

(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2018-13863

(P2018-13863A)

(43) 公開日 平成30年1月25日(2018.1.25)

(51) Int.Cl.		F I			テーマコード (参考)
G06F 17/27	(2006.01)	G06F 17/27	640		5B091
G06F 17/30	(2006.01)	G06F 17/30	170A		
		G06F 17/30	330C		

審査請求 未請求 請求項の数 12 O L (全 37 頁)

(21) 出願番号	特願2016-141598 (P2016-141598)	(71) 出願人	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
(22) 出願日	平成28年7月19日 (2016.7.19)	(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	片岡 正弘 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】 符号化プログラム、検索プログラム、符号化装置、検索装置、符号化方法、及び検索方法

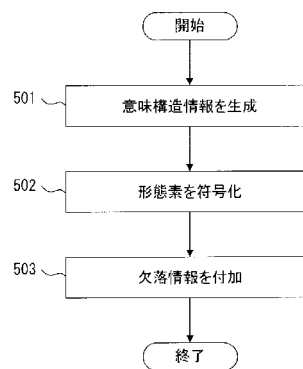
(57) 【要約】

【課題】意味構造検索において、検索対象文書から欠落している形態素の情報を補完する処理の処理量を削減する。

【解決手段】コンピュータは、文書に含まれる複数の文それぞれについて、文に含まれる複数の形態素が属する意味構造を表す意味構造情報を生成する(ステップ501)。次に、コンピュータは、複数の文それぞれについて、複数の形態素それぞれを符号化した複数の符号を生成する(ステップ502)。そして、コンピュータは、複数の文それぞれについて、各文に含まれる複数の形態素のうち、意味構造において対応する形態素の少なくとも一部が欠落している可能性のある形態素に対応する符号を特定し、特定された符号それぞれに対し、欠落情報をそれぞれ付加する(ステップ503)。

【選択図】図5

符号化処理のフローチャート



【特許請求の範囲】**【請求項 1】**

文書に含まれる複数の文それぞれについて、文に含まれる複数の形態素が属する意味構造を表す意味構造情報を生成し、

前記複数の文それぞれについて、前記複数の形態素それぞれを符号化した複数の符号を生成し、

前記複数の文それぞれについて、各文に含まれる複数の形態素のうち、意味構造において対応する形態素の少なくとも一部が欠落している可能性のある形態素に対応する符号を特定し、

前記特定された符号それぞれに対し、欠落情報をそれぞれ付加する、
処理をコンピュータに実行させる符号化プログラム。

10

【請求項 2】

前記複数の形態素が属する前記意味構造は、前記複数の形態素それぞれに対応する複数の意味記号と、前記複数の意味記号の間の接続関係とを用いて表され、前記欠落情報は、前記文書内において、前記欠落情報が付加された符号に対応する形態素と同じ意味記号に対応する形態素が、前記欠落情報が付加された前記符号に対応する形態素よりも前に出現することを示し、前記複数の文のうち前記同じ意味記号に対応する形態素が出現する文を特定する情報を含むことを特徴とする請求項 1 記載の符号化プログラム。

【請求項 3】

前記コンピュータは、前記複数の符号各々に対して、前記欠落情報、又は前記文書内において同じ意味記号に対応する形態素が前に出現しないことを示す非出現情報を付加することを特徴とする請求項 2 記載の符号化プログラム。

20

【請求項 4】

前記意味構造情報は、前記複数の意味記号各々について、前記複数の意味記号のうち他の意味記号に対する接続距離を表す接続距離情報と、前記他の意味記号に対する接続関係を表す接続関係情報とを含むことを特徴とする請求項 2 又は 3 記載の符号化プログラム。

【請求項 5】

検索要求に含まれる複数の形態素が属する意味構造を表す検索要求意味構造情報を生成し、

検索対象文書に含まれる複数の文それぞれについて文に含まれる複数の形態素が属する意味構造を表す検索対象意味構造情報と、前記複数の文それぞれについて前記複数の形態素それぞれを符号化することで生成された複数の符号と、前記複数の文それぞれについて、各文に含まれる複数の形態素のうち、意味構造において対応する形態素の少なくとも一部が欠落している可能性のある形態素に対応する符号それぞれに対して付加された欠落情報とを含む、符号化文書から、前記検索要求意味構造情報に基づいて前記検索要求に対応する文を検索し、

30

検索結果を出力する、

処理をコンピュータに実行させる検索プログラム。

【請求項 6】

前記検索要求に含まれる前記複数の形態素が属する前記意味構造は、前記検索要求に含まれる前記複数の形態素それぞれに対応する複数の意味記号と、前記検索要求に含まれる前記複数の意味記号の間の接続関係とを用いて表され、

40

前記文に含まれる前記複数の形態素が属する前記意味構造は、前記文に含まれる前記複数の形態素それぞれに対応する複数の意味記号と、前記文に含まれる前記複数の意味記号の間の接続関係とを用いて表され、

前記欠落情報は、前記検索対象文書内において、前記欠落情報が付加された符号に対応する形態素と同じ意味記号に対応する形態素が、前記欠落情報が付加された前記符号に対応する形態素よりも前に出現することを示し、前記複数の文のうち前記同じ意味記号に対応する形態素が出現する文を特定する情報を含み、

前記コンピュータは、前記符号化文書から、前記検索要求に含まれる前記複数の形態素

50

のうち1つの形態素を検索し、検索した形態素が、前記欠落情報が付加された前記符号に対応する場合、前記欠落情報を参照して、前記同じ意味記号に対応する形態素が出現する前記文を特定し、前記検索対象意味構造情報において、特定した前記文に出現する前記同じ意味記号に対応する形態素が属する第1意味構造と、前記検索対象意味構造情報において、前記欠落情報が付加された前記符号に対応する形態素が属する第2意味構造と、前記検索要求意味構造情報とを用いて、前記検索要求に対応する前記文を検索することを特徴とする請求項5記載の検索プログラム。

【請求項7】

前記複数の符号各々に対して、前記欠落情報、又は前記検索対象文書内において同じ意味記号に対応する形態素が前に出現しないことを示す非出現情報が付加されており、

前記コンピュータは、前記検索した形態素が、前記非出現情報が付加された符号に対応する場合、前記検索対象意味構造情報において、前記非出現情報が付加された前記符号に対応する形態素が属する第3意味構造と、前記検索要求意味構造情報とを用いて、前記検索要求に対応する前記文を検索することを特徴とする請求項6記載の検索プログラム。

【請求項8】

前記検索要求意味構造情報及び前記検索対象意味構造情報は、前記複数の意味記号各々について、前記複数の意味記号のうち他の意味記号に対する接続距離を表す接続距離情報と、前記他の意味記号に対する接続関係を表す接続関係情報とを含み、

前記コンピュータは、前記検索要求意味構造情報に含まれる接続距離情報及び接続関係情報と、前記検索対象意味構造情報に含まれる接続距離情報及び接続関係情報とを比較することで、前記検索要求に対応する前記文を検索することを特徴とする請求項6又は7記載の検索プログラム。

【請求項9】

文書を記憶する記憶部と、

前記文書に含まれる複数の文それぞれについて、文に含まれる複数の形態素が属する意味構造を表す意味構造情報を生成し、前記複数の文それぞれについて、前記複数の形態素それぞれを符号化した複数の符号を生成し、前記複数の文それぞれについて、各文に含まれる複数の形態素のうち、意味構造において対応する形態素の少なくとも一部が欠落している可能性のある形態素に対応する符号を特定し、前記特定された符号それぞれに対し、欠落情報をそれぞれ付加する、

符号化部と、

を備えることを特徴とする符号化装置。

【請求項10】

検索対象文書に含まれる複数の文それぞれについて文に含まれる複数の形態素が属する意味構造を表す検索対象意味構造情報と、前記複数の文それぞれについて前記複数の形態素それぞれを符号化することで生成された複数の符号と、前記複数の文それぞれについて、各文に含まれる複数の形態素のうち、意味構造において対応する形態素の少なくとも一部が欠落している可能性のある形態素に対応する符号それぞれに対して付加された欠落情報とを含む、符号化文書を記憶する記憶部と、

検索要求に含まれる複数の形態素が属する意味構造を表す検索要求意味構造情報を生成し、前記符号化文書から、前記検索要求意味構造情報に基づいて前記検索要求に対応する文を検索する検索部と、

検索結果を出力する出力部と、

を備えることを特徴とする検索装置。

【請求項11】

コンピュータが、

文書に含まれる複数の文それぞれについて、文に含まれる複数の形態素が属する意味構造を表す意味構造情報を生成し、

前記複数の文それぞれについて、前記複数の形態素それぞれを符号化した複数の符号を生成し、

10

20

30

40

50

前記複数の文それぞれについて、各文に含まれる複数の形態素のうち、意味構造において対応する形態素の少なくとも一部が欠落している可能性のある形態素に対応する符号を特定し、

前記特定された符号それぞれに対し、欠落情報をそれぞれ付加する、
ことを特徴とする符号化方法。

【請求項 1 2】

コンピュータが、

検索要求に含まれる複数の形態素が属する意味構造を表す検索要求意味構造情報を生成し、

検索対象文書に含まれる複数の文それぞれについて文に含まれる複数の形態素が属する意味構造を表す検索対象意味構造情報と、前記複数の文それぞれについて前記複数の形態素それぞれを符号化することで生成された複数の符号と、前記複数の文それぞれについて、各文に含まれる複数の形態素のうち、意味構造において対応する形態素の少なくとも一部が欠落している可能性のある形態素に対応する符号それぞれに対して付加された欠落情報とを含む、符号化文書から、前記検索要求意味構造情報に基づいて前記検索要求に対応する文を検索し、

検索結果を出力する、
ことを特徴とする検索方法。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、符号化プログラム、検索プログラム、符号化装置、検索装置、符号化方法、及び検索方法に関する。

【背景技術】

【0002】

近年、テキストデータの量が爆発的に増大しており、テキスト検索の重要性が増している。特に、秘書機能アプリケーションソフトウェア等のための意味処理の研究が盛んになってきており、自然文の意味構造の検索がますます重要になってきている。

【0003】

テキスト検索で用いられる自然文の解析には、字句解析、形態素解析、意味解析等が利用されている。字句解析は、文字列を単語に分割する処理であり、形態素解析は、文字列を形態素に分割し、各形態素に対して品詞、属性等の情報を付与する処理である。形態素解析により得られる形態素は、単語として扱われることもある。

【0004】

意味解析は、自然文の形態素解析結果を用いて、その自然文の意味構造を求める処理である。意味解析結果である意味構造を用いることで、自然文が何を意味するかを、コンピュータが扱うデータとして表現することができる。

【0005】

意味構造は、形態素解析結果に含まれる複数の単語の意味をそれぞれ表す複数の意味記号と、2つの意味記号の間の接続関係を表す情報とを含む。1つの意味記号が複数の単語に対応する場合もある。意味構造は、例えば、複数の意味記号を表す複数のノードと、2つのノード間の接続関係を表すアークとからなる有向グラフにより表すことができる。意味構造の最小の部分構造は、意味最小単位と呼ばれ、2つのノードとそれらのノードの間のアークとからなる。

【0006】

複数の文書に含まれるテキストデータに対して形態素解析及び意味解析を行うことで、自然文の検索要求の意味構造を用いて複数の文書を検索する意味構造検索を実現することができる。

【先行技術文献】

【特許文献】

10

20

30

40

50

【 0 0 0 7 】

【特許文献 1】特開 2 0 1 3 - 1 8 6 7 6 6 号公報

【特許文献 2】特開 2 0 1 0 - 9 3 4 1 4 号公報

【特許文献 3】国際公開第 2 0 1 2 / 1 1 1 0 7 8 号パンフレット

【特許文献 4】国際公開第 2 0 1 1 / 1 4 8 5 1 1 号パンフレット

【特許文献 5】特開 2 0 1 2 - 2 2 5 9 9 号公報

【特許文献 6】特開 2 0 1 2 - 1 5 0 5 8 6 号公報

【発明の概要】

【発明が解決しようとする課題】

【 0 0 0 8 】

10

主語、目的語等が省略されている等の、意味構造上で欠落のある不完全な文を対象として意味構造検索を行う場合、省略された語、すなわち、意味構造上欠落のある形態素を補完する処理が必要となるため、補完処理の分の計算量が増加する。

【 0 0 0 9 】

1つの側面において、本発明は、意味構造検索において、検索対象文書から欠落している形態素の情報を補完する処理の処理量を削減することを目的とする。

【課題を解決するための手段】

【 0 0 1 0 】

1つの案では、符号化プログラムは、以下の処理をコンピュータに実行させる。

(1) コンピュータは、文書に含まれる複数の文それぞれについて、文に含まれる複数の形態素が属する意味構造を表す意味構造情報を生成する。

20

(2) コンピュータは、複数の文それぞれについて、複数の形態素それぞれを符号化した複数の符号を生成する。

(3) コンピュータは、複数の文それぞれについて、各文に含まれる複数の形態素のうち、意味構造において対応する形態素の少なくとも一部が欠落している可能性のある形態素に対応する符号を特定し、特定された符号それぞれに対し、欠落情報をそれぞれ付加する。

【発明の効果】

【 0 0 1 1 】

1つの実施形態によれば、意味構造検索において、検索対象文書から欠落している形態素の情報を補完する処理の処理量を削減することができる。

30

【図面の簡単な説明】

【 0 0 1 2 】

【図 1】文書に含まれる第 1 文の意味構造を示す図である。

【図 2】文書に含まれる第 2 文の意味構造を示す図である。

【図 3】照応解析を用いた意味構造検索処理のフローチャートである。

【図 4】符号化装置の機能的構成図である。

【図 5】符号化処理のフローチャートである。

【図 6】検索装置の機能的構成図である。

【図 7】検索処理のフローチャートである。

40

【図 8】符号化装置の具体例を示す機能的構成図である。

【図 9】符号化文書を示す図である。

【図 10】符号化処理の具体例を示すフローチャートである。

【図 11】文書の解析結果を示す図である。

【図 12】第 1 文の意味構造に対応する接続表を示す図である。

【図 13】第 2 文の意味構造に対応する接続表を示す図である。

【図 14】アークの符号を示す図である。

【図 15】第 1 文の意味構造情報を示す図である。

【図 16】第 2 文の意味構造情報を示す図である。

【図 17】符号化文書の具体例を示す図である。

50

- 【図18】 検索装置の具体例を示す機能的構成図である。
- 【図19】 検索処理を示す図である。
- 【図20】 検索処理の具体例を示すフローチャートである。
- 【図21】 検索要求の意味構造を示す図である。
- 【図22】 検索要求の意味構造に対応する接続表を示す図である。
- 【図23】 検索要求意味構造情報を示す図である。
- 【図24】 意味構造情報マッチング処理のフローチャートである。
- 【図25】 第1文の「サーバー」に対する意味構造情報マッチング処理を示す図である。
- 【図26】 第1文の「起動し」に対する意味構造情報マッチング処理を示す図である。
- 【図27】 第2文の「起動し」に対する意味構造情報マッチング処理を示す図である。
- 【図28】 否定フラグを含む付加情報を示す図である。
- 【図29】 文の末尾から順に符号を並べた符号化文書を示す図である。
- 【図30】 一連の符号の後に付加情報を並べた符号化文書を示す図である。
- 【図31】 情報処理装置の構成図である。
- 【発明を実施するための形態】

