１６０は、名詞の単語が述語に変換された単語又は特定された述語の単語と、複数の品詞付単語の中の単語（例えば、第１の単語とも言う。）と、補助情報１１６とに基づいて、当該単語（すなわち、第１の単語）と関係がある単語である関係単語を特定する。作成検索算出部１６０は、重要文と関係単語とに基づいてクエリを生成する。
【００９０】
例えば、重要文が“証明が落ちる”であるものとする。作成検索算出部１６０は、述語の単語“落ちる”を特定する。作成検索算出部１６０は、述語の単語“落ちる”と、単語“証明”と、補助情報１１６とに基づいて、単語“証明”と関係がある単語“照明”を特定する。作成検索算出部１６０は、重要文“証明が落ちる”と単語“照明”とに基づいてクエリを生成する

。例えば、作成検索算出部１６０は、クエリ“証明　落ちる　照明”を作成する。このように、作成検索算出部１６０は、クエリ拡張によって、クエリを作成する。

【００９１】

ここで、重要文“証明が落ちる”の“証明”は、“照明”の誤りである。例えば、重要文（すなわち、テキストデータ）が音声認識によって作成された場合、音声認識の誤りによって、重要文“証明が落ちる”が作成される。クエリが“証明が落ちる”に基づいて、作成された場合、“証明”に関係する文が、検索される。“証明”に関係する文に基づいて、選択された要約文は、正確性が低い。そこで、作成検索算出部１６０は、“照明”を含むクエリを作成する。これにより、“照明”に関係する文も、検索される。これにより、情報処理装置１００は、“照明”に関係する文を要約文として、選択できる。

【００９２】

図２３は、実施の形態３の情報処理装置が実行する処理の例を示すフローチャートである。図２３の処理では、ステップＳ１７がステップＳ１７ｂに変更される。そこで、図２３では、ステップＳ１７ｂを説明する。そして、ステップＳ１７ｂ以外の処理の説明は、省略する。

【００９３】

（ステップＳ１７ｂ）作成検索算出部１６０は、クエリ拡張によって、クエリを作成する。作成検索算出部１６０は、知識データベース１１４の中から、クエリにより得られる文を検索する。これにより、複数の文が検索される。作成検索算出部１６０は、検索された複数の文のそれぞれのスコアを算出する。これにより、複数の文に対応する複数のスコアが、算出される。

【００９４】

実施の形態３によれば、情報処理装置１００は、正確性の高い要約文を選択できる。

ここで、補助情報１１６は、作成装置２００により、作成される。補助情報１１６の作成について説明する。

【００９５】

図２４は、実施の形態３の作成装置の機能を示すブロック図である。作成装置２００は、さらに、補助情報作成部２５０を有する。

補助情報作成部２５０は、補助情報１１６を作成する。

【００９６】

図２５は、実施の形態３の作成装置が実行する処理の例を示すフローチャートである。図２５の処理は、ステップＳ２４ｂが実行される点が図１５の処理と異なる。そこで、図２５では、ステップＳ２４ｂを説明する。そして、ステップＳ２４ｂ以外の処理の説明は、省略する。

【００９７】

（ステップＳ２４ｂ）例えば、補助情報作成部２５０は、対話データベース２１１に含まれている１つの述語と１つの名詞を抽出する。補助情報作成部２５０は、報告書データベース２１２に含まれている１つの名詞を抽出する。補助情報作成部２５０は、抽出された、対話データベース２１１の述語と名詞と、報告書データベース２１２の名詞とを用いて、自己相互情報量を算出する。補助情報作成部２５０は、自己相互情報量に基づいて、補助情報１１６を作成する。また、対話データベース２１１（詳細には、対話データベース２１１の述語と名詞）に誤りが含まれていても、報告書データベース２１２に基づいて作成された知識データベース１１４内の名詞が検索されることで、当該誤りが回復される。

【００９８】

実施の形態４．

次に、実施の形態４を説明する。実施の形態４では、実施の形態３と相違する事項を主に説明する。そして、実施の形態４では、実施の形態３と共通する事項の説明を省略する。

図２６は、実施の形態４の情報処理装置の機能を示すブロック図である。情報処理装置１００は、さらに、抽出更新部１９２を有する。

取得部１２０は、修正情報を取得する。修正情報は、ユーザに修正された要約文の情報

である。

【００９９】

抽出更新部１９２は、出力部１９０が出力した要約テキスト内の要約文と、修正情報とを比較し、差分を抽出する。抽出更新部１９２は、重要文と要約文の差分とに基づいて、重要文の単語と要約文の差分に対応する単語との関係度を補助情報１１６から特定し、特定された関係度を、現状の当該関係度（すなわち、値）よりも低くする。抽出更新部１９２は、重要文と修正情報の差分とに基づいて、重要文の単語と修正情報の差分に対応する単語との関係度を補助情報１１６から特定し、特定された関係度を、現状の当該関係度（すなわち、値）よりも高くする。

