(19)日本国特許庁(JP)　(12)特　許　公　報(B2)　(11)特許番号
特許第5234992号
(P5234992)
(45)発行日　平成25年7月10日(2013.7.10)　(24)登録日　平成25年4月5日(2013.4.5)

(51)Int.Cl.　F I
G06F 17/30 (2006.01)　G06F 17/30 180A
G06F 17/30 210D
G06F 17/30 210A

請求項の数 8　(全 21 頁)

(21)出願番号　特願2009-121438 (P2009-121438)
(22)出願日　平成21年5月19日 (2009.5.19)
(65)公開番号　特開2010-271800 (P2010-271800A)
(43)公開日　平成22年12月2日 (2010.12.2)
審査請求日　平成23年10月12日 (2011.10.12)

(73)特許権者　000004226
日本電信電話株式会社
東京都千代田区大手町二丁目３番１号
(74)代理人　100087446
弁理士　川久保　新一
(72)発明者　川島　晴美
東京都千代田区大手町二丁目３番１号　日本電信電話株式会社内
(72)発明者　甲谷　優
東京都千代田区大手町二丁目３番１号　日本電信電話株式会社内

審査官　野崎　大進

最終頁に続く

(54)【発明の名称】回答文書分類装置、回答文書分類方法及びプログラム

(57)【特許請求の範囲】
【請求項１】
質問文書に対する回答文書を、グループを代表するキーワードである分類キーワードを含むグループ毎に分類する回答文書分類装置において、
単語と品詞との組み合わせによって構成されているパターンであって、所定の条件を示すパターンである条件パターンを蓄積している条件パターン蓄積手段と；
入力テキスト集合に含まれている文が、上記条件パターン蓄積手段に蓄積されている条件パターンを含んでいるか否かを判定する条件パターン判定手段と；
上記条件パターンを含んでいれば、上記条件パターンの前方に配置されている単語の配列である条件キーワードを抽出する条件キーワード抽出手段と；
上記条件パターンを含んでいなければ、主格になる文節に含まれている単語である主格キーワードを抽出する主格キーワード抽出手段と；
上記条件キーワードおよび上記主格キーワードに含まれるキーワード毎に、着目しているキーワードが上記条件キーワードとして出現する回数である出現回数と、上記主格キーワードとして出現する回数である出現回数とを集計し、上記条件キーワードの出現回数と上記主格キーワードの出現回数とが所定の条件を満たす場合に、上記着目しているキーワードを分類キーワードであると判定する分類キーワード判定手段と；
を有することを特徴とする回答文書分類装置。
【請求項２】
質問文書に対する回答文書を、グループを代表するキーワードである分類キーワードを

含むグループ毎に分類する回答文書分類装置において、
単語と品詞との組み合わせによって構成されているパターンであって、所定の条件を示すパターンである条件パターンを蓄積している条件パターン蓄積手段と；
回答文書集合に含まれている文が、上記条件パターン蓄積手段に蓄積されている条件パターンを含んでいるか否かを判定する条件パターン判定手段と；
上記条件パターンを含んでいれば、上記条件パターンの前方に配置されている単語の配列である条件キーワードを抽出する条件キーワード抽出手段と；
上記条件パターンを含んでいなければ、主格になる文節に含まれている単語である主格キーワードを抽出する主格キーワード抽出手段と；
回答文書の文章毎に、条件キーワードであることを示す情報または主格キーワードであることを示す情報が追加されている解析結果と、回答文書を一意に示す回答文章ＩＤとを対応付けた分類キーワード候補を蓄積する分類キーワード候補蓄積手段と；
回答文章ＩＤが複数与えられた際に、回答文章ＩＤに該当する回答文書の解析結果の集合を、上記分類キーワード候補蓄積手段から取得し、上記解析結果の集合に含まれている条件キーワードおよび主格キーワードに含まれるキーワード毎に、着目しているキーワードが上記条件キーワードとして出現する回数である出現回数と、上記主格キーワードとして出現する回数である出現回数とを集計し、上記条件キーワードの出現回数と上記主格キーワードの出現回数とが所定の条件を満たす場合に、上記着目しているキーワードを分類キーワードであると判定する分類キーワード判定手段と；
を有することを特徴とする回答文書分類装置。
【請求項３】
請求項１または請求項２において、
上記条件キーワード抽出手段は、上記条件パターンを含む文節から、条件パターンを除いた語句に、条件パターンを含む文節へ係る０個以上の文節を追加した語句を、条件キーワードとして抽出する手段であることを特徴とする回答文書分類装置。
【請求項４】
請求項１または請求項２において、
上記主格キーワード抽出手段は、主格になる文節に含まれている名詞句に、主格になる文節へ係る０個以上の文節を追加した語句を、上記条件キーワードとして抽出する手段であることを特徴とする回答文書分類装置。
【請求項５】
請求項１～請求項４のいずれか１項において、
上記分類キーワード判定手段は、上記条件キーワードとしての出現回数が多いほど、上記分類キーワードとして判定され易くなる条件を用いて、上記分類キーワードを判定する手段であることを特徴とする回答文書分類装置。
【請求項６】
質問文書に対する回答文書を、グループを代表するキーワードである分類キーワードを含むグループ毎に分類する回答文書分類方法において、
単語と品詞との組み合わせによって構成されているパターンであって、所定の条件を示すパターンである条件パターンが蓄積されている条件パターン蓄積手段に蓄積されている条件パターンを、入力テキスト集合に含まれている文が含んでいるか否かを、条件パターン判定手段が判定し、記憶装置に記憶する条件パターン判定工程と；
上記条件パターンを含んでいれば、上記条件パターンの前方に配置されている単語の配列である条件キーワードを、条件キーワード抽出手段が抽出し、記憶装置に記憶する条件キーワード抽出工程と；
上記条件パターンを含んでいなければ、主格になる文節に含まれている単語である主格キーワードを、主格キーワード抽出手段が抽出し、記憶装置に記憶する主格キーワード抽出工程と；
上記条件キーワードおよび上記主格キーワードに含まれるキーワード毎に、着目しているキーワードが上記条件キーワードとして出現する回数である出現回数と、上記主格キー

ワードとして出現する回数である出現回数とを集計し、上記条件キーワードの出現回数と上記主格キーワードの出現回数とが所定の条件を満たす場合に、上記着目しているキーワードを分類キーワードであると、分類キーワード判定手段が判定し、記憶装置に記憶する分類キーワード判定工程と；

を有することを特徴とする回答文書分類方法。

【請求項７】

質問文書に対する回答文書を、グループを代表するキーワードである分類キーワードを含むグループ毎に分類する回答文書分類方法において、

単語と品詞との組み合わせによって構成されているパターンであって、所定の条件を示すパターンである条件パターンを蓄積している条件パターン蓄積手段に蓄積されている条件パターンを、回答文書集合に含まれている文が含んでいるか否かを、条件パターン判定手段が判定し、記憶装置に記憶する条件パターン判定工程と；

上記条件パターンを含んでいれば、上記条件パターンの前方に配置されている単語の配列である条件キーワードを、条件キーワード抽出手段が抽出し、記憶装置に記憶する条件キーワード抽出工程と；

上記条件パターンを含んでいなければ、主格になる文節に含まれている単語である主格キーワードを、主格キーワード抽出手段が抽出し、記憶装置に記憶する主格キーワード抽出工程と；

回答文書の文章毎に、条件キーワードであることを示す情報または主格キーワードであることを示す情報が追加されている解析結果と、回答文書を一意に示す回答文章ＩＤとを対応付けた分類キーワード候補とを、分類キーワード候補蓄積手段が蓄積し、記憶装置に記憶する分類キーワード候補蓄積工程と；

回答文章ＩＤが複数与えられた際に、回答文章ＩＤに該当する回答文書の上記解析結果の集合を、上記分類キーワード候補蓄積工程から取得し、上記解析結果の集合に含まれている条件キーワードおよび主格キーワードに含まれるキーワード毎に、着目しているキーワードが上記条件キーワードとして出現する回数である出現回数と、上記主格キーワードとして出現する回数である出現回数とを集計し、上記条件キーワードの出現回数と上記主格キーワードの出現回数とが所定の条件を満たす場合に、上記着目しているキーワードを分類キーワードであると、分類キーワード判定が判定し、記憶装置に記憶する分類キーワード判定工程と；