10

【0013】

以下、図面を参照しながら、実施形態を詳細に説明する。

先願である特願2015-8936号に記載された技術（以下では、先願技術と記載する）では、文書に対する形態素解析により得られた形態素と、意味解析により得られた意味記号とを併せた一連の符号を用いて、文書が符号化される。これにより、形態素解析結果と意味解析結果との対応関係を符号化文書に効果的に埋め込むことができ、意味記号を用いて直接符号化文書を検索することが可能になる。

20

【0014】

しかしながら、先願技術では、検索対象文書に含まれる文及び検索要求に含まれる文が、主語、目的語等が省略されていない完全な文であることを前提としている。このため、主語、目的語等が省略されている不完全な文を対象として意味構造検索を行う場合、何らかの問題が生じる。

【0015】

例えば、「パッチはサーバーに当てて再起動した。すると、起動しなかった。」という文書が検索対象である場合、この文書は以下の2文を含んでいる。

30

第1文：「パッチはサーバーに当てて再起動した。」

第2文：「すると、起動しなかった。」

【0016】

図1は、第1文の意味構造の例を示している。第1文は、「パッチ」、「は」、「サーバー」、「に」、「当て」、「て」、「再」、「起動し」、「た」、及び「。」の10個の形態素を含む。このうち、「パッチ」、「サーバー」、「当て」、「再」、及び「起動し」に対応する意味記号は、それぞれ、PA001、SE001、AP001、AG001、及びBO001である。

【0017】

BO001とAG001の間の接続関係は<修飾>であり、BO001とAP001の間の接続関係は<順接>である。AP001とSE001の間の接続関係は<目標>であり、AP001とPA001の間の接続関係は<対象>である。さらに、BO001は、<中心>を表すアークを有する。

40

【0018】

図1の意味構造においては、「サーバー」は「起動し」の直接の目的語ではない。このため、例えば、「サーバーを再起動する」という自然文の検索要求に対して単純にマッチしない。

【0019】

図2は、第2文の意味構造の例を示している。第2文は、「すると」、「起動し」、「なかつ」、「た」、及び「。」の5個の形態素を含む。このうち、「すると」及び「起動

50

し」に対応する意味記号は、それぞれ、TH001及びB0001である。B0001とTH001の間の接続関係は<順接>である。さらに、B0001は、<中心>、<過去>、及び<否定>を表すアークを有する。

【0020】

第2文の「起動し」の主語は、第1文の「サーバー」であるが、第2文では省略されている。しかし、先願技術における意味構造検索の単位は文であるため、第2文内では主語が見つからない。

【0021】

そこで、意味構造検索で検索要求として与えられる自然文を想定して、事前にその自然文に含まれる主語、目的語等を意味構造に展開しておく方法が考えられる。しかし、すべての検索要求のバリエーションを事前に推定することは極めて困難であり、仮にバリエーションが推定できたとしても、展開された意味構造の情報量が膨大になるため、この方法は現実的とは言えない。

【0022】

一方、検索要求に応じて検索対象となる意味構造を柔軟にマッチングできるようにする方法も考えられる。この場合、省略された主語、目的語等は、照応解析によって補完される。

【0023】

照応解析は、文に含まれる代名詞の指す語句、文中で省略されている省略語等を同定する処理である。例えば、動詞に対する主語が省略されている場合、その文よりも前の文に出現する単語の中から、主語になり得る単語が検索される。そして、検索した単語と動詞の解析結果を再度検証して、その単語が動詞の主語であるか否かを判定することで、省略された主語を同定することができる。

【0024】

しかし、照応解析は難しい処理であり、正しい解析結果が得られる確率は約60～70%にとどまり、計算量も膨大になる。このため、意味構造検索と照応解析とを並行して行うと、処理に用いるリソースが膨大になる。

【0025】

図3は、照応解析を用いた意味構造検索処理の例を示すフローチャートである。まず、検索装置は、検索要求に含まれる意味最小単位が検索対象文書の意味構造にマッチするかどうかをチェックする(ステップ301)。意味最小単位が検索対象文書の意味構造にマッチした場合(ステップ301, YES)、検索装置は、意味最小単位のスコアを所定値に設定し(ステップ302)、その意味最小単位のスコアを検索対象文書のスコアに加算する(ステップ303)。

【0026】

一方、意味最小単位が検索対象文書の意味構造にマッチしない場合(ステップ301, NO)、検索装置は、照応解析を行って(ステップ304)、単語が省略されている否かをチェックする(ステップ305)。単語が省略されている場合(ステップ305, YES)、検索装置は、意味最小単位のスコアを所定値×0.5に設定して(ステップ306)、ステップ303の処理を行う。単語が省略されていない場合(ステップ305, NO)、検索装置は、意味最小単位のスコアを加算せずに、処理を終了する。

【0027】

このように、主語の省略等の現象に対応するためには、検索要求に対するマッチングとは別に照応解析が行われ、マッチングと照応解析の処理が分離される。検索対象文書にN個の意味記号が含まれている場合、1回の照応解析の計算量はN×Nのオーダーになる。図3は、検索キーとなる意味最小単位が1個の場合の意味構造検索を示しているが、実際の意味構造検索では数十個の検索キーが用いられることもある。その場合、ステップ301～ステップ306の処理が検索キーの個数と同じ回数だけ繰り返され、計算量がさらに増加する。

【0028】

10

20

30

40

50

図4は、実施形態の符号化装置の機能的構成例を示している。符号化装置401は、記憶部411及び符号化部412を含む。記憶部411は、文書421を記憶する。符号化部412は、記憶部411が記憶する文書421を符号化する。

【0029】

図5は、図4の符号化装置401が行う符号化処理の例を示すフローチャートである。まず、符号化部412は、文書421に含まれる複数の文それぞれについて、文に含まれる複数の形態素が属する意味構造を表す意味構造情報を生成する(ステップ501)。次に、符号化部412は、複数の文それぞれについて、複数の形態素それぞれを符号化した複数の符号を生成する(ステップ502)。そして、符号化部412は、複数の文それぞれについて、各文に含まれる複数の形態素のうち、意味構造において対応する形態素の少なくとも一部が欠落している可能性のある形態素に対応する符号を特定し、特定された符号それぞれに対し、欠落情報をそれぞれ付加する(ステップ503)。

10

【0030】

図4の符号化装置401によれば、意味構造検索において、検索対象文書から欠落している形態素の情報を補完する処理の処理量を削減することができる。

【0031】

図6は、実施形態の検索装置の機能的構成例を示している。検索装置601は、記憶部611、検索部612、及び出力部613を含む。記憶部611は、検索対象意味構造情報、複数の符号、及び欠落情報を含む符号化文書621を記憶する。

【0032】

検索対象意味構造情報は、検索対象文書に含まれる複数の文それぞれについて、文に含まれる複数の形態素が属する意味構造を表す。複数の符号は、複数の文それぞれについて、複数の形態素それぞれを符号化することで生成される。欠落情報は、複数の文それぞれについて、各文に含まれる複数の形態素のうち、意味構造において対応する形態素の少なくとも一部が欠落している可能性のある形態素に対応する符号それぞれに対して付加される。

20

【0033】

例えば、図4の文書421が検索対象文書である場合、符号化装置401によって文書421を符号化することで、符号化文書621が生成される。検索部612は、検索要求に基づいて、記憶部411が記憶する符号化文書621を検索し、出力部613は、検索結果を出力する。

30

【0034】

図7は、図6の検索装置601が行う検索処理の例を示すフローチャートである。まず、検索部612は、検索要求に含まれる複数の形態素が属する意味構造を表す検索要求意味構造情報を生成し(ステップ701)、符号化文書621から、検索要求意味構造情報に基づいて検索要求に対応する文を検索する(ステップ702)。そして、出力部613は、検索結果を出力する(ステップ703)。

【0035】

図6の検索装置601によれば、意味構造検索において、検索対象文書から欠落している形態素の情報を補完する処理の処理量を削減することができる。

40

【0036】

図8は、図4の符号化装置401の具体例を示している。図8の符号化装置401は、記憶部411、符号化部412、解析部801、及び出力部802を含む。記憶部411は、文書421及び符号化情報811を記憶する。符号化情報811は、形態素と符号との対応関係を表し、符号化辞書と呼ばれることもある。

【0037】

解析部801は、記憶部411が記憶する文書421に対して、形態素解析及び意味解析を行い、形態素解析結果及び意味解析結果を含む解析結果812を生成して、記憶部411に格納する。符号化部412は、符号化情報811及び解析結果812を用いて文書421を符号化することで、符号化文書621を生成し、記憶部411に格納する。出力

50

部 8 0 2 は、符号化文書 6 2 1 を図 6 の検索装置 6 0 1 へ出力する。

【 0 0 3 8 】

図 9 は、符号化文書 6 2 1 の例を示している。図 9 の符号化文書 6 2 1 は、符号 9 0 1 - 1 ~ 符号 9 0 1 - N、付加情報 9 0 2 - 1 ~ 付加情報 9 0 2 - N、及び意味構造情報 9 0 3 - 1 ~ 意味構造情報 9 0 3 - M を含む。N は、文書 4 2 1 に含まれる形態素の個数を表し、M は、意味構造におけるノードに対応する形態素の個数を表す。したがって、M N である。

【 0 0 3 9 】

符号 9 0 1 - i (i = 1 ~ N) は、符号化情報 8 1 1 によって各形態素に対応付けられた符号であり、付加情報 9 0 2 - i は、各符号に対して付加された情報である。意味構造情報 9 0 3 - j (j = 1 ~ M) は、各形態素が属する意味構造を表す情報である。

10

【 0 0 4 0 】

符号 9 0 1 - i は、文書 4 2 1 内の先頭の文から末尾の文へ向かって、文毎にまとめて配置され、各文に対応する複数の符号 9 0 1 - i は、その文の先頭の符号 9 0 1 - i から順に並べられている。そして、符号 9 0 1 - 1 ~ 符号 9 0 1 - N に続いて、意味構造情報 9 0 3 - j が符号 9 0 1 - i と同じ順序で配置される。

【 0 0 4 1 】

図 1 0 は、図 8 の符号化装置 4 0 1 が行う符号化処理の具体例を示すフローチャートである。まず、解析部 8 0 1 は、文書 4 2 1 に対して形態素解析を行って形態素解析結果を生成し (ステップ 1 0 0 1)、文書 4 2 1 に対して意味解析を行って意味解析結果を生成する (ステップ 1 0 0 2)。そして、解析部 8 0 1 は、形態素解析結果及び意味解析結果を含む解析結果 8 1 2 を記憶部 4 1 1 に格納する。

20

【 0 0 4 2 】

次に、符号化部 4 1 2 は、文書 4 2 1 内の各文に含まれる各形態素の意味構造情報を生成し (ステップ 1 0 0 3)、符号化情報 8 1 1 を用いて、各形態素を対応する符号に置き換えることで、各形態素を符号化する (ステップ 1 0 0 4)。そして、符号化部 4 1 2 は、各形態素の付加情報を生成し、対応する符号に対して付加する (ステップ 1 0 0 5)。これにより、符号、付加情報、及び意味構造情報を含む符号化文書 6 2 1 が生成される。

【 0 0 4 3 】

図 1 1 は、解析結果 8 1 2 の例を示している。図 1 1 の解析結果 8 1 2 は、上述した「パッチはサーバーに当てて再起動した。すると、起動しなかった。」という文書 4 2 1 の解析結果を表している。文番号は、文書 4 2 1 に含まれる文の識別情報であり、文番号“ 1 ”及び文番号“ 2 ”は、それぞれ、第 1 文及び第 2 文に対応する。以下では、第 1 文及び第 2 文を、それぞれ、文 1 及び文 2 と記載することがある。

30

【 0 0 4 4 】

文頭からの位置は、各文に含まれる形態素の識別情報であり、文頭から何番目の形態素であるかを表す。例えば、文 1 の「パッチ」は、1 番目の形態素であり、「サーバー」は、3 番目の形態素であり、文 2 の「起動し」は、2 番目の形態素である。

【 0 0 4 5 】

以下では、文番号 p の文の文頭から q 番目の位置に存在する形態素を、[p , q] 「形態素」と表記することがある。例えば、文 1 の「パッチ」は、[1 , 1] 「パッチ」と表記され、「サーバー」は、[1 , 3] 「サーバー」と表記され、文 2 の「起動し」は、[2 , 2] 「起動し」と表記される。

40

【 0 0 4 6 】

付加情報は、図 1 0 のステップ 1 0 0 5 において生成される付加情報であり、参照フラグ及び照応情報を含む。参照フラグは、論理“ 1 ”のとき、符号化文書 6 2 1 内において、各形態素に対応する意味構造情報が存在することを示し、論理“ 0 ”のとき、その意味構造情報が存在しないことを示す。

【 0 0 4 7 】

例えば、[1 , 1] 「パッチ」の参照フラグ“ 1 ”は、対応する意味構造情報が存在す

50

ることを示し、[1 , 2] 「は」の参照フラグ“ 0 ”は、対応する意味構造情報が存在しないことを示す。

【 0 0 4 8 】

照応情報は、照応フラグ f 及び位置情報 X を用いて、“ f ”又は“ $f : X$ ”の形式で記述される。照応フラグ f は、論理“ 1 ”のとき、文書 4 2 1 内において、符号に対応する形態素と同じ意味記号に対応する形態素が、その符号に対応する形態素よりも前に出現することを示す。この場合、照応情報は“ $f : X$ ”の形式で記述され、位置情報 X を含む。

【 0 0 4 9 】

位置情報 X は、文書 4 2 1 内において、同じ意味記号に対応する形態素が出現する文を特定する情報であり、 X 文前の文に同じ意味記号に対応する形態素が出現することを表す。照応情報“ $1 : X$ ”は、欠落情報として符号に付加される。

10

【 0 0 5 0 】

また、照応フラグ f は、論理“ 0 ”のとき、同じ意味記号に対応する形態素が前に出現しないことを示す。この場合、照応情報は“ f ”の形式で記述され、照応情報“ 0 ”は、非出現情報として符号に付加される。

【 0 0 5 1 】

例えば、[1 , 1] 「パッチ」の照応情報“ 0 ”は、「パッチ」と同じ意味記号に対応する形態素が前に出現しないことを示す。また、[2 , 2] 「起動し」の照応情報“ $1 : 1$ ”は、「起動し」と同じ意味記号に対応する形態素が 1 文前に出現することを示す。

【 0 0 5 2 】

図 1 0 のステップ 1 0 0 3 において、符号化部 4 1 2 は、各文の意味構造に対応する接続表を生成し、その接続表に基づいて各形態素の意味構造情報を生成する。

20

【 0 0 5 3 】

図 1 2 は、図 1 に示した第 1 文の意味構造に対応する接続表の例を示している。図 1 2 の接続表の各行及び各列の意味記号は、意味構造に含まれるノードを表す。各セル内の“ : ”の左側の数字 d は、セルに対応する行が表すノードと列が表すノードとの間の接続距離を表す接続距離情報である。 $d = 2$ は、2 つのノードが直接接続していることを示し、 $d = 3$ は、2 つのノードが $(d - 2)$ 個のノードを介して間接的に接続していることを示す。