【０１００】

例えば、重要文が“証明が落ちる”であるものとする。当該要約テキストが“賞名が落ちる”であるとする。修正情報が“照明が落ちる”であるとする。当該要約テキストと修正情報との差分は、“賞名”と“照明”である。抽出更新部１９２は、重要文と要約文の差分とに基づいて、重要文の単語“証明”と要約文の差分に対応する単語“賞名”との関係度を補助情報１１６から特定し、特定された関係度を低くする。抽出更新部１９２は、重要文と修正情報の差分とに基づいて、重要文の単語“証明”と修正情報の差分に対応する単語“照明”との関係度を補助情報１１６から特定し、特定された関係度を高くする。

これにより、より正確な単語が、クエリに含まれる。よって、情報処理装置１００は、正確性の高い要約文を選択することができる。

【０１０１】

図２７は、実施の形態４の情報処理装置が実行する処理の例を示すフローチャートである。図２７の処理では、ステップＳ２０ａが実行される点が図２３の処理と異なる。そこで、図２７では、ステップＳ２０ａを説明する。そして、ステップＳ２０ａ以外の処理の説明は、省略する。

【０１０２】

（ステップＳ２０ａ）抽出更新部１９２は、出力部１９０が出力した要約テキスト内の要約文と、取得部１２０により取得された修正情報とを比較し、差分を抽出する。抽出更新部１９２は、差分に基づいて、補助情報１１６を更新する。

【０１０３】

実施の形態４によれば、情報処理装置１００は、正確性の高い要約文を選択することができる。

【０１０４】

以上に説明した各実施の形態における特徴は、互いに適宜組み合わせることができる。

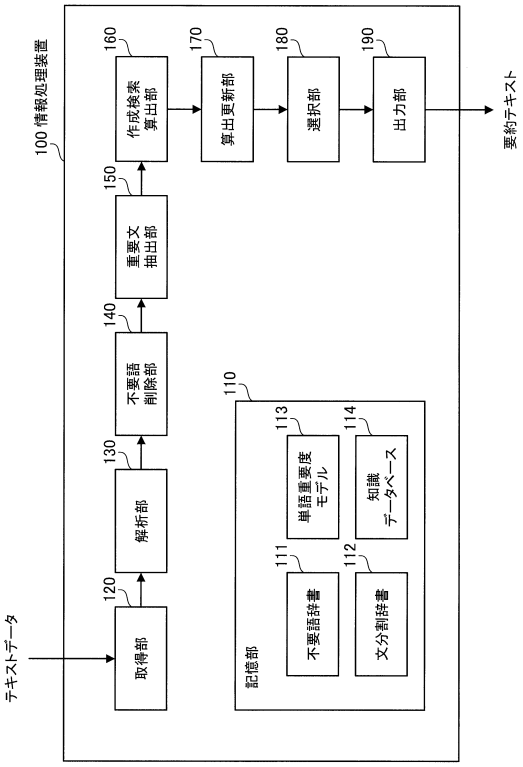
【符号の説明】

【０１０５】

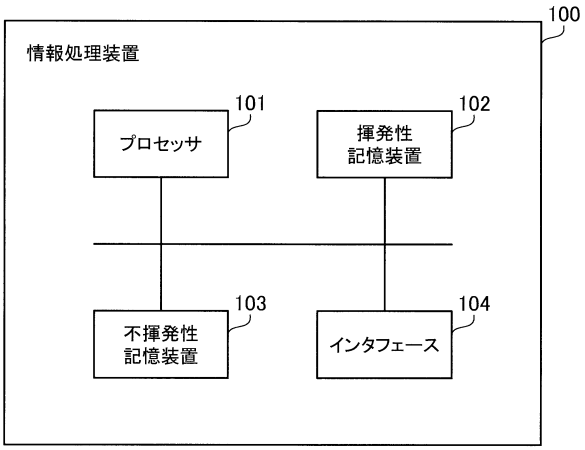
１００　情報処理装置、　１０１　プロセッサ、　１０２　揮発性記憶装置、　１０３　不揮発性記憶装置、　１０４　インタフェース、　１１０　記憶部、　１１１　不要語辞書、　１１２　文分割辞書、　１１３　単語重要度モデル、　１１４　知識データベース、　１１５　カテゴリ推定モデル、　１１６　補助情報、　１２０　取得部、　１３０　解析部、　１４０　不要語削除部、　１５０　重要文抽出部、　１６０　作成検索算出部、　１７０　算出更新部、　１８０　選択部、　１９０　出力部、　１９１　カテゴリ推定部、　１９２　抽出更新部、　２００　作成装置、　２１０　記憶部、　２１１　対話データベース、　２１２　報告書データベース、　２２０　単語重要度学習部、　２３０　データベース作成部、　２４０　カテゴリ推定学習部、　２５０　補助情報作成部。