を有することを特徴とする回答文書分類方法。

【請求項８】

請求項１～請求項５のいずれか１項に記載の回答文書分類装置を構成する各手段としてコンピュータを機能させるプログラム。

【発明の詳細な説明】

【技術分野】

【０００１】

本発明は、質問についての回答を分類する装置に関し、特に、質問についての複数の回答を、キーワード毎に分類して提供する回答文書分類装置に関する。

【背景技術】

【０００２】

従来、インターネット等のネットワーク上で公開されている情報を探す手段として、キーワード検索がよく利用されている。キーワード検索は、利用者から１つまたは複数の検索キーワードを入力し、入力された検索キーワードの全て、またはいずれかを含む文書を検索して出力するサービスである。検索結果として出力される文書の数は、検索キーワードが一般的な語であるほど数が多く、利用者が目的に合った情報を見つけるのは困難である。そのために、利用者は検索キーワードを追加したり、調べたい分野に特徴的な検索キーワードを指定したりして、検索結果を絞り込む作業を行なっている。

【０００３】

また、キーワード検索でうまく目的に合った情報を見つけられない場合や、人に相談し

たいような場合に、利用者同士が質問と回答をやり取りするＱ＆Ａコミュニティーサイトを利用する場合がある。Ｑ＆Ａコミュニティーサイトでは、図５に示すように、質問したい利用者が投稿した１つの質問３２に対して、複数の回答者によって、回答３３、３４、３５が投稿され、サイト上でコミュニケーションが行われる。Ｑ＆Ａコミュニティーサイトにおいてもキーワード検索する機能が提供され、検索する対象を質問だけ、質問・回答の両方等と指定して検索することができる。この場合も、検索キーワードが一般的な語であれば検索結果の数は多くなり、目的に合った情報を探すのは容易ではない。

【０００４】

目的に合った情報を探しやすくするために、検索結果をグループ化して提示する手法がある。たとえば、利用者が入力した検索キーワードに対して、過去に入力された検索キーワードの履歴の中から、利用者が入力した検索キーワードに隣接して良く入力されるキーワードを関連語として抽出し、該関連語毎にグループ化して検索結果を表示する発明が知られている（たとえば、特許文献１参照）。この発明によれば、利用者が自分で検索結果を絞り込むためのキーワードを考える必要がなく、選択するだけで良いという利点がある。

【０００５】

また、回答文書を分類する方法として、クラスタリング技術を利用することができる。クラスタリング技術は、文章集合が与えられると、文章間の類似度を算出し、類似した文章毎にグループを作成する方法である。文章間の類似度は、単語出現頻度に基づく文章ベクトルで文章を表し、文章ベクトル間のコサイン類似度を適用する手法が広く用いられている。すなわち、文章d_nを文章ベクトル

【０００６】

【数１】

$$\vec{X}_n = (x_{n1}, x_{n2}, \cdots, x_{nv})$$

【０００７】

により表す場合、ｖは、単語集合Ｗ＝｛w_1，w_2，…，w_v｝中の単語の総数を示し、x_{ni}は、文章d_nにおける単語w_iの重みを示す。このとき、文章d_jと文章d_kの類似度は、各文章ベクトルがなす角

【０００８】

【数２】

$$\cos\theta_{j,k} = \frac{\vec{X}_j \cdot \vec{X}_k}{\left|\vec{X}_j\right|\left|\vec{X}_k\right|}$$

【０００９】

で表される。またw_iの重みは、単語の文章内での出現頻度ｔｆ（term frequency）をそのまま利用する場合や、出現頻度ｔｆに、単語出現数を全文章数で割った値の対数ｉｄｆを乗算したｔｆ－ｉｄｆ（term frequency/inverse document）を利用する。つまり、類似度の高い文章同士は、この単語の重みの傾向が似通っている文章同士であることを意味する。

【先行技術文献】

【特許文献】

【００１０】

【特許文献１】特許第４００９９３７号公報

【発明の概要】

【発明が解決しようとする課題】

【００１１】

本来、Ｑ＆Ａコミュニティーサイトの利用者の目的は、質問についての回答を得ることであるので、回答文書を分類して提供することが望ましい。上記特許文献１記載の従来技術を、Ｑ＆Ａコミュニティーサイトに適用する場合、検索する利用者は知りたい情報の分野に詳しくない場合が多く、適切なキーワードを入力しているとは限らない。このために、検索キーワードを分類用のキーワードとして利用するだけでは、回答文書を精度良く分類できないという問題がある。

【００１２】

また、クラスタリング技術によって回答文書を分類する場合、文章数の二乗に比例して処理時間がかかるので、キーワード検索結果の文章集合が多い場合、分類結果が出力されるまでに時間がかかるという問題がある。

【００１３】

本発明は、検索キーワードに含まれないキーワードも利用して、短時間で回答文書を分類することができる回答文書分類装置、回答文書分類方法およびプログラムの提供を目的とする。

【課題を解決するための手段】

【００１４】

本発明は、質問文書に対する回答文書を、グループを代表するキーワードである分類キーワードを含むグループ毎に分類する回答文書分類装置において、単語と品詞との組み合わせによって構成されているパターンであって、所定の条件を示すパターンである条件パターンを蓄積している条件パターン蓄積手段と、入力テキスト集合に含まれている文が、上記条件パターン蓄積手段に蓄積されている条件パターンを含んでいるか否かを判定する条件パターン判定手段と、上記条件パターンを含んでいれば、上記条件パターンの前方に配置されている単語の配列である条件キーワードを抽出する条件キーワード抽出手段と、上記条件パターンを含んでいなければ、主格になる文節に含まれている単語である主格キーワードを抽出する主格キーワード抽出手段と、上記条件キーワードおよび上記主格キーワードに含まれるキーワード毎に、着目しているキーワードが上記条件キーワードとして出現する回数である出現回数と、上記主格キーワードとして出現する回数である出現回数とを集計し、上記条件キーワードの出現回数と上記主格キーワードの出現回数とが所定の条件を満たす場合に、上記着目しているキーワードを分類キーワードであると判定する分類キーワード判定手段とを有する回答文書分類装置である。

【発明の効果】

【００１５】

本発明によれば、質問についての回答文書を検索したい利用者が、検索キーワードとして入力しないキーワードも利用して、回答文書を分類することができ、質問者は、分類された情報を選択するだけで、所望の情報を取得することができるという効果を奏する。

