

(19) 日本国特許庁(JP)

(12) 公開特許公報(A)

(11) 特許出願公開番号

特開2007-241908

(P2007-241908A)

(43) 公開日 平成19年9月20日(2007.9.20)

(51) Int. Cl.	F I	テーマコード (参考)
G06F 17/30 (2006.01)	G06F 17/30 170A	5B075
G06F 17/28 (2006.01)	G06F 17/28 X	5B091

審査請求 未請求 請求項の数 6 O L (全 19 頁)

(21) 出願番号 特願2006-66829 (P2006-66829)
 (22) 出願日 平成18年3月11日 (2006.3.11)

(71) 出願人 504139662
 国立大学法人名古屋大学
 愛知県名古屋市千種区不老町1番
 (74) 代理人 100085361
 弁理士 池田 治幸
 (72) 発明者 松原 茂樹
 愛知県名古屋市千種区不老町1番 国立大
 学法人名古屋大学内
 (72) 発明者 加藤 芳秀
 愛知県名古屋市千種区不老町1番 国立大
 学法人名古屋大学内
 Fターム(参考) 5B075 ND03 PP23 UU06
 5B091 AA15 CA02 CA05 CC15

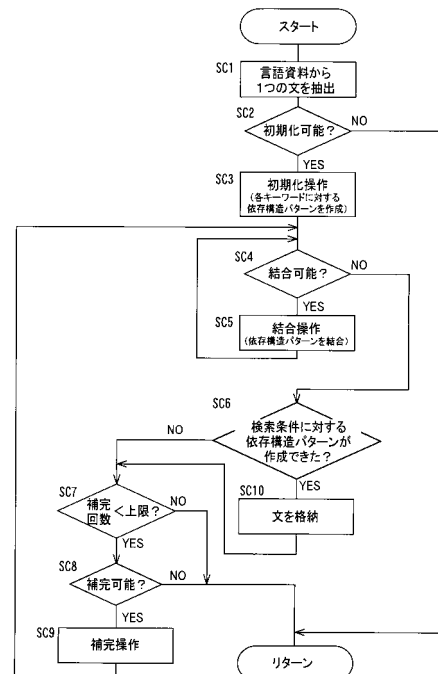
(54) 【発明の名称】 依存構造に基づく例文検索方法、プログラム、および例文検索プログラムを記録した記録媒体、ならびに例文検索装置

(57) 【要約】

【課題】 特別な表現による検索条件を必要とすることなく、蓄積された複数の例文から、検索条件に含まれる単語を構文的に含む例文を検索する例文検索方法、装置およびプログラムを提供する。

【解決手段】 予め蓄積された複数の例文と、そのそれぞれの例文を構成する単語間の依存関係を予め解析して得られた依存構造情報とをあわせて蓄積する言語資料蓄積工程と、検索しようとする文に含まれる複数の単語と該単語の順序とに基づく検索条件を設定する検索条件設定工程と、複数の単語の依存構造を表現する依存構造パターンを生成する依存構造パターン生成工程と、前記依存構造パターン生成工程によって、前記複数の例文のうち、前記検索条件に含まれるすべての単語間の依存構造を一の依存構造パターンで生成できたものを前記検索条件に該当するものとして選択する文選択工程とを有する。

【選択図】 図9



【特許請求の範囲】

【請求項 1】

蓄積された言語資料中の複数の例文と、該例文を構成する単語間の依存関係についての情報である依存構造情報とから、検索条件に該当する例文を検索する例文検索方法であって、

検索しようとする文に含まれる複数の単語と該単語の順序とに基づく検索条件を設定する検索条件設定工程と、

複数の単語の依存構造を表現する依存構造パターンを生成する依存構造パターン生成工程と、

前記依存構造パターン生成工程によって、前記複数の例文のうち、該複数の例文のそれぞれに対し、前記検索条件に含まれるすべての単語間の依存構造を一の依存構造パターンで生成できたものを前記検索条件に該当するものとして選択する文選択工程と

を有することを特徴とする例文検索方法。

10

【請求項 2】

蓄積された言語資料中の複数の例文と、該例文を構成する単語間の依存関係についての情報である依存構造情報、および、該例文を構成する単語の品詞とから、検索条件に該当する例文を検索する例文検索方法であって、

前記検索条件設定工程は、前記検索しようとする文中の複数の単語の品詞を、前記複数の単語に加えて、あるいは該複数の単語に替えて前記検索条件として設定するものであり

20

、前記依存構造パターン生成工程は、前記例文中の単語の品詞の一致性についても考慮するものであること

を特徴とする請求項 1 の例文検索方法。

【請求項 3】

請求項 1 または 2 に記載の方法をコンピュータに実行させる例文検索プログラム。

【請求項 4】

請求項 1 または 2 に記載の方法をコンピュータに実行させる例文検索プログラムを記録したコンピュータ読み取り可能な記録媒体。

【請求項 5】

蓄積された言語資料中の複数の例文と該例文を構成する単語間の依存関係についての情報である依存構造情報とから、検索条件に該当する例文を検索する例文検索装置であって

30

、検索しようとする文に含まれる複数の単語と該単語の順序とに基づく検索条件を設定する検索条件設定手段と、

複数の単語の依存構造を表現する依存構造パターンを生成する依存構造パターン生成手段と、

前記依存構造パターン生成手段によって、前記複数の例文のうち、該複数の例文のそれぞれに対し、前記検索条件に含まれるすべての単語間の依存構造を一の依存構造パターンで生成できたものを前記検索条件に該当するものとして選択する文選択手段と

を有することを特徴とする例文検索装置。

40

【請求項 6】

蓄積された言語資料中の複数の例文と、該例文を構成する単語間の依存関係についての情報である依存構造情報、および、該例文を構成する単語の品詞とから、検索条件に該当する例文を検索する例文検索装置であって、

前記検索条件設定手段は、前記検索しようとする文中の複数の単語の品詞を、前記複数の単語に加えて、あるいは該複数の単語に替えて前記検索条件として設定するものであり

、前記依存構造パターン生成手段は、前記例文中の単語の品詞の一致性についても考慮するものであること

を特徴とする請求項 5 の例文検索装置。

50

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、蓄積された言語資料中の複数の例文から、検索条件に該当する例文を検索する例文検索方法に関するものである。

【背景技術】

【0002】

正しい英文を作成するために、用例を参照し模倣することが効果的であり、大量の例文の中から、参照に値する英文を容易に見つけることは、英文作成者にとって重要である。そのため、英文作成者は、適切な英文用例を見つけるために、「手がかりとなるフレーズ」を入力し、それを含む英文が出力される英文用例検索の環境を必要としている。 10

【0003】

このような要求のもと、近年、大量の文が集積された大規模コーパス(c o r p u s ; 言語資料)の重要性はますます高まっており、言語現象の調査、外国語学習、自然言語処理システムの開発など様々な場面で言語資源として活用されている。そして、コーパスを効果的に活用するために、コーパスから任意の検索条件に該当する文を検索する様々なコーパス検索システムが提案されている。

【0004】

しかしながら、多くのシステムは、キーワードベースの用例検索を実現するにすぎず、このキーワードベースの検索は、単純で直感的であるという利点があるものの、単純なマ 20
atching、すなわち、コーパス中の各文に、検索条件として設定された単語が単に含まれるか否かによってなされる検索であり、構文構造などの言語的構造を活用した検索はできない。このため、複数の単語から成る構文を含む例文を検索したい場合であっても、その複数を単語を単に含むのみであって、目的とする構文として含まれるわけではない文も検索結果に含まれてしまうこととなる。

【0005】

【特許文献1】特開平9-265476号公報

【非特許文献1】Corley, S., Corley, M., Keller, F., Crocker, M., and Trewin, S., "Finding Syntactic Structure in Unparsed Corpora: The Gsearch Corpus Query System", Computers and the Humanities, Springer Netherlands, (オランダ), 2001年, 35巻, 2号, p. 81-94 30