【図面】

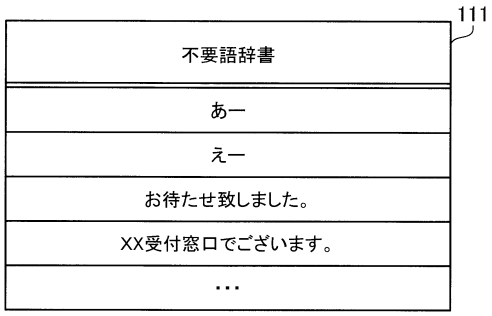
【図１】


100 情報処理装置
テキストデータ
要約テキスト
取得部
120
解析部
130
不要語削除部
140
重要文抽出部
150
作成検索算出部
160
算出更新部
170
選択部
180
出力部
190
記憶部
110
不要語辞書
111
文分割辞書
112
単語重要度モデル
113
知識データベース
114

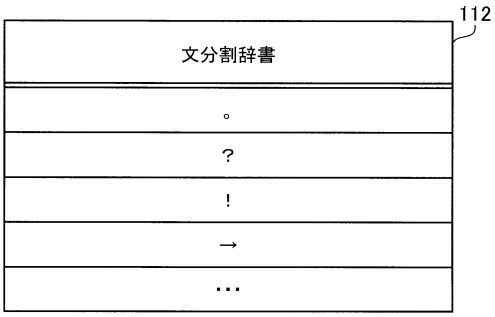
【図２】


情報処理装置
100
プロセッサ
101
揮発性記憶装置
102
不揮発性記憶装置
103
インタフェース
104

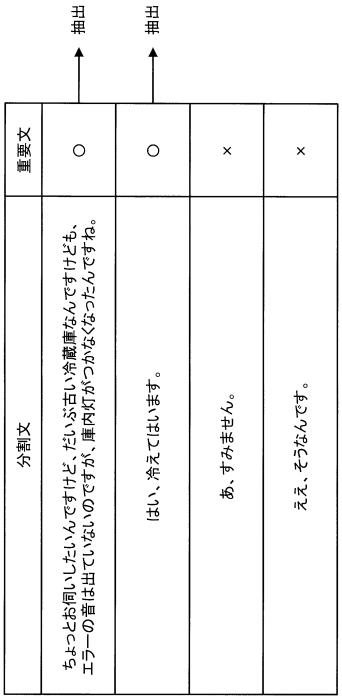
【図３】


111
不要語辞書
あー
えー
お待たせ致しました。
XX受付窓口でございます。
…

【図４】


112
文分割辞書
。
？
！
→
…

【図５】



分割文	重要文	
ちょっとお伺いしたいんですけど、だいぶ古い冷蔵庫なんですけども、エラーの音は出ていないのですが、庫内灯がつかなくなったんですね。	○	抽出
はい、冷えてはいます。	○	抽出
あ、すみません。	×	
ええ、そうなんです。	×	

【図６】

重要文	クエリ
ちょっとお伺いしたいんですけど、 だいぶ古い冷蔵庫なんですけども、 エラーの音は出ていないのですが、 庫内灯がつかなくなったんですね。	伺う/古い 古い/冷蔵庫 冷蔵庫/エラー エラー/音 音/出る 出る/居る_ない 居る_ない/庫内 庫内/灯 灯/付く_ない

【図７】

114

知識データベース			
文	インデックス登録クエリ	ステップ	カテゴリ
お手入れ押してもピッと言わない	手入れ/押す 押す/ぴっ ぴっ/言う_ない	受付	エアコン
お湯が出ない、エラーは出てない	湯/出る_ない 出る_ない/エラー エラー/出る_ない	受付	給湯器
エアコン取り付け依頼したい	エアコン/取り付け 取り付け/依頼	受付	エアコン
床暖房が暖かくならない	床暖房/温かい_ない	受付	床暖房
蛍光灯が急に消える	蛍光灯/急 急/消える	受付	照明
…	…	…	…

