

(19) 日本国特許庁(JP)

(12) 特許公報(B2)

(11) 特許番号

特許第4336404号
(P4336404)

(45) 発行日 平成21年9月30日(2009.9.30)

(24) 登録日 平成21年7月3日(2009.7.3)

(51) Int.Cl.

G06F 17/30 (2006.01)

F 1

G06F 17/30 350C
G06F 17/30 330C
G06F 17/30 170A

請求項の数 7 (全 26 頁)

(21) 出願番号 特願平10-355039
 (22) 出願日 平成10年12月14日(1998.12.14)
 (65) 公開番号 特開2000-181923(P2000-181923A)
 (43) 公開日 平成12年6月30日(2000.6.30)
 審査請求日 平成17年12月5日(2005.12.5)

(73) 特許権者 000001007
 キヤノン株式会社
 東京都大田区下丸子3丁目30番2号
 (74) 代理人 100076428
 弁理士 大塚 康徳
 (74) 代理人 100112508
 弁理士 高柳 司郎
 (74) 代理人 100115071
 弁理士 大塚 康弘
 (74) 代理人 100116894
 弁理士 木村 秀二
 (72) 発明者 池田 和世
 東京都大田区下丸子3丁目30番2号 キ
 ャノン株式会社内

最終頁に続く

(54) 【発明の名称】情報検索装置及びその方法、コンピュータ可読メモリ

(57) 【特許請求の範囲】

【請求項1】

複数種類の文を管理するデータベースから検索条件に合致する文を検索する情報検索装置であって、

複数の単語から構成される検索条件を入力する入力手段と、

前記検索条件に対して形態素解析を行い、解析結果として、単語とそのカテゴリとが定義されている形態素リストを生成する形態素解析手段と、

前記検索条件に対して構文解析を行い、解析結果として、前記検索条件中の2つの注目単語の左側の単語と右側の単語と、それらのそれぞれのカテゴリから決定される前記2つの注目単語の間の関係の種類を示すロールとが定義されているロールリストを生成する構文解析手段と、

前記形態素リストで定義されているカテゴリの優先順位から、前記検索条件を構成する各単語の内、中心となる単語である中心語の候補を絞り込み、複数の中心語の候補がある場合は、更に、前記ロールリストを用いて、各候補と他の単語の関係によって、中心語の候補を絞り込み、最終的に存在する中心語の候補を、前記検索条件を構成する各単語の中心語として決定し、その中心語から前記ロールリスト中で関係がある単語をたどって、各単語に順位を付与することで、前記中心語をルートとする順位のツリー構造である順位ツリーを生成し、その順位ツリーを構成する各単語の順位に応じて、前記検索条件を構成する各単語の重要度を決定する決定手段と、

前記決定手段によって決定された重要度に応じて、検索対象文と前記検索条件との一致

度を算出する算出手段と、

前記算出手段で算出された一致度に基づいて、検索結果とする検索対象文を出力する出力手段と

を備えることを特徴とした情報検索装置。

【請求項 2】

前記カテゴリは、品詞である

ことを特徴とする請求項 1 に記載の情報検索装置。

【請求項 3】

前記算出手段は、前記決定手段によって決定された重要度に応じて、前記検索対象文を構成する各単語及び単語間の関係と、前記検索条件を構成する各単語及び単語間の関係との一致度を算出する

10

ことを特徴とする請求項 1 に記載の情報検索装置。

【請求項 4】

前記算出手段は、前記検索対象文を構成する各単語の重要度を決定する決定手段と

を備えることを特徴とする請求項 1 に記載の情報検索装置。

【請求項 5】

前記出力手段は、前記決定手段で決定された重要度の内、少なくとも最高の重要度の単語を含む検索対象文を検索結果として出力する

ことを特徴とする請求項 1 に記載の情報検索装置。

【請求項 6】

20

複数種類の文を管理するデータベースから検索条件に合致する文を検索する情報検索装置における情報検索方法であって、

前記情報検索装置の入力手段が、複数の単語から構成される検索条件を入力する入力工程と、

前記情報検索装置の形態素解析手段が、前記検索条件に対して形態素解析を行い、解析結果として、単語とそのカテゴリとが定義されている形態素リストを生成する形態素解析工程と、

前記情報検索装置の構文解析手段が、前記検索条件に対して構文解析を行い、解析結果として、前記検索条件中の 2 つの注目単語の左側の単語と右側の単語と、それらのそれぞれのカテゴリから決定される前記 2 つの注目単語の間の関係の種類を示すロールとが定義されているロールリストを生成する構文解析工程と、

30

前記情報検索装置の決定手段が、前記形態素リストで定義されているカテゴリの優先順位から、前記検索条件を構成する各単語の内、中心となる単語である中心語の候補を絞り込み、複数の中心語の候補がある場合は、更に、前記ロールリストを用いて、各候補と他の単語の関係によって、中心語の候補を絞り込み、最終的に存在する中心語の候補を、前記検索条件を構成する各単語の中心語として決定し、その中心語から前記ロールリスト中で関係がある単語をたどって、各単語に順位を付与することで、前記中心語をルートとする順位のツリー構造である順位ツリーを生成し、その順位ツリーを構成する各単語の順位に応じて、前記検索条件を構成する各単語の重要度を決定する決定工程と、

前記情報検索装置の算出手段が、前記決定工程によって決定された重要度に応じて、検索対象文と前記検索条件との一致度を算出する算出工程と、

40

前記情報検索装置の出力手段が、前記算出工程で算出された一致度に基づいて、検索結果とする検索対象文を出力する出力工程と

を備えることを特徴とした情報検索方法。

【請求項 7】

複数種類の文を管理するデータベースから検索条件に合致する文を検索する情報検索をコンピュータに機能させるためのプログラムが格納されたコンピュータ可読メモリであつて、

前記コンピュータを、

複数の単語から構成される検索条件を入力する入力手段と、

50

前記検索条件に対して形態素解析を行い、解析結果として、単語とそのカテゴリとが定義されている形態素リストを生成する形態素解析手段と、

前記検索条件に対して構文解析を行い、解析結果として、前記検索条件中の2つの注目単語の左側の単語と右側の単語と、それらのそれぞれのカテゴリから決定される前記2つの注目単語の間の関係の種類を示すロールとが定義されているロールリストを生成する構文解析手段と、

前記形態素リストで定義されているカテゴリの優先順位から、前記検索条件を構成する各単語の内、中心となる単語である中心語の候補を絞り込み、複数の中心語の候補がある場合は、更に、前記ロールリストを用いて、各候補と他の単語の関係によって、中心語の候補を絞り込み、最終的に存在する中心語の候補を、前記検索条件を構成する各単語の中心語として決定し、その中心語から前記ロールリスト中で関係がある単語をたどって、各単語に順位を付与することで、前記中心語をルートとする順位のツリー構造である順位ツリーを生成し、その順位ツリーを構成する各単語の順位に応じて、前記検索条件を構成する各単語の重要度を決定する決定手段と、

前記決定手段によって決定された重要度に応じて、検索対象文と前記検索条件との一致度を算出する算出手段と、

前記算出手段で算出された一致度に基づいて、検索結果とする検索対象文を出力する出力手段と

として機能させるためのプログラムが格納されたコンピュータ可読メモリ。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】

本発明は、複数種類の情報を管理し、その管理される情報より所望の情報を検索する情報検索装置及びその方法、コンピュータ可読メモリに関するものである。

【0002】

【従来の技術】

従来、複数種類の情報を管理する情報検索装置における検索手法として、検索の対象となる個々の情報に対してキーワードを付与し、キーワードを組み合わせたクエリを発行することによって検索するキーワード検索や、検索対象がテキストである場合は、キーワードを付与することなく、クエリで指定した単語の組み合わせを直接テキスト中の単語と比較することによって行う全文検索などが行われている。

【0003】

検索対象の情報が複数存在する場合、このような検索手法では、検索条件を満たす情報が大量に存在することがあり、検索結果の全てに目を通すことが困難になる。このような問題を解決する方法として、キーワード検索の場合は、検索対象の情報に付与するキーワードに重要度を反映した重みを付与し、その重みを利用して検索結果に順位をつけて出力する方法などが採用されている。また、全文検索などでは、テキスト中の個々の単語に対して、そのテキストにおける重要度を反映した重みを自動的に算出して、その重みを利用して検索結果に順位をつけて出力する方法などが採用されている。

【0004】

これらのキーワード検索や全文検索などでは、検索条件が適切でない場合や、検索条件を構成する単語の曖昧性などから、検索意図と合致する情報が存在するにも関わらず、検索結果の中にその情報が含まれていない、いわゆる、検索漏れが発生する場合がある。このような検索漏れの対処方法としては、検索要求の中の条件と部分一致する情報も含めて検索結果として出力する方法が知られている。また、検索対象の情報が大量に存在する場合、検索条件と部分一致する情報も大量に存在することになり、検索結果に順位をつけて出力する必要が生じてくる。そのため、このような場合は、検索対象中の単語に重要度を付与し、その重要度に基づいて検索結果の出力順を決定していた。

【0005】

【発明が解決しようとする課題】

10

20

30

40

50

しかしながら、上記従来の情報検索装置の検索対象中の単語に重要度を付与して、その重要度に基づいて検索結果の出力順位を決定する方法において、検索条件と部分一致した情報を検索結果としてを出力する場合には、その検索結果につけられる出力順位は必ずしも適切でなく、適切な検索結果が得られないといふ問題点があった。

【0006】

例えば、検索意図が「白い犬」である場合、検索条件として「白い」かつ「犬」ととらえることができるが、この条件と部分一致する情報として、「犬」の条件のみを満たした情報の方が、「白い」の条件のみを満たした情報よりも、検索意図に近いといえるが、従来の方法では、この観点から検索結果に順位をつけることができなかった。

【0007】

本発明は上記の問題点に鑑みてなされたものであり、検索条件を満たす情報を適切に検索することができる情報検索装置及びその方法、コンピュータ可読メモリを提供することを目的とする。