【００１６】

また、本発明によれば、回答文書集合に含まれるキーワードの出現回数に基づいて分類キーワードの判定を行うので、処理が高速であるという効果を奏する。

【図面の簡単な説明】

【００１７】

【図１】本発明の原理を説明する図である。

【図２】本発明の実施例１である回答文書分類装置１００の構成を示す図である。

【図３】条件パターン蓄積手段１１が蓄積している条件パターンの例を示す図である。

【図４】条件パターン判定手段１２の処理を示すフローチャートである。

【図５】Ｑ＆Ａコミュニティサイトの構成を示す図である。

【図６】条件パターン判定手段１２が判定した形態素解析結果を示す図である。

【図７】実施例１において、条件キーワード抽出手段１３が抽出した係り受け解析した結果の例を示す図である。

【図８】主格キーワード抽出手段１４の動作を示す図である。

【図９】分類キーワード判定手段１５の動作の説明図である。

【図１０】分類キーワード判定手段１９の動作を示すフローチャートである。
【図１１】本発明の実施例２である回答文書分類装置２００を示すブロック図である。
【図１２】実施例２における分類キーワード蓄積手段１８の蓄積例を示す図である。
【発明を実施するための形態】
【００１８】
発明を実施するための形態は、以下の実施例である。
【実施例１】
【００１９】
本発明では、グループを代表するキーワードであり、回答文書を分類するためのキーワードである分類キーワードを、回答文書の中から抽出する。質問された内容に詳しい回答者は、質問文書中に含まれていない語句を用いて、詳細な情報を提供することが考えられるので、上記分類キーワードを回答文書から抽出する。質問文書についての回答文書を、分類キーワード毎にグループ化して利用者に提供する。これによって、利用者は、分類キーワードを手がかりに、所望の情報を容易に取得することができる。
【００２０】
本発明は、質問文書についての回答文書を、分類キーワード毎にグループ化する回答文書分類装置において、与えられた回答文書集合に含まれているキーワード（単語）から、グループを代表するキーワードである分類キーワードを判定する。
【００２１】
まず、回答文書の特徴について説明する。たとえば、次のような質問文書があった場合について考える。
【００２２】
「６月にテーマパークＸに遊びに行こうと計画しています。雨に備えて傘か合羽を用意しようと思っていますが、どちらがいいでしょう？」
この質問に対して３人の回答者がそれぞれ以下の回答をしたとする。
【００２３】
回答者Ａ：「傘は人が多いと迷惑になるので、合羽がお勧めです。」
回答者Ｂ：「大人だけなら、傘で大丈夫ですよ。小さい子供が一緒なら、合羽が楽です。」
回答者Ｃ：「大人は傘で良いと思うよ。」
回答者Ａは、質問文書に記載された内容の範囲で自分の意見を述べている。回答者Ｂは、質問文書には記載されていない「大人だけ」なのか「子供が一緒」なのかの情報を追加して、それぞれの場合について自分の意見を述べている。
【００２４】
質問者は、「大人だけ」か「子供が一緒」かによって、お勧めの情報が異なることを、質問した時点では知らないが、回答文書の分類キーワードとして、「大人だけ」、「子供が一緒」というキーワードが提示されれば、自分の状況に合わせて分類キーワードを選択することができ、的確な回答文書を参照することが可能になる。
【００２５】
回答者Ｃは「大人は」と述べ、条件パターンである「○○なら」を用いていない。なお、上記「条件パターン」は、明示的に条件であることが分かる文節である。
【００２６】
回答者Ｃによる回答文書中、「大人は」という主格となる文節に、暗黙的に「大人の場合」という条件を含めて記述している。回答文書において、ハ格である「○○は」は、○○について説明をする場合に用いられる場合が多い。そこで、「○○は」という文節の○○の中にも、回答文書を分類する場合に適したキーワードが含まれていると考える。
【００２７】
図１は、本発明の原理を説明するフローチャートである。
【００２８】
まず、入力されたテキスト集合に含まれている各文に対して、上記条件パターンが記述

されているか否かを判定する（Ｓ１）。上記「条件パターン」は、たとえば「○○なら、××です。」という文における「○○なら」中の「なら」である。

【００２９】

なお、条件パターンとして、次のパターンもが考えられる。

【００３０】

「○○であれば」中の「であれば」、

「○○の場合」中の「の場合」、

「○○ですと」中の「ですと」。

【００３１】

文中に、上記条件パターンが記述されていると判定さると、「○○なら」という条件を示す文節中の単語「○○」に含まれているキーワードを、「条件キーワード」として抽出する（Ｓ２）。つまり、「条件キーワード」は、条件パターンを含む文節から、条件パターンを除いた単語に含まれているキーワードである。また、条件パターンが記述されていないと判定されると、「○○は」のように、主格になる文節が存在するかどうかを調べ、主格になる文節が存在すれば、主格になる文節「○○」に含まれているキーワードを、「主格キーワード」として抽出する（Ｓ３）。

【００３２】

つまり、「主格キーワード」は、主格になる文節に含まれている名詞句である。

【００３３】

条件キーワードと主格キーワードとは、少なくとも１つの単語によって構成されている。

【００３４】

次に、与えられている入力テキスト集合全体から抽出したキーワードのうちで、着目しているキーワードが、上記条件キーワードとして出現する回数である出現回数と、上記主格キーワードとして出現する回数である出現回数とを求め、上記条件キーワードの出現回数と上記主格キーワードの出現回数とが所定の条件を満たす場合に、上記着目しているキーワードを分類キーワードであると判定する（Ｓ４）。

【００３５】

つまり、「分類キーワード」は、条件キーワードの出現回数と、主格キーワードの出現回数とが、所定の条件を満たす場合におけるキーワードである。

【００３６】

図２は、本発明の実施例１である回答文書分類装置１００の構成を示す図である。

【００３７】

回答文書分類装置１００は、テキスト入力手段２０とテキスト出力手段４０とに、接続され、条件パターン蓄積手段１１と、条件パターン判定手段１２と、条件キーワード抽出手段１３と、主格キーワード抽出手段１４と、分類キーワード判定手段１５とを有する。

【００３８】

条件パターン判定手段１２は、テキスト入力手段２０から回答文書の集合を入力すると、入力した回答文書を文単位に分割し、各文に条件パターンが含まれているかどうかを判定し、そして、条件パターンが含まれている文を、条件キーワード抽出手段１３へ送り、条件パターンが含まれていない文を、主格キーワード抽出手段１４へ送る。

【００３９】

条件キーワード抽出手段１３は、条件パターン判定手段１２から、条件パターンを含む形態素解析結果の文と、検出された条件パターンを示す情報とを受け取り、条件パターンに一致する箇所の前方に存在するキーワードを、条件キーワードとして抽出し、記憶装置に記憶する手段である。

【００４０】

主格キーワード抽出手段１４は、条件パターン判定手段１２から、条件パターンを含まないと判定された文を受け取り、係り受け解析し、記憶装置に記憶する。この解析結果中に、主格となる文節が存在するかどうかを調べ、主格となる文節が存在すれば、主格とな

る文節に含まれているキーワード（名詞句）を、「主格キーワード」として抽出し、記憶装置に記憶する。

【００４１】

分類キーワード判定手段１５は、条件キーワード抽出手段１３が抽出した条件キーワードと、主格キーワード抽出手段１４が抽出した主格キーワードとを入力し、キーワード毎に、所定の条件（後述の式（１）、式（２）に示す条件）を満たすかどうかを判定し、この予め設定された条件を満たすキーワードを、分類キーワードとして、記憶装置に記憶し、この分類キーワードを出力する。分類キーワード判定手段１５は、具体的には、分類キーワードと、このキーワードを含む回答文章ＩＤと、文番号の情報とを組みにして、テキスト出力手段４０に送る。

【００４２】

なお、テキスト入力手段２０とテキスト出力手段４０との間に、質問・回答文書蓄積手段３０が接続されている。質問・回答文書蓄積手段３０は、Ｑ＆Ａコミュニティーサイトにおいて投稿された質問文書、回答文書を多数、蓄積している。

【００４３】

テキスト入力手段２０は、質問・回答文書蓄積手段３０から、後述する特定の条件に合致する回答文書の集合を取得し、条件パターン判定手段１２に送る。

【００４４】

後述する特定の条件に合致する回答文書の集合を取得するための条件は、特定のキーワードを含む質問に回答した回答文書の集合等である。

【００４５】

テキスト出力手段４０は、分類キーワード判定手段１５が出力した分類キーワードを入力し、分類キーワード毎に、回答文書を分類し、記憶装置に記憶し、出力する。

【００４６】

図３は、条件パターン蓄積手段１１が蓄積している条件パターンの例を示す図である。

【００４７】

条件パターン蓄積部１１は、条件パターンとして抽出したい単語と、この単語の出現順序パターンとを、図３に示すように予め蓄積する。図３に示す蓄積例では、１行に１つの条件パターンが記述されている。１つの条件パターンは、「品詞：読み」がカンマで接続され、左から順に連続して単語が出現するパターンである。４行目の単語パターン、「判定詞：デ、動詞語幹：ア、動詞活用語尾：レ、動詞接尾辞：バ」は、「であれば」という条件パターンを抽出するための記述である。

【００４８】

条件パターン判定手段１２は、テキスト入力手段２０から回答文書の集合を入力すると、入力した回答文書を文単位に分割し、各文に条件パターンが含まれているかどうかを判定する。そして、条件パターンが含まれている文を、条件キーワード抽出手段１３へ送り、条件パターンが含まれていない文を、主格キーワード抽出手段１４へ送る。なお、各回答文書には、回答文書を一意に特定する回答文章ＩＤがそれぞれ付与される。

【００４９】

図４は、条件パターン判定手段１２の動作を示すフローチャートである。

【００５０】

まず、条件パターン蓄積手段１１に蓄積されている条件パターンを全て読み込み（Ｓ５１）、テキスト入力手段２０から受け取った回答文書について、回答文書毎に、回答文章ＩＤに対応する回答文書を文単位に分割し、この分割された文に、回答文書内での出現番号を付与する（Ｓ５２、Ｓ５３）。

【００５１】

続いて、未処理の文番号が存在すれば（Ｓ５４）、未処理の文番号を１つ選択し、形態素解析処理を行う（Ｓ５５）。形態素解析結果の単語の品詞と読みとが、条件パターンに一致する箇所があるかどうかを調べ（Ｓ５６）、条件パターンに一致する箇所があれば（Ｓ５６、ＹＥＳ）、処理対象の回答文章ＩＤと、文番号と、文と条件パターンとが一致し