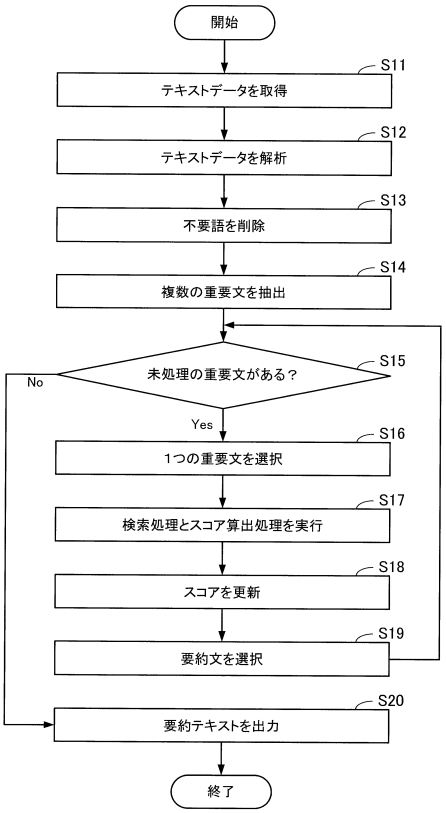
【図８】

検索された文	スコア
業務用の冷蔵庫、エラー音がなった	22.0
特に時間指定はでていません	21.9
庫内灯がつかず先ほど相談したら、10分切ってまた入れたらつくだろうと言われたが、つかなかった	21.4
加熱しようとしたらキーも受け付けず庫内灯も付かずターンテーブルも動かない	21.4
リコールは出ているか⇒出ていません	20.0
テレビのカタログの1月号はまだですか⇒まだ、出ていません	18.4
庫内灯が点かない	18.1
…	…

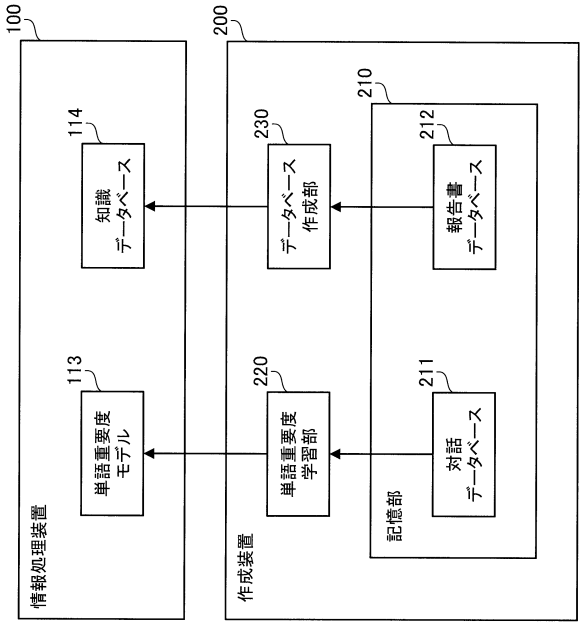
【図９】

検索された文	スコア
庫内灯が点かない	0.15
業務用の冷蔵庫、エラー音がなった	0.08
加熱しようとしたらキーも受け付けず庫内灯も付かずターンテーブルも動かない	0.06
庫内灯がつかず先ほど相談したら、10分切ってまた入れたらつくだろうと言われたが、つかなかった	0.06
特に時間指定はでていません	0.04
リコールは出ているか⇒出ていません	0.03
テレビのカタログの1月号はまだですか⇒まだ、出ていません	0.03