【0008】

【課題を解決するための手段】

上記の目的を達成するための本発明による情報検索装置は以下の構成を備える。即ち、複数種類の文を管理するデータベースから検索条件に合致する文を検索する情報検索装置であって、

複数の単語から構成される検索条件を入力する入力手段と、

前記検索条件に対して形態素解析を行い、解析結果として、単語とそのカテゴリとが定義されている形態素リストを生成する形態素解析手段と、

前記検索条件に対して構文解析を行い、解析結果として、前記検索条件中の2つの注目単語の左側の単語と右側の単語と、それらのそれぞれのカテゴリから決定される前記2つの注目単語の間の関係の種類を示すロールとが定義されているロールリストを生成する構文解析手段と、

前記形態素リストで定義されているカテゴリの優先順位から、前記検索条件を構成する各単語の内、中心となる単語である中心語の候補を絞り込み、複数の中心語の候補がある場合は、更に、前記ロールリストを用いて、各候補と他の単語の関係によって、中心語の候補を絞り込み、最終的に存在する中心語の候補を、前記検索条件を構成する各単語の中心語として決定し、その中心語から前記ロールリスト中で関係がある単語をたどって、各単語に順位を付与することで、前記中心語をルートとする順位のツリー構造である順位ツリーを生成し、その順位ツリーを構成する各単語の順位に応じて、前記検索条件を構成する各単語の重要度を決定する決定手段と、

前記決定手段によって決定された重要度に応じて、検索対象文と前記検索条件との一致度を算出する算出手段と、

前記算出手段で算出された一致度に基づいて、検索結果とする検索対象文を出力する出力手段と

を備える。

【0009】

また、好ましくは、前記カテゴリは、品詞である。

【0011】

また、好ましくは、前記算出手段は、前記決定手段によって決定された重要度に応じて、前記検索対象文を構成する各単語及び単語間の関係と、前記検索条件を構成する各単語及び単語間の関係との一致度を算出する。

【0012】

また、好ましくは、前記算出手段は、前記算出手段は、前記検索対象文を構成する各単語の重要度を決定する決定手段と

を備える。

【0013】

また、好ましくは、前記出力手段は、前記決定手段で決定された重要度の内、少なくと

10

20

30

40

50

も最高の重要度の単語を含む検索対象文を検索結果として出力する。

【0014】

上記の目的を達成するための本発明による情報検索方法は以下の構成を備える。即ち、複数種類の文を管理するデータベースから検索条件に合致する文を検索する情報検索装置における情報検索方法であって、

前記情報検索装置の入力手段が、複数の単語から構成される検索条件を入力する入力工程と、

前記情報検索装置の形態素解析手段が、前記検索条件に対して形態素解析を行い、解析結果として、単語とそのカテゴリとが定義されている形態素リストを生成する形態素解析工程と、

10

前記情報検索装置の構文解析手段が、前記検索条件に対して構文解析を行い、解析結果として、前記検索条件中の2つの注目単語の左側の単語と右側の単語と、それらのそれぞれのカテゴリから決定される前記2つの注目単語の間の関係の種類を示すロールとが定義されているロールリストを生成する構文解析工程と、

前記情報検索装置の決定手段が、前記形態素リストで定義されているカテゴリの優先順位から、前記検索条件を構成する各単語の内、中心となる単語である中心語の候補を絞り込み、複数の中心語の候補がある場合は、更に、前記ロールリストを用いて、各候補と他の単語の関係によって、中心語の候補を絞り込み、最終的に存在する中心語の候補を、前記検索条件を構成する各単語の中心語として決定し、その中心語から前記ロールリスト中で関係がある単語をたどって、各単語に順位を付与することで、前記中心語をルートとする順位のツリー構造である順位ツリーを生成し、その順位ツリーを構成する各単語の順位に応じて、前記検索条件を構成する各単語の重要度を決定する決定工程と、

20

前記情報検索装置の算出手段が、前記決定工程によって決定された重要度に応じて、検索対象文と前記検索条件との一致度を算出する算出工程と、

前記情報検索装置の出力手段が、前記算出工程で算出された一致度に基づいて、検索結果とする検索対象文を出力する出力工程と

を備える。

【0015】

上記の目的を達成するための本発明によるコンピュータ可読メモリは以下の構成を備える。即ち、

30

複数種類の文を管理するデータベースから検索条件に合致する文を検索する情報検索をコンピュータに機能させるためのプログラムが格納されたコンピュータ可読メモリであって、

前記コンピュータを、

複数の単語から構成される検索条件を入力する入力手段と、

前記検索条件に対して形態素解析を行い、解析結果として、単語とそのカテゴリとが定義されている形態素リストを生成する形態素解析手段と、

前記検索条件に対して構文解析を行い、解析結果として、前記検索条件中の2つの注目単語の左側の単語と右側の単語と、それらのそれぞれのカテゴリから決定される前記2つの注目単語の間の関係の種類を示すロールとが定義されているロールリストを生成する構文解析手段と、

40

前記形態素リストで定義されているカテゴリの優先順位から、前記検索条件を構成する各単語の内、中心となる単語である中心語の候補を絞り込み、複数の中心語の候補がある場合は、更に、前記ロールリストを用いて、各候補と他の単語の関係によって、中心語の候補を絞り込み、最終的に存在する中心語の候補を、前記検索条件を構成する各単語の中心語として決定し、その中心語から前記ロールリスト中で関係がある単語をたどって、各単語に順位を付与することで、前記中心語をルートとする順位のツリー構造である順位ツリーを生成し、その順位ツリーを構成する各単語の順位に応じて、前記検索条件を構成する各単語の重要度を決定する決定手段と、

前記決定手段によって決定された重要度に応じて、検索対象文と前記検索条件との一致

50

度を算出する算出手段と、

前記算出手段で算出された一致度に基づいて、検索結果とする検索対象文を出力する出力手段と

として機能させるためのプログラムが格納されたコンピュータ可読メモリ。

【0016】

【発明の実施の形態】

以下、図面を参照して本発明の好適な実施形態を詳細に説明する。

【0017】

図1は本発明の実施形態の情報検索装置の機能構成を示す図である。

【0018】

10

図1において、入力部101は文字キー、ファンクションキーなどを備えたキー・ボード、マウス等から構成され、検索条件であるクエリを制御部102へ入力する。制御部102はマイクロプロセッサからなり、重要度決定部103においてクエリ中に含まれる個々の検索条件の重要度を求める。また、記憶部105の情報記憶部106に格納された検索対象の個々の情報とクエリを比較する。そして、その比較結果に基づいて、一致度算出部104において個々の情報とクエリの一一致度を、クエリの個々の条件の重要度に応じて算出し、一致度の高い情報から表示部107へ出力する。

【0019】

次に、実施形態の情報検索装置のハードウェア構成について、図2を用いて説明する。

【0020】

20

図2は本発明の実施形態の情報検索装置のハードウェア構成を示す図である。

201はCPU(中央処理装置)であり、ROM202に格納されている、例えば、図4に示すような制御手順プログラムに従って各種処理を実行する。後述するフローチャートに示す処理も、CPU201の制御に基づいて実行される。また、本発明で実行される処理は、その処理を実行するためのプログラムを記憶したCD-ROM205を本情報検索装置300のCD-ROMドライブ204に搭載し、そのCD-ROM205から供給される場合を例に挙げて説明する。

RAM203は、CD-ROMドライブ204に搭載されるCD-ROM205に記憶されている情報検索プログラム501、関連データである電子図鑑502、単語辞書503、文法辞書504等のデータを読み取り、後述する処理で用いる各種パラメータ等を記憶する。また、処理中のデータを格納するためのワーキングメモリの領域を備える。また、本実施形態の処理が実行可能となった状態のメモリマップは、図6に示すように、情報検索プログラム603、電子図鑑604、単語辞書605、文法辞書606以外に、情報検索プログラムが使用するメモリとして、形態素リスト607、ロールリスト608、および検索結果リスト609の領域が確保され、初期化される。

【0021】

30

206はハードディスクドライブ(HDドライブ)であり、内蔵されるハードディスク(不図示)に検索対象である情報や制御プログラム等の各種データを記憶する。

【0022】

40

208はCRTや液晶表示器等のディスプレイであり、画像、テキスト、処理途中のメッセージ等を表示する。また、キーボード207やマウス209により入力されたデータや操作画面もディスプレイ208により表示する。

【0023】

210はLBП(レーザビームプリンタ)やインクジェットプリンタ等のプリンタであり、画像やテキスト等を記録する。207はキーボード、209はマウスであり、オペレータの操作により、各種コマンドや文字コード等を入力することができる。211は制御バスであり、各種構成要素間でのデータあるいはコマンド等を相互に通信可能とする。

【0024】

次に、電子図鑑604の構成の詳細について、図7を用いて説明する。

【0025】

50

図7は本発明の実施形態の電子図鑑の構成の詳細を示す図である。

【0026】

電子図鑑604には、見出し語とその見出し語に対する説明文とその見出し語を表す画像ファイルの名称を対応づけて構成した1つの項目を、複数種類格納している。例えば、図7中では、1つの項目として、見出し語「d a c h s f u n d」と、説明文「d o g w i t h s h o r t l e g s . . .」と、画像ファイル「00001002.jpg」を対応づけて格納されている。各項目は、見出し語の先頭文字によって昇順にソートされる。

【0027】

次に、単語辞書605の構成の詳細について、図8を用いて説明する。

10

【0028】

図8は本発明の実施形態の単語辞書の構成の詳細を示す図である。

【0029】

単語辞書605には、見出し語とカテゴリをそれぞれ1つずつ対応づけて構成した1つの単語の情報を、複数種類格納している。例えば、単語辞書605には、1つの単語の情報として、見出し語「d o g」とカテゴリ「N o u n」を対応づけて格納している。各単語の情報は、見出し語の先頭文字によって昇順にソートされる。