【非特許文献2】Resnik, P. and Elkins, A. "The Linguist's Search Engine: An Overview" the ACL Interactive Poster and Demonstration Sessions 予稿集, (米国), Association for Computational Linguistics, 2005年, p. 33-36

【発明の開示】

【発明が解決しようとする課題】 40

【0006】

一方、これまでに、構文構造情報を利用したコーパス検索システムとしてGsearch(非特許文献1)やLinguist's Search Engine(LSE)(非特許文献2)といったシステムが提案されている。これらのシステムでは、後述する句構造を構成する品詞や当該句構造のタイプについての情報である句構造パターンを用いて、すなわち検索条件において設定されたまたは、解析によって得られた、複数の単語からなる単語列のまとまりを示す構造である句構造と、コーパスに付されたコーパスを構成する文における句構造の一致性を考慮してコーパスを検索する。Gsearchでは、ユーザは句構造パターンと文法をシステムに入力する。システムは入力された文法を用いてコーパス中の文を構文解析し、与えられた句構造パターンを持つ文を検索結果として提示す 50

る。LSEでは、ユーザはまず、探したい文の例を入力する。システムは、入力された例文を統計的構文解析により解析し、その解析結果をユーザに提示する。ユーザはこの構文解析結果を編集し、構造的なクエリを作成する。最終的にシステムは、このクエリにマッチする句構造を持つ文を検索結果として返す。これらのシステムでは、構文的情報を利用したコーパス検索を実現できる。しかし、検索にあたり、所定のパターンを入力を必要としたり、あるいは例文を入力する必要があるため、キーワードベースの検索システムのような、簡単で、直感的な検索を実現しているとは言いがたい。また、これらのシステムでは、クエリおよびコーパス中の文における複数の単語間の修飾および被修飾関係である係り受け関係が考慮されていない。

【0007】

本発明は、以上の事情を背景として為されたものであり、その目的とするところは、特別な表現による検索条件を必要とすることなく、蓄積された言語資料中の複数の例文から、検索条件に含まれる単語を構文的に含む例文を検索する例文検索方法、コンピュータが実行可能な例文検索プログラム、および、その例文検索プログラムが記憶された記録媒体ならびに例文検索装置を提供するところにある。

【課題を解決するための手段】**【0008】**

かかる目的を達成するために、請求項1に係る方法発明の要旨とするところは、蓄積された言語資料中の複数の例文と、該例文を構成する単語間の依存関係についての情報である依存構造情報とから、検索条件に該当する例文を検索する例文検索方法であって、(a)検索しようとする文に含まれる複数の単語と該単語の順序とに基づく検索条件を設定する検索条件設定工程と、(b)複数の単語の依存構造を表現する依存構造パターンを生成する依存構造パターン生成工程と、(c)前記依存構造パターン生成工程によって、前記複数の例文のうち、該複数の例文のそれぞれに対し、前記検索条件に含まれるすべての単語間の依存構造を一の依存構造パターンで生成できたものを前記検索条件に該当するものとして選択する文選択工程とを有することを特徴とする。

【発明の効果】**【0009】**

このようにすれば、検索条件は検索しようとする文に含まれる複数の単語と該単語の順序とに基づいて設定され、複数の単語の依存構造を表現する依存構造パターンの生成が試みられ、前記複数の例文のうち、該複数の例文のそれぞれに対し、前記検索条件に含まれるすべての単語間の依存構造を一の依存構造パターンで生成できたものを前記検索条件に該当するものとして選択されるので、好適に構文的に検索を行うことができる。

【0010】

また、請求項2に係る発明によれば、好適には、蓄積された言語資料中の複数の例文と、該例文を構成する単語間の依存関係についての情報である依存構造情報と、該例文を構成する単語の品詞から、検索条件に該当する例文を検索する例文検索方法であって、(d)前記検索条件設定工程は、前記検索しようとする文中の複数の単語の品詞を、前記複数の単語に加えて、あるいは該複数の単語に替えて前記検索条件として設定するものであり、(e)前記依存構造パターン生成工程は、前記例文中の単語の品詞の一致性についても考慮するものであることを特徴とする。このようにすれば、例文の検索において品詞の一致性についても考慮されることから、一層正確な例文の検索が可能となる。

【0011】

また、請求項3に係る発明の要旨とするところは、上記請求項1または2に係る方法発明をコンピュータに実行させる例文検索プログラムであることを特徴とする。

【0012】

また、請求項4に係る発明の要旨とするところは、上記請求項1または2に係る方法発明をコンピュータに実行させる例文検索プログラムが記憶された記録媒体であることを特徴とする。

【0013】

10

20

30

40

50

また、請求項 5 に係る発明の要旨とするところは、蓄積された言語資料中の複数の例文と、該例文を構成する単語間の依存関係についての情報である依存構造情報とから、検索条件に該当する例文を検索する例文検索装置であって、(a) 検索しようとする文に含まれる複数の単語と該単語の順序とに基づく検索条件を設定する検索条件設定手段と、(b) 複数の単語の依存構造を表現する依存構造パターンを生成する依存構造パターン生成手段と、(c) 前記依存構造パターン生成手段によって、前記複数の例文のうち、該複数の例文のそれぞれに対し、前記検索条件に含まれるすべての単語間の依存構造を一の依存構造パターンで生成できたものを前記検索条件に該当するものとして選択する文選択手段とを有することを特徴とする。このようにすれば、検索条件は検索しようとする文に含まれる複数の単語と該単語の順序とに基づいて設定され、複数の単語の依存構造を表現する依存構造パターンの生成が試みられ、前記複数の例文のうち、該複数の例文のそれぞれに対し、前記検索条件に含まれるすべての単語間の依存構造を一の依存構造パターンで生成できたものを前記検索条件に該当するものとして選択されるので、好適に構文的に検索を行うことができる。

10

【0014】

また、請求項 6 に係る発明によれば、好適には、蓄積された言語資料中の複数の例文と、該例文を構成する単語間の依存関係についての情報である依存構造情報と、該例文を構成する単語の品詞から、検索条件に該当する例文を検索する例文検索装置であって、(d) 前記検索条件設定手段は、前記検索しようとする文中の複数の単語の品詞を、前記複数の単語に加えて、あるいは該複数の単語に替えて前記検索条件として設定するものであり、(e) 前記依存構造パターン生成手段は、前記例文中の単語の品詞の一致性についても考慮するものであることを特徴とする。このようにすれば、例文の検索において品詞の一致性についても考慮されることから、一層正確な例文の検索が可能となる。

20

【発明を実施するための最良の形態】**【0015】**

以下、本発明の好適な実施の形態について図面を参照しつつ詳細に説明する。

【実施例】**【0016】**

図 1 は、本発明の一実施例である所謂コンピュータから主体的に構成される例文検索装置 10 を示している。この例文検索装置 10 は、よく知られた CPU、ROM、RAM、HDD、入出力インターフェース等を有するコンピュータ本体 12、キーボードなどの入力操作装置 14 および CRT 等の画像表示装置 16 等を備えたコンピュータであり、CPU は入力操作装置 14 の操作に回答して予め記憶されたプログラムを実行し、演算結果を画像表示装置 16 の画面に表示させる。

30

【0017】

入力操作装置 14 は、キーボードあるいはマウス等により構成され、後述する検索条件設定手段 30 において検索条件として設定入力される複数の単語の列をユーザが入力するのに用いられる。出力装置 16 は、例えば画像表示装置であり、後述する検索結果格納手段 36 に検索結果として格納された例文を適宜表示する。

【0018】

図 3 は、その例文検索装置 10 を、予め媒体 20 に記憶された例文検索プログラムを該媒体 20 から読み込み実行可能とすることにより得られる制御機能の要部を表す機能ブロック線図を模式的に示している。