【図１０】



【図１１】

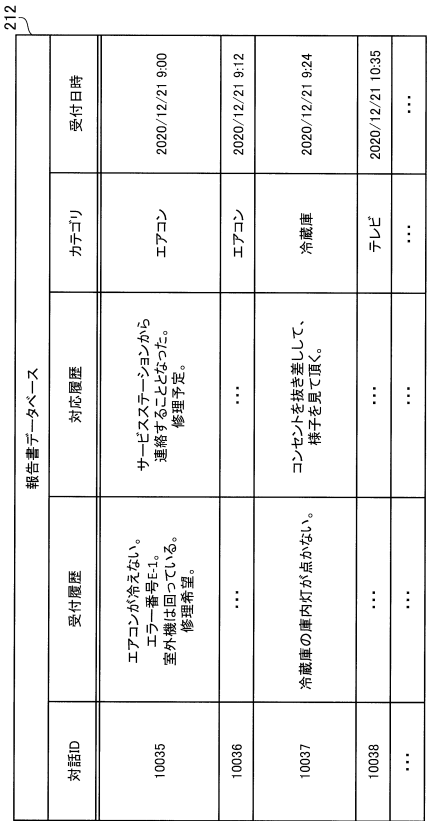


【図１２】

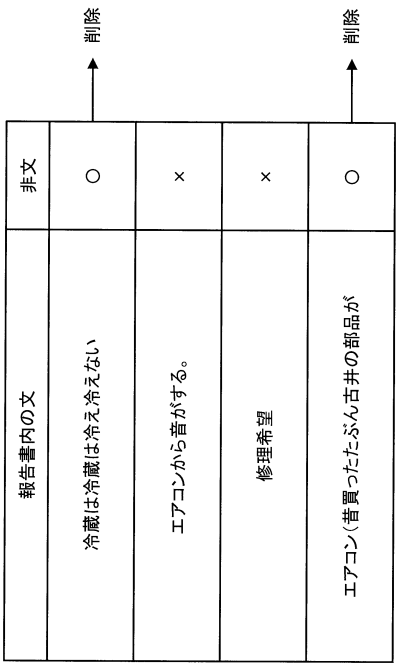
211 対話データベース

対話ID	音声認識結果	カテゴリ	受付日時
10035	お客様、お待たせいたしました。 XX受付のYYでございます。 どのようなご用件でしょうか？ エアコンが冷えないということでお間違えないでしょうか。 ご不便をおかけし申し訳ございません。	エアコン	2020/12/21 9:00
10036	…	エアコン	2020/12/21 9:12
10037	冷蔵庫の型番をお伺いしてもよろしいでしょうか。 はい、はい、はい。 裏面に記載されております。 はい。	冷蔵庫	2020/12/21 9:24
10038	…	テレビ	2020/12/21 10:35
…	…	…	…