た箇所とを、条件キーワード抽出手段１３へ送る（Ｓ５７）。条件パターンと一致する箇所がなければ（Ｓ５６、ＮＯ）、処理対象の回答文章ＩＤと、文番号と、文とを、主格キーワード抽出手段１４へ送る（Ｓ５８）。

【００５２】

図５は、Ｑ＆Ａコミュニティサイトの構成例を示す図である。

【００５３】

図５に示す回答文書３３、３４、３５を、条件パターン判定手段１２が入力した場合における具体的な動作について説明する。回答文書３３、３４、３５の文章ＩＤを、それぞれ、３３、３４、３５とする。まず、未処理の回答文章ＩＤ３３が選択され（Ｓ５２、Ｓ５３）、この選択された回答文章ＩＤ３３は、１文から構成されているので（Ｓ５３）、１文が選択され、形態素解析処理が行われる（Ｓ５４、Ｓ５５）。

【００５４】

図６は、条件パターン判定手段１２が判定した形態素解析結果を示す図である。

【００５５】

形態素解析結果６１において、１行目に回答文章ＩＤが記載され、２行目に文番号が記載され、３行目以降に、形態素毎の表記、品詞、読みが記載されている。形態素解析結果６１の中に、条件パターンと一致する単語出現パターン（単語が出現するパターン）があるかどうかを調べる（Ｓ５６）。図６に示す形態素解析結果６１には、条件パターンと一致する箇所がないので（Ｓ５６、ＮＯ）、形態素解析結果６１を主格キーワード抽出手段１４へ送る。

【００５６】

回答文章ＩＤ３３には他に文がないので（Ｓ５４、ＮＯ）、次の未処理の回答文書３４を選択する（Ｓ５２）。回答文章ＩＤ３４の文章は、２つの文に分割され（Ｓ５３）、１番の文が選択され、形態素解析処理される（Ｓ５５）。形態素解析結果６２に、条件パターンに一致する単語出現パターンがあるかどうかを調べると（Ｓ５６）、条件パターン「判定詞：ナラ」と一致する箇所６３が発見される。

【００５７】

条件パターンと一致する箇所６３があるので（Ｓ５６、ＹＥＳ）、形態素解析結果６２中で、条件パターンと一致する箇所を明示するために、「＊」を付与し、条件キーワード抽出部１３へ送る（Ｓ５７）。続いて、回答文章ＩＤ３４の文章の２番目の文（形態素解析結果６４）について、Ｓ５４～Ｓ５６の処理を実行し、条件パターンに一致する箇所６５が発見される。

【００５８】

次に、回答文章ＩＤ３５の文章について、Ｓ５２～Ｓ５６の処理を実行し、形態素解析結果６６に、条件パターンが存在しないので（Ｓ５６、ＮＯ）、形態素解析結果６６を、主格キーワード抽出手段１４へ送る。未処理の回答文章ＩＤがなくなったので（Ｓ５２、ＮＯ）、処理を終了する。

【００５９】

条件キーワード抽出手段１３は、条件パターン判定手段１２から、条件パターンを含む形態素解析結果の文と、条件パターンに一致する箇所（検出された条件パターン）を示す情報とを受け取り、条件パターンに一致する箇所の前方に存在するキーワードを、条件キーワードとして抽出する。

【００６０】

条件キーワードを抽出する方法として、次の［方法１］～［方法３］が考えられる。

［方法１］検出された条件パターンに最も近い前方の名詞を、条件キーワードとして抽出する方法、

［方法２］検出された条件パターンを含む文節から、条件パターンを除いた語句を、条件キーワードとして抽出する方法、

［方法３］上記［方法２］において、検出された条件パターンを含む文節に係る文節を、Ｎ個追加したものを、条件キーワードとして抽出する方法。

【００６１】
上記［方法３］における整数値Ｎを、予め設定するようにしてもよく、利用者が設定できるようにしてもよい。
【００６２】
図７は、実施例１において、条件キーワード抽出手段１３が抽出した係り受け解析した結果の例を示す図である。
【００６３】
つまり、図７は、回答文章ＩＤ３４の文章の文番号１の形態素解析結果６２を、係り受け解析した結果７１と、文番号２の形態素解析結果６４を、係り受け解析した結果７２とを示す図である。係り受け解析した結果７１において、上記［方法１］では、「大人」が、条件キーワードとして抽出され、上記［方法２］では、「大人だけ」が、条件キーワードとして抽出され、上記［方法３］では、「大人だけなら」に係る文節が存在しないので、「大人だけ」が、条件キーワードとして抽出される。
【００６４】
係り受け解析した結果７２において、上記［方法１］では、「一緒」が、条件キーワードとして抽出され、上記［方法２］でも、「一緒」が、条件キーワードとして抽出され、上記［方法３］では、条件パターンを含む文節に係る文節の数Ｎ＝１であれば、「子供が一緒」が、条件キーワードとして抽出される。抽出された条件キーワードは、抽出された回答文章ＩＤと文番号とに対応付けられ、分類キーワード判定手段１５へ渡される。
【００６５】
主格キーワード抽出手段１４は、条件パターン判定手段１２から、条件パターンを含まないと判定された文を受け取り、係り受け解析を行う。この解析結果中に、主格となる文節が存在するかどうかを調べ、主格となる文節が存在すれば、主格となる文節に含まれているキーワード（名詞句）を、「主格キーワード」として抽出する。本発明では、主格となる文節として、ハ格「ＯＯは」を抽出する。ハ格「○○は」は、動作主体となる場合もあるが、名詞句「○○」について説明する場合にも用いられる。このために、回答文書の中で、名詞句「○○」の説明をしている箇所は、質問者にとって有益な情報となる可能性があると考え、ハ格「○○は」を抽出することとする。
【００６６】
主格キーワードとして、ハ格の名詞句部分「○○」だけを抽出する場合と、条件キーワードを抽出する場合における上記［方法３］のように、主格となる文節に係る文節を、Ｎ個追加したものを、主格キーワードとして抽出するようにしてもよい。
【００６７】
図８は、主格キーワード抽出手段１４の動作を示す図である。
【００６８】
図８（１）は、たとえば、回答文章ＩＤ３３、３５の文が、主格キーワード抽出部１４に入力された場合、回答文章ＩＤ３３を、係り受け解析した結果を示す図である。図８（２）は、回答文章ＩＤ３５の文を、係り受け解析した結果を示す図である。
【００６９】
図８（１）に示す例において、主格となる文節は、「傘は」であり、この文節から、名詞「傘」を、主格キーワードとして抽出する。これと同様に、図８（２）に示す例において、主格となる文節は、「大人は」であり、この文節から名詞「大人」を、主格キーワードとして抽出する。抽出された主格キーワード「傘」、「大人」を、それぞれ回答文章ＩＤと文番号とを対応付け、「分類キーワード」として分類キーワード判定手段１５へ送る。なお、分類キーワードと回答文章ＩＤと文番号を対応づけて分類キーワード判定手段１５へ送るのであり、回答文章ＩＤと文番号が分類キーワードに含まれるわけではない。
【００７０】
分類キーワード判定手段１５は、条件キーワード抽出手段１３が抽出した条件キーワードと回答文章ＩＤと文番号、主格キーワード抽出手段１４が抽出した主格キーワードと回答文章ＩＤと文番号を入力し、キーワード毎に、後述の式（１）、式（２）に示す条件を

満たすかどうかを判定し、この予め設定された条件を満たすキーワードを、分類キーワードとして出力する。

【００７１】

つまり、上記「分類キーワード」は、条件キーワードと、主格キーワードとが、予め設定された条件を満たすキーワードである。

【００７２】

上記予め設定された条件は、キーワードｗが条件キーワードとして出現する回数である出現回数Ｘ（ｗ）と、キーワードｗが主格キーワードとして出現する回数である出現回数Ｙ（ｗ）とに応じた条件である。たとえば、次の式（１）を条件とし、判定式Ｆ（ｗ）を求める。

【００７３】

Ｆ（ｗ）＝αＸ（ｗ）＋（１－α）Ｙ（ｗ）…式（１）

Ｆ（ｗ）≧ｔｈ…式（２）

ここで、０≦α≦１であり、αを１に近い値に設定すると、上記式（１）は、条件キーワードとしての出現回数が多いキーワードほど、高い値になる。さらに、上記式（２）によって、予め設定された閾値ｔｈ以上の値をとるキーワードのみを、分類キーワードとして判定することができる。

【００７４】

出現回数を集計するときに、厳密に一致するキーワードだけを、出現回数としてカウントする場合と、一部が一致する場合（部分一致）も、出現回数としてカウントする場合とが考えられる。以下では、部分一致した場合に出現回数としてカウントする場合について説明する。