【 0 0 5 4 】

また、各セル内の“ : ”の右側の文字列は、2 つのノードの間のアークに対応する接続関係を表す接続関係情報である。文字列の先頭の“ + ”は、行が表すノードから列が表すノードへアークが向かうことを示し、“ - ”は、列が表すノードから行が表すノードへアークが向かうことを示す。接続関係情報は、 $d = 2$ のセルのみに対して付加される。

30

【 0 0 5 5 】

以下では、接続表内におけるセルの位置を、(行が表すノードの意味記号 , 列が表すノードの意味記号) を用いて表記する。例えば、(S E 0 0 1 , P A 0 0 1) のセルの“ 3 : ”は、S E 0 0 1 (サーバー) と P A 0 0 1 (パッチ) とが 1 個のノードを介して間接的に接続されていることを表す。

【 0 0 5 6 】

(S E 0 0 1 , A P 0 0 1) のセルの“ 2 : - 目標 ”は、S E 0 0 1 と A P 0 0 1 (当て) とが < 目標 > を表すアークによって直接接続されており、アークが A P 0 0 1 から S E 0 0 1 へ向かうことを表す。(S E 0 0 1 , A G 0 0 1) のセルの“ 4 : ”は、S E 0 0 1 と A G 0 0 1 (再) とが 2 個のノードを介して間接的に接続されていることを表す。(S E 0 0 1 , B O 0 0 1) のセルの“ 3 : ”は、S E 0 0 1 と B O 0 0 1 (起動し) とが 1 個のノードを介して間接的に接続されていることを表す。

40

【 0 0 5 7 】

(A P 0 0 1 , P A 0 0 1) のセルの“ 2 : + 対象 ”は、A P 0 0 1 (当て) と P A 0 0 1 (パッチ) とが < 対象 > を表すアークによって直接接続されており、アークが A P 0 0 1 から P A 0 0 1 へ向かうことを表す。(A P 0 0 1 , S E 0 0 1) のセルの“ 2 : + ”

50

目標”は、A P 0 0 1とS E 0 0 1（サーバー）とが<目標>を表すアークによって直接接続されており、アークがA P 0 0 1からS E 0 0 1へ向かうことを表す。

【0058】

（A P 0 0 1，A G 0 0 1）のセルの“3：”は、A P 0 0 1とA G 0 0 1（再）とが1個のノードを介して間接的に接続されていることを表す。（A P 0 0 1，B O 0 0 1）のセルの“2：-順接”は、A P 0 0 1とB O 0 0 1（起動し）とが<順接>を表すアークによって直接接続されており、アークがB O 0 0 1からA P 0 0 1へ向かうことを表す。

【0059】

（B O 0 0 1，P A 0 0 1）のセルの“3：”は、B O 0 0 1（起動し）とP A 0 0 1（パッチ）とが1個のノードを介して間接的に接続されていることを表す。（B O 0 0 1，S E 0 0 1）のセルの“3：”は、B O 0 0 1とS E 0 0 1（サーバー）とが1個のノードを介して間接的に接続されていることを表す。

【0060】

（B O 0 0 1，A P 0 0 1）のセルの“2：+順接”は、B O 0 0 1とA P 0 0 1（当て）とが<順接>を表すアークによって直接接続されており、アークがB O 0 0 1からA P 0 0 1へ向かうことを表す。（B O 0 0 1，A G 0 0 1）のセルの“2：+修飾”は、B O 0 0 1とA G 0 0 1（再）とが<修飾>を表すアークによって直接接続されており、アークがB O 0 0 1からA G 0 0 1へ向かうことを表す。

【0061】

図13は、図2に示した第2文の意味構造に対応する接続表の例を示している。例えば、（T H 0 0 1，B O 0 0 1）のセルの“2：-順接”は、T H 0 0 1（すると）とB O 0 0 1（起動し）とが<順接>を表すアークによって直接接続されており、アークがB O 0 0 1からT H 0 0 1へ向かうことを表す。（B O 0 0 1，T H 0 0 1）のセルの“2：+順接”も、（T H 0 0 1，B O 0 0 1）のセルと同じ情報を表す。

【0062】

一方、（B O 0 0 1，B O 0 0 1）のセルの“否定”は、B O 0 0 1（起動し）が<否定>を表すアークを有することを表す。このように、<否定>を表す接続関係情報は、接続表内において、行と列が同じ意味記号に対応する対角セルに格納される。

【0063】

図14は、アークの接続関係に対応する符号の例を示している。図14の例では、<動作主>、<条件>、<対象>、<逆接>、<順接>、<目標>、及び<修飾>に対して、符号“1”～符号“7”がそれぞれ割り当てられており、<否定>に対して符号“-1”が割り当てられている。

【0064】

図15は、図12の接続表から生成される第1文の意味構造情報の例を示している。図15の意味構造情報は、図12の接続表における各行の情報を表すビット列を含む。各行のビット列は、接続距離を表す3ビットの情報を列の個数だけ含み、接続距離が2である列に対して、接続関係を表す8ビットの情報を列の個数だけ含む。空白セルの接続距離は0に設定される。

【0065】

例えば、S E 0 0 1（サーバー）のビット列は、5個の接続距離を表す15ビットのビット列“000 011 010 100 011”と、1個の接続関係を表す8ビットのビット列“11111101”とを含む。

【0066】

A P 0 0 1（当て）のビット列は、5個の接続距離を表す15ビットのビット列“010 010 000 011 010”と、3個の接続関係を表す24ビットのビット列“00000011 000000110 11111011”とを含む。

【0067】

B O 0 0 1（起動し）のビット列は、5個の接続距離を表す15ビットのビット列“0

10

20

30

40

50

1 1 0 1 1 0 1 0 0 0 1 0 0 0 ”と、2個の接続関係を表す16ビットのビット列“0 0 0 0 0 1 0 1 0 0 0 0 0 1 1 1”とを含む。

【0068】

図15の意味構造情報では、図12の接続表の各セルの接続距離情報が3ビットで表されるため、5×5の接続表全体の接続距離情報を格納するための記憶容量は、3ビット×5×5=75ビット=約9バイトである。また、各セルの接続関係情報は8ビットで表されるため、接続表の接続関係情報を格納するための記憶容量は、1バイト×アークの個数×2=1×4×2=8バイトである。したがって、意味構造情報を格納するための記憶容量は、9+8=17バイトである。

【0069】

これに対して、上述した先願技術の符号化文書では、1個の意味最小単位に対して6バイトの符号が割り当てられ、意味最小単位の個数に比例して符号量が増加する。したがって、図12の接続表に対応する5個の意味最小単位を格納するための記憶容量は、6バイト×5=30バイトである。

【0070】

このように、図15の意味構造情報を用いれば、先願技術の約56.7%の記憶容量で、同じ意味構造を格納することができる。

【0071】

図16は、図13の接続表から生成される第2文の意味構造情報の例を示している。〈否定〉を表すセルの接続距離は-1に設定される。

【0072】

例えば、TH001(すると)のビット列は、2個の接続距離を表す6ビットのビット列“0 0 0 0 1 0”と、1個の接続関係を表す8ビットのビット列“1 1 1 1 1 0 1 1”とを含む。BO001(起動し)のビット列は、2個の接続距離を表す6ビットのビット列“0 1 0 1 1 1”と、2個の接続関係を表す16ビットのビット列“0 0 0 0 0 1 0 1 1 1 1 1 1 1 1 1”とを含む。

【0073】

図17は、「パッチはサーバーに当てて再起動した。すると、起動しなかった。」という文書421から生成される符号化文書621の具体例を示している。この具体例では、第1文の「サーバーに当てて再起動し」に対応する符号、付加情報、及び意味構造情報が示されており、その前後の文字列及び第2文に対応する情報は省略されている。

【0074】

符号1701-1~符号1701-6は、それぞれ、「サーバー」、「に」、「当て」、「て」、「再」、及び「起動し」に対応する符号である。付加情報1702-1~付加情報1702-6は、それぞれ、符号1701-1~符号1701-6に対して付加された付加情報であり、図11に示した付加情報に対応している。

【0075】

意味構造情報1703-1~意味構造情報1703-4は、それぞれ、「サーバー」、「当て」、「再」、及び「起動し」の意味構造情報であり、図15に示した意味構造情報に対応している。付加情報1702-i(i=1~6)の参照フラグが論理“1”である4個の符号(i=1,3,5,6)に対して、それらの符号と同じ順序で、意味構造情報1703-j(j=1~4)が格納されている。

【0076】

このように、各形態素の符号の直後に付加情報を格納することで、検索処理を行う際に、意味構造情報を参照するか否かを効率的に判定することができるとともに、照応情報を直ちに参照することができる。また、複数の形態素の符号と、それらの形態素の意味構造情報とを、同じ順序で並べて格納することで、符号から対応する意味構造情報へ高速にアクセスすることが可能になる。

【0077】

各形態素に対応する意味構造情報は、その形態素と他の形態素との接続距離及び接続関

10

20

30

40

50

係を表す最小限の情報にあらかじめ絞り込むことができ、意味構造に属するどの形態素からでも意味構造情報へアクセスすることができる。

【0078】

図18は、図6の検索装置601の具体例を示している。図18の検索装置601は、記憶部611、検索部612、出力部613、及び解析部1801を含む。記憶部611は、図8の符号化装置401から出力される符号化文書621を記憶する。

【0079】

検索部612は、ユーザによって自然文の検索要求1811が入力されると、入力された検索要求1811を受け付けて、記憶部611に格納する。解析部1801は、検索要求1811に対して形態素解析及び意味解析を行い、形態素解析結果及び意味解析結果を含む解析結果1812を生成して、記憶部611に格納する。

10

【0080】

検索部612は、解析結果1812を用いて、検索要求1811の意味構造を表す検索要求意味構造情報1813を生成し、記憶部611に格納する。また、検索部612は、解析結果1812を用いて、検索要求1811に含まれる2つの形態素の組み合わせのスコアを表すスコア情報1814を生成し、記憶部611に格納する。そして、検索部612は、解析結果1812及び検索要求意味構造情報1813を用いて、符号化文書621を検索し、検索結果1815を生成して、記憶部611に格納する。出力部613は、検索結果1815を出力する。

【0081】

図19は、図9の符号化文書621に対する検索処理の例を示している。この場合、意味構造情報903-1～意味構造情報903-Mは、検索対象意味構造情報に対応する。

20

【0082】

まず、検索部612は、検索要求1811に含まれる形態素の符号を符号化文書621の先頭から順に検索し、符号901-2及び符号901-3が検索要求1811の符号とマッチする。そこで、検索部612は、符号901-2及び符号901-3に付加されている付加情報902-2及び付加情報902-3を参照する。

【0083】

付加情報902-2の参照フラグが論理“1”であり、付加情報902-3の参照フラグが論理“0”である場合、検索部612は、符号901-2に対応する意味構造情報903-2を参照する。そして、検索部612は、検索要求意味構造情報1813と意味構造情報903-2とを比較することで、検索要求1811の意味構造が、符号901-2に対応する形態素が属する意味構造にマッチするか否かをチェックする。

30

【0084】

そして、検索部612は、スコア情報1814を用いて、符号901-2に対応する形態素が属する意味構造に対する検索要求1811のスコアを計算し、計算したスコアを用いて検索結果1815を生成する。

【0085】

この場合、符号901-2に対応する形態素が属する意味構造に、符号901-1に対応する形態素の意味記号が含まれていたとしても、検索部612は、符号901-1に対応する意味構造情報903-1を参照する必要はない。

40

【0086】

このような検索処理によれば、意味構造のみに基づいて検索を行うのではなく、まず、高速な形態素マッチングが行われ、マッチした形態素に対してのみ、計算量の多い意味構造情報マッチングが行われる。これにより、検索処理の負荷が削減される。

【0087】

図20は、図18の検索装置601が行う検索処理の具体例を示すフローチャートである。まず、検索部612は、符号化文書621のスコアSC1を0に設定する(ステップ2001)。次に、検索部612は、解析結果1812を用いて、検索要求1811の意味構造に対応する接続表を生成し、その接続表に基づいて検索要求意味構造情報1813

50

を生成する（ステップ2002）。

【0088】

図21は、検索要求1811の意味構造の例を示している。図21の意味構造は、「サーバーが起動しない。」という検索要求1811に対応する。この検索要求1811は、「サーバー」、「が」、「起動し」、「ない」、及び「。」の5個の形態素を含む。このうち、「サーバー」及び「起動し」に対応する意味記号は、それぞれ、“SE001”及び“B0001”である。“B0001”と“SE001”の間の接続関係は<動作主>である。さらに、“B0001”は、<中心>及び<否定>を表すアークを有する。

【0089】

図22は、図21の意味構造に対応する接続表の例を示している。図22の接続表は、図12及び図13に示した文書421の接続表と同様に、接続距離情報及び接続関係情報を含む。ただし、検索要求1811の接続距離の定義は、文書421の接続距離とは異なっている。検索要求1811の接続表では、2つのノードが直接接続していることを示す接続距離dが用いられ、 $d = 1$ に設定される。

10

【0090】

意味構造情報マッチングでは、文書421の接続距離から、同じ2つのノードの間における検索要求1811の接続距離を減算することで、2つの接続距離が比較される。このとき、検索要求1811の接続距離として、直接接続を示す“1”を用いることで、文書421の接続距離dが直接接続を示す“2”である場合、dから“1”を減算することにより、直接接続を示す差分“1”を得ることができる。

20

【0091】

また、文書421の接続距離dが間接接続を示す“3”以上の値である場合、dから“1”を減算することにより、間接接続を示す“2”以上の差分を得ることができる。得られた差分から、文書421の意味構造において、2つのノードがどのくらい離れているかを計算することも可能である。

【0092】

例えば、(SE001, B0001)のセルの“1: - 動作主”は、SE001(サーバー)とB0001(起動し)とが<動作主>を表すアークによって直接接続されており、アークがB0001からSE001へ向かうことを表す。(B0001, SE001)のセルの“1: + 動作主”も、(SE001, B0001)のセルと同じ情報を表す。

30

【0093】

また、(B0001, B0001)のセルの“否定”は、B0001(起動し)が<否定>を表すアークを有することを表す。<否定>を表すセルの接続距離は-1に設定される。

【0094】

図23は、図22の接続表から生成される検索要求意味構造情報1813の例を示している。例えば、SE001(サーバー)のビット列は、2個の接続距離を表す6ビットのビット列“000 001”と、1個の接続関係を表す8ビットのビット列“11111111”とを含む。B0001(起動し)のビット列は、2個の接続距離を表す6ビットのビット列“001 111”と、2個の接続関係を表す16ビットのビット列“00000001 11111111”とを含む。

40

【0095】

検索要求意味構造情報1813を生成した後、検索部612は、検索要求1811の接続表に含まれる複数の意味記号を表す、検索要求1811の意味記号リストを生成する（ステップ2003）。検索要求1811の意味記号リストに含まれる各意味記号は、検索要求意味構造情報1813が表す接続表の各列に対応している。

【0096】

次に、検索部612は、図8の符号化情報811と同様の情報を用いて、検索要求1811に含まれる各形態素を対応する符号に置き換え、符号を用いて形態素マッチングを行う（ステップ2004）。形態素マッチングにおいて、検索部612は、検索要求181