40

【0019】

検索条件設定手段 30 は、入力操作装置 14 によって入力された単語の列およびその品詞を、所定の形式に配列し、検索条件として後述する依存構造パターン生成手段 34 に渡す。ここで、検索条件として入力される単語の一部は単語が特定されることなく単語の品詞のみが与えられてもよい。すなわち、検索条件設定手段 30 によって設定される検索条件 q は、

$$q = (q_1, q_2, \dots, q_m)$$

50

のように表現され、ここで、 q_1, q_2, \dots, q_m は m 個の単語とその品詞との組をそれぞれ示している。また、単語を特定せず品詞のみを特定する場合には単語部分にはその旨を示すデータが記載される。

【0020】

言語資料蓄積手段32は、言語資料、いわゆるコーパスが記憶されたハードディスク等の記憶装置に相当し、そこには複数の例文が蓄積されている。ここで、この複数の例文のそれぞれは、その例文を構成する単語間の依存関係が予め解析されており、依存構造情報としてその例文と共に蓄積されている。また、その例文を構成する単語の品詞についても予め解析されており、その例文と共に蓄積されている。上記複数の例文としては、たとえば、Marcus, M. P. and Santorini, B. and Marcinkiewicz, M. 著 Building a Large Annotated Corpus of English: the Penn Treebank, (Computational Linguistics誌, Vol. 19, No.2, pp.310-330, 1993) による英語新聞記事のコーパスなどが該当し、前記依存構造情報としては、このコーパスに対し、たとえば、Collins, M. 著 Head-Driven Statistical Models for Natural Language Parsing, Ph.D Dissertation, University of Pennsylvania, 1999に提案の方法に従ってコーパス中の各文を構成する単語間の依存関係を解析した結果が該当する。このようにして、言語資料蓄積手段32に蓄積された言語資料Cには複数の例文 s と、それに対応する依存構造情報 D が蓄積されている。すなわち、言語資料Cに p 個の例文が蓄積されているとき、

$$C = (s_1, D_1, s_2, D_2, \dots, s_p, D_p)$$

のように表記される。ここで例文 s は、 n 個の単語により構成されるとき、

$$s = (w_1, w_2, \dots, w_n)$$

のように表記され、ここで、 w_1, \dots, w_n は単語とその品詞との組である。

【0021】

上記言語資料Cに含まれるある例文 s について、その例文 s と依存構造情報 D 、およびその例文を構成する単語の品詞の組 w の例を図3に示す。図3のように、1つの例文に対し、その例文を構成する複数の単語間の依存関係が依存構造として矢印で示されている。また、便宜的に文頭から単語毎に1から始まる番号がその例文を構成する単語に付される。このとき、この依存構造の一つが、たとえば左から i 番目の単語が左から j 番目の単語に依存するものであるとき、この番号を用いて $i \rightarrow j$ の様に表記する。このようにすれば、図3の矢印で表された依存構造情報 D は、 $1 \rightarrow 2, 3 \rightarrow 2, 4 \rightarrow 2, 5 \rightarrow 6, 6 \rightarrow 4, 7 \rightarrow 6, 8 \rightarrow 9, 9 \rightarrow 7$ のように表現される。

【0022】

依存構造パターン生成手段34は、言語資料蓄積手段32において蓄積される複数の例文のそれぞれと、そのそれぞれの例文に対応する依存構造情報に基づいて、検索条件設定手段30において設定された検索条件を構成する複数の単語について、その複数の単語間の依存関係を、後述する依存構造パターンによって最終的に一の依存構造パターンによって記述できるか否かを試みる。

【0023】

依存構造パターンとは、検索条件設定手段30において設定された検索条件 q に含まれる複数の単語 q_1, \dots, q_m が、言語資料蓄積手段32中の言語資料Cに含まれる各例文 s_1, \dots, s_p のそれぞれにおいて、どのような依存関係を有しつつ用いられているかを示すものであり、 $d = (h, L, R)$ のように表記される。ここで、 h は単語の位置、すなわち、各例文における先頭(左)から何番目の単語であることを示す整数値であり、 L および R は依存構造パターンのリストである。 L に記載された依存構造パターンがある場合、その主辞が左から h に依存することを意味しており、 R に記載された依存構造パターンがある場合、その主辞が右から h に依存することを意味する。このとき、 L および R にそれぞれ依存構造パターンが重疊的に記載されることが可能であり、そのように記載されることにより、複数の単語の重疊的な依存関係を記載することができる。また、 h に対してそれぞれ左または右から依存する単語が存在しないときには、 L または R にはその旨を表す \emptyset が記載される。

10

20

30

40

50

【0024】

依存構造パターン生成手段34は、以下に述べる初期化操作、結合操作、および補完操作の3つの操作を行う初期化操作手段、結合操作手段、および補完操作手段を備え、それら3つの手段により、依存構造パターンの生成を試みる。初期化操作手段は、言語資料中の各例文において検索条件に含まれる複数の単語および品詞の組の間の依存関係を記述する準備として、各単語および品詞の組に対応する初期依存構造パターンを各文ごとに生成する(初期化操作)。具体的には、言語資料C中のある例文sに対して、検索条件qに含まれる単語および品詞の組 $q_i(1 \leq i \leq m)$ のそれぞれがその例文s中に一致するものがあるかを探し、その結果、例文s中の単語とその品詞との組 $w_j(1 \leq j \leq n)$ と一致するならば、 q_i に対する初期依存構造パターンとして (j, \quad, \quad) を生成する。これを検索条件qに含まれるすべての単語とその品詞との組 q_i について行う。この結果、すべての単語とその品詞との組 q_i について初期依存構造パターンの生成を行うことができれば、その例文sは、少なくとも検索条件qに含まれる単語および品詞の組をその文中に含むものであると判断され、続いて結合操作および補完操作が試みられる。一方、すべての単語とその品詞との組 q_i について初期依存構造パターンの生成を行うことができなかった場合には、その例文sは検索条件qに含まれる単語とその品詞との組 q_i の全てを含むものではないため、検索条件に合致する例文ではないと判断される。

10

【0025】

結合操作手段は、2つの依存構造パターンが存在し、それらの主辞が依存関係を有する場合に、それら2つの依存構造パターンを結合することにより1つの依存構造パターンとする(結合操作)。この操作の様子を図示したのが図4である。図4において、三角形で表された記号50は依存構造パターンを表している。具体的には、いま、検索条件の一部分である q_i, \dots, q_j および q_{j+1}, \dots, q_k に対する依存構造パターンとして、それぞれ $d = (h, L, R)$ および $d' = (h', L', R')$ が存在し、かつ、依存構造パターンdの右端の枝を可能な限りたどっていき、たどり先がなくなったところにある主辞の位置が、構造パターンd'の左端の枝を可能な限りたどっていき、たどり先がなくなったところにある主辞の位置よりも左にある場合において、(図4(a))、それらの主辞であるhおよびh'の関係が、hがh'に依存する関係である場合、すなわち $h \leq h'$ で、かつ $R' = \quad$ である場合、には、これらの2つの依存構造パターンを結合し、新たに、検索条件の一部分である q_i, \dots, q_k に対する構造パターンとして $d'' = (h', \quad, \quad)$ を生成する(図4(b))。また、h'がhに依存する関係、すなわち $h' \leq h$ の場合には、検索条件の一部分である q_i, \dots, q_k に対する構造パターンとして $d'' = (h, L, R d')$ を生成する(図4(c))。ここで、結合操作が行われる条件として $R' = \quad$ が必要とされるのは、同じ依存構造パターンを重複して生成しないようにするためである。この結合操作を繰り返すことにより、依存関係を有する複数の依存構造パターンは順次結合され、最終的に1つの依存構造パターンで記述されることとなる。

20

30

【0026】

このように、結合操作により、検索条件q中の各単語とその品詞との組のそれぞれについての初期依存構造パターンが、他の依存構造パターンと順次結合されることにより、最終的に1つの依存構造パターンとして記述される場合、それは、検索条件q中に含まれるすべての単語とその品詞との組が、言語資料C中のその例文において、依存関係を持つ形で含まれることを意味している。したがって、すべての初期依存構造パターンを最終的に1つの依存構造パターンとして記述される場合には、その文は、検索条件に合致するものとされる。