【００７５】

図９は、分類キーワード判定手段１５の動作の説明図である。

【００７６】

たとえば、条件キーワード抽出手段において「大人だけ」「小さな子供」の２語が抽出され、主格キーワード抽出手段１４において「傘」「大人」の２語が主格キーワードとして抽出された場合について考える。

【００７７】

条件キーワード情報９１は、１行に「条件キーワード　回答文章ＩＤ　文番号」の順にスペース区切りで記載されている。主格キーワード情報９２は、１行に「主格キーワード　回答文章ＩＤ　文番号」の順にスペース区切りで記載されている。

【００７８】

図１０は、分類キーワード判定手段１５の動作を示すフローチャートである。

【００７９】

分類キーワード判定手段１５は、分類キーワード候補を選ぶ（Ｓ１０１）。分類キーワード候補は、入力された条件キーワードと主格キーワードとの中から、重複を除去したキーワードの集合である。

【００８０】

集計結果９３における分類キーワード候補の列に記載する４種類が選ばれる。次に、未処理の分類キーワード候補が存在すれば（Ｓ１０２、ＹＥＳ）、分類キーワード候補を１つ選択し、この分類キーワード候補の条件キーワード出現回数と主格キーワード出現回数とを０に初期化する（Ｓ１０３）。

【００８１】

次に、入力された条件キーワードを、１つずつ「比較キーワード」として選択し、この選択された比較キーワードを、分類キーワード候補と比較する。なお、分類キーワード候補と比較する条件キーワードを、便宜上、「比較キーワード」と表現する。

【００８２】

未処理の条件キーワードが存在すれば（Ｓ１０４）、未処理の条件キーワードを、比較キーワードとして、１つ選択し、分類キーワード候補と比較する（Ｓ１０５）。

【００８３】
この比較の結果、完全一致の場合、または分類キーワード候補に比較キーワードを含む場合（部分一致の場合）には（Ｓ１０６、ＹＥＳ）、条件キーワード出現回数を１増やす（Ｓ１０７）。これ以外の場合は（Ｓ１０６、ＮＯ）、Ｓ１０４に戻り、未処理の条件キーワードがなくなるまで、Ｓ１０４～Ｓ１０７の処理を実行する。未処理の条件キーワードがなくなれば（Ｓ１０４、ＮＯ）、主格キーワード似ついて比較する処理に移行する。
【００８４】
未処理の主格キーワードが存在すれば（Ｓ１０８、ＹＥＳ）、未処理の主格キーワードを、比較キーワードとして、１つ選択し、分類キーワード候輔と比較する（Ｓ１０９）。この比較の結果、完全一致である場合、または分類キーワード候補に比較キーワードを含む場合（部分一致の場合）には（Ｓ１１０、ＹＥＳ）、主格キーワード出現回数を、１増やす（Ｓ１１１）。これ以外の場合（Ｓ１１０、ＮＯ）、Ｓ１０８に戻り、未処理の主格キーワードがなくなるまで、Ｓ１０８～Ｓ１１１の処理を実行する。
【００８５】
未処理の主格キーワードがなくなると（Ｓ１０８、ＮＯ）、次の分類キーワード候補の比較処理に移行する。未処理のキーワードを選択し、条件キーワード出現回数と主格キーワード出現回数とをカウントする処理（Ｓ１０２～Ｓ１１１）を、分類キーワード候補の全てについて実行し、処理を終了する。
【００８６】
上記比較によって、図９に示す集計結果９３を得ることができる。
【００８７】
図９に示す条件キーワード情報９１と、主格キーワード情報９２との例では、まず、分類キーワード候補「大人だけ」を選択し（Ｓ１０２）、入力された条件キーワードと比較する。条件キーワード「大人だけ」を選択し、比較すると（Ｓ１０５）、完全一致するので、条件キーワード出現回数を、１増やす（Ｓ１０７）。
【００８８】
次に、条件キーワード「小さな子供」と比較するが、Ｓ１０６の条件を満たさないので、未処理の条件キーワードが存在するかどうかを調べる（Ｓ１０４）。未処理の条件キーワードが存在しなければ（Ｓ１０４、ＮＯ）、主格キーワードとの比較処理を実行する。未処理の主格キーワード「傘」を選択し、分類キーワード候補「大人だけ」と比較する（Ｓ１０９）。この比較の結果、完全一致または部分一致であるというＳ１１０の条件を満たさないので、処理Ｓ１０８に戻る。未処理の主格キーワード「大人」が存在するので、「大人」を選択し、分類キーワード候補「大人だけ」と比較する（Ｓ１０９）。
【００８９】
分類キーワード候補「大人だけ」に、上記比較キーワード「大人」が含まれているので、主格キーワード出現回数を、１増やす。未処理の主格キーワードが無くなるので（Ｓ１０８、ＮＯ）、未処理の分類キーワード候補「小さな子供」を選択し（Ｓ１０３）、（Ｓ１０４～Ｓ１１１）の処理を実行する。この結果、分類キーワード候補「小さな子供」について、条件キーワード出現回数が、１増え、１になり、分類キーワード候補「傘」について、主格キーワード出現回数が１増え、１になる。
【００９０】
また、分類キーワード候補「大人」と、条件キーワード「大人だけ」とを比較すると、部分一致はするが、「大人」に「大人だけ」が含まれていないので、条件キーワード出現回数をカウントすることができない。分類キーワード候補「大人」と、主格キーワード「大人」とを比較すると、完全一致するので、主格キーワード出現回数を、１増やし、１とする。このように処理した結果、図９に示す集計結果９３を得ることができる。
【００９１】
実施例１では、分類キーワード候補と比較キーワードとが部分一致である場合にも、出現回数をカウントするが、完全一致する場合にのみ、カウントするようにしてもよい。
【００９２】

このようにして求めた条件キーワード出現回数と主格出現回数とが所定の条件を満たすキーワードを、分類キーワードとして判定する。上記所定の条件は、たとえば、上記式（１）、式（２）に示す条件である。

【００９３】

上記式（１）、式（２）において、α＝０．８、ｔｈ＝１であるとした場合、各キーワードについての式（１）の値は、以下のようになり、「大人だけ」が、分類キーワードとして抽出される。

【００９４】

ｆ（大人だけ）＝０．８＊１＋０．２＊１＝１．０

ｆ（小さな子供）＝０．８＊１＝０．８

ｆ（傘）＝０．２＊１＝０．２

ｆ（大人）＝０．２＊１＝０．２

【００９５】

実施例１では、説明を簡単にするために、３つの回答文書から、分類キーワードを求めているが、本来、大量の回答文書を入力し、分類キーワードを求める。このために、式（２）の閾値ｔｈの値によっては、多くのキーワードが分類キーワードとして判定されることがある。この場合、式（１）の値の上位何件等のように、件数による条件を追加することによって、分類キーワードを判定するようにしてもよい。

【００９６】

分類キーワード判定手段１５は、分類キーワードと、このキーワードを含む回答文章ＩＤと、文番号の情報とを組みにして、テキスト出力手段４０に送る。

【００９７】

テキスト出力手段４０は、分類キーワード判定手段１５が出力した分類キーワードを入力し、分類キーワード毎に、回答文書を分類して出力する。分類キーワードとなるキーワード抽出の処理を、文単位に行うので、出力も文単位で行う例について説明する。

【００９８】

入力として受け取ったキーワードと、このキーワードを含む回答文章ＩＤと、文番号の情報都に基づいて、質問・回答文書蓄積手段３０から、該当する文を検索して取得し、分類キーワード毎に、文の集合を生成する。分類キーワードとこのキーワードを含む文の集合とを一度に出力してもよく、まず、分類キーワードだけを表示し、利用者が選択した分類キーワードを含む文の集合を表示するようにしてもよい。

【００９９】

つまり、回答文書分類装置１００は、質問文書に対する回答文書を、グループを代表するキーワードである分類キーワードを含むグループ毎に分類する回答文書分類装置である。

【０１００】

条件パターン蓄積手段１１は、単語と品詞との組み合わせによって構成されているパターンであって、所定の条件を示すパターンである条件パターンを蓄積している条件パターン蓄積手段の例である。条件パターン判定手段１２は、入力テキスト集合に含まれている文が、上記条件パターン蓄積手段に蓄積されている条件パターンを含んでいるか否かを判定する条件パターン判定手段の例である。条件キーワード抽出手段１４は、上記条件パターンを含んでいれば、上記条件パターンの前方に配置されている単語の配列である条件キーワードを抽出する条件キーワード抽出手段の例である。主格キーワード抽出手段１４は、上記条件パターンを含んでいなければ、主格になる文節に含まれている単語である主格キーワードを抽出する主格キーワード抽出手段の例である。