50

1 に含まれる各形態素の符号を、符号化文書 6 2 1 から検索する。

【0097】

図 1 1 に示した文書 4 2 1 の符号化文書 6 2 1 に対して、「サーバーが起動しない。」という検索要求 1 8 1 1 を用いて形態素マッチングを行う場合を想定する。この場合、[1 , 3] 「サーバー」が検索要求 1 8 1 1 の「サーバー」にマッチし、[1 , 8] 「起動し」及び [2 , 2] 「起動し」が検索要求 1 8 1 1 の「起動し」にマッチする。また、[2 , 3] 「なかつ」が検索要求 1 8 1 1 の「ない」にマッチする。

【0098】

形態素マッチングを行った後、検索部 6 1 2 は、マッチした符号に対応する形態素毎に、ステップ 2 0 0 5 ~ ステップ 2 0 1 5 の処理を繰り返す。まず、検索部 6 1 2 は、マッ 10
チした符号に付加されている付加情報を参照して (ステップ 2 0 0 5)、その参照フラグをチェックする (ステップ 2 0 0 6)。

【0099】

参照フラグが論理 “ 1 ” である場合 (ステップ 2 0 0 6 , YES)、検索部 6 1 2 は、その符号に対応する意味構造情報を参照して、意味構造情報マッチングを行う (ステップ 2 0 0 7)。意味構造情報マッチングにおいて、検索部 6 1 2 は、参照した意味構造情報が表す接続表と検索要求意味構造情報 1 8 1 3 が表す接続表に共通する列について、2 つの意味構造情報間における接続距離の差分を計算する。

【0100】

一方、参照フラグが論理 “ 0 ” である場合 (ステップ 2 0 0 6 , NO)、検索部 6 1 2 は、次の形態素についてステップ 2 0 0 5 以降の処理を繰り返す。 20

【0101】

図 1 1 に示した文書 4 2 1 では、[1 , 3] 「サーバー」、[1 , 8] 「起動し」、及び [2 , 2] 「起動し」の参照フラグが論理 “ 1 ” であるため、これらの形態素について意味構造情報マッチングが行われる。一方、[2 , 3] 「なかつ」の参照フラグは論理 “ 0 ” であるため、この形態素については意味構造情報マッチングが行われない。

【0102】

図 2 4 は、図 2 0 のステップ 2 0 0 7 における意味構造情報マッチング処理の例を示すフローチャートである。まず、検索部 6 1 2 は、符号化文書 6 2 1 から、マッチした符号に対応する意味構造情報を取得して、意味構造情報 A に設定する (ステップ 2 4 0 1)。 30

【0103】

また、検索部 6 1 2 は、符号化文書 6 2 1 において、マッチした符号に対応する形態素を含む文から、論理 “ 1 ” の参照フラグが付加された複数の符号を抽出する。そして、検索部 6 1 2 は、それらの符号に対応する複数の形態素それぞれを意味記号に置き換えることで、検索対象の意味記号リストを生成する。検索対象の意味記号リストに含まれる各意味記号は、意味構造情報 A が表す接続表の各列に対応している。

【0104】

次に、検索部 6 1 2 は、マッチした符号に付加されている付加情報を参照して、その照 40
応情報をチェックする (ステップ 2 4 0 2)。照応情報に含まれる照応フラグ f が論理 “ 0 ” である場合 (ステップ 2 4 0 2 , NO)、検索部 6 1 2 は、検索要求意味構造情報 1 8 1 3 から、マッチした符号に対応する意味構造情報を抽出して、意味構造情報 B に設定する (ステップ 2 4 0 6)。

【0105】

次に、検索部 6 1 2 は、検索要求 1 8 1 1 の意味記号リストと、検索対象の意味記号リストとに共通する意味記号を特定し、意味構造情報 A 及び意味構造情報 B から、特定した意味記号に対応する列の接続距離を抽出する (ステップ 2 4 0 7)。そして、検索部 6 1 2 は、抽出した接続距離を用いて、共通する意味記号に対応する列のみを含む意味構造情報 A ' 及び意味構造情報 B ' を生成する (ステップ 2 4 0 8)。

【0106】

次に、検索部 6 1 2 は、意味構造情報 A ' 及び意味構造情報 B ' の各列について、意味 50

構造情報 A' の接続距離から意味構造情報 B' の接続距離を減算し、接続距離の差分を生成する (ステップ 2409)。

【0107】

一方、照応情報に含まれる照応フラグ f が論理 “ 1 ” である場合 (ステップ 2402, YES)、検索部 612 は、照応情報から位置情報 X を取得する (ステップ 2403)。次に、検索部 612 は、符号化文書 621 において、マッチした符号に対応する文よりも X 文前の文に含まれる符号のうち、マッチした符号に対応する形態素と同じ意味記号に対応する形態素の符号を特定する (ステップ 2404)。そして、検索部 612 は、符号化文書 621 から、特定した符号に対応する意味構造情報を取得して、意味構造情報 Z に設定する。

10

【0108】

また、検索部 612 は、X 文前の文から、論理 “ 1 ” の参照フラグが付加された複数の符号を抽出し、それらの符号に対応する複数の形態素それぞれを意味記号に置き換えることで、X 文前の文の意味記号リストを生成する。

【0109】

次に、検索部 612 は、意味構造情報 Z と意味構造情報 A とをマージすることで、意味構造情報 A を更新し、X 文前の文の意味記号リストと検索対象の意味記号リストとをマージすることで、検索対象の意味記号リストを更新する (ステップ 2405)。そして、検索部 612 は、ステップ 2406 以降の処理を行う。

【0110】

図 25 は、第 1 文の [1, 3] 「サーバー」に対する意味構造情報マッチング処理の例を示している。まず、図 15 に示した第 1 文の意味構造情報のうち、[1, 3] 「サーバー」に対応する SE001 のビット列が意味構造情報 A に設定される。そして、第 1 文の意味構造情報に含まれる 5 個の意味記号が、検索対象の意味記号リスト (A の意味記号リスト) に設定される。この場合、[1, 3] 「サーバー」の照応フラグ f は論理 “ 0 ” であるため、意味構造情報 A 及び A の意味記号リストは更新されない。

20

【0111】

また、図 23 に示した検索要求意味構造情報 1813 のうち、「サーバー」に対応する SE001 のビット列が意味構造情報 B に設定される。そして、検索要求意味構造情報 1813 に含まれる 2 個の意味記号が、検索要求 1811 の意味記号リスト (B の意味記号リスト) に設定される。

30

【0112】

この場合、A の意味記号リストの 2 列目の SE001 が B の意味記号リストの 1 列目の SE001 と一致しており、A の意味記号リストの 5 列目の BO001 が B の意味記号リストの 2 列目の BO001 と一致している。

【0113】

そこで、意味構造情報 A から 2 列目及び 5 列目の接続距離が抽出され、それらの接続距離を用いて意味構造情報 A' を表すビット列 “ 000 011 ” が生成される。また、意味構造情報 B から 1 列目及び 2 列目の接続距離が抽出され、それらの接続距離を用いて意味構造情報 B' を表すビット列 “ 000 001 ” が生成される。

40

【0114】

そして、意味構造情報 A' の各列の接続距離から意味構造情報 B' の各列の接続距離を減算することで、接続距離の差分を表すビット列 “ 000 010 ” が生成される。このビット列の 2 列目の “ 010 ” は、接続距離 “ 2 ” を表す。したがって、検索要求 1811 に含まれる SE001 及び BO001 が間接的に接続された意味構造が、第 1 文に存在することが分かる。

【0115】

図 26 は、第 1 文の [1, 8] 「起動し」に対する意味構造情報マッチング処理の例を示している。まず、図 15 に示した第 1 文の意味構造情報のうち、[1, 8] 「起動し」に対応する BO001 のビット列が意味構造情報 A に設定される。そして、第 1 文の意味

50

構造情報に含まれる5個の意味記号が、Aの意味記号リストに設定される。この場合、[1, 8]「起動し」の照応フラグfは論理“0”であるため、意味構造情報A及びAの意味記号リストは更新されない。

【0116】

また、図23に示した検索要求意味構造情報1813のうち、「起動し」に対応するB0001のビット列が意味構造情報Bに設定される。そして、検索要求意味構造情報1813に含まれる2個の意味記号が、Bの意味記号リストに設定される。

【0117】

この場合、Aの意味記号リストの2列目のSE001がBの意味記号リストの1列目のSE001と一致しており、Aの意味記号リストの5列目のB0001がBの意味記号リストの2列目のB0001と一致している。

10

【0118】

そこで、意味構造情報Aから2列目及び5列目の接続距離が抽出され、それらの接続距離を用いて意味構造情報A'を表すビット列“011 000”が生成される。また、意味構造情報Bから1列目及び2列目の接続距離が抽出され、それらの接続距離を用いて意味構造情報B'を表すビット列“001 111”が生成される。このビット列の2列目の“111”は、<否定>の接続距離である-1を表している。

【0119】

そして、意味構造情報A'の各列の接続距離から意味構造情報B'の各列の接続距離を減算することで、接続距離の差分を表すビット列“010 001”が生成される。このビット列の1列目の“010”は、接続距離“2”を表し、2列目の“001”は、意味構造情報A'の2列目の接続距離が-1ではなく、<否定>に対応していないことを表す。したがって、検索要求1811に含まれるSE001及びB0001が間接的に接続された意味構造が、第1文に存在するが、<否定>が一致しないことが分かる。

20

【0120】

図27は、第2文(文2)の[2, 2]「起動し」に対する意味構造情報マッチング処理の例を示している。まず、図16に示した文2の意味構造情報のうち、[2, 2]「起動し」に対応するB0001のビット列2701が意味構造情報Aに設定される。そして、文2の意味構造情報に含まれる2個の意味記号が、Aの意味記号リストに設定される。

【0121】

この場合、[2, 2]「起動し」の照応フラグfは論理“1”であり、位置情報Xは1であるため、文2よりも1文前の文1に含まれる形態素のうち、B0001に対応する[1, 8]「起動し」が特定される。そして、図15に示した文1の意味構造情報のうち、[1, 8]「起動し」に対応するB0001のビット列2702が意味構造情報Zに設定され、文1の意味構造情報に含まれる5個の意味記号が、文1の意味記号リストに設定される。

30

【0122】

次に、意味構造情報Zと意味構造情報Aとをマージすることで、意味構造情報Aが更新される。このとき、ビット列2702の5列目の斜体で示される“000”に対応する意味記号はB0001であり、ビット列2701の2列目の“010”に対応する意味記号と重複している。そこで、ビット列2702の5列目の“000”が削除されて、ビット列2703が生成され、ビット列2703とビット列2701とをマージすることで、ビット列2704が生成される。これにより、意味構造情報Aがビット列2704に更新される。

40

【0123】

また、文1の意味記号リストとAの意味記号リストとをマージすることで、Aの意味記号リストが更新される。このとき、2つの意味記号リストに重複して含まれるB0001が文1の意味記号リストから削除され、B0001を除く文1の意味記号がAの意味記号リストにマージされる。したがって、更新後のAの意味記号リストは、“PA001, SE001, AP001, AG001, TH001, B0001”となる。

50

【 0 1 2 4 】

また、図 2 3 に示した検索要求意味構造情報 1 8 1 3 のうち、「起動し」に対応する B 0 0 0 1 のビット列が意味構造情報 B に設定される。そして、検索要求意味構造情報 1 8 1 3 に含まれる 2 個の意味記号が、B の意味記号リストに設定される。

【 0 1 2 5 】

この場合、A の意味記号リストの 2 列目の S E 0 0 1 が B の意味記号リストの 1 列目の S E 0 0 1 と一致しており、A の意味記号リストの 6 列目の B O 0 0 1 が B の意味記号リストの 2 列目の B O 0 0 1 と一致している。

【 0 1 2 6 】

そこで、意味構造情報 A から 2 列目及び 6 列目の接続距離が抽出され、それらの接続距離を用いて意味構造情報 A ' を表すビット列 “ 0 1 1 1 1 1 ” が生成される。このビット列の 2 列目の “ 1 1 1 ” は、< 否定 > の接続距離である - 1 を表している。また、意味構造情報 B から 1 列目及び 2 列目の接続距離が抽出され、それらの接続距離を用いて意味構造情報 B ' を表すビット列 “ 0 0 1 1 1 1 ” が生成される。このビット列の 2 列目の “ 1 1 1 ” も、< 否定 > の接続距離である - 1 を表している。

10

【 0 1 2 7 】

そして、意味構造情報 A ' の各列の接続距離から意味構造情報 B ' の各列の接続距離を減算することで、接続距離の差分を表すビット列 “ 0 1 0 0 0 0 ” が生成される。このビット列の 1 列目の “ 0 1 0 ” は、接続距離 “ 2 ” を表し、2 列目の “ 0 0 0 ” は、意味構造情報 A ' 及び意味構造情報 B ' の 2 列目がともに < 否定 > に対応していることを表す。したがって、検索要求 1 8 1 1 に含まれる S E 0 0 1 及び B O 0 0 1 が間接的に接続された意味構造が、文 2 に存在し、< 否定 > が一致することが分かる。

20

【 0 1 2 8 】

意味構造情報マッチングを行った後、検索部 6 1 2 は、形態素マッチングでマッチした符号のスコア S C 2 を 0 に設定し、そのスコア S C 2 の係数 V を 1 に設定する（ステップ 2 0 0 8 ）。そして、検索部 6 1 2 は、接続距離の差分を計算した列毎に、ステップ 2 0 0 9 ~ ステップ 2 0 1 4 の処理を繰り返す。

【 0 1 2 9 】

まず、検索部 6 1 2 は、処理対象の列が < 否定 > の接続距離を用いた減算の結果を表す列（< 否定 > の列）であるか否かをチェックする（ステップ 2 0 0 9 ）。処理対象の列が < 否定 > の列である場合（ステップ 2 0 0 9 , Y E S ）、検索部 6 1 2 は、その列の減算の結果から、< 否定 > が一致するか否かをチェックする（ステップ 2 0 1 0 ）。< 否定 > が一致する場合（ステップ 2 0 1 0 , Y E S ）、検索部 6 1 2 は、係数 V を変更せず、< 否定 > が一致しない場合（ステップ 2 0 1 0 , N O ）、検索部 6 1 2 は、係数 V を - 1 に変更する（ステップ 2 0 1 1 ）。

30

【 0 1 3 0 】

次に、検索部 6 1 2 は、処理対象の列の差分を 1 と比較する（ステップ 2 0 1 2 ）。差分が 1 以上である場合（ステップ 2 0 1 2 , Y E S ）、検索部 6 1 2 は、形態素マッチングでマッチした符号に対応する形態素と、処理対象の列に対応する形態素との組み合わせのスコアを、スコア情報 1 8 1 4 から取得する（ステップ 2 0 1 3 ）。そして、検索部 6 1 2 は、取得したスコアを S C 3 に設定する。

40

【 0 1 3 1 】

次に、検索部 6 1 2 は、S C 2 に S C 3 を加算し、加算結果を S C 2 に設定して（ステップ 2 0 1 4 ）、次の列を処理対象として、ステップ 2 0 0 9 以降の処理を繰り返す。処理対象の列が < 否定 > の列ではない場合（ステップ 2 0 0 9 , N O ）、検索部 6 1 2 は、ステップ 2 0 1 2 以降の処理を行う。差分が 0 以下である場合（ステップ 2 0 1 2 , N O ）、検索部 6 1 2 は、次の列を処理対象として、ステップ 2 0 0 9 以降の処理を繰り返す。