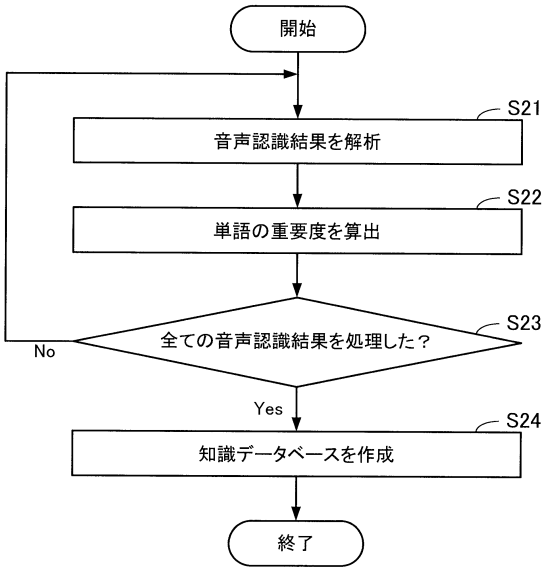
【図１３】



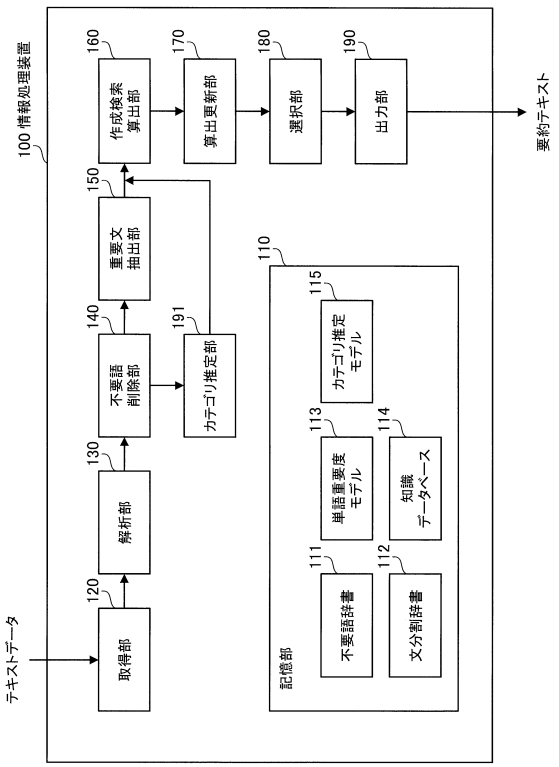
【図１４】



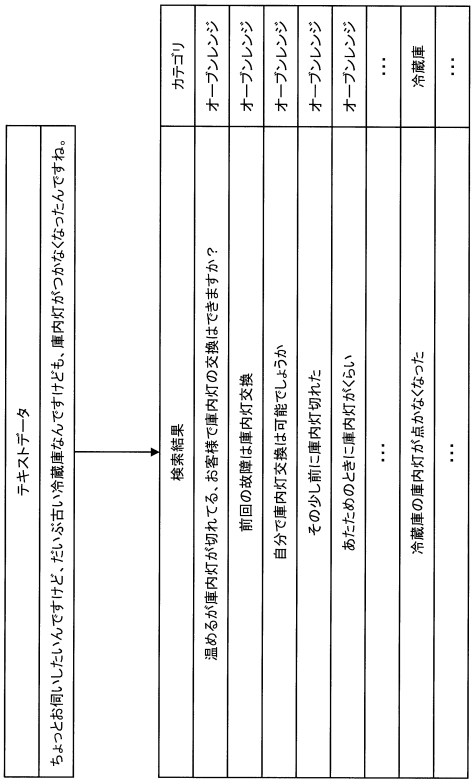
【図１５】



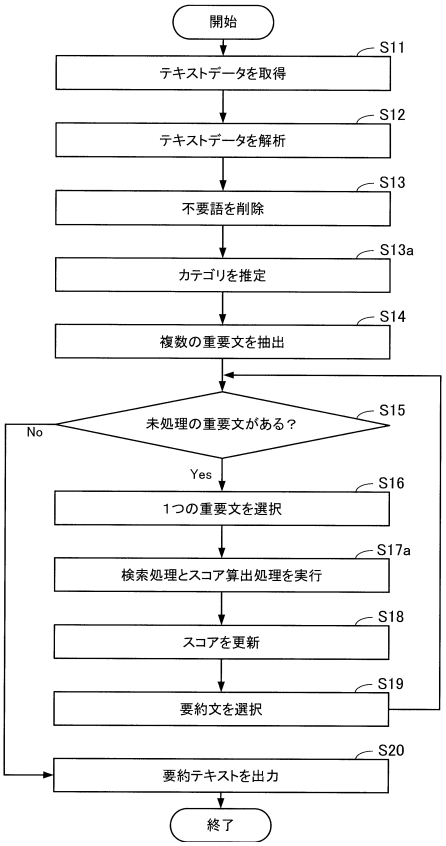
【図１６】



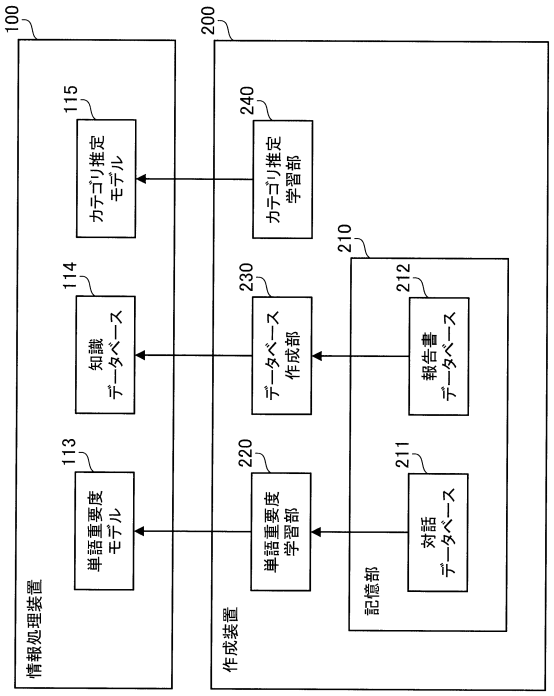
【図１７】



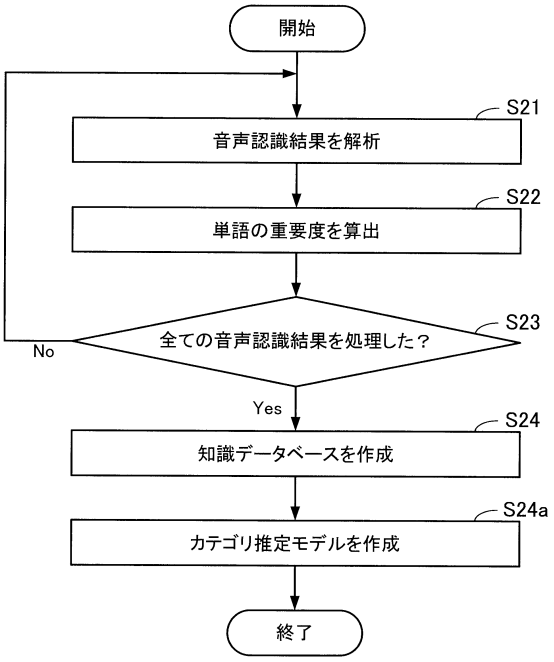
【図１８】



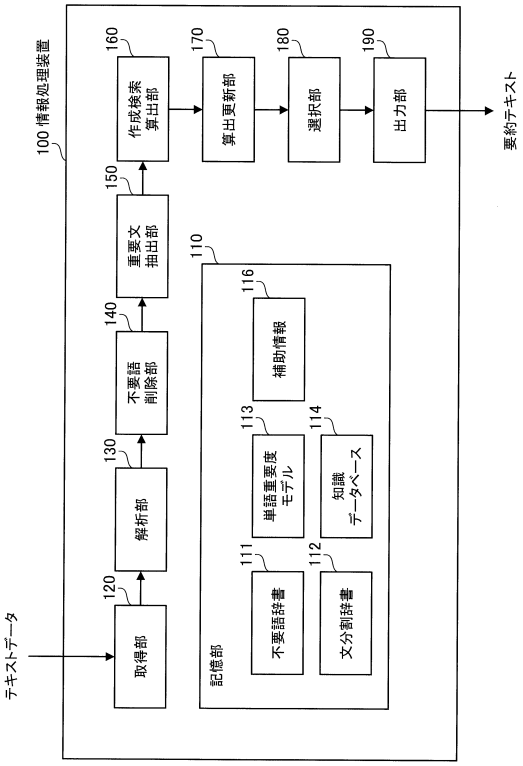
【図１９】



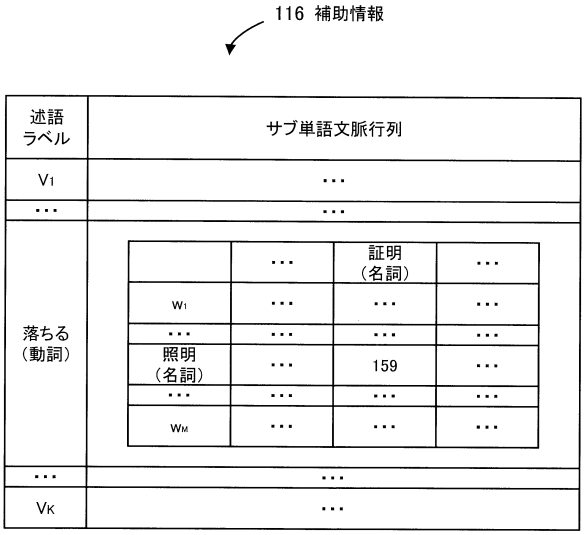
【図２０】



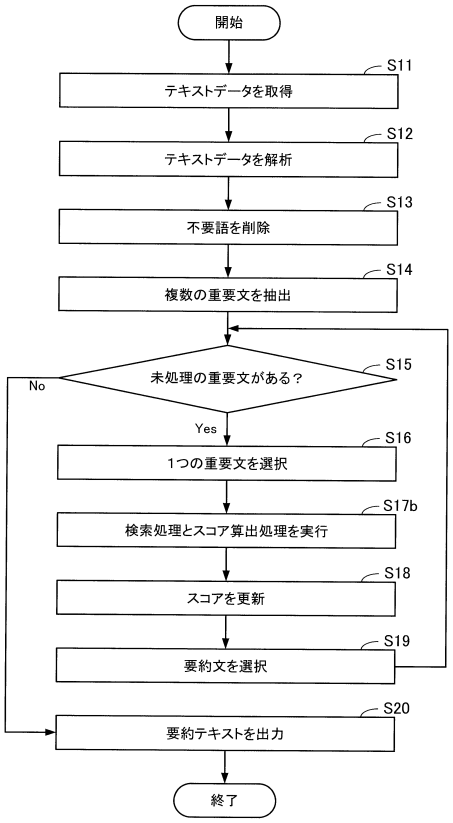