【０１０１】

また、分類キーワード判定手段１５は、上記条件キーワードおよび上記主格キーワードに含まれるキーワード毎に、着目しているキーワードが上記条件キーワードとして出現する回数である出現回数と、上記主格キーワードとして出現する回数である出現回数とを集計し、上記条件キーワードの出現回数と上記主格キーワードの出現回数とが所定の条件を

満たす場合に、上記着目しているキーワードを分類キーワードであると判定する分類キーワード判定手段の例である。

【実施例２】

【０１０２】

図１１は、本発明の実施例２である回答文書分類装置２００を示すブロック図である。

【０１０３】

回答文書分類装置２００は、条件キーワード抽出手段１６と、主格キーワード抽出手段１７とが抽出した分類キーワード候補を、分類キーワード候補蓄積手段１８に蓄積し、利用者からの検索要求に応じて、検索結果に含まれている回答文章ＩＤを持つ文における分類キーワード候補を、分類キーワード蓄積手段１８から取得し、各分類キーワード候補の条件キーワード出現回数と主格キーワード出現回数とを集計し、分類キーワードを判定する装置である。

【０１０４】

つまり、回答文書分類装置２００は、テキスト入力手段２１とテキスト出力手段４０とに、接続され、条件パターン蓄積手段１１と、条件パターン判定手段１２と、条件キーワード抽出手段１６と、主格キーワード抽出手段１７と、分類キーワード候補蓄積手段１８と、分類キーワード判定手段１９とを有する。なお、実施例１における構成要素と同一の構成要素には、同一符号を付してある。

【０１０５】

テキスト入力手段２１は、質問・回答文書蓄積手段３０に蓄積されている未処理の回答文書を定期的に取得し、条件パターン判定手段１２に送る。

【０１０６】

条件パターン判定手段１２と条件パターン蓄積手段１１とは、実施例１の構成と同じであるので、その説明を省略する。

【０１０７】

条件キーワード抽出手段１６は、条件パターン判定手段１２から条件パターンを含む形態素解析結果の文と、条件パターンに一致する箇所とを受け取り、条件パターンに一致する箇所の前方に存在するキーワードを、条件キーワードとして抽出し、記憶装置に記憶する。

【０１０８】

実施例１において、抽出した条件キーワードに、回答文章ＩＤと文番号とを対応付けて分類キーワード判定手段１５へ渡したが、実施例２では、入力された形態素解析結果に、条件キーワードを示す記号を追加し、分類キーワード侯補蓄積手段１８に記録する。なお、上記のように、入力された形態素解析結果に、条件キーワードを示す記号を追加する理由は、条件キーワードを明示するためである。

【０１０９】

図１２は、実施例２における分類キーワード蓄積手段１８の蓄積例を示す図である。

【０１１０】

条件キーワードを示す記号１２２、１２３は、分類キーワード候補蓄積手段１８に蓄積されている分類キーワード候補が条件キーワードであることを示す記号であり、分類キーワード候補の行末に、記号「Ｘ」が付与されている。記号「Ｘ」が複数行に渡って付与されていれば、単語ではなく文節が抽出されていることを示す。

【０１１１】

主格キーワード抽出手段１７は、条件パターン判定手段１２から、条件パターンを含まないと判定された文を受け取り、係り受け解析を行い、この解析結果から、文の中に主格となる文節が存在するかどうかを調べ、存在すれば、主格となる文節に含まれているキーワード（単語、名詞、名詞句）を、主格キーワードとして抽出し、記憶装置に記憶する。

【０１１２】

実施例１において、抽出した主格キーワードに、回答文章ＩＤと文番号とを対応付け、分類キーワード判定手段１５へ渡したが、実施例２では、入力された形態素解析結果に、

主格キーワードを示す記号を追加し、分類キーワード候補蓄積手段１８に記録する。主格キーワードを示す記号１２１、１２４は、分類キーワード候補蓄積手段１８に蓄積されている分類キーワードが主格キーワードであることを示す記号であり、分類キーワード候補の行末に、記号Ｙが付与されている。

【０１１３】

テキスト入力手段２１において、処理を定期的に実行する度に、条件パターン判定手段１２、条件キーワード抽出手段１６、主格キーワード抽出手段１７が処理し、分類キーワード候補蓄積手段１８に、分類キーワード候補を含む文を蓄積する。

【０１１４】

次に、利用者が、キーワード検索手段５０に検索キーワードを入力すると、キーワード検索手段５０は、質問・回答文書蓄積手段３０を検索し、検索キーワードを含む質問文書について回答している回答文書の集合を取得する。取得した回答文書集合の回答文章ＩＤの集合を、分類キーワード判定手段１９に渡す。

【０１１５】

分類キーワード判定手段１９は、処理対象となるキーワード検索手段５０から、回答文章ＩＤを受け取ると、回答文章ＩＤが一致する文を、分類キーワード候補蓄積手段１８から取得する。各文から、条件キーワードまたは主格キーワードを取得し、「条件キーワード　回答文章ＩＤ　文番号」の組からなる条件キーワード情報９１の集合と、「主格キーワード　回答文章ＩＤ　文番号」の組からなる主格キーワード情報９２の集合とを取得し、図１０に示す分類キーワード判定処理を実行し、分類キーワードを判定する。判定した結果、分類キーワードと、このキーワードを含む回答文章ＩＤと、文番号の情報とを組みにして、テキスト出力手段４０に送る。

【０１１６】

実施例１では、テキスト入力手段２０から、回答文章ＩＤが与えられてから、分類キーワードを出力するまでに、形態素解析処理、係り受け処理、分類キーワード判定処理の時間を加算した処理時間が必要である。これらの処理は、文章数に比例する処理であるので、従来技術で説明したクラスタリング処理（文章数の２乗に比例する）に比較すれば高速である。

【０１１７】

実施例２では、分類キーワード候補蓄積手段１８が、分類キーワード候補を予め抽出して蓄積するので、分類キーワード判定処理のみの時間で、分類キーワードを出力することができ、大変高速である。

【０１１８】

分類キーワード候補蓄積手段１８は、回答文書の文章毎に、条件キーワードであることを示す情報または主格キーワードであることを示す情報が追加されている解析結果と、回答文書を一意に示す回答文章ＩＤとを対応付けた分類キーワード候補を蓄積する分類キーワード候補蓄積手段の例である。

【０１１９】

分類キーワード判定手段１９は、回答文章ＩＤが複数与えられた際に、回答文章ＩＤに該当する回答文書の解析結果の集合を、上記分類キーワード候補蓄積手段から取得し、上記解析結果の集合に含まれている条件キーワードおよび主格キーワードに含まれるキーワード毎に、着目しているキーワードが上記条件キーワードとして出現する回数である出現回数と、上記主格キーワードとして出現する回数である出現回数とを集計し、上記条件キーワードの出現回数と上記主格キーワードの出現回数とが所定の条件を満たす場合に、上記着目しているキーワードを分類キーワードであると判定する分類キーワード判定手段の例である。

【０１２０】

また、上記実施例において、上記条件キーワード抽出手段は、上記条件パターンを含む文節から、条件パターンを除いた語句に、条件パターンを含む文節へ係る０個以上の文節を追加した語句を、条件キーワードとして抽出する手段である。また、上記主格キーワー

ド抽出手段は、主格になる文節に含まれている名詞句に、主格になる文節へ係る０個以上の文節を追加した語句を、上記条件キーワードとして抽出する手段である。そして、上記分類キーワード判定手段は、上記条件キーワードとしての出現回数が多いほど、上記分類キーワードとして判定され易くなる条件を用いて、上記分類キーワードを判定する手段である。