【 0 1 3 2 】

すべての列についてステップ 2 0 0 9 ~ ステップ 2 0 1 4 の処理を行った後、検索部 6

50

12は、SC2と係数Vの積をSC1に加算し、加算結果をSC1に設定する(ステップ2015)。そして、検索部612は、次の形態素についてステップ2005以降の処理を繰り返す。すべての形態素についてステップ2005~ステップ2015の処理を行った後、検索部612は、スコアSC1を検索結果1815として記憶部611に格納し、処理を終了する。

【0133】

列毎にステップ2009~ステップ2014の処理を繰り返すことで、検索要求1811において、マッチした符号に対応する形態素と接続された複数の形態素について、組み合わせのスコアSC3がマッチした符号のスコアSC2に加算される。また、形態素毎にステップ2005~ステップ2015の処理を繰り返すことで、マッチした複数の符号について、スコアSC2が符号化文書621のスコアSC1に加算される。このとき、<否定>が一致しない意味構造については、スコアSC1からスコアSC2が減算される。

10

【0134】

こうして計算されたスコアSC1は、検索要求1811の意味構造と符号化文書621の意味構造との間の類似度を表す。したがって、スコアSC1が大きいほど、符号化文書621に含まれる意味構造は検索要求1811の意味構造と類似している。

【0135】

スコア情報1814に含まれる、2つの形態素の組み合わせのスコアSは、例えば、次式により計算することができる。

$$S = idf1 * N1 + idf2 * N2 \quad (1)$$

20

【0136】

idf1は、組み合わせに含まれる1番目の形態素の逆文書頻度(inverse document frequency)を表し、idf2は、2番目の形態素の逆文書頻度を表す。N1は、1番目の形態素の検索要求1811内における出現回数を表し、N2は、2番目の形態素の検索要求1811内における出現回数を表す。

【0137】

2つの形態素の組み合わせのスコアSは、式(1)の代わりに、次式により計算することもできる。

$$S = idf11 * N11 + idf12 * N12 \quad (2)$$

【0138】

idf11は、組み合わせに含まれる1番目の形態素に対応する意味記号の逆文書頻度を表し、idf12は、2番目の形態素に対応する意味記号の逆文書頻度を表す。N11は、1番目の形態素に対応する意味記号の検索要求1811内における出現回数を表し、N12は、2番目の形態素に対応する意味記号の検索要求1811内における出現回数を表す。

30

【0139】

例えば、図25の接続距離の差分を表すビット列“000 010”の場合、2列目の“010”が接続距離“2”を表すため、SE001とBO001の組み合わせのスコアS(SE001, BO001)がSC2に設定される。そして、S(SE001, BO001)がSC1に加算される。

40

【0140】

一方、図26の接続距離の差分を表すビット列“010 001”の場合、1列目の“010”が接続距離“2”を表すため、スコアS(SE001, BO001)がSC2に設定される。しかし、2列目の<否定>が一致しないため、S(SE001, BO001)に-1が乗算されて、SC1に加算される。

【0141】

また、図27の接続距離の差分を表すビット列“010 000”の場合、1列目の“010”が接続距離“2”を表すため、スコアS(SE001, BO001)がSC2に設定される。そして、2列目の<否定>が一致するため、S(SE001, BO001)がそのままSC1に加算される。

50

【 0 1 4 2 】

なお、図 20 のステップ 2014 において、検索部 612 は、接続距離の差分 D を用いて、次式により SC2 を更新してもよい。

$$SC2 = SC2 + SC3 \times (1 / D) \quad (3)$$

【 0 1 4 3 】

これにより、直接接続を示す $D = 1$ の場合、SC3 がそのまま SC2 に加算され、間接接続を示す $D > 1$ の場合、SC3 と $1 / D$ の積が SC2 に加算される。したがって、接続距離が短いほど SC2 が大きくなり、符号化文書 621 のスコア SC1 も大きくなる。

【 0 1 4 4 】

複数の文書 421 に対応する複数の符号化文書 621 が検索対象として与えられた場合、検索部 612 は、それぞれの符号化文書 621 に対して、図 20 の検索処理を行う。これにより、それぞれの符号化文書 621 に対するスコア SC1 が計算される。

10

【 0 1 4 5 】

図 20 の検索処理によれば、検索対象の文書 421 に含まれる文が不完全であっても、検索要求 1811 と類似する文を検索することが可能になる。このとき、符号化文書 621 に付加情報と意味構造情報とを含めることで、形態素マッチングと意味構造情報マッチングとが統合される。このため、すべての検索要求 1811 のバリエーションを事前に想定して意味構造に展開しておかなくても、照応解析の負荷を大幅に削減して、スコア演算を高速化することができる。

【 0 1 4 6 】

文書 421 に N 個の形態素が含まれている場合、図 10 の符号化処理及び図 20 の検索処理において、欠落している形態素の情報を補完する処理の計算量は、いずれも N のオーダーで済む。符号化処理では、形態素を用いて照応解析が行われるので、計算量は N に比例する。また、検索処理では、マッチした符号に付加された照応情報が位置情報 X を含む場合に限って、意味構造情報が参照されるので、計算量は N に比例する。したがって、 $N \times N$ のオーダーの計算量を有する図 3 の意味構造検索処理と比較して、より高速な意味構造検索が実現される。

20

【 0 1 4 7 】

ところで、符号化文書 621 において、<否定>を表すアークの情報を意味構造情報に含める代わりに、否定フラグとして付加情報に含めることも可能である。

30

【 0 1 4 8 】

図 28 は、否定フラグを含む付加情報の例を示している。図 28 の否定フラグは、図 11 に示した文書 421 の符号化文書 621 において、各符号に対応する形態素の意味構造が<否定>を表すアークを有するか否かを示す。この例では、図 2 に示したように、[2, 2]「起動し」の意味記号である B0001 が<否定>を表すアークを有するため、[2, 2]「起動し」の否定フラグに論理“1”が設定されている。

【 0 1 4 9 】

この場合、図 16 に示した第 2 文の意味構造情報において、<否定>を表す接続距離は設定されない。したがって、B0001 (起動し) のビット列の 2 列目の“111”は、“000”に変更される。そして、図 20 の検索処理において、検索部 612 は、ステップ 2009 ~ ステップ 2011 の処理を行う代わりに、否定フラグを参照して係数 V の値を決定する。

40

【 0 1 5 0 】

図 29 は、文の末尾から順に符号を並べた符号化文書 621 の例を示している。図 29 の符号化文書 621 では、文書 421 内の P 個の文について、先頭の文 2901 - 1 から末尾の文 2901 - P へ向かって、符号が文毎にまとめて配置される。そして、各文に含まれる Q 個の形態素について、文の末尾の形態素 2902 - 1 から文の先頭の形態素 2902 - Q へ向かって、符号 2903 及び付加情報 2904 が順に並べられている。この場合、それぞれの符号 2903 に対応する意味構造情報も、符号 2903 と同じ順序で配置される。

50

【 0 1 5 1 】

図 29 の符号の配置は、多数の形態素と接続される形態素が文の末尾近くに出現する言語に対して特に有効である。例えば、日本語、韓国語等の場合、複数の名詞等と接続される動詞が文の末尾近くに出現することが多い。図 11 に示した文書 4 2 1 では、[1 , 8] 「起動し」がこのような動詞に該当する。

【 0 1 5 2 】

この場合、文の末尾から順に符号を並べることで、検索処理の早い段階で、多数の形態素と接続される形態素の意味構造情報を参照して、意味構造検索を実施することが可能になる。これにより、早い段階で、最終的なスコアに近い符号化文書 6 2 1 のスコア S C 1 を求めることができる。

10

【 0 1 5 3 】

図 30 は、一連の符号の後に付加情報を並べた符号化文書 6 2 1 の例を示している。図 30 (a) の符号化文書 6 2 1 では、図 9 の符号 9 0 1 - 1 ~ 符号 9 0 1 - N に続いて、付加情報 9 0 2 - 1 ~ 付加情報 9 0 2 - N が符号 9 0 1 - i と同じ順序で配置される。そして、付加情報 9 0 2 - 1 ~ 付加情報 9 0 2 - N に続いて、意味構造情報 9 0 3 - 1 ~ 意味構造情報 9 0 3 - M が符号 9 0 1 - i と同じ順序で配置される。

【 0 1 5 4 】

一方、図 30 (b) の符号化文書 6 2 1 では、図 9 の符号 9 0 1 - 1 ~ 符号 9 0 1 - N に続いて、付加情報 9 0 2 - N ~ 付加情報 9 0 2 - 1 が符号 9 0 1 - i とは逆の順序で配置される。そして、付加情報 9 0 2 - N ~ 付加情報 9 0 2 - 1 に続いて、意味構造情報 9 0 3 - M ~ 意味構造情報 9 0 3 - 1 が符号 9 0 1 - i とは逆の順序で配置される。

20

【 0 1 5 5 】

図 29 又は図 30 の符号化文書 6 2 1 を用いた場合も、付加情報及び意味構造情報を参照しながら、意味構造検索を高速に行うことができる。

【 0 1 5 6 】

図 4 及び図 8 の符号化装置 4 0 1 の構成は一例に過ぎず、符号化装置 4 0 1 の用途又は条件に応じて、一部の構成要素を省略又は変更してもよい。例えば、解析結果 8 1 2 が外部の装置で生成される場合は、図 8 の解析部 8 0 1 を省略することができる。

【 0 1 5 7 】

図 6 及び図 18 の検索装置 6 0 1 の構成は一例に過ぎず、検索装置 6 0 1 の用途又は条件に応じて、一部の構成要素を省略又は変更してもよい。例えば、解析結果 1 8 1 2 が外部の装置で生成される場合は、図 18 の解析部 1 8 0 1 を省略することができる。検索装置 6 0 1 に、図 8 の符号化部 4 1 2 及び解析部 8 0 1 を追加して、検索装置 6 0 1 内で符号化処理を行ってもよい。

30

【 0 1 5 8 】

図 3、図 5、図 7、図 10、図 20、及び図 24 のフローチャートは一例に過ぎず、符号化装置 4 0 1 又は検索装置 6 0 1 の構成又は条件に応じて一部の処理を省略又は変更してもよい。例えば、図 10 の符号化処理において、解析結果 8 1 2 が外部の装置で生成される場合は、ステップ 1 0 0 1 及びステップ 1 0 0 2 の処理を省略することができる。

【 0 1 5 9 】

図 20 の検索処理において、符号化文書 6 2 1 のスコアを計算しない場合は、ステップ 2 0 0 8 ~ ステップ 2 0 1 5 の処理を省略することができる。この場合、ステップ 2 0 0 7 の意味構造情報マッチング処理の結果が、検索結果 1 8 1 5 として記憶部 6 1 1 に格納される。

40

【 0 1 6 0 】

図 1、図 2、図 21 の意味構造、図 9、図 17、図 29、及び図 30 の符号化文書、図 11 及び図 28 の解析結果及び付加情報、図 12、図 13、及び図 22 の接続表、図 14 のアーク及び符号、図 15、図 16、及び図 23 の意味構造情報は一例に過ぎない。符号化装置 4 0 1 又は検索装置 6 0 1 の構成又は条件に応じて、別のデータ構造の情報を用いてもよい。また、各情報の内容は、文書 4 2 1 及び検索要求 1 8 1 1 の内容に応じて変化

50

する。文書 4 2 1 及び検索要求 1 8 1 1 の言語は、英語、中国語、韓国語等のように、日本語以外の言語であってもよい。

【 0 1 6 1 】

図 1 9 の検索処理及び図 2 5 ~ 図 2 7 の意味構造情報マッチング処理は一例に過ぎず、各処理の内容は、文書 4 2 1 及び検索要求 1 8 1 1 の内容に応じて変化する。式 (1) ~ 式 (3) は一例に過ぎず、別の計算式によりスコア S 又はスコア S C 2 を計算してもよい。

【 0 1 6 2 】

図 3 1 は、図 4 及び図 8 の符号化装置 4 0 1、又は図 6 及び図 1 8 の検索装置 6 0 1 として用いられる情報処理装置 (コンピュータ) の構成例を示している。図 3 1 の情報処理装置は、CPU 3 1 0 1、メモリ 3 1 0 2、入力装置 3 1 0 3、出力装置 3 1 0 4、補助記憶装置 3 1 0 5、媒体駆動装置 3 1 0 6、及びネットワーク接続装置 3 1 0 7 を備える。これらの構成要素はバス 3 1 0 8 により互いに接続されている。

10

【 0 1 6 3 】

メモリ 3 1 0 2 は、例えば、Read Only Memory (R O M)、Random Access Memory (R A M)、フラッシュメモリ等の半導体メモリであり、処理に用いられるプログラム及びデータを格納する。メモリ 3 1 0 2 は、図 4 及び図 8 の記憶部 4 1 1、又は図 6 及び図 1 8 の記憶部 6 1 1 として用いることができる。

【 0 1 6 4 】

CPU 3 1 0 1 (プロセッサ) は、例えば、メモリ 3 1 0 2 を利用してプログラムを実行することにより、図 4 及び図 8 の符号化部 4 1 2 及び解析部 8 0 1 として動作する。CPU 3 1 0 1 は、メモリ 3 1 0 2 を利用してプログラムを実行することにより、図 6 及び図 1 8 の検索部 6 1 2 及び解析部 1 8 0 1 としても動作する。

20

【 0 1 6 5 】

入力装置 3 1 0 3 は、例えば、キーボード、ポインティングデバイス等であり、オペレータ又はユーザからの指示及び情報の入力に用いられる。入力される情報は、検索要求 1 8 1 1 であってもよい。

【 0 1 6 6 】

出力装置 3 1 0 4 は、例えば、表示装置、プリンタ、スピーカ等であり、オペレータ又はユーザへの問い合わせ又は指示、及び処理結果の出力に用いられる。出力装置 3 1 0 4 は、図 6 及び図 1 8 の出力部 6 1 3 として用いることができる。処理結果は、検索結果 1 8 1 5 であってもよい。

30

【 0 1 6 7 】

補助記憶装置 3 1 0 5 は、例えば、磁気ディスク装置、光ディスク装置、光磁気ディスク装置、テープ装置等である。補助記憶装置 3 1 0 5 は、ハードディスクドライブ又はフラッシュメモリであってもよい。情報処理装置は、補助記憶装置 3 1 0 5 にプログラム及びデータを格納しておき、それらをメモリ 3 1 0 2 にロードして使用することができる。補助記憶装置 3 1 0 5 は、図 4 及び図 8 の記憶部 4 1 1、又は図 6 及び図 1 8 の記憶部 6 1 1 として用いることができる。

【 0 1 6 8 】

媒体駆動装置 3 1 0 6 は、可搬型記録媒体 3 1 0 9 を駆動し、その記録内容にアクセスする。可搬型記録媒体 3 1 0 9 は、メモリデバイス、フレキシブルディスク、光ディスク、光磁気ディスク等である。可搬型記録媒体 3 1 0 9 は、Compact Disk Read Only Memory (C D - R O M)、Digital Versatile Disk (D V D)、Universal Serial Bus (U S B) メモリ等であってもよい。オペレータ又はユーザは、この可搬型記録媒体 3 1 0 9 にプログラム及びデータを格納しておき、それらをメモリ 3 1 0 2 にロードして使用することができる。