40

【0027】

続いて、補完操作手段について説明する。たとえば、単語とその品詞との組の2つを検索条件とする場合であって、その2つの単語には直接依存関係を持たない文を検索したい場合を考える。このような場合には、言語資料中のある例文について2つの初期依存構造パターンを作成できたとしても、それらが直接的には依存関係を有さないため、上述の結合操作を行うことができず、それらを1つの依存構造パターンで記述することはできない

50

。しかしながら、それら2つの依存構造パターンの主辞が、例文には含まれるものの検索条件に含まれない単語を介して依存している場合には、その単語を主辞とする依存構造パターンの生成を行うことで、結合操作が可能となることがある。このような操作を行うのが補完操作手段である。また、この補完操作の様子を示したのが図5である。具体的には、検索条件の一部である q_i, \dots, q_j に対する依存構造パターン d が存在し、その主辞を h とする。このとき、依存構造パターン d の主辞 h が依存する関係にある h' が存在するとき、すなわち $h = h'$ であるとき、 $h < h'$ 、すなわち、例文 s において h に対応する単語が h' に対応する単語よりも左にあるならば、 q_i, \dots, q_j に対する依存構造パターンとして $d^* = (h', *, d, \dots)$ を生成する(図5(a))。また、 $h > h'$ すなわち、例文 s において h に対応する単語が h' に対応する単語よりも右にあるならば、 q_i, \dots, q_j に対する依存構造パターンとして $d^* = (h', *, \dots, d)$ を生成する(図5(b))。このとき、添字 $*$ は、主辞 h' が補完操作により導入されたものであることを示す。このようにすれば、検索条件中の単語間に直接の依存関係がないような場合であっても補完操作により生成された依存構造パターンを用いて上記の結合操作を行うことによって、初期依存構造パターンから最終的に一つの依存構造パターンを生成することができることが期待できる。

【0028】

ただし、補完操作を行った場合であっても、補完操作によって生成された依存構造パターンの主辞である h' と、他の依存構造パターンの主辞との間に依存関係がない場合には、補完操作を1回行ってその後結合操作を行うことはできない。

【0029】

一方で、この補完操作を無制限に繰り返すことができれば、複数個の単語を介して間接的に依存する2つの単語についても結合操作を行うことが可能となる。このように、複数回の補完操作を経て可能となった結合操作によって一の依存構造パターンが生成された場合、たとえ最終的に一の依存構造パターンが生成されたとしても、検索条件 q に含まれるすべての単語とその品詞との組とが言語資料 C の中のその例文に含まれるとしても、検索条件 q において意図した依存関係とは異なる依存関係を持つ形で含まれる可能性がある。そして、この可能性は、補完操作を多く行う程高くなる。

【0030】

そこで、この補完操作を行う回数を制限することが行われる。この制限は、たとえば、予め最大値のみを与えておき、依存構造パターン生成手段34が自動的に結合操作を行いつつ、その制限の範囲内で適宜補完操作を行うようにすればよい。

【0031】

図6は、例文検索装置10の制御作動の要部を表すフローチャートである。以下本フローチャートに沿って例文検索装置10の作動を説明する。

【0032】

前記検索条件設定手段30および検索条件設定肯定に対応するステップ(以下「ステップ」を省略する。) $SA1$ においては、検索条件となる複数の単語が入力操作装置14等により入力され、これを上述の様式である検索条件 q とされる。本実施例においては、たとえば、入力操作装置14等により入力された検索条件 q が「 $i t$ (代名詞), $i s$ ($b e$ 動詞), $f o r$ (前置詞), $t o$ (前置詞)」であった場合を考える。これは、これら4つの単語とその品詞との組を依存関係を有する状態で文中に含む例文を言語資料蓄積手段32における言語資料 C から検索することを意味する。

【0033】

$SA2$ においては、図7または図9に示す依存構造パターン生成ルーチンが実行される。ここで、図7は依存構造パターン生成手段34が上述の初期化操作工程および結合操作工程のみを行う場合の依存構造パターン生成ルーチンであり、図9は依存構造パターン生成手段34が初期化操作工程、結合操作工程に加え、補完操作工程を行う場合の依存構造パターン生成ルーチンである。まず、図7の依存構造パターン生成ルーチンが採用される場合の例について説明する。図7における $SB1$ では言語資料蓄積手段32に対応する記

憶装置において、言語資料Cとして蓄積されている複数の例文の中から1の例文sが抽出される。また、併せて保存蓄積されている、その例文の依存構造情報およびその例文を構成する単語の品詞についての情報についても抽出される。たとえば、「It is important for us to have such technology.」という文sがその依存構造情報Dと共に抽出される。この例文と依存構造情報、および品詞についての情報は図3に示されるものである。

【0034】

続くSB2～SB6は依存構造パターン生成手段34に対応する。SB2においては、上述の初期化操作が可能であるかが判定される。すなわち、検索条件を構成する単語とその品詞との組の全てが、抽出された例文に含まれるか否かが判断される。そして、このSB2の判断が肯定されれば、前記初期化操作工程に対応するSB3において初期化操作が行われ、SB2の判断が否定された場合には、その抽出された例文は依存構造パターンを生成し得ないものであるから検索条件を満たさないものと判断され、依存構造パターン生成ルーチンが終了させられる。本実施例においては、検索条件qに含まれる4つの単語およびその品詞の組はいずれも文sに含まれることから、SB2の判断は肯定され、SB3に移る。

10

【0035】

SB3の初期化操作では、検索条件を構成する単語とその品詞との組の全てについて、初期依存構造パターンが生成される。本実施例においては、検索条件qに含まれる単語のうち、「it(代名詞)」については、文sを構成する左から1番目の単語である「It(代名詞)」と一致することから、これに対応する初期依存構造パターンとして、

20

(1, ,) . . . (1)

が生成される。同様にして、検索条件q中の「is(動詞)」、「for(前置詞)」、「to(前置詞)」に対してそれぞれ初期依存構造パターン

(2, ,) . . . (2)

(4, ,) . . . (3)

(6, ,) . . . (4)

が生成される。なお、このとき、検索条件qに含まれる単語およびその品詞と、文sに含まれる単語およびその品詞との一致を照合する際には、単語の変化、例えば、三人称単数による動詞の変化や時制による動詞の変化、単数および複数による名詞の変化についても考慮し、これらの差異は同じ単語であるように認識する。

30

【0036】

SB4においては、依存構造パターンについて、結合操作が可能かどうか判断される。すなわち、存在する依存構造パターンのそれぞれの主辞と、文sの依存構造情報とが比較され、依存関係にある2つの主辞の組があるかが判断される。SB4において依存関係にある主辞の組があると判断されれば、その依存関係にある2つの主辞についての依存構造パターンを結合すべくSB5に移る。なお、依存関係にある主辞の組が複数ある場合には、依存関係の末端にある単語、すなわち、他の単語から依存されることのない単語に相当する主辞が依存する依存関係から結合操作を実行する。一方、依存関係にある主辞がないと判断された場合には、なし得るすべての結合操作を完了したとして、SB6に移る。

40

【0037】

本実施例においては、SB3において生成された4つの初期依存構造パターンの主辞1、2、4、6について依存構造情報Dを参照すると、1 2、4 2、6 4の3つの依存関係があることがわかる。ここで、この3つの依存関係のうち、1 2および6 4の依存関係において依存元に相当する1および6には他の依存関係の依存先とはなっていない一方で、4 2については6 4の依存関係の依存先となっているため4 2の依存関係についての結合操作よりも1 2および6 4の依存関係の結合操作のほうが先に行われることとなる。この場合において、1 2の依存関係と6 4の依存関係に対応する結合操作はいずれが先に行われても良いが、例えば、文sのより左に依存元を有する依存関係に対応する結合操作を先に行うよう定めるようにすればよい。以上より、SB4におい