【０１２１】

上記実施例において、手段を工程に置き換えれば、上記実施例を方法の発明として把握することができる。つまり、上記実施例は、質問文書に対する回答文書を、グループを代表するキーワードである分類キーワードを含むグループ毎に分類する回答文書分類方法において、単語と品詞との組み合わせによって構成されているパターンであって、所定の条件を示すパターンである条件パターンが蓄積されている条件パターン蓄積手段に蓄積されている条件パターンを、入力テキスト集合に含まれている文が含んでいるか否かを、条件パターン判定手段が判定し、記憶装置に記憶する条件パターン判定工程と、上記条件パターンを含んでいれば、上記条件パターンの前方に配置されている単語の配列である条件キーワードを、条件キーワード抽出手段が抽出し、記憶装置に記憶する条件キーワード抽出工程と、上記条件パターンを含んでいなければ、主格になる文節に含まれている単語である主格キーワードを、主格キーワード抽出手段が抽出し、記憶装置に記憶する主格キーワード抽出工程と、上記条件キーワードおよび上記主格キーワードに含まれるキーワード毎に、着目しているキーワードが上記条件キーワードとして出現する回数である出現回数と、上記主格キーワードとして出現する回数である出現回数とを集計し、上記条件キーワードの出現回数と上記主格キーワードの出現回数とが所定の条件を満たす場合に、上記着目しているキーワードを分類キーワードであると、分類キーワード判定手段が判定し、記憶装置に記憶する分類キーワード判定工程とを有する回答文書分類方法の例である。

【０１２２】

また、上記実施例は、質問文書に対する回答文書を、グループを代表するキーワードである分類キーワードを含むグループ毎に分類する回答文書分類方法において、単語と品詞との組み合わせによって構成されているパターンであって、所定の条件を示すパターンである条件パターンを蓄積している条件パターン蓄積手段に蓄積されている条件パターンを、回答文書集合に含まれている文が含んでいるか否かを、条件パターン判定手段が判定し、記憶装置に記憶する条件パターン判定工程と、上記条件パターンを含んでいれば、上記条件パターンの前方に配置されている単語の配列である条件キーワードを、条件キーワード抽出手段が抽出し、記憶装置に記憶する条件キーワード抽出工程と、上記条件パターンを含んでいなければ、主格になる文節に含まれている単語である主格キーワードを、主格キーワード抽出手段が抽出し、記憶装置に記憶する主格キーワード抽出工程と、回答文書の文章毎に、条件キーワードであることを示す情報または主格キーワードであることを示す情報が追加されている解析結果と、回答文書を一意に示す回答文章ＩＤとを対応付けた分類キーワード候補とを、分類キーワード候補蓄積手段が蓄積し、記憶装置に記憶する分類キーワード候補蓄積工程と、回答文章ＩＤが複数与えられた際に、回答文章ＩＤに該当する回答文書の上記解析結果の集合を、上記分類キーワード候補蓄積工程から取得し、上記解析結果の集合に含まれている条件キーワード毎にまたは主格キーワード毎に、着目しているキーワードが上記条件キーワードとして出現する回数である出現回数と、上記主格キーワードとして出現する回数である出現回数とを集計し、上記条件キーワードの出現回数と上記主格キーワードの出現回数とが所定の条件を満たす場合に、上記着目しているキーワードを分類キーワードであると、分類キーワード判定が判定し、記憶装置に記憶する分類キーワード判定工程とを有する回答文書分類方法の例である。

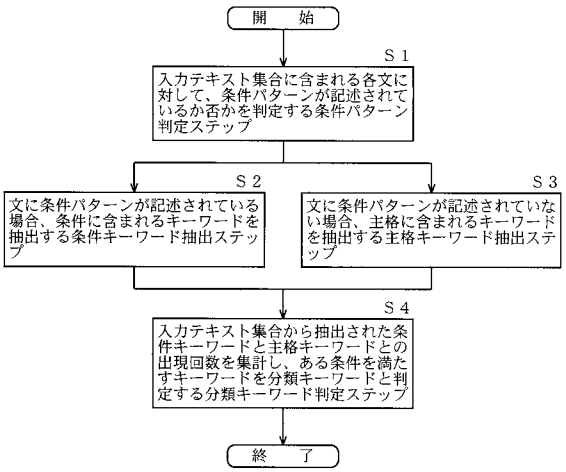
【０１２３】

さらに、上記実施例をプログラムの発明として把握することができる。つまり、上記実施例は、請求項１～請求項５のいずれか１項に記載の回答文書分類装置を構成する各手段としてコンピュータを機能させるプログラムの例である。

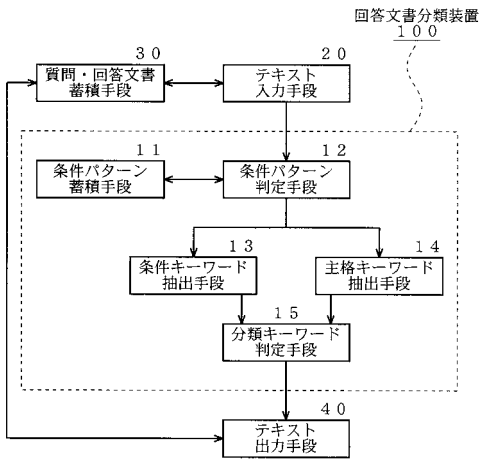
【符号の説明】

【０１２４】
１００…回答文書分類装置、
１１…条件パターン蓄積手段、
１２…条件パターン判定手段、
１３…条件キーワード抽出手段、
１４…主格キーワード抽出手段、
１５…分類キーワード判定手段、
２０…テキスト入力手段、
３０…質問・回答文書蓄積手段、
４０…テキスト出力手段、
２００…回答文書分類装置、
１６…条件キーワード抽出手段、
１７…主格キーワード抽出手段、
１８…分類キーワード候補蓄積手段、
１９…分類キーワード判定手段、
５０…キーワード検索手段。

