

(19) 日本国特許庁 (JP)

(12) 特 許 公 報 (B2)

(11) 特許番号

特許第6152711号
(P6152711)

(45) 発行日 平成29年6月28日 (2017.6.28)

(24) 登録日 平成29年6月9日 (2017.6.9)

(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 4 A

請求項の数 7 (全 30 頁)

(21) 出願番号 特願2013-118248 (P2013-118248)
 (22) 出願日 平成25年6月4日 (2013.6.4)
 (65) 公開番号 特開2014-235664 (P2014-235664A)
 (43) 公開日 平成26年12月15日 (2014.12.15)
 審査請求日 平成28年3月10日 (2016.3.10)

(73) 特許権者 000005223
 富士通株式会社
 神奈川県川崎市中原区上小田中4丁目1番
 1号
 (74) 代理人 100074099
 弁理士 大菅 義之
 (74) 代理人 100133570
 弁理士 ▲徳▼永 民雄
 (72) 発明者 大倉 清司
 神奈川県川崎市中原区上小田中4丁目1番
 1号 富士通株式会社内
 (72) 発明者 潮田 明
 神奈川県川崎市中原区上小田中4丁目1番
 1号 富士通株式会社内

最終頁に続く

(54) 【発明の名称】 情報検索装置および情報検索方法

(57) 【特許請求の範囲】

【請求項 1】

複数の検索用単語を含む情報の入力を受け付ける入力部と、

前記複数の検索用単語を含む情報から2つの検索用単語を分離し、検索対象文に含まれる複数の単語および前記複数の単語と他の単語間の前記検索対象文における関係を示す意味情報を前記検索対象文と関連付けて記憶した記憶部から、前記2つの検索用単語と対応する2つの単語とその意味情報を検索して抽出する演算処理部と、

抽出された前記意味情報を出力する出力部と、
 を有し、

前記入力部は、少なくとも一つの文の入力の受け付けも行い、

前記入力部が受け付けた入力が入力文の場合には、前記演算処理部は、前記文を意味解析することにより意味情報を生成し、前記記憶部から前記意味情報と関連付けて記憶された文を検索する

ことを特徴とする情報検索装置。

【請求項 2】

前記意味情報は、前記2つの単語のそれぞれに対応する意味記号を含み、

前記演算処理部は、

分離した前記検索用単語を意味記号に変換し、変換された前記意味記号のうちの2つを検索キーとし、前記検索キーを含む前記意味情報を前記記憶部において検索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 3】

前記出力部は、前記意味情報を表層文字列に変換して出力することを特徴とする請求項 1 または請求項 2 のいずれかに記載の情報検索装置。

【請求項 4】

前記演算処理部は、

前記記憶部に前記意味情報と関連付けて記憶された前記検索対象文において、前記意味情報に含まれる前記 2 つの単語の少なくともいずれかが出現する出現位置を参照し、前記出現位置に基づき前記文の少なくとも一部を抽出し、

前記出力部は、

抽出された前記検索対象文の少なくとも一部を出力することを特徴とする請求項 1 から請求項 3 のいずれかに記載の情報検索装置。

10

【請求項 5】

前記入力部は、抽出された前記意味情報を絞り込むための指示を受け付け、

前記出力部は、前記入力部により受け付けられた指示に応じて絞り込まれた前記意味情報のみを出力することを特徴とする請求項 4 に記載の情報検索装置。

【請求項 6】

前記演算処理部は、入力された文を意味解析することにより、前記記憶部に、前記意味情報と前記文とを関連付けて記憶させることを特徴とする請求項 1 から請求項 5 のいずれかに記載の情報検索装置。

【請求項 7】

20

情報検索装置が、

複数の検索用単語を含む情報の入力を受け付け、

前記複数の検索用単語を含む情報から前記 2 つの検索用単語を分離し、

検索対象文に含まれる複数の単語および前記複数の単語と他の単語間の前記検索対象文における関係を示す意味情報を前記検索対象文と関連付けて記憶した記憶部から、前記 2 つの検索用単語と対応する 2 つの単語とその意味情報を検索して抽出し、

抽出された前記意味情報を出力し、

前記複数の検索用単語を含む情報の入力の受け付けでは、少なくとも一つの文の入力の受け付けも行い、

前記受け付けた入力が入力された文の場合には、前記文を意味解析することにより意味情報を生成し、前記記憶部から前記意味情報と関連付けて記憶された文を検索することを特徴とする情報検索方法。

30

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、情報検索装置および情報検索方法に関する。

【背景技術】

【0002】

例えば、インターネット上で何らかの情報を得たい場合に、検索サイトにおいてキーワードを入力して、入力したキーワードを含む文書を抽出する技術が知られている。また、このようなキーワード検索を行うための言語処理に関しては、様々な技術が知られている。（例えば、非特許文献 1 ～ 3 参照）

40

【先行技術文献】

【非特許文献】

【0003】

【非特許文献 1】「自然言語理解」、田中穂積、辻井潤一共編、オーム社、1988 年

【非特許文献 2】「入門 自然言語処理」、Steven Bird、Ewan Klein、Edward Loper 著、萩原正人、中山敬広、水野貴明 訳、O'Reilly Japan、2010 年

【非特許文献 3】Steven Bird、Ewan Klein、Edward Loper 著、萩原正人、中山敬広、水野貴明 訳、“Python による日本語自然言語処理”、[online]、インターネット

50

(<http://nltk.googlecode.com/svn/trunk/doc/book-jp/ch12.html>)

【発明の概要】

【発明が解決しようとする課題】

【0004】

ところで、上記のようなよく知られたキーワードによる検索では、キーワード単位のクエリを用いるため、複数のキーワード間の関係を検索条件に含めることができない。よって、キーワード単位のクエリには、あいまい性が含まれ、各キーワードを組み合わせで表される意味が特定されないことがある。このため、キーワード検索では、ユーザが意図したとおりに検索されない場合がある。また、キーワードを含んでいても意図しない文書が検索される場合がある。つまり、たとえ抽出された文書中のキーワードにヒットした部分を出力しても、必ずしもユーザが欲しい情報ではないことがある。従って、ユーザは、有用な情報を抽出する判断に時間を費やすことになる。

10

【0005】

ひとつの側面によれば、本発明は、意図する検索結果を効率よく取得することが可能な情報検索装置および情報検索方法である。

【課題を解決するための手段】

【0006】

ひとつの態様である情報検索装置は、入力部、演算処理部、出力部を有している。入力部は、複数の検索用単語を含む情報の入力を受け付ける。演算処理部は、受け付けた情報から2つの検索用単語を分離し、検索対象文に含まれる複数の単語および前記複数の単語と他の単語間の前記検索対象文における関係を示す意味情報を検索対象文と関連付けて記憶した記憶部から、2つの検索用単語と対応する2つの単語とその意味情報を検索して抽出する。出力部は、抽出された意味情報を出力する。ここで、入力部は、少なくとも一つの文の入力の受け付けも行う。入力部が受け付けた入力が文の場合には、演算処理部は、文を意味解析することにより意味情報を生成し、記憶部から意味情報と関連付けて記憶された文を検索する。

20

【0007】

別の態様である情報検索方法は、情報検索装置が、複数の検索用単語を含む情報の入力を受け付け、複数の検索用単語を含む情報から2つの検索用単語を分離する。また、情報検索装置が、記憶部から、2つの検索用単語と対応する2つの単語とその意味情報を検索して抽出し、抽出された意味情報を出力する。このとき、記憶部は、検索対象文に含まれる複数の単語および前記複数の単語と他の単語間の検索対象文における関係を示す意味情報を検索対象文と関連付けて記憶している。複数の検索用単語を含む情報の入力の受け付けでは、少なくとも一つの文の入力の受け付けも行う。ここで、受け付けた入力が文の場合には、文を意味解析することにより意味情報を生成し、記憶部から意味情報と関連付けて記憶された文を検索する。

30

【発明の効果】

【0008】

上述した態様の情報検索装置および情報検索方法によれば、意図する検索結果を効率よく取得することが可能となる。

40

【図面の簡単な説明】

【0009】

【図1】情報検索装置の構成の一例を示すブロック図である。

【図2】文の解析例を示す図である。

【図3】文の解析例を示す図である。

【図4】文の解析例を示す図である。

【図5】文字オフセット例および意味記号の一例を示す図である。

【図6】インデックステーブルの一例を示す図である。

【図7】評価値テーブルの一例を示す図である。

【図8】クエリが文の場合の検索処理を示すフローチャートである。

50

【図 9】クエリから分割された単語を含む単語テーブルの一例を示す図である。

【図 10】辞書テーブルの一例を示す図である。

【図 11】検索キーの一例を示す図である。

【図 12】検索結果の一例を示す図である。

【図 13】検索結果の表示例を示す図である。

【図 14】検索結果を示す表の変換例を示す図である。

【図 15】検索結果を示す表の変換例を示す図である。

【図 16】検索結果を示す表の変換例を示す図である。

【図 17】検索結果を示す表の変換例を示す図である。

【図 18】選択例を示す図である。

10

【図 19】キーワードによる検索処理を示すフローチャートである。

【図 20】表の変換処理の一例を示すフローチャートである。

【図 21】変形例 1 による検索結果の表示例を示す図である。

【図 22】変形例 1 による検索結果の表示例を示す図である。

【図 23】変形例 1 による検索結果の表示例を示す図である。

【図 24】変形例 1 による検索結果の表示例を示す図である。

【図 25】変形例 1 による検索結果の表示例を示す図である。

【図 26】変形例 1 による検索結果の表示例を示す図である。

【図 27】変形例 2 による文の解析例を示す図である。

【図 28】変形例 2 による文の解析例を示す図である。

20

【図 29】変形例 2 による文の解析例を示す図である。

【図 30】変形例 2 による文字オフセット例および意味記号を示している。

【図 31】変形例 2 による意味解析を説明する図である。

【図 32】変形例 2 による辞書テーブルの一例を示す図である。

【図 33】変形例 2 による意味解析を説明する図である。

【図 34】変形例 2 による表示例を示す図である。

【図 35】変形例 2 による検索結果の一例を示す図である。

【図 36】標準的なコンピュータのハードウェア構成の一例を示す図である。

【発明を実施するための形態】

【0010】

30

(第 1 の実施の形態)

以下、図面を参照しながら、第 1 の実施の形態による情報検索装置 1 について説明する。図 1 は、情報検索装置 1 の構成の一例を示すブロック図である。情報検索装置 1 は、クエリとして、少なくとも一つの単語、または文を入力することにより検索を行うシステムである。情報検索装置 1 は、対象文書 Data Base (DB) 11、検索用インデックス 13、評価値テーブル 15、評価値計算部 39、ランキング部 41 を有している。また、情報検索装置 1 は、クエリ入力部 23、キーワード入力部 25、キーワード変換部 27、検索キー生成部 29、文章入力部 31、意味解析部 33、意味最小単位生成部 35、検索部 37、出力部 43、辞書 51、記憶部 53 を有している。検索部 37 は、キーワード検索部 45、および自然文検索部 47 を備えている。