50

ては、1 2の依存関係に対応する初期依存構造パターン、すなわち1および2を主辞とする初期依存構造パターンを結合することができるとしてS B 5に移る。

【0038】

結合操作工程に対応するS B 5においては、S B 4において結合操作を行う様に判断された2つの依存構造パターンが結合される。本実施例においては、上記(1)で記述された1を主辞とする初期依存構造パターンと上記(2)で記述された2を主辞とする初期依存パターンが結合され、依存構造パターン

(2, (1, ,),) . . . (5)
が生成される。

【0039】

続いてS B 4に戻り、再度結合可能な依存構造パターンがないかが判断される。本実施例においては、S B 5において(1)、(2)の初期依存パターンが結合された結果、(3)、(4)の初期依存パターンおよび(5)の依存構造パターンが存在している。すなわち、その主辞は4、6、2であるから、文sにおける依存構造情報Dに基づいて6 4、4 2の依存関係があるが、上述のように4 2の依存関係よりも6 4の依存関係に対応する依存構造パターンの結合操作が先に行われる。したがって、6 4の依存関係に対応する初期依存構造パターン、すなわち6および4を主辞とする初期依存構造パターンを結合することができるとしてS B 5に移る。

【0040】

S B 5においては、先の場合と同様にして結合操作が行われる。すなわち、(3)で記述された4を主辞とする初期依存構造パターンと(4)で記述された6を主辞とする初期依存パターンが結合され、依存構造パターン

(4, (6, ,),) . . . (6)
が生成される。

【0041】

そして、再度S B 4に戻り、結合可能な依存構造パターンがないかが判断される。この段階においては、(5)および(6)で記述された2つの依存構造パターンが存在しており、その主辞は2および4である。ここで、文sの依存構造情報Dによれば4 2の依存関係があるので、この2つの依存構造パターンを結合することができるとしてS B 5に移る。

【0042】

S B 5においては、先の場合と同様にして結合操作が行われる。すなわち、(5)で記述された2を主辞とする依存構造パターンと(6)で記述された4を主辞とする依存パターンが結合され、依存構造パターン

(2, (1, ,), (4, (6, ,),)) . . . (7)
が生成される。

【0043】

S B 4に戻り、再度結合可能な依存構造パターンがないかが判断される。この段階においては、(7)で記述された1つの依存構造パターンしか残っておらず、これ以上の結合操作を行うことはできない。したがって、S B 4の判断が否定され、S B 6に移る。

【0044】

S B 6においては、S B 4 ~ S B 5における結合操作の結果、検索条件qに対する依存構造パターンが生成できたかが判断される。具体的には、複数の初期依存構造パターンから結合操作により1つの依存構造パターンが生成されたか否かによって判断される。この判断が肯定された場合には、文sは検索条件qを満たすものとして、S B 7に移る。一方、この判断が否定された場合には、文sは検索条件qを満たさないものとして、依存構造パターン生成ルーチンを終了する。本実施例においては、上記(1) ~ (4)として生成された初期依存構造パターンが、結合操作の結果、(7)で記述された1つの依存構造パターンとして生成されているため、この判断が工程され、S B 7に移る。

【0045】

10

20

30

40

50

文選択手段 36 および文選択工程に対応する S B 7 においては、S B 6 における判断が肯定された文 s について、検索条件 q を満たすものとして、その文 s が検索結果格納手段 38 に格納される。

【0046】

図 6 に戻って、S A 3 においては、言語資料 C 中の全ての文について S A 2 の依存構造パターン生成ルーチンが実行されたかが判断され、この判断が否定される場合には S A 2 に戻る。一方、この判断が肯定された場合には S A 4 に移る。これにより、言語資料 C 中のすべての文について依存構造パターンの生成を試みる。

【0047】

S A 4 においては、検索条件 q に合致する文として検索結果格納手段 38 に格納された文が出力装置 16 を通して出力される。 10

【0048】

例文検索装置 10 がこのように作動することにより、検索条件 q に含まれる単語と品詞の組が単に含まれるのみならず、それらの単語がその内部において依存関係を有する文を検索結果とすることができる。例えば、上記 S B 1 において、別の文 s ' 「It is clear whether support for the proposal will be broad enough to a serious challenge.」が抽出された場合を考える。この文を構成する単語間の依存関係と単語の品詞を表したのが図 8 である。この場合、検索条件 q の 4 つの単語および品詞の組はいずれも文 s ' に含まれることから S B 2 の判断が肯定される。したがって、S B 3 において初期化操作が行われ、 20

(1 , ,) . . . (8)
 (2 , ,) . . . (9)
 (6 , ,) . . . (10)
 (13 , ,) . . . (11)

の 4 つの初期依存構造パターンが生成される。続く S B 4 において、これらの主辞と文 s ' の依存構造情報 D を比較すると 1 2 の依存構造があることがわかるので、S B 5 において (8) および (9) の初期依存構造パターンが結合操作され、

(2 , (1 , ,) ,) . . . (12)

が生成される。 30

【0049】

続いて再度 S A 4 に戻って、他に結合可能な依存構造パターンがないかが判断されるが、このとき (10)、(11)、(12) の主辞である 6、13、2 には依存関係がなく、これ以上結合操作を行うことはできないと判断される。従って、S B 6 に移り、検索条件に対応する依存構造パターンが生成できたかが判断されるが、このとき、依存構造パターンは (10)、(11)、(12) の 3 つが存在しており、1 つの依存構造パターンで記述できていない。従って、S B 6 の判断が否定され、文 s ' は検索結果に含まれることなく依存構造パターン生成ルーチンが終了させられる。すなわち、検索条件 q に含まれるすべての単語を単に含むのみであって、それらの単語が文中において依存関係を有さない状態で存在する場合には、その文は検索結果に含まれることがない。 40

【0050】

続いて、別の実施例について説明する。以下の説明において、実施例相互に共通する部分には同一の符号を付して説明を省略する。

【0051】

図 9 は、図 6 における依存構造パターン生成ルーチンの別の作動を表すフローチャートであり、図 7 のフローチャートに代えて用いられるものである。本図のフローチャートにおいては、依存構造パターン生成ルーチンが初期化操作、結合操作に加え、補完操作を行う点において図 7 のフローチャートと相違する。なお、上述のように、補完操作を行う場合には、操作者が補完操作を行う回数の上限を予め設定しておく必要がある。本実施例では、この上限として例えば 2 回が設定される。 50

【 0 0 5 2 】

検索条件設定手段 3 0 および検索条件設定工程に対応する S A 1 においては、検索条件となる複数の単語が入力操作装置 1 4 等により入力され、これを上述の様式である検索条件 q とされる。本実施例においては、たとえば、入力操作装置 1 4 等により入力された検索条件 q ' が「 combines (動詞), and (接続詞) 」であった場合を考える。これは、これら 2 つの単語と品詞の組を依存関係を有する状態で文中に含む例文を言語資料蓄積手段 3 2 における言語資料 C から検索することを意味する。

【 0 0 5 3 】

続く S A 2 においては、依存構造パターン生成ルーチンとして図 9 のフローチャートが実行される。図 9 における S C 1 は、前述の S B 1 同様、言語資料蓄積手段 3 2 に対応し、言語資料 C として蓄積されている複数の例文の中から 1 の例文 s ' ' が抽出される。また、併せて保存蓄積されている、その例文の依存構造情報およびその例文を構成する単語の品詞についての情報についても抽出される。たとえば、「 Opera combine s music and drama . 」という文 s ' ' がその依存構造情報 D と共に抽出される。この例文と依存構造情報、および品詞についての情報は図 1 0 に示されるものである。