【0030】

次に、文法辞書606の構成の詳細について、図9を用いて説明する。

【0031】

20

図9は本発明の実施形態の文法辞書の構成の詳細を示す図である。

【0032】

文法辞書606には、左カテゴリと右カテゴリとロールとをそれぞれ一つずつ対応づけて構成した文法ルールを、複数種類格納している。各文法ルールにおいて、左カテゴリと右カテゴリは、単語辞書605に格納されているカテゴリに対応する。各文法ルールの意味は、文中において、注目単語の左の単語のカテゴリと右の単語のカテゴリが、それぞれ、文法辞書606の1つの文法ルールの左カテゴリと右カテゴリに一致した場合、その2つの単語の間の関係として、左カテゴリと右カテゴリに対応したロールの関係にある可能性があることを意味している。例えば、文中において、注目単語の左の単語のカテゴリが「A d v e r b」で、右の単語のカテゴリが「A d j e c t i v e」であれば、その2つの単語の間の関係は「M o d i f i e r」の可能性があることが判明する。

30

【0033】

次に、形態素リスト607の構成の詳細について、図10を用いて説明する。

図10は本発明の実施形態の形態素リストの構成の詳細を示す図である。

【0034】

形態素リスト607には、キーボード207から入力されたクエリ、電子図鑑604中の説明文を形態素解析した結果、各単語の情報がそれぞれ、1つの項目として格納される。

【0035】

各項目には、IDと単語とカテゴリと順位と重みが、それぞれ1つずつ対応づけられて格納される。IDには、文の先頭の番号を1とした文の先頭からの単語の番号が格納される。単語には、各単語の語幹が格納される。カテゴリには、各単語のカテゴリが格納される。順位には、文の中心となる単語の順位を1とし、各単語の重要度に応じて順位を定めた結果が格納される。重みには、各単語の重要度に応じた重みが格納される。また、この重みは、文中の全ての単語に対応した重みの和をとると、必ず1になるように構成されている。

40

【0036】

例えば、表T101は、文「d o g w i t h v e r y s h o r t l e g」に対して、形態素解析を行った結果を格納した形態素リストを示している。表T101において、「w i t h」に対応した順位の値が「-1」になっているが、これは、順位を付与する対象外の単語であることを意味している。他の例として、表T104は、文「d o g s

50

「*short leg*」に対して、形態素解析を行った結果を格納した形態素リストを示している。表T104において、順位が「1」の単語として、「*dog*」と「*short*」の2つが存在するように、一つの文中で、順位として、同じ値が出現することもある。形態素リスト607には、クエリに対して形態素解析を行った結果を格納する領域と、説明文に対して形態素解析を行った結果を格納する領域との2つの領域が区別されて存在する。

【0037】

次に、ロールリスト608の構成の詳細について、図11を用いて説明する。

図11は本発明の実施形態のロールリストの構成の詳細を示す図である。

【0038】

ロールリスト608には、キーボード207から入力されたクエリ、電子図鑑604中の説明文に対して構文解析を行った結果、2つの単語間の各関係が、それぞれ1つの項目として格納される。各項目には、左単語IDとロールと右単語IDが、それぞれ1つずつ対応づけられて格納される。左単語IDと右単語IDには、文中の2つの注目単語のそれぞれ左側の単語と右側の単語が対応し、形態素リスト607のIDの値が使用される。ロールには、2つの注目単語の間の関係の種類、すなわち、一方の単語が果たす他方の単語へのロール(役割)が格納される。ロールの種類は、文法辞書606のロールの種類と同じである。

【0039】

例えば、表T201は、文「*dog with very short leg*」に対して、構文解析を行った結果を格納したロールリストを示している。この文に対する形態素リストは図10の表T101であり、これによれば、左単語ID「1」に相当する単語「*dog*」は、右単語ID「2」に相当する単語「*with*」に対して、ロールとして「*Preposition Subject*」すなわち、「*Preposition Subj ect*」の関係にあることが判明する。ロールリスト608には、クエリに対して構文解析を行った結果を格納する領域と、説明文に対して構文解析を行った結果を格納する領域との2つの領域が区別されて存在する。

【0040】

次に、検索結果リスト609の構成の詳細について、図12を用いて説明する。

【0041】

図12は本発明の実施形態の検索結果リストの構成の詳細を示す図である。

検索結果リスト609には、キーボード207から入力されたクエリに従って、電子図鑑604を検索した結果、クエリの条件を満たす各項目が格納される。検索結果リスト609には、一つの項目として、電子図鑑604の見出し語と、検索結果として得られるスコアとが、それぞれ一つずつ対応づけられて格納され、複数の項目が格納される。

【0042】

次に、キーボード207よりクエリを入力し、検索を指示した後、電子図鑑604を検索した結果がディスプレイ208に表示された画面の例について、図13を用いて説明する。

【0043】

図13は本発明の実施形態のディスプレイに表示された画面の例を示す図である。

【0044】

図13において、1301はクエリを入力するためのテキストボックスである。テキストボックス1301内にキーボード207からクエリを入力し、OKボタン1302をマウス209でクリックすると、1303に、クエリの条件を満たす電子図鑑604中の見出し語と、検索の際に得られたスコアが表示される。また、1304には、電子図鑑604中の見出し語に対応した画像ファイルから得られる画像が表示される。また、1305には、見出し語に対応した説明文が表示される。検索結果は、スコアの大きい順に画面左より右に向かって表示され、1画面中に収まらないときは、1306の次画面ボタンをマウス209でクリックすることにより、他の検索結果も画面に表示することができる。

【0045】

10

20

30

40

50

図13の例では、クエリとして「dog with short leg」を指定した場合の検索後の画面を表示しており、「dachsund」がスコア「1.00」で、「skyeterrier」がスコア「1.00」で、「shih tzu」がスコア「1.00」で、それぞれ検索された場合を示している。

【0046】

次に、情報検索プログラム603で実行される処理について、図14を用いて説明する。

【0047】

図14は本発明の実施形態の情報検索プログラムで実行される処理を示す図である。

【0048】

まず、CD-ROM205に格納された情報検索プログラム501、およびその関連データの電子図鑑502、単語辞書503、文法辞書504をCD-ROMドライブ204からRAM203にロードする。また、RAM203に形態素リスト607、ロールリスト608、検索結果リスト609の領域を確保し、その他必要な初期化を行う(ステップS101)。次に、オペレータが、テキストボックス1301にキーボード207を用いてクエリを入力し、マウス209を用いてOKボタン1302をクリックする(ステップS102)。

【0049】

次に、ステップS102で入力されたクエリと一致する見出し語を電子図鑑604の中から検索し、見つかった見出し語を検索結果リスト609の先頭からつめて格納する。その際、スコアの値は1.00としておく(ステップS103)。尚、電子図鑑604中の各項目は、見出し語によって昇順にソートされている。また、本処理は、同種の電子図鑑や電子辞書において、一般に行われている処理である。

【0050】

次に、ステップS102で入力されたクエリの解析を行い、形態素リスト607およびロールリスト608のクエリ用の領域に、解析結果を格納する(ステップS104)。尚、この処理の詳細については、図15を用いて後述する。

【0051】

次に、電子図鑑604の先頭から、順次、説明文を一つずつ取り出す(ステップS105)。そして、取り出すことができる説明文があるか否かを判定する(ステップS106)。取り出すことができる説明文がある場合(ステップS107でYES)、ステップS107へ進み、ステップS105で取り出した説明文の解析を行い、形態素リスト607およびロールリスト608の説明文用の領域に、解析結果を格納する。尚、この処理の詳細についても、図15を用いて後述する。

【0052】

ステップS107の処理を終えたら、ステップS108で、ステップS104で得られたクエリの解析結果と、ステップS107で得られた説明文の解析結果を比較する。そして、クエリの条件を含む説明文に対する見出し語を、条件を含む度合いに応じて算出したスコアとともに、検索結果リスト609につめて格納する。尚、この処理の詳細については、図16を用いて後述する。処理を終えると、ステップS105へ戻る。

【0053】

一方、ステップS107において、全ての説明文を取り出し終えている場合(ステップS107でNO)、ステップS109へ進み、検索結果リスト609に格納された見出し語とその見出し語に対応した情報を電子図鑑604から取り出し、ディスプレイ208に表示する。この表示前には、検索結果リスト609に格納されている各項目を、スコアの値によって降順にソートした後、検索結果リスト609の先頭から、すなわち、スコアの値が大きいものから順番に、図13で説明したような手順で表示する。また、ソートする際、スコアの値が同じ場合は、リストの先頭に出現する項目が先頭側になるように配置する。これによって、ステップS103で検索された項目、すなわち、クエリと一致する見出し語を持つ項目が、一番最初に表示されることが保証される。

【0054】

10

20

30

40

50

次に、図14のステップS104のクエリ解析、ステップS107の説明文解析の処理の詳細について、図15を用いて説明する

図15は本発明の実施形態のステップS104のクエリ解析、ステップS107の説明文解析の処理の詳細を示すフローチャートである。

【0055】

本処理は、形態素リスト607およびロールリスト608に解析結果を出力するが、ステップS104の場合は、クエリ用の形態素リストとロールリストに解析結果が出力される。また、ステップS107の場合は、説明文用の形態素リストとロールリストに解析結果が出力される。以降の説明においては、これらの区別を明示的には説明していないが、実際には区別されていることをあらかじめ断っておく。

10

【0056】

図15のフローチャートにおいて、まず、単語辞書605を用いて、クエリまたは説明文の形態素解析を行い、解析結果を形態素リスト607に格納する（ステップS201）。本処理によって、形態素リスト607には、例えば、図10に示したような情報が格納される。但し、この段階では、形態素リスト607の中で、ID、単語、カテゴリのフィールドのみに値が格納され、順位と重みには格納されていない。また、本処理は、一般に行われている処理であり、特に詳細な説明は行わない。

【0057】

次に、形態素リスト607、文法辞書606を用いて、クエリまたは説明文の構文解析を行い、解析結果をロールリスト608に格納する（ステップS202）。本処理によって、ロールリスト608には、例えば、図11に示したような情報が格納される。本処理は、依存文法を用いた構文解析であり、日本語の係り受け解析などで、一般に行われているので、特に詳細な説明は行わない。