40

【 0 1 6 9 】

このように、処理に用いられるプログラム及びデータを格納するコンピュータ読み取り可能な記録媒体は、メモリ 3 1 0 2、補助記憶装置 3 1 0 5、又は可搬型記録媒体 3 1 0

50

9のような、物理的な（非一時的な）記録媒体である。

【0170】

ネットワーク接続装置3107は、Local Area Network、Wide Area Network等の通信ネットワークに接続され、通信に伴うデータ変換を行う通信インタフェースである。情報処理装置は、プログラム及びデータを外部の装置からネットワーク接続装置3107を介して受け取り、それらをメモリ3102にロードして使用することができる。

【0171】

情報処理装置が図8の符号化装置401である場合、ネットワーク接続装置3107は、符号化文書621を検索装置601へ送信することができる。この場合、ネットワーク接続装置3107は、出力部802として用いられる。

10

【0172】

情報処理装置が図6及び図18の検索装置601である場合、ネットワーク接続装置3107は、ユーザ端末から検索要求1811を受信し、検索結果1815をユーザ端末へ送信することができる。この場合、ネットワーク接続装置3107は、出力部613として用いられる。

【0173】

なお、情報処理装置が図31のすべての構成要素を含む必要はなく、用途又は条件に応じて一部の構成要素を省略することも可能である。例えば、情報処理装置がユーザ端末から通信ネットワーク経由で検索要求1811を受信する場合は、入力装置3103及び出力装置3104を省略してもよい。また、可搬型記録媒体3109又は通信ネットワーク

20

【0174】

情報処理装置がスマートフォンのような通話機能を有する携帯端末である場合、マイク及びスピーカのような通話用の装置を含んでいてもよく、カメラのような撮像装置を含んでいてもよい。

【0175】

開示の実施形態とその利点について詳しく説明したが、当業者は、特許請求の範囲に明確に記載した本発明の範囲から逸脱することなく、様々な変更、追加、省略をすることができるであろう。

30

【0176】

図1乃至図31を参照しながら説明した実施形態に関し、さらに以下の付記を開示する。

(付記1)

文書に含まれる複数の文それぞれについて、文に含まれる複数の形態素が属する意味構造を表す意味構造情報を生成し、

前記複数の文それぞれについて、前記複数の形態素それぞれを符号化した複数の符号を生成し、

前記複数の文それぞれについて、各文に含まれる複数の形態素のうち、意味構造において対応する形態素の少なくとも一部が欠落している可能性のある形態素に対応する符号を

40

特定し、前記特定された符号それぞれに対し、欠落情報をそれぞれ付加する、処理をコンピュータに実行させる符号化プログラム。

(付記2)

前記複数の形態素が属する前記意味構造は、前記複数の形態素それぞれに対応する複数の意味記号と、前記複数の意味記号の間の接続関係とを用いて表され、前記欠落情報は、前記文書内において、前記欠落情報が付加された符号に対応する形態素と同じ意味記号に対応する形態素が、前記欠落情報が付加された前記符号に対応する形態素よりも前に出現することを示し、前記複数の文のうち前記同じ意味記号に対応する形態素が出現する文を特定する情報を含むことを特徴とする付記1記載の符号化プログラム。

50

(付記 3)

前記コンピュータは、前記複数の符号各々に対して、前記欠落情報、又は前記文書内において同じ意味記号に対応する形態素が前に出現しないことを示す非出現情報を付加することを特徴とする付記 2 記載の符号化プログラム。

(付記 4)

前記意味構造情報は、前記複数の意味記号各々について、前記複数の意味記号のうち他の意味記号に対する接続距離を表す接続距離情報と、前記他の意味記号に対する接続関係を表す接続関係情報とを含むことを特徴とする付記 2 又は 3 記載の符号化プログラム。

(付記 5)

検索要求に含まれる複数の形態素が属する意味構造を表す検索要求意味構造情報を生成し、

検索対象文書に含まれる複数の文それぞれについて文に含まれる複数の形態素が属する意味構造を表す検索対象意味構造情報と、前記複数の文それぞれについて前記複数の形態素それぞれを符号化することで生成された複数の符号と、前記複数の文それぞれについて、各文に含まれる複数の形態素のうち、意味構造において対応する形態素の少なくとも一部が欠落している可能性のある形態素に対応する符号それぞれに対して付加された欠落情報とを含む、符号化文書から、前記検索要求意味構造情報に基づいて前記検索要求に対応する文を検索し、

検索結果を出力する、

処理をコンピュータに実行させる検索プログラム。

(付記 6)

前記検索要求に含まれる前記複数の形態素が属する前記意味構造は、前記検索要求に含まれる前記複数の形態素それぞれに対応する複数の意味記号と、前記検索要求に含まれる前記複数の意味記号の間の接続関係とを用いて表され、

前記文に含まれる前記複数の形態素が属する前記意味構造は、前記文に含まれる前記複数の形態素それぞれに対応する複数の意味記号と、前記文に含まれる前記複数の意味記号の間の接続関係とを用いて表され、

前記欠落情報は、前記検索対象文書内において、前記欠落情報が付加された符号に対応する形態素と同じ意味記号に対応する形態素が、前記欠落情報が付加された前記符号に対応する形態素よりも前に出現することを示し、前記複数の文のうち前記同じ意味記号に対応する形態素が出現する文を特定する情報を含み、

前記コンピュータは、前記符号化文書から、前記検索要求に含まれる前記複数の形態素のうち 1 つの形態素を検索し、検索した形態素が、前記欠落情報が付加された前記符号に対応する場合、前記欠落情報を参照して、前記同じ意味記号に対応する形態素が出現する前記文を特定し、前記検索対象意味構造情報において、特定した前記文に出現する前記同じ意味記号に対応する形態素が属する第 1 意味構造と、前記検索対象意味構造情報において、前記欠落情報が付加された前記符号に対応する形態素が属する第 2 意味構造と、前記検索要求意味構造情報とを用いて、前記検索要求に対応する前記文を検索することを特徴とする付記 5 記載の検索プログラム。

(付記 7)

前記複数の符号各々に対して、前記欠落情報、又は前記検索対象文書内において同じ意味記号に対応する形態素が前に出現しないことを示す非出現情報が付加されており、

前記コンピュータは、前記検索した形態素が、前記非出現情報が付加された符号に対応する場合、前記検索対象意味構造情報において、前記非出現情報が付加された前記符号に対応する形態素が属する第 3 意味構造と、前記検索要求意味構造情報とを用いて、前記検索要求に対応する前記文を検索することを特徴とする付記 6 記載の検索プログラム。

(付記 8)

前記検索要求意味構造情報及び前記検索対象意味構造情報は、前記複数の意味記号各々について、前記複数の意味記号のうち他の意味記号に対する接続距離を表す接続距離情報と、前記他の意味記号に対する接続関係を表す接続関係情報とを含み、

10

20

30

40

50

前記コンピュータは、前記検索要求意味構造情報に含まれる接続距離情報及び接続関係情報と、前記検索対象意味構造情報に含まれる接続距離情報及び接続関係情報とを比較することで、前記検索要求に対応する前記文を検索することを特徴とする付記 6 又は 7 記載の検索プログラム。

(付記 9)

文書を記憶する記憶部と、

前記文書に含まれる複数の文それぞれについて、文に含まれる複数の形態素が属する意味構造を表す意味構造情報を生成し、前記複数の文それぞれについて、前記複数の形態素それぞれを符号化した複数の符号を生成し、前記複数の文それぞれについて、各文に含まれる複数の形態素のうち、意味構造において対応する形態素の少なくとも一部が欠落している可能性のある形態素に対応する符号を特定し、前記特定された符号それぞれに対し、欠落情報をそれぞれ付加する符号化部と、
備えることを特徴とする符号化装置。

10

(付記 10)

前記複数の形態素が属する前記意味構造は、前記複数の形態素それぞれに対応する複数の意味記号と、前記複数の意味記号の間の接続関係とを用いて表され、前記欠落情報は、前記文書内において、前記欠落情報が付加された符号に対応する形態素と同じ意味記号に対応する形態素が、前記欠落情報が付加された前記符号に対応する形態素よりも前に出現することを示し、前記複数の文のうち前記同じ意味記号に対応する形態素が出現する文を特定する情報を含むことを特徴とする付記 9 記載の符号化装置。

20

(付記 11)

前記符号化部は、前記複数の符号各々に対して、前記欠落情報、又は前記文書内において同じ意味記号に対応する形態素が前に出現しないことを示す非出現情報を付加することを特徴とする付記 10 記載の符号化装置。

(付記 12)

コンピュータが、

文書に含まれる複数の文それぞれについて、文に含まれる複数の形態素が属する意味構造を表す意味構造情報を生成し、

前記複数の文それぞれについて、前記複数の形態素それぞれを符号化した複数の符号を生成し、

30

前記複数の文それぞれについて、各文に含まれる複数の形態素のうち、意味構造において対応する形態素の少なくとも一部が欠落している可能性のある形態素に対応する符号を特定し、

前記特定された符号それぞれに対し、欠落情報をそれぞれ付加する、
ことを特徴とする符号化方法。

(付記 13)

前記複数の形態素が属する前記意味構造は、前記複数の形態素それぞれに対応する複数の意味記号と、前記複数の意味記号の間の接続関係とを用いて表され、前記欠落情報は、前記文書内において、前記欠落情報が付加された符号に対応する形態素と同じ意味記号に対応する形態素が、前記欠落情報が付加された前記符号に対応する形態素よりも前に出現することを示し、前記複数の文のうち前記同じ意味記号に対応する形態素が出現する文を特定する情報を含むことを特徴とする付記 12 記載の符号化方法。

40

(付記 14)

前記コンピュータは、前記複数の符号各々に対して、前記欠落情報、又は前記文書内において同じ意味記号に対応する形態素が前に出現しないことを示す非出現情報を付加することを特徴とする付記 13 記載の符号化方法。

【符号の説明】

【0177】

401 符号化装置

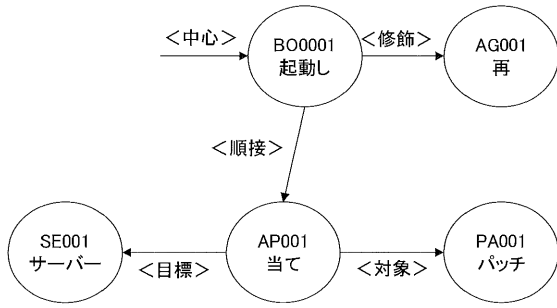
411、611 記憶部

50

4 1 2	符号化部	
4 2 1	文書	
6 0 1	検索装置	
6 1 2	検索部	
6 1 3、8 0 2	出力部	
6 2 1	符号化文書	
8 0 1、1 8 0 1	解析部	
8 0 2	出力部	
8 1 1	符号化情報	
8 1 2、1 8 1 2	解析結果	10
9 0 1 - 1 ~ 9 0 1 - 3、1 7 0 1 - 1 ~ 1 7 0 1 - 6、2 9 0 3	符号	
9 0 2 - 1 ~ 9 0 2 - 3、1 7 0 2 - 1 ~ 1 7 0 2 - 6、2 9 0 4	付加情報	
9 0 3 - 1 ~ 9 0 3 - 3、1 7 0 3 - 1 ~ 1 7 0 3 - 4	意味構造情報	
1 8 1 1	検索要求	
1 8 1 3	検索要求意味構造情報	
1 8 1 4	スコア情報	
1 8 1 5	検索結果	
2 7 0 1 ~ 2 7 0 4	ビット列	
2 7 0 2	ビット列	
2 7 0 3	ビット列	20
2 7 0 4	ビット列	
2 9 0 1 - 1 ~ 2 9 0 1 - P	文	
2 9 0 2 - 1 ~ 2 9 0 2 - Q	形態素	
3 1 0 1	C P U	
3 1 0 2	メモリ	
3 1 0 3	入力装置	
3 1 0 4	出力装置	
3 1 0 5	補助記憶装置	
3 1 0 6	媒体駆動装置	
3 1 0 7	ネットワーク接続装置	30
3 1 0 8	バス	
3 1 0 9	可搬型記録媒体	

