

(19) 日本国特許庁(JP)

(12) 特 許 公 報(B2)

(11) 特許番号

特許第5259462号
(P5259462)

(45) 発行日 平成25年8月7日(2013.8.7)

(24) 登録日 平成25年5月2日(2013.5.2)

(51) Int.Cl.

F I

G 0 6 F 17/30 (2006.01)

G 0 6 F 17/30 3 3 0 C

G 0 6 F 17/30 4 1 9 A

G 0 6 F 17/30 1 7 0 A

G 0 6 F 17/30 3 5 0 C

請求項の数 6 (全 36 頁)

(21) 出願番号 特願2009-59013 (P2009-59013)
 (22) 出願日 平成21年3月12日 (2009.3.12)
 (65) 公開番号 特開2010-211664 (P2010-211664A)
 (43) 公開日 平成22年9月24日 (2010.9.24)
 審査請求日 平成23年9月16日 (2011.9.16)

(73) 特許権者 000003078
 株式会社東芝
 東京都港区芝浦一丁目1番1号
 (74) 代理人 100089118
 弁理士 酒井 宏明
 (74) 代理人 100112656
 弁理士 宮田 英毅
 (72) 発明者 浜田 伸一郎
 東京都港区芝浦一丁目1番1号 株式会社
 東芝内
 審査官 伊知地 和之

最終頁に続く

(54) 【発明の名称】 検索を支援する装置、方法およびプログラム

(57) 【特許請求の範囲】

【請求項 1】

入力文書と、前記入力文書に含まれる文字のうち指定された文字の位置との入力を受け付ける受付部と、

前記入力文書の複数のパッセージ間の階層構造を解析する構造解析部と、

前記指定された文字を含む前記パッセージより上位の前記パッセージである上位パッセージに含まれる単語それぞれについて、文書記憶部に記憶された、前記入力文書と関連する関連文書を含む複数の文書内での前記単語の重要性の度合いを表す第1単語重要度を算出する第1単語重要度算出部と、

前記上位パッセージに含まれる単語それぞれについて、前記入力文書内での前記単語の重要性の度合いを表し、より上位の前記上位パッセージに含まれる単語ほど小さい値となる第2単語重要度を算出する第2単語重要度算出部と、

前記関連文書を前記文書記憶部から検索するための条件であって、前記上位パッセージに含まれる単語それぞれについて算出された前記第1単語重要度と前記第2単語重要度とを含む検索条件を出力する出力部と、

を備えることを特徴とする検索支援装置。

【請求項 2】

前記検索条件に基づいて前記関連文書を前記文書記憶部から検索し、前記関連文書が前記入力文書と関連する度合いを表す関連度を算出する文書検索部と、

前記関連文書を複数の前記パッセージに分割する分割部と、

10

20

分割された前記パッセージそれぞれについて、前記関連文書内での前記パッセージの重要性の度合いを表す第 1 パッセージ重要度を算出する第 1 パッセージ重要度算出部と、

分割された前記パッセージそれぞれについて、前記入力文書内での前記パッセージの重要性の度合いを表す第 2 パッセージ重要度を算出する第 2 パッセージ重要度算出部と、

分割された前記パッセージそれぞれについて、前記パッセージが含まれる前記関連文書の前記関連度と、前記第 1 パッセージ重要度と、前記第 2 パッセージ重要度と、を重み付け加算した総合重要度を算出する総合重要度算出部と、

分割された前記パッセージのうち、前記総合重要度が大きい前記パッセージを優先して選択するパッセージ選択部と、をさらに備えたこと、

を特徴とする請求項 1 に記載の検索支援装置。

10

【請求項 3】

前記指定された文字の品詞に応じて予め定められた検索キーワードを生成する生成部をさらに備え、

前記出力部は、さらに、生成された前記検索キーワードを含む前記検索条件を出力すること、

を特徴とする請求項 1 に記載の検索支援装置。

【請求項 4】

前記第 2 単語重要度算出部は、ユーザが編集した単語ほど、ユーザが編集していない単語より大きい前記第 2 単語重要度を算出すること、

を特徴とする請求項 1 に記載の検索支援装置。

20

【請求項 5】

受付部が、入力文書と、前記入力文書に含まれる文字のうち指定された文字の位置との入力を受付ける受付ステップと、

構造解析部が、前記入力文書の複数のパッセージ間の階層構造を解析する構造解析ステップと、

第 1 単語重要度算出部が、前記指定された文字を含む前記パッセージより上位の前記パッセージである上位パッセージに含まれる単語それぞれについて、文書記憶部に記憶された、前記入力文書と関連する関連文書を含む複数の文書内での前記単語の重要性の度合いを表す第 1 単語重要度を算出する第 1 単語重要度算出ステップと、

第 2 単語重要度算出部が、前記上位パッセージに含まれる単語それぞれについて、前記入力文書内での前記単語の重要性の度合いを表し、より上位の前記上位パッセージに含まれる単語ほど小さい値となる第 2 単語重要度を算出する第 2 単語重要度算出ステップと、

30

出力部が、前記関連文書を前記文書記憶部から検索するための条件であって、前記上位パッセージに含まれる単語それぞれについて算出された前記第 1 単語重要度と前記第 2 単語重要度とを含む検索条件を出力する出力ステップと、

を備えることを特徴とする検索支援方法。

【請求項 6】

コンピュータを、

入力文書と、前記入力文書に含まれる文字のうち指定された文字の位置との入力を受付ける受付部と、

40

前記入力文書の複数のパッセージ間の階層構造を解析する構造解析部と、

前記指定された文字を含む前記パッセージより上位の前記パッセージである上位パッセージに含まれる単語それぞれについて、文書記憶部に記憶された、前記入力文書と関連する関連文書を含む複数の文書内での前記単語の重要性の度合いを表す第 1 単語重要度を算出する第 1 単語重要度算出部と、

前記上位パッセージに含まれる単語それぞれについて、前記入力文書内での前記単語の重要性の度合いを表し、より上位の前記上位パッセージに含まれる単語ほど小さい値となる第 2 単語重要度を算出する第 2 単語重要度算出部と、

前記関連文書を前記文書記憶部から検索するための条件であって、前記上位パッセージに含まれる単語それぞれについて算出された前記第 1 単語重要度と前記第 2 単語重要度と

50

を含む検索条件を出力する出力部と、
として機能させるための検索支援プログラム。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、検索を支援する装置、方法およびプログラムに関する。

【背景技術】

【0002】

様々な文書に含まれる情報を統合することにより、知りたい情報に関する知識を得られるようにするタスクに対するニーズの高まりが指摘されている。このようなタスクは、探
求検索 (Exploratory Search) タスクとして知られている。

10

【0003】

例えば、Web からサブプライム問題の概要に関する情報を得る場合には、ユーザは、サブプライム問題に関連する検索キーワードを入力してWeb を検索する。しかし、Web 上に概要がまとめられている文書が存在するとは限らない。また、存在するとしても、ユーザが望む詳細度で記載された文書であるとも限らない。すなわち、ユーザの情報要求が1つの文書で満たされることは、一般的に保証されていない。このような問題に対しては、ユーザは多数の文書を収集し、文書間の情報差分等をもとに知識統合する対話的作業が必要になる。この作業はベリー摘みモデルと呼ばれる。

【0004】

20

探求検索タスクを実現するためにキーワード検索システムを用いる場合、閲覧した文書群では不足する情報に対応するクエリをユーザが適切に作成する必要がある。しかし、キーワードを並べて作成されるクエリでは、検索要求を適切に表現することができず、所望の情報を得られない場合がある。このような問題を解決する技術として、ユーザが作成したクエリを用いずに現在閲覧している文書に関連する文書を返す連想検索システムが知られている。

【0005】

例えば、特許文献1では、ユーザが入力したキーワードやテキストなどの問い合わせ用の情報源に応じてサーバで生成されたクエリーベクトルを用いて関連情報を検索する情報
検索システムに関する技術が提案されている。

30

【先行技術文献】

【特許文献】

【0006】

【特許文献1】特開2003-16089号公報

【発明の概要】

【発明が解決しようとする課題】

【0007】

しかしながら、上記のような連想検索システムでは、システムにより生成される検索要求が本来の要求から逸れて探求の方向性を見失うという問題 (徘徊問題) が指摘されている。

40

【0008】

本発明は、上記に鑑みてなされたものであって、連想検索システム等で生じうる徘徊問題の発生を低減できる装置、方法およびプログラムを提供することを目的とする。

【課題を解決するための手段】

【0009】

上述した課題を解決し、目的を達成するために、本発明は、入力文書と、前記入力文書に含まれる文字のうち指定された文字の位置との入力を受け付ける受付部と、前記入力文書の構成単位である複数のパッセージ間の階層構造を解析する構造解析部と、前記位置の文字を含む前記パッセージより上位の前記パッセージである上位パッセージに含まれる単語それぞれについて、文書記憶部に記憶された複数の文書内での前記単語の重要性の度合い

50

を表す第1単語重要度を算出する第1単語重要度算出部と、前記上位パッセージに含まれる単語それぞれについて、前記入力文書内での前記単語の重要性の度合いを表し、より上位の前記上位パッセージに含まれる単語ほど小さい値となる第2単語重要度を算出する第2単語重要度算出部と、前記入力文書と関連する関連文書を前記文書記憶部から検索するための条件であって、前記上位パッセージに含まれる単語それぞれについて算出された前記第1単語重要度と前記第2単語重要度とを含む検索条件を出力する出力部と、を備えることを特徴とする。

【0010】

また、本発明は、上記装置で実行することができる方法およびプログラムである。

【発明の効果】

10

【0011】

本発明によれば、連想検索システム等で生じうる徘徊問題の発生を低減できるという効果を奏する。

【図面の簡単な説明】

【0012】

【図1】図1は、第1の実施の形態の検索支援装置で表示される編集画面の一例を示す図である。

【図2】図2は、第1の実施の形態にかかる検索支援装置の構成の一例を示すブロック図である。

【図3】図3は、第1の実施の形態における検索支援処理の全体の流れを示すフローチャートである。

20

【図4】図4は、第1の実施の形態における構造解析処理の全体の流れを示すフローチャートである。

【図5】図5は、第1の実施の形態における一般重要度算出処理の全体の流れを示すフローチャートである。

【図6】図6は、第1の実施の形態における個別重要度算出処理の全体の流れを示すフローチャートである。

【図7】図7は、第1の実施の形態における既知度算出処理の全体の流れを示すフローチャートである。

【図8】図8は、第1の実施の形態における文書検索処理の全体の流れを示すフローチャートである。

30

【図9】図9は、第1の実施の形態におけるパッセージ分割処理の全体の流れを示すフローチャートである。

【図10】図10は、第1の実施の形態における文書内重要度算出処理の全体の流れを示すフローチャートである。

【図11】図11は、第1の実施の形態におけるサマリ内重要度算出処理の全体の流れを示すフローチャートである。

【図12】図12は、第1の実施の形態におけるパッセージ選択処理の全体の流れを示すフローチャートである。

【図13】図13は、第1の実施の形態における多段階パッセージ検索処理の全体の流れを示すフローチャートである。

40

【図14】図14は、第1の実施の形態の変形例の検索支援装置で表示される編集画面の一例を示す図である。

【図15】図15は、第1の実施の形態の変形例における構造解析処理の全体の流れを示すフローチャートである。

【図16】図16は、第2の実施の形態の検索支援装置で表示される編集画面の一例を示す図である。

【図17】図17は、第2の実施の形態にかかる検索支援装置の構成の一例を示すブロック図である。

【図18】図18は、第3の実施の形態の検索支援装置で表示される編集画面の一例を示

50

す図である。

【図 19】図 19 は、第 3 の実施の形態にかかる検索支援装置の構成の一例を示すブロック図である。

【図 20】図 20 は、第 3 の実施の形態における検索支援処理の全体の流れを示すフローチャートである。

【図 21】図 21 は、第 3 の実施の形態における質問生成処理の全体の流れを示すフローチャートである。

【図 22】図 22 は、第 3 の実施の形態における文書検索処理の全体の流れを示すフローチャートである。

【図 23】図 23 は、第 3 の実施の形態におけるサマリ内重要度算出処理の全体の流れを示すフローチャートである。

10

【図 24】図 24 は、第 4 の実施の形態にかかる検索支援装置の構成の一例を示すブロック図である。

【図 25】図 25 は、第 4 の実施の形態における個別重要度算出処理の全体の流れを示すフローチャートである。

【図 26】図 26 は、第 4 の実施の形態におけるサマリ内重要度算出処理の全体の流れを示すフローチャートである。

【図 27】図 27 は、第 1 ～ 第 4 の実施の形態にかかる検索支援装置のハードウェア構成を示す説明図である。

【発明を実施するための形態】

20

【0013】

以下に添付図面を参照して、この発明にかかる検索支援装置、検索支援方法、および検索支援プログラムの最良な実施の形態を詳細に説明する。

【0014】

(第 1 の実施の形態)

上述のように、連想検索システムでは、徘徊問題が発生しうる。また、連想検索システムを含むいずれの検索システムも検索条件を満たす文書を返すことができるが、それだけではユーザの要求を満たすには不十分である。すなわち、探求検索タスクでは、文書の中に記載されている、ユーザにとっての新規情報が重要であって、単に条件を満たす文書が得られればよいものではない。しかし、従来のシステムでは、返された各文書の中からユーザが新規情報を探す作業が強いられていた。

30

【0015】

一方、複数文書の新規情報を統合する技術として複数文書要約システムが知られている。しかし要約の観点は十人十色であり、このようなシステムがユーザの望む要約を生成できるとは限らない。これに対し、事実優先・意見優先など要求タイプを事前入力するシステムが存在する。しかし要求タイプの種類が少ないため、情報要求としては十分ではない。また現在の技術的限界として、単数文書要約における文選定は実用レベルであるが、異なる文書の文脈統合は実用レベルでない。

【0016】

一方、探求検索タスクでは、その情報収集プロセスの複雑さゆえ、調査途中結果をメモなどにまとめることが多い(以下、このようにしてまとめた文書をユーザサマリという)。第 1 の実施の形態にかかる検索支援装置は、この点に着眼し、サマリ作成作業を含む探求検索タスクを支援する装置であって、その一部機能として、サマリ文脈に基づいて必要な情報を検索・提示する装置を実現する。これにより、上記のような各問題を包括的に解決しうる装置を実現できる。なお、このような装置は一例であって、上記徘徊問題を解消するためには、例えばサマリ作成機能を備える必要はない。

40

【0017】

ここで、本実施の形態の検索支援装置の具体的な機能構成の概要について説明する。検索支援装置は、コピーペースト機能および構造テキスト編集機能を有する。構造テキスト編集機能とは、箇条書きなどにより構造化されたテキスト(以下、構造テキストという)