20

【0058】

次に、形態素リスト607およびロールリスト608を用いて、形態素リスト607に格納されている単語のうち、中心となる単語、すなわち、中心語を決定する（ステップS203）。例えば、「Noun」が一番強く、次に「Verb」、「Adjective」といったように、中心語としてのカテゴリの優先順位をつけて中心語の候補を絞り込む。また、形態素リスト607中に、複数の中心語の候補が存在すれば、ロールリスト608を用いて、その候補と他の単語との関係によって、中心語を決定する。更に、形態素リスト607中に「Noun」が2つ以上存在する場合、他の単語と「Object」の関係にある候補は、中心語の候補からはずす等として中心語の候補を絞り込む。このようにして、中心語の候補を絞り込んでいく、最終的に複数の中心語の候補が存在する場合は、すべての候補を中心語とする。中心語として決定された単語に対しては、形態素リスト607の順位のフィールドに、「1」を格納する。

30

【0059】

次に、形態素リスト607中の中心語以外の単語に対して、順位の値を格納する（ステップS204）。ロールリスト608を利用して、中心語と関係がある単語に対して順位の値として「2」を格納する。また、順位「2」の単語と関係がある単語の順位の値を「3」とするようにして、中心語からの関係を次々とたぐっていくことによって、順位を付与していく。但し、カテゴリが、「Preposition」など、特定のカテゴリを持つ単語は、順位の付与対象外となり、順位の値として、「-1」が格納される。また、中心語からの関係をたどって、順位を付与する場合に、付与対象外の単語が出現した場合は、その単語をスキップして、順位の値が連続するようにする。中心語から関係する単語をたどっていく場合、一つの単語が複数の単語と関係する場合があるので、このような場合は、その単語に関する単語は、同じ順位の値が付与されることになる。

40

【0060】

以上説明した処理によって、形態素リスト607中の全ての単語に対する順位が付与されることになる。このようにして付与された順位は、中心語をルートとするツリー構造を構成することになる。このツリー構造を順位ツリーと呼ぶことにする。

50

【0061】

次に、形態素リスト607中の全ての単語に重みを付与する（ステップS205）。重みは、中心語に一番大きい値が付与され、順位が大きくなるに従って、値が小さくなるよう付与する。但し、全ての重みの値の和が、必ず「1.0」になるようにする。また、順位付与対象外の単語に対しては、重みの値として、「0」を格納する。重みの付与の手順としては、まず、形態素リスト607中の中心語の個数を求める。その際、順位付与対象外の単語に対しては、重みの値として、「0」を格納しておく。次に、「1.0」を、中心語の個数で割り、その値を、各中心語から連なる順位ツリーの値とする。そして、各順位ツリーにおいて、順位ツリーの値に「0.7」を掛けた値を、順位ツリーのルートの単語（すなわち、中心語）の重みの値として格納する。

10

【0062】

但し、順位ツリーが中心語のみから構成される場合は、順位ツリーの値を中心語の重みとして格納する。また、ルートの下の枝の個数によって、順位ツリーの値からルートに付与した値を引いた値を割った値を、各枝（すなわち、サブ順位ツリー）の順位ツリーの値とする。そして、同様にして、各サブ順位ツリーにおいて、順位ツリーの値に「0.7」を掛けた値を、ルートの単語の重みの値として格納する。

【0063】

但し、順位ツリーがルートの単語のみから構成される場合は、順位ツリーの値をルートの単語の重みとして格納する。また、ルートの下の枝の個数によって、順位ツリーの値からルートに付与した値を引いた値を割った値を、各枝（すなわち、サブ順位ツリー）の順位ツリーの値とする。この処理を繰り返すことによって、全ての単語に、前述の条件を満たす重みを付与することができる。

20

【0064】

次に、図14のステップS108の処理の詳細について、図16を用いて説明する。

【0065】

図16は本発明の実施形態の図14のステップS108の処理の詳細を示すフローチャートである。

【0066】

まず、スコアの値に初期値の「0」を格納する等の本処理を実行するために必要な初期化処理を行う。

30

【0067】

その後、クエリ用の形態素リスト607中の中心語を順次一つずつ取り出す（ステップS301）。取り出すことができる中心語があるか否かを判定する（ステップS302）。次に、取り出すことができる中心語がある場合（ステップS302でNO）、ステップS303に進む。一方、全ての中心語を取り出し終えている場合（ステップS302でYES）、ステップS313へ進む。

【0068】

ステップS302において、取り出すことができる中心語がある場合、説明文用の形態素リスト607の中にステップS301で取り出した単語（中心語）と同じ単語が存在するか否かを判定する（ステップS303）。存在する場合（ステップS303でYES）、ステップS304へ進む。一方、存在しない場合（ステップS303でNO）、ステップS307へ進む。

40

【0069】

ステップS303において、説明文用の形態素リスト607の中にステップS301で取り出した単語（中心語）と同じ単語が存在する場合、ステップS301で取り出した単語（中心語）と同じ単語が説明文用の形態素リスト607中でも中心語であるか否かを判定する（ステップS304）。中心語である場合（ステップS304でYES）、ステップS305へ進み、ステップS301で取り出した単語（中心語）に対応した重みをクエリ用の形態素リスト607から取り出し、その値をスコアに加える。この処理を終えると、ステップS307へ進む。一方、中心語でない場合（ステップS304でNO）、ステッ

50

PS306へ進み、ステップS301で取り出した単語（中心語）に対応した重みをクエリ用の形態素リスト607から取り出し、その値を2分の1にした値をスコアに加える。この処理を終えると、ステップS307へ進む。

【0070】

ステップS307では、クエリ用のロールリスト608を利用して、ステップS301で取り出した単語（中心語）を順次たどって、中心語から連なっている単語を一つずつ取り出す。そして、取り出すことができる単語があるか否かを判定する（ステップS308）。取り出すことができる単語がある場合（ステップS308でNO）、ステップS309に進む。一方、中心語から連なる単語を全て取り出し終えている場合（ステップS308でYES）、ステップS301へ戻る。

10

【0071】

ステップS308において、取り出すことができる単語がある場合、ステップS307で取り出した単語と同じ単語が説明文用の形態素リスト607に存在するか否かを判定する（ステップS309）。存在する場合（ステップS309でYES）、ステップS310へ進む。一方、存在しない場合（ステップS309でNO）、ステップS307へ戻る。

【0072】

ステップS309において、ステップS307で取り出した単語と同じ単語が説明文用の形態素リスト607に存在する場合、ステップS307で取り出した単語と中心語側への単語との関係が説明文でも一致している否かを判定する（ステップS310）。この判定は以下のように行う。まず、ステップS307で取り出した単語（以降、後単語と呼ぶ）と、それより順位の値が1だけ小さい単語（以降、前単語と呼ぶ）との間のロールをクエリ用のロールリスト608から求める。説明文用の形態素リスト607中に、前単語と同じ単語が存在し、かつ、説明文用のロールリスト608中に、前単語と同じ単語と後単語と同じ単語に対応したロールが、先ほど求めたクエリのロールと一致する場合にのみ、単語の関係が説明文でも一致していると見なす。一致している場合（ステップS310でYES）、ステップS311へ進み、ステップS307で取り出した単語に対応した重みをクエリ用の形態素リスト607から取り出し、その値をスコアに加える。この処理を終えると、ステップS307へ戻る。一方、一致していない場合（ステップS310でNO）、ステップS307で取り出した単語に対応した重みをクエリ用の形態素リスト607から取り出し、その値を2分の1にした値をスコアに加える。この処理を終えると、ステップS307へ戻る。

20

【0073】

一方、ステップS302において、全ての中心語を取り出し終えている場合、スコアの値が0より大きいか否かを判定する（ステップS313）。スコアの値が0である場合（ステップS313でNO）、本処理を終了する。一方、スコアの値が0より大きい場合（ステップS313でYES）、ステップS314へ進み、検索された結果を検索結果リスト609へ格納する。これは、対象となっている説明文に対応した見出し語とスコアの値を対応させて、検索結果リスト609へつめて格納する。処理を終えると、本処理を終了する。

30

【0074】

以上説明した動作を、具体例を挙げてより詳細に説明する。

40

【0075】

まず、クエリとして、「dachsfund」が指定された場合の例を説明する。

【0076】

「dachsfund」が指定されると（ステップS102）、この場合、図7に示されるように、電子図鑑604には見出し語として「dachsfund」が格納されているので、検索結果リスト609の先頭に、見出し語「dachsfund」とスコア「1.0」を対応させた項目が格納される（ステップS103）。スコアの最大値は「1.0」である。そして、ステップS109では、スコアの値が大きい物から順番に表示し、スコアの値が同じ場合はリストの先頭に出現する項目が最初に表示されるので、ステップS1

50

03で検索された「d a c h s f u n d」が検索結果として一番最初にディスプレイ208に表示される(ステップS109)。

【0077】

次に、クエリとして、「d o g w i t h s h o r t l e g」が指定された場合の例を説明する。

【0078】

ステップS201において、クエリに対する形態素解析が行われると、図8に示されるように、各単語のカテゴリが対応づけられているので、クエリ用の形態素リスト607には図10の表T102に示されるような情報が格納される。但し、この時点では、順位と重みには値は格納されていない。

10

【0079】

続いて、ステップS202において、クエリに対する構文解析が行われると、図9に示されるように、各カテゴリ間の関係が記載されているので、クエリ用のロールリスト608には図11の表T202に示されるような情報が格納される。

【0080】

続いて、ステップS203において、中心語が決定される。表T102からわかるように、カテゴリ「Noun」の単語として「d o g」と「l e g」が存在するので、この2つの単語に中心語が絞り込まれる。また、表T202からわかるように、「l e g」は、ロール「P O b j e c t」を持つので、中心語の対象から外される。そして、最終的に「d o g」がクエリの中心語であると決定され、表T102に示されるように、クエリ用の形態素リストの「d o g」に対応した順位として「1」が格納される。