40

【0011】

検索対象文書 DB 11、検索用インデックス 13、評価値テーブル 15 は、検索を実行する前に準備処理として生成されるものとする。辞書 51 は、予め用意されているものとするが、必要に応じて追加記憶、または修正が可能であるようにしてもよい。検索対象文書 DB 11 は、検索対象となる文書を記憶したデータベースである。例えば、検索対象文書 DB 11 に記憶される各文書は、文書を識別する識別情報と関連付けて記憶されることが好ましい。

【0012】

検索用インデックス 13 は、検索対象文書に含まれる各文の意味最小単位、ノードの位置等を格納しているデータベースである。意味最小単位は、文中の 2 つの概念間の関係ま

50

たは概念の役割を表すものである。ノードとは、文中の単語の概念を表すものである。予め行う準備処理では、複数の検索対象文書に対して意味解析を行い、それらの文書中の各文について意味最小単位を生成し、始点、終点の各ノードの位置および文字列長等を含む検索用インデックス 13 を生成する。意味最小単位については後述する。

【0013】

評価値テーブル 15 は、検索用インデックス 13 に含まれる各意味最小単位に関する評価値を格納したテーブルである。評価値とは、例えば、各意味最小単位を含む文書の数を示す検索数に基づき計算される値とすることができる。一例としては、下記の式 1 のような idf 値を用いることができる。

$$idf = \log (\text{総文書数} / \text{当該意味最小単位を含む文書数}) \cdots (\text{式 1})$$

10

【0014】

ここで、総文書数とは、検索対象文書 DB 11 に記憶された文書の総数である。当該意味最小単位を含む文書数とは、総文書数の中で、 idf 値算出の対象の意味最小単位を含む文書の数である。 idf 値は、意味最小単位を含む検索対象文書の数が多い場合に、より大きい。意味最小単位の評価値は、他の値でもよいが、意味最小単位の有用性を示す値とすることが好ましい。評価値は、評価値計算部 39 により計算される。

【0015】

上述したように、検索は、自然言語文（以下、単に文という）が入力される場合と、単語（以下、キーワードという）が入力される場合が想定される。クエリ 21 は、検索のための少なくとも一つのキーワード、または少なくとも一つの文、あるいはそれらの組み合わせ等である。クエリ入力部 23 は、ユーザのキーボード、マウス、タッチパネル等の操作、またはネットワークを介して入力されるクエリ 21 を受け付け、クエリ 21 が、文であるかキーワードであるかを判別する。文であるかキーワードであるかの判別は、例えば句点や読点の有無により行うようにしてもよい。

20

【0016】

キーワード入力部 25 は、クエリ 21 が少なくとも一つのキーワードを含む場合に、クエリ 21 のキーワード文字列を受け付け、キーワードをスペースなどのデリミタで分割する。キーワード変換部 27 は、分割したそれぞれのキーワードに対し、辞書 51 を参照しながら、単語から意味記号へ変換する。辞書 51 は、単語と意味記号とを関連付けた情報である。意味記号とは、意味を表す記号である。

30

【0017】

検索キー生成部 29 は、変換された意味記号から 2 つ組を生成し、検索キーとする。検索部 37 は、検索キーに基づいて検索対象文書 DB 11、検索用インデックス 13 等のデータベースを検索する。このとき、検索キーにマッチした意味最小単位についての頻度情報も検索する。検索結果表示部に検索結果が表示される。

【0018】

文章入力部 31 は、クエリ入力部 23 に入力されたクエリ 21 が文の場合にクエリ 21 を受け付け、例えば、句点等により一文ずつに分割する。意味解析部 33 は、例えば、クエリ 21 の 1 文 1 文に関して意味解析を行う。意味解析の出力は、単語の意味（意味記号）をノード、2 つの意味記号の関係をアークとした有向グラフである。

40

【0019】

意味最小単位生成部 35 は、1 文の意味を表す有向グラフから、2 つの意味記号の関係を表す「意味最小単位」を抽出する。意味最小単位は、アークごとに、（アークが発するノード（始点ノード）、アークが到達するノード（終点ノード）、アーク名とする。アークが発・到達するノードがない場合は、“NIL”で示す。

【0020】

検索部 37 のキーワード検索部 45 は、クエリ 21 がキーワードの場合に、クエリ 21 から生成された検索キーを条件として、検索用インデックス 13 を検索する。自然文検索部 47 は、クエリ 21 が文の場合に、クエリ 21 から生成された意味最小単位を条件として、検索用インデックス 13 を検索する。複数の意味最小単位が検索条件となっている場

50

合には、いずれか少なくとも一つを含む場合に検索結果として抽出される。検索にマッチした意味最小単位に対応する文書が、検索用インデックス 13 から選出される。

【0021】

評価値計算部 39 は、評価値テーブル 15 と検索用インデックス 13 を参照し、マッチした意味最小単位に基づいて抽出した各文を含む文書の評価値を計算する。ランキング部 41 は、抽出した各文書をランク付けする。つまり、ランキング部 41 は、評価値計算部 39 により計算された文書の評価値をソートキーとしてソートする。

【0022】

出力部 43 は、ランク付けされた結果、キーワード検索部 45 による後述する検索結果等を出力する。出力の形態は、例えば表示、印刷、又は送信である。抽出された文書は、例えば、有用な順、または分類された順等に並べられ、ユーザに提示される。例えば、抽出された文が表示される。辞書 51 は、単語と意味記号とを関連付けて記憶した情報である。記憶部 53 は、例えば、各種処理を行う際に必要に応じて情報を随時読み出し及び書き込み可能な記憶装置である。

【0023】

ここで、準備処理として、検索対象文書 DB 11、検索用インデックス 13、評価値テーブル 15 を生成する処理について、図 2 から図 6 を参照しながら説明する。この処理は、クエリ 21 として文が入力された場合の処理と類似しており、文章入力部 31、意味解析部 33、意味最小単位生成部 35 により実行することができるため、これらを用いて処理を行うとして説明する。なお、準備処理は、実際に情報検索装置 1 により検索実行前に行ってもよい。また、準備処理は、文章入力部 31、意味解析部 33、意味最小単位生成部 35 などを有する別の装置で行って、検索時には、準備処理を行った装置で生成された検索対象文書 DB 11、検索用インデックス 13、評価値テーブル 15 を利用するようにしてもよい。

【0024】

図 2 から図 4 は、文の解析例を示す図である。図 5 は、文字オフセット例、および意味記号の一例を示す図である。図 6 は、インデックステーブル 81 の一例を示す図である。文章入力部 31 は、検索対象文書 DB 11 に記憶しようとする文書が入力されると、入力された文書を一文ずつに分割する。意味解析部 33 は、分割されたそれぞれの文に関して、意味解析を行う。このとき、意味解析部 33 は、文を単語に分割してノードとし、単語間の関係を解析することによりノード間の関係、始点ノード、終点ノード、および文におけるそれぞれのノードの位置および文字列長を抽出する。意味最小単位生成部 35 は、意味解析の結果に基づき意味最小単位を生成する。

【0025】

図 2 の例で説明すると、入力された原文 71 が、「太郎は花子に本をあげた。」という文であるとする、意味解析部 33 により、意味解析が行われ、有向グラフ 73 および意味最小単位 75 が生成される。

【0026】

ここで、有向グラフと意味最小単位について説明する。意味最小単位は、意味解析の結果として得られる有向グラフの部分構造を表す。有向グラフは、ノードとアークとを含む。図 2 の有向グラフ 73 は、有効グラフの一例を示し、意味最小単位 75 は、意味最小単位の一例を示す。有向グラフの生成は、例えば、非特許文献 1～3 等に記載されている技術を用いて行うことができる。

【0027】

ノードは、入力文中の単語の概念（意味）を表す。「あげる」、「本」、「太郎」、及び「花子」は、ノードの例である。それぞれのノードには、その概念を表す記号（意味記号という。）が付加される。「GIVE」、「BOOK」、「TARO」、及び「HANAKO」は、意味記号の例である。

【0028】

アークは、ノード間の関係又はノードの役割を表す。アークが 2 つのノード間にある場

10

20

30

40

50

合には、そのアークは2つのノード間の関係を表す。例えば、図中「G I V E」ノードから「B O O K」ノードに引かれているアークは、「対象」と名付けられている。この表示は、「B O O K」は「G I V E」の対象であることを意味している。一方、終点のノードがないアークは、起点のノードが有する役割を表す。例えば、図中起点のノードが「G I V E」であり、かつ終点がノードのないアークの一方は、「過去」と名付けられている。「G I V E」は、過去の役割であることを意味している。なお、アークが出るノードを始点ノード、アークが向かうノードを終点ノードという。

【0029】

意味解析部33は、意味最小単位の生成において、有向グラフからアークを抽出し、以下のように処理する。

10

(a) アークが2つのノードをつないでいる場合には、意味最小単位として、それぞれのアークに対して、(始点ノード、終点ノード、アークの名前)を出力する。

(b) 始点ノードがない場合には、意味最小単位として、(「N I L」、終点ノード、アークの名前)を出力する。

(c) 終点ノードがない場合には、意味最小単位として、(始点ノード、「N I L」、アークの名前)を出力する。

【0030】

このように、例えば、原文71からは、意味最小単位75に記載の意味最小単位が抽出される。同様に、「花子は太郎に本をあげるだろう。」という原文に基づき、図3の解析例76が抽出され、「太郎は棚に本をあげた。」という原文に基づき、図4の解析例77が生成される。

20

【0031】

図5は、文字オフセット例78および意味記号79を示している。この例は、検索対象文書DB11に格納する文の一例であり、文書ID=21、文番号=3の文の例である。オフセットとは、文の先頭から数えた文字数である。文の先頭のオフセットを「0」とし、1文字ごとにオフセットを増やしていくと、文字オフセット例78に示すように、1文字ごとにオフセットが対応付けられる。例えば意味解析部33により意味解析が行われた際、文字列と意味記号との対応がとれる。また、「太郎」に相当する意味記号は「T A R O」、のようになる。

【0032】

30

図6に示すように、インデックステーブル81は、意味最小単位を検索用インデックス13に格納した例である。インデックステーブル81は、意味最小単位83、文書ID85、文ID87、始点ノード位置89、始点ノード文字列長91、終点ノード位置93、終点ノード95を有している。文書ID85は、意味最小単位83が抽出された文書の識別情報である。文ID87は、意味最小単位83が抽出された文の識別情報である。

【0033】

始点ノード位置89は、意味最小単位83における始点ノードのはじめの文字の、文ID87の先頭から数えた文字数を表す。始点ノード文字列長91は、始点ノードの文字数である。終点ノード位置93は、意味最小単位83における終点ノードのはじめの文字の、文ID87の先頭から数えた文字数を表す。終点ノード文字列長95は、終点ノードの文字数である。

40

【0034】

インデックステーブル81において、最初の3行が、図3の意味最小単位75のうちの3つを格納した例である。(G I V E、H A N A K O、目標)を例にとると、文書ID=23、文ID=3である。そして始点ノード(=「G I V E」)の位置は、図6を参照すると、始点ノード位置89=8であり、始点ノード文字列長91=2である。同様に、終点ノード(=「H A N A K O」)の位置は終点ノード位置93=3、長さは、終点ノード文字列長95=2である。このようにして、解析した全ての意味最小単位等が、検索用インデックス13に格納される。