50

を編集する機能をいう。

【 0 0 1 8 】

そして、検索支援装置は、テキスト編集に任意のカーソル位置で特定キーが押下されたときに情報提示機能と呼び出す。情報提示機能は、後述する文書検索処理および重要文選定処理により獲得したパッセージ群の候補を提示する。パッセージとは、文、章、節、および段落などの意味単位で文書を分割した文書の構成単位をいう。ユーザが候補の1つを選択すると、カーソル位置に当パッセージが挿入される。

【 0 0 1 9 】

文書検索処理では、検索支援装置は、サマリ内のカーソルより構造上位のパッセージ群から内容語を抽出し、 $t f - i d f$ (Term Frequency - Inverse Document Frequency) などによる一般的な重要度をベースにカーソルにより近い位置に存在する文に含まれる内容語に傾斜加重した重みつき要求ベクトル(文書検索用ベクトル)を生成する。そして、検索支援装置は、この文書検索用ベクトルを元に関連文書を検索する。

【 0 0 2 0 】

重要文選定処理では、検索支援装置は、検索した各文書から、タイトル、位置情報、出現頻度、語彙結束性、カバレッジ、および文構造制約を素性としたエドムンソン(H.P. Edmundson. New methods in automatic abstracting. Journal of ACM, Vol.16, No.2, pp264-285, 1969)の手法などにより、重要文(重要パッセージ)を選定する。ただし、語彙結束性評価への入力の前段で求めた文書検索用ベクトルとし、カバレッジ性評価への入力は、ユーザサマリのカーソルと構造同位のパッセージ群(後述)とする点が従来と異なる。これにより、従来と同様の単文書内での重要パッセージ判定結果に、ユーザサマリに対する情報の親和性・新規性が加味される。以上によって得た文スコアを所属文書のスコアと合算し、スコア上位のパッセージを返却する。

【 0 0 2 1 】

次に、画面イメージを参照しながら(図1)、本実施の形態の検索支援装置の機能の概要について説明する。図1は、第1の実施の形態の検索支援装置で表示される編集画面の一例を示す図である。

【 0 0 2 2 】

同図は、サマリを編集するためのサマリ編集ウィンドウ10と、サマリ編集ウィンドウ10内で指定されたカーソル位置に応じて検索されたパッセージの候補を表示するパッセージ候補ウィンドウ11とを含む編集画面の例が示されている。

【 0 0 2 3 】

ユーザは、サマリ編集ウィンドウ10内でキーボード・マウスなどを用いることにより構造テキストを作成および編集することができる。このテキストが上述のユーザサマリに相当する。ユーザサマリはユーザによる調査結果をまとめた文書となる。

【 0 0 2 4 】

サマリ編集ウィンドウ10は、テキストを編集できるほか、少なくとも箇条書きなどパッセージ同士の構造的関係を表す表現を入力することができる。好ましくは、太字・色変更などの文字修飾、および図表表現などの、例えばHTML(HyperText Markup Language)が有するプレゼンテーション表現機能をサポートしてもよい。その場合、ユーザは調査結果をよりわかりやすくまとめることができよう。

【 0 0 2 5 】

また好ましくは、外部の文書からのコピーペースト機能をサポートとしてもよい。その場合、ユーザは当該ツール以外のツールを用いて調査した結果を当該サマリに反映することもできる。すなわちユーザは、様々な検索ツールを併用して調査を進めることができるようになる。

【 0 0 2 6 】

ユーザは、パッセージ候補ウィンドウ11によって、文献DB(文書記憶部(後述))に記憶されている複数の文書内のパッセージ群を検索し取り込むことができる。例えば、ユーザがサマリ編集ウィンドウ10上で空パッセージ(例えば箇条書き形式の表現ならば

10

20

30

40

50

、中点だけが存在する行)にカーソルを置いて特定キーを押下するとパッセージ候補ウィンドウ 11 が表示される。なお、この操作を以降の説明の簡便化のため「パッセージ検索呼び出し操作」と呼ぶことにする。パッセージ候補ウィンドウ 11 には、ユーザサマリの文脈に沿って文書記憶部の文書群から検索されたパッセージ群の候補が表示される。ユーザが候補の 1 つを選択すると、サマリ編集ウィンドウ 10 のカーソル位置に選択されたパッセージが挿入される。

【0027】

ユーザは、サマリ編集ウィンドウ 10 とパッセージ候補ウィンドウ 11 とを用いた以上の作業を繰り返すことで、探求検索タスクを進めることができるようになっている。パッセージ候補ウィンドウ 11 によって挿入されたパッセージ群も、一般的なテキストとして

10

【0028】

なお、同図は、以下のようなユーザ操作により作成されたコンテンツ(文書)の例を示している。

(1)「株価」を手入力する。

(2)「2008/8/28」を手入力する。

(3)その構造下位で「パッセージ検索呼び出し操作」を行い、表示された候補の 1 つを挿入する。この作業を反復する。

(4)ただし(3)の作業で、興味のない内容を一部削る、パッセージ構造を変更するなど、調査目的や興味に基づいて適宜修正を加えるものとする。

20

(5)また(3)の作業で、所望のパッセージ候補が表示されないときは、構造下位で自分の興味ある話題語を記載してから「パッセージ検索呼び出し操作」を行う。

【0029】

また、本実施の形態では、オプションな機能として多段階パッセージ検索機能を利用することができる。「多段階パッセージ検索機能」とは、パッセージ候補ウィンドウ 11 内で、さらに関連するパッセージ群を検索して別のウィンドウに表示する機能をいう。例えば、パッセージ候補ウィンドウ 11 に表示された候補パッセージ群の特定の候補パッセージ上で特定キーが押下されたときに、その候補パッセージをサマリに挿入したと仮定した上で、その候補パッセージに関連するパッセージ群を検索して、さらなるポップアップウィンドウに検索されたパッセージ候補を提示するように構成することができる。このポップアップウィンドウ上でもまた同様の操作を行うことができる。すなわち、再帰的な操作が可能である。

30

【0030】

また、2 種類の特定キーを用意し、特定キーの一方が押下されたときに、表示済み候補パッセージの子供となる候補パッセージを検索し、他方が押下されたときに、表示済み候補パッセージの兄弟となる候補パッセージを検索するように構成してもよい。子孫のポップアップウィンドウ上で候補の 1 つを選択すると、それまでに選択した祖先のパッセージ群すべてが組み立てられた状態で一括してサマリ編集ウィンドウ 10 内のユーザサマリ本文へ挿入される。

【0031】

40

次に、本実施の形態の検索支援装置の構成の詳細について図 2 を用いて説明する。図 2 は、第 1 の実施の形態にかかる検索支援装置 100 の構成の一例を示すブロック図である。同図に示すように、検索支援装置 100 は、文脈解析部 110 と、パッセージ検索部 120 と、編集部 130 と、文書記憶部 140 と、を備えている。

【0032】

なお、パッセージ検索(取込含む)機能は、編集部 130、文脈解析部 110、およびパッセージ検索部 120 の循環的なデータのやり取りとして実現される。文書記憶部 140 は、検索対象となる複数の文書を記憶する記憶部である。

【0033】

編集部 130 は、上述のような各種ユーザ操作を受け、受け付けたユーザ操作に応じた

50

処理を各構成部に指示する。例えば、編集部 130 は、「パッセージ検索呼び出し操作」を受理すると、内部に保存しているユーザサマリおよびカーソル位置を文脈解析部 110 へ渡す。また、編集部 130 は、表示装置（図示せず）を備え、表示装置に表示された画面上でテキストを編集する方法などの、従来から用いられているあらゆる方法により、ユーザサマリを編集する機能を備えている。

【0034】

文脈解析部 110 は、受理したユーザサマリおよびカーソル位置をユーザの情報要求として解析し、関連文書の検索に用いる形式（ベクトル形式）で表した情報要求、および、検索結果の重要度を判定するための情報を出力する。具体的には、文脈解析部 110 は、単語一般重要度ベクトル A および単語個別重要度ベクトル B を、情報要求としてパッセージ検索部 120 に出力する。また、文脈解析部 110 は、単語既知度ベクトル C を、重要度を判定するための情報としてパッセージ検索部 120 に出力する。

10

【0035】

本実施の形態では、文脈解析部 110 が、ユーザサマリ内の単語の重要度やパッセージ間の階層構造に応じた情報要求を作成するため、ユーザの要求をより適切に反映した情報要求を作成することができる。すなわち、本実施の形態によれば、重みが徐々に小さくなるが、過去に編集した上位階層のパッセージ内の単語も常に情報要求に含めるため、情報要求の一貫性を保つことができる。これにより、システムにより生成される検索要求がユーザが望む本来の要求から逸れて探求の方向性を見失う徘徊問題を解消することが可能となる。なお、上記各ベクトル A～C の作成方法の詳細は後述する。

20

【0036】

パッセージ検索部 120 は、受理した情報要求表現であるベクトル A～C を元に、文書記憶部 140 内に格納されている文書群から、情報要求を満たす適切なパッセージ群を順位つきで取得し、パッセージ候補として編集部 130 に出力する。なお、編集部 130 は、このパッセージ候補を上述のパッセージ候補ウィンドウ 11 に表示する。

【0037】

次に、文脈解析部 110 の機能構成の詳細について説明する。同図に示すように、文脈解析部 110 は、受付部 111 と、構造解析部 112 と、第 1 単語重要度算出部 113 と、第 2 単語重要度算出部 114 と、既知度算出部 115 と、出力部 116 と、を備えている。

30

【0038】

受付部 111 は、構造解析の対象となる入力文書であるユーザサマリと、ユーザサマリ内で指定された文字の位置（カーソル位置）との入力を編集部 130 から受付ける。なお、カーソル位置は、ユーザサマリ内のパッセージのいずれか 1 つを特定できればよい。例えば、ユーザサマリ内で指定された 1 文字以上の文字（文字列）の位置をカーソル位置として受付けるように構成してもよい。

【0039】

構造解析部 112 は、入力されたユーザサマリをパッセージに分割し、各パッセージ間の階層構造を解析し、パッセージ間の親子関係および兄弟関係などの構造関係が付与されたパッセージ群を生成する。

40

【0040】

第 1 単語重要度算出部 113 は、生成された構造関係付きパッセージ群を元に、文書群全体の単語分布から単語一般重要度ベクトル A を算出する。具体的には、第 1 単語重要度算出部 113 は、カーソル位置の文字が含まれるパッセージより階層構造が上位のパッセージ群（以下、構造上位パッセージという）から内容語の一般重要度（第 1 単語重要度）を算出し、算出した一般重要度を重みとする重み付き単語ベクトルである単語一般重要度ベクトル A を作成する。

【0041】

一般重要度は、文書記憶部 140 に記憶された複数の文書内での単語の重要性の度合いを表す。すなわち、一般重要度は、ユーザサマリの文脈と関係なく、文書記憶部 140 内

50

に含まれる単語の単語分布を元にした「人間一般として何を知りたいか」に関する手がかりである。また、内容語とは、検索語として有用と思われる単語を意味する。1つの例としては、日本語ならば「は」「を」「が」など助詞を除いたすべての形態素を内容語とする方法がある。また、例えば経済専用の検索ツールとして構成するのであれば、予め用意してある経済用語集に含まれる語を内容語とする方法もある。

【0042】

第2単語重要度算出部114は、生成された構造関係付きパッセージ群を元に、構造上位パッセージから単語個別重要度ベクトルBを算出する。具体的には、第2単語重要度算出部114は、構造上位パッセージ群に含まれる内容語について、ユーザ個人にとっての重要度を表す個別重要度（第2単語重要度）を算出し、算出した個別重要度を重みとする重み付き単語ベクトルである単語個別重要度ベクトルBを作成する。

10

【0043】

個別重要度は、ユーザサマリ内での単語の重要性の度合いを表す。すなわち、個別重要度は、ユーザサマリのカーソル位置での文脈をバイアスとして用いた、「次にユーザは何を知りたいか」（ユーザの注目度）に関する手がかりである。なお、個別重要度は、より上位のパッセージほど小さい値となるように算出される（後述）。

【0044】

既知度算出部115は、生成された構造関係付きパッセージ群を元に、カーソル位置の文字が含まれるパッセージと階層構造上の階層位置が同位のパッセージ（以下、構造同位パッセージという）から、単語既知度ベクトルCを算出する。具体的には、既知度算出部115は、構造同位パッセージ群に含まれる内容語について、ユーザサマリ内でユーザが既に知っている度合いを表す既知度を算出し、算出した既知度を重みとする重み付き単語ベクトルである単語既知度ベクトルCを作成する。既知度は、ユーザサマリのカーソル位置での文脈をバイアスとして用いた、「次にユーザは何を知りたくないか」に関する手がかりである。

20

【0045】

出力部116は、作成された単語一般重要度ベクトルA、単語個別重要度ベクトルB、および単語既知度ベクトルCをパッセージ検索部120に出力する。後述するように、単語一般重要度ベクトルAおよび単語個別重要度ベクトルBは、文書記憶部140に記憶された文書を検索するための検索条件として利用される。また、単語既知度ベクトルCは、検索した文書内の各パッセージの重要度（サマリ内重要度）を算出するための情報として利用される。

30

【0046】

次に、パッセージ検索部120の機能構成の詳細について説明する。同図に示すように、パッセージ検索部120は、文書検索部121と、パッセージ分割部122と、第1パッセージ重要度算出部123と、第2パッセージ重要度算出部124と、総合重要度算出部125と、パッセージ選択部126と、を備えている。

【0047】

文書検索部121は、第1単語重要度算出部113によって生成された単語一般重要度ベクトルAと、第2単語重要度算出部114によって生成された単語個別重要度ベクトルBと、を用いて、ユーザサマリに関連する文書を文書記憶部140から検索し、検索した文書ごとにユーザサマリに関連する度合いを表す関連度を算出する。

40

【0048】

より具体的には、文書検索部121は、単語一般重要度ベクトルAと単語個別重要度ベクトルBとから検索用のベクトルである文書検索用ベクトル（詳細は後述）を生成し、文書記憶部140内の文書の特徴を表す文書ベクトルと文書検索用ベクトルとの間のコサイン尺度を関連度として算出する。なお、文書の文書ベクトルとは、例えば単語ごとの出現頻度またはtf-idfなどをベクトル化したものである。

【0049】

パッセージ分割部122は、検索された各文書をパッセージ単位に分割する。パッセー

50

ジ分割部 1 2 2 は、さらに、分割したパッセージの照応解決処理を実行する。すなわち、パッセージ分割部 1 2 2 は、分割したパッセージに、他のパッセージに含まれる対象を指示す照応表現が含まれる場合に、当該他のパッセージから照応表現の指示対象を表す名詞句を取得し、名詞句で照応表現を置換する。