【図２１】



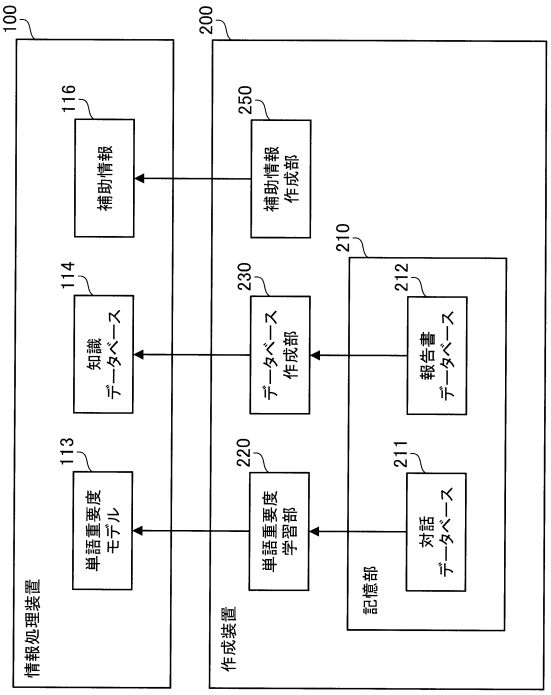
【図２２】



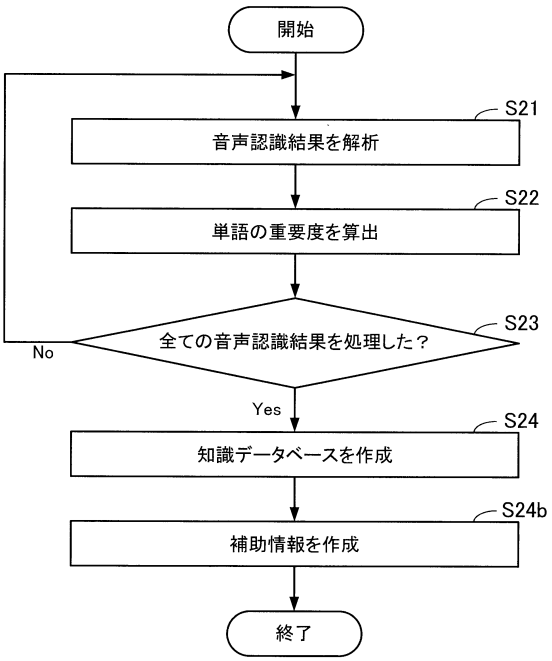
【図２３】



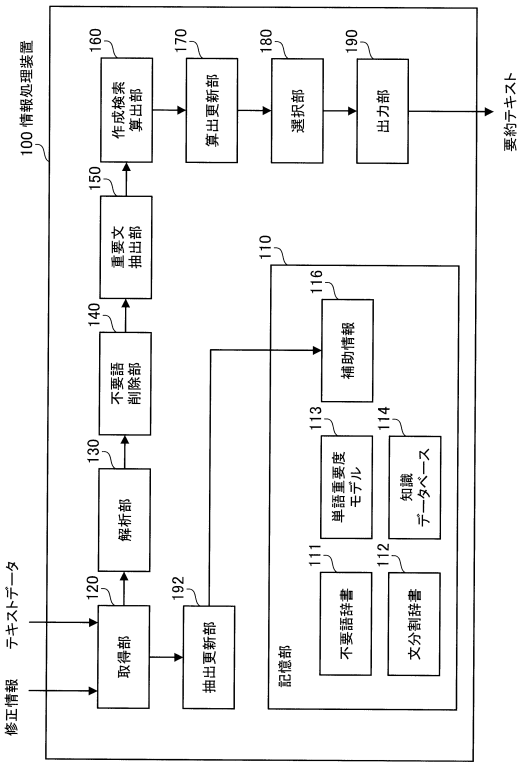
【図２４】



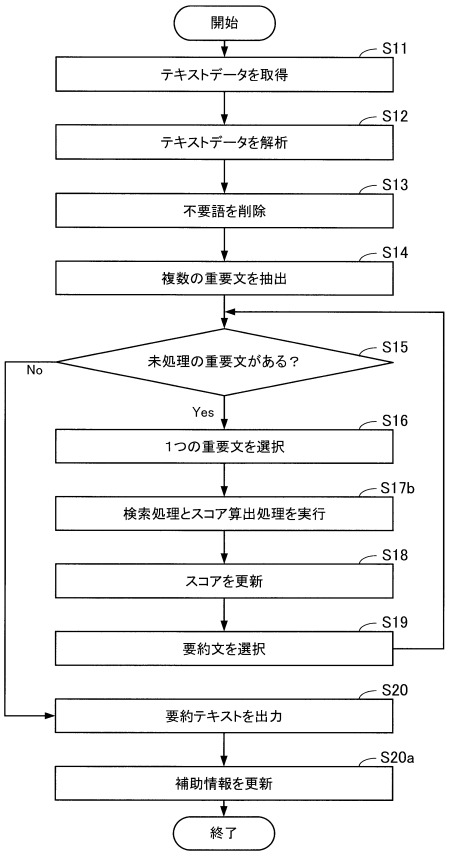
【図２５】


開始
S21
音声認識結果を解析
S22
単語の重要度を算出
S23
全ての音声認識結果を処理した？
No
Yes
S24
知識データベースを作成
S24b
補助情報を作成
終了

【図２６】


100 情報処理装置
160 作成検索式算出部
170 算出更新部
180 選択部
190 出力部
要約テキスト
150 重要文抽出部
140 不要語削除部
130 解析部
120 取得部
192 抽出更新部
テキストデータ
修正情報
110 記憶部
116 補助情報
113 単語重要度モデル
114 知識データベース
111 不要語辞書
112 文分割辞書

【図２７】


開始
S11
テキストデータを取得
S12
テキストデータを解析
S13
不要語を削除
S14
複数の重要文を抽出
S15
未処理の重要文がある？
No
Yes
S16
1つの重要文を選択
S17b
検索処理とスコア算出処理を実行
S18
スコアを更新
S19
要約文を選択
S20
要約テキストを出力
S20a
補助情報を更新
終了

フロントページの続き

三菱電機株式会社内
(72)発明者　伍井 啓恭
東京都千代田区丸の内二丁目７番３号 三菱電機株式会社内
審査官　甲斐 哲雄
(56)参考文献　特開２０２０－０７１６７６（ＪＰ，Ａ）
特開２００５－０８４８５８（ＪＰ，Ａ）
特開２０２０－０３５１３５（ＪＰ，Ａ）
(58)調査した分野　(Int.Cl.，ＤＢ名)
Ｇ０６Ｆ　１６／００－１６／９５８