【0035】

50

全ての意味最小単位が格納されると、例えば評価値計算部 39 により頻度情報が計算される。頻度情報とは、各意味最小単位がデータベース中に出現する回数である。頻度情報は、例えば、評価値テーブル 15 に格納される。さらに、頻度情報に基づき、上述した i d f 値が計算される。評価値計算部 39 は、算出した i d f 値を、意味最小単位と関連付けて評価値テーブル 15 に記憶するようにしてもよい。

【0036】

図 7 は、評価値テーブル 99 の一例を示す図である。評価値テーブル 99 は、最小意味単位と、対応する i d f 値とを対応付けた情報である。さらに、最小意味単位ごとの頻度情報を記憶するようにしてもよい。

【0037】

以上のように、準備処理においては、文章入力部 31 が、検索対象文書 DB 11 に含まれる文書を文に分割する。意味解析部 33 は意味解析を行って有向グラフを生成し、有向グラフに基づき、例えば、インデックステーブル 81 のように、検索用インデックス 13 に情報を追加する。意味解析部 33 は、全ての文書および文について意味解析を行うとともに、検索用インデックス 13 へ解析結果を記憶させる。評価値計算部 39 は、頻度情報、i d f 値を計算する。これにより、検索対象文書 DB 11 と、対応する検索用インデックス 13、評価値テーブル 15 が生成される。検索用インデックス 13 によれば、意味最小単位から、文書 ID 85、文 ID 87、ノードの文中の位置が検索可能となる。

【0038】

次に、図 8 を参照しながら、文による検索処理について説明する。この検索処理においては、クエリ及び各検索対象文書に含まれる各文に関して意味解析を行い、それぞれの意味最小単位を得て、意味最小単位を検索キーとして検索が行われる。また、意味最小単位の i d f 値を使って抽出された文書の評価値を計算することにより、ランク付けが行われる。

【0039】

図 8 は、クエリが文の場合の検索処理を示すフローチャートである。図 8 に示すように、文章入力部 31 は、クエリとして入力された文を受け付け (S111)、文が複数の場合には、一つずつの文に分割する (S112)。意味解析部 33 は、それぞれの文の意味解析を行い、例えば有向グラフを生成する。意味最小単位生成部 35 は、意味解析の結果に基づき、上記準備処理と同様に、意味最小単位を生成する (S113)。ただし、意味最小単位のクエリを受け付けることにより、意味最小単位を特定するようにしてもよい。自然文検索部 47 は、抽出された意味最小単位を検索キーとする。例えば、検索キーは、図 2 に示した意味最小単位 75 に含まれる意味最小単位 (G I V E、T A R O、目標) 等とすることができる。

【0040】

自然文検索部 47 は、検索用インデックス 13 において、検索キーに合致する意味最小単位 83、その意味最小単位 83 を含む文の文 ID 87 等を抽出し、例えば記憶部 53 に記憶する (S115)。すなわち、自然文検索部 47 は、始点ノード、終点ノード、アークが検索キーと一致する意味最小単位を検索用インデックス 13 から抽出する。

【0041】

自然文検索部 47 は、S115 の処理を、クエリ 21 から抽出された全ての検索キーについて行うまで繰り返す (S116: NO)。S115 の処理が全ての検索キーについて行われると (S116: YES)、評価値計算部 39 は、抽出された文書ごとに、評価値テーブル 15 を参照しながら、文書の評価値を算出する (S117)。ランキング部 41 は、計算された評価値に基づき抽出された文書をソートし (S118)、出力部 43 により結果を出力する (ステップ 119)。

【0042】

ここで、クエリが文の場合の評価値計算の例について説明する。まず、評価値計算部 39 は、全文書の評価値を「0」に設定し、検索キーが検索用インデックス 13 に記憶された意味最小単位とマッチした場合に、文毎に評価値を算出する。評価値計算部 39 は、そ

10

20

30

40

50

の文を含む文書の評価値にその文の評価値を足す。評価値計算部 39 がマッチした文のすべてを処理することにより、文書の評価値が得られる。文書の評価値は、その文書に含まれる文の評価値の総和である。

【0043】

一つの検索対象文 n の評価値は、例えば、下記の式 2 で表される。

文 n の評価値 $S_n = (\text{クエリの意味最小単位の集合}(K_1, K_2, \dots, K_i, \dots))$
 のうち、(文 n に出現する K_i の idf 値 \times 文 n における K_i の出現回数) の総和)
 $\times M^2$

ただし、 M は、文 n において検索キーとして特定された意味最小単位の種類数

・・・(式 2)

10

【0044】

種類数 M は、クエリの全体を網羅していることを評価することに役立つ。また、 M の二乗値を用いることにより、その評価の程度が増す。文 n における K_i の出現回数は、一つの検索対象文に含まれる意味最小単位のうち、検索キーとして特定された意味最小単位と一致した意味最小単位の数である。

【0045】

文書の評価値は、例えば、下記の式 3 で表される。

文書の評価値 (D) = 文 n の評価値 (S_n) の総計・・・(式 3)

このように、評価値計算部 39 は、文書に含まれる文の評価値を合算する。

【0046】

20

一例として、ある文 m が、6 つの意味最小単位を含んでおり、それぞれの idf 値 = 2.0 であり、それぞれの意味単位の出現回数が 1 回の場合、この文 m の評価値 (S_m) は、下記の式 4 で算出される。

評価値 (S_m) = $(2 \times 1 + 2 \times 1 + 2 \times 1 + 2 \times 1 + 2 \times 1 + 2 \times 1) \times 6^2$
 $= 432.0 \dots$ (式 4)

上記評価値は、クエリ 21 に応じた意味最小単位が多く含まれる文ほど、高評価となる。

【0047】

文書の評価値の計算例は、以下のようになる。例えば、文書 A が文 1 と文 m の 2 文からなると仮定する。文 1 の評価値 (S_1) = 18.0、文書 A の評価値は $18.0 + 432.0 = 450.0$ となる。

30

【0048】

ランキング部 41 は、例えば、文書の評価値で昇順または降順にランキングすることができる。出力部 43 は、並び替えられた文書のデータを出力する。このとき、抽出された文の評価値をソートキーとして、抽出された文をソートし、ソート順に表示するようにしてもよい。

【0049】

以上のように、クエリ入力部 23 により、入力が文と判別された場合、文章入力部 31 は、クエリ 21 に含まれる少なくとも一つの文を文に分割する。意味解析部 33 は、それぞれの文に対して意味解析を行い、有向グラフを生成する。意味最小単位生成部 35 は、生成された有向グラフに基づき、意味最小単位を生成する。自然文検索部 47 は、生成された意味最小単位を検索キーとして、検索用インデックス 13 を対象に検索を行う。評価値計算部 39 は、検索結果に基づき文書の評価値を計算し、ランキング部 41 は、評価値に基づき文書をソートする。出力部 43 は、検索結果を出力する。

40

【0050】

次に、キーワードがクエリ 21 として入力された場合について、図 9 から図 18 を参照しながら説明する。図 9 は、クエリ 21 から分割された単語を含む単語テーブル 131 の一例を示す図である。図 10 は、辞書テーブル 133 の一例を示す図である。図 11 は、検索キー 135 の一例を示す図である。

【0051】

図 9 の例では、ユーザがクエリ 21 として「あげる 太郎 本」と入力し、検索を行う

50

場合を示している。ユーザの意図は、「誰かが誰かに本をあげる（プレゼントする）」という文を検索することである。「誰か」、には「太郎」を含むとする。

【 0 0 5 2 】

図 9 に示すように、単語テーブル 1 3 1 は、上記クエリ 2 1 から分割された単語を示す例であり、「あげる」、「太郎」、「本」が含まれる。単語テーブル 1 3 1 は、例えば、キーワード入力部 2 5 において生成される。

【 0 0 5 3 】

図 1 0 に示すように、辞書テーブル 1 3 3 は、辞書 5 1 に含まれる情報の一例である。辞書テーブル 1 3 3 には、「あげる」に対応する意味記号「G I V E」、「L I F T」、「太郎」に対応する意味記号「T A R O」等が含まれている。辞書テーブル 1 3 3 は、キーワード変換部 2 7 が、単語テーブル 1 3 1 に含まれる単語を辞書テーブル 1 3 3 に含まれる意味記号に変換する際に参照される。

【 0 0 5 4 】

図 1 1 に示すように、検索キー 1 3 5 は、抽出された単語に対応する意味記号の組み合わせにより生成される。すなわち、「あげる」、「太郎」、「本」の 3 つの単語から、対応する意味記号が 4 つ「G I V E」、「L I F T」、「T A R O」、「B O O K」が検索された場合、4 つの意味記号から異なる 2 つを選択した 1 2 個の検索キーが抽出される。このとき検索キーは、2 つの意味記号と、一つのアークで表され、例えば、(G I V E、T A R O、*)、(G I V E、B O O K、*)、・・・、である。なお、「*」は任意のアークを表す。

【 0 0 5 5 】

一般的に、検索キーは、(意味記号 A、意味記号 B、*) で表される。ただし、意味記号 A 意味記号 B である。また、検索時には、(意味記号 A、意味記号 B、*) および (意味記号 B、意味記号 A、*) について検索を行うものとする。このとき、名詞と動詞との組み合わせのみを抽出するようにしてもよい。検索キー 1 3 5 は、検索キー生成部 2 9 で生成される。

【 0 0 5 6 】

図 1 2 は、検索結果 1 4 1 の一例を示す図である。検索結果 1 4 1 は、検索結果の一例を示す情報である。検索結果 1 4 1 は、検索キー 1 4 3、検索結果 1 4 5、検索結果を含む文 I D 1 4 7、検索件数 1 4 9 を含んでいる。検索キー 1 4 3 は、検索キー生成部 2 9 で生成された例えば検索キー 1 3 5 である。検索結果 1 4 5 は、検索用インデックス 1 3 から抽出された、検索キー 1 3 5 のそれぞれと一致する意味最小単位である。検索結果 1 4 5 を含む文 I D 1 4 7 は、検索結果 1 4 5 の意味最小単位が含まれる文書および文の識別情報である。検索件数 1 4 9 は、検索の結果抽出された文の数である。

【 0 0 5 7 】

例えば、検索キーとして (G I V E、T A R O、*) を検索する場合、図 6 のインデックステーブル 8 1 において、検索結果 9 7、検索結果 9 8 がマッチする。検索結果 9 7、検索結果 9 8 を参照すると、文書 I D 8 5、文 I D 8 7 に基づき以下の情報が抽出される。

【 0 0 5 8 】

すなわち、検索キー (G I V E、T A R O、動作主) が含まれる文は、(文書 I D 2 1、文 I D 3) であり、検索キー (G I V E、T A R O、目標) が含まれる文は、(文書 I D 3 2、文 I D 5 3) である。同様に他の全ての組み合わせについても検索が行われる。