【 0 0 5 0 】

第 1 パッセージ重要度算出部 1 2 3 は、分割された各パッセージについて、パッセージが所属する文書の中での当該パッセージの重要性の度合いを表すスコアである文書内重要度（第 1 パッセージ重要度）を算出する。文書内重要度は、ユーザにとって重要かという観点に含まれていない。文書内重要度の算出方法としては、例えば上述のエドムンソンで用いられているような、単文書要約の文選定処理の 1 つとして知られる手法を用いることができる。

10

【 0 0 5 1 】

具体的には、第 1 パッセージ重要度算出部 1 2 3 は、パッセージ i の文書内重要度 $P S c o r e _ i$ を、以下の (1) 式により算出する。

$$P S c o r e _ i = W 1 1 \times C 1 _ i + W 1 2 \times C 2 _ i + W 1 3 \times C 3 _ i + W 1 4 \times C 4 _ i \quad \cdots (1)$$

【 0 0 5 2 】

なお、パッセージ i とは、分割された n 個のパッセージのうち i 番目 ($1 \leq i \leq n$) のパッセージを意味する。 $W 1 1 \sim W 1 4$ は、予め定められた重み付け係数を表す。また、 $C 1 _ i$ 、 $C 2 _ i$ 、 $C 3 _ i$ 、および $C 4 _ i$ は、それぞれ位置情報、手がかり表現情報、タイトル関連性、および出現頻度を表す。これらの各変数の算出方法については後述する。

20

【 0 0 5 3 】

第 2 パッセージ重要度算出部 1 2 4 は、分割された各パッセージについて、ユーザサマリ内のカーソル位置における当該パッセージの重要性の度合いを表すスコアであるサマリ内重要度（第 2 パッセージ重要度）を算出する。算出方法の詳細は後述する。

【 0 0 5 4 】

総合重要度算出部 1 2 5 は、文書内重要度とサマリ内重要度との重み付け線形和（重み付け加算）により、各パッセージの総合重要度を算出する。具体的には、総合重要度算出部 1 2 5 は、以下の (2) 式により総合重要度 $S c o r e _ i$ を算出する。

30

$$S c o r e _ i = W 2 1 \times D S c o r e _ i + W 2 2 \times P S c o r e _ i + W 2 3 \times S S c o r e _ i \quad \cdots (2)$$

【 0 0 5 5 】

ただし、 $D S c o r e _ i$ はパッセージ i が含まれる文書の関連度、 $P S c o r e _ i$ はパッセージ i の文書内重要度、 $S S c o r e _ i$ はパッセージ i のサマリ重要度、 $W 2 1 \sim W 2 3$ は予め定められた正数を表す。

【 0 0 5 6 】

パッセージ選択部 1 2 6 は、総合重要度が付与されたパッセージ群を元に、ユーザに提示すべきパッセージ群を順位付きで選択する。本実施の形態では、パッセージ選択部 1 2 6 は、できるだけ重要度の高いパッセージ群を優先して選択するとともに、パッセージ間の内容ができるだけ重ならないように選択するパッセージ候補を決定する。

40

【 0 0 5 7 】

次に、編集部 1 3 0 の機能構成の詳細について説明する。同図に示すように、編集部 1 3 0 は、候補選択部 1 3 1 と、多段階検索部 1 3 2 と、編集結果記憶部 1 3 3 と、を備えている。

【 0 0 5 8 】

編集結果記憶部 1 3 3 は、編集部 1 3 0 の編集機能により編集された結果であるユーザサマリのテキストと現在のカーソル位置とを記憶する。

【 0 0 5 9 】

なお、上述の文書記憶部 1 4 0 および編集結果記憶部 1 3 3 は、H D D (Hard Disk Dr

50

ive)、光ディスク、メモリカード、R A M (Random Access Memory) などの一般的に利用されているあらゆる記憶媒体により構成することができる。

【 0 0 6 0 】

候補選択部 1 3 1 は、パッセージ選択部 1 2 6 によって選択されたパッセージ候補を、例えば図 1 に示すパッセージ候補ウィンドウ 1 1 にリスト形式で表示する。そして、候補選択部 1 3 1 は、ユーザによりパッセージ候補のいずれか 1 つが選択された場合に、サマリ編集ウィンドウ 1 0 上のカーソル位置に選択されたパッセージを挿入する。また、候補選択部 1 3 1 は、挿入処理に応じて編集結果記憶部 1 3 3 のユーザサマリの内容を更新する。なお、多段階検索部 1 3 2 により多段階パッセージ検索機能が実行されていた場合は、多段階に検索されたパッセージ群をすべて合成した内容を、サマリ編集ウィンドウ 1 0 上のカーソル位置に挿入し、編集結果記憶部 1 3 3 の格納内容を更新する。

10

【 0 0 6 1 】

多段階検索部 1 3 2 は、上述の多段階パッセージ検索機能を実現する。

【 0 0 6 2 】

次に、このように構成された第 1 の実施の形態にかかる検索支援装置 1 0 0 による検索支援処理について図 3 を用いて説明する。図 3 は、第 1 の実施の形態における検索支援処理の全体の流れを示すフローチャートである。

【 0 0 6 3 】

同図の検索支援処理は、ユーザがユーザサマリの任意のカーソル位置で特定キーを押下したときに開始される。これにより、編集部 1 3 0 は、ユーザサマリと、指定されたカーソル位置とを文脈解析部 1 1 0 に入力する。

20

【 0 0 6 4 】

受付部 1 1 1 は、入力されたユーザサマリとカーソル位置とを受付ける (ステップ S 3 0 1)。次に、構造解析部 1 1 2 が、ユーザサマリの階層構造を解析する構造解析処理を実行する (ステップ S 3 0 2)。構造解析処理の詳細は後述する。

【 0 0 6 5 】

次に、第 1 単語重要度算出部 1 1 3 が、構造解析部 1 1 2 による解析結果を元に単語一般重要度ベクトル A を算出する一般重要度算出処理を実行する (ステップ S 3 0 3)。また、第 2 単語重要度算出部 1 1 4 が、構造解析部 1 1 2 による解析結果を元に単語個別重要度ベクトル B を算出する個別重要度算出処理を実行する (ステップ S 3 0 4)。さらに、既知度算出部 1 1 5 が、構造解析部 1 1 2 による解析結果を元に単語既知度ベクトル C を算出する既知度算出処理を実行する (ステップ S 3 0 5)。

30

【 0 0 6 6 】

一般重要度算出処理、個別重要度算出処理、および既知度算出処理の詳細については後述する。なお、同図ではこれらの各処理を順次実行するように記載しているが、実行順序はこれに限られず、各処理を任意の順序で実行することができる。また、各処理の 2 つ以上を並列に実行してもよい。各処理の処理結果は、出力部 1 1 6 によりパッセージ検索部 1 2 0 に出力される。

【 0 0 6 7 】

次に、パッセージ検索部 1 2 0 の文書検索部 1 2 1 が、単語一般重要度ベクトル A および単語個別重要度ベクトル B を元に、文書記憶部 1 4 0 からユーザサマリに関連する文書を検索する文書検索処理を実行する (ステップ S 3 0 6)。次に、パッセージ分割部 1 2 2 が、検索された文書を各パッセージに分割するパッセージ分割処理を実行する (ステップ S 3 0 7)。次に、第 1 パッセージ重要度算出部 1 2 3 が、分割された各パッセージの文書内重要度を算出する文書内重要度算出処理を実行する (ステップ S 3 0 8)。さらに、第 2 パッセージ重要度算出部 1 2 4 が、分割された各パッセージのサマリ内重要度を算出するサマリ内重要度算出処理を実行する (ステップ S 3 0 9)。

40

【 0 0 6 8 】

文書検索処理、文書内重要度算出処理、およびサマリ内重要度算出処理の詳細については後述する。なお、文書内重要度算出処理およびサマリ内重要度算出処理の実行順序は同

50

図に示すものに限られるものではない。両処理を並列に実行してもよい。

【0069】

文書内重要度算出処理およびサマリ内重要度算出処理の後、総合重要度算出部125が、上記(2)式により、検索文書のスコア、文書内重要度、およびサマリ内重要度の重み付け線形和である総合重要度を算出する(ステップS310)。

【0070】

次に、パッセージ選択部126が、総合重要度を元にユーザに提示すべきパッセージ群(パッセージ候補)を選択するパッセージ選択処理を実行する(ステップS311)。パッセージ選択処理の詳細は後述する。選択されたパッセージ候補は、編集部130の候補選択部131に送出される。

10

【0071】

次に、候補選択部131が、パッセージ選択処理によって選択されたパッセージ候補を表示する(ステップS312)。ユーザが表示されたパッセージ候補からいずれかのパッセージ候補を選択した場合は、候補選択部131が、選択が指定されたパッセージを選択し、編集画面の表示を更新するとともに、編集結果記憶部133に更新内容を保存する(ステップS313)。

【0072】

次に、ステップS302の構造解析処理の詳細について図4を用いて説明する。図4は、第1の実施の形態における構造解析処理の全体の流れを示すフローチャートである。

【0073】

20

まず、構造解析部112は、ユーザサマリを読み込みパッセージ単位に分解する(ステップS401)。構造解析部112によるパッセージ分解処理は従来から用いられているあらゆる方法を適用できる。例えば、図1のような箇条書きテキストに対しては、構造解析部112は、1つの箇条書き項目を1つのパッセージとして分解する方法を適用することができる。この場合、パッセージが1つの文に相当するとは限らない。

【0074】

入力テキストとして、図1のような箇条書きテキストではなく、単なるプレーンテキストを受付けた場合は、文字列の並びを解析することによりパッセージ単位に分解するように構成してもよい。例えば、行頭に「*」などの所定の記号が存在する場合に、この行は箇条書きとして記載されていると判断するといったルールをベースに構造を抽出することができる。

30

【0075】

次に、構造解析部112は、分割して得られたパッセージ群について、パッセージ同士の兄弟関係および親子関係を抽出する(ステップS402)。例えば、図1のような箇条書きテキストに対しては、構造解析部112は、箇条書きなどの書式表現を分析することにより、兄弟関係および親子関係を抽出できる。例えば図1のテキストの場合、最上位のパッセージが「株価」、その子供のパッセージが「2008/8/28」、さらにその子供のパッセージ群が「株価平均は小反発・・・」であることを抽出できる。

【0076】

以上の処理により、パッセージ間の階層構造関係すなわち親子関係および兄弟関係が付与されたパッセージ群を得ることができる。

40

【0077】

次に、ステップS303の一般重要度算出処理の詳細について図5を用いて説明する。図5は、第1の実施の形態における一般重要度算出処理の全体の流れを示すフローチャートである。

【0078】

まず、第1単語重要度算出部113は、カーソルが所属するパッセージより上位の構造上位パッセージ群の全体に対して形態素解析を行い、1つの形態素集合を取得する(ステップS501)。次に、第1単語重要度算出部113は、形態素集合内の各形態素(各内容語)について、以下の(3)式により一般重要度 GW_j を算出する(ステップS50

50

2)。

$$GW_j = idf(j) \times bExist_j \cdots (3)$$

【0079】

ただし、 j は $0 \leq j \leq m$ を満たす整数 (m は内容語の種類数)、 $idf(j)$ は内容語 j の全文書中における逆出現頻度 (Inverse Document Frequency)、 $bExist_j$ は形態素集合内に内容語 j が存在する場合は 1 を、存在しない場合は 0 となる変数を表す。ここで全文書というのは、文書記憶部にある文書群である。あるいは Web 文書群などですでに算出済の idf 値を用いても良い。なお、 idf の代わりに別の重要度指標を用いてもよい。

【0080】

次に、第 1 単語重要度算出部 113 は、算出した GW_j を要素とする以下の (4) 式のベクトルを、単語一般重要度ベクトル A として算出し (ステップ S503)、一般重要度算出処理を終了する。

$$A = (GW_0, GW_1, \cdots, GW_m) \cdots (4)$$

【0081】

次に、ステップ S304 の個別重要度算出処理の詳細について図 6 を用いて説明する。図 6 は、第 1 の実施の形態における個別重要度算出処理の全体の流れを示すフローチャートである。

【0082】

まず、第 2 単語重要度算出部 114 は、カーソルが所属するパッセージより上位の構造上位パッセージ群の全体に対して形態素解析を行い、1 つの形態素集合を取得する (ステップ S601)。次に、第 2 単語重要度算出部 114 は、形態素集合内の各内容語について、以下の (5) 式により個別重要度 IW_j を算出する (ステップ S602)。

【数 1】

$$IW_j = \sum_{t \in T_j} \left(\log \left(\frac{1}{\text{dist}(t)} \right) + 1 \right) \cdots (5)$$

【0083】

ただし、 j は $0 \leq j \leq m$ を満たす整数 (m は内容語の種類数)、 T_j は形態素集合のうち内容語 j である形態素を集めたサブ集合、 $\text{dist}(t)$ は形態素 t が出現したパッセージと、カーソルが存在するパッセージとの階層的距離を表す。階層的距離とは、階層構造上で、一方のパッセージから他方のパッセージに達するまでの階層数をいう。例えば、形態素 t が出現したパッセージとカーソルが存在するパッセージとが同一の場合は、階層的距離は 0 となる。

【0084】

図 1 の例では、最上位から第 6 位のパッセージにカーソルが存在する。この場合、「NYMEX」から始まるパッセージ内に含まれる各形態素 t に対する $\text{dist}(t)$ の算出結果は 1 となる。また、最上位のパッセージ内に含まれる形態素 t (すなわち「株価」) の $\text{dist}(t)$ の算出結果は 5 となる。

【0085】

なお、階層的距離を、パッセージ階層間に含まれるテキストの文字数などをベースに算出するように構成してもよい。例えば、パッセージ階層間に含まれる文字数が多い場合に、階層的距離を大きくするように構成してもよい。このような構成により、パッセージのテキスト量に応じて重みを変化させることができる。

【0086】

次に、第 2 単語重要度算出部 114 は、算出した IW_j を要素とする以下の (6) 式のベクトルを、単語個別重要度ベクトル B として算出し (ステップ S603)、個別重要度算出処理を終了する。

$$B = (IW_0, IW_1, \cdots, IW_m) \cdots (6)$$

【0087】

10

20

30

40

50

次に、ステップ S 3 0 5 の既知度算出処理の詳細について図 7 を用いて説明する。図 7 は、第 1 の実施の形態における既知度算出処理の全体の流れを示すフローチャートである。