20

【0081】

続いて、ステップS204において、クエリ用のロールリストである表T202を利用して、クエリ用の形態素リストの「d o g」以外の単語に対して、順位の値が格納される。表T202によると、「d o g」に対応した左単語ID「1」に対して、右単語ID「2」が対応している。しかし、単語ID「2」の単語「w i t h」は、表T102によれば、カテゴリが「P r e p o s i t i o n」であるので、順位付与対象外の単語となり、順位には「-1」が格納される。また、「w i t h」に対応した左単語ID「2」に対して、右単語ID「4」が対応している。単語ID「4」の単語「l e g」は、順位付与対象外の単語ではなく、中心語の「d o g」からたどって、最初の順位付与の対象となる単語なので、順位として「2」が格納される。さらに、「l e g」に対応した右単語ID「4」に対して、左単語ID「3」が対応している。単語ID「3」の単語「s h o r t」は、順位付与対象外の単語ではなく、中心語の「d o g」からたどって、「d o g」を含めて3番目の順位付与の対象となる単語なので、順位として「3」が格納される。以上によって、クエリ用の形態素リストに表T102に示されるような順位が格納される。

30

【0082】

続いて、ステップS205において、クエリ用の形態素リストの各単語に対して、重みの値が格納される。まず、順位付与対象外の単語に対して、重みの値として「0」を格納しながら、中心語の個数を求める。表T102によれば、順位の値として「1」を持つのは、「d o g」の1個だけであることがわかる。また、順位の値として「-1」を持つ「w i t h」に対する重みの値として、「0」が格納される。この場合、中心語の個数が1個であるので、順位ツリーの値は「1.0」を「1」で割った「1.0」になる。次に、中心語の「d o g」を取り出し、順位ツリーの値「1.0」に「0.7」を掛けた値「0.7」を「d o g」に対応する重みに格納する。そして、中心語「d o g」に続く順位の「2」を持つ単語は「l e g」のみなので、「1.0」から「0.7」を引いた値「0.3」を「1」で割った値「0.3」が、「l e g」をルートとするサブ順位ツリーの値となる。そして、順位ツリーの値「0.3」に「0.7」を掛けた値「0.21」が、サブ順位ツリーのルートである「l e g」に対応した重みとして格納される。そして、「l e g」に続く順位の「3」を持つ単語は「s h o r t」のみなので、「0.3」から「0.21」を引いて「1」で割った値「0.09」が、「s h o r t」をルートとするサブ順位

40

50

ツリーの値となる。「short」をルートとするサブ順位ツリーには、「short」しか残っていないので、順位ツリーの値「0.09」が、そのまま、「short」に対応した重みとして格納される。このようにして、クエリ用の形態素リストに表T102に示されるような重みが格納される。

【0083】

続いて、ステップS105において、説明文「dog with very short legs」が取り出された場合を説明する。

【0084】

ステップS107において、説明文用の形態素リスト607とロールリスト608に、解析結果が格納される。解析は、前述のクエリの解析と同様にして行われ、形態素リストは図10の表T101に、ロールリストは図11の表T201に、それぞれ示される状態になる。

10

【0085】

続いて、ステップS301において、クエリの中心語「dog」が取り出される。一方、説明文の形態素リストである表T101に「dog」が中心語として格納されている。そのため、ステップS305において、スコアの初期値「0」に、クエリの「dog」に対応した重みの値が加算され、スコアの値は「0.7」になる。

【0086】

続いて、ステップS307において、クエリの中心語「dog」に関係のある単語を取り出す。表T202によれば、「dog」と関係する単語は、「with」だけであるので、「with」を取り出す。説明文の形態素リストである表T101には「with」が存在し、クエリと説明文の両方において、「dog」と「with」の間の関係であるロールの値が「PSub ject」であるので、関係も一致したことになる。これにより、ステップS311において、スコアの値「0.7」に、クエリの「with」に対応した重みの値「0」が加算され、スコアの値は、「0.7」になる。

20

【0087】

続いて、ステップS307において、クエリの「with」に関係のある単語を取り出す。表T202によれば、すでに取り出された単語以外で「with」と関係する単語は、「leg」だけであるので、「leg」を取り出す。説明文の形態素リストには「leg」が存在し、クエリと説明文の両方において、「with」と「leg」の間の関係であるロールの値が「PObject」であるので、関係も一致したことになる。これにより、ステップS311において、スコアの値「0.7」に、クエリの「leg」に対応した重みの値「0.21」が加算され、スコアの値は、「0.91」になる。

30

【0088】

続いて、ステップS307において、クエリの「leg」に関係のある単語を取り出す。表T202によれば、すでに取り出された単語以外で「leg」と関係する単語は、「short」だけであるので、「short」を取り出す。説明文の形態素リストである表T101には「short」が存在し、クエリと説明文の両方において、「leg」と「short」の間の関係であるロールの値が「Modifier」であるので、関係も一致したことになる。これにより、ステップS311において、スコアの値「0.91」に、クエリの「short」に対応した重みの値「0.09」が加算され、スコアの値は、「1.00」になる。

40

【0089】

続いて、中心語も含めて、クエリの全ての単語を取り出し終え、スコアの値も「0」よりも大きいので、検索結果リスト609に、説明文「dog with very short legs」に対応した見出し語が、スコアの値「1.00」とともに格納される。スコアの値が「1.0」であるということは、クエリの条件を完全に満たしていることを意味している。

【0090】

次に、ステップS105において、説明文「short dog with long

50

「leg」が取り出された場合を説明する。

【0091】

ステップS107において、説明文用の形態素リスト607とロールリスト608に、解析結果が格納される。解析は、前述のクエリの解析と同様にして行われ、形態素リストは図10の表T105に、ロールリストは図11の表T205に、それぞれ示される状態になる。

【0092】

続いて、ステップS301において、クエリの中心語「dog」が取り出される。一方、説明文の形態素リストである表T105に「dog」が中心語として格納されている。そのため、ステップS305において、スコアの初期値「0」に、クエリの「dog」に対応した重みの値が加算され、スコアの値は「0.7」になる。

【0093】

続いて、ステップS307において、クエリの中心語「dog」に関係のある単語を取り出す。表T202によれば、「dog」と関係する単語は、「with」だけであるので、「with」を取り出す。説明文の形態素リストである表T105には「with」が存在し、クエリと説明文の両方において、「dog」と「with」の間の関係であるロールの値が「Subj ect」であるので、関係も一致したことになり、ステップS311において、スコアの値「0.7」に、クエリの「with」に対応した重みの値「0」が加算され、スコアの値は、「0.7」になる。

【0094】

続いて、ステップS307において、クエリの「with」に関係のある単語を取り出す。表T202によれば、すでに取り出された単語以外で「with」と関係する単語は、「leg」だけであるので、「leg」を取り出す。説明文の形態素リストである表T105には「leg」が存在し、クエリと説明文の両方において、「with」と「leg」の間の関係であるロールの値が「Obj ect」であるので、関係も一致したことになる。これにより、ステップS311において、スコアの値「0.7」に、クエリの「leg」に対応した重みの値「0.21」が加算され、スコアの値は、「0.91」になる。

【0095】

続いて、ステップS307において、クエリの「leg」に関係のある単語を取り出す。表T202によれば、すでに取り出された単語以外で「leg」と関係する単語は、「short」だけであるので、「short」を取り出す。説明文の形態素リストである表T105には「short」が存在するが、表T205によれば、説明文では、「leg」と「short」の間に関係が存在しない。そのため、ステップS312において、スコアの値「0.91」に、クエリの「short」に対応した重みの値「0.09」を2分の1した値「0.045」が加算され、スコアの値は、「0.955」になる。

【0096】

続いて、中心語も含めて、クエリの全ての単語を取り出し終え、スコアの値も「0」よりも大きいので、検索結果リスト609に、説明文「short dog with long legs」に対応した見出し語が、スコアの値「0.955」とともに格納される。

【0097】

続いて、同様に様々な説明文に対して、検索を行った後、ステップS109において、検索結果がディスプレイに表示されることになるが、その中には、説明文「dog with very short legs」に対応した電子図鑑604の項目が、スコアの値「1.00」として表示される。また、それに続いて、説明文「short dog with long legs」に対応した電子図鑑604の項目が、スコアの値「0.955」とともに表示される。

【0098】

この場合、「dog with short legs」というクエリに対して、「do

10

20

30

40

50

g with very short legs」という説明文のスコアは「1.0」になり、「dog with long legs」という説明文のスコアは「0.955」になるので、前者の説明文の方がクエリとの一致度が高いことがわかる。

【0099】

次に、クエリとして、「dog with thick legs」が指定された場合の例を説明する。

【0100】

ステップS104において、クエリに対する解析が行われ、クエリ用の形態素リストは図10の表T103に示される状態になり、クエリ用のロールリストは図11の表203に示される状態になる。

10

【0101】

続いて、ステップS105において、説明文「dog with very short legs」が取り出された場合を説明する。

【0102】

ステップS107において、説明文用の形態素リスト607とロールリスト608に、解析結果が格納される。解析は、前述のクエリの解析と同様にして行われ、形態素リストは図10の表T101に、ロールリストは図11の表T201に、それぞれ示される状態になる。

【0103】

続いて、ステップS301において、クエリの中心語「dog」が取り出される。一方、説明文の形態素リストである表T101に「dog」が中心語として格納されている。そのため、ステップS305において、スコアの初期値「0」に、クエリの「dog」に対応した重みの値が加算され、スコアの値は「0.7」になる。

20

【0104】

続いて、ステップS307において、クエリの中心語「dog」に関係のある単語を取り出す。表T203によれば、「dog」と関係する単語は、「with」だけであるので、「with」を取り出す。説明文の形態素リストである表T101には「with」が存在し、クエリと説明文の両方において、「dog」と「with」の間の関係であるロールの値が「PSubj ect」であるので、関係も一致したことになる。これにより、ステップS311において、スコアの値「0.7」に、クエリの「with」に対応した重みの値「0」が加算され、スコアの値は、「0.7」になる。