【 0 0 5 9 】

図 1 3 は、検索結果の表示例 1 5 1 を示す図である。図 1 3 に示すように、表示例 1 5 1 は、検索結果 1 4 1 の検索結果を含む文 I D 1 4 7 に重複して抽出されたものを排除することにより、3 種類の文が検索結果として抽出されたことを示している。すなわち、(文書 I D 2 1、文 I D 3)、(文書 I D 3 2、文 I D 5 3)、(文書 I D 8 1、文 I D 3) である。

【 0 0 6 0 】

10

20

30

40

50

図 1 2 の検索結果 1 4 1、および、図 1 3 に示した表示例 1 5 1 は、例えば、ユーザの意図とは異なる「L I F T」に相当する検索結果を含んでいる。そこで、よりユーザの意図に近い検出結果の表示、または意図する結果の絞込みを容易にする表示のための表変換について、図 1 4 から図 1 7 を参照しながら説明する。図 1 4 から図 1 7 は、検索結果を示す表の変換例を示す図である。

【 0 0 6 1 】

図 1 4 に示すように、表変換例 1 5 3 は、検索キー 1 5 5、検索結果 1 5 7、検索件数 1 4 9、検索結果を含む文 I D 1 4 7、文例 1 5 9 を示している。検索キー 1 5 5 は、検索キー 1 3 5 の意味記号の部分単語表示したものである。単語表示は、検索時にユーザが入力したクエリ 2 1 に含まれる各単語を意味記号に変換したときの対応をキーワード変換部 2 7 が例えば記憶部 5 3 に記憶させておき、意味記号を対応する単語に置換することにより可能である。それぞれの意味最小単位は、2 つの単語に置換される。

10

【 0 0 6 2 】

検索結果 1 5 7 は、検索結果 1 4 5 を表層文字列に変換した文である。変換は、検索用インデックス 1 3 の始点ノード位置 8 9、終点ノード位置 9 3 等に基づき行うことができる。文例 1 5 9 は、検索結果を含む文 I D 1 4 7 における文 I D に相当する文である。文 I D が複数ある場合、その 1 文をある基準で選択してもよいし、ランダムに選択してもよい。なお、検索結果 1 5 4 は、ユーザの意図とは異なる「L I F T」に相当する検索結果である。

【 0 0 6 3 】

20

図 1 5 の表変換例 1 6 1 は、検索キー 1 5 5 で表変換例 1 5 3 をソートした例である。表変換例 1 6 1 は、検索キー 1 5 5、検索結果 1 5 7、検索件数 1 4 9、文例 1 5 9 を含んでいる。検索結果を含む文 I D 1 4 7 は、表変換例 1 6 1 からは削除されているが、対応は例えば記憶部 5 3 に記憶されることが好ましい。表変換例 1 6 1 では、同じ検索キー 1 5 5 を含む複数のセルを一つにまとめている。

【 0 0 6 4 】

図 1 6 は、表示例 1 6 3 を示している。表示例 1 6 3 は、表変換例 1 6 1 において、文例 1 5 9 を削除し、検索結果 1 5 7 毎にまとめて表示した例である。すなわち、同じ検索結果 1 5 7 が複数行ある場合、例えば、先頭行のみを残して他の行を削除されている。このとき検索件数 1 4 9 は、まとめた行に対応する数の合計数である。また、表示例 1 6 3 は、チェックボックス 1 6 5 および絞り込みボタン 1 6 7 を有している。チェックボックス 1 6 5 は、それぞれの行を選択する選択欄であり、絞り込みボタン 1 6 7 は、クリックやタッチなどにより選択を行うことで、チェックされたチェックボックス 1 6 5 に対応する行に絞り込むボタンである。

30

【 0 0 6 5 】

例えば、図 1 5 の検索結果 1 5 7 では、「太郎はあげる」に対応する行は 2 行あり、検索件数は 1 件ずつである。図 1 6 の表示例 1 6 3 の検索結果 1 5 7 では、検索件数 1 4 9 は合計して 2 件とし、1 行にまとめられている。例えば、表示例 1 6 3 において、検索結果 1 5 7 に下線 1 6 2 で示したようにリンクをつけ、このリンクを選択すると、検索された文書中の文を表示するようにしてもよい。

40

【 0 0 6 6 】

図 1 7 は、表展開例 1 7 1 を示す図である。図 1 7 に示すように、表展開例 1 7 1 は、表示例 1 6 3 において、「本をあげる」の欄のチェックボックス 1 6 5 が選択され、絞り込みボタン 1 6 7 が押下された状態を示している。このとき、選択された行は 2 行に展開され、チェックボックス 1 7 3、1 7 5 は、展開された行毎に表示されるが、ともに選択された状態を示している。チェックボックスは展開した行数だけ表示するが、全て選択された状態にする。このように選択を行うことにより、より詳細な抽出結果が表示される。ここで、「本をあげる」に対応する検索キー 1 5 5 は「あげる 本」であり、表展開例 1 7 1 の例では、斜めの文字で表示されている。

【 0 0 6 7 】

50

図18は、選択例181を示す図である。本実施の形態では、ユーザの意図は、「誰かが誰かに本をあげる（プレゼントする）」という文を検索することなので、チェックボックス183により、「本をあげる」が選択されている。すなわちユーザは、「太郎は花子に本をあげた」、「太郎は棚に本をあげた」の2つの文例を見て、意図しているのは「太郎は花子に本をあげた」と判断する。そして、「太郎は花子に本をあげた」の行のチェックボックス183が選択され、絞り込みボタンが押下される。

【0068】

以下、図19を参照しながら、クエリ21がキーワードの場合の検索処理について説明する。図19は、キーワードによる検索処理を示すフローチャートである。まず、クエリ入力部23は、クエリ21を受付ける。このときクエリ入力部23は、クエリ21が、少なくとも一つの単語を含む単語列であると判別する（S191）。

10

【0069】

キーワード入力部25は、クエリ21の単語列を単語に分割する（S192）。また、キーワード入力部25は、辞書51を参照し、それぞれの単語を意味記号に変換する（S193）。検索キー生成部29は、変換された意味記号の組み合わせを生成し、検索キーを生成する（S194）。

【0070】

キーワード検索部45は、検索用インデックス13から検索キーを含む文書および文の文書ID、文IDを取得する（S195）。キーワード検索部45は、全ての検索キーについてS195の処理が終了するまでS195を実行し（S196：NO）、終了すると（S196：YES）、検索結果の件数を計算する（S197）。

20

【0071】

出力部43は、検索結果を検索件数順に表示する（S198）。キーワード検索部45は、出力結果において、ユーザによる絞り込みが行われたことを検出すると（S199：YES）、S197に戻って処理を繰り返す。例えば、一定時間内に絞り込みが行われない場合には（S199：NO）、キーワード検索部45は、処理を終了する。

【0072】

以下、図20を参照しながら、表変換処理について説明する。図20は、表の変換処理の一例を示すフローチャートである。図20に示すように、出力部43は、表示結果の表において、検索キーの列をキーワードに変換する（S201）。例えば、出力部43は、図12の検索キー143を、図14の検索キー155のように変換する。出力部43は、検索結果の列を、表層文字列に変換する（S202）。例えば、出力部43は、図12の検索結果145を、図14の検索結果157に変換する。

30

【0073】

出力部43は、表において、文例を追加する（S203）。例えば、出力部43は、図14の表変換例153における文例159を追加する。出力部43は、検索キーで表をソートする（S204）。例えば、出力部43は、図14の検索キー155を、図15の検索キー155のようにソートする。出力部43は、例えば、表変換例161において、同じ検索キーの行については、同じ検索キーを含む複数の行を一つにまとめる（S205）。また、出力部43は、表変換例161において、各行に付き、対応する文例を例えば、記憶部53に記憶させる（S206）。出力部43は、表変換例161において、文例を削除し（S207）、各検索キー155について、検索結果157によりソートする（S208）。出力部43は、同じ検索結果157について複数の行が存在している場合には、先頭行のみを残して、他の行を削除するとともに、検索件数149を合計する（S209）。さらに、出力部43は、必要なリンク、およびチェックボックスを追加し、例えば図16の表示例163を生成する。（S210）。

40

【0074】

以上説明したように、本実施の形態による情報検索装置1によれば、クエリ入力部23が、入力されたクエリ21が、単語列であるか文であるかを判別し、それぞれに応じた処理を選択する。クエリ21が単語列である場合には、キーワード入力部25は、クエリ2

50

1の単語列を単語に分割する。キーワード変換部27は、辞書51を参照し、分割された単語をそれぞれ意味記号に変換する。検索キー生成部29は、変換された意味記号の組み合わせを生成し、検索キーを生成する。キーワード検索部45は、検索キーにマッチする意味最小単位を検索用インデックス13から抽出し、検索結果とする。出力部43は、検索結果を例えば表の形態にして出力する。出力部43は、ユーザが結果から絞込みが可能な形態で結果を出力し、ユーザの選択に基づき、結果の表示を変更する。

【0075】

クエリ21が文章の場合には、文章入力部31は、クエリ21を文に分割する。意味解析部33は、分割されたそれぞれの文について意味解析を行う。意味最小単位生成部35は、意味解析の結果に基づき、それぞれの文について意味最小単位を生成する。自然文検索部47は、意味最小単位生成部35で生成された意味最小単位を検索用インデックス13において検索し、文書ID、文ID等の検索結果を抽出する。評価値計算部39は、抽出された結果および評価値テーブル15に基づき、抽出結果の文、または文書の評価値を算出する。ランキング部41は、算出された評価値に基づき抽出結果の文または文書をソートする。出力部43は、結果を出力する。

【0076】

また、情報検索装置1は、検索対象文書DB11に新たな文書を登録し、登録した文書について意味解析を行って、意味最小単位を生成し、検索用インデックス13に登録するとともに評価値テーブル15に評価値を記憶させる機能を有する。

【0077】

以上のように、情報検索装置1によれば、クエリ21が文であっても、単語であっても、自動的に判別して検索を行うことができる。情報検索装置1によれば、クエリ21の意味解析結果に応じて、意図した文書の検索が可能である。これにより、検索の精度が向上する。また、クエリ21に含まれるキーワードの数が増えたり、文が入力されたりすることにより、ユーザの意図があいまいになり、検索結果にユーザが意図しないものが入ってしまうことを防止できる。本実施の形態では単純な例を挙げたが、この構成、アルゴリズムにより、キーワードが増えても対応できる。

【0078】

ユーザに検索結果として提示される表に、検索結果およびそれに対する検索件数が表示される。また、提示される表において、評価値や、検索件数を用いて検索結果をソートした上で表示を行うことができる。これにより、検索結果から意図した情報を抽出するためにかかる時間を削減でき、意図した情報がより簡易に検索可能となる。

【0079】

また、文に関する評価値を導入することにより、例えば、同一文内で繰り返される意味最小単位に着目した優先付けができる。例えば、特定のテーマに専門化した文が抽出されやすい。また、文書単位の評価値を導入することにより、検索対象文書全体に対する意味最小単位の評価と、意味最小単位の文中の出現様を、総合的に加味した重み付けができる。