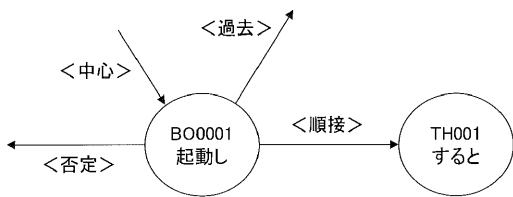
【 図 1 】

文書に含まれる第1文の意味構造を示す図



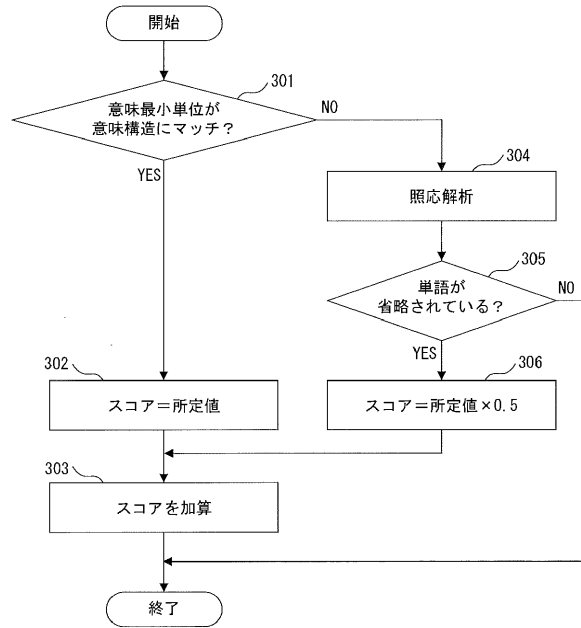
【 図 2 】

文書に含まれる第2文の意味構造を示す図



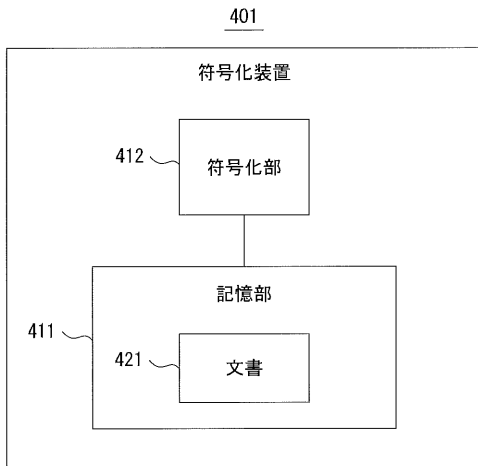
【 図 3 】

照応解析を用いた意味構造
検索処理のフローチャート



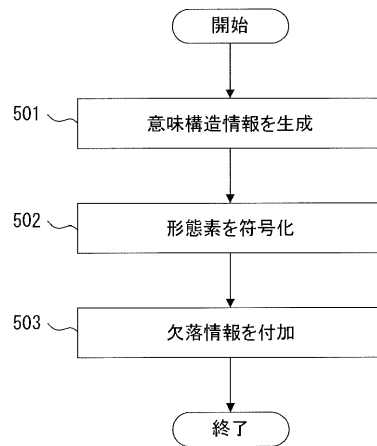
【 図 4 】

符号化装置の機能的構成図



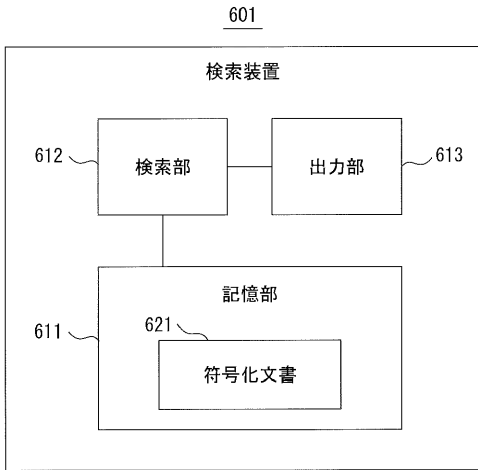
【 図 5 】

符号化処理のフローチャート



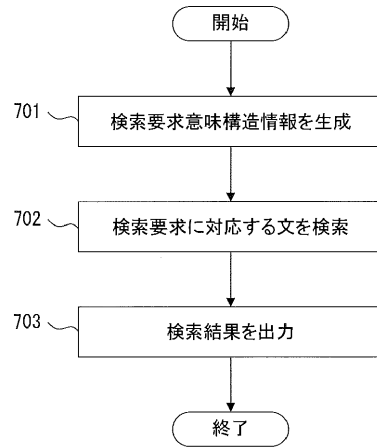
【 図 6 】

検索装置の機能的構成図



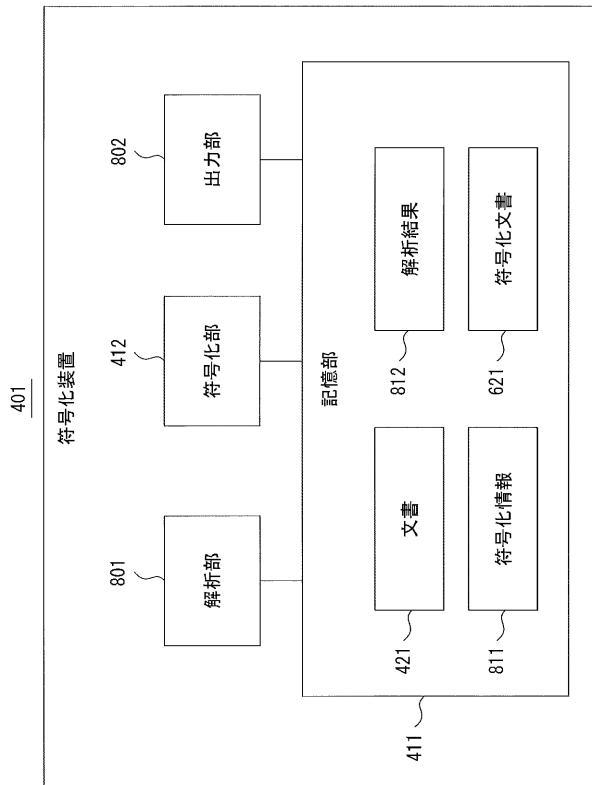
【 図 7 】

検索処理のフローチャート



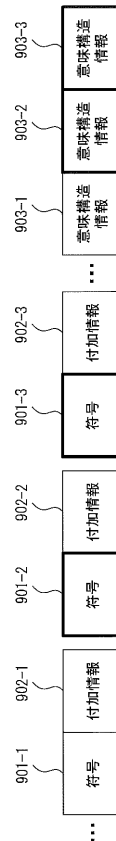
【 図 8 】

符号化装置の具体例を示す機能的構成図



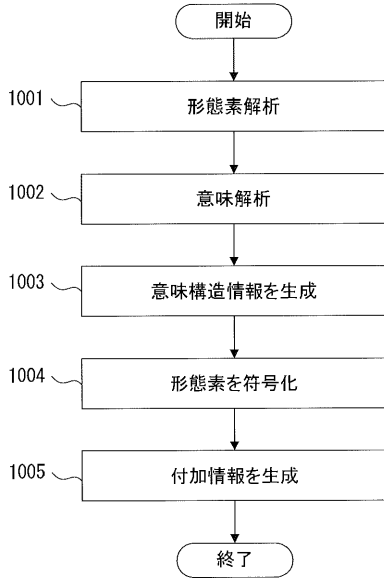
【 図 9 】

符号化文書を示す図



【図 1 0】

符号化処理の具体例を示すフローチャート



【図 1 1】

文書の解析結果を示す図

文番号	1										2				
	文頭からの位置	1	2	3	4	5	6	7	8	9	10	1	2	3	4
形態素	バツ子	は	サーバー	に	当て	て	再	起動し	た	。	ずると	起動し	なかつ	た	。
参照フラグ	1	0	1	0	1	0	1	1	0	0	1	1	0	0	0
付加情報	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
照応情報											1:1				

【図 1 2】

第1文の意味構造に対応する接続表を示す図

文1	PA001	SE001	AP001	AG001	BO001
PA001		3:	2:-対象	4:	3:
SE001	3:		2:-目標	4:	3:
AP001	2:+対象	2:+目標		3:	2:-順接
AG001	4:	4:	3:		2:-修飾
BO001	3:	3:	2:+順接	2:+修飾	

【図 1 3】

第2文の意味構造に対応する接続表を示す図

文2	TH001	BO001
TH001		2:-順接
BO001	2:+順接	否定

【図 1 4】

アークの符号を示す図

アーク	符号
動作主	1
条件	2
対象	3
逆接	4
順接	5
目標	6
修飾	7
否定	-1

【 図 1 5 】

第1文の意味構造情報を示す図

PA001: 000 011 010 100 011 11111101 (=0 3 2 4 3 -3)
 SE001: 011 000 010 100 011 11111010 000 (=3 0 2 4 3 -6)
 AP001: 010 010 000 011 010 00000011 00000110 11111011 (=2 2 0 3 2 3 6 -5)
 AG001: 100 100 011 000 001 11111001 (=4 4 3 0 1 -7)
 BO001: 011 011 010 001 000 00000101 00000111 (=3 3 2 1 0 5 7)

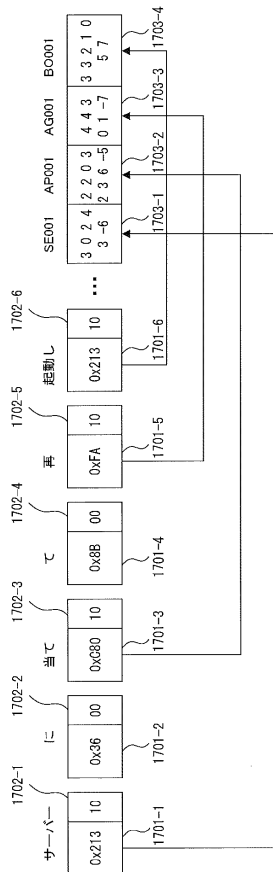
【 図 1 6 】

第2文の意味構造情報を示す図

TH001: 000 010 11111011 (=0 2 -5)
 BO001: 010 111 0000101 11111111 (=2 -1 5 -1)

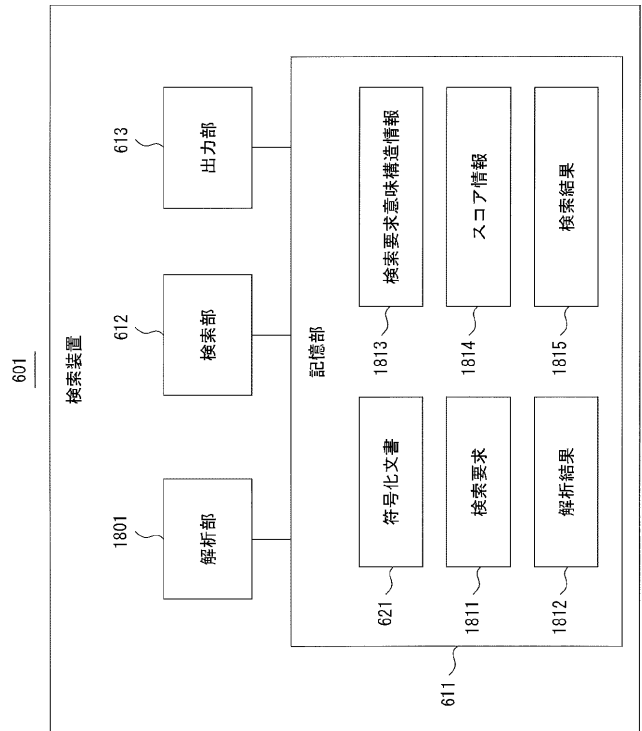
【 図 1 7 】

符号化文書の具体例を示す図



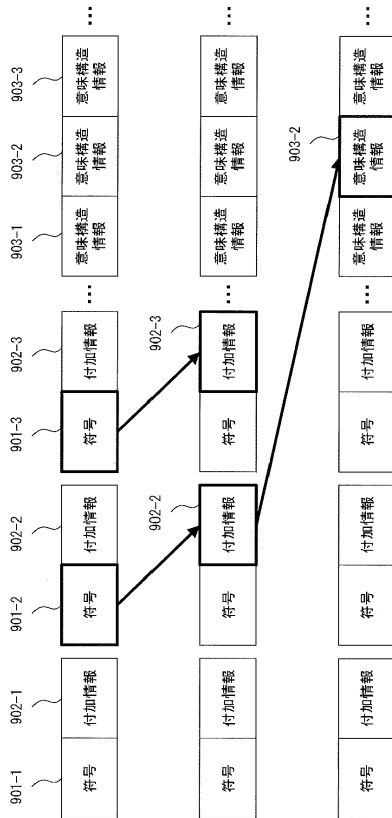
【 図 1 8 】

検索装置の具体例を示す機能的構成図



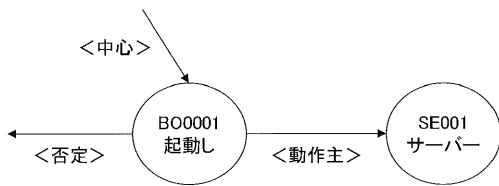
【図19】

検索処理を示す図



【図21】

検索要求の意味構造を示す図



【図22】

検索要求の意味構造に対応する接続表を示す図

検索要求	SE001	BO001
SE001		1:-動作主
BO001	1:+動作主	否定

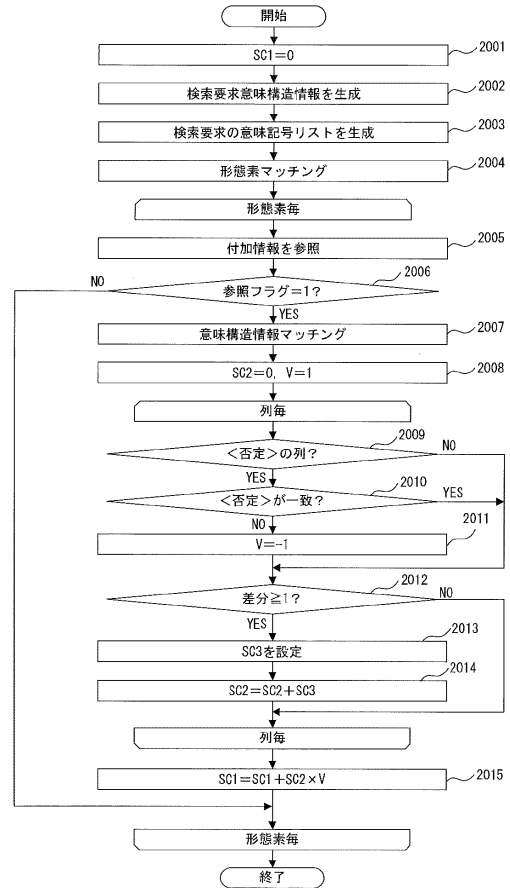
【図23】

検索要求意味構造情報を示す図

SE001: 000 001 11111111 (=0 1 -1)
 BO001: 001 111 00000001 11111111 (=1 -1 1 -1)