30

【0105】

続いて、ステップS307において、クエリの「with」に関係のある単語を取り出す。表T203によれば、すでに取り出された単語以外で「with」と関係する単語は、「leg」だけであるので、「leg」を取り出す。説明文の形態素リストである表T101には「leg」が存在し、クエリと説明文の両方において、「with」と「leg」の間の関係であるロールの値が「PObject」であるので、関係も一致したことになる。これにより、ステップS311において、スコアの値「0.7」に、クエリの「leg」に対応した重みの値「0.21」が加算され、スコアの値は、「0.91」になる。

40

【0106】

続いて、ステップS307において、クエリの「leg」に関係のある単語を取り出す。表T203によれば、すでに取り出された単語以外で「leg」と関係する単語は、「thick」だけであるので、「thick」を取り出す。説明文の形態素リストである表T101には「thick」は存在しない。そのため、ステップS312において、スコアの値「0.91」に、クエリの「thick」に対応した重みの値「0.09」を2分の1した値「0.045」が加算され、スコアの値は、「0.955」になる。

【0107】

続いて、中心語も含めて、クエリの全ての単語を取り出し終え、スコアの値も「0」よりも大きいので、検索結果リスト609に、説明文「dog with very shor

50

t legs」に対応した見出し語が、スコアの値「0.955」と共に格納される。

【0108】

続いて、同様に様々な説明文に対して、検索を行った後、ステップS109において、検索結果がディスプレイに表示されることになるが、その中には、説明文「dog with very short legs」に対応した電子図鑑604の項目が、スコアの値「0.955」として表示される。

【0109】

このように、クエリの条件と部分一致する説明文において、重要度の低い条件のみが一致していない場合は、一致度が高いことがわかる。

【0110】

次に、クエリとして、「dog short legs」が指定された場合の例を説明する。この例は、文法的に正しくないクエリが入力されても、構文解析で使用する文法ルールに不備があり、正しく解析できない場合に実行される処理を説明するものである。

【0111】

ステップS201において、クエリに対する形態素解析が行われると、クエリ用の形態素リスト607には、図10の表T104に示されるような情報が格納される。但し、この時点では、順位と重みには、値は格納されていない。

【0112】

続いて、ステップS202において、クエリに対する構文解析が行われる。ここで、文法辞書606に、左カテゴリNounと右カテゴリAdjectiveに対する文法ルールが記述されておらず、また、左カテゴリNounと右カテゴリNounに対する文法ルールが記述されていないものとする。すると、クエリ用のロールリスト608には、図11の表T204に示されるような情報が格納される。表T204に示されるように、文法的に正しくないクエリを解析した場合、あるいは、文法的に正しいクエリが入力されても、構文解析で使用する文法ルールに不備があり、正しく解析できない場合、構文解析の結果として、1文中の単語と単語の間の関係の連鎖がとぎれることになる。例えば、表T204によれば、「dog」は、クエリ中の他の単語との関係がなく、連鎖がとぎれている。

【0113】

続いて、ステップS203において、中心語が決定される。表T204からわかるように、カテゴリ「Noun」の単語として「dog」と「leg」が存在するので、この2つの単語に中心語が絞り込まれる。しかし、「dog」はロールを持たないし、「leg」はロールとして「Modifier」を持つだけなので、これ以上の中心語の絞り込みはできない。そのため、最終的に「dog」と「leg」がクエリの中心語であると決定され、表T104に示されるように、クエリ用の形態素リストの「dog」と「leg」に対応した順位として「1」が格納される。

【0114】

続いて、ステップS204において、クエリ用のロールリストである表T204を利用して、クエリ用の形態素リストの「dog」と「leg」以外の単語に対して、順位の値が格納される。表T204によると、「leg」に対応した右単語ID「3」に対して、左単語ID「2」が対応している。単語ID「2」の単語「short」は、順位付与対象外の単語でなく、中心語の「leg」からたどって、「leg」を含めて2番目の順位付与の対象となる単語なので、順位として「2」が格納される。以上によって、クエリ用の形態素リストに表T104に示されるような順位が格納される。

【0115】

続いて、ステップS205において、クエリ用の形態素リストの各単語に対して、重みの値が格納される。まず、順位付与対象外の単語に対して、重みの値として「0」を格納しながら、中心語の個数を求める。表T104によれば、順位の値として「1」を持つのは、「dog」と「leg」の2個であることがわかる。中心語の個数が2個であるので、各順位ツリーの値は「1.0」を「2」で割った「0.5」になる。次に最初の中心語の「dog」を取り出す。中心語「dog」の順位ツリーには、「dog」しか存在しない

10

20

30

40

50

ので、順位ツリーの値「0.5」が、そのまま、「dog」に対応した重みとして格納される。続いて、次の中心語の「leg」を取り出し、順位ツリーの値「0.5」に「0.7」を掛けた値「0.35」を「leg」に対応する重みに格納する。そして、中心語「leg」に続く順位の「2」を持つ単語は「short」のみなので、「0.5」から「0.35」を引いて「1」で割った値「0.15」が、「short」をルートとするサブ順位ツリーの値となる。「short」をルートとするサブ順位ツリーには、「short」しか残っていないので、順位ツリーの値「0.15」が、そのまま、「short」に対応した重みとして格納される。このようにして、クエリ用の形態素リストに表T104に示されるような重みが格納される。

【0116】

10

以上説明したように、文法的に正しくないクエリを解析した場合、あるいは、文法的に正しいクエリが入力されても、構文解析で使用する文法ルールに不備があり、正しく解析できない場合においても、各単語の順位と重みを付与することができる。また、説明文に対しても、同様に、文法的に正しくない場合や、文法ルールに不備がある場合でも、各単語の順位と重みを付与することができる。

【0117】

続いて、ステップS105において、説明文「dog with very short legs」が取り出された場合を説明する。ステップS107において、説明文用の形態素リスト607とロールリスト608に、解析結果が格納される。解析は、前述のクエリの解析と同様にして行われ、形態素リストは図10の表T101に、ロールリストは図11の表T201に、それぞれ示される状態になる。

20

【0118】

続いて、ステップS301において、クエリの最初の中心語「dog」が取り出される。一方、説明文の形態素リストである表T101に「dog」が中心語として格納されている。そのため、ステップS305において、スコアの初期値「0」に、クエリの「dog」に対応した重みの値が加算され、スコアの値は「0.5」になる。

【0119】

続いて、ステップS307において、クエリの中心語「dog」に関係のある単語を取り出す。表T204によれば、「dog」と関係する単語は存在しないので、ステップS308において分岐し、ステップS301へ進む。

30

【0120】

続いて、ステップS301において、クエリの次の中心語「leg」が取り出される。一方、説明文の形態素リストである表T101には「leg」は中心語としては格納されていない。そのため、ステップS306において、スコアの初期値「0」に、クエリの「leg」に対応した重みを2分の1した値「0.175」が加算され、スコアの値は「0.675」になる。

【0121】

続いて、ステップS307において、クエリの「leg」に関係のある単語を取り出す。表T204によれば、すでに取り出された単語以外で「leg」と関係する単語は、「short」だけであるので、「short」を取り出す。説明文の形態素リストである表T101には「short」は存在し、説明文のロールリストである表201には「short」のロールとして「Modifier」とあり、クエリの「short」のロールと一致する。そのため、ステップS311において、スコアの値「0.675」に、クエリの「short」に対応した重みの値「0.15」が加算され、スコアの値は、「0.825」になる。

40

【0122】

続いて、中心語も含めて、クエリの全ての単語を取り出し終え、スコアの値も「0」よりも大きいので、検索結果リスト609に、説明文「dog with very short legs」に対応した見出し語が、スコアの値「0.825」とともに格納される。

【0123】

50

続いて、同様に様々な説明文に対して、検索を行った後、ステップ S 1 0 9 において、検索結果がディスプレイに表示されることになるが、その中には、説明文「d o g w i t h v e r y s h o r t l e g s」に対応した電子図鑑 6 0 4 の項目が、スコアの値「0.825」として表示される。

【0124】

以上説明したように、本実施形態によれば、個々の検索条件の重要度（重み）を求め、情報記憶部 1 0 6 に格納された検索対象の情報との一致度に、検索条件の重要度を用いて、検索結果を一致度の大きい順に出力する。そのため、検索条件に部分一致する情報の中から、検索の意図に近い状況をより早く出力でき、検索結果のチェック作業を大幅に軽減できる。

10

【0125】

また、自然な言葉による文で入力された検索条件を構文解析することによって、検索条件を構成する単語の個々の重要度を求め、情報記憶部 1 0 6 に格納された検索対象の情報との一致度に、それらの重要度を用いて、検索結果を一致度の大きい順に出力する。そのため、検索条件の入力が容易になるとともに、検索条件に部分一致する情報の中から、検索の意図に近い情報をより早く出力でき、検索結果のチェック作業を大幅に軽減できる。

【0126】

また、自然な言葉による文で入力された検索条件を構文解析することによって、検索条件を構成する単語の個々の重要度を求める。また、情報記憶部 1 0 6 に格納された検索対象の情報（文章）を構文解析し、その情報に含まれる単語および各単語間の関係も含めて検索条件と比較し、その一致度に検索条件の重要度を用いて、検索結果を一致度の大きい順に出力する。そのため、検索条件の入力が容易になるとともに、検索の精度が向上し、検索条件に部分一致する情報の中から、検索の意図に近い情報をより早く出力でき、検索結果のチェック作業を大幅に軽減できる。

20

【0127】

また、自然な言葉による文で入力された検索条件を構文解析することによって、検索条件を構成する単語の個々の重要度を求める。また、情報記憶部 1 0 6 に格納された検索対象の情報（文章）を構文解析してその情報中の個々の重要度を求める。そして、検索条件に含まれる単語および各単語間の関係も含めて検索条件と比較し、その一致度に検索条件の重要度とその情報との重要度の両方を用いて、検索結果を一致度の大きい順に出力する。そのため、検索条件の入力が容易になるとともに、検索の精度が向上し、検索条件に部分一致する情報の中から、検索の意図に近い情報をより早く出力でき、検索結果のチェック作業を大幅に軽減できる。