【0080】

意味最小単位は、有向グラフの部分構造に基づくので、意味最小単位のマッチングによる検索は、有向グラフのマッチングによる検索に比べて、より柔軟に行うことができる。よって、効率よく文書を絞り込むことができ、意味的に意図通りの文書を簡単に選択できる。なお、上記実施の形態による情報検索装置1は、例えば論文や特許の検索、あるいは一般のWebページの検索などに特に有用である。

【0081】

(変形例1)

以下、図21から図26を参照しながら、変形例1について説明する。本変形例1は、検索結果の表示における変形例である。図21から図26は、検索結果の表示例を示す図である。本変形例1では、「低気圧を観測することにより、日本の気象を予測する」という文書を探したいとする。ユーザは、例えば「低気圧 観測 日本 気象 予測」という

10

20

30

40

50

キーワードを入力する。

【0082】

図21は、検索結果221を示す図である。検索結果221は、上記キーワードによる検索結果の一例である。図22は、別の検索結果223を示す図である。検索結果223は、検索結果221において、一つの検索キーに対して最も検索件数の多い抽出結果のみを表示した例である。これにより、ユーザが見る検索結果の量が削減される。検索結果223は、データベースに頻繁に出現するものを表示しているため、ユーザが欲するであろう情報をもれなく提示することができる。

【0083】

図23は、検索結果225を示す図である。検索結果225は、検索結果221において、検索キー毎の検索件数が1000以上の結果のみを表示した例である。これによっても、ユーザが見る検索結果の項目数が削減される。

【0084】

図24は、検索結果227を示す図である。検索結果227では、検索キー毎の検索件数が最も多く、かつ1000以上の項目のみを表示した例である。図25は、検索結果229を示す図である。検索結果229は、検索結果227において、全ての項目が選択された状態を示している。すなわちチェックボックス231が全て選択された状態である。検索結果229においては、ユーザは、選択を外すだけでよく、結果的にユーザの選択数が多い場合は、この表示の方が効率がよくなる。

【0085】

図26は、表示例233を示す図である。表示例233は、ユーザの意図（「低気圧を観測することにより、日本の気象を予測する」）にそって、チェックボックス235のように選択が行われた例である。これにより、ユーザの意図が正しく反映された検索結果を取得することができる。

【0086】

以上説明したように、変形例1によれば、ユーザにわかりやすく検索結果を表示し、絞り込みが容易になるような画面インタフェースが提供される。また、キーワード間の関係に基づく絞り込みが可能になり、意図する検索結果にさらに効率よくたどりつける。すなわち、単語間の意味関係に着目し、ユーザがその関係に基づいて画面インタフェースにより絞り込むことができるようになる。

【0087】

（変形例2）

以下、図27から図35を参照しながら、他言語に本発明を適用した例について説明する。本変形例2では、英語を例にして説明する。変形例2による情報検索装置1の構成および動作については、上記一実施の形態および変形例1と同様であるので、重複説明を省略する。

【0088】

図27から図29は、例えば、検索用インデックス13を生成するための準備処理としての文の解析例を示す図である。文章入力部31は、検索対象文書DB11に記憶しようとする文書が入力されると、入力された文書を一文ずつに分割する。意味解析部33は、分割されたそれぞれの文に関して、意味解析を行う。このとき、意味解析部33は、文を単語に分割してノードとし、単語間の関係を解析することによりノード間の関係、始点ノード、終点ノード、および文におけるそれぞれのノードの位置および文字列長を抽出する。意味最小単位生成部35は、意味解析の結果に基づき意味最小単位を生成する。

【0089】

図27の例では、原文263は、「She took care of Mary.」という文である。意味解析部33は、意味解析を行い、有向グラフ265および意味最小単位267を生成する。図27の例では、「SHE」、「TAKE CARE OF」、「MARY」は、ノードの例である。英語の場合には、意味記号は、文中の単語と同一としてもよい。また、英語の場合には、2語以上で一つの意味を持つ場合があるので、文を単語へ変換する場合には

10

20

30

40

50

、一語、または二語以上として変換を行う。

【0090】

図27のように、「TAKE CARE OF」のノードから「SHE」のノードに引かれているアークは、「動作主」、「TAKE CARE OF」のノードから「MARY」のノードに引かれているアークは、「対象」である。起点のノードが「TAKE CARE OF」であり、終点のノードがないアークは、「過去」および「述語」である。起点のノードがなく、終点のノードが「TAKE CARE OF」のアークは、「中心」である。

【0091】

意味解析部33は、意味最小単位の生成において、有向グラフからアークを抽出し、例えば、意味最小単位267を生成する。生成の方法は、上記一実施の形態における生成の方法と同様である。

【0092】

このように、例えば、原文263からは、意味最小単位267に記載の意味最小単位が抽出される。同様に、「Mary took a bus for San Francisco.」という原文に基づき、図28の解析例268が抽出され、「He took Mary to the school.」という原文に基づき、図29の解析例269が生成される。

【0093】

図30は、文字オフセット例271および意味記号273を示している。この例は、図27の原文263の解析例であり、例えば、文書ID=21、文番号=3の文の例である。文字オフセット例271では、「SHE」のオフセットは、「0」、文字列長は、「3」である。「TAKE CARE OF」のオフセットは、「4」、文字列長は「12」である。上記のようにして、英文についても和文と同様に、例えば原文263などを検索対象文書DB11に格納し、検索対象文書DB11に格納された文書を文毎に意味解析することにより、検索用インデックス13が作成される。

【0094】

続いて、図31から図35を参照しながら、クエリ21として英語が入力された場合の検索処理について説明する。図31は、クエリ21として、「Mary take」が入力された場合の意味解析を説明する図である。図32は、辞書テーブル279の一例を示す図である。

【0095】

図31に示すように、クエリ入力部23が、クエリ21がキーワードであることを判別すると、キーワード入力部25は、クエリ21を単語に分割する。このとき、英語の場合には、2語以上で一つの意味を持つ場合があるので、キーワード入力部25は、単語への変換を1単語または2単語以上として行う。図31の例では、キーワード入力部25は、「Mary take」を、「Mary」、「Mary take」、「take」の3つに展開する。キーワード変換部27は、展開された単語を辞書51に記憶された辞書テーブル279において参照する。参照の結果、「Mary take」は辞書テーブル279にないため、検索キー生成部29は、検索キー277のように「Mary」と「take」による意味最小単位を生成する。

【0096】

図33は、クエリ21として、「Mary take care」が入力された場合の意味解析を説明する図である。図33に示すように、クエリ入力部23が、クエリ21がキーワードであることを判別すると、キーワード入力部25は、クエリ21を単語に分割する。図33の例では、キーワード入力部25は、「Mary take care」を、「Mary」、「Mary take」、「take」、「take care」、「care」の5つに展開する。キーワード変換部27は、展開された単語を辞書51に記憶された辞書テーブル279において参照する。参照の結果、「Mary take」は辞書テーブル279にないため、検索キー生成部29は、検索キー283のように意味最小単位を生成する。

【0097】

図34は、検索結果285の一例を示す図である。図34に示すように、検索結果28

10

20

30

40

50

5 は、クエリ 2 1 が「Mary take」の場合の検索結果を示している。キーワード検索部 4 5 が、検索キー 2 7 7 に対応する文を検索対象文書 DB 1 1 から検索した結果を示している。検索結果 2 8 5 は、2 つの文が抽出されたことを示している。図 3 5 は、表示例 2 8 7 を示す図である。図 3 5 に示すように、表示例 2 8 7 は、クエリ 2 1 と、検索結果、検索件数を示し、さらに絞込みを行うためのボタンを含んでいる。

【0098】

以上説明したように、変形例 2 による情報検索装置 1 によれば、英語の少なくとも一つの単語を含むクエリ 2 1 により、英語の文書の検索が可能である。このとき、クエリ 2 1 が英語の文であっても、単語であっても自動的に判別し、和文の場合と同様に、クエリ 2 1 の意味解析を行うことにより検索が可能である。これにより、クエリ 2 1 に含まれるキーワードの数が増えたり、文が入力されたりすることにより、ユーザの意図があいまいになり、検索結果にユーザが意図しないものが入ってしまうことを防止できる。本実施の形態では単純な例を挙げたが、この構成、アルゴリズムにより、キーワードが増えても対応できる。

【0099】

また、情報検索装置 1 により、英語の文書の意味解析を行い検索用インデックス 1 3 を生成することもできる。さらに、ユーザに検索結果として提示される表において、上記位置実施の形態による情報検索装置 1 と同様、評価値を用いて検索結果をソートした上で表示を行うことができる。これにより、意図した情報がより簡易に検索可能となる。

【0100】

ここで、上記実施の形態および変形例 1、2 による情報検索方法の動作をコンピュータに行わせるために共通に適用されるコンピュータの例について説明する。図 3 6 は、標準的なコンピュータのハードウェア構成の一例を示すブロック図である。図 3 6 に示すように、コンピュータ 3 0 0 は、Central Processing Unit (CPU) 3 0 2、メモリ 3 0 4、入力装置 3 0 6、出力装置 3 0 8、外部記憶装置 3 1 2、媒体駆動装置 3 1 4、ネットワーク接続装置等がバス 3 1 0 を介して接続されている。

【0101】

CPU 3 0 2 は、コンピュータ 3 0 0 全体の動作を制御する演算処理装置である。メモリ 3 0 4 は、コンピュータ 3 0 0 の動作を制御するプログラムを予め記憶したり、プログラムを実行する際に必要に応じて作業領域として使用したりするための記憶部である。メモリ 3 0 4 は、例えば Random Access Memory (RAM)、Read Only Memory (ROM) 等である。入力装置 3 0 6 は、コンピュータの使用者により操作されると、その操作内容に対応付けられている使用者からの各種情報の入力を取得し、取得した入力情報を CPU 3 0 2 に送付する装置であり、例えばキーボード装置、マウス装置などである。出力装置 3 0 8 は、コンピュータ 3 0 0 による処理結果を出力する装置であり、表示装置などが含まれる。例えば表示装置は、CPU 3 0 2 により送付される表示データに応じてテキストや画像を表示する。

【0102】

外部記憶装置 3 1 2 は、例えば、ハードディスクなどの記憶装置であり、CPU 3 0 2 により実行される各種制御プログラムや、取得したデータ等を記憶しておく装置である。媒体駆動装置 3 1 4 は、可搬記録媒体 3 1 6 に書き込みおよび読み出しを行うための装置である。CPU 3 0 2 は、可搬記録媒体 3 1 6 に記録されている所定の制御プログラムを、記録媒体駆動装置 3 1 4 を介して読み出して実行することによって、各種の制御処理を行うようにすることもできる。可搬記録媒体 3 1 6 は、例えば Compact Disc (CD) - ROM、Digital Versatile Disc (DVD)、Universal Serial Bus (USB) メモリ等である。ネットワーク接続装置 3 1 8 は、有線または無線により外部との間で行われる各種データの授受の管理を行うインタフェース装置である。バス 3 1 0 は、上記各装置等を互いに接続し、データのやり取りを行う通信経路である。