【図１】



【図２】



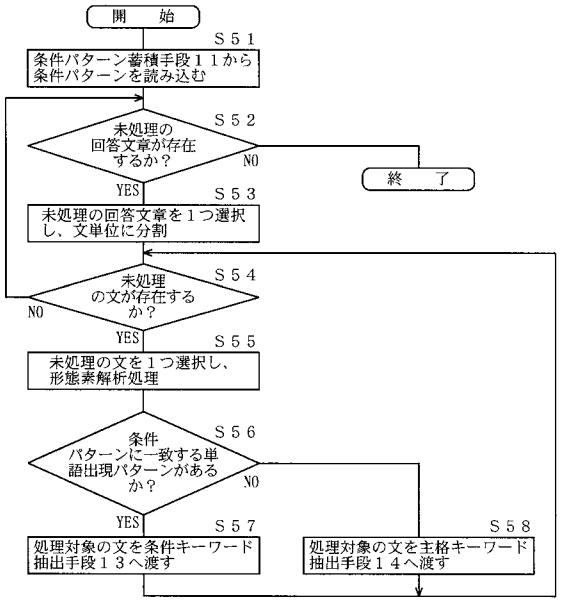
【図３】

条件パターン蓄積手段１１が蓄積している条件パターンの例

接続接尾辞：ナラ 判定詞：ナラ 判定詞：ナラバ 判定詞：デ、動詞語幹：ア、動詞活用語尾：レ、動詞接尾辞：バ 格助詞：ノ、名詞：バアイ 判定詞：デス、引用助詞：ト

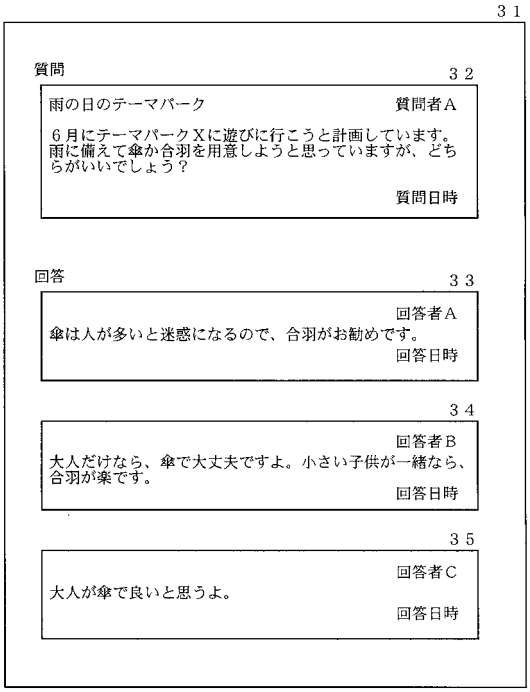
【図４】

条件パターン判定手段１２の動作



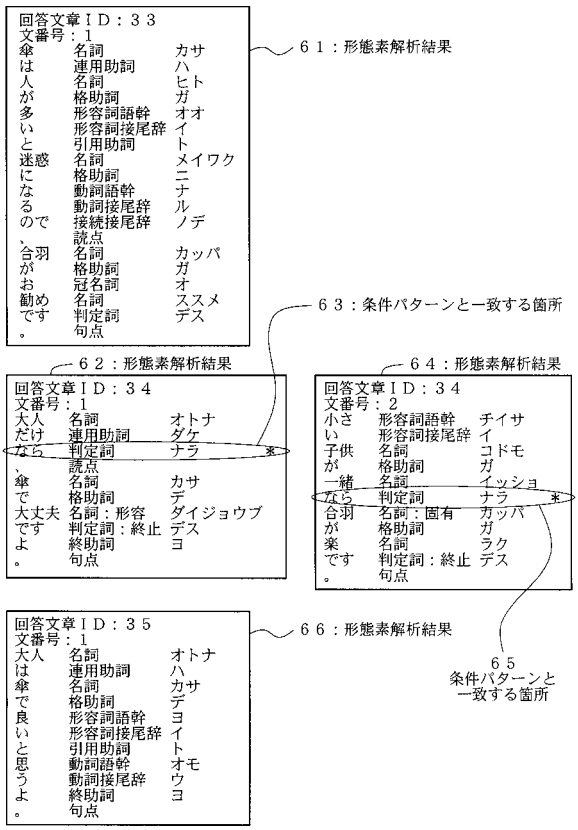
【図５】

Ｑ＆Ａコミュニティサイトの構成



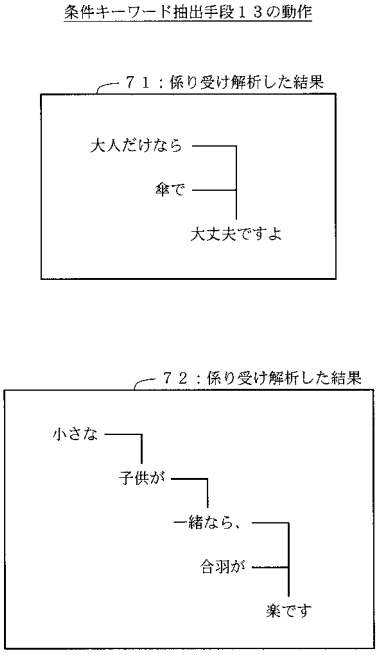
【図６】

条件パターン判定手段１２が判定した形態素解析結果の例



【図７】

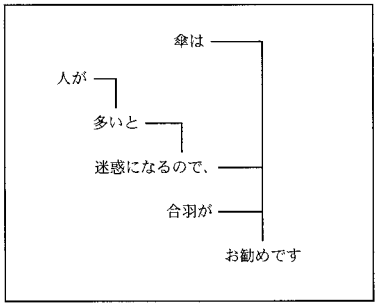
条件キーワード抽出手段１３の動作



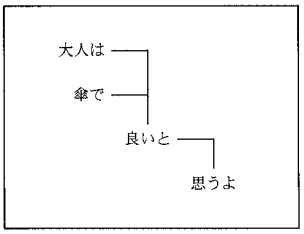
【図８】

主格キーワード抽出手段１４の動作

(1)



(2)



【図９】

分類キーワード判定手段１５の動作説明

９１：条件キーワード情報

大人だけ	３４	１
小さな子供	３４	２

９２：主格キーワード情報

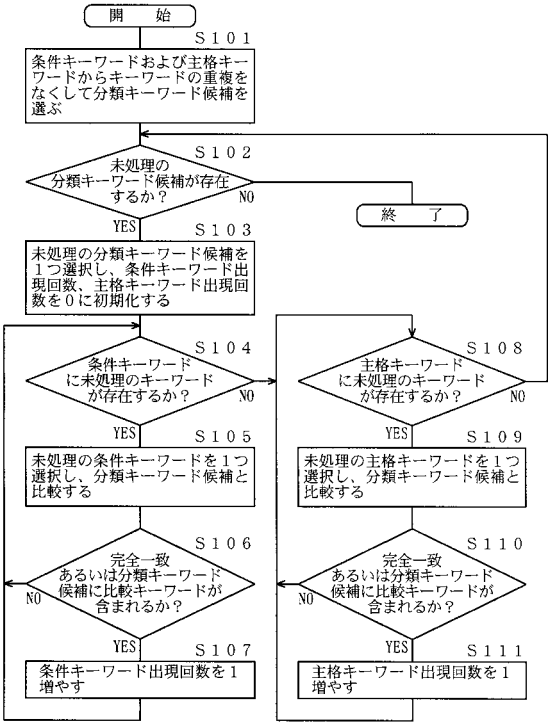
傘	３３	１
大人	３５	１

集計結果
９３

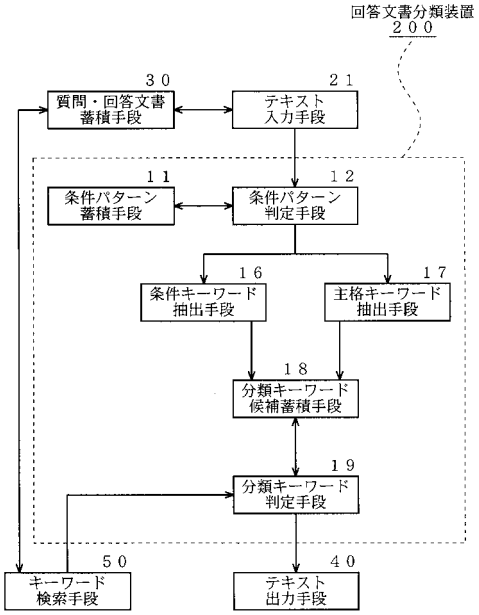
分類キーワード候補	条件キーワード出現回数	主格キーワード出現回数	条件キーワード	主格キーワード
大人だけ	1	1	大人だけ　３４　１	大人　３５　１
小さな子供	1	0	小さな子供　３４　２	—
傘	0	1	—	傘　３３　１
大人	0	1	—	大人　３５　１

【図１０】

分類キーワード判定手段１５の動作

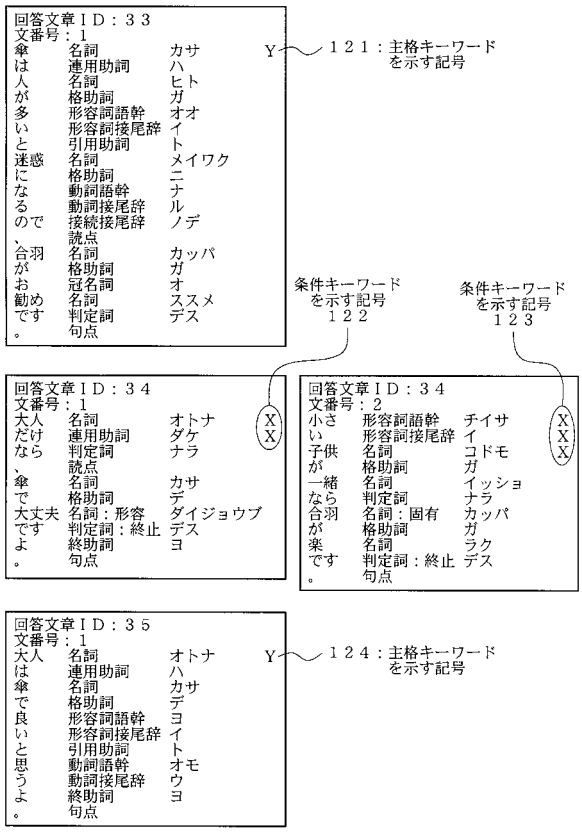


【図１１】



【図１２】

分類キーワード候補蓄積手段１８における蓄積例



フロントページの続き

(56)参考文献 特開平０５－１２０３４５（ＪＰ，Ａ）
特開平０１－１２１９２８（ＪＰ，Ａ）
特開平１１－２０３３１８（ＪＰ，Ａ）
特開２０００－２５９６６６（ＪＰ，Ａ）
別所 克人 他，自然言語検索システムにおける分野推論方式，電子情報通信学会論文誌，日本，電子情報通信学会，１９９８年 6月２５日，Vol.J81-D-II,No.6，PP.1317-1327.
佐々木 裕 他，ＳＶＭを用いた学習型質問応答システムＳＡＩＱＡ－ＩＩ，情報処理学会論文誌，日本，情報処理学会，２００４年 2月１５日，Vol.45,No.2，PP.635-646.

(58)調査した分野(Int.Cl.，ＤＢ名)
Ｇ０６Ｆ １７／３０
ＪＳＴＰｌｕｓ（ＪＤｒｅａｍＩＩ）