【 0 0 8 8 】

まず、既知度算出部 1 1 5 は、カーソルが所属するパッセージと同位の構造同位パッセージ群の全体に対して形態素解析を行い、1 つの形態素集合を取得する（ステップ S 7 0 1 ）。次に、既知度算出部 1 1 5 は、形態素集合内の各内容語について、以下の（ 7 ）式により既知度 KW_k を算出する（ステップ S 7 0 2 ）。

$$KW_k = tf(k) \cdots (7)$$

【 0 0 8 9 】

ただし、 k は 0 i 1 を満たす整数（1 は内容語の種類数）、 tf は単語の出現頻度（term frequency）を示す。なお、 tf の代わりに、別の重要度指標を用いてもよい。

【 0 0 9 0 】

次に、既知度算出部 1 1 5 は、算出した KW_k を要素とする以下の（ 8 ）式のベクトルを、単語既知度ベクトル C として算出し（ステップ S 7 0 3 ）、既知度算出処理を終了する。

$$C = (KW_0, KW_1, \cdots, KW_1) \cdots (8)$$

【 0 0 9 1 】

次に、ステップ S 3 0 6 の文書検索処理の詳細について図 8 を用いて説明する。図 8 は、第 1 の実施の形態における文書検索処理の全体の流れを示すフローチャートである。

【 0 0 9 2 】

まず、文書検索部 1 2 1 は、単語一般重要度ベクトル A および単語個別重要度ベクトル B の重み付け線形和である文書検索用ベクトル Q を、以下の（ 9 ）式を用いて算出する（ステップ S 8 0 1 ）。文書検索用ベクトル Q は、重み付きの単語ベクトルとなる。

$$Q = W_{31} \times A + W_{32} \times B \quad (W_{31}, W_{32} \text{ は正数}) \cdots (9)$$

【 0 0 9 3 】

なお、文脈解析部 1 1 0 の出力部 1 1 6 が単語一般重要度ベクトル A および単語個別重要度ベクトル B から文書検索用ベクトル Q を生成し、パッセージ検索部 1 2 0 に出力するように構成してもよい。

【 0 0 9 4 】

次に、文書検索部 1 2 1 は、文書記憶部 1 4 0 内に格納されている各文書の文書ベクトルと文書検索用ベクトル Q とのコサイン尺度を関連度として算出する（ステップ S 8 0 2 ）。これにより文書ごとに関連度が得られる。すなわち関連度付きの文書群が得られる。なお、後続の処理効率化のため、関連度が所定の閾値を下回っている文書を出力に含めないように構成してもよい。また、関連度が上位の所定数の文書のみを関連文書として検索するように構成してもよい。

【 0 0 9 5 】

次に、ステップ S 3 0 7 のパッセージ分割処理の詳細について図 9 を用いて説明する。図 9 は、第 1 の実施の形態におけるパッセージ分割処理の全体の流れを示すフローチャートである。

【 0 0 9 6 】

まず、パッセージ分割部 1 2 2 は、検索された各文書を、表層レベルでパッセージ単位に分割する（ステップ S 9 0 1 ）。例えば、文書がプレーンテキストの場合、パッセージ分割部 1 2 2 は、「。」、「、」、「？」、「！」など終末記号に基づいて文書を文単位に分解する。なお、分割方法はこれに限られず、従来から用いられているあらゆる方法を適用できる。例えば、改行記号に基づいてパラグラフとみなせる単位で文書を分解するように構成してもよい。

【 0 0 9 7 】

また、例えば HTML 形式の文書の場合、終末記号による分割が適用できない場合がある。例えば、レイアウト的に距離を開けること、または、罫線や配色などを用いることに

10

20

30

40

50

より、テキストが相互に分割されていることを示すようにHTML文書が記載されている場合が挙げられる。このような場合は、例えば、HTMLの表示イメージをレイアウト解析することによりテキストの分割位置を特定し、パッセージ単位に分割するように構成すればよい。

【0098】

このようにして得られたパッセージは、文法的解釈を行っていないため、そのパッセージのみでは内容として完結していない場合がある。すなわち、当該パッセージ以外のパッセージ等から情報が得られなければ意味がわからない状態になっている可能性がある。

【0099】

例えば「その結果上昇した。」という文では、主語が何なのか（何が上昇したのか）が省略されている。また、「上昇」の原因となった「その」が何を指すのかを表す照応詞の解決がなされていない。このため、この文のみでは意味を理解することができない。

【0100】

そこで、パッセージ分割部122は、構文解析や照応解析を実行し、必要に応じて抽出したパッセージを修正する（ステップS902～ステップS903）。

【0101】

具体的には、パッセージ分割部122は、分割したパッセージに、「その」、「あの」、および「この」などの照応詞が含まれる場合は、1つ前のパッセージの中から、先行詞である可能性が高い名詞句を抽出し、照応詞を抽出した名詞句で置き換える（ステップS902）。また、パッセージ分割部122は、分割したパッセージを構文解析し、主語が省略されていることが判明した場合、1つ前のパッセージの中から対応する名詞句を検索して主語として追加する（ステップS903）。

【0102】

次に、ステップS308の文書内重要度算出処理の詳細について図10を用いて説明する。図10は、第1の実施の形態における文書内重要度算出処理の全体の流れを示すフローチャートである。

【0103】

まず、第1パッセージ重要度算出部123は、パッセージ分割処理により得られたパッセージ群の集合（パッセージ群集合）から、未処理のパッセージ（パッセージ*i*とする）を取得する（ステップS1001）。次に、第1パッセージ重要度算出部123は、パッセージ*i*の位置情報C1__*i*を算出する（ステップS1002）。

【0104】

位置情報C1__*i*は、当該パッセージが、所属する文書内のいずれの位置に存在しているかを表す指標である。例えば、ニュース記事などでは、文書の先頭から3文以内に存在する文は非常に重要度が高いと言われている。位置情報C1__*i*はこのような知見を元に、例えば文書の先頭から所定数のパッセージである場合に高い値となり、それ以外の場合は文書の後ろに存在するほど小さくなる値となるように算出される。

【0105】

次に、第1パッセージ重要度算出部123は、パッセージ*i*の手がかり表現情報C2__*i*を算出する（ステップS1003）。

【0106】

手がかり表現情報C2__*i*は、予め定められた表層表現（手がかり表現）を含むか否かによりパッセージの重要度を決定するための指標である。例えば論文などでは「本研究（で）は」、「まとめると」、および「我々は」などの表現を含む文は、論文の主題を表すと考えられる。第1パッセージ重要度算出部123は、このような知見を元に、例えば所定の手がかり表現とスコアとを対応付ける辞書を用いて、パッセージに含まれる手がかり表現に対応づけられたスコアを、手がかり表現情報C2__*i*として算出する。

【0107】

次に、第1パッセージ重要度算出部123は、パッセージ*i*のタイトル関連性C3__*i*を算出する。タイトル関連性C3__*i*は、タイトルと一致するパッセージまたはタイトル

10

20

30

40

50

と関連するパッセージは重要であるとする指標である。第1パッセージ重要度算出部123は、文書にタイトル（またはHTMLのメタタグなどのようなタイトルに準じるコンテンツ）が存在する場合、この指標の一例として、タイトルの文書ベクトルと評価対象のパッセージの文書ベクトルとのコサイン尺度をタイトル関連性 $C3_i$ として算出する（ステップS1004）。タイトルの文書ベクトルおよびパッセージの文書ベクトルは、 $tf-idf$ などにより算出する。

【0108】

次に、第1パッセージ重要度算出部123は、パッセージ i の出現頻度 $C4_i$ を算出する。出現頻度 $C4_i$ は、所属する文書内の他のパッセージと強い関連性を持つパッセージが重要であることを表す指標である。語彙的結束性とも呼ばれる。第1パッセージ重要度算出部123は、この指標の一例として、評価対象のパッセージと他のパッセージとの単語共起数の総和を出現頻度 $C4_i$ として算出する（ステップS1005）。

10

【0109】

次に、第1パッセージ重要度算出部123は、各パッセージについて、 $C1_i \sim C4_i$ を予め定めた重み付け係数で重み付け加算することにより、当パッセージの文書内重要度を算出する（ステップS1006）。具体的には、第1パッセージ重要度算出部123は、パッセージ i の文書内重要度 $PScore_i$ を上記（1）式により算出する。

【0110】

次に、第1パッセージ重要度算出部123は、未処理のパッセージが存在するか否かを判断し（ステップS1007）、存在する場合は（ステップS1007：YES）、未処理のパッセージを取得して処理を繰り返す（ステップS1001）。存在しない場合は（ステップS1007：NO）、文書内重要度算出処理を終了する。

20

【0111】

次に、ステップS309のサマリ内重要度算出処理の詳細について図11を用いて説明する。図11は、第1の実施の形態におけるサマリ内重要度算出処理の全体の流れを示すフローチャートである。

【0112】

まず、第2パッセージ重要度算出部124は、単語一般重要度ベクトルA、単語個別重要度ベクトルB、および単語既知度ベクトルCを用いて、サマリ内重要度評価用ベクトルQを以下の（10）式により算出する。

30

$$Q = W41 \times A + W42 \times B - W43 \times C \quad \cdots (10)$$

【0113】

なお、 $W41$ 、 $W42$ 、および $W43$ は予め定められた正数を表す。結果として得られるサマリ内重要度評価用ベクトルQは、重み付きの単語ベクトルとなる。

【0114】

次に、第2パッセージ重要度算出部124は、パッセージ分割処理により得られたパッセージ群集合から、未処理のパッセージ（パッセージ i とする）を取得する（ステップS1102）。次に、第2パッセージ重要度算出部124は、サマリ内重要度評価用ベクトルQと、パッセージ i の文書ベクトルとのコサイン尺度を、パッセージ i のサマリ内重要度として算出する（ステップS1103）。なお、第2パッセージ重要度算出部124は、パッセージ i の文書ベクトルを例えば $tf-idf$ を用いて算出する。

40

【0115】

次に、第2パッセージ重要度算出部124は、未処理のパッセージが存在するか否かを判断し（ステップS1104）、存在する場合は（ステップS1104：YES）、未処理のパッセージを取得して処理を繰り返す（ステップS1101）。存在しない場合は（ステップS1104：NO）、サマリ内重要度算出処理を終了する。

【0116】

次に、ステップS311のパッセージ選択処理の詳細について図12を用いて説明する。図12は、第1の実施の形態におけるパッセージ選択処理の全体の流れを示すフローチャートである。

50

【 0 1 1 7 】

まず、パッセージ選択部 1 2 6 は、パッセージ分割処理により得られたパッセージ群集合から、未処理のパッセージ（パッセージ i とする）を取得する（ステップ S 1 2 0 1）。次に、パッセージ選択部 1 2 6 は、以下の（ 1 1 ）式により、パッセージ i の提示有効度を算出する（ステップ S 1 2 0 2）。

【数 2】

$$CScore_i = Score_i - \sum_{j \in NS} sim(i, h) \quad \cdots (11)$$

【 0 1 1 8 】

ただし、 $Score_i$ はパッセージ i の総合重要度、 NS は未選択のパッセージ群、 $sim(i, h)$ はパッセージ i とパッセージ h との類似度を表す。提示有効度は、各パッセージの重要度と新規性（例えば、すでに提示決定されたパッセージと内容がかぶる場合は新規性はないと判断される）とを考慮に入れた指標である。

10

【 0 1 1 9 】

次に、パッセージ選択部 1 2 6 は、算出した提示有効度のうち、最も高い提示有効度を示すパッセージを提示候補として決定する（ステップ S 1 2 0 3）。次に、パッセージ選択部 1 2 6 は、一定数（例えば 10 件）の候補が決定されたか否かを判断する（ステップ S 1 2 0 4）。決定されていない場合（ステップ S 1 2 0 4：NO）、未処理のパッセージを選択して処理を繰り返す（ステップ S 1 2 0 1）。

【 0 1 2 0 】

20

一定数の候補が決定された場合（ステップ S 1 2 0 4：YES）、パッセージ選択部 1 2 6 は、提示有効度の高い順に決定されたパッセージ候補を出力し（ステップ S 1 2 0 5）、パッセージ選択処理を終了する。

【 0 1 2 1 】

次に、多段階検索部 1 3 2 による多段階パッセージ検索機能の詳細について図 1 3 を用いて説明する。図 1 3 は、第 1 の実施の形態における多段階パッセージ検索処理の全体の流れを示すフローチャートである。以下では、子供となるパッセージを検索するための入力キーを特定キー 1 とし、兄弟となるパッセージを検索するための入力キーを特定キー 2 とする。

【 0 1 2 2 】

30

まず、多段階検索部 1 3 2 は、パッセージ候補ウィンドウで、特定キー 1 が押下されたか否かを判断する（ステップ S 1 3 0 1）。押下された場合（ステップ S 1 3 0 1：YES）、多段階検索部 1 3 2 は、編集結果記憶部 1 3 3 から、ユーザサマリとカーソル位置とを読み出し、メモリバッファ等の記憶部（図示せず）上にロードする（ステップ S 1 3 0 3）。

【 0 1 2 3 】

次に、多段階検索部 1 3 2 は、メモリバッファ内で、パッセージ候補ウィンドウ上でフォーカスされているパッセージを、サマリのカーソル位置へ挿入する（ステップ S 1 3 0 4）。次に、多段階検索部 1 3 2 は、メモリバッファ内で、挿入したパッセージの子供として空パッセージを生成し、その空パッセージ上にカーソルを置く（ステップ S 1 3 0 5）。次に、多段階検索部 1 3 2 は、メモリバッファ上で生成されたサマリとカーソル位置とを文脈解析部 1 1 0 に送信する（ステップ S 1 3 0 6）。

40

【 0 1 2 4 】

ステップ S 1 3 0 1 で特定キー 1 が押下されていない場合（ステップ S 1 3 0 1：NO）、多段階検索部 1 3 2 は、さらに、特定キー 2 が押下されたか否かを判断する（ステップ S 1 3 0 2）。押下された場合（ステップ S 1 3 0 2：YES）、多段階検索部 1 3 2 は、編集結果記憶部 1 3 3 から、ユーザサマリとカーソル位置とを読み出し、メモリバッファ上にロードする（ステップ S 1 3 0 7）。

【 0 1 2 5 】

次に、多段階検索部 1 3 2 は、メモリバッファ内で、パッセージ候補ウィンドウ上でフ

50

オーカスされているパッセージを、サマリのカーソル位置へ挿入する（ステップS 1 3 0 8）。次に、多段階検索部 1 3 2 は、メモリバッファ内で、挿入したパッセージの兄弟として空パッセージを生成し、その空パッセージ上にカーソルを置く（ステップS 1 3 0 9）。次に、多段階検索部 1 3 2 は、メモリバッファ上で生成されたサマリとカーソル位置とを文脈解析部 1 1 0 に送信する（ステップS 1 3 1 0）。