【0103】

上記実施の形態および変形例1、2による情報検索方法をコンピュータに実行させるプログラムは、例えば外部記憶装置312に記憶させる。CPU302は、外部記憶装置312からプログラムを読み出し、コンピュータ300に情報検索の動作を行なわせる。このとき、まず、情報検索の処理をCPU302に行わせるための制御プログラムを作成して外部記憶装置312に記憶させておく。そして、入力装置306から所定の指示をCPU302に与えて、この制御プログラムを外部記憶装置312から読み出させて実行させるようにする。また、このプログラムは、可搬記録媒体316に記憶するようにしてもよい。

【0104】

なお、本発明は、以上に述べた実施の形態に限定されるものではなく、本発明の要旨を逸脱しない範囲内で種々の構成または実施形態を採ることができる。例えば、情報検索装置1の機能は、1台のコンピュータあるいは複数台のコンピュータで実現するようにしてもよい。処理フローは一例であって、処理結果が変化しない限り変更してもよい。

【0105】

情報検索装置1の構成要素は、演算処理装置上で実行されるプログラムにより実現される、機能モジュールであってもよい。図1の機能ブロック分けは一例であり、実際のプログラムモジュール構成と一致していない場合もある。更に、これらの各構成要素は、一部若しくは全部が集積された集積回路として実装されてもよい。処理の少なくとも一部を専用のモジュールとして備えた装置として実現してもよい。

【0106】

あるいは、情報検索装置1は、例えば、入出力部分をクライアント側に備え、処理および利用する情報をサーバ側に備え、ネットワークを介して接続されたシステムにより実現するようにしてもよい。さらに、サーバ側において、処理を行う装置と、情報を蓄積する装置とを別個に備えるようにしてもよい。また、情報検索装置1は、例えば、情報検索装置1の機能のそれぞれ一部を備えた複数の情報処理装置を備えたシステムでもよい。

【0107】

検索対象文書DB11、検索用インデックス13等は、例えば、検索処理を行うコンピュータとは別に設けるようにしてもよい。また、検索対象文書DB11、検索用インデックス13を生成する装置を、検索装置とは別に設けるようにしてもよい。このように別々の構成にすることにより、それぞれの装置の構成を簡易にすることが可能になる。

【0108】

上記実施の形態においては、クエリ21が文の場合に評価値を導入する例について説明したが、キーワードによる検索の場合にも、文書の評価値を算出し、ランク付けするようにしてもよい。

【0109】

なお、上記実施の形態および変形例1、変形例2において、クエリ入力部23、入力装置306は、入力部の一例である。キーワード入力部25、キーワード変換部27、検索キー生成部29、文章入力部31、意味解析部33、意味最小単位生成部35、キーワード検索部45、自然文検索部47、CPU302は、演算処理部、またはその機能の一例である。記憶部53、外部記憶装置312、可搬記憶媒体316は、記憶部の一例である。意味最小単位は、意味情報の一例である。

【0110】

以上の実施形態に関し、さらに以下の付記を開示する。

(付記1)

複数の検索用単語を含む情報の入力を受け付ける入力部と、

前記複数の検索用単語を含む情報から2つの検索用単語を分離し、検索対象文に含まれる複数の単語および前記複数の単語と他の単語間の前記検索対象文における関係を示す意味情報を前記検索対象文と関連付けて記憶した記憶部から、前記2つの検索用単語と対応する2つの単語とその意味情報を検索して抽出する演算処理部と、

抽出された前記意味情報を出力する出力部と、
を有することを特徴とする情報検索装置。

(付記2)

前記意味情報は、前記2つの単語のそれぞれに対応する意味記号を含み、
前記演算処理部は、

分離した前記検索用単語を意味記号に変換し、変換された前記意味記号のうちの2つを
検索キーとし、前記検索キーを含む前記意味情報を前記記憶部において検索することを特
徴とする付記1に記載の情報検索装置。

(付記3)

前記出力部は、前記意味情報を表層文字列に変換して出力することを特徴とする付記1
または付記2のいずれかに記載の情報検索装置。

10

(付記4)

前記演算処理部は、

前記記憶部に前記意味情報と関連付けて記憶された前記検索対象文において、前記意味
情報に含まれる前記2つの単語の少なくともいずれかが出現する出現位置を参照し、前記
出現位置に基づき前記文の少なくとも一部を抽出し、

前記出力部は、

抽出された前記検索対象文の少なくとも一部を出力する
ことを特徴とする付記1から付記3に記載の情報検索装置。

(付記5)

20

前記入力部は、抽出された前記意味情報を絞り込むための指示を受け付け、

前記出力部は、前記入力部により受け付けられた指示に応じて絞り込まれた前記意味情
報のみを出力することを特徴とする付記4に記載の情報検索装置。

(付記6)

前記入力部は、2つの検索用単語を含む情報、または少なくとも一つの文の入力を受け
付け、

前記演算処理部は、前記入力部が受け付けた入力が入力文の場合には、前記文を意味解
析することにより意味情報を生成し、前記記憶部から前記意味情報と関連付けて記憶され
た文を検索する

ことを特徴とする付記1から付記5に記載の情報検索装置。

30

(付記7)

検索対象文に含まれる複数の単語および前記複数の単語と他の単語間の前記検索対象文
における関係を示す意味情報を前記検索対象文と関連付けて記憶した前記記憶部、
をさらに有し、

前記演算処理部は、入力された文を意味解析することにより、前記記憶部に、前記意味
情報と前記文とを関連付けて記憶することを特徴とする付記1から付記6に記載の情報検
索装置。

(付記8)

情報検索装置が、

複数の検索用単語を含む情報の入力を受け付け、

40

前記複数の検索用単語を含む情報から前記2つの検索用単語を分離し、

検索対象文に含まれる複数の単語および前記複数の単語と他の単語間の前記検索対象文
における関係を示す意味情報を前記検索対象文と関連付けて記憶した記憶部から、前記2
つの検索用単語と対応する2つの単語とその意味情報を検索して抽出し、

抽出された前記意味情報を出力する、
を有することを特徴とする情報検索方法。

(付記9)

さらに、

前記意味情報は、前記2つの単語のそれぞれに対応する意味記号を含み、

分離した前記検索用単語を意味記号に変換し、

50

変換された前記意味記号のうちの２つを検索キーとし、
前記検索キーを含む前記意味情報を前記記憶部において検索する
ことを特徴とする付記 ８ に記載の情報検索方法。

(付記 １ ０)

さらに、
前記意味情報を表層文字列に変換して出力する、
ことを特徴とする付記 ８ または付記 ９ に記載の情報検索方法。

(付記 １ １)

さらに、
前記記憶部に、前記意味情報と関連付けて記憶された前記検索対象文において、前記意味情報に含まれる前記 ２ つの単語の少なくともいずれかが出現する出現位置を参照し、前記出現位置に基づき前記文の少なくとも一部を抽出し、抽出された前記検索対象文の少なくとも一部を出力する
ことを特徴とする付記 ８ または付記 ９ に記載の情報検索方法。

10

(付記 １ ２)

さらに、
抽出された前記意味情報を絞り込むための指示を受け付け、
受け付けた前記指示に応じて絞り込まれた前記意味情報のみを出力する
ことを特徴とする付記 １ １ に記載の情報検索方法。

(付記 １ ３)

20

さらに、
２つの検索用単語を含む情報、または少なくとも一つの文の入力を受け付け、
受け付けた入力の前記文の場合には、前記文を意味解析することにより意味情報を生成し、
前記記憶部から前記意味情報と関連付けて記憶された文を検索する
ことを特徴とする付記 ８ から付記 １ ２ に記載の情報検索方法。

(付記 １ ４)

さらに、
入力された文を意味解析し、前記意味解析により得られる前記文に含まれる複数の単語および前記複数の単語と他の単語間の前記文における関係を示す意味情報を前記文と関連付けて前記記憶部に記憶する
ことを特徴とする付記 ８ から付記 １ ３ に記載の情報検索方法。

30

【符号の説明】

【 ０ １ １ １ 】

- １ 情報検索装置
- １ １ 検索対象文書 Ｄ Ｂ
- １ ３ 検索用インデックス
- １ ５ 評価値テーブル
- ２ １ クエリ
- ２ ３ クエリ入力部
- ２ ５ キーワード入力部
- ２ ７ キーワード変換部
- ２ ９ 検索キー生成部
- ３ １ 文章入力部
- ３ ３ 意味解析部
- ３ ５ 意味最小単位生成部
- ３ ７ 検索部
- ３ ９ 評価値計算部
- ４ １ ランキング部
- ４ ３ 出力部

40

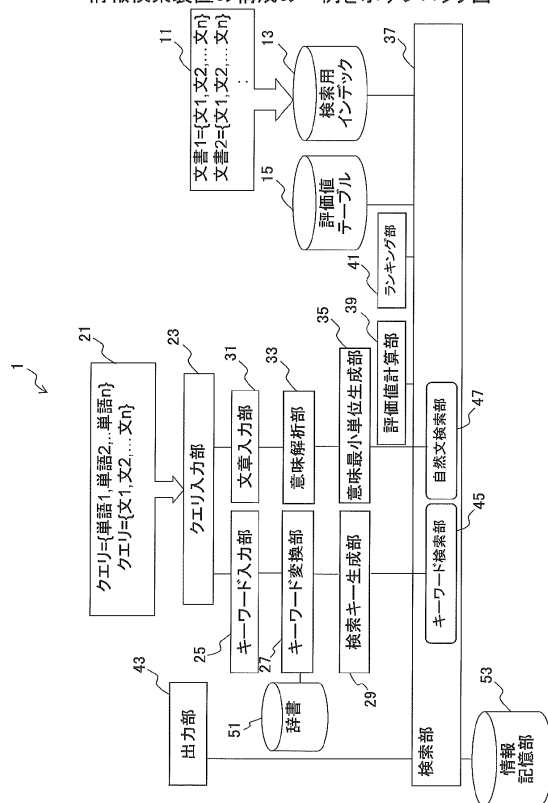
50

4 5	キーワード検索部
4 7	自然文検索部
5 1	辞書
5 3	記憶部
7 1	原文
7 3	有向グラフ
7 5	意味最小単位
7 6	解析例
7 7	解析例
7 8	文字オフセット例
7 9	意味記号
8 1	インデックステーブル
8 3	意味最小単位
8 5	文書 I D
8 7	文 I D
8 9	始点ノード位置
9 1	始点ノード文字列長
9 3	終点ノード位置
9 5	終点ノード文字列長