30

【0128】

また、検索条件を構成する単語の個々の重要度を求め、情報記憶部 1 0 6 に格納された検索対象の情報との一致度に、それらの重要度を用いて、検索結果を一致度の大きい順に出力するが、検索条件の最も重要な条件を満たさない検索結果は出力しないので、検索条件に部分一致する情報の中から、検索の意図に近い情報をより早く出力でき、検索結果のチェック作業を大幅に軽減できる。

40

【0129】

また、文法的に正しくないクエリを解析した場合、あるいは文法的に正しいクエリが入力されても、構文解析で使用する文法ルールに不備があり、正しく解析できない場合においても、説明文との一致度を求めることができることがわかる。また、説明文に対しても、同様に、文法的に正しくない場合や、文法ルールに不備がある場合でも、クエリとの一致度を求めることができる。

【0130】

上記実施形態では、外部記憶装置として C D - R O M 2 0 5 から情報検索プログラム 5 0 1 および関連データを直接 R A M 2 0 3 にロードして実行させる例を示したが、この他に、C D - R O M 2 0 5 から情報検索プログラム 5 0 1 および関連データをいったんハードディスクドライブ 2 0 6 に内蔵されるハードディスク（不図示）に格納（インストール）

50

しておき、本情報検索プログラムを動作させる時点で、そのハードディスクから情報検索プログラムをRAM203にロードするようにしてよい。

【0131】

また、本情報検索プログラムを記録する媒体は、CD-ROM以外にFD(フロッピーディスク)、CD-R、ICメモリカード等であってもよい。

【0132】

更に、本情報検索プログラムを、ROM202に記憶しておき、これをメモリマップの一部となるように構成し、直接CPU201で実行することも可能である。

【0133】

また、以上の説明において、クエリおよび説明文として使用する言語として、英語を使用した例を示したが、本発明は英語のみに適用可能な発明ではなく、日本語を含めたあらゆる言語に適用可能な発明である。

10

【0134】

また、上記実施形態では、電子図鑑の説明文を検索の対象とし、説明文に対する見出し語と画像と説明文自身を出力する例を説明したが、本発明は、電子図鑑の説明文の検索のみに適用されるものではない。ワードプロセッサ等で作成された一般的な文章を検索の対象とし、大量の文章を格納した文書管理システムなどのデータベースの中から、検索意図に合致した文書を取り出す目的にも、本発明は適用可能である。同様に、インターネットのWWW(World Wide Web)上の文章を検索する検索エンジンにも、本発明は適用可能である。その場合、電子図鑑における見出し語を文書の見出しに対応させ、電子図鑑における説明文を文書の本文に対応させて適用することも可能である。

20

【0135】

また、上記実施形態では、検索結果として出力する電子図鑑の項目として、スコアの値が0よりも大きい説明文に対する項目を出力していたが、このしきい値をオペレータが自由に設定できるようにし、オペレータが設定したしきい値よりも大きいスコアの値を持つ説明文に対応した見出し語のみを、検索結果リストに出力するようにしても良い。

【0136】

また、上記実施形態では、電子図鑑には、説明文のみが格納された例を示したが、説明文をあらかじめ構文解析した結果を説明文と対応させて格納しておき、図14のステップS107で説明文の解析を行う代わりに、電子図鑑に格納されている説明文の構文解析結果を取り出すようにしても良い。

30

【0137】

また、上記実施形態では、ステップS103において、電子図鑑の見出し語を検索対象とし、検索条件と見出し語が完全に一致する場合のみ、スコアの値を1.0として、検索結果リストに出力していたが、これに限定されない。例えば、ステップS104からステップS108までに示される説明文とクエリとのマッチングのように、クエリと見出し語の両方の解析を行った後、ステップS108で行うマッチングと同様のマッチングを行った結果を検索結果リストに出力しても良い。

【0138】

また、上記実施形態では、検索条件を自然な言葉によって表現していたが、単語の論理演算式で表現しても良い。そして、検索条件の個々の重要度を、論理演算式の中での単語の出現順序で決定しても良い。例えば、検索条件「白い犬」を表現するのに、『「犬」「白い」』と表現し、重要度は、「犬」の方が「白い」より大きいと決定しても良い。あるいは、重要度を決定するために、検索要求者が、個々の条件の重要度の重みを明示的に指定しても良い。例えば、検索条件「白い犬」を表現するのに、『「犬」「白い」』と表現した後、重要度の重みとして、検索要求者が「0.7, 0.3」と入力することにより、「犬」の重要度の重みが「0.7」であり、「白い」の重要度の重みが「0.3」であるというように決定しても良い。

40

【0139】

また、上記実施形態では、クエリと説明文を比較する場合、単語と単語との間の関係まで

50

含めて比較していたが、単語と単語との間の関係を含めて比較する必要はなく、クエリ中の単語が説明文に存在すれば、そのクエリの単語の重みをスコアに加算するようにしてもよい。

【0140】

また、上記実施形態では、構文解析した結果、単語に対してのみ重要度の重みを付与し、その重みを利用して一致度を算出したが、これに限定されない。例えば、単語間の関係であるロールに対しても、重みを付与し、一致度を算出する際に、クエリと説明文のロールを比較する際に、単語の重みだけでなく、ロール重みも利用して一致度を算出するようにしても良い。

【0141】

また、上記実施形態では、電子図鑑には、説明文のみが格納されたが、これに限定されない。例えば、説明文をあらかじめ構文解析した結果を説明文と対応させて格納しておき、図14のステップS107で説明文の解析を行う代わりに、電子図鑑に格納されている説明文の構文解析結果を取り出すようにしても良い。

また、上記実施形態では、図16のステップS303において、説明文に中心語が含まれない場合、ステップS307へ進んで、中心語以外の単語に対してもマッチングを行っていたが、この場合にステップS307へ進む代わりに、ステップS301へ進んで、その中心語をルートとする順位ツリーに存在する単語のマッチングは行わないようにも良い。また、この場合に、ステップS307へ進む代わりに、ステップS313へ進むようにしても良い。前者の場合は、クエリの中心語が一つでも説明文に存在すれば、スコアの値は0より大きくなり、検索結果リストに出力されるが、後者の場合は、クエリの中心語が説明文中に一つでも存在しなければ、検索結果リストに出力されない。このように、用途に応じて、検索結果リストの内容を変更することも可能である。

【0142】

尚、本発明は、複数の機器（例えばホストコンピュータ、インターフェース機器、リーダ、プリンタなど）から構成されるシステムに適用しても、一つの機器からなる装置（例えば、複写機、ファクシミリ装置など）に適用してもよい。

【0143】

また、本発明の目的は、前述した実施形態の機能を実現するソフトウェアのプログラムコードを記録した記憶媒体を、システムあるいは装置に供給し、そのシステムあるいは装置のコンピュータ（またはCPUやMPU）が記憶媒体に格納されたプログラムコードを読み出し実行することによっても、達成されることは言うまでもない。

【0144】

この場合、記憶媒体から読み出されたプログラムコード自体が前述した実施形態の機能を実現することになり、そのプログラムコードを記憶した記憶媒体は本発明を構成することになる。

【0145】

プログラムコードを供給するための記憶媒体としては、例えば、フロッピディスク、ハードディスク、光ディスク、光磁気ディスク、CD-ROM、CD-R、磁気テープ、不揮発性のメモリカード、ROMなどを用いることができる。

【0146】

また、コンピュータが読み出したプログラムコードを実行することにより、前述した実施形態の機能が実現されるだけでなく、そのプログラムコードの指示に基づき、コンピュータ上で稼働しているOS（オペレーティングシステム）などが実際の処理の一部または全部を行い、その処理によって前述した実施形態の機能が実現される場合も含まれることは言うまでもない。

【0147】

更に、記憶媒体から読み出されたプログラムコードが、コンピュータに挿入された機能拡張ボードやコンピュータに接続された機能拡張ユニットに備わるメモリに書込まれた後、そのプログラムコードの指示に基づき、その機能拡張ボードや機能拡張ユニットに備わるC