【0 1 2 6】

なお、同図では省略しているが、ステップS 1 3 0 6 およびステップS 1 3 1 0 の後は、図 3 と同様の処理により、送信されたユーザサマリとカーソル位置とを元に関連するパッセージが文書記憶部 1 4 0 の文書から選択され、編集部 1 3 0 に出力される。

【0 1 2 7】

ステップS 1 3 0 2 で特定キー 2 が押下されていないと判断された場合（ステップS 1 3 0 2：NO）、多段階パッセージ検索処理を終了する。

【0 1 2 8】

（変形例）

上記第 1 の実施の形態では、構造解析部 1 1 2 は、箇条書きなどで表現されたユーザサマリを解析することにより、階層構造を求めていた。本変形例では、ユーザによって編集されたサマリテキストが、箇条書き表現などにより構造が明示化されていない場合であっても、第 1 の実施の形態と同等のパッセージ検索機能を提供可能とする。具体的には、本変形例の構造解析部 1 1 2 は、入力されたテキスト（ユーザサマリ）を修辞解析することで論理構造を認識する。

【0 1 2 9】

図 1 4 は、本変形例の検索支援装置で表示される編集画面の一例を示す図である。同図に示すように、本変形例のサマリ編集ウィンドウ 1 4 1 0 内のテキストの内容は、図 1 のサマリ編集ウィンドウ 1 0 内のテキストの内容と同じである。しかし、本変形例では、箇条書きが示されていないこと、および、文と文との接続する語句として「・・・の背景としては」などをユーザが加筆している点が異なっている。

【0 1 3 0】

本変形例では、構造解析部 1 1 2 は、入力されたユーザサマリに対して、R S T（修辞構造理論）に基づく論理構造解析を行う。R S Tでは、パッセージ間の関係をツリー構造で表現するとともに、ツリー構造の各ノード間に 2 4 種類のラベルを付与する。本変形例ではラベルを利用しないため、解析結果としてラベルが得られても破棄する。

【0 1 3 1】

なお、第 1 の実施の形態の構造解析手法と、本変形例の構造解析手法とを組み合わせることで、箇条書きなどの明示的に構造が示された記述と、明示的に構造が示されない記述とが混在したテキストを取り扱うように構成することが可能である。具体的には、最初に明示的に示された箇条書き記述を元にツリー構造（階層構造）を作成した後、ツリーの各ノードに対して、R S Tに基づく論理構造解析を行えばよい。

【0 1 3 2】

次に、本変形例の構造解析処理の詳細について図 1 5 を用いて説明する。図 1 5 は、第 1 の実施の形態の変形例における構造解析処理の全体の流れを示すフローチャートである。

【0 1 3 3】

まず、構造解析部 1 1 2 は、第 1 の実施の形態と同様に（図 4 のステップS 4 0 1）、プレーンテキスト形式で表現されたユーザサマリを読み込みパッセージ単位に分解する（ステップS 1 5 0 1）。プレーンテキストに対しては、構造解析部 1 1 2 は、「。」、「、」、「？」、「！」など終末記号に基づいてテキストを文単位に分解する。なお、分割方法はこれに限られず、従来から用いられているあらゆる方法を適用できる。例えば、上述のパッセージ分割部 1 2 2 と同様に、改行記号に基づいてパラグラフとみなせる単位で文書を分解するように構成してもよい。また、例えばHTML形式の文書の場合、HTMLの表示イメージをレイアウト解析することによりテキストの分割位置を特定し、

10

20

30

40

50

パッセージ単位に分割するように構成してもよい。

【 0 1 3 4 】

次に、構造解析部 1 1 2 は、R S T 解析用の手がかり語辞書を参照して、パッセージ間の上下関係を同定する（ステップ S 1 5 0 2）。例えば、「・・・（パッセージ A）。その結果、（パッセージ B）・・・」のように 2 つのパッセージが記載されている場合、「その結果」という手がかり語を判断材料として、パッセージ A が親、パッセージ B が子であり、両者の関係は「証拠 / 原因」であることが解析される。なお、R S T では親を核、子を衛星と呼ぶ。手がかり語辞書とは、このようにパッセージ間の関係の同定材料になる特別な言い回しを集めたデータベースである。

【 0 1 3 5 】

本ステップでは、隣接するパッセージ間に手がかり語が存在するか否かを照合し、照合した場合は、両者の関係を同定する処理を反復的にを行い、最終的にツリー構造を作成する。

【 0 1 3 6 】

このように、本変形例は、例えば他人に配布する清書された報告書などのようなプレーンテキストを作成する場合に、第 1 の実施の形態よりも有用となる可能性がある。すなわち、第 1 の実施の形態が想定しているテキスト表現は、他人へも配布可能な清書されたレポートというよりは、自分専用の調査メモという色彩が強い。これに対して、本変形例が想定しているテキスト表現は、他人へも配布可能なレポートと似た表現が許されている。このため、最終的な生成物の形に向かって調査を進めていくことが可能となり、作業効率化が期待される。

【 0 1 3 7 】

なお、第 1 の実施の形態の実現方法と、本変形例の実現方法とを組み合わせることで、箇条書きなど明示的に構造が示された記述と、上記で説明した明示的に示されない記述とが混在したテキストを取り扱うことも可能である。

【 0 1 3 8 】

このように、第 1 の実施の形態にかかる検索支援装置では、ユーザサマリ内の単語の重要度やパッセージ間の階層構造に応じた文書検索用ベクトルを作成できるため、キーワード入力の検索システムでは不可能な精度の高い情報要求をかけることができるとともに、ユーザはクエリ作成作業から解放される。また、ユーザサマリを検索要求の源とすることで、関連文書検索の徘徊問題を回避することができる。また、パッセージを検索結果として取得できるため、ユーザが文書中から新規情報を探す作業を低減できる。また、パッセージ挿入後のサマリ編集はユーザが行うため、複数文書要約技術では文脈統合が実用レベルに達していないという問題を回避可能となる。

【 0 1 3 9 】

（第 2 の実施の形態）

第 1 の実施の形態では、探求検索を目的として、ユーザがサマリを作成することを前提としていた。第 2 の実施の形態は、ユーザがサマリを作成することなく探求検索を行う。具体的には、第 2 の実施の形態の検索支援装置は、ユーザが W e b 上の文書等を閲覧するときに効率的に探求検索を実行可能とする装置である。

【 0 1 4 0 】

次に、画面イメージを参照しながら（図 1 6）、本実施の形態の検索支援装置の機能の概要について説明する。図 1 6 は、第 2 の実施の形態の検索支援装置で表示される編集画面の一例を示す図である。

【 0 1 4 1 】

ユーザは、閲覧ウィンドウ 1 6 1 0 によって、例えば W e b 文書などの所望の文書を閲覧することができる。第 1 の実施の形態と同様に、閲覧ウィンドウ 1 6 1 0 内の任意の箇所にカーソルを設定することができる。また、特定キーが押下されると、パッセージ候補ウィンドウ 1 1 が表示される。

【 0 1 4 2 】

パッセージ候補ウィンドウ 11 には、閲覧文書のカーソル位置の文脈に沿って文書群から検索されたパッセージ群の候補が表示される。ユーザがパッセージ候補の 1 つを選択すると、そのパッセージの原文書が閲覧ウィンドウ 1610 に表示される。

【0143】

また、第 1 の実施の形態と同様に、多段階パッセージ検索機能も利用できる。多段階パッセージ検索機能を用いて、ウィンドウを複数呼び出し、最終的にパッセージを選択した場合、選択したパッセージの原文書が閲覧ウィンドウ 1610 に表示される。

【0144】

なお、キーワードの代わりに文やパッセージをユーザに入力させ、入力された文やパッセージに適合する文書群を返すフレーズ検索システム（関連文書検索システム）が知られている。フレーズ検索システムは、文書のうち範囲指定内の指定パッセージのみを用いて情報要求を生成する。これに対して、本実施の形態の検索支援装置は、カーソル近辺に存在するパッセージ、すなわち指定パッセージだけでなく、階層構造や文字的距離に基づく傾斜加重を施した上で、文書全体にわたって指定パッセージの前後に存在するパッセージも含めて情報要求を生成する。このため、本実施の形態の検索支援装置による検索結果は、文書全体の内容や文脈を織り込んだ内容となる点が、従来のフレーズ検索システムと異なっている。

【0145】

次に、本実施の形態の検索支援装置の構成の詳細について図 17 を用いて説明する。図 17 は、第 2 の実施の形態にかかる検索支援装置 200 の構成の一例を示すブロック図である。同図に示すように、検索支援装置 200 は、文脈解析部 110 と、パッセージ検索部 120 と、閲覧部 230 と、文書記憶部 140 と、を備えている。

【0146】

第 2 の実施の形態では、編集部 130 の代わりに閲覧部 230 を追加したことが第 1 の実施の形態と異なっている。その他の構成および機能は、第 1 の実施の形態にかかる検索支援装置 100 の構成を表すブロック図である図 2 と同様であるので、同一符号を付し、ここでの説明は省略する。

【0147】

閲覧部 230 は、表示装置（図示せず）に、検索されたテキストを閲覧可能に表示する機能を備えている。閲覧部 230 は、編集部 130 のように表示されたテキストを編集する機能を備える必要はない。

【0148】

閲覧部 230 は、候補選択部 231 と、多段階検索部 232 と、閲覧結果記憶部 233 とを備えている。

【0149】

閲覧結果記憶部 233 は、検索された文書と現在のカーソル位置とを記憶する。

【0150】

候補選択部 231 および多段階検索部 232 は、ユーザによりパッセージ候補が選択された後の動作が、それぞれ第 1 の実施の形態の候補選択部 131 および多段階検索部 132 と異なる。すなわち、候補選択部 231 および多段階検索部 232 は、ユーザがパッセージ候補の 1 つを選択すると、選択されたパッセージ候補の原文書の内容を文書記憶部 140 から読み出し、読み出した文書内容を閲覧結果記憶部 233 に格納する。

【0151】

このように、第 2 の実施の形態にかかる検索支援装置では、ユーザが Web 上の文書等を閲覧するときにも、第 1 の実施の形態と同様の手法により、効率的に探求検索を実行することが可能となる。

【0152】

（第 3 の実施の形態）

第 3 の実施の形態にかかる検索支援装置は、指定された形態素の品詞に応じて予め定められた修辭語を検索キーワードとして検索要求に付加する。これにより、形態素の品詞に

10

20

30

40

50

応じてより適切なパッセージ候補を検索することが可能となる。

【 0 1 5 3 】

次に、画面イメージを参照しながら（図 1 8）、本実施の形態の検索支援装置の機能の概要について説明する。図 1 8 は、第 3 の実施の形態の検索支援装置で表示される編集画面の一例を示す図である。

【 0 1 5 4 】

ユーザは、サマリ編集ウィンドウ 1 8 1 0 内でキーボード・マウスなどを用いることにより構造テキスト（ユーザサマリ）を作成および編集することができる。本実施の形態では、ユーザサマリ内の各形態素が、その品詞に応じて異なる表示態様で強調表示される点が、第 1 および第 2 の実施の形態と異なる。

10

【 0 1 5 5 】

具体的には、動作性名詞または動詞は矩形の枠で囲まれて表示される。また、動作性名詞以外の名詞は下線が付されて表示される。編集などによりユーザサマリが更新されるたびに（例えば文字が入力されるごとに）、適切なタイミングで名詞の品詞の判定処理および表示態様の変更処理が行われ、常に正しい表示態様により表示されるものとする。このハイライト（強調表示）は後述のパッセージ検索機能で利用される。

【 0 1 5 6 】

なお、上記のような強調表示のための表示態様は一例であり、従来から用いられているあらゆる強調表示方法を適用できる。例えば、文字色、フォントの種類、およびフォントサイズ等の表示態様を品詞ごとに変更するように構成してもよい。以下では、動作性名詞または動詞に対する強調表示を第 1 ハイライトといい、動作性名詞以外の名詞に対する強調表示を第 2 ハイライトという。

20

【 0 1 5 7 】

ユーザは、パッセージ候補ウィンドウ 1 8 1 1 によって、文献 D B に記憶されている複数の文書内のパッセージ群を検索し取り込むことができる。本実施の形態では、検索機能を利用するための操作方法として 2 種類の操作方法を用いる。

【 0 1 5 8 】

1 つは、第 1 ハイライト上でマウスクリックなどの特定キーを押下する操作である。第 1 ハイライトで表示された形態素は、動作性名詞または動詞であり、この操作により、動作性名詞または動詞が表す動作がなぜ起こったのか、を説明するパッセージが検索される。具体的には、上記操作により「それはなぜ？」という質問、すなわち原因や背景などを要求する質問が、検索要求に付加され、この結果、原因や背景を説明するパッセージが検索される。例えば、ユーザが、サマリ編集ウィンドウ 1 8 1 0 内の最下層のパッセージ（図 1 8 参照）に含まれる単語「上昇」の上で特定キーを押下すると、「上昇」という動作の主語が「原油先物価格」であることが同定された上で、原油先物価格がなぜ上昇したのか、に関する説明として適切なパッセージ群がパッセージ候補として検索され、パッセージ候補ウィンドウ 1 8 1 1 に提示される。

30

【 0 1 5 9 】

もう 1 つは、第 2 ハイライト上でマウスクリックなどの特定キーを押下する操作である。第 2 ハイライトで表示された形態素は、動作性名詞以外の名詞であり、この操作により、その名詞がどのようなものなのかを説明するパッセージが検索される。具体的には、上記操作により「それは何？」という質問、すなわち定義などを要求する質問が、検索要求に付加され、この結果、名詞の定義等を説明するパッセージが検索される。例えば、ユーザが、サマリ編集ウィンドウ 1 8 1 0 内の最下層のパッセージ（図 1 8 参照）に含まれる単語「N Y M E X」の上で特定キーを押下すると、ニューヨーク・マーカンタイル取引所に関する説明として適切なパッセージ群がパッセージ候補として検索され、パッセージ候補ウィンドウ 1 8 1 1 に提示される。

40

【 0 1 6 0 】

第 1 および第 2 の実施の形態の情報要求は、「もっと詳しく」というような漠然とした要求と言えるが、第 3 の実施の形態の情報要求は、「それは何？」「それはなぜ？」など

50

のように的を絞った要求である。探求検索は、情報の収集過程でユーザの精通度が動的に変化するため、情報要求を変更する操作（コマンド）を適宜使い分けられることにより、ユーザの利便性を向上させることができる。