10

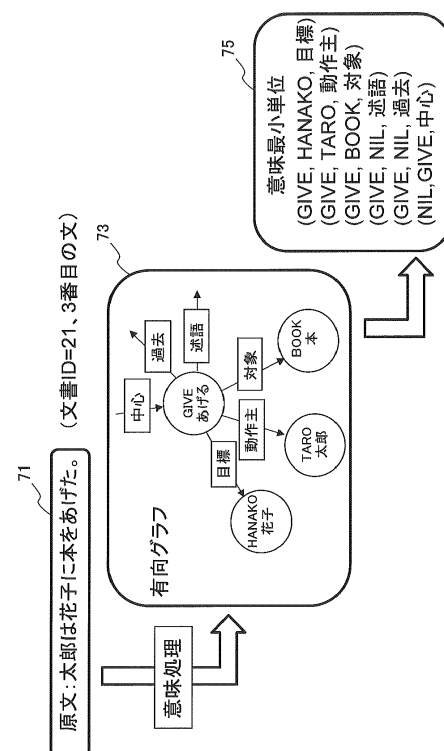
【図 1】

情報検索装置の構成の一例を示すブロック図

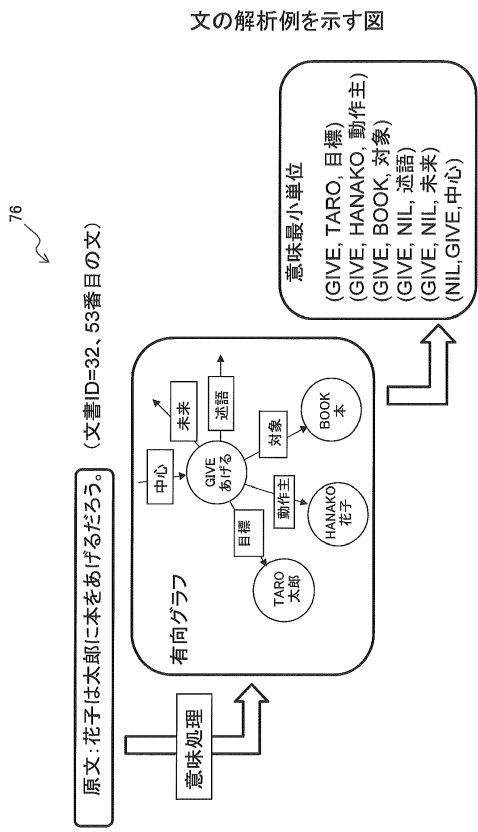


【図 2】

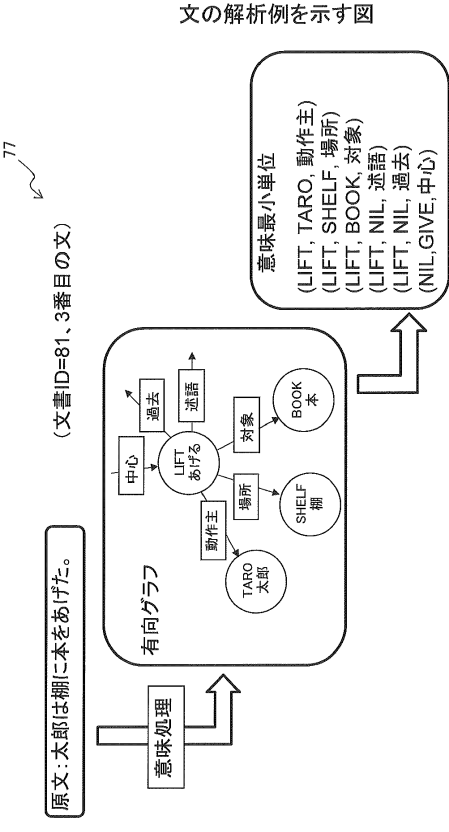
文の解析例を示す図



【図 3】

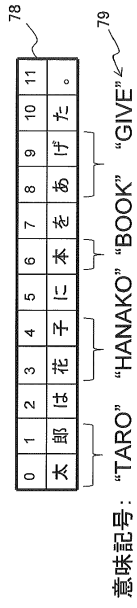


【図 4】



【図 5】

文字オフセット例および意味記号の一例を示す図



【図 6】

インデクステーブルの一例を示す図

81

意味最小単位	文書	文ID	始点ノードの位置	始点ノードの文字列長	終点ノードの位置	終点ノードの文字列長
(GIVE, HANAKO, 目標)	21	3	8	2	3	2
(GIVE, TARO, 動作主)	21	3	8	2	0	2
(GIVE, BOOK, 対象)	21	3	8	2	6	1
:	:	:	:	:	:	:
(GIVE, TARO, 目標)	32	53	8	2	3	2
(GIVE, HANAKO, 動作主)	32	53	8	2	0	2
(GIVE, BOOK, 対象)	32	53	8	2	6	1
(LIFT, TARO, 動作主)	81	3	7	2	0	2
(LIFT, SHELF, 場所)	81	3	7	2	3	1
(LIFT, BOOK, 対象)	81	3	7	2	5	1

97

98

【図 7】

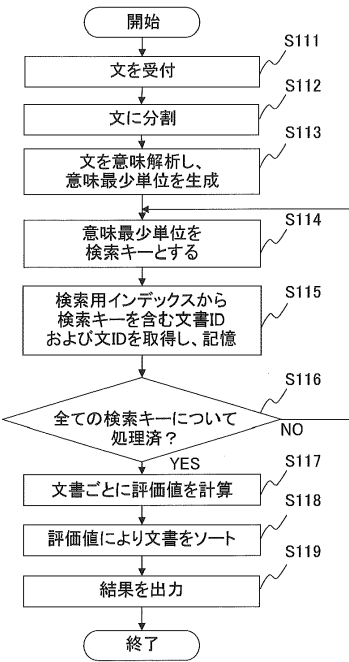
評価値テーブルの一例を示す図

97

意味最小単位	Idf値
(GIVE, HANAKO, 目標)	x . x x x x x x
(GIVE, TARO, 動作主)	x . x x x x x x
(GIVE, BOOK, 対象)	x . x x x x x x
(GIVE, NIL, 述語)	x . x x x x x x
(GIVE, NIL, 過去)	x . x x x x x x
(NIL, GIVE, 中心)	x . x x x x x x

【図 8】

クエリが文の場合の検索処理を示すフローチャート



【図 9】

クエリから分割された単語を含む単語テーブルの一例を示す図

131

単語
あげる
太郎
本

【図 1 1】

検索キーの一例を示す図

135

(GIVE, TARO, *)
(GIVE, BOOK, *)
(LIFT, TARO, *)
(LIFT, BOOK, *)
(TARO, GIVE, *)
(TARO, LIFT, *)
(TARO, BOOK, *)
(BOOK, GIVE, *)
(BOOK, LIFT, *)
(BOOK, TARO, *)

【図 1 0】

辞書テーブルの一例を示す図

51

133

辞書	
単語	意味記号
あげる	GIVE
あげる	LIFT
太郎	TARO
花子	HANAKO
本	BOOK
:	:

【図 1 2】

検索結果の一例を示す図

141 149 検索キー	143 検索結果	145 検索結果を含む文ID	147 検索件数
(GIVE, TARO, *)	(GIVE, TARO, 動作主)	(文書ID21, 文ID3)	1
(GIVE, BOOK, *)	(GIVE, TARO, 目標) (GIVE, BOOK, 対象)	(文書ID32, 文ID53) (文書ID21, 文ID3) (文書ID32, 文ID53)	1 2
(LIFT, TARO, *)	(LIFT, TARO, 動作主)	(文書ID81, 文ID3)	1
(LIFT, BOOK, *)	(LIFT, BOOK, 対象)	(文書ID81, 文ID3)	1
(TARO, GIVE, *)	なし	なし	0
(TARO, BOOK, *)	なし	なし	0
(TARO, LIFT, *)	なし	なし	0
(BOOK, GIVE, *)	なし	なし	0
(BOOK, LIFT, *)	なし	なし	0
(BOOK, TARO, *)	なし	なし	0

【図 1 3】

検索結果の表示例を示す図

「文書3件、文が3件ヒットしました。」

(文書ID21, 文ID3)
(文書ID32, 文ID53)
(文書ID81, 文ID3)

【図 1 4】

検索結果を示す表の変換例を示す図

153 154 検索キー	155 検索結果	149 検索件数	147 検索結果を含む文ID	159 文例
あげる 太郎	太郎はあげる 太郎にあげる	1 1	(文書ID21, 文ID3) (文書ID32, 文ID53)	太郎は花子に本をあげた。 花子は太郎に本をあげる だろう。
あげる 本	本をあげる	2	(文書ID21, 文ID3) (文書ID32, 文ID53)	太郎は花子に本をあげた
あげる 太郎	太郎はあげる 本をあげる	1 1	(文書ID81, 文ID3) (文書ID81, 文ID3)	太郎は棚に本をあげた。 太郎は棚に本をあげた。

【図 1 5】

検索結果を示す表の変換例を示す図

161 154 検索キー	155 検索結果	149 検索件数	159 文例
あげる 太郎	太郎はあげる 太郎にあげる	1 1	太郎は花子に本をあげた。 花子は太郎に本をあげるだろう。
あげる 本	本をあげる 本をあげる	1 2	太郎は棚に本をあげた。 太郎は花子に本をあげた
	太郎はあげる	1	太郎は棚に本をあげた。

【図 16】

検索結果を示す表の変換例を示す図

163		
155	157	149
検索キー	検索結果	検索件数
あげる 太郎	太郎はあげる	2
	太郎にあげる	1
あげる 本	本をあげる	3
		162
		167
		絞り込み

【図 17】

検索結果を示す表の変換例を示す図

149		
155	157	159
検索キー	検索結果	文例
あげる 太郎	太郎はあげる	2
	太郎にあげる	1
あげる 本	本をあげる	2太郎は花子に本をあげた。
		1太郎は棚に本をあげた。
		173
		171
		167
		絞り込み
		175

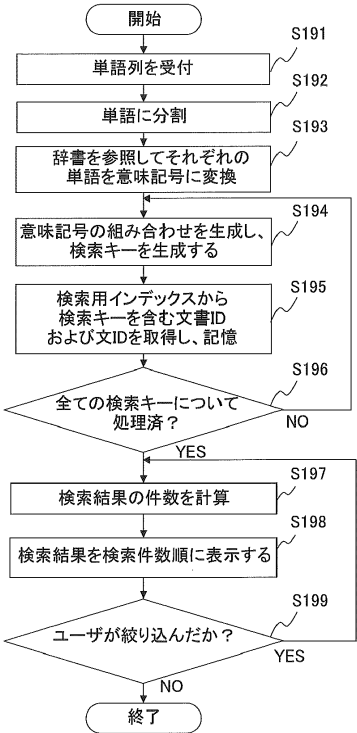
【図 18】

選択例を示す図

181		
155	157	149
検索キー	検索結果	文例
あげる 太郎	太郎はあげる	2
	太郎にあげる	1
あげる 本	本をあげる	2太郎は花子に本をあげた。
		1太郎は棚に本をあげた。
		183
		167
		絞り込み