10

20

30

40

50

PUなどが実際の処理の一部または全部を行い、その処理によって前述した実施形態の機能が実現される場合も含まれることは言うまでもない。

【0148】

【発明の効果】

以上に説明したように、本発明によれば、検索条件を満たす情報を適切に検索することができる情報検索装置及びその方法、コンピュータ可読メモリを提供できる。

【0149】

【図面の簡単な説明】

【図1】本発明の実施形態の情報検索装置の機能構成を示す図である。

【図2】本発明の実施形態の情報検索装置のハードウェア構成を示す図である。 10

【図3】本発明の実施形態で実行される処理のプログラムが記憶媒体から供給されることを示す図である。

【図4】本発明の実施形態のROMの構成を示す図である。

【図5】本発明の実施形態のCD-ROMの構成を示す図である。

【図6】本発明の実施形態のRAM上のメモリマップの構造を示す図である。

【図7】本発明の実施形態の電子図鑑の構成の詳細を示す図である。

【図8】本発明の実施形態の単語辞書の構成の詳細を示す図である。

【図9】本発明の実施形態の文法辞書の構成の詳細を示す図である。

【図10】本発明の実施形態の形態素リストの構成の詳細を示す図である。

【図11】本発明の実施形態のロールリストの構成の詳細を示す図である。 20

【図12】本発明の実施形態の検索結果リストの構成の詳細を示す図である。

【図13】本発明の実施形態のディスプレイに表示された画面の例を示す図である。

【図14】本発明の実施形態の情報検索プログラムで実行される処理を示す図である。

【図15】本発明の実施形態のステップS104のクエリ解析、ステップS107の説明文解析の処理の詳細を示すフローチャートである。

【図16】本発明の実施形態の図14のステップS108の処理の詳細を示すフローチャートである。

【符号の説明】

101 入力部

102 制御部

103 重要度決定部

104 一致度算出部

105 記憶部

106 情報記憶部

107 表示部

201 CPU

202 ROM

203 RAM

204 CD-ROMドライブ

205 CD-ROM

206 ハードディスクドライブ

207 キーボード

208 ディスプレイ

209 マウス

210 プリンタ

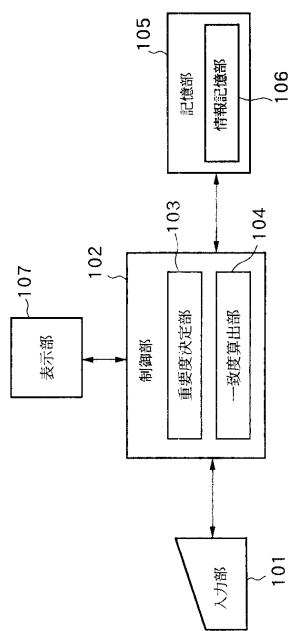
211 制御バス

20

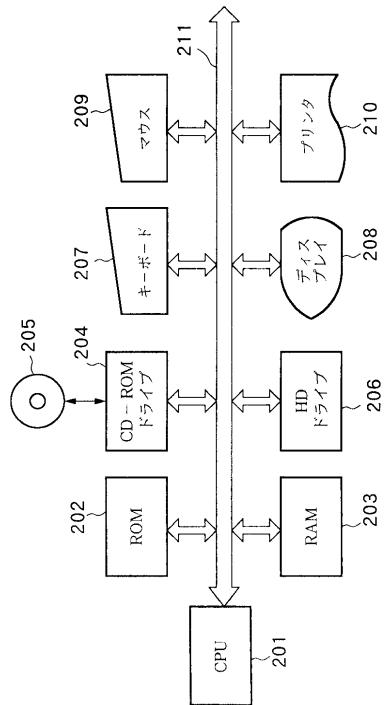
30

40

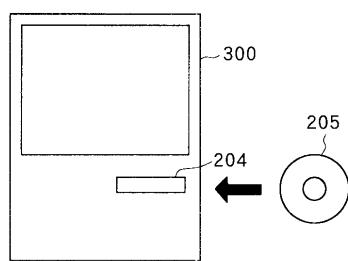
【図1】



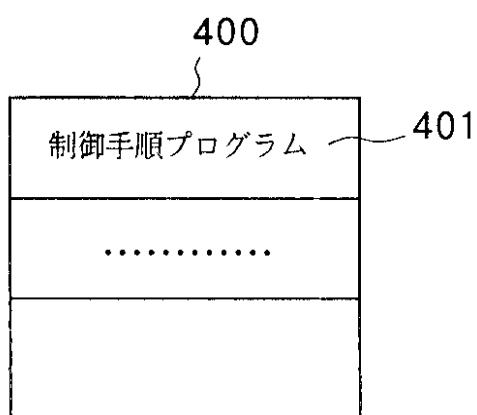
【図2】



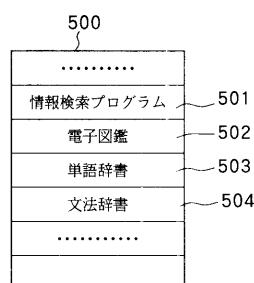
【図3】



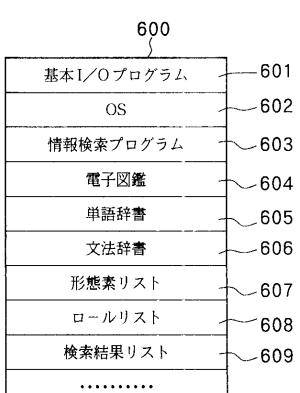
【図4】



【図5】



【図6】



【図7】

見出し語	画像ファイル	説明文
:	:	:
dachshund	00001002.jpg	dog with short legs and
:	:	:
shih tzuh	00002002.jpg	dog with long hair and
:	:	:
skye terrier	00003002.jpg	dog with long body and
:	:	:

【図8】

見出し語	カテゴリ
:	:
dog	Noun
:	:
leg	Noun
:	:
short	Adjective
:	:
thick	Adjective
:	:
very	Adverb
:	:
with	Preposition
:	:

【図9】

左カテゴリ	右カテゴリ	ロール
:	:	:
adverb	Adjective	Modifier
:	:	:
Adjective	Noun	Modifier
Preposition	Noun	PObject
Verb	Noun	Object
Verb	Noun	Goal
:	:	:
Noun	Preposition	PSubject
:	:	:

【図10】

(T101)	ID	単語	カテゴリ	順位	重み
1	dog	Noun	1	0.70	
2	with	Preposition	-1	0	
3	very	Adverb	4	0.03	
4	short	Adjective	3	0.06	
5	leg	Noun	2	0.21	

(T102)	ID	単語	カテゴリ	順位	重み
1	dog	Noun	1	0.70	
2	with	Preposition	-1	0	
3	short	Adjective	3	0.09	
4	leg	Noun	2	0.21	

(T103)	ID	単語	カテゴリ	順位	重み
1	dog	Noun	1	0.70	
2	with	Preposition	-1	0	
3	thick	Adjective	3	0.09	
4	leg	Noun	2	0.21	

(T104)	ID	単語	カテゴリ	順位	重み
1	dog	Noun	1	0.50	
2	short	Adjective	2	0.15	
3	leg	Noun	1	0.35	

(T105)	ID	単語	カテゴリ	順位	重み
1	short	Adjective	2	0.15	
2	dog	Noun	1	0.70	
3	with	Preposition	-1	0	
4	long	Adjective	3	0.045	
5	leg	Noun	2	0.105	

【図11】

(T201)	左単語 ID	ロール	右単語 ID
1	PSubject	2	
2	PObject	5	
3	Modifier	4	
4	Modifier	5	

(T202)	左単語 ID	ロール	右単語 ID
1	PSubject	2	
2	PObject	4	
3	Modifier	4	

(T203)	左単語 ID	ロール	右単語 ID
1	PSubject	2	
2	PObject	4	
3	Modifier	4	

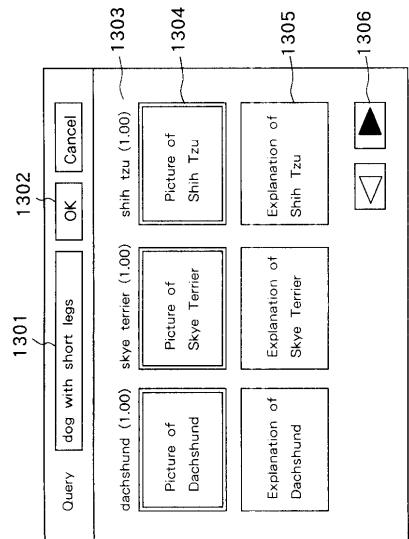
(T204)	左単語 ID	ロール	右単語 ID
2	Modifier	3	

(T205)	左単語 ID	ロール	右単語 ID
1	Modifier	2	
2	PSubject	3	
3	PObject	5	
4	Modifier	5	

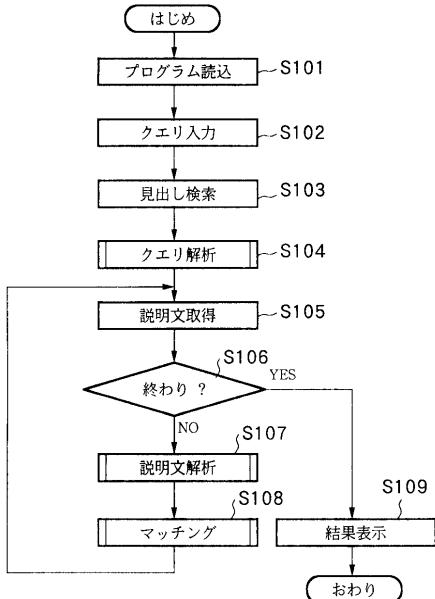
【図12】

見出し語	スコア
dachshund	1.00
sky terrier	1.00
shih tzu	1.00
:	:
:	:

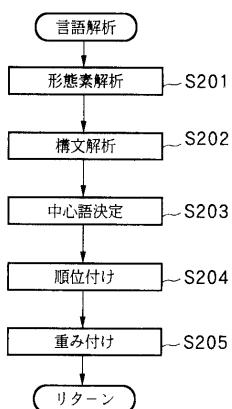
【図13】



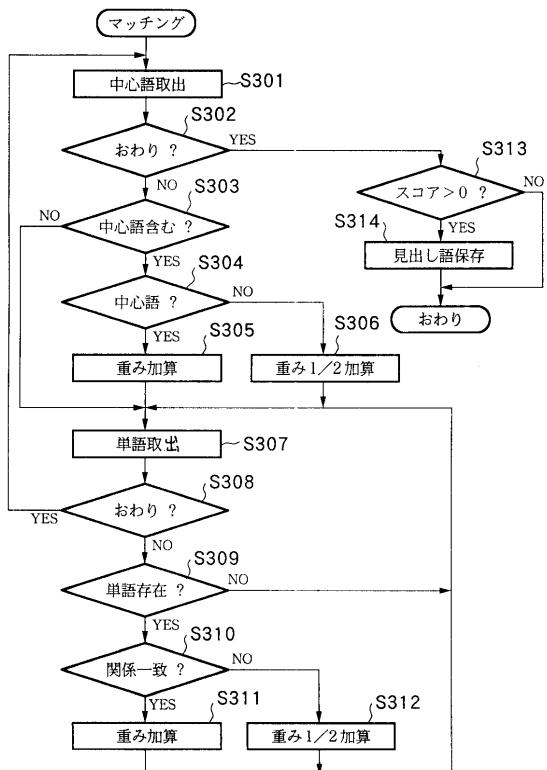
【図14】



【図15】



【図16】



フロントページの続き

審査官 長 由紀子

(56)参考文献 特開平06-208588 (JP, A)

特開平07-085101 (JP, A)

特開平05-342255 (JP, A)

特開平02-027476 (JP, A)

特開平04-139580 (JP, A)

富田 準二, 竹野 浩, 主題グラフ及び関連度情報からの単語重要度付与を用いた情報検索システムの提案, 情報処理学会研究報告, 日本, 社団法人情報処理学会, 1998年11月30日, 第98巻第109号, 17-24

岡 満美子, キーワード検索に基づくテキスト検索, 情報処理学会研究報告, 日本, 社団法人情報処理学会, 1994年 9月16日, 第94巻第77号, 89-96

(58)調査した分野(Int.Cl., DB名)

G06F 17/30