【 0 1 6 1 】

なお、上記説明では、品詞種別に応じた質問の種類（質問タイプ）として、定義（それは何？）および理由（それはなぜ？）の2つのみを挙げた。しかし、質問タイプはこれらに限られるものではなく、この他にも様々な質問タイプを用いることができる。例えば、動作性名詞または動詞に対する質問タイプとして、結果（それでどうなる？）を用いるように構成してもよい。上記例の場合、「原油先物価格が上昇した結果どうなるのか？」という質問が追加される。RSTでは24種類のパッセージ間関係を定義しているため、RSTに基づいて24種類の質問を利用することもできる。本実施の形態では、説明の簡便性のため、2種類の質問のみを取り扱う。

10

【 0 1 6 2 】

次に、本実施の形態の検索支援装置の構成の詳細について図19を用いて説明する。図19は、第3の実施の形態にかかる検索支援装置300の構成の一例を示すブロック図である。同図に示すように、検索支援装置300は、文脈解析部310と、パッセージ検索部320と、編集部330と、文書記憶部140と、を備えている。

【 0 1 6 3 】

第3の実施の形態では、文脈解析部310に係り受け解析部317と生成部318とを追加したこと、パッセージ検索部320の文書検索部321と第2パッセージ重要度算出部324の機能、および、編集部330の機能が第1の実施の形態と異なっている。その他の構成および機能は、第1の実施の形態にかかる検索支援装置100の構成を表すブロック図である図2と同様であるので、同一符号を付し、ここでの説明は省略する。

20

【 0 1 6 4 】

編集部330は、各種ユーザ操作を受け付け、受け付けたユーザ操作に応じた処理を各構成部に指示する。本実施の形態では、編集部330は、ユーザサマリおよびカーソル位置に加え、カーソル位置の単語を表す指定形態素と、指定形態素の品詞を表す品詞情報とを文脈解析部310へ渡す点が、第1の実施の形態の編集部130と異なっている。

【 0 1 6 5 】

なお、カーソル位置の代わりに、指定形態素の位置を表す情報を渡すように構成してもよい。すなわち、指定された位置に対応するパッセージが特定可能な情報であればどのような情報を渡してもよい。また、品詞情報を渡さずに、文脈解析部310内で指定形態素の品詞を判定するように構成してもよい。

30

【 0 1 6 6 】

係り受け解析部317は、カーソル位置の単語を含む文の句構造（係り受け構造）を解析する。日本語のテキストに対しては、係り受け解析部317は、例えばKNPやCabochaなどの解析手法を適用して句構造解析を実行することができる。

【 0 1 6 7 】

以下に、句構造解析結果の例を示す。以下の（入力）に示すテキストに対して、係り受け解析部317は、以下の（出力）に示す解析結果を出力する。

40

（入力）「NYMEXの時間外取引で原油先物価格が1バレル=119ドル台に上昇」

（出力）（（（（（NYMEX）の）時間外取引）で[adv]）（（（1バレル=119ドル台）に[adv]）（（（原油先物価格）が[np]）（上昇した[vp]））））

【 0 1 6 8 】

なお、記号「（）」はツリー構造を示す。すなわち、「上昇した[vp]」がツリーのルートであり、「原油先物価格が[np]」がその直下ノードとなる。「原油価格が上昇した」に対して2つの副詞句(adv)がぶら下がる。また（）内の最後尾に記載される[]は、句の種別を示す。なお、上記以外の句構造の表現を用いるように構成してもよい。

【 0 1 6 9 】

50

生成部 318 は、指定形態素と句構造解析結果とを用いて、文書検索用ベクトルに追加する質問を表すベクトルである質問ベクトルを生成する。生成部 318 は、例えば以下の (12) 式のような質問ベクトル D を生成する。

$$D = (QW_0, QW_1, \dots, QW_n) \quad \dots (12)$$

【0170】

QW_i は、指定形態素が属する文に含まれる形態素および修辞語（後述）に含まれる形態素のうち、 i 番目の形態素 ($1 \leq i \leq n$) に対する重み、 n は指定形態素が属する文に含まれる形態素および修辞語に含まれる形態素の総数を表す。重み QW_i の算出方法、および質問ベクトル D の算出方法の詳細は後述する。

【0171】

文書検索部 321 は、単語一般重要度ベクトル A、単語個別重要度ベクトル B、に加え、質問ベクトル D を用いて、関連文書を検索する点が、第 1 の実施の形態の文書検索部 121 と異なっている。

【0172】

第 2 パッセージ重要度算出部 324 は、単語一般重要度ベクトル A、単語個別重要度ベクトル B、および単語既知度ベクトル C に加え、質問ベクトル D を用いて、サマリ内重要度評価用ベクトル Q を算出する点が、第 1 の実施の形態の第 2 パッセージ重要度算出部 124 と異なっている。

【0173】

次に、このように構成された第 3 の実施の形態にかかる検索支援装置 300 による検索支援処理について図 20 を用いて説明する。図 20 は、第 3 の実施の形態における検索支援処理の全体の流れを示すフローチャートである。

【0174】

第 3 の実施の形態では、ステップ S2001 の受付処理、ステップ S2006 およびステップ S2007 が追加されたこと、文書検索処理（ステップ S2008）、および、サマリ内重要度算出処理（ステップ S2011）が第 1 の実施の形態と異なっている。その他のステップは、第 1 の実施の形態にかかる検索支援装置 100 における検索支援処理（図 3）と同様の処理なので、その説明を省略する。

【0175】

ステップ S2001 では、受付部 311 が、ユーザサマリおよびカーソル位置とともに、指定形態素および品詞情報の入力を受付ける（ステップ S2001）。また、ステップ S2006 では、係り受け解析部 317 が、入力されたユーザサマリの指定形態素を含む文の句構造解析（係り受け解析）を実行する（ステップ S2006）。その後、生成部 318 が、句構造解析結果を元に質問ベクトル D を生成する質問生成処理を実行する（ステップ S2007）。質問生成処理の詳細については後述する。

【0176】

なお、係り受け解析処理（ステップ S2006）および質問生成処理（ステップ S2007）は、ステップ S2003～ステップ S2005 の各処理の後に実行する必要はない。ステップ S2003～ステップ S2005 の各処理の前に実行してもよいし、並列的に実行するように構成してもよい。

【0177】

ステップ S2008 の文書検索処理、および、ステップ S2011 のサマリ内重要度算出処理の詳細は後述する。

【0178】

次に、ステップ S2007 の質問生成処理の詳細について図 21 を用いて説明する。図 21 は、第 3 の実施の形態における質問生成処理の全体の流れを示すフローチャートである。

【0179】

まず、生成部 318 は、スコア X を 1.0 に初期化する（ステップ S2101）。次に、生成部 318 は、句構造解析結果を参照して、指定形態素が句構造のいずれのノードに

10

20

30

40

50

含まれるかを照合し、照合したノードを現在ノードとして選択する（ステップS 2 1 0 2）。

【0 1 8 0】

次に、生成部3 1 8は、現在ノード内の内容語を抽出し、内容語に対する重み QW_i としてスコア X を付与する（ステップS 2 1 0 3）。なお、 i は処理ごとに1加算されるカウンタ値（0以上）とする。

【0 1 8 1】

次に、生成部3 1 8は、スコア X の値を、現在の値の9割の値に更新する（ステップS 2 1 0 4）。次に、生成部3 1 8は、現在ノードの子ノード群を取得する（ステップS 2 1 0 5）。そして、生成部3 1 8は、子ノードが取得できたか否かを判断する（ステップS 2 1 0 6）。取得できた場合（ステップS 2 1 0 6：YES）、生成部3 1 8は、子ノードを現在ノードとして選択し（ステップS 2 1 0 7）、処理を繰り返す（ステップS 2 1 0 3）。なお、子ノードが複数取得された場合は、各子ノードに対して処理を繰り返す。

【0 1 8 2】

子ノードが取得できなかった場合（ステップS 2 1 0 6：NO）、生成部3 1 8は、指定形態素の品詞に対応する修辭語を、予め定められた修辭語辞書（図示せず）等を参照して取得する（ステップS 2 1 0 8）。指定形態素の品詞は、入力された品詞情報から判別することができる。

【0 1 8 3】

例えば、生成部3 1 8は、指定形態素が動作性名詞または動詞の場合、理由に相当する修辭語として予め定められた修辭語群を修辭語辞書から取得する。また、生成部3 1 8は、指定形態素が動作性名詞以外の名詞の場合は、定義に相当する修辭語として予め定められた修辭語群を修辭語辞書から取得する。なお、修辭語を、「^その結果・*」および「^背景として・*」のように正規表現で記述するように構成してもよい。

【0 1 8 4】

なお、生成部3 1 8は、修辭語に含まれる形態素のスコアとして「1.0」を付与する。また、修辭語に含まれる形態素と抽出された内容語とが一致する場合は、当該形態素のスコアを「1.0」で更新する。

【0 1 8 5】

次に、生成部3 1 8は、抽出した各内容語および修辭語のスコア（重み QW_i ）を要素とする質問ベクトル D を生成する（ステップS 2 1 0 9）。これにより、修辭語を検索キーワードとして含む検索要求（文書検索用ベクトル）を生成可能となる。

【0 1 8 6】

例えば、「NYMEXの時間外取引で原油先物価格が1バレル=119ドル台に上昇」が入力された上記例では、「原油先物価格」、「上昇」、および「NYMEX」が内容語として抽出され、それぞれの重みとして「1.0」、「0.9」、および「0.81」が与えられる。さらに、修辭語群として、「^その結果」および「^背景として・*」が取得されたとする。この場合、生成部3 1 8は、これらの内容語および修辭語の各重みを要素とする質問ベクトル D を生成する。

【0 1 8 7】

次に、ステップS 2 0 0 8の文書検索処理の詳細について図22を用いて説明する。図22は、第3の実施の形態における文書検索処理の全体の流れを示すフローチャートである。

【0 1 8 8】

まず、文書検索部3 2 1は、単語一般重要度ベクトル A 、単語個別重要度ベクトル B 、および質問ベクトル D の重み付け線形和である文書検索用ベクトル Q を、以下の(13)式を用いて算出する（ステップS 2 2 0 1）。

$$Q = W_{31} \times A + W_{32} \times B + W_{33} \times D \quad (W_{31}, W_{32}, W_{33} \text{ は正数}) \quad \cdots (13)$$

10

20

30

40

50

【 0 1 8 9 】

次に、文書検索部 3 2 1 は、文書記憶部 1 4 0 内に格納されている各文書の文書ベクトルと文書検索用ベクトル Q とのコサイン尺度を関連度として算出し（ステップ S 2 2 0 2 ）、文書検索処理を終了する。

【 0 1 9 0 】

次に、ステップ S 2 0 1 1 のサマリ内重要度算出処理の詳細について図 2 3 を用いて説明する。図 2 3 は、第 3 の実施の形態におけるサマリ内重要度算出処理の全体の流れを示すフローチャートである。

【 0 1 9 1 】

まず、第 2 パッセージ重要度算出部 3 2 4 は、質問ベクトル D に含まれる正規表現で記述された修辭パタン（修辭語）に適合するパッセージ群のみを取り出す（ステップ S 2 3 0 1 ）。

【 0 1 9 2 】

次に、第 2 パッセージ重要度算出部 3 2 4 は、単語一般重要度ベクトル A、単語個別重要度ベクトル B、単語既知度ベクトル C、および質問ベクトル D を用いて、サマリ内重要度評価用ベクトル Q を以下の（ 1 4 ）式により算出する（ステップ S 2 3 0 2 ）。

$$Q = W 4 1 \times A + W 4 2 \times B - W 4 3 \times C + W 4 4 \times D \quad \cdots (14)$$

【 0 1 9 3 】

なお、W 4 1、W 4 2、W 4 3、および W 4 4 は予め定められた正数を表す。結果として得られるサマリ内重要度評価用ベクトル Q は、重み付きの単語ベクトルとなる。

【 0 1 9 4 】

ステップ S 2 3 0 3 からステップ S 2 3 0 5 までの、コサイン尺度算出処理は、第 1 の実施の形態にかかる検索支援装置 1 0 0 におけるステップ S 1 1 0 2 からステップ S 1 1 0 4 までと同様の処理なので、その説明を省略する。

【 0 1 9 5 】

このように、第 3 の実施の形態にかかる検索支援装置では、指定された形態素の品詞に応じて予め定められた修辭語を検索要求に付加することができる。これにより、形態素の品詞に応じてより適切なパッセージ候補を検索することが可能となる。

【 0 1 9 6 】

（第 4 の実施の形態）

第 4 の実施の形態にかかる検索支援装置は、ユーザがサマリを編集した履歴を表す編集履歴情報を用いることにより、より高精度に関連するパッセージを検索する。

【 0 1 9 7 】

図 2 4 は、第 4 の実施の形態にかかる検索支援装置 4 0 0 の構成の一例を示すブロック図である。図 2 4 に示すように、検索支援装置 4 0 0 は、文脈解析部 4 1 0 と、パッセージ検索部 4 2 0 と、編集部 4 3 0 と、文書記憶部 1 4 0 と、を備えている。

【 0 1 9 8 】

第 4 の実施の形態では、文脈解析部 4 1 0 の第 2 単語重要度算出部 4 1 4 の機能、パッセージ検索部 4 2 0 の第 2 パッセージ重要度算出部 4 2 4 の機能、および、編集部 4 3 0 に履歴記憶部 4 3 4 を追加したことが第 1 の実施の形態と異なっている。その他の構成および機能は、第 1 の実施の形態にかかる検索支援装置 1 0 0 の構成を表すブロック図である図 2 と同様であるので、同一符号を付し、ここでの説明は省略する。

【 0 1 9 9 】

履歴記憶部 4 3 4 は、ユーザによるユーザサマリの編集履歴を記憶する。具体的には、履歴記憶部 4 3 4 は、ユーザサマリのテキストのうち、ユーザが独自に記述した部分と、外部からコピーペースト（引用）した部分とを判別する判定情報、および、コピーペーストした部分については、いずれの文書から引用したかを特定可能な引用元情報（例えば URL など）を記憶する。

【 0 2 0 0 】

編集部 4 3 0 は、この引用元情報を用いて、引用元の原文書を表示する機能を備える。

10

20

30

40

50

具体的には、ユーザが任意のテキスト上でダブルクリックなど所定操作を行い、そのテキストが引用により追加されたテキストであった場合、編集部 430 は、引用元情報を参照して当該テキストの引用元となる原文書を取得し、編集画面に表示する。

【0201】

第2単語重要度算出部 414 は、構造上位パッセージ群に含まれる各形態素（内容語）のうち、ユーザが独自に記述した形態素（すなわち引用された記述ではない部分）については、重み（個別重要度）を高める処置をする点が、第1の実施の形態の第2単語重要度算出部 114 と異なっている。ユーザ自身が記述した部分は、ユーザの意思が強く現れていることを織り込むためである。