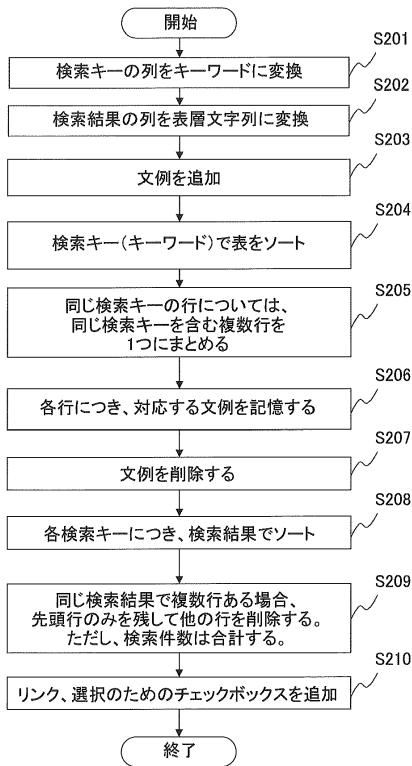
【図 19】

キーワードによる検索処理を示すフローチャート



【図 2 0】

表の変換処理の一例を示すフローチャート



【図 2 1】

変形例1による検索結果の表示例を示す図

検索結果		検索件数
低気圧 観測	低気圧を観測する	5024
低気圧 日本	低気圧から観測	35
低気圧 気象	日本にある低気圧	8009
低気圧 予想	低気圧の気象	1
観測 日本	低気圧の予想	215
	低気圧により予想する	31
	日本を観測する	910
	日本で観測する	8049
	日本が観測する	5
気象 観測	気象を観測する	10052
気象 予想	観測により予想する	1575
	観測して予想する	1056
	予測して観測する	2
日本 気象	日本の気象	10518
日本 予想	日本による予測	1
気象 予想	気象を予測する	4823
	気象により予測する	1024

【図 2 2】

変形例1による検索結果の表示例を示す図

検索結果		検索件数
低気圧 観測	低気圧を観測する	5024
低気圧 日本	日本にある低気圧	8009
低気圧 気象	低気圧の気象	1
低気圧 予想	低気圧の予想	215
観測 日本	日本で観測する	8049
観測 気象	気象を観測する	10052
観測 予想	観測により予想する	1575
日本 気象	日本の気象	10518
日本 予想	日本による予測	1
気象 予想	気象を予測する	4823

【図 2 3】

変形例1による検索結果の表示例を示す図

検索結果		検索件数
低気圧 観測	低気圧を観測する	5024
低気圧 日本	日本にある低気圧	8009
観測 日本	日本で観測する	8049
観測 気象	気象を観測する	10052
観測 予想	観測により予想する	1575
	観測して予想する	1056
日本 気象	日本の気象	10518
気象 予想	気象を予測する	4823
	気象により予測する	1024

【図 2 4】

変形例1による検索結果の表示例を示す図

検索キー		検索結果	検索件数
低気圧	観測	低気圧を観測する	5024
低気圧	日本	日本にある低気圧	8009
観測	日本	日本で観測する	8049
観測	気象	気象を観測する	10052
観測	予測	観測により予測する	1575
日本	気象	日本の気象	10518
気象	予測	気象を予測する	4823

【図 2 5】

変形例1による検索結果の表示例を示す図

検索キー	検索結果	検索件数
低気圧 観測	低気圧を観測する	5024
低気圧 日本	日本にある低気圧	8009
観測 日本	日本で観測する	8049
観測 気象	気象を観測する	10052
観測 予測	観測により予測する	1575
日本 気象	日本の気象	10518
気象 予測	気象を予測する	4823

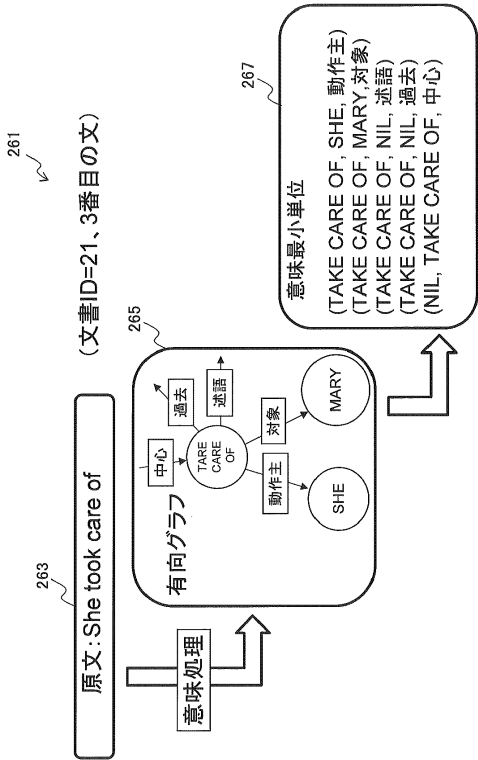
【図 2 6】

変形例1による検索結果の表示例を示す図

検索キー	検索結果	検索件数
低気圧 観測	低気圧を観測する	5024
低気圧 日本	日本にある低気圧	8009
観測 日本	日本で観測する	8049
観測 気象	気象を観測する	10052
観測 予測	観測により予測する	1575
日本 気象	日本の気象	10518
気象 予測	気象を予測する	4823

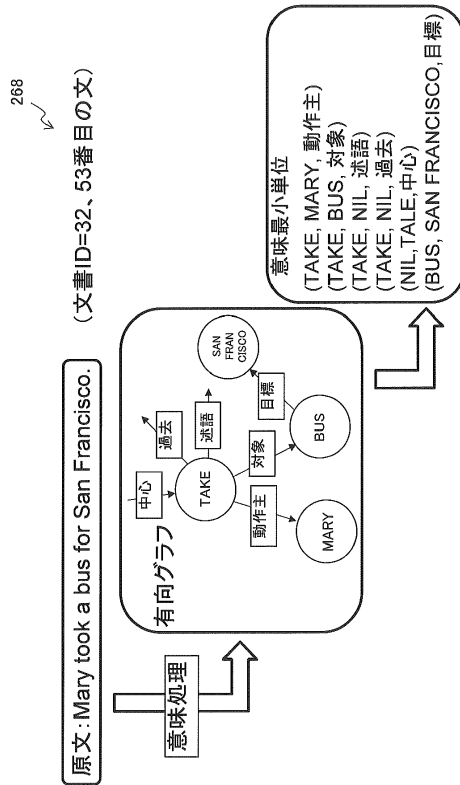
【図 2 7】

変形例2による文の解析例を示す図



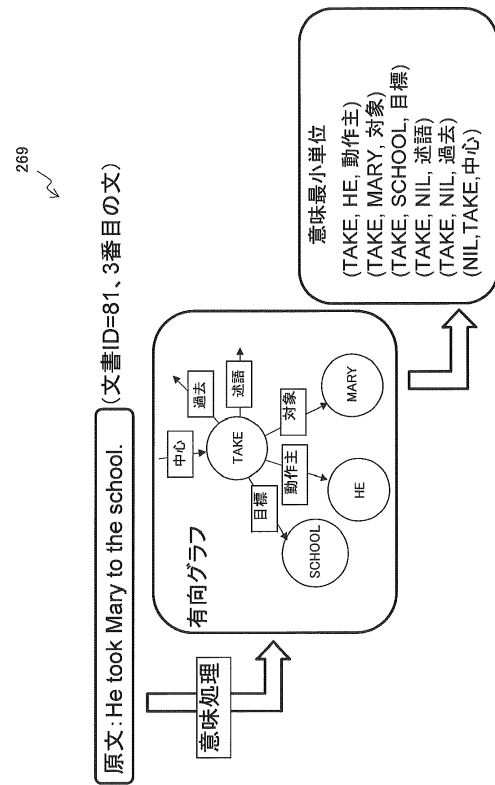
【図 28】

変形例2による文の解析例を示す図



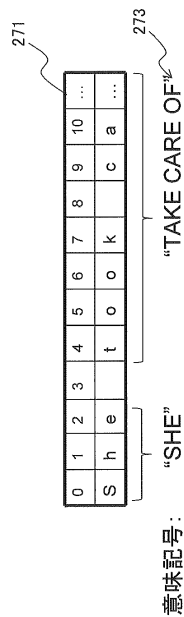
【図 29】

変形例2による文の解析例を示す図



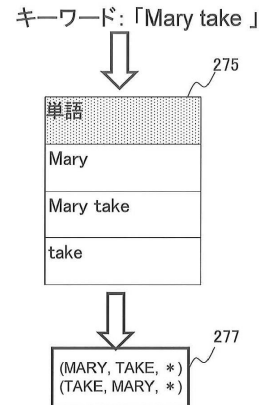
【図 30】

変形例2による文字オフセット例および意味記号を示している



【図 31】

変形例2による意味解析を説明する図



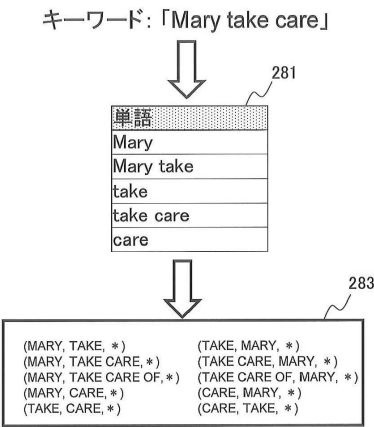
【図 3 2】

変形例2による辞書テーブルの一例を示す図

辞書	
単語	意味記号
Mary	MARY
take	TAKE
bus	BUS
San Francisco	SAN FRANCISCO
take care	TAKE CARE
take care	TAKE CARE OF
care	CARE
:	:

【図 3 3】

変形例2による意味解析を説明する図



【図 3 4】

変形例2による表示例を示す図

検索キー(意味最小単位)	検索結果	検索結果を含む文ID	頻度
(MARY, TAKE, *)	なし	なし	0
(TAKE, MARY, *)	(TAKE, MARY, 動作主)	(文書ID32, 文ID53)	1
	(TAKE, MARY, 対象)	(文書ID81, 文ID3)	1

【図 3 5】

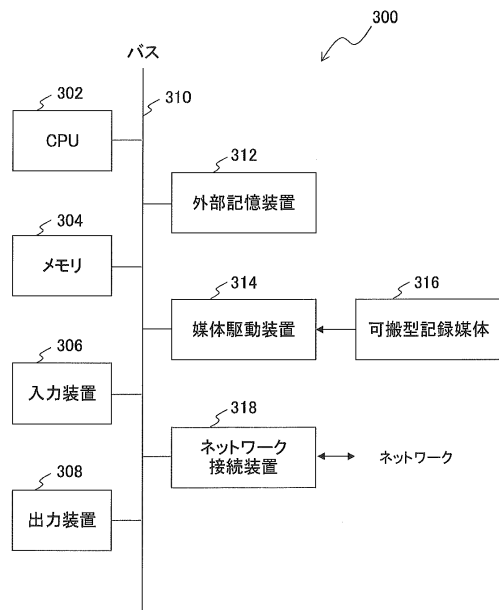
変形例2による検索結果の一例を示す図

search keyword	search result	number of matches	
Mary take	Mary take	1	<input type="checkbox"/>
	take Mary	1	<input type="checkbox"/>

select

【図 36】

標準的なコンピュータのハードウェア構成を示す図



フロントページの続き

審査官 川 崎 博章

(56)参考文献 特開 2 0 1 1 - 1 1 3 5 3 6 (J P , A)
特開 2 0 0 9 - 1 9 9 2 8 0 (J P , A)
特開 2 0 0 3 - 0 9 1 5 4 1 (J P , A)
特開 2 0 0 7 - 0 8 7 4 0 1 (J P , A)

(58)調査した分野(Int.Cl. , D B 名)
G 0 6 F 1 7 / 3 0