10

【 0 0 5 4 】

続く S C 2 ~ S C 9 は依存構造パターン生成手段 3 4 に対応する。S C 2 においては、S B 2 同様、初期化操作が可能であるかが判定される。S C 2 における判断が肯定されれば、S C 3 において初期化操作が実行され、S C 2 における判断が否定されれば、その文は検索条件を満たさないものと判断され、依存構造パターン生成ルーチンが終了させられる。本実施例においては、検索条件 q ' に含まれる 2 つの単語および品詞の組はいずれも文 s ' ' に含まれることから S C 2 の判断は肯定され、S C 3 に移る。

20

【 0 0 5 5 】

初期化操作工程に対応する S C 3 においては、S B 3 同様、初期化操作が行われる。すなわち、検索条件 q ' に含まれる単語である combines と and のそれぞれについて、初期依存構造パターン

(2 , ,) . . . (1 3)

(4 , ,) . . . (1 4)

が生成される。

30

【 0 0 5 6 】

S C 4 においては、S B 4 同様、複数の依存構造パターンの中に、結合操作が可能なものがあるかどうか判断される。本判断が肯定される場合は S C 5 に移り、結合操作工程に対応する S C 5 においては、S B 5 同様、依存構造パターンの結合操作が行われる。また、本判断が否定されれば、S C 6 に移る。本実施例においては、文 s ' ' の依存構造情報 D を参照すると、2 4 も 4 2 の依存関係もない。すなわち、初期依存パターンの主辞である 2 と 4 の間には直接の依存関係は存在しない。したがって、S C 4 の判断が否定され、S C 6 に移る。

【 0 0 5 7 】

S C 6 においては、検索条件 q ' に対する依存構造パターンが生成できたかが、複数の初期依存構造パターンから 1 つの依存構造パターンが結合操作により生成されたかによって判断される。この判断が肯定されれば、文 s ' ' は検索条件 q ' を満たすものとして S C 1 0 に移る。一方、この判断が否定された場合には、続く S C 7 に移る。本実施例においては、この時点で (1 3) および (1 4) で記述される 2 つの依存構造パターンが存在するので、S C 6 の判断は否定され S C 7 に移る。

40

【 0 0 5 8 】

S C 7 においては、これまでに行った補完操作の回数とその回数の制限値よりも小さいかを判断する。本判断が肯定されれば、さらに補完操作を行うことが可能であるとして S C 8 に移る。一方本判断が否定される場合は、文 s ' ' に対してそれ以上の補完操作は行うことができないとして、依存構造パターン生成ルーチンは終了させられる。本実施例に

50

おいては、この時点で未だ補完操作を行っておらず、また、補完操作の上限は2回と設定されていることから、S A 7における判断は肯定されS C 8に移る。

【0059】

S C 8においては、依存構造パターンにおいて、補完操作が可能かどうかを判定する。これは、文s'の依存構造情報Dをもとに、依存構造パターンの主辞ではない文s'の1つの単語を介して間接的に依存構造パターンの主辞が依存関係をとるような単語が文s'中に存在するか否かによって判断する。具体的には、本実施例においては、依存構造パターンの主辞は2および4であるが、文s'の依存構造情報Dにおいては、2も4も存在しない。しかしながら、依存構造情報Dには4-3という依存関係が存在するため、補完操作が可能であると判断される。以上のように、本実施例においては、S C 8の判断が肯定され、S C 9に移る。

10

【0060】

補完操作工程に対応するS C 9においては、S C 8において補完が可能であると判断される根拠となった依存関係に基づいて、上述の補完操作が行われる。本実施例においては、初期依存構造パターン(14)に対して、4-3という依存関係に基づいて補完操作を行うと、依存構造パターン(3*, (4,))・・・(15)が生成される。

【0061】

続いて、再びS C 4に戻り、依存構造パターンについて結合操作が可能かが判断される。本実施例においては(13)および(15)の依存構造パターンが存在しており、その主辞は2および3である。ここで、文s'の依存構造情報Dには3-2が含まれることから、これらの依存構造パターンは結合操作が可能であるとしてS C 4の判断は肯定される。

20

【0062】

S C 5においては、S C 4において結合可能であると判断された2つの依存構造パターンが結合操作により結合される。本実施例においては、(13)で記述された2を主辞とする依存構造パターンと(15)で記述された3を主辞とする依存構造パターンが結合され、依存構造パターン(2, (3*, (4,)))・・・(16)が生成される。

30

【0063】

その後、再度S C 4に戻り、さらに結合可能な依存構造パターンがないかが判断される。この段階においては、(16)で記述された1つの依存構造パターンしか残っていないため、これ以上の結合操作を行うことはできないとして、S C 4の判断は否定され、S C 6に移る。

【0064】

S C 6においては、検索条件q'に対する依存構造パターンが生成できたかが、複数の初期依存構造パターンから1つの依存構造パターンが結合操作により生成されたかによって再度判断される。本実施例においては、この時点においては、(13)および(14)で記述される2つの初期依存構造パターンから、補完操作および結合操作によって(16)で記述される1つの依存構造パターンが生成されたので、この判断が肯定され、文s'は検索条件q'を満たすものとしてS C 10に移る。

40

【0065】

文選択手段36に対応するS C 10においては、S C 6における判断が肯定された文s'について、検索条件q'を満たすものとして、その文s'を検索結果格納手段38に格納する。

【0066】

続いて、S C 7に移り、更に補完操作が可能かどうかについて、すでに行った補完回数と予め設定されたその上限値の大小により判断される。本実施例においては、すでに行っ

50

た補完操作の回数が1回である一方、補完操作の上限回数が2回と設定されていることから、本判断は肯定され、SC8に移る。

【0067】

SC8においては、先に行ったSC8と同様、依存構造パターンにおいて、補完操作が可能かどうかを判定し、本判断が肯定されればSC9に移り、再度の補完操作を実行する。一方本判断が否定されれば、これ以上の補完操作は行うことができないとして、本ルーチンは終了させられる。本実施例においては、依存構造パターンの主辞は2であるが、文s'の依存構造情報Dにおいては、2が依存する単語が存在しないことから、補完操作の対象となる主辞h'*が存在せず、補完操作をなし得ない。したがって、SC8の判断が否定され、これ以上の補完操作は行うことができないとして、本ルーチンが終了させられる。

10

【0068】

図6に戻って、SA3においては、言語資料C中の全ての文についてSA2の依存構造パターン生成ルーチンが実行されたかが判断され、この判断が否定される場合にはSA2に戻る。一方、この判断が肯定された場合にはSA4に移る。これにより、言語資料C中のすべての文について依存構造パターンの生成を試みる。

【0069】

SA4においては、検索条件q'に合致する文として検索結果格納手段38に格納された文が出力装置16を通して出力される。

【0070】

表1および表2は、本実施例の例文検索装置10と、従来技術である、検索条件qに含まれる単語を含む文を取りだす単純な方法による例文検索装置とについて、その検索結果の精度、再現率およびそれらの調和平均の値を比較するものである。ここで、この比較における言語資料Cとしては、上述のthe Penn Treebankを使用し、その言語資料中の文の依存構造情報は、上述のCollins, M. 著 Head-Driven Statistical Models for Natural Language Parsing, Ph.D Dissertation, University of Pennsylvania, 1999に提案の方法に従って解析した結果を用いた。

20

【0071】

【表1】

補完操作の上限	精度	再現率	調和平均
0回	100.0% (17/17)	81.0% (17/21)	89.5% (34/38)
1回	90.5% (19/21)	90.5% (19/21)	90.5% (38/42)
2回	70.0% (21/30)	100.0% (21/21)	82.4% (42/51)
3回	60.0% (21/35)	100.0% (21/21)	75.0% (42/56)
(従来の方法)	27.3% (21/77)	100.0% (21/21)	42.9% (42/98)

30

【0072】

表1は、検索条件を「it(代名詞), is(動詞), for(前置詞), to(前置詞)」とした場合の検索結果について、その精度、再現率およびそれらの調和平均を算出し、比較したものである。ここで、精度、再現率およびそれらの調和平均はそれぞれ