【0202】

第2パッセージ重要度算出部 424 は、ユーザサマリ内のカーソル直前のパッセージと同じ文書に所属するパッセージ（すなわち引用元が同じパッセージ）については、スコア（サマリ内重要度）を高める処置をする点が、第1の実施の形態の第2パッセージ重要度算出部 124 と異なっている。パッセージが同じ文書からの引用であれば、話題の結束性（つながり）がより高いと期待されるためである。なお、カーソル直前のパッセージとは、カーソルが所属するパッセージの1つ上位のパッセージを意味する。

【0203】

次に、このように構成された第4の実施の形態にかかる検索支援装置 400 による検索支援処理について説明する。本実施の形態では、個別重要度算出処理およびサマリ内重要度算出処理の内容が第1の実施の形態と異なっている。その他の処理は、第1の実施の形態にかかる検索支援装置 100 の検索支援処理を表す図3と同様であるため、その説明を省略する。

【0204】

次に、本実施の形態の個別重要度算出処理の詳細について図25を用いて説明する。図25は、第4の実施の形態における個別重要度算出処理の全体の流れを示すフローチャートである。

【0205】

本実施の形態では、ステップ S2502 で算出する個別重要度 IW_j の算出式が第1の実施の形態と異なっている。その他のステップは、第1の実施の形態にかかる検索支援装置 100 における個別重要度算出処理（図6）と同様の処理なので、その説明を省略する。

【0206】

ステップ S2502 で、第2単語重要度算出部 414 は、形態素集合内の各内容語について、以下の(15)式により個別重要度 IW_j を算出する（ステップ S602）。

【数3】

$$IW_j = \sum_{t \in T_j} \left(\log \left(\frac{\text{org}(t)}{\text{dist}(t)} \right) + 1 \right) \quad \cdots (15)$$

【0207】

ただし、 j は $0 \leq j < m$ を満たす整数（ m は内容語の種類数）、 T_j は形態素集合のうち内容語 j である形態素を集めたサブ集合、 $\text{dist}(t)$ は形態素 t が出現したパッセージと、カーソルが存在するパッセージとの階層的距離を表す。また、 $\text{org}(t)$ は、形態素 t がユーザ自身によって記述された形態素である場合に $WORG1$ を返し、それ以外の場合に $WORG2$ を返す関数を表す（ $WORG1$ および $WORG2$ は正数、かつ、 $WORG1 > WORG2$ ）。このような関数 $\text{org}(t)$ を用いることにより、第2単語重要度算出部 414 は、ユーザが独自に記述した形態素の個別重要度を高めることができる。

【0208】

次に、本実施の形態のサマリ内重要度算出処理の詳細について図26を用いて説明する。図26は、第4の実施の形態におけるサマリ内重要度算出処理の全体の流れを示すフロ

10

20

30

40

50

ーチャートである。

【0209】

本実施の形態では、ステップS2604およびステップS2605が追加されたことが第1の実施の形態と異なっている。その他のステップは、第1の実施の形態にかかる検索支援装置100におけるサマリ内重要度算出処理(図11)と同様の処理なので、その説明を省略する。

【0210】

ステップS2603でコサイン尺度(サマリ内重要度)を算出後、第2パッセージ重要度算出部424は、パッセージiが、ユーザサマリ内のカーソル直前のパッセージと同じ文書から引用されたパッセージであるか否かを判断する(ステップS2604)。同じ文書から引用されたパッセージであった場合(ステップS2604: YES)、第2パッセージ重要度算出部424は、算出したサマリ内重要度に予め定められた正数W44を加算する(ステップS2605)。正数W44を加算後、または、パッセージiがユーザサマリ内のカーソル直前のパッセージと同じ文書から引用されたパッセージでないと判断した場合(ステップS2604: NO)、第2パッセージ重要度算出部424は、未処理のパッセージが存在するか否かを判断する(ステップS2606)。この後の処理は、図11と同様である。

【0211】

このように、第4の実施の形態にかかる検索支援装置では、ユーザがサマリを編集した履歴を表す編集履歴情報を用いることにより、ユーザの意思が強く現れている内容語に関連するパッセージを高精度に検索できる。また、パッセージの引用関係を参照することにより、話題が関連するパッセージを高精度に検索できる。これにより、ユーザの要求を適切に反映した関連パッセージをより高精度に検索することができる。

【0212】

次に、第1～第4の実施の形態にかかる検索支援装置のハードウェア構成について図27を用いて説明する。図27は、第1～第4の実施の形態にかかる検索支援装置のハードウェア構成を示す説明図である。

【0213】

第1～第4の実施の形態にかかる検索支援装置は、CPU(Central Processing Unit)51などの制御装置と、ROM(Read Only Memory)52やRAM53などの記憶装置と、ネットワークに接続して通信を行う通信I/F54と、HDD(Hard Disk Drive)、CD(Compact Disc)ドライブ装置などの外部記憶装置と、ディスプレイ装置などの表示装置と、キーボードやマウスなどの入力装置と、各部を接続するバス61を備えており、通常のコンピュータを利用したハードウェア構成となっている。

【0214】

第1～第4の実施の形態にかかる検索支援装置で実行される検索支援プログラムは、インストール可能な形式又は実行可能な形式のファイルでCD-ROM(Compact Disk Read Only Memory)、フレキシブルディスク(FD)、CD-R(Compact Disk Recordable)、DVD(Digital Versatile Disk)等のコンピュータで読み取り可能な記録媒体に記録されて提供される。

【0215】

また、第1～第4の実施の形態にかかる検索支援装置で実行される検索支援プログラムを、インターネット等のネットワークに接続されたコンピュータ上に格納し、ネットワーク経由でダウンロードさせることにより提供するように構成してもよい。また、第1～第4の実施の形態にかかる検索支援装置で実行される検索支援プログラムをインターネット等のネットワーク経由で提供または配布するように構成してもよい。

【0216】

また、第1～第4の実施の形態の検索支援プログラムを、ROM等に予め組み込んで提供するように構成してもよい。

【0217】

第１～第４の実施の形態にかかる検索支援装置で実行される検索支援プログラムは、上述した各部（文脈解析部等）を含むモジュール構成となっており、実際のハードウェアとしてはＣＰＵ５１（プロセッサ）が上記記憶媒体から検索支援プログラムを読み出して実行することにより上記各部が主記憶装置上にロードされ、上述した各部が主記憶装置上に生成されるようになっている。

【０２１８】

なお、本発明は、上記実施の形態そのままに限定されるものではなく、実施段階ではその要旨を逸脱しない範囲で構成要素を変形して具体化することができる。また、上記実施の形態に開示されている複数の構成要素の適宜な組み合わせにより、種々の発明を形成することができる。例えば、実施の形態に示される全構成要素からいくつかの構成要素を削除してもよい。さらに、異なる実施の形態にわたる構成要素を適宜組み合わせてもよい。

【０２１９】

本発明の他の第１の発明は、請求項２に記載の発明において、前記位置の文字を含む前記パッセージと前記階層構造が同位の前記パッセージに含まれる単語それぞれについて、前記入力文書内でのユーザが既に知っている度合いを表す既知度を算出する既知度算出部をさらに備え、前記第２パッセージ重要度算出部は、前記既知度が大きい前記単語を含む前記パッセージに対して、前記既知度が小さい前記単語を含む前記パッセージより小さい前記第２パッセージ重要度を算出すること、を特徴とする。

【０２２０】

本発明の他の第２の発明は、請求項２に記載の発明において、前記第２パッセージ重要度算出部は、前記位置の文字を含む前記パッセージの１つ上位の前記上位パッセージである直前パッセージが含まれる文書に属する前記パッセージに対して、前記直前パッセージが含まれる文書に属さない前記パッセージより大きい前記第２パッセージ重要度を算出すること、を特徴とする。

【０２２１】

本発明の他の第３の発明は、請求項２に記載の発明において、前記分割部は、さらに、分割した前記パッセージに、他の前記パッセージに含まれる対象を指し示す照応表現が含まれるか否かを判断し、前記照応表現が含まれる場合に、他の前記パッセージから前記照応表現の指示対象を表す名詞句を取得し、前記名詞句で前記照応表現を置換すること、を特徴とする。

【０２２２】

本発明の他の第４の発明は、請求項２に記載の発明において、前記分割部は、さらに、分割した前記パッセージに、主語が省略された文が含まれるか否かを判断し、前記主語が省略された文が含まれる場合に、他の前記パッセージから前記主語を表す名詞句を取得し、前記主語が省略された文に前記名詞句を主語として追加すること、を特徴とする。

【０２２３】

本発明の他の第５の発明は、請求項２に記載の発明において、前記パッセージ選択部は、分割された前記パッセージのうち、前記総合重要度が大きく、かつ、選択済みの前記パッセージとの間の類似度が小さい前記パッセージを優先して選択すること、を特徴とする。

【符号の説明】

【０２２４】

５１ ＣＰＵ

５２ ＲＯＭ

５３ ＲＡＭ

５４ 通信Ｉ／Ｆ

６１ バス

１００、２００、３００、４００ 検索支援装置

１１０、３１０、４１０ 文脈解析部

１１１、３１１ 受付部

10

20

30

40

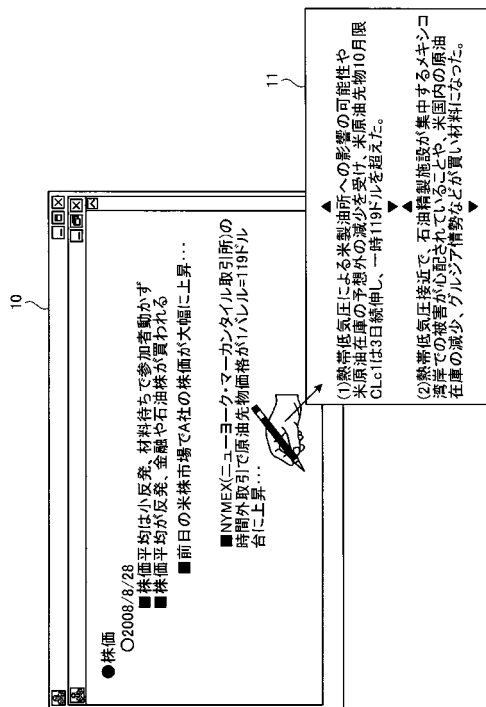
50

- 1 1 2 構造解析部
- 1 1 3 第1単語重要度算出部
- 1 1 4、4 1 4 第2単語重要度算出部
- 1 1 5 既知度算出部
- 1 1 6 出力部
- 1 2 0、3 2 0、4 2 0 パッセージ検索部
- 1 2 1、3 2 1 文書検索部
- 1 2 2 パッセージ分割部
- 1 2 3 第1パッセージ重要度算出部
- 1 2 4、3 2 4、4 2 4 第2パッセージ重要度算出部
- 1 2 5 総合重要度算出部
- 1 2 6 パッセージ選択部
- 1 3 0、3 3 0、4 3 0 編集部
- 1 3 1、2 3 1 候補選択部
- 1 3 2、2 3 2 多段階検索部
- 1 3 3 編集結果記憶部
- 1 4 0 文書記憶部
- 2 3 0 閲覧部
- 2 3 3 閲覧結果記憶部
- 3 1 7 係り受け解析部
- 3 1 8 生成部
- 4 3 4 履歴記憶部