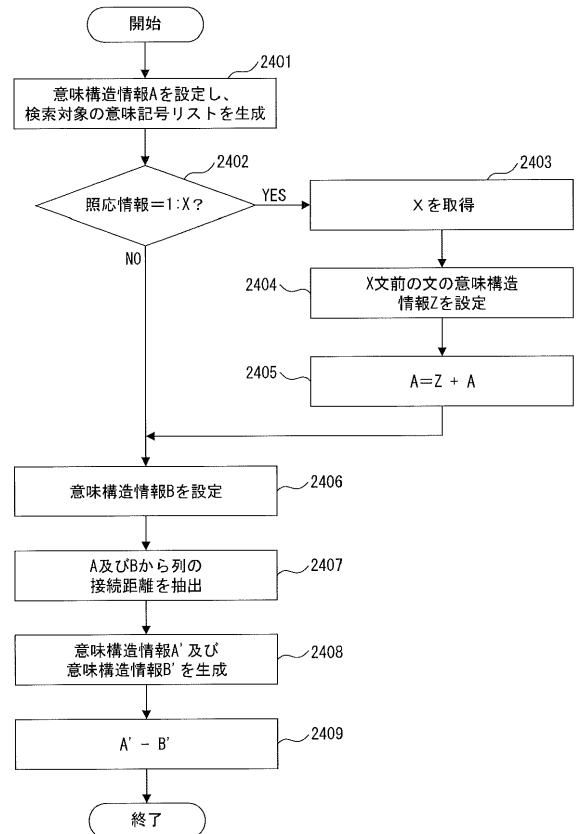
【図20】

検索処理の具体例を示すフローチャート



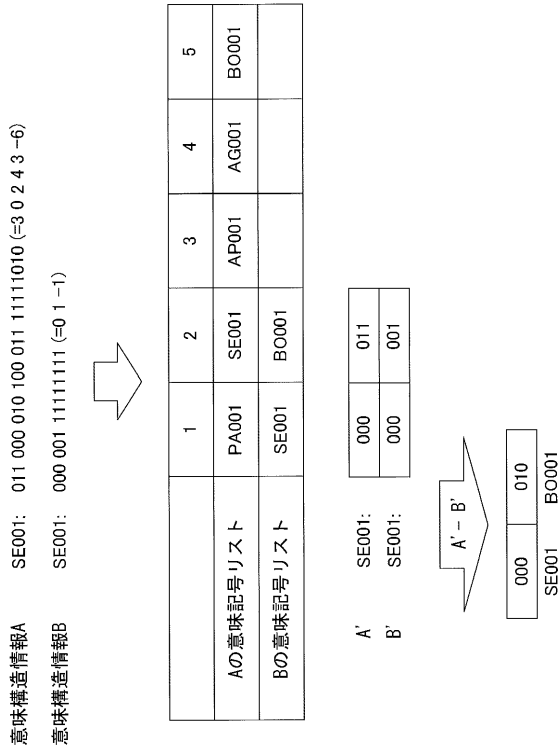
【図24】

意味構造情報マッチング処理のフローチャート



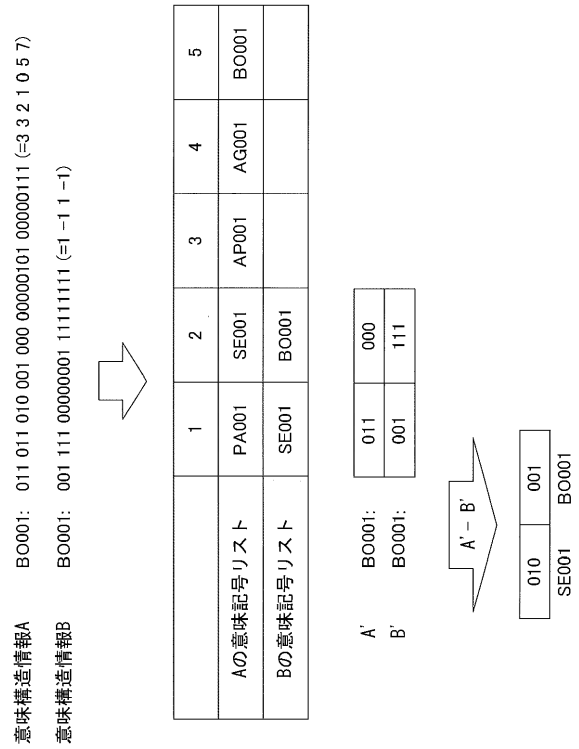
【 図 2 5 】

第1文の「サーバー」に対する意味構造
情報マッチング処理を示す図



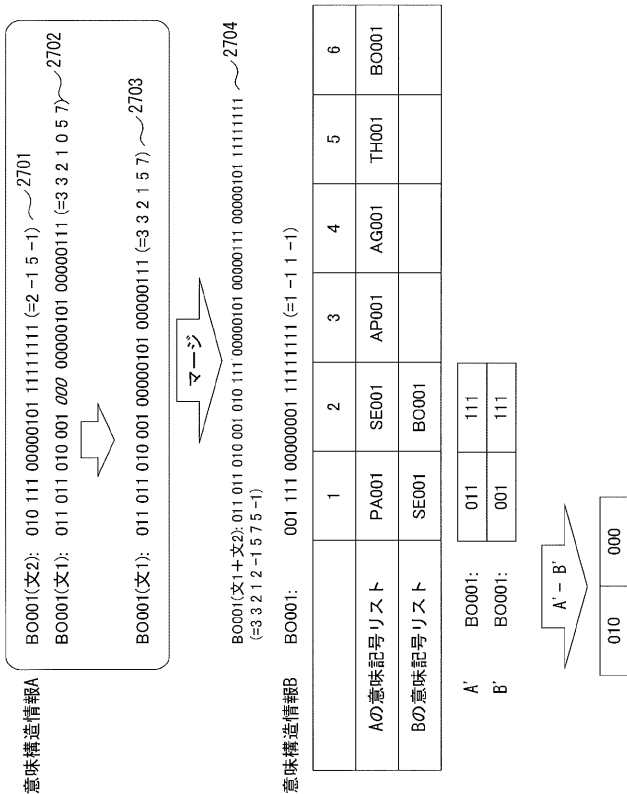
【 図 2 6 】

第1文の「起動し」に対する意味構造
情報マッチング処理を示す図



【 図 2 7 】

第2文の「起動し」に対する意味構造
情報マッチング処理を示す図



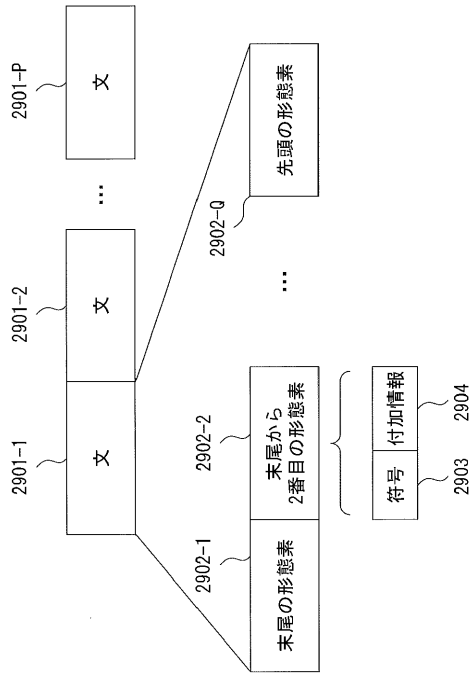
【 図 2 8 】

否定フラグを含む付加情報を示す図

文番号	1										2				
	1	2	3	4	5	6	7	8	9	10	1	2	3	4	5
文面からの位置															
形態素	ハッチ	は	サーバー	に	当て	て	再	起動し	た	。	ずる	と	起動し	な	か
否定フラグ	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0
参照フラグ	1	0	1	0	1	0	1	1	0	0	1	1	0	0	0
無応情報	0	0	0	0	0	0	0	0	0	0	1:1	0	0	0	0

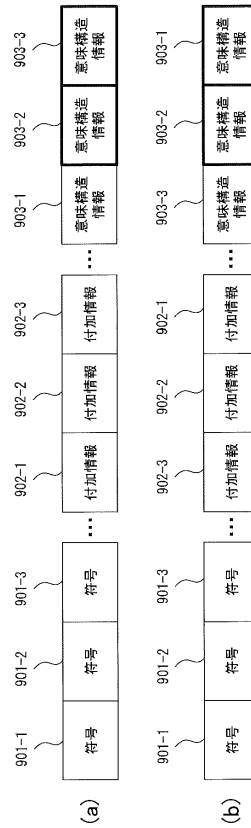
【 図 2 9 】

文の末尾から順に符号を並べた
符号化文書を示す図



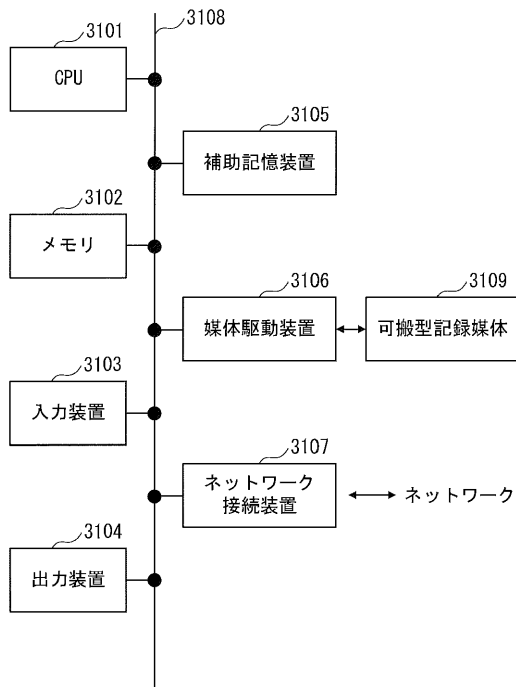
【 図 3 0 】

一連の符号の後に付加情報を
並べた符号化文書を示す図



【 図 3 1 】

情報処理装置の構成図



【手続補正書】

【提出日】平成29年8月30日(2017.8.30)

【手続補正1】

【補正対象書類名】明細書

【補正対象項目名】0033

【補正方法】変更

【補正の内容】

【0033】

例えば、図4の文書421が検索対象文書である場合、符号化装置401によって文書421を符号化することで、符号化文書621が生成される。検索部612は、検索要求に基づいて、記憶部611が記憶する符号化文書621を検索し、出力部613は、検索結果を出力する。

【手続補正2】

【補正対象書類名】明細書

【補正対象項目名】0172

【補正方法】変更

【補正の内容】

【0172】

情報処理装置が図18の検索装置601である場合、ネットワーク接続装置3107は、ユーザ端末から検索要求1811を受信し、検索結果1815をユーザ端末へ送信することができる。この場合、ネットワーク接続装置3107は、出力部613として用いられる。

【手続補正3】

【補正対象書類名】明細書

【補正対象項目名】0176

【補正方法】変更

【補正の内容】

【0176】

図1乃至図31を参照しながら説明した実施形態に関し、さらに以下の付記を開示する。

(付記1)

文書に含まれる複数の文それぞれについて、文に含まれる複数の形態素が属する意味構造を表す意味構造情報を生成し、

前記複数の文それぞれについて、前記複数の形態素それぞれを符号化した複数の符号を生成し、

前記複数の文それぞれについて、各文に含まれる複数の形態素のうち、意味構造において対応する形態素の少なくとも一部が欠落している可能性のある形態素に対応する符号を特定し、

前記特定された符号それぞれに対し、欠落情報をそれぞれ付加する、
処理をコンピュータに実行させる符号化プログラム。

(付記2)

前記複数の形態素が属する前記意味構造は、前記複数の形態素それぞれに対応する複数の意味記号と、前記複数の意味記号の間の接続関係とを用いて表され、前記欠落情報は、前記文書内において、前記欠落情報が付加された符号に対応する形態素と同じ意味記号に対応する形態素が、前記欠落情報が付加された前記符号に対応する形態素よりも前に出現することを示し、前記複数の文のうち前記同じ意味記号に対応する形態素が出現する文を特定する情報を含むことを特徴とする付記1記載の符号化プログラム。

(付記3)

前記コンピュータは、前記複数の符号各々に対して、前記欠落情報、又は前記文書内において同じ意味記号に対応する形態素が前に出現しないことを示す非出現情報を付加する

ことを特徴とする付記 2 記載の符号化プログラム。

(付記 4)

前記意味構造情報は、前記複数の意味記号各々について、前記複数の意味記号のうち他の意味記号に対する接続距離を表す接続距離情報と、前記他の意味記号に対する接続関係を表す接続関係情報とを含むことを特徴とする付記 2 又は 3 記載の符号化プログラム。

(付記 5)

検索要求に含まれる複数の形態素が属する意味構造を表す検索要求意味構造情報を生成し、

検索対象文書に含まれる複数の文それぞれについて文に含まれる複数の形態素が属する意味構造を表す検索対象意味構造情報と、前記複数の文それぞれについて前記複数の形態素それぞれを符号化することで生成された複数の符号と、前記複数の文それぞれについて、各文に含まれる複数の形態素のうち、意味構造において対応する形態素の少なくとも一部が欠落している可能性のある形態素に対応する符号それぞれに対して付加された欠落情報とを含む、符号化文書から、前記検索要求意味構造情報に基づいて前記検索要求に対応する文を検索し、

検索結果を出力する、

処理をコンピュータに実行させる検索プログラム。

(付記 6)

前記検索要求に含まれる前記複数の形態素が属する前記意味構造は、前記検索要求に含まれる前記複数の形態素それぞれに対応する複数の意味記号と、前記検索要求に含まれる前記複数の意味記号の間の接続関係とを用いて表され、

前記文に含まれる前記複数の形態素が属する前記意味構造は、前記文に含まれる前記複数の形態素それぞれに対応する複数の意味記号と、前記文に含まれる前記複数の意味記号の間の接続関係とを用いて表され、

前記欠落情報は、前記検索対象文書内において、前記欠落情報が付加された符号に対応する形態素と同じ意味記号に対応する形態素が、前記欠落情報が付加された前記符号に対応する形態素よりも前に出現することを示し、前記複数の文のうち前記同じ意味記号に対応する形態素が出現する文を特定する情報を含み、

前記コンピュータは、前記符号化文書から、前記検索要求に含まれる前記複数の形態素のうち 1 つの形態素を検索し、検索した形態素が、前記欠落情報が付加された前記符号に対応する場合、前記欠落情報を参照して、前記同じ意味記号に対応する形態素が出現する前記文を特定し、前記検索対象意味構造情報において、特定した前記文に出現する前記同じ意味記号に対応する形態素が属する第 1 意味構造と、前記検索対象意味構造情報において、前記欠落情報が付加された前記符号に対応する形態素が属する第 2 意味構造と、前記検索要求意味構造情報とを用いて、前記検索要求に対応する前記文を検索することを特徴とする付記 5 記載の検索プログラム。

(付記 7)

前記複数の符号各々に対して、前記欠落情報、又は前記検索対象文書内において同じ意味記号に対応する形態素が前に出現しないことを示す非出現情報が付加されており、

前記コンピュータは、前記検索した形態素が、前記非出現情報が付加された符号に対応する場合、前記検索対象意味構造情報において、前記非出現情報が付加された前記符号に対応する形態素が属する第 3 意味構造と、前記検索要求意味構造情報とを用いて、前記検索要求に対応する前記文を検索することを特徴とする付記 6 記載の検索プログラム。

(付記 8)

前記検索要求意味構造情報及び前記検索対象意味構造情報は、前記複数の意味記号各々について、前記複数の意味記号のうち他の意味記号に対する接続距離を表す接続距離情報と、前記他の意味記号に対する接続関係を表す接続関係情報とを含み、

前記コンピュータは、前記検索要求意味構造情報に含まれる接続距離情報及び接続関係情報と、前記検索対象意味構造情報に含まれる接続距離情報及び接続関係情報とを比較することで、前記検索要求に対応する前記文を検索することを特徴とする付記 6 又は 7 記載

の検索プログラム。

(付記 9)

文書を記憶する記憶部と、

前記文書に含まれる複数の文それぞれについて、文に含まれる複数の形態素が属する意味構造を表す意味構造情報を生成し、前記複数の文それぞれについて、前記複数の形態素それぞれを符号化した複数の符号を生成し、前記複数の文それぞれについて、各文に含まれる複数の形態素のうち、意味構造において対応する形態素の少なくとも一部が欠落している可能性のある形態素に対応する符号を特定し、前記特定された符号それぞれに対し、欠落情報をそれぞれ付加する符号化部と、
を備えることを特徴とする符号化装置。

(付記 10)

前記複数の形態素が属する前記意味構造は、前記複数の形態素それぞれに対応する複数の意味記号と、前記複数の意味記号の間の接続関係とを用いて表され、前記欠落情報は、前記文書内において、前記欠落情報が付加された符号に対応する形態素と同じ意味記号に対応する形態素が、前記欠落情報が付加された前記符号に対応する形態素よりも前に出現することを示し、前記複数の文のうち前記同じ意味記号に対応する形態素が出現する文を特定する情報を含むことを特徴とする付記 9 記載の符号化装置。

(付記 11)

前記符号化部は、前記複数の符号各々に対して、前記欠落情報、又は前記文書内において同じ意味記号に対応する形態素が前に出現しないことを示す非出現情報を付加することを特徴とする付記 10 記載の符号化装置。

(付記 12)

コンピュータが、

文書に含まれる複数の文それぞれについて、文に含まれる複数の形態素が属する意味構造を表す意味構造情報を生成し、

前記複数の文それぞれについて、前記複数の形態素それぞれを符号化した複数の符号を生成し、

前記複数の文それぞれについて、各文に含まれる複数の形態素のうち、意味構造において対応する形態素の少なくとも一部が欠落している可能性のある形態素に対応する符号を特定し、

前記特定された符号それぞれに対し、欠落情報をそれぞれ付加する、
ことを特徴とする符号化方法。

(付記 13)

前記複数の形態素が属する前記意味構造は、前記複数の形態素それぞれに対応する複数の意味記号と、前記複数の意味記号の間の接続関係とを用いて表され、前記欠落情報は、前記文書内において、前記欠落情報が付加された符号に対応する形態素と同じ意味記号に対応する形態素が、前記欠落情報が付加された前記符号に対応する形態素よりも前に出現することを示し、前記複数の文のうち前記同じ意味記号に対応する形態素が出現する文を特定する情報を含むことを特徴とする付記 12 記載の符号化方法。

(付記 14)

前記コンピュータは、前記複数の符号各々に対して、前記欠落情報、又は前記文書内において同じ意味記号に対応する形態素が前に出現しないことを示す非出現情報を付加することを特徴とする付記 13 記載の符号化方法。

フロントページの続き

(72)発明者 出内 将夫

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 中村 文昭

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

Fターム(参考) 5B091 AA15 AB19 CA05 CA12