40

精度 = 検出された正解の数 / 検出結果の数

再現率 = 検出された正解の数 / 真の正解の数

調和平均 = (検出された正解の数 × 2) / (検出結果の数 + 真の正解の数)

で表される数値をパーセント表示したものであり、検索装置の検索結果の精度を表す指標として通常用いられるものである。特に調和平均(「F値」ともいう。)を用いて評価される。なお、補完操作の上限を0回としたものは、依存構造パターン生成ルーチンとして図7のフローチャートを用いる場合に相当する。

50

【0073】

表1によれば、補完操作の上限を0～3回のいずれとした場合であっても従来の手法による例文検索装置よりはよいF値が得られており、精度のよい検索が行われていることが確認される。

【0074】

【表2】

補完操作の上限	精度	再現率	調和平均
0回	0.0% (0/2)	0.0% (0/11)	0.0% (0/13)
1回	55.0% (11/20)	100.0% (11/11)	71.0% (22/31)
2回	42.3% (11/26)	100.0% (11/11)	59.5% (22/37)
3回	32.4% (11/24)	100.0% (11/11)	62.9% (22/35)
(従来の方法)	24.5% (11/45)	100.0% (11/11)	39.3% (22/56)

10

【0075】

表2は、検索条件を「combines(動詞), and(接続詞)」とした場合の検索結果について、表1と同様に比較したものである。この場合も従来の週報による例文検索装置よりは補完操作の上限に関わらずよいF値が得られ、精度のよい検索が行われていることが確認される。

【0076】

20

このようにすれば、検索条件設定手段30および検索条件設定工程SA1において設定される検索条件qは検索しようとする文に含まれる複数の単語と該単語の順序とに基づいて設定され、依存構造パターン生成手段34および依存構造パターン生成工程SB2～SB6またはSC2～SC9によって複数の単語の依存構造を表現する依存構造パターンdの生成が試みられ、前記複数の例文のうち、該複数の例文のそれぞれに対し、前記検索条件に含まれるすべての単語間の依存構造を一の依存構造パターンで生成できたものが文選択手段36および文選択工程SB7またはSC10によって前記検索条件qに該当するものとして選択されるので、好適に、予め蓄積された言語資料C中の複数の例文とその例文を構成する単語間の依存関係についての情報である依存構造情報Dとから、構文的に例文の検索を行うことができる。

30

【0077】

また、言語資料蓄積手段32は、蓄積された複数の例文のそれぞれについてその複数の例文を構成する各単語の品詞について予め解析し、その複数の例文のそれぞれと、その例文を構成する単語の品詞および単語間の依存構造Dについての情報をあわせて蓄積するものであり、検索条件設定手段30および検索条件設定工程SA1は、検索しようとする文中の複数の単語の品詞を前記検索条件として設定するものであり、さらには、該複数の単語の一部については、単語が特定されることなく品詞のみが与えられること、依存構造パターン生成手段34および依存構造パターン生成工程SB2～SB6またはSC2～SC9は、前記例文中の単語の品詞の一致性についても考慮するものであるようにすれば、例文の検索において品詞の一致性についても考慮されることから、予め蓄積された言語資料C中の複数の例文とその例文を構成する単語間の依存関係についての情報である依存構造情報D、およびその単語の品詞から、一層正確な例文の検索が可能となる。

40

【0078】

以上、本発明の実施例を図面に基づいて詳細に説明したが、本発明はその他の態様においても適用される。

【0079】

例えば、本実施例においては、検索条件として単語およびその品詞が与えられたが、その一部については単語は必要とせず品詞のみを指定してもよい。検索条件の一部として品詞のみが与えられた場合には、上述の依存構造パターン生成手段34における初期化操作手段において、その検索条件の一部である品詞が一致する単語が例文中に存在する場合に

50

、初期依存構造パターンを生成し、その後の操作は同様に行うことによって、検索を行うことができる。このようにすれば、検索条件の一部を、単語を特定せず、品詞の一致のみとすることにより、検索条件の一部をあいまいにした検索をすることが可能となる。また、検索結果として、操作者が意図しない依存構造情報を有するにもかかわらず検索結果として抽出された例文が多い場合には、その検索における検索条件に加えて、単語を特定せず品詞のみを特定する検索条件を付加して再度検索することにより、単語を特定せず品詞のみに基づく単語間の依存関係に基づいた検索が可能となり、特定の構文構造を有する例文のみに検索結果を絞り込むことができる。その結果として所望の検索結果を得ることが可能となる。

【0080】

また、逆に、品詞は与えられず、単語のみから成る検索条件であってもよい。この場合、上述の依存構造パターン生成手段34における初期化操作手段において、検索条件中に含まれる単語が例文中に含まれれば、その品詞の一致を要件とせず初期化操作を行い、その後の操作は同様に行うことによって検索を行うことができる。このとき、言語資料蓄積手段32において蓄積される言語資料Cには、その言語資料に含まれる複数の文をそれぞれ構成する単語の品詞についての情報が含まれていたが、その単語の品詞についての情報を含まないものであってもよい。

【0081】

上述の補完操作を行うにあたり、操作者が補完操作の回数の上限を予め定めておき、依存構造パターン生成手段34がその範囲内において適宜補完操作を行うようにしたが、これに限られず、例えば、依存構造パターン生成手段34がその上限を定めるようにしてもよく、あるいは、操作者が一旦定めた上限の条件下で例文検索装置10を実行し、その結果に応じて適宜その上限の値を増減させることも可能である。

【0082】

また、本実施例においては、言語資料蓄積手段32は例文検索装置10としてのコンピュータ12の内部に配置されたが、これに限られず、例えば、ネットワークを介して別の場所に設けられた他のコンピュータ内に配置されてもよい。また、入力操作装置14および出力装置16についても、例文検索装置10としてのコンピュータ12に取り付けられたものが用いられたが、これに限られず、ネットワークを介して接続された他のコンピュータの入力操作装置や出力装置が用いられてもよい。

【0083】

また、本実施例においては、例文検索装置10は英語について例文の検索を行ったが、これに限られず、上述の依存構造情報のように単語間の依存関係を記述することができる言語であれば他の言語についても適用することができる。

【図面の簡単な説明】

【0084】

【図1】本発明の実施例による例文検索装置の構成の概略を示す図である。

【図2】本発明の実施例による例文検索装置の制御機能の要部の概要を表す機能ブロック線図である。

【図3】言語資料蓄積手段に蓄積される例文と、その例文を構成する単語の依存関係および単語の品詞を表した図である。

【図4】結合操作の概要を表した図である。

【図5】補完操作の概要を表した図である。

【図6】本発明の例文検索装置の作動を表すフローチャートである。

【図7】図5における依存構造パターン生成ルーチンを表すフローチャートである。

【図8】言語資料蓄積手段に蓄積される別の例文と、その例文を構成する単語の依存関係および単語の品詞を表した図である。

【図9】図6のフローチャートに代えて用いられる、別の依存構造パターン生成ルーチンを表すフローチャートである。

【図10】言語資料蓄積手段に蓄積される別の例文と、その例文を構成する単語の依存関

10

20

30

40

50

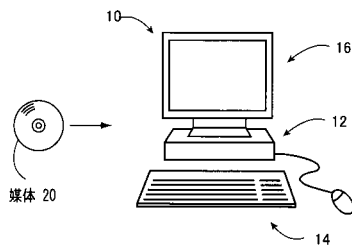
係および単語の品詞を表した図である。

【符号の説明】

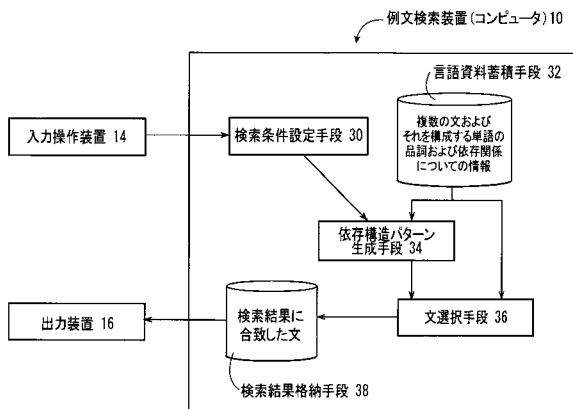
【0085】

- 10：例文検索装置（コンピュータ）
- 20：媒体
- 30：検索条件設定手段
- 34：依存構造パターン生成手段
- 36：文選択手段
- SA1：検索条件設定工程
- SB2～SB6，SC2～S9：依存構造パターン生成工程
- SB7，SC10：文選択工程