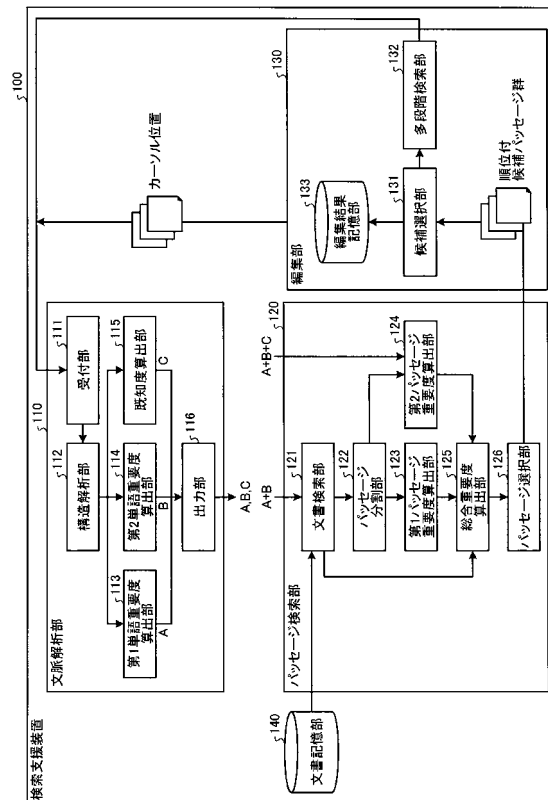
10

20

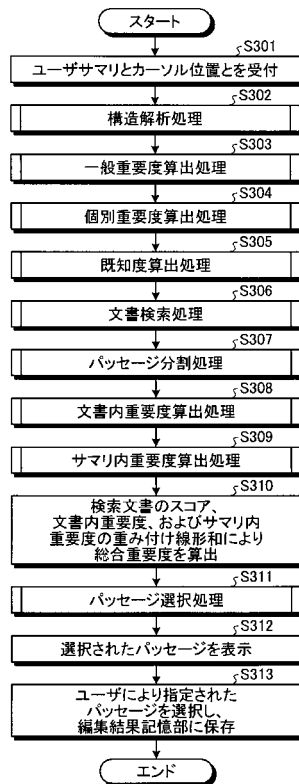
【図1】



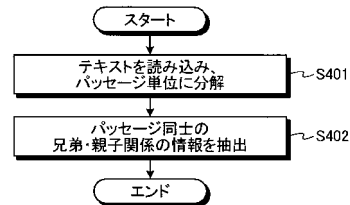
【図2】



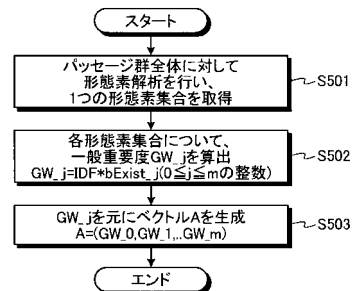
【図 3】



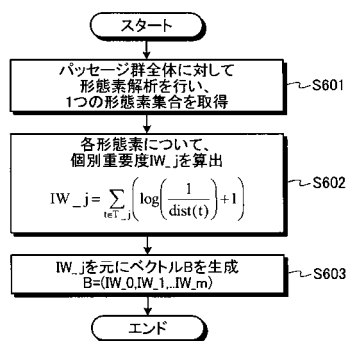
【図 4】



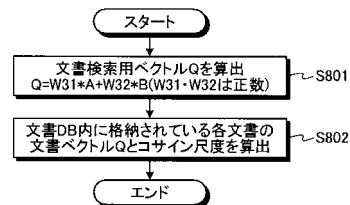
【図 5】



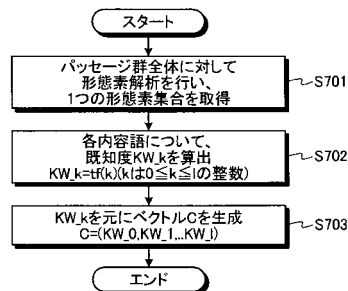
【図 6】



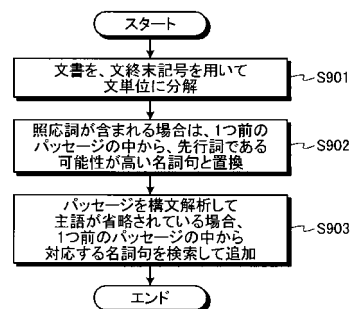
【図 8】



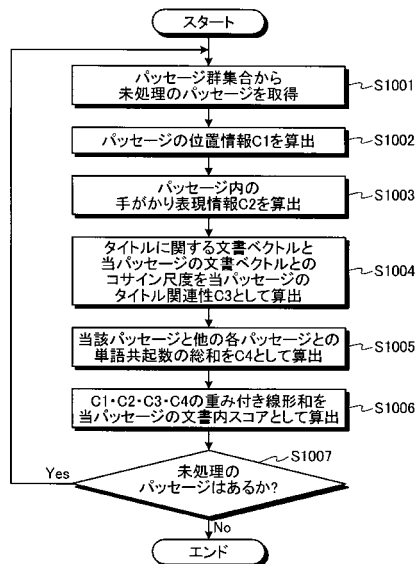
【図 7】



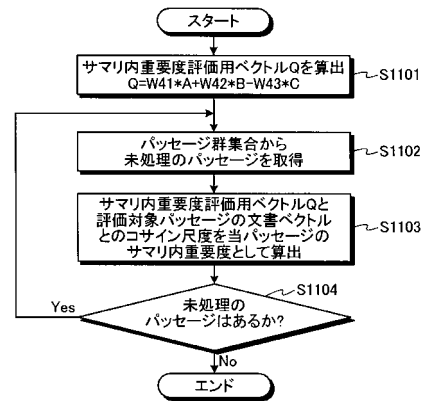
【図 9】



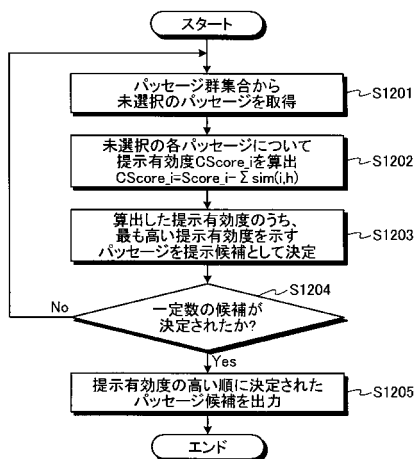
【図 10】



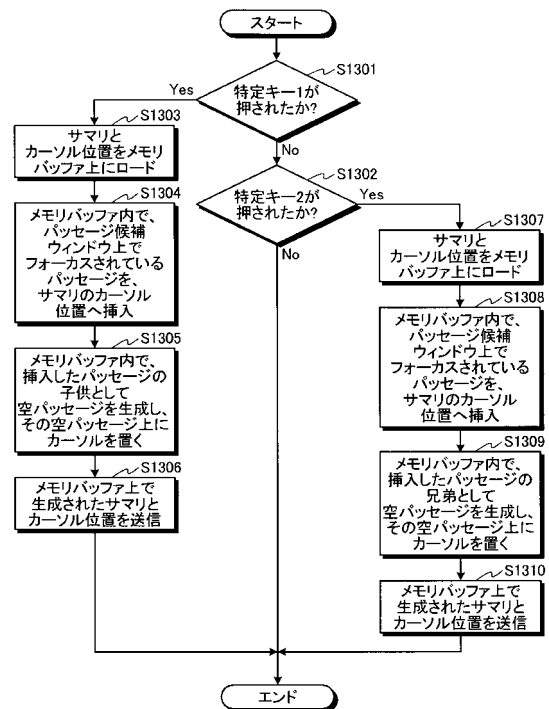
【図 11】



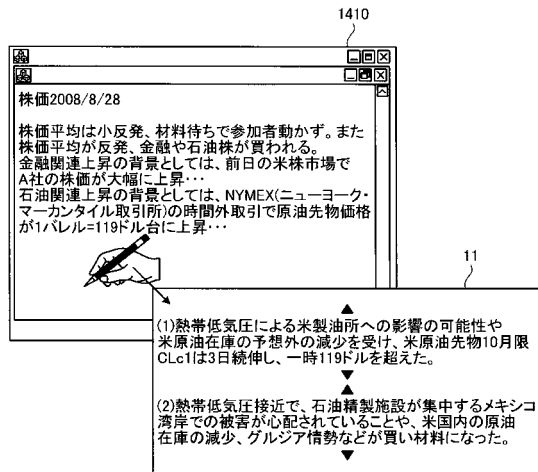
【図 12】



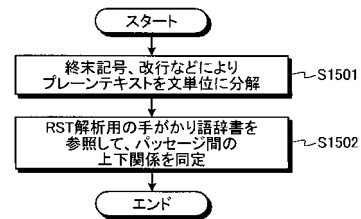
【図 13】



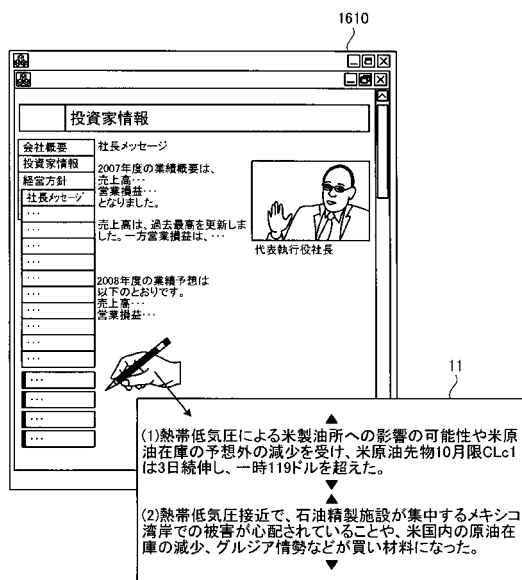
【図 14】



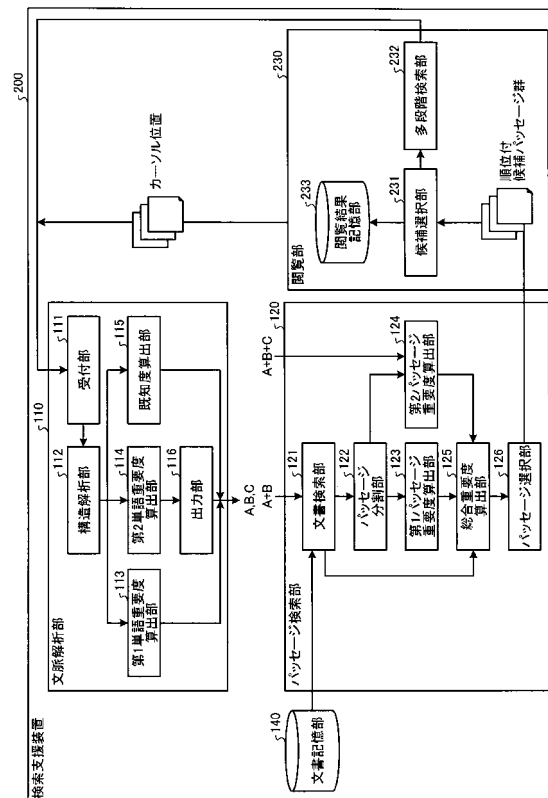
【図 15】



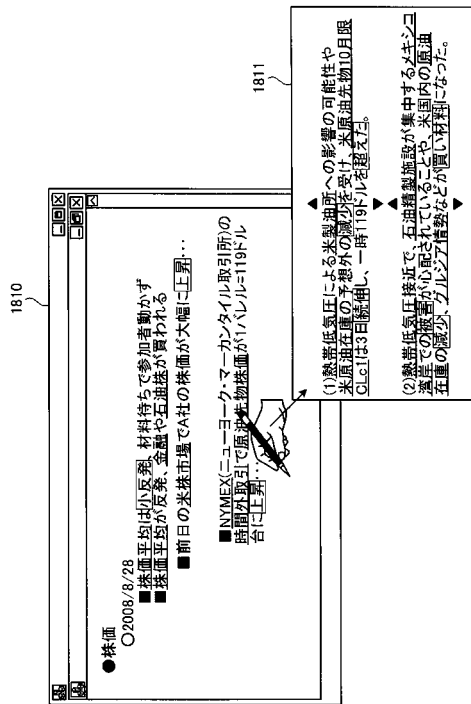
【図 16】



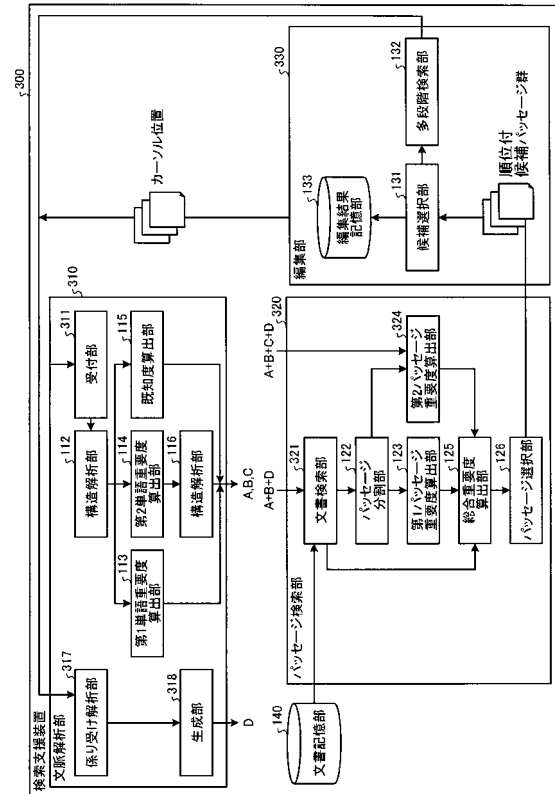
【図 17】



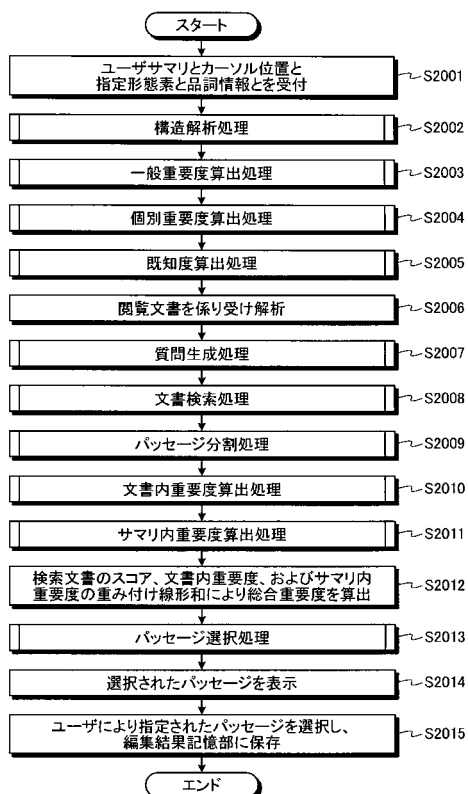
【 図 1 8 】



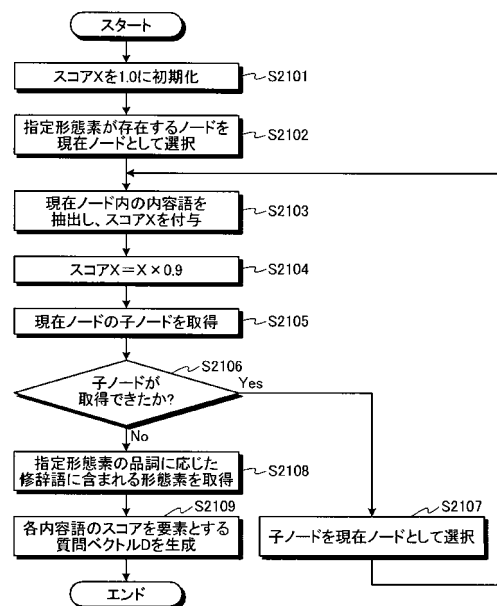
【 図 1 9 】



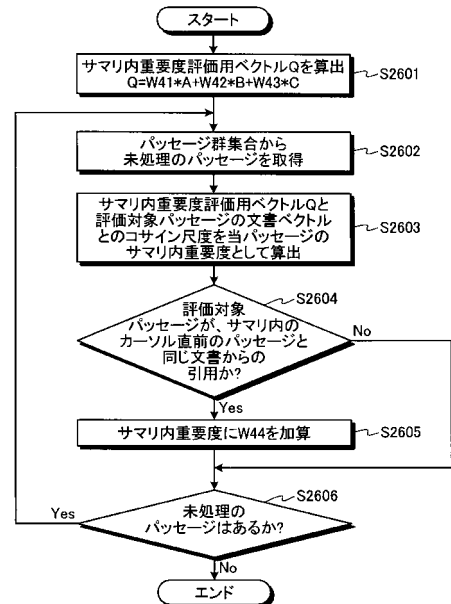
【 図 2 0 】



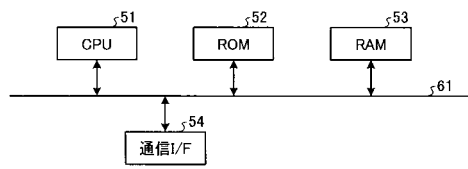
【 図 2 1 】



【 図 2 2 】



【図 27】



フロントページの続き

(56)参考文献 特開平10-207911(JP,A)

特開平11-045255(JP,A)

特開平11-102374(JP,A)

特開平03-040067(JP,A)

特開平11-045252(JP,A)

特開2002-358315(JP,A)

麻野間 直樹 外2名, “文書構造と言語表現の分析に基づく方法説明抽出”, 言語処理学会第12回年次大会発表論文, 日本, 言語処理学会, 2006年 3月13日, p. 324-327

品川 徳秀 外1名, “内容解析に基づく文書構造の自動抽出”, 情報処理学会研究報告, 日本, 社団法人情報処理学会, 1998年 7月10日, 第98巻, 第58号, p. 157-164

(58)調査した分野(Int.Cl., DB名)

G06F 17/30

G06F 17/27

CSD B(日本国特許庁)