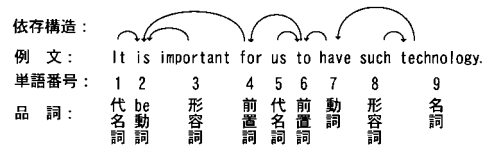
【図1】



【図2】



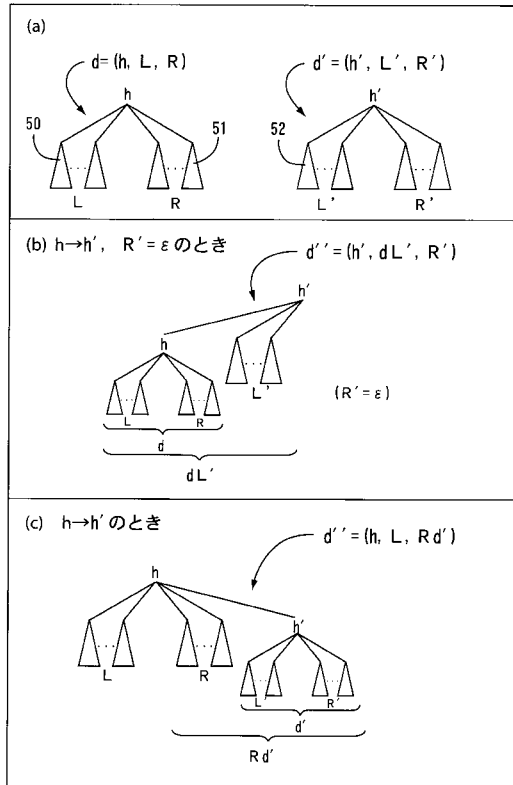
【図3】



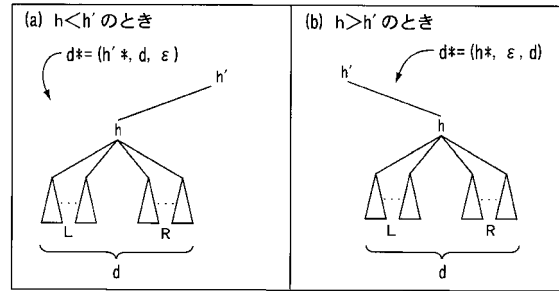
依存構造情報D: 1→2
 3→2
 4→2
 5→6
 6→4
 7→6
 8→9
 9→7

文s=(w1, w2, w3, w4, w5, w6, w7, w8, w9)
 w1= it, 代名詞
 w2= is, be動詞
 w3= important, 形容詞
 w4= for, 前置詞
 w5= us, 代名詞
 w6= to, 前置詞
 w7= have, 動詞
 w8= such, 形容詞
 w9= technology, 名詞

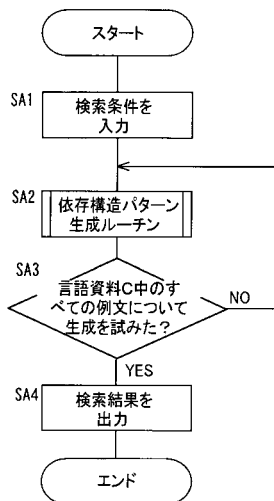
【 図 4 】



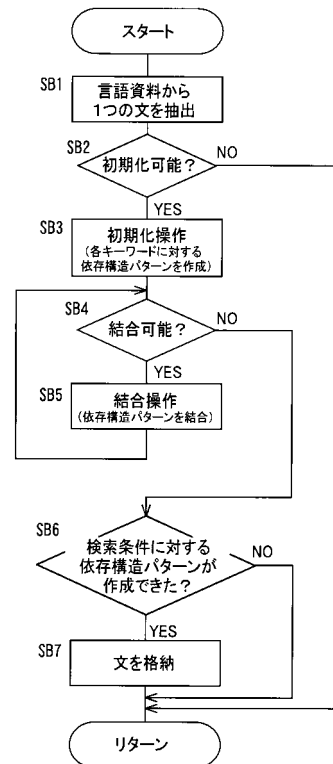
【 図 5 】



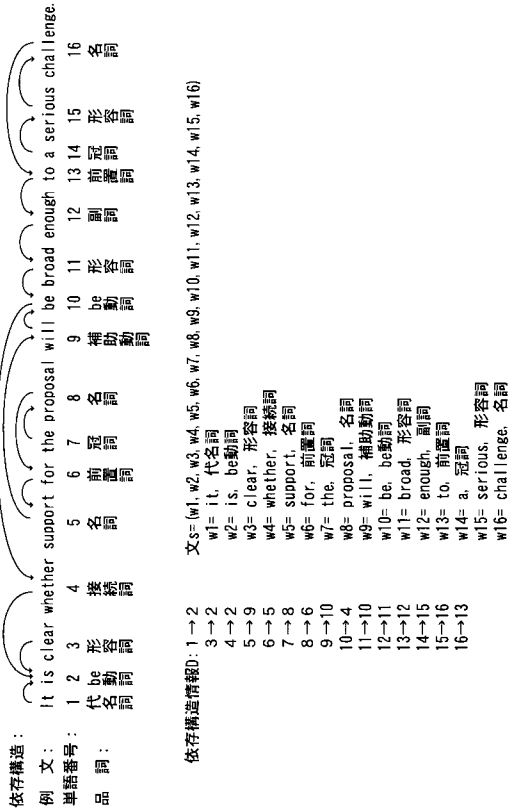
【 図 6 】



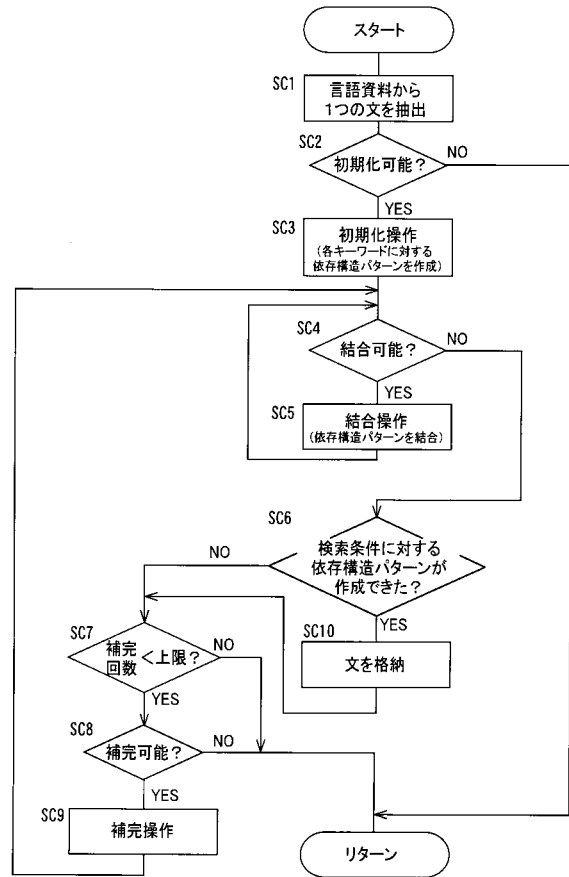
【 図 7 】



【 図 8 】



【 図 9 】



【 